University of Twente
Graduate University

Thesis submitted for the degree

Master Business and Information Technology

Heineken-Global Analytics

An Enterprise Architecture based Big Data Analytics Capability Deployment Reference Architecture to improve Business Value

by

Jesús Contecha

Supervisor: **M. João and F.Erwin**
Co-Supervisor: **T. Deeepti**

Jul 2023

Abstract

The increasing architectural complexity poses a significant challenge for digital leaders as organizations are at risk of being overwhelmed by data floods, complexity, and rising costs. As companies transition to become AI-driven entities, the architectural complexity and IT fragmentation increase. Furthermore, while structured data storage is projected to increase, a significant portion of data stored remains unused. Moreover, only half of Data & Analytics teams effectively contribute value to their organizations, suggesting that a fraction of available structured data is used to create incremental business value, indicating data and resources underutilization. To overcome these challenges, Enterprise Architecture (EA) artifacts are proposed to serve as a "blue print roadmap representation" to guide the deployment of Big Data analytics (BDA) initiatives.

However, further research is needed to understand the role of EA in adopting big data analytics. This study aims to integrate the Big Data Analytics capabilities theory, EA frameworks, and empirical organizational resources by exploring how Enterprise Architecture can improve the deployment of big data analytics initiatives. EA plays is theorized to play a critical role in representing digital transformation's building blocks and processes to align Information Systems with business strategy. This multifaceted approach implies a reference architecture that captures business, applications, and information and technology architectures changes.

The present thesis emphasizes the importance of EA practice in planning, guiding, and assessing the transformations required to leverage current and future BD capabilities and resources to develop DB Analytics capability. By effectively leveraging EA artifacts, organizations can orchestrate big data resources (People, Process, Tech) and BDA capabilities (Business Infrastructure alignment, seizing/reconfiguration, and Infrastructure Flexibility) to optimize deployment processes. For instance, a firm/function where "data validates business experience," versus "business experience complements data" would benefit from a specific deployment architecture tailored to its capabilities and resources context. To this end, three architectural levels are developed to represent essential deployment processes and core building blocks that serve as a blueprint for big data deployment initiatives. Finally, concrete architecture levels are instantiated and evaluated within Heineken's Advanced Analytical product global deployment context. The results indicate that the architecture effectively encompasses core components, enables cross-functional teams, and reduces deployment time and improve resource optimization.

Contents

List of Figures

List of Tables

Acknowledgment

This submission of my master's thesis signifies the culmination of my academic journey at the University of Twente, representing a significant milestone in my life. Over the span of more than two years, I made the decision to leave my home country and reside in the Netherlands, immersing myself in a multitude of challenges that taught me invaluable lessons in patience, grit, and determination. The present research work originated from my internship with the Global Analytics team in August 2022. During my internship, I actively contributed to the implementation of the Global Analytics assessment across over ten countries on various continents and played a key role in driving the analytics transformation under the guidance of my manager, Sandra Oudshoff, in our goal to make Heineken the "best-connected brewer".

I would like to express my deep gratitude to my thesis supervisor at Heineken, Deepti TonpaeUmesh, for their constant support, confidence, and guidance throughout this complex and visionary project. Your expertise in business and technology, along with your invaluable words of support and advice, have contributed immensely to my learning experience and provided a professional perspective on this journey. I am also grateful to Sandra Oudshoff for her support, trust, and valuable insights in the field of Engagement and Analytics. Furthermore, I want to share my gratitude to Roelef Schoeman, Kliment Markovski, Lynise Esterhuizen, Olle Dahlen, and the entire Global Analytics team, who have shared their time and expertise from diverse domains and guided me throughout this research. Your collective talent is truly remarkable, and I sincerely wish you all the best in your respective endeavors. I would also like to express my sincere gratitude to my supervisors, João Moreira and Erwin Folmer, which have been invaluable guidance and contributions for this research realization.

To Dea Savitri and Danniar Reza, I sincerely appreciate your support and comforting presence during the most challenging times. Your kindness and companionship have made you incredibly special to me, and I genuinely wish you both success and good fortune in all your personal and professional goals and dreams.

Last but certainly not least, I am eternally grateful to my family, whose teachings, hope, support, and love have been indispensable throughout this process. To my sister Daniela, my father Alvaro, my mother Magali, my grandma Tata, Moñita, Ponchito, and my uncles and aunts from my extended family in Costa Rica and family in Colombia, you have all played an integral role in this journey. Your guidance, examples, unwavering support, and unconditional love have illuminated the beauty of life for me. Despite the challenges, setbacks, and learnings, you have always provided encouraging words and encouraged a drive to continue on my growth path. You are an inseparable part of me, and I am immensely fortunate to know that you are always there for me when I need it the most. I would not be where I am today without you. Thank you for everything you have done for me; this research is dedicated to each of you.

Jesús Antonio Contecha Montes

Amsterdam, June 2023

Chapter 1

Introduction

In the contemporary era, global organizations and businesses operations have been affected by various factors such as globalization, technological advancements, and economic pressures. These factors have undeniably impacted the companies value chains with several unforeseen disrupting events. These include situations such as the blockage of the Suez Canal and subsequent increase in container shipment costs, Chinese port disturbances due to Covid-19, historical inflation and energy price fluctuations, risks to world food supply caused by the Ukraine war, or future technology events from regulations like the future Artificial Intelligence Act, or the launch of artificial intelligence tools such ChatGPT4. As these events illustrate, that "change is the only constant". For this reason, the ability to swiftly adapt and orchestrate organizational resources and capabilities to face those circumstances is becoming more crucial for achieving and maintaining a competitive advantage.

The rapid advancement of technology has resulted in considerable operations improvements to address some the challenges present in various industries. One of the approaches are related to the increased world connectivity and significant upsurge in the use and analysis of both large structured and unstructured big data assets. Multiple practical applications in different sectors are found is recent academia research, such as: water management [34], finance [64], agriculture [32], healthcare [61], manufacturing [14], and transportation [59]. Despite these advancements, a considerable proportion of big data investments fail to produce the desired outcomes due to companies' lack of preparedness or the inability to act on insights obtained from data [54]. According to a BCG (2023) [7], the volume of data generated has doubled from 2018 to 2021, reaching approximately 84 zettabytes. This growth is projected to continue, with a compound annual growth rate of 21% from 2021 to 2024, on which 40% of these data is being stored in the cloud, and over 95% of this data is considered unstructured data (video, voice, and text). However, currently small in proportion, the complementary 5% represented structured data is being stored at a faster paced, a trend attributed to the expansion of Analytics business intelligence (BI) use cases. However, over 50% of this structure data is classified as **dark data**, meaning that it is not used to generate any type of Analytics insights or value.

Furthermore, a 2023 report from Gartner's revealed that "less than 50% of data and analytics teams effectively contribute value to their organizations" and implying that only a portion of the data utilized in organizations genuinely is used to create value. This situation is attributed to several factors, such as a shortage of skilled personnel and talent, inadequate use of resources, cultural challenges, or poor data literacy. Given this scenario, a critical question is how to orchestrate effectively the existing organizational Big Data analytics resources and capabilities to create value from the enormous business data.

1.1 Research context and Motivation

Big Data is defined by Gartner as "technologies that are targeting processes with high: volume, velocity, and variety data (sets/assets) to extract intended data value and ensure high veracity of original data and obtained information that demand cost-effective, innovative forms of data and information processing (analytics) for enhanced insight, decision making, and processes control; all of those demand (should be supported by) new data models (supporting all data states and stages during the whole data lifecycle) and new infrastructure services and tools that allow obtaining (and processing) data from a variety of sources [...]". Another complementary definition [17] assess that Big Data comprises five encompassing characteristics commonly referred to as the 5V properties: Volume, Velocity, and Variety, which are inherent to Big Data, and Value and Veracity are also part of Big Data are acquired through data classification and processing within a specific process or model. While databases are the essential resource for large-scale data processing and analytics, Big Data encompasses a more complex range of processes, including the ability to storage, processing, visualization, and delivery of results to target applications. As such, Big Data is defined as the "fuel" resource for all data-related processes, and as the main outcome.

A resource is as defined as (Tangible and intangible) assets that are owned and controlled by a firm [25], while a capability is the ability to make use of the resources in the most strategically way [46]. In the case of Big Data Analytics capability (BDAC), this definition includes the ability to "acquire, assemble, integrate, and deploy" big data specific resources [25]. As such, the ability to utilize BD resources in the implementation of Big Data Analytics (BDA) "entails a multifaceted procedure that relies on a company's capacity to leverage different resources and capabilities while orchestrating them synergistically "[5]. Numerous studies have confirmed that the utilization of Big Data Resources (BDR) is instrumental to enhance the efficiency and effectiveness of **Big Data strategic Capabilities**, which in turn facilitates the development of competitive advantages [57].

The process of developing a (BDA) is intricate and requires the utilization of various tangible and intangible resources as well as human skills at the organizational level [25]. However, these resources and their combination have been identified as a primary constraint for deploying Big Data Analitycs initiatives. According to the McKinsey Global Institute [15], reported that firms are presently unable to fully leverage the potential value of big data due to three main factors: (1) Information technology (IT) infrastructure, (2) Organizational strategies, leadership, and talent, and (3) Organizational structure and processes. Consequently, limitations and current challenges within these types of resources and capabilities impede the successful implementation of big data analytic initiatives.

On top of the limitations to deploy big data analytics initiatives, the enormous data volume and data process optimization in the cloud, combined with limited technical knowledge and complex analytical processes are creating additional organizational challenges. According to a recent report from BCG [7], 50% of data leaders reported that architectural complexity is a significant pain point. As a consequence, many companies are at risk of being overwhelmed by a flood of data, burdened by complexity and expenses. As organizations grow and companies evolve from data-driven to AI-driven entities, the architectural complexity and fragmentation naturally escalates. Naturally, EA practice, by design, should be considered to play a critical role which requires alterations in business architecture/ processes, and information and technology architecture to address the digital transformation imperatives [49]. From the architecture point of view, Enterprise Architecture (EA) would contribute to support the development of the future capabilities, products and services with controlled effort and adequate timing [11].

In that sense, limited research has been develop in the enterprise architecture's role for adopting big data analytics [22]. EA research is still sparse in the Big Data Analytics Capabilities context [49]. Despite significant endeavors to define the building blocks of a firm's BDA capability, little is known about the processes and structures necessary to orchestrate BD resources into a firm-wide Analytics capability. In other words, the literature provided extensive coverage of the process of selecting resources for BDA yet, relatively limited insights about the activities that need to be put into place to develop the capability [46].

Future research can be conducted to comprehensively assess the contributions of Enterprise Architecture (EA) to generate value in the context of BDA. This encompasses not only EA artifacts but also the concerns of associated stakeholders. Given that diverse stakeholders possess varying perspectives, different resources, and capabilities, and produce an array of artifacts, a broad spectrum of EA roles can be utilized to support them [5]. Therefore, future research is required to bridge this gap between the resources theory and an empirical Big Data Analytics processes approach of orchestrating Big data Resources/capabilities through the use of EA artifacts to leverage business value and increase competitive advantage. In other words, despite the various obstacles organizations encounter, this research aims to explore how Enterprise Architecture can assist BDA professionals through a common blueprint and road map reference architecture to support the deploy Big data analytics initiatives (i.e., Advanced Analytics products) based local BDA resources and capabilities. This research aims to provide practical and up-to-date artifacts and methodologies that can benefit both the academic and corporate realms.

1.2 Research Questions

The high level goal of this research is to develop an BDAC integrated reference architecture to support the deployment advanced analytical products.

How can an BDAC reference architecture improve the deployment of advanced analytical products?

1.2.1 Research sub-questions

As previously mentioned, the high level goal of this research is to develop an Big Data Analytics Capability Enterprise Architecture reference Architecture to support and optimally orchestrate the different BDA capabilities and resources to deploy the BD analytics product initiatives. In order to provide a more comprehensive answer, the main research question will be subdivided into the following sub-questions:

1. **How can Enterprise Architecture be used to identify big data analytics capabilities (BDAC) to improve business value?** To answer this question, the state-of-the-art alignment between Big Data Analytics Capabilities and Enterprise Architecture (frameworks, architecture patterns and methodologies) will be presented. For this chapter, a set constructs of these fields will be derived from the literature collected through the SLR, as discussed in chapter 3. Subsequently, the academic literature presented in this chapter will summarize the results, research and empirical gaps through the following research questions:
   (a) What is the state of the art of big data analytics capabilities?
   (b) What business value is generated using big data analytics capabilities?
   (c) Which EA tools or techniques are used to leverage business value?
   (d) What are the most pressing challenges in the deployment of BDA capability projects?
2. **How can Big Data Analytics capabilities and Enterprise Architecture integrate into a BD capability deployment reference architecture?** The aim of this chapter is to create a theoretic BDA Capability Deployment architecture that integrates state-of-the-art BDA capabilities dimensions, and EA methods and patterns presented in the previous chapter 1. The architecture includes multiple methods to integrate the BDA capabilities levels and BD resources maturity to serve as a big data analytics deployment building block processes road-map. The architecture objective is to define a method to integrate the BDA capabilities levels and BD resources maturity into a propose a BDAC deployment architecture to server as a big data analytics deployment building block processes road-map. The following are some additional sub questions that will be answer in the chapter:
   (a) What are the most relevant BDA capabilities required to create a BDAC deployment architecture?

(b) How can BDA Capabilities and Advanced Analytics products deployment process be integrated into a reference architecture?

(c) What are the different BDA deployment architecture layers and transformation processes to deploy Analytical products?

3. **How to operationalize the BDA Capability Deployment architecture to support the deployment of advanced Analytical products in Global Analytics, Heineken?** Chapter 4 will elaborate on the DBA capability deployment method design for advanced analytics based on in a real-case scenario context using the design obtained in first subsection of chapter 3. This procedure will be based on a real-world analytics product deployment that can serve as a case study for implementing the BDA Capability reference architecture into a concrete architecture.

4. **What are the effects of the implementation of the BDA Capability Deployment architecture in Heineken´s local Operational Company context?** On this chapter, the goal is to validate the extent to which the proposed method, along with its functional prototype, successfully fulfills the predefined objectives and satisfies the identified requirements. To accomplish this, a semi-structure session will be organized, involving cross functional stakeholders involve in the anayltical product deployment. During this session, the proposed architecture and its application prototype will be presented and demonstrated.

   (a) Does the BDAC Deployment architecture levels represent the most important building blocks, layers and processes of the deployed Advanced analytics product?

   (b) How are the BDAC (Business Infrastructure alignment Capability,Seizing & Configuration Capability, and Information Transformation Capability) present in the deployment architecture levels?

   (c) What would be the effects if the DBAC Deployment architecture had been instantiated back in the past for the analytical product deployment?

   (d) To what extent is the deployment architecture expected to contribute in the cross- domains teams collaboration efforts to deploy analytical product (i.e., Final user, Product Owner, Translator, Local Data team, Data Scientist, and engineers)?

1.3 Research Scope

The fundamental objective of this document is to propose a Big Data Analytics capability reference architecture in the context of Heineken Global Analytics initiatives. This artifact will be based in the current state-of-the-art application of Enterprise Architecture to facilitate the development and orchestration of big data analytics capabilities. To achieve this purpose, the research will gather and examine the underlying methods, drivers, challenges, and activities required for creating and adopting the big data analytics capability. Additionally, the present study aims to extract the architectural components and patterns utilized, including the enabling resources, capabilities, technologies, and iteration processes. The ultimate aim is to identify and consolidate the essential architectural components and activities derived from previous studies and establish a BDA Capability Deployment reference architecture that serves as a target architecture in the different Heineken local context.

The intended artifact will also require to be adaptable and reusable by involving multiple organization stakeholders, functions/ departments from different backgrounds (i.e., Data Science, Data Engineering, Sales, Supply, Commerce, and Revenue management) through a Reference Architecture that measures and assesses the specific local context. The Big Data Analytics instance will comprise a variety of organizational resources types (people, processes, and technologies) from different domains, and their highly complex business processes will be considered and further integrated. For this reason, the current research simplifies some of the relevant activities implemented in the EA prototype to address this complexity while still providing significant stakeholders value.

1.4 Definition and methodologies

To enhance the clarity of the research topic, it is necessary to provide a unified definition by examining several studies, concepts and frameworks, as explained in the next subsection. As this research involves designing and delivering a reference architecture, the knowledge domain and a frameworks that serves as a development guide and the objective goal will be examined. The research methodology that governs the direction and structure of this research will also be discussed in the following subsections.

1.4.1 Design Science Research Methodology

In order to address the research questions outlined in the previous section, a design science project [52] will conducted. Design science is "the design and investigation of artifacts in context" that aims to create and validate prescriptive knowledge in information science. This approach emphasizes the development and evaluation of a designed artifact, with the explicit goal of improving its functional performance in a specific context. The design process involves solving design problems and answering knowledge questions through an iterative sequence of activities, leading to the creation of an innovative product. As part of the design science project, the stakeholder requirements and those who may affect or be affected by the project will be considered.

In order to ensure the artifact's usefulness to the identified problem and context, it must be evaluated. As such, the artifact must either solve a previously unsolved problem or offer a more efficient solution to make a new research contribution. The design process is divided into three parts: problem investigation, treatment design, and treatment validation. These three parts are commonly known as the design cycle, and researchers often repeat these steps multiple times during a design science research project. Likewise the design cycle is only a part of a larger cycle, which involves taking the validated treatment and its application within a real-world context. This larger cycle, known as the engineering cycle, entails using and evaluating the treatment in a real-world context. Typically, this build-and-evaluate cycle loop is performed several times before the final design product is developed. (Please refer Figure C Reference Architecture Archimate meta-models layers views in the Global Analytics context.

1.4.2 Research Structure-Design Science Research Methodology (DSRM)

This study follows the Design Science Research Methodology (DSRM) process iteration in the field of Information systems (IS) [51]. To achieve this, the research follows the corresponding six iterative steps as the central methodology presented in Figure 1.1 .

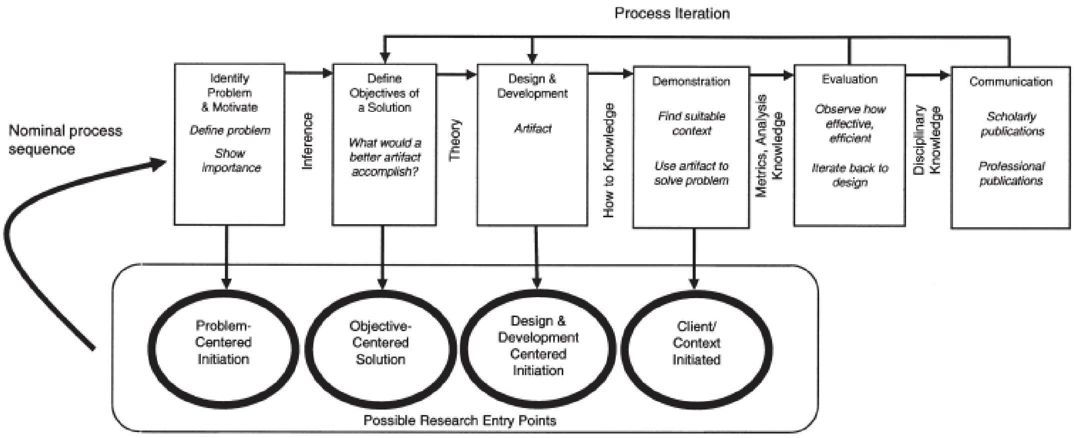


Figure 1.1: Design science research methodology [51]

1. **Problem identification & motivation:** The primary problem and motivations are identified at the first step as the research starting point. To support this process, an SLR is conducted to aggregate and filter out existing research and content related to the main research question and objectives to support the development of an evidence-based approach [35]. During the first phase the main focus is to establish the objective of the artifact by considering SLR theory stakeholder requirements, aims, processes, and consequences.

2. **Define objectives of a solution:** This step establishes the primary research objectives to create an artifact that addresses the problem identified in the initial phase. As such, it involves outlining the stakeholder's requirements, specifications, and desired outcomes that the artifact should fulfill from the previous section SLR.

3. **Design & development:** In the design and development step, the focus is to create a theoretical BDA capability reference Architecture. This process involves the use of insights gained from the SLR to identify the requirements and design patterns that should be considered in the artifact design. For this step, multiple requirements are gathered for the BDAC reference architecture and MLOps framework architecture patterns, and EA frameworks. First, the stakeholders identified, the collect functional and non-functional requirements, models, and challenges are

collected from the systematic literature review to be used within the artifact design. These specifications are use as the baseline to the develop the artifact design, on top of the state-of-the-art Machine Learning & Operations (MLOps) architecture [38] (Please refer to Appendix: A.2. The final BDA Capability reference architecture in the context of advanced analytical product global deployment views and local BDA maturity resources and capabilities levels.

4. **Demonstration:** The prototype architecture will be set in place in the context of a Big Data Analytics Deployment initiative for the Global Analytical team, the artifact's purpose is derived from Heineken's Evergreen strategy and the goal of becoming "the best-connected brewery by digitally transforming our business end-to-end" through Data-driven insights and foresight. Analytical solutions are then created and implemented to assist in fact-based decision-making across the Heineken value chain and assist this data-driven objective. To understand the deployment and development context and difficulties of Big Data initiatives, various Heineken stakeholders from multiple functions and cross-departmental teams (i.e., Data Engineers, Data Scientists, Translators, product owners, and solutions architects) were interviewed to determine deployment and development contexts and challenges (Please refer to section 5). In addition, a thorough literature review was conducted in section 3 to determine potential research gaps from an academic perspective. The current artifact represents partially some of the relevant activities implemented in the EA prototype to address this complexity while still providing significant stakeholders value.

5. **Evaluation:** On this step, the specific reference architecture prototype will be instantiate and evaluated into a concrete (GA) Advance Analytics product deployment through the Expert Opinions in a context through a set of interviews. The method and architecture artifact prototype developed in the previous chapter will be validated in the context of a case study context by multiple domain experts/ stakeholders who will be asked to provide feedback on the real problem in the context. As such, the validation model for this research will consist of domain experts' semi-structured interviews over reliable effects predictions of the BDAC Reference Architecture Archimate meta-models layers views in the Global Analytics context (Chapter 5).

6. **Communication:** On the final step, the conclusions derived from the design, demonstration, and evaluation of the proposed reference architecture in an real-world scenario are communicated.

The current research will follow the proposed DSRM methodology with the corresponding sequential phases in order to validate the potential advantages of the proposed reference architecture in Heineken's Global Analytics context. The first one, 2 will introduce the foundational constructs and frameworks that supports the research´s conceptual models and artifacts methodology. Chapter 3, will address the first research objective and conduct a systematic literature review to identify preliminary BDA capabilities & resources, EA frameworks, methods, and reference architectures composed of different components from relevant research publications. The second phase, chapter 4-section: 4 entails identifying business stakeholders' requirements for further development, implementation, and validation of the MLOps-BDAC integrated Reference architecture artifact. Subsequently, in chapter: 4, a further artifact requirements are defined based on the literature analysis, domain experts interviews, and knowledge gaps. The fourth step, in chapter 5, demonstrates through a specific reference architecture application into a concrete architecture instantiate into a working prototype in the context of Big Data Analytics initiative. In step phase, chapter 6, the evaluation of the artifact's perceived effects will be assessed through expert's semi-structure interview context. Finally, the sixth step, chapter 7, will present the conclusion, limitations, and recommendations for future work.

Chapter	**DSRM phase**	**Research Method**	**Subquestion**
1-Introduction	Problem Identification & Motivate	-	-
2-Theoretical Background	-	Systematic Literature Review (SLR) [35]	SQ:1
3-Systematic Literature Review	Problem Identification & Motivate	Systematic Literature Review (SLR) [35]	SQ:1
4-Artifact BDAC Reference Architecture section	Define Objectives of a solution - Design & Development	Systematic Literature Review (SLR)- TOGAF (ADM) [1]	SQ:2
5- Demonstration BDAC Deployment Reference Architecture section	Demonstration	Multi-Criteria and Model-Based Analysis method [6]- GA Global initiative	SQ:3
6-Evaluation BDAC Deployment Reference Architecture	Evaluation	Expert Opinions, Expert interviews [51]	SQ:4
7-Conclusions	Communication	-	SQ:-1,2,3,4

Table 1.1: Thesis Structure

Chapter 2

Theoretical Frameworks Background

In this chapter, the theoretical EA definition, frameworks and methods are introduced in order to achieve a clearer topic comprehension and structure as results of a common unified definition obtained from several knowledge question in the research SLR (Chapter 3). This definitions and frameworks selection will be further explained upon the next subsection. Additionally, since this research aims to design and deliver a reference architecture, the knowledge domain and foundational guiding frameworks will be further operationalize in chapter 4 and instantiate in chapter 5.

2.0.1 Enterprise Architecture

Digital progressions have led to increased customer expectations, demanding fast and seamless services and products enriched with digital information. Despite this, organizations struggle to adjust their operations and outcomes to these expectations, for instance, leading to the vertical specialization of products/services design without further considering the impact on the organization as a whole [53]. At a high level, Enterprise Architecture (EA) is a comprehensive approach that incorporates people, processes, and technology resources with methods, rules, models and tools to guide organizations future improvements according to its digital operational vision and accomplish digital strategic goals. Expanding to this definition, [42] defines EA as a coherent whole of principles, methods, and models used to design and realize an enterprise's organizational structure, business process, information systems, and infrastructure.

From a modeling perspective, Enterprise Architecture (EA) is a discipline that enhances strategic alignment by planning, designing, and executing organizational changes [44]. EA establishes the organizing logic for business processes and IT infrastructure, reflecting the integration and standardization requirements of the company's operating model [21]. Moreover, EA offers a long-term view of a company's procedures, systems, and technologies, allowing projects to create capabilities instead of only satisfying immediate needs. Essentially, EA can function as a blueprint to specify a target implementation or provide guidelines for implementation on a higher level [22]. Additionally, an EA model enables the representation of a perspective that clarifies various organizational levels and assists in business-IT alignment. Through this process, Enterprise Architecture drives in, more or less degree, some of the following organizational core organizational benefits [58]:

1. **Organisational Alignment:** the extent to which an organisation‘s sub-units share a common understanding of its strategic goals, and contribute towards achieving these goals as a hole.

2. **Information Availability:** the extent of useful, high-quality information accessible to organisational decision makers.

3. **Resource Portfolio Optimisation:** the extent to which an organisation leverages current resources, invests in resources that target performance gaps, and reduce unnecessary investments in duplicated resources. For instance, optimisation could involve the removal of duplicated or non-value-adding technology or human resources, and/or replacing them with resources that are more efficient in assisting with the achievement of organisational goals.

4. **Resource Complementarity:** refers to the degree to which an organization's resources synergistically work together effectively to achieve its strategic goals. The organization's capabilities (human resources, IT, and organisational processes) are developed over time through the exchange, retention, and creation of information, and rely on the skills, knowledge, and organization´s processes.

2.0.2 The Open Goup Architecture Framework- TOGAF

An EA framework could be understood as a "[...] tool which can be used for developing a broad range of different architectures. It should describe a method for designing an information system in terms of a set of building blocks, and for showing how the building blocks fit together" [1]. Organizations around the globe have adopted some of the current well-established frameworks, such as DoDAF, TOGAF, Zachman, the Department of Defense Architecture Framework, or MODAF to implement best industry practices in Enterprise Architecture. These frameworks allow companies to gain implementation momentum in simplifying and accelerating architecture development through a comprehensive artifact that effectively addresses key stakeholders' interests [42]. Furthermore, EA frameworks should also focus on managing the gradual EA changes over time. This continual iteration approach will benefit the gradual transition from the baseline to the target EA by scheduling changes according to each of the versioned EA requirements [11].

As it will be discussess in Systematic Literature Review (SRL) section: 2, subsection: 3.4.3 professionals have identified TOGAF, with its ADM methodology, as the most commonly used by professionals [6]. Particularly, this framework introduces basic notions of capabilities and capability-based Planning which allows a business-driven approach that combines the efforts of multiple business areas to achieve a desired capability development [30] , [6]. As a widely used architecture framework, TOGAF offers a standard model for representing current and future business needs. Its most current version TOGAF 10, and its previous 9.2 version include the TOGAF EA capability and Governance Framework, ADM Guidelines & Techniques, Architecture Content Framework, Enterprise Continuum and Tools, and Architecture Capability Framework, all of which are designed to showcase the enterprise architecture capability (Please refer to Figure 2.1). In the context of the present study, version 9.2 will be used instead of version 10, given the recent launch of version 10. However, it is worth to mention that this version includes valuable concepts such as agility, micro-services architecture, and digital strategy. Additionally, the guideline includes the creation of five architecture domains, which consist of the Vision, Business, IS Data, IS Application, and Technology Architecture.

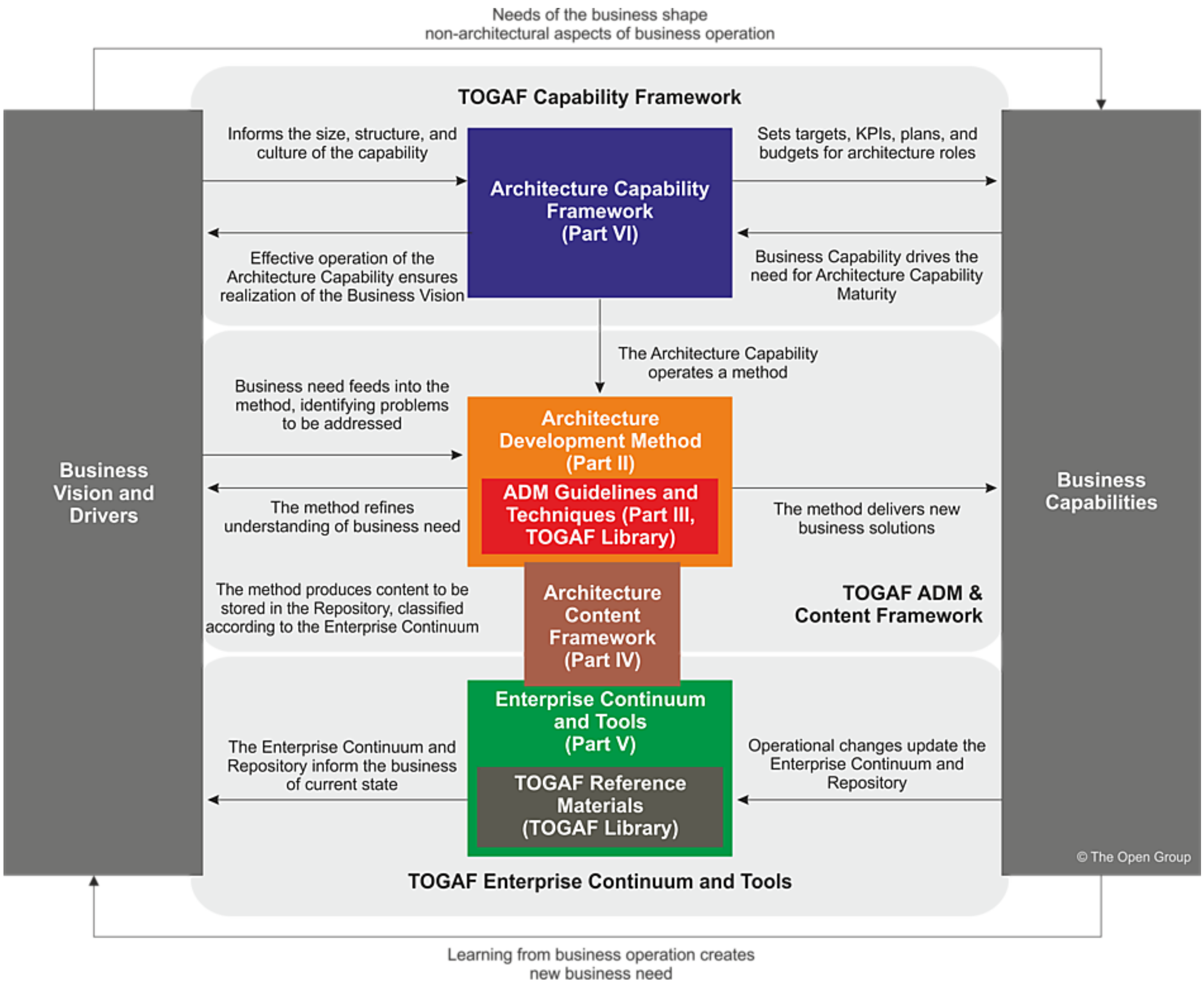


Figure 2.1: Structure of the TOGAF Standard [1]

TOGAF Architecture Development Method (ADM)

The TOGAF Architecture Development Method (ADM) is a reliable and replicable process for architectures development. It comprises the creation an architecture framework, producing architecture content, transitioning, and governing the implementation of architectures [1]. A core key concept within an ADM cycle is iteration which "describes the integrated process of developing an architecture where the activities described in different

ADM phases interact to produce an integrated architecture". These activities are conducted through an iterative cycle consistent of architecture definitions and implementations which enables organizations to systematically control the Enterprise transformations in response to their business goals and opportunities. (Please refer to figure 2.3). Each of the ADM iteration phases and corresponding ADM cycle phase definitions are further explain next followed by the corresponding ADM phases in table 2.1.

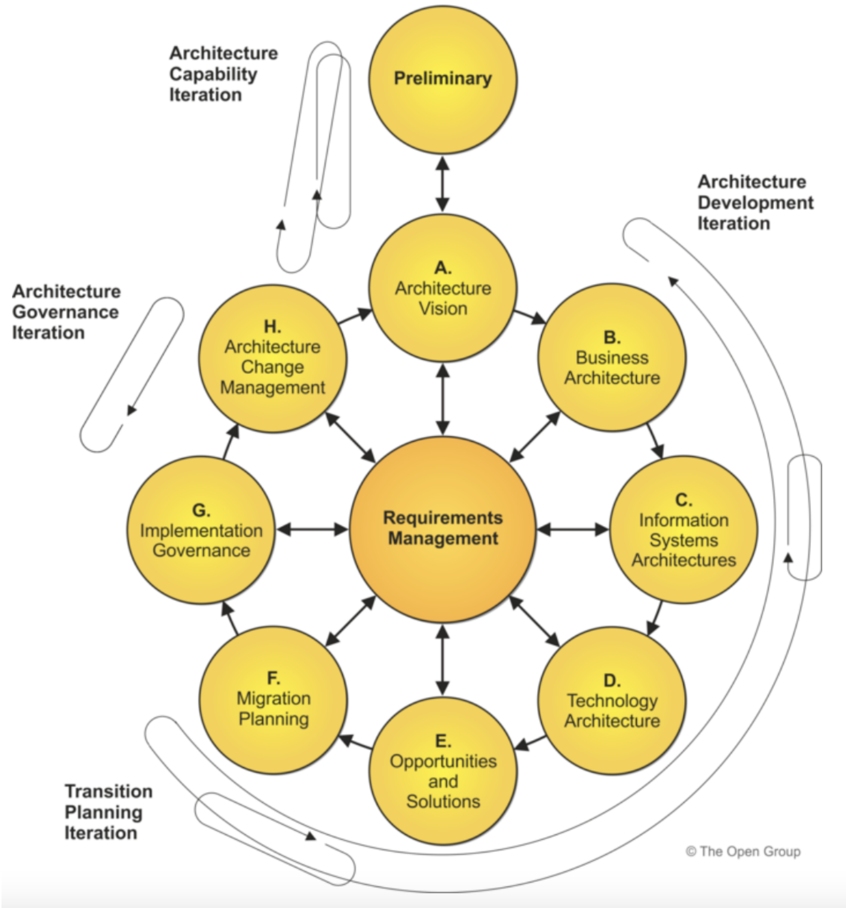


Figure 2.2: TOGAF Iteration within an ADM cycle [1]

1. **Architecture Capability:** iterations support the creation and evolution of the required Architecture Capability (i.e., defining or adjusting architecture approach, principles, scope, vision, and governance).

2. **Architecture Development:** iterations allow the creation of architecture content by cycling through, or integrating, Business, Information Systems, and Technology Architecture phases.

3. **Transition Planning:** iterations support the creation of formal change road-maps for a defined architecture.

4. **Architecture Governance:** refers to the iterations that supports change activity progress towards a defined Target Architecture.

Iteration Cycle	**ADM Phase**	**Description**
1-Architecture Capability	Preliminary	Includes the review the organizational context for conducting Enterprise Architecture,Identify and scope the elements of the artifact capability and identifying the established frameworks, methods, and processes that intersect with the Architecture Capability.
1-Architecture Capability	A-Architecture Vision	Establishes the boundaries, limitations, and anticipations of a TOGAF initiative. It identifies stakeholders and verifies the business context and business value to be delivered based on the proposed Enterprise Architecture.
2-Architecture Development	B-Business Architecture	Outlines and the basic road-map structure of the business process, goals and stakeholders that will support the agreed architectural vision.
2-Architecture Development	C-Information Systems Architecture	Develop the target IS architecture vision with the major information and application systems that support the business architecture and address the stakeholders concerns. It includes data and application architecture.
2-Architecture Development	D-Technology Architecture	Capture the software and hardware Architecture vision, target business, data, and application building blocks systems that support the preceding architecture stages, and represents the basic and target enterprise IT systems.
3-Transition Planning	E-Opportunities and solutions	The projects are grouped into work packages to deliver the target architectures based upon the gap analysis and candidate Architecture road-map components from previous Phases B, C, and D
3-Transition Planning	F-Migration Planning	Outlines the transition steps to transition from the current enterprise to the desired future state as defined in the target architecture by finalizing a detailed Implementation and Migration Plan.
4-Architecture Governance	G-Implementation Governance	Oversee the implementation and conformance with the target architecture.
4-Architecture Governance	H-Architecture Change Management	Establish procedures for managing the new architecture change and verify the architecture responds to the enterprise needs and requirements.

Table 2.1: ADM iteration cycles, phases and descriptions [1]

Archimate Modeling Language

TOGAF standard specification defines ArchiMate as the standard Enterprise Architecture modeling language (Please refer to 2.3). It uses a predefined set of visual symbols to describe and communicate different aspects of enterprise architectures as they evolve through different model layers and stakeholder views. The language includes various entities, relationships, generic meta-models, layers dependencies, and corresponding symbols to create specific architectural process descriptions. The framework to structure allows for the modeling of the enterprise from different viewpoints and distinguishes between Aspects (Active Structure, Behavior, Passive), the Connection between these elements using realization relationships to relate elements across these layers, and Dimensions layers (Strategy & Motivation, Business, Application, and Technology Layers of Enterprise Architectures, Implementation & Migration). (Please refer to Figure 3.6) The following

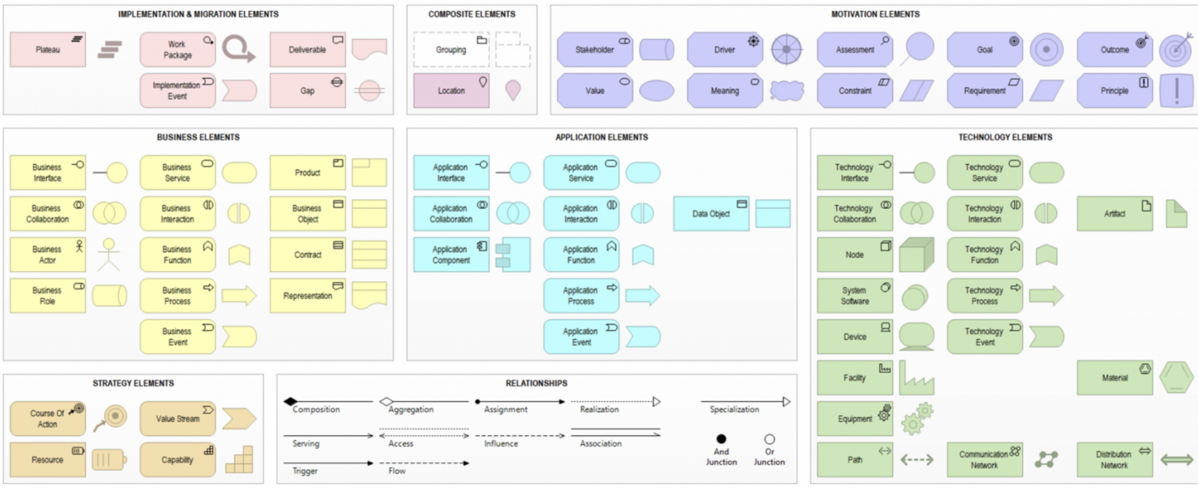


Figure 2.3: ArchiMate 3.2 standard Archi

are the descriptions for the 6 layers views:

1. **Strategy & Motivation layer:** Motivation elements are utilized to represent the drivers or incentives that motivate the development or adjustments of an Enterprise Architecture.

2. **Business layer:** The Business Layer represent business services realized by business processes performed by business actors or products and services to external customers. These are realized in the organization by, i.e., business processes, functions, and internal or external business services.

3. **Application layer:** describes the application services that support the business, the applications components that realize them.

4. **Technology layer:** illustrates the technology services needed to run the applications such as the physical or virtual infrastructural hardware/ software services.

5. **Implementation and migration concepts:** Present the main Imp. and migration concepts, portfolio, and project management including the migration planning.

6. **Physical Layer:** Support them modeling of the physical world such as equipment, materials, locations, assets (i.e., transport, trucks).

2.0.3 Multi-Criteria and Model-Based Analysis for Project Selection

Project Portfolio Management (PPM) is becoming increasingly important in organizations as they evaluate, choose, and prioritize project proposals based on resource allocation and reallocation. To enhance traditional PPM, Multi-Criteria Decision Analysis (MCDA) techniques, and EA model-based analyses, a new method called Multi-Criteria and Model-Based Analysis for Project Selection has been proposed [6]. This method combines capability-based assessments and EA model-based analyses to provide a comprehensive view of how organizations can make better investment decisions based on risk, cost, and benefit analyses. The method's primary objective is to guide project selection based on strategic concerns and their impact on the Enterprise Architecture. The CPB method consists of Capability-based analysis (steps 1-3) and EA model-based analysis (steps 4-7), which can be developed concurrently. The eight method steps are presented in table 2.2.

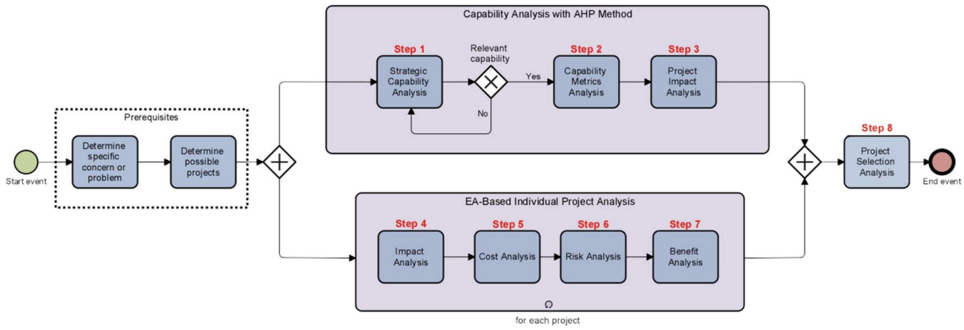


Figure 2.4: Multi-Criteria and Model-Based Project Selection Method [6]

Analysis part	**Step**	**Description**
Analysis based on capabilities	1-Strategic Capability Analysis	In this step, the Analytical Hierarchy Process (AHP) is utilized to examine capabilities and determine which one has the most significant impact in addressing an organization's problem or implementing its strategic shift. The AHP produces a capability prioritization based on their direct effects on the identified problems.
Analysis based on capabilities	2-Capability Metrics Analysis	The aim of this stage is to define the criteria to evaluate the selected capabilities, and assign a weight to each criterion that corresponds to its relevance level.
Analysis based on capabilities	3-Project Impact Analysis	The step aim is to select projects that can contribute to the improvement of a capability based on the metrics identified in step 2. Each capability improvement can be represented by a particular plateau in the capability evolution from an Enterprise Architecture perspective.
Analysis based on EA models	4-Impact Analysis	In this step, an objective analysis is conducted to determine the impact of implementing a project on the enterprise architecture. This analysis complements the subjective project impact analysis done in Step 3, previously based on stakeholders' opinions. The aim is to assess the projects interdependence's and effects on the EA.
Analysis based on EA models	5-Cost Analysis	In this step, the aim is to evaluate the costs associated in the project implementation, considering the modifications that would occur in the Enterprise Architecture, as assessed in the impact analysis.
Analysis based on EA models	6-Risk Analysis	This step aims to assess the risk level associated with the enterprise architecture modifications specified in step 4.
Analysis based on EA models	7-Benefit Analysis	This step aims to estimate the potential benefits of executing a project. This step is typically conducted concurrently with cost analysis as a component of cost analysis as part of a project evaluation method.
Analysis based on EA models	8-Project Selection and Analysis with DEA	As the last method step, the main goal is to select (the optimal) projects based on the analysis and criteria defined in the previous steps, for instance, by orchestrating the resources to minimize the cost, time and risk required while maximizing the project benefit and capability improvement.

Table 2.2: Multi-Criteria and Model-Based Analysis for Project Selection steps

Chapter 3

Systematic Literature Review

This research section aims to conduct a thorough and systematic review of literature, known as a Systematic Literature Review (SLR), to investigate the impact Enterprise Architecture on business value creation through Big Data Analytics Capabilities identification. The SLR process employed in this review follows [35] methodology. This methodology comprises three phases: Planning, Selection, and Result Analysis, executed in sequence. The detailed methodology activities in each phase are presented in table 3.1 and will be further elaborated in the following sections.

3.1 Planning

This section aims to establish the review's objectives and outline the methodology employed to accomplish these goals. The primary focus will be defining the research questions, identifying relevant scientific databases, and formulating search queries. In addition, criteria for the inclusion and exclusion of search results will be defined in this process.

3.1.1 Scientific Databases

This section uses different scientific databases to access and review relevant academic publications to answer the research questions. Particularly for this report, the following two scientific peer-review databases were selected:

1. **Scopus** (https://www.scopus.com)
2. **Web of Science** (https://www.webofscience.com/)
3. **IEEE** (https://ieeexplore.ieee.org/Xplore/home.jsp)
4. **Taylor & Francis** (https://www.tandfonline.com/)
5. **AIS** (https://aisel.aisnet.org/jais/)

Table 1: SLR Phases
Planning
1-Define main the Research Question and its Sub-Questions
2-Select scientific databases
3-Formulate search query based on the main Research Question
4-Define inclusion and exclusion criteria
Execution
5-Execution of the formulated search query for each scientific database
6-Article selection to each query results by inclusion criteria
7-Remove duplicate studies across scientific databases
8-Exclusion of irrelevant articles based on title and abstract assessment
9-Exclusion based on full text availability and its assessment
Result Analysis
10-Data extraction according to defined main Research Question
11-Synthesis of the extracted data (Discussion)

Figure 3.1: Systematic Literature Review [16]

The first two databases were selected as two of the most extensive peer-review scientific databases worldwide. First, Scopus and Web of Science possess a bast coverage of multiple multidisciplinary academic works of literature about the topic. The third and the third IEEE were related to specific topics on computer science, electronics, and, electrical engineering. Taylor & Francis is a well-known academic and scientific publisher. Finally, AIS is recognized as a top peer-to-peer scientific journal in information systems and technology. Additionally, other databases were consulted indirectly through Scopus, such as MDPI.

3.1.2 Search Queries

The process of creating an advanced search query involves the selection of a set of keywords that are relevant to the most relevant concepts of the research questions. These primary keywords are identified based on their relevance in answering the main and subquestions. For instance, performance was projected to facilitate answering the influence of remote work on teamwork. Additionally, synonyms were designated for each main keyword to expand the pool of articles that can be retrieved. Nonetheless, given the total amount of results, more specific and broader keywords were added to one of the specific groups: (i) **Subject**, (ii) **Requirements**, (iii) **Problem** and, (iv) **Contexts**. On each of these groups additional synonyms were incorporated in order to verify a proper literature examination. (Please refer to 3.2) Using the previously mentioned keywords groups, multiple search queries were created in the three scientific databases by grouping synonymous keyword concepts through the use of the logical operators "OR" and combining them with

	OR			
	Subject	**Requirements**	**Problem**	**Context**
	big data	capabilities	Value	EA
	big data Analytics	capabilitiy model	Gap	enterprise architecture
	big data dynamic	requirements	opportunity	reference architecture
	big data analytics factors	project portfolio management	metrics	model-based analysis
	advance Analytics	capability-based Planning	improvement	data Pattern
AND	machine learning	capability modeling	innovation	data architecture
	artificial Intelligence	business capability model	strategic business value	big data infrastructure
	high-volume data analysis		maturity model	
	data-intensive analysis		benefit analysis	
	large-scale data analysis		strategic business value	
	data Envelopment Analysis			

Figure 3.2: Query Keywords

other groups using the "AND" operators. The final search queries are as presented below:

Scopus:

TITLE-ABS-KEY (("Big data" OR "Big Data Analytics" OR "big data dynamic" OR "big data analytics factors" OR "Advance Analytics" OR "Machine learning" OR "Artificial Intelligence" OR "High-volume data analysis" OR "Data-intensive analysis" OR "Large-scale data analysis" OR "data Envelopment Analysis ") AND ("Project Portfolio Management" OR "capability modeling" OR "Business capability model" OR "Capabilities" OR "Capability model" OR "Capability-based Planning") AND (value OR gap OR opportunity OR metrics OR improvement OR innovation OR "Benefit analysis" OR "strategic business value" OR "Maturity model") AND (EA OR "Enterprise Architecture" OR "Reference architecture" OR "Enterprise architecture Pattern" OR "reference architecture" OR "data architecture" OR "Data Pattern" OR "Model-Based Analysis" OR “big data infrastructure”)) AND (EXCLUDE (SUBJAREA,"MATE") OR EXCLUDE (SUBJAREA,"EART") OR EXCLUDE (SUBJAREA,"SOCI"))

Web of science:

ALL=(("Big data" OR "Analytics" OR "Big Data Analytics" OR "big data dynamic" OR "big data analytics factors" OR "Advance Analytics" OR "Machine learning" OR "Artificial Intelligence" OR "High-volume data analysis" OR "Data-intensive analysis" OR "Large-scale data analysis" OR "data Envelopment Analysis ") AND ("Project Portfolio Management" OR "Capabilities" OR "Capability model" OR "Capability-based Planning") AND (value OR gap OR opportunity OR metrics OR improvement OR innovation OR "Benefit analysis" OR "strategic business value" OR "Maturity model") AND (ea OR "Enterprise Architecture" OR "Reference architecture" OR "Enterprise architecture Pattern" OR "reference architecture" OR "data architecture" OR "Data Pattern" OR "Model-Based Analysis" OR "big data infrastructure"))

IEEE:

(("Document Title": "Big data" OR "Analytics" OR "Big Data Analytics" OR "big data dynamic" OR "big data analytics factors" OR "Advance Analytics" OR "Machine learning" OR "Artificial Intelligence" OR "High-volume data analysis" OR "Data-intensive analysis" OR "Large-scale data analysis" OR "data Envelopment Analysis ") AND ("Document Title": "Project Portfolio Management" OR "Capabilities" OR "Capability model" OR "Capability-based Planning") AND ("Document Title": value OR gap OR opportunity OR metrics OR improvement OR innovation OR "Benefit analysis" OR "strategic business value" OR "Maturity model") AND ("Document Title": ea OR "Enterprise Architecture" OR "Reference architecture" OR "Enterprise architecture Pattern" OR "reference architecture" OR "data architecture" OR "Data Pattern" OR "Model-Based Analysis"))

Taylor & Francis - AIS:

("Big data" OR "Analytics" OR "Big Data Analytics" OR "big data dynamic" OR "big data analytics factors" OR "Advance Analytics" OR "Machine learning" OR "Artificial Intelligence" OR "High-volume data analysis" OR "Data-intensive analysis" OR "Large-scale data analysis" OR "data Envelopment Analysis " AND ("Project Portfolio Management" OR "Capabilities" OR "Capability model" OR "Capability-based Planning") AND (value OR gap OR opportunity OR metrics OR improvement OR innovation OR "Benefit analysis" OR "strategic business value" OR "Maturity model") AND (ea OR "Enterprise Architecture" OR "Reference architecture" OR "Enterprise architecture Pattern" OR "reference architecture" OR "data architecture" OR "Data Pattern" OR "Model-Based Analysis")

3.1.3 Inclusion and Exclusion Criteria

The criteria for determining the relevance of the literature for this review included freely available documents or those accessible through the University of Twente credentials. Furthermore, the literature should to be written in English and pertain to the subject areas of Computer Science, Business, and Information Technologies, with the publication dates from 2013 to 2023.This current review additionally adopts a systematic approach that excludes studies that lack an explicit connection to the central research topic or the themes of Enterprise Architecture, Big data Analytics capabilities, and business value by reviewing their titles, abstracts, and contents. Furthermore, redundant articles with identical titles, characteristics, or content across multiple scholarly databases will be eliminated to ensure the quality of the review. Finally, incomplete articles, limited access, and those consisting only of preliminary pages once retrieved through online search will also be excluded. (Please refer to figure 3.3)

3.1.4 Research Questions

The critical component of a systematic literature review is the precise definition of the research questions. These question(s) are the pillar for the entire systematic review

Inclusion	Exclusion
Subject area related to Business Management, Social Sciences	Duplicated articles by content, abstract or titles
Peer Review journals	Papers not available for limited access
English based papers	Studies that do not align to the research question, topics or keywords (synonyms)
	Articles published years between 2013 and 2023

Figure 3.3: Inclusion & Exclusion criteria

methodology[35]: search process, data extraction, and data analysis. The critical component of a systematic literature review is the precise definition of the research questions. These question (s) are the pillar for the entire systematic review methodology: search process, data extraction, and data analysis. As previously mentioned, this paper high level goal is to develop an Big Data Analytics Enterprise Architecture reference Architecture to support and optimally orchestrate the different Big data analytic capabilities and resources deploy the different BDA projects. Therefore, the following is the the main research question:

How can Enterprise Architecture be used to identify big data analytics capabilities (BDAC) to improve business value?

This main main research question is divided in the following three sub-questions:

Sub-questions:

1. **What is the state of the art of big data analytics capabilities?**
2. **What business value is generated using big data analytics capabilities?**
3. **Which EA tools or techniques are used to leverage business value?**

3.2 Selection

In order to improve the relevance of the study's literature review and focus resources and time on the most pertinent publications, a series of steps must be taken. Initially, the gathered articles must be reviewed, which involves executing specific search queries on scientific databases, applying established inclusion and exclusion criteria, exporting search results to EndNote, removing duplicates, and selecting full-text articles while discarding incomplete documents, those that cannot be found or limited access. Finally, the selected articles' full text is evaluated, and only those that adequately address the main questions are chosen. Following these steps resulted in selection of 34 articles 179 available articles (L). Additionally, ten articles were extracted through the snowballing method (SB) (Please refer to figure 3.4)

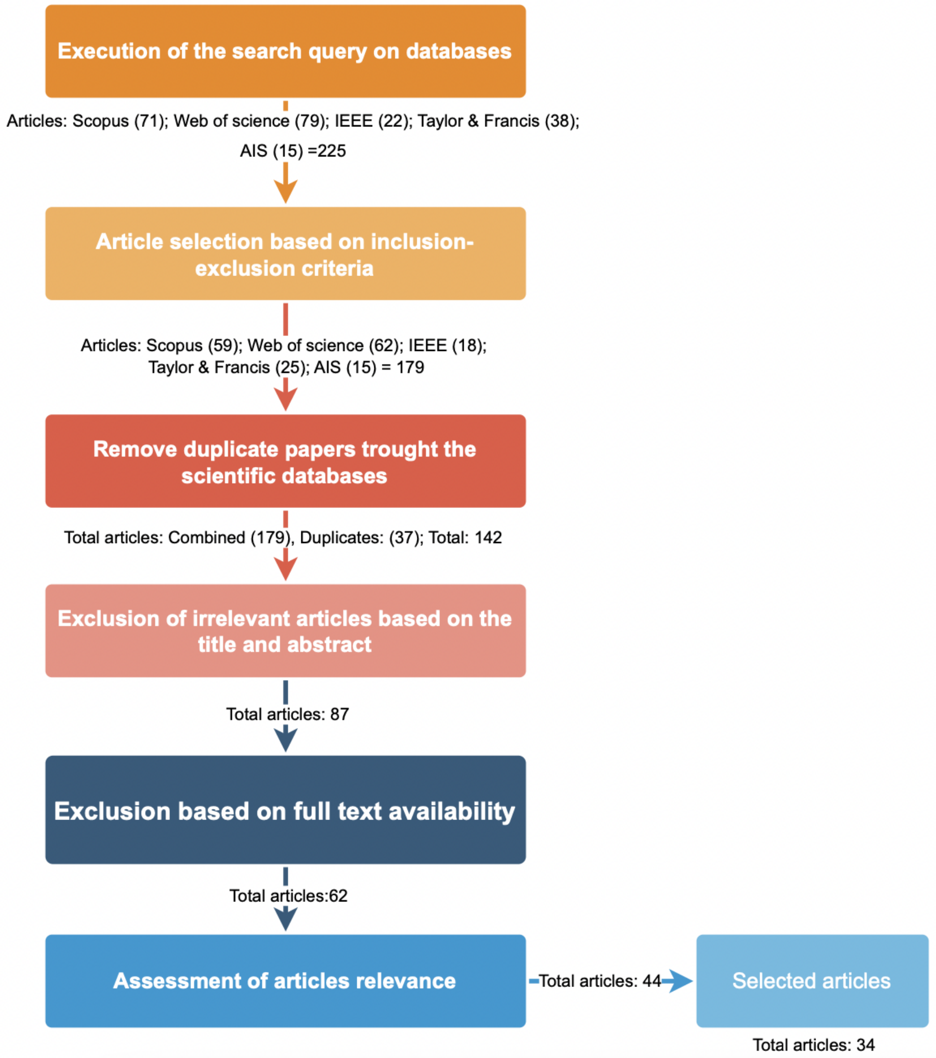


Figure 3.4: SLR Selection process

3.3 Data Extraction

Once the most relevant articles have been identified and selected, the crucial next step is gathering essential information directly addressing the defined research question. This collected content will be instrumental in comprehending how Enterprise Architecture can be utilized to identify Big Data analytics capabilities (BDAC) for enhancing business value. In order to achieve this, a systematic process of information collection is necessary, where each article is reviewed and pertinent information is extracted. The research type and category type are presented in the first and second columns of the table, respectively, while the remaining three columns represent the sub-questions related to Big Data Analytics Capabilities state the art (BDAC), Enterprise Architecture frameworks, and Big data-AI business value cases. These sub-questions and research content provide the foundation for answering the research question. The chosen articles and the research purpose are presented in Table 3.1 that serves as a reference for the information collection process.

3.4 (SLR) Results

In this chapter, the outcomes of the data extraction process are presented to address the research questions. The first sub-question focuses on the up-to-date reasons for adopting Big Data Analytics Capabilities, and the current constructs, dimensions, and sub-dimensions are discussed from a Resource-based view. The second section presents multi-industry cases demonstrating the business value derived from big data capabilities. Lastly, various current industry Enterprise Architecture Frameworks and methods are presented and discussed within the context of Big Data Analytics Capability regarding Capabilities and resource orchestration, highlighting opportunities for value creation and competitive advantage.

3.4.1 Big Data Analytics Capabilities

The resource-based view (RBV) proposes that a company can be perceived as a set of resources and capabilities [49]. To effectively incorporate emerging technologies and integrate big data into their enterprise architectures, firms must possess mature technology and business teams, develop relevant capabilities, and adopt appropriate strategies [10]. The strategic plans outlines the enterprise's direction, influencing its products/ services, competencies, capabilities, and behaviors. In dynamic and turbulent contexts, companies can gain a significant competitive advantage by enhancing their organizational capabilities through the targeted application of big data and business analytics. The ability to manage and coordinate big data-related resources is referred to as Big Data Analytics capability[46].

Big data poses a challenge to existing enterprise analytics systems while also offering new opportunities for creating competitive advantage and generating novel knowledge. Companies have capitalized on these opportunities by incorporating big data analytics capabilities into their operations by implementing and operating big data analytics capabilities [47]. However, big data analytics (BDA) is not solely about analyzing the data per se but also involves the tools, infrastructure, and means of presenting insights. BDA's capability encompasses all organizational resources necessary to leverage organizational big data resources to their full strategic potential, including tangible, intangible, and human skills.[46]. Firms require a combination of particular tangible, human, and intangible resources to build a BDA capability for building a BDA capability that is unique to each firm and generates a competitive edge [46]. Therefore, more than big data is needed to create a competitive advantage but requires a combination of financial, physical, human, and organizational resources that are difficult for competitors to match over time [25].

According to various scholars, the management, infrastructure, and talent capabilities related to big data analytics (BDA) are widely agreed upon as key dimensions of big data analytics capability (BDAC)[8]. On the same perspective, similar approaches capabilities within the different resources are found at a firm IT level, such as (1) big data technology resources; (2) big data analytics skills; and (3) organizational BDR [8]. In order to understand the impact of IT/business data resources (BDR) on business value, it is essential to comprehend their influence on firm big data capabilities (management, infrastructure capability, and talent)[8]. Hence, they can be leveraged for enterprise transformations and to optimize the potential of BDA, as different capabilities must be developed[4]. From these three categories, multiple dimensions have been used and found in the academic literature.

[8] identified three primary capability dimensions and thrir corresponding subdimentions that reflect BDAC: First, **BDA management capability** which ensures solid business decisions are made applying proper management framework with the following sub-capabilities dimensions: BDA planning, investment, coordination, and control). The second, **BDA technology capability**: refers to the flexibility of the BDA platform (connectivity, compatibility and modularity) and, **BDA talent capability**: refers to the ability of an analytics professional (e.g., someone with analytics skills or knowledge) to perform assigned tasks in the big data environment. The first one, technical knowledge (e.g., database management); technology management knowledge (e.g., visualization tools, and techniques management and deployment); business knowledge (e.g., understanding of short-term and long-term goals); and relational knowledge). In other words, the key for company managers is to understand what they can do to maximize the likelihood that their firms benefit from investments in BDR [57]. Strengthening these capabilities by virtue of big data is what will lead to competitive performance gains, and is contingent upon multiple internal and external factors[46]. As organizations that know where they are in terms of analytics adoption are better prepared to turn challenges into opportunities.[43].

As multiple academic papers and research have agree upon these three categories, researcher's have shown inconsistencies in BDAC conceptualisation and dimensions. [49] framework proposes fifteen BDAC dimensions that contribute to Functional **(F)** or Evolutionary **(E)** capabilities/resources **(R)** required to manage and deliver insights and value from data. In this framework Functional **(F)** capabilities are meant to help deliver a specific functionality focus on efficiency aspects by 'doing the things right'. In contrast, capabilities that support a firm's growth and evolution are known as evolutionary **(E)** capabilities, and they focus on "doing the right things" to achieve evolutionary fitness. This perspective that combines functional and evolutionary capabilities can help firms pursue efficiency and growth objectives in a changing business environment (Please refer to the Integrated framework 3.5). In case BDAC lacks functional fitness, it may not be able to perform its intended information processing function. Conversely, if BDAC's evolutionary fitness is weak, the benefits of its excellent functional capabilities may be diminished. Each dimension possess different type of properties that defined them as a type People **(P)**, Process **(PR)** or Technology **(T)**. The following are the nine **Functional dimensions**: (1) Technology infrastructure, (2) Financial resources, (3) Big Data and information, (4) Technical skills, (5) Analytical skills, (6) Managerial skills, (7) Tools capability, (8) Big Data management capability, and (9) Information processing capability. The other six **Evolutionary dimensions** were divided in two groups, the first called "what to change", which supports the organization achieve reconfiguration and modify its core resources and capabilities: (10) Business process integration, (11) Infrastructure

Flexibility (technology, process, and people aspects), and (12) Strategic alignment. The second group "how to change", allows the organization to seizing and re-configuring by making "fact-based decisions using BDA systems" and are represented by (13) Relationship Infrastructure , (14)Learning capability, and (15) Driven decision making. (Please refer to the table 3.2)

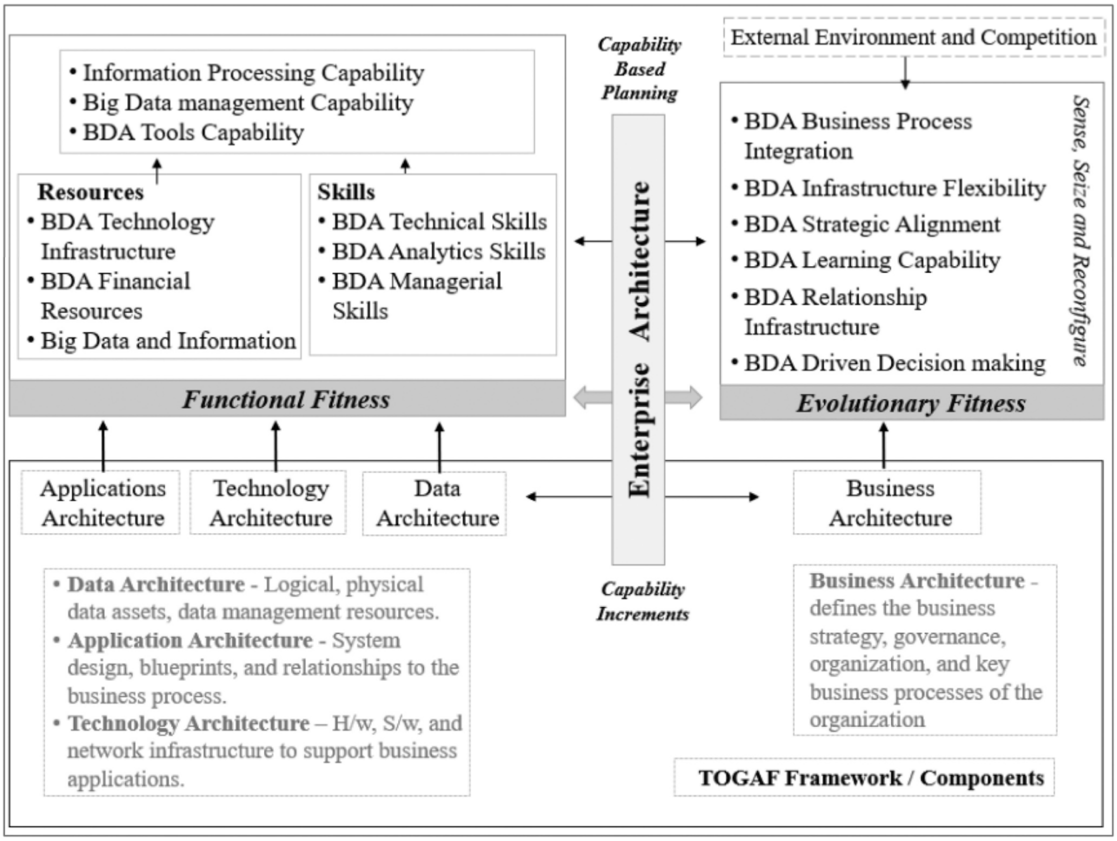


Figure 3.5: Integrated BDAC fitness framework [49]

The Resources Based View (RBV) perspective enables companies to be regarded as composition of multiple resources (such as people, processes, and technology) and with the capabilities (Functional or Evolutionary). As companies evolve and adjust to changing circumstances, the identification, orchestration and effectiveness of these resources and capabilities will determine the success of Big Data Analytics capability outcomes and incremental value creation.

Organizations are increasingly using data-driven projects to achieve their strategic value objectives, and multiple domains have shown valuable approaches to project evaluation and selection. For instance, Project Portfolio Management (PPM) plays a central role in the decision-making process of organizations by analyzing initial project proposals, prioritizing, selecting, and allocating resources based on priority. Similarly, Capability-

Based Planning (CBP) focuses on the planning, engineering, and delivery of strategic business capabilities required by the enterprise, such as skills and systems. However, these frameworks fail to consider two important project selection aspects: **1) Project interdependencies, and 2) The organizational project execution impact** [6].

Similarly, another discipline that can be considered complementary to PPM is Enterprise Architecture (EA). EA 2.0.1 can be used to explicitly define the project interdependencies and to analyze the impact, cost, risk and benefit of executing a project [6]. EA, in practice by design, is considered critical in achieving transformation objectives, requiring changes in business architecture, processes, and information and technology architecture to address digital transformation needs. Therefore, the role of EA in aligning BDAC functional and evolutionary business objectives must be examined [49].

3.4.2 Big Data Business Value

Organizations that utilize business information and analytics to make decisions are more successful than those that do not. In that sense, The business use of information and analytics differentiates them within their industry, where twice as likely to be top performers as lower performers. Top performers are twice as likely as lower performers to differentiate themselves within their industry by making decisions based on rigorous analysis [43]. The value created by BDA can be categorized into functional and symbolic. Functional value denotes the direct performance improvement achieved by the adoption of BDA, such as market share and financial performance, whereas symbolic value is largely derived from the positive image and reputation created by investment in BDA, thereby mitigating environmental/ contextual pressures. However, the success of BDA strategies goes beyond the data asset, techniques to collect and manage big data, and knowledge and implementation experience of analytics methods and tools. It requires an understanding of the mediating process and mechanisms so that BDA can serve as a resource to harness strategic business value and keep firms competitive. Simply having BDA as a valuable resource may not be enough to sustain competitive advantage [24].

Regarding business value impact and consistency with academia approach, [57] study results show that enhancing Big Data Resources (BDR) result in an 11–12% improvement in firm performance, meaning that big data resources primarily improve firm performance by enhancing the market-directed capabilities. In other words, BDR is found to play a vital role in improving the efficiency and effectiveness of strategic Market Capabilities in achieving competitive advantage. Furthermore, these results are supported by study results [25] showing the positive effect and showing that BDA capability accounted for 46.2% of the variance in Market Performance (MP) and 74.4% of the variance in Operational Performance (OP)[25]. Enterprise Architecture (EA) plays various roles in the Big Data Analytics (BDA) value creation [5]. From a top-down approach, EA identifies business requirements, aligns analytics with business goals, and sets priorities for integrating BDA into business processes and strategy. Conversely, it also employs a bottom-up approach that leverages data analytics to transform business processes, models, and product offerings and promote stakeholders' transparency. As the primary outcome, the EA roles support BDA Value Creation in four ways: 1) Enacting strategy to achieve organizational goals; 2) guiding technology selection for data sharing and analytics; 3) promoting collaboration among stakeholders; and 4) governing BDA implementation processes from planning to decision-making.

From a similar standpoint, [8] develop a BDAC comprehensive and multifaceted model to measure the capacity of big data environment and its impact on firm performance (FPER). According to the results, the higher-order BDAC constructs (management, technology, and talent capability), similar to those found by [25], demonstrated a significant standardized beta impact of 0.71 on the relationship between BDAC-FPER path, indicating that having a Big Data Analytics Capability (BDAC) will positively affect the firm's performance (FPER). This provides a compelling reason why being proficient in mobilizing different BDAC resources distinguishes FPER and generates a competitive advantage. This represents a strong argument that competence in mobilizing and deploying various BDAC resources differentiates (FPER) and creates competitive advantage. One way to accomplish this is suggested by [49] by utilizing the Coevolutionary process of adaptation or "learning by doing" cycles, which is a crucial process in enhancing BDA capabilities. Additionally, there are various value creation mechanisms that mediate the value chain between capabilities and their realization, and if identified, they can be developed to enhance the different targets of value through an integrative BDA strategy and strong leadership.

A significant portion of the selected articles mentions directly or indirectly different impacts of BDA capabilities and resources within the big data analytic capability implementation. Most of these articles focus on organizational processes opportunities/ challenges and the different phases of data collection, data processing, transformation, and outcomes within their Industries. The table 3.3 presents the compilation of several Big Data Analytics from a variety of industries and solutions applications found in the systematic literature Review (Section 3.1). The cases of Big Data implementation vary across industries, including Water Management, Metal, Telco, Energy, Logistics, Marketing & Sales, Healthcare, and Finance. Each of these industry solutions proved the multiple opportunities for value creation. On each of the different approaches to day-to-day processes, these industries and corporations are leveraging their processes through the orchestration of BDAC to generate Functional or Symbolic value that ultimately will lead to value creation and long-term competitive advantage.

Big Data implementation challenges

The implementation of big data analytics initiatives are consider in some cases are need, and in another a future competitive advantage actions are required to adapt to necessary transformations and technology disruptions not only to survive but to thrive, by leveraging current organizational processes data. In today's highly competitive and fast-paced industry context, the continuous operational improvement through data analytics allows organizations to quickly adapt to market disruptions and innovation opportunities through data drive decisions. Despite the fact that organizations are deploying big data projects, multiple challenges in terms or resources and capabilities are found in the literature and business cases, the following are some of these:

1. Fragmented and siloed applications, Privacy, Management capabilities, Authority and legitimacy [22].

2. Difficulty in integrating BDA into the current IT landscape [22].

3. Over parameterization, high computational requirements,extensive data preparation, data interoperability [34]

4. The data generated along the network systems is complex and lack of shared data standard because of its sensitivity within organisations [59]

5. Security and privacy (sensitive data), data access and storage, multiple copies of data and storing on different nodes as these nodes have to be synchronized to retrieve data efficiently [10]

6. Extracting data from multiple data sources and combining in a format that can be easily imported for analysis [10].

7. The skillset (advanced statistical techniques, data optimization methods, machine learning algorithms and thorough understanding of business value) required to extract meaningful information from big data is limited [32] [10].

8. The training and upgrading of skills and competences but also by designing AI-based systems, devices, and robots with full consideration of determining perceived characteristics of AI adoption [32], [10].

9. Finding the real cause for model outcome is complex given the underlying variables that best describe a customer's behavior [10].

10. Merging online data with offline transaction data as these datasets may not be managed by a single entity [10].

11. Customers' security and privacy concerns [10].

12. Data abnormalities due to the faulty behaviors caused either by natural conditions or by human interference [10].

13. Data leakage caused by third party intervention [10].

14. The high volume digital flood of information that is being generated at ever-higher velocities and varieties (IoT devices in Healthcare) [61].

15. Individual employee or worker, not only through training and upgrading of skills and competences but also by designing AI-based systems, devices, and robots with full consideration of determining perceived characteristics of AI adoption [32].

16. Deployment of data processes, and change management challenges [5].

3.4.3 Enterprise Architecture Frameworks

The alignment between an organization's information systems and their business infrastructure, strategies, and needs represents a critical obstacle. Alignment difficulties often arise due to multiple factors, such as insufficient knowledge of how to utilize analytics to enhance business performance, limited management resources as a result of other pressing tasks, inadequate expertise within the business department, or difficulty in acquiring the necessary data [43]. As numerous projects are unsuccessful because they do not adequately comprehend and develop BDAC competencies that guarantee BDA success. [43] suggests implementing an information agenda that aligns the goals of the IT department with those of the business through the use of enterprise information plans and deployment roadmaps. This agenda is designed to bridge the gap between the individuals who establish the organization's priorities and strategy within each business department and those who oversee the management of data and information [43].

The enterprise architecture (EA) serves as the underlying framework for the coordination of business operations and IT infrastructure, taking into account the integration and standardization demands of the company's operating model. By offering a comprehensive outlook on a company's procedures, systems, and technologies, the enterprise architecture enables individual projects to develop capabilities beyond the "short term" immediate requirements [8]. Enterprise Architecture (EA) enables the clear definition of the interrelationships and cost, risk and benefit impact analysis over project implementation [61]. Advocates of EA view it as a strategic tool for handling and overseeing the intricacy present in contemporary organizations by means of organized representation of the enterprise and its relationships[55]. Naturally, by design, EA practice should be considered as playing a crucial role in achieving digital transformation initiatives which involves modifications of the business architecture, business processes, and information, technology architecture and, the role of EA in aligning BDAC functional and evolutionary business objectives [49]. From the EA´s point of view, knowledge concerning future capabilities, products and services could contribute in designing EA changes gradually, with the goal of providing support for those targeted capabilities, products and services with well-managed effort and suitable timing [11]. As such, Enterprise Architecture (EA) can be used as a instrumental tool to facilitate the incorporation of big data analytics into the pre-existing information technology (IT) framework, thereby supporting the cultivation of capabilities necessary for deriving business value from these technologies[22].

Typically, an EA framework delineates a systematic approach and prescribed guidelines for devising and executing an EA strategy in a corporate setting through a sequence of stages. As such, multiple EA frameworks are available, each tailored by various authorities to meet particular organizations' specific requirements, thus emphasizing distinct aspects of an EA deployment [33]. Several prominent EA frameworks are currently in use, including the Zachman framework, the Department of Defense Architecture Framework, and the technology-neutral international consortium-developed The Open Group Architecture Framework (TOGAF) developed by the Open Group – a technology-neutral international consortium [1]. This EA framework has already introduced basic concepts of capabilities and Capability-based Planning (CBP) with a focus on achieving business outcomes [6]. This implies that professionals can adapt to forthcoming alterations in the functional components of the enterprise architecture while simultaneously achieving more matured capabilities and resources. Consequently, establishing a "loose coupling" between the matured capabilities and other elements of the operating enterprise architecture [12].

The TOGAF framework involves adoption the Architecture Development Method (ADM) (Please refer to TOGAF ADM and Archimate layers 3.6), a standardized cycle approach to building enterprise architecture in a structured manner through nine stages. The initial stages of the ADM involve defining and illustrating the design of the EA for a given organization. Conversely, the later stages (beginning from Phase E) primarily focus on the practical implementation of the EA.[33]. Additional to this Method, multiple specific strategic and project based methods are present in the academia and the industry. For instance, the Enterprise Strategic Alignment Method (ESAM), takes advantage of a comprehensive and interdisciplinary approach to enhance the strategic planning process through its eleven phases phases, were each stage is form by at least one strategy model. The company's business model within its operational activities as a primary focal point, which aids in synchronizing all strategic phases of the organization's development and maturity in accordance with any further modifications [9]. On the other hand, taking a project-oriented approach, the Multi-Criteria and Model-Based Analysis for Project Selection integrates conventional PPM Multi-Criteria Decision Analysis (MCDA) methods with capability-based assessments and EA model-based analyses. This approach offers a multifaceted outlook on how organizations can make more effective investment decisions by leveraging a wide range of interdisciplinary knowledge. It can be used as a reference for selecting projects based on specific strategic considerations, with the impacts being expressed in terms of risk, cost, and benefit model-based analyses. The approach is divided in two parts, both of which can be conducted simultaneously. The first segment is based on capability analysis, and it consists of the first three steps. The second segment is based on EA models, and it covers steps four to seven. The results of these two analyses are combined to determine which projects should be chosen for implementation as a methodology Enterprises may also utilize to develop business cases in more impartial and detail manner [6] (Please refer to figure 2.4).

Despite this multiple methods approaches, several research studies have identified TOGAF, with its ADM methodology, as the most commonly used by professionals ([33], [6], [49]). The Practitioners preference for TOGAF in explain because is the only framework that is backed up by a formal modeling language ArchiMate, which is widely used for enterprise architecture modeling in various industries including government, finance, healthcare, retail, telecommunications, and information technology [18]. ArchiMate is a modeling tool that enables organizations to represent and analyze their structure based on six layers: Strategy, Business, Application, Technology, Physical, and Implementation and Migration. The Strategy layer is utilized to illustrate an organization's resources, capabilities, and plans, while the Implementation and Migration layer is employed to depict the projects and their respective architectural modifications [12].

[4] proposes three ways in which EA can assist in implementing BDA: building strategic

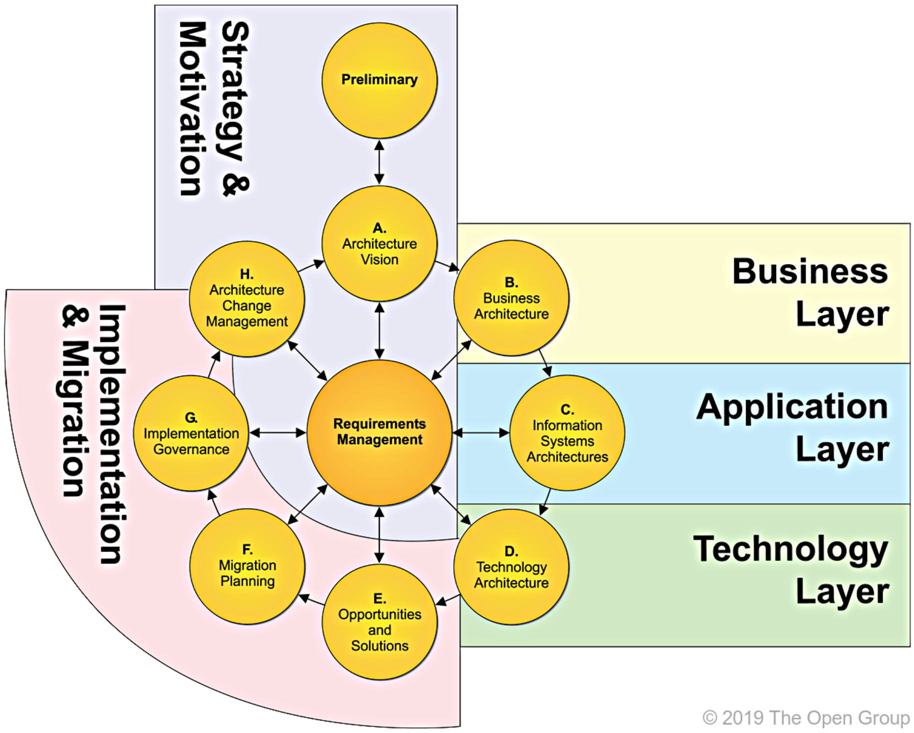


Figure 3.6: TOGAF ADM and Archimate layers [1]

capabilities, managing BDA challenges, and using EA resources to govern BDA deployment. Therefore, it is possible to connect resources and capabilities to the architecture fragments they are derived from, which enables complete traceability from strategic decisions to architecture changes and implementations and offers sufficient decision-making aid for intricate business/IT contexts; it is crucial to identify the affected modifications and their associated quickly adaptable architecture, for instance, by making visible the effects of these changes evident in the comprehensive scope of influenced Enterprise Architecture Management capabilities. However, little academic consensus on the role of enterprise architecture (EA) in implementing big data analytics (BDA), and the specific ways in which EA can support value creation through BDA are not yet well understood [5].

These results indicate some of the commonly used EA tools and techniques used in practice, along with the most relevant dimensions, roles, industry frameworks, and methods to identify and develop Big Data capabilities and resources that optimally lead to achieving higher performance and gaining business value amidst organizational internal/ external changes. However, these findings also suggest a need for more empirical evidence on using EA frameworks in coordinating BDAC within the industry to develop a competitive advantage and create business value.

3.5 SRL-Conclusion

Although Big Data is commonly associated with large-scale data processing and analytics, it encompasses a broad range of components and processes such as storage, processing, visualization, and delivery of results to target applications [17]. As such, Big Data serves as the fuel for all data-related processes from source to target and require a combination of tangible, human, and intangible resources to build a unique Big Data Analytics (BDA) capability at each firm to generate a competitive edge. Combining BDA with other organizational resources and capabilities provides a new way to sustain competitive advantage [46]. However, successfully leveraging big data to achieve its strategic business value requires a significant investment not only in data infrastructure and analytic technologies but also in skilled analysts and strategic positioning. Delineating the identified BDAC dimensions based on their fitness objectives helps organizations to examine individual dimensions' role in contributing to the firm's functional or evolutionary needs [49].

Empirical studies have demonstrated that People, Process, and Technology resources and capabilities dimensions 3.2 are pivotal to firm performance and value creation. For instance, enhancing Big Data Resources (BDR) impacted from 11 to 12% in firm performance improvement [57], and BDAC explained 74.4% of the variance in Operational Performance (OP)[25]. As such, BD's capabilities and resources are increasingly fundamental to the current company's functional and symbolic value. Therefore, organizations across various industries continuously develop localized big data business cases to improve their operations and adopt a data-driven decision-making approach. However, the success of these big data analytics (BDA) strategies is not solely determined by data assets, techniques, and analytics tools but also requires an understanding of the mediating processes and mechanisms that allow BDA to serve as a resource for harnessing strategic business value and ensuring firms' competitiveness [24].

The coordination of business operations and IT infrastructure is facilitated by the enterprise architecture (EA), which considers the integration and standardization requirements of the company's operating model. EA is regarded as a strategic tool for managing and overseeing the complexity found in modern organizations by providing an organized representation of the enterprise and its relationships [55]. Furthermore, from the EA perspective, knowledge of future capabilities, products, and services can assist in gradually designing EA changes, aiming to support targeted capabilities, products, and services with well-managed effort and appropriate timing [11]. Consequently, Enterprise Architecture (EA) serves as an effective tool to enable the integration of big data analytics into the existing information technology (IT) framework, thereby fostering the development of capabilities necessary for deriving business value from these technologies [22].

Various Enterprise Architecture (EA) frameworks are available in the market, each with specific features tailored by different authorities to meet the unique requirements of organizations and emphasizing different aspects and approaches of an EA deployment [33]. Some popular EA frameworks currently used include the Zachman, the Department of Defense Architecture, and the consortium-developed Open Group Architecture (TOGAF). The TOGAF framework, through its ADM method, is considered the most commonly used by professionals in multiple industries through the use of its modeling language ArchiMate. This language is used in multiple industries, and possess six different view layers including: Strategy, Business, Application, Technology, Physical, and Implementation and Migration layer. Through its ADM cycle and Archimate modeling, managers can strategically align business objectives with Big Data Analytics Capability (BDAC) in various business contexts and address the gaps to achieve desired goals. In other words, by identifying the capability and resources gaps, companies can manage local BDAC and resources continuously [49].

3.6 SLR Discussion

Despite the hype that surrounds big data, the business potential and mechanisms through which it results translate in competitive performance gains have remained largely unexplored to date in empirical studies [46]. Yet, there is limited research about the role of enterprise architecture in adopting big data analytics [22].While few EA studies have examined the role of EA in digital transformation, EA research is still sparse in the BDAC context [49]. The lack of consensus regarding the role of EA in BDA implementation and currently little is known about how EA role can be played out to support BDA value creation. [4]. While considerable effort has been made to define the building blocks of a firm's BDA capability, little is known so far about the processes and structures necessary to orchestrate these resources into a firm-wide capability [46].

The examination process to investigate the capability-building process has been proposed, as firms with similar BDA capabilities levels might develop them differently [46]. As a foundation for future research, [57], [46] proposes the study underscores and better understanding/building of the big data–related resources impact on the firm´s performance. Similarly, a related theme could explore how firms navigate the multiple adverse, undesirable, or contingent BDAC outcomes and examine the role of EA in governing such transitions [49]. Another research opportunity is the study of the effects on EA maturity for BDA capabilities by examining the development of required capabilities based on different stages of EA, for instance, using EA capability-based framework, such as TOGAF, to identify capability gaps and manage BDA capabilities continuously [4].

Therefore, the Resource-Based View (RBV) can be used as a solid starting point for verifying the required Big Data Analytics (BDA) resources and capability dimensions needed to achieve value creation, firm performance, and competitive advantage. The BDA integrated fitness framework (3.5) will serve as a baseline theoretical framework that combines the Enterprise Architecture (EA) TOGAF framework, its ADM methodology, and BDA dimensions. Following the Integrated-TOGAF framework and ADM methodology, the a propose EA artifact will model and identify the BDAC and dimensions. Additionally, as novel empirical approach, the artifact Multi-Criteria and Model-Based Analysis Project Selection method will be use to assess these dimensions together with their maturity level measurement in a real-world case studies within a Bid Data Analytics context. This evaluation process will assist executive and interdisciplinary teams in further identifying, communicating and orchestrating the BDA capabilities and resources to enhance the the Big Data Analytics products deployment agility and business value.

S/SB	Year	Article	Conf/Journal	Type	BDAC	EA	BDAC business values
S	2019	[6]	Journal	EA Method		T	
S	2015	[10]	Journal	Use case	T		T
S	2015	[11]	Journal	EA Model	T	T	
S	2013	[12]	Use case		T		
S	2022	[14]	Conference	Use case			T
S	2021	[22]	Journal	Case study	T	T	T
S	2022	[32]	Journal	Use case	T		T
S	2022	[34]	Journal	Use case			T
S	2016	[36]	Journal	Use case	T	T	
S	2022	[41]	Journal	Use case	T		T
S	2019	[30]	Conference	Method	T	T	
S	2023	[49]	Journal	Framework	T	T	
S	2017	[47]	Conference	Ref. Arch	T		T
S	2020	[57]	Journal	Conceptual	T		T
S	2013	[65]	Journal	Conceptual			
S	2015	[66]	Conference	Ref. Arch. Model	T		T
S	2018	[24]	Conference	Arch. Ref. Model		T	
S	2022	[5]	Conference	Lit. review			
SB	2016	[25]	Conference	Lit. review	T	T	
SB	2021	[37]	Conference	Use case	T	T	T
SB	2018	[61]	Conference	Experiment	T		T
SB	2018	[46]	Conference	Ref. Architecture		T	T
SB	2022	[45]	Journal	Use case	T	T	T
SB	2011	[43]	Journal	Lit. review	T		T
SB	2016	[8]	Journal	Lit. review	T		T
S	2016	[21]	Journal	Use case	T	T	T
S	2014	[48]	Journal	Use case	T	T	T
S	2019	[13]	Journal	Use case			T
S	2022	[18]	Conference	Use case	T	T	T
SB	2013	[50]	Consultant Research	Report	T		
SB	2016	[33]	Conference paper	Archi. Ref. Model		T	
SB	2017	[9]	PhD Thesis	Strategic method		T	

Table 3.1: Data extraction literature

Dimensions	**F/E**	**R/C**	**P/PR/T**	**Definition**
(1)Technology infrastructure	F	R	T	Specific hardware, software, and other physical assets to implement BDA systems for storing, processing, analysing, and visualising data
(2)Financial resources	F	R	Finance*	Resources needed for building, acquiring, investing, training, and supporting BDA systems
(3)Big Data and information	F	R	T	Data or information are critical resources that a firm possesses or acquires from the external environment and exploits to create value and differentiation
(4)Technical skills	F	R	P	Know-how required to build and implement new forms of technology to acquire, assimilate, integrate and extract information from big data
(5)Analytical skills	F	R	P	skills possessed by human resources to process, manage and analyse big data
(6)Managerial skills	F	R	P	Planning, investment, coordination, and control of BDA implementation and its use. Managerial skills help identify the potential of new information in the current and future needs of other business units, customers, and other partners
(7)Tools capability	F	C	T	ability and availability of various Big Data analytics tools to support day-to-day operations.
(8)Big Data management capability	F	C	T	addresses big data management issues, such as data quality, data policy compliance, regulatory requirements, and data governance
(9)Information processing capability	F	C	T	organisation's ability to process information to gather, interpret and synthesise information and enable decision-makers to process a significant amount of data
(10)Business process integration	E	C	PR	firm's ability to streamline existing business processes using IT systems
(11)Infrastructure Flexibility	E	C	PR	refers to the BDA platforms' flexibility to quickly develop, deploy and support their firms' resources
(12)Strategic alignment	E	C	PR	refers to the importance placed by the firm's top leadership on big data initiatives in achieving business objectives
(13) Relationship Infrastructure (P,PR,T)	E	C	PR	refers to the collaboration and relationships among different functional units in a firm put aside functional silos that helped achieve organisational goals.
(14)Learning capability	E	C	P	learn about continually evolving BDA tools, the firm's information needs, the latest analytical practices, and sense-making from the data
(15)Data driven decision making	E	C	P*	ability to use data-driven and fact-based decision making, creating new services or products and routine operations

Table 3.2: Big Data Analytics Dimensions [49]

Solution	**Industry**	**Department**	**Value**
Sentiment Analysis (Various)	Retail	Sales, Marketing	Investigation of users' opinions or sentiments about any product or service, expressed in textual form, on these websites/blogs
Preventing customer churn- offerings	Telecom (Celcom)	Sales, Marketing	Predictive personalized analytics to predict churn probability of its customers, personalized incentives and geolocation based cross brand promotional offers and coupons and offers
Enhancing online shopping experience (Amazon, eBay, Walter)	Retail	Sales, Marketing	Identify different customer segments and predict customers' preference and spending abilities
Smart utility meters (EnerNoc, Comverge)	Energy	Operations	Demand tariff plans, Analyze consumers peak patterns
Improving Security	-	Operations	Network monitoring, authentication and control, identity management, fraud detection, data loss prevention and control
Predictive Maintenance (Hema Industry)	Industrial	Operations	Data-driven decision making possibilities of the tool wearing and optimise breakage costs with using artificial intelligence
Various	Health care	Operations	Reduce system redundancy,Avoid unnecessary IT costs, data transfer speed, reduction patient travel time
Fraud and anomaly detection (Deutsche Bank) - trading	Finance	Operations	Consumer data, mortgages, bank accounts- predictable profitable trading
Water Management	Water	Operations	Planing and Demand (climate change and population growth in households, districts or urban levels or agriculture), prediction of surface or groundwater levels, Water quality parameters

Table 3.3: Compilation Big Data Analytics Business Value Cases from section 3.1

Chapter 4

BDAC Deployment Reference Architecture

Upon examining the fundamental concepts, current big data project initiatives, and practical implications concerning the research problem, this section derives the objectives of the proposed solution. Consistent with the stages outlined in the Design Science Research Methodology (DSRM) of sub-section: 1.4.1, this chapter focuses on the sub-research question: 2.

4.1 Define Objectives of a solution

4.1.1 Big Data Analytics Deployment Architecture goals

In this thesis section, the goals of the solution artifact for the analytical products deployment process are presented based on the results of the Systematic Literature Review (RO-1), the research frameworks and methods (chapter 2) and state of the art BDA Capabilities, Methods, Frameworks and architectural patterns (chapter 3). The aim of this solution design is to provide a blueprint and guideline to improve deployment analytical products based on different local resources and capabilities with the following three objectives of the Integrated BDAC Deployment reference architecture:

1. Demonstrate empirically the firm's BDA capabilities building blocks of a firm's BDA capability.

2. Portrait the different deployment architecture layers and transformations process to deploy big data analytical projects.

3. Identify the BDA deployment interdepencies, mediating processes and mechanisms in different local resources and capabilities.

Demonstrate empirically the firm's BDA capabilities building blocks of a firm's BDA capability

As an organization evolves in its utilization of different analytics resources and capabilities, it becomes essential to align, define, and understand the building blocks of Big Data Analytics. Due to its complex and interdisciplinary nature, it is important to establish clear boundaries and responsibilities for team collaboration in various processes. The BDA reference architecture introduces specific layers, roles, and capabilities for each step, offering guidance for the implementation and comprehension of future Big Data Analytical projects. By delineating these elements, the reference architecture aims to facilitate the efficient execution and comprehension of Big Data Analytics initiatives within the organization.

Portrait the different deployment architecture layers and transformations process to deploy Big Data Analytical projects

The deployment reference architecture merges the technical and business processes required for deploying architectures in big data analytics projects. It incorporates a range of methods, frameworks, assessments, and patterns related to MLOps, as well as the integration and transformations of the various views, including those related to business, application, technology, and data. The model utilizes the Archimate modeling language to accomplish this, which offers a high-level depiction of multiple MLOps and business processes while reducing architectural complexity.

Identify the BDA deployment interdepencies, mediating processes and mechanisms in different local resources and capabilities

The reference models will encompass fundamental processes that empower multidisciplinary teams to enhance collaboration and adapt to the unique resources and capabilities of Big Data Analytics (BDA). By incorporating various academic methodologies, the reference architecture facilitates the local BDA maturity assessment, resources, and capabilities orchestration. This alignment of multiple deployment interdependencies promotes clarity in teamwork and the overall road-map process. Furthermore, through these tailored adjustments based on local BD analytics projects, the cost and time required to articulate BDA resources, capabilities, and business analytics are reduced, resulting in an optimal architecture design that aligns with the organization's specific needs.

4.2 Design & development

Upon exploring the fundamental BDAC principles and state-of-the-art dimensions, EA Frameworks and methodologies, ML-Ops architectural patterns presented in the former chapters, this chapter will design and specify the BDAC Deployment reference architecture artifact for advanced analytical products. This objective correspond to the stages of the Design Science Research Methodology (DSRM) introduced in section subsection Design & development section (EDM phases until step 6), the additional phases will not be develop as they are out of the scope of the current research. Particularly, the subsection b (Design & development) outcome will present the designed artifact that later on will be instantiated in the chapter 5.

4.3 TOGAF ADM Methodology

The BDAC Deployment architecture (artifact) will serve two purposes: First, as a foundational path that provides a common cross-functional blueprint path to the target architecture from the current base or "as it is" architecture. Second, as a communication tool that facilitates the Business and Information Systems (IS) strategy integration in the deployment of advanced analytical products based on the local BDA resources and capabilities orchestration to distinguish: What needs to change, how to change and the data transformation required in the process. To achieve this purpose, the section will follow the TOGAF ADM framework methodology phases (Subsection 2.0.2) by:

4.3.1 Preliminary Phase:

This phase "describes the preparation and initiation activities required to meet the business directive for a new enterprise architecture, including the definition of an Organization-Specific Architecture framework and the definition of principles". The Preliminary Phase defines "where, what, why, who, and how we do architecture" in the enterprise concerned. Additionally, this phase includes the review the organizational context, identify and scope the artifact elements, capability and the established frameworks, methods, and processes that intersect with the Architecture Capability.

The BDAC Deployment Reference Architecture is situated in organizations that possess multiple internal operational systems that store and manage the retrieval of operational information from multiple industries (e.g., Retail, Telecom, Energy, Healthcare, Finance or Water-management -Please refer to table: 3.3). Commonly these organizations possesses a set of strategic goals that includes operational and financial efficiency that aim to increase the product/service market share, generate competitive advantage or maintain/ improve operational processes (Please refer to Table: 3.3).

4.3.2 Business Intelligence and Big Data Analytics product

Business intelligence and analytics refers to the field that encompasses the various technologies, applications, and processes used to collect, store, and analyze data in order to facilitate informed decision-making within a business settings [40]. On top of this analytical and BI process, multiple companies and organizations are leveraging their historic data and data patterns in the search for optimal decision-making from complex data integration's. Gartner refers to this Advanced Analytics defines it as "the autonomous or semi-autonomous examination of data or content using sophisticated techniques and tools to discover deeper insights, make predictions, or generate recommendations" through the use of advanced analytics techniques such as "[...] data/text mining, machine learning, pattern matching, forecasting, visualization, semantic analysis, sentiment analysis, network and cluster analysis, multivariate statistics, graph analysis, simulation, complex event processing, neural networks" [3].

The effective use of big data and advanced analytics can provide a more comprehensive understanding of customer needs, business operations, and customer service, insights that may have otherwise gone undiscovered or been unachievable without these tools [19]. However, to achieve this goal, corresponding mature technical and non-technical capabilities must be achieved corresponding with a target architecture to support the advance Analitycs product deployment.

Moreover, higher analytics maturity levels enables a dynamic response to the changing environment. More and more organizations wish to develop their analytics strategy besides spreadsheets or simple management dashboards. A growing number of them are attempting to build a widespread analytics culture in which data analysis will play an important role in the decision-making process[39]. However, despite the increasing relevance of data analytics, limited organizations possess the necessary analytics capabilities, including the required infrastructure, human, and skills resources in effectively managing them and meet current analytics needs [26]. Moreover, few organizations are able to accurately gauge their utilization of data analytics or to determine how to enhance the effectiveness of business processes through analytics[26].

4.3.3 Digital Maturity

Digital Maturity presents a systematic approach for organizations to transform digitally [63]. It reflects the company's current state by describing **WHAT** the company has achieved in terms of digital transformation efforts and **HOW** is systematically preparing to adapt to increasingly demanding digital environments to compete effectively. Digital Maturity encompasses not only a technological interpretation of a company's ability to perform tasks and handle information flows through IT but additionally interprets and describes past achievements in terms of digital transformations actions, including products, processes, services, skills, cultures, and capabilities to manage change management processes effectively [60].

Digital maturity combines two independent but highly related dimensions: The first, **Digital Intensity** represents the investment in technology-enabled initiatives to change how the company operates – its customer engagements, internal operations, and even business models [2]. The second, **Transformation Management Intensity** refers to creating leadership capabilities necessary to drive digital transformation in the company. Organizations developing Transformation intensity are shaping their vision to shape the new future governance and engagement towards a defined course, and the relationship between IT and business in implementing technology-based change [62].

4.3.4 Analytics Maturity Models

A **model** is a schematic and simplified higher level representation of a specific complex reality. Typically, models are based on hypothesis over essential elements that are selected and abstracted to create a general concept understating through the process of theory development and validation. In turn, **Maturity Model (MM)** is a tool designed to assess and represent the domain/objects maturity (i.e., capability, people, culture, processes,structures, technology) by evaluating the level of development based on a comprehensive set of parameters and criteria. In other words, MMs represents the foreseen evolution path of the evaluated objects or domains through a set of discrete phases [39].

The underlying concept of all models is that circumstances & objects change over time and that most of the inherited changes can be predicted and controlled [29]. As multiple changes for different domains are constantly occurring, Maturity Models are used to describe, explain and evaluate domains' evolution and transformations. Maturity Models function as a means of assessing an entity's progression along a developmental trajectory, defined by milestones/phases criteria that must be fulfilled in order to reach specific maturity levels. Some of the most important characteristics in Maturity Models are the maturity domain/object, the dimensions, the maturity levels and the assessment methods [29].

An organization's analytics maturity can be assessed in various traditional methods of measuring Analytics capabilities including self-assessment, quantitative studies, and qualitative interviews. However, traditional approaches to the analytics capacities assessment possess certain limitations due to the lack of an opportunity to verify them using what is referred to as the "depth and width" of analytics capacity. **Self-assessment** and **Quantitative** studies commonly involve using a checklist to evaluate the extent to which specific technologies and tools have been implemented within an organization. This type of assessment typically involves a review of the current state of technology implementation. However, the first two methods do not provide and limit the insights into whether an organization effectively utilizes these technologies and tools to make business decisions or how these impact the organization's operations.

In turn, management **Qualitative interviews** might only cover a small portion of the organization's Analytics Maturity and may be anecdotal. Furthermore, these studies could miss variances in Analytics maturity amongst multiple employee groups within the organization [39]. From these former three options, only the Analytics Maturity Assessment Model represents an alternative method to define and reach a desired level of maturity; as a way for an organization to progress along a transformative path from an "as-it-is" state to a "target stage." The following represent the four maturity models and assessments related to the current Global Analytics Maturity Assessment development: Capability Maturity Model Integration-CMMI, Analytic Processes Maturity Model-APMM, Blast Analytics Maturity Assessment Framework, or Gartner's Maturity Model for Data and Analytics. (Please refer to the appendix section: A.1)

4.3.5 Analytics Maturity path

The evolution of the enterprise's use of analytics has been dynamic and gradual. As organizational changes might vary in terms of intensity and order, given a particular company's specificity and business context [39]. In Analytics 1.0 era, used data warehouses and analytics based on operational data copies. Analytics 2.0 supported distributed processing for large data sets across computers clusters (Hadoop cluster) and non relational databases(NoSQL). Finally, Analytics 3.0 uses "Agile" analytics methods and scalable Machine Learning techniques and new technologies integration (refer to Figure 4.1) [39]. Assessing the organization's capabilities to utilize data analytics to drive innovation and

Figure 4.1: Analytics continuum[39]

gain competitive advantage requires an assessment on the current position on the known as "Analytics Continuum". In the present, data Analytics can be divided into three categories: by tools, techniques, and the approaches: Traditional Analytics, (1) Descriptive Analytics, (2) Diagnostic Analytics; Advance Analytics (3), Predictive Analytics, (4) Prescriptive Analytics, and (5) Cognitive Analytics. In practice each of these five approaches coexist and complement each other (refer to Figure 4.2). Each step on the analytics continuum brings the organization closer to solutions that enable faster, data-driven decision-making (on-demand enterprise)[31].

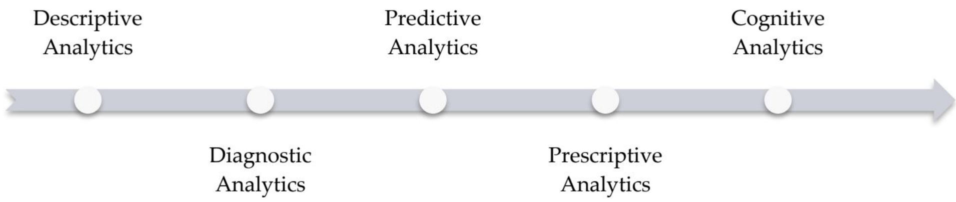


Figure 4.2: Analytics continuum[39]

1. **Descriptive Analytics** is the process of using data or content to understand past/current events through the use of traditional Business Intelligence (BI) techniques and visualizations (pie charts, bar charts, line graphs, tables, or generated narratives). Typically, it involves the manual examination of data and isolated patterns to answer the question "What happened?" or "What is happening?". It is commonly applied to discover economic insights and increase the operational process's effectiveness.

2. **Diagnostic Analytics** is a type of advanced analytics that uses data or content to determine the cause of an event. It involves examining data to answer the question "Why did it happen?" using historic data defined in specific time intervals to detect irregularities or quantitative relationships from multiple variables. Some of the most common Diagnostic techniques are drill-down, data discovery, data mining, and correlations.

3. **Predictive Analytics** is categorized as advanced analytics, which involves predicting modeling, simulation, forecasting, and machine learning to analyze current and historical data to make predictions about future or otherwise unknown future events. It groups multiple data analysis methods to forecast future outcomes and gain insight by answering "What would happen in happen in the future?" by predicting future events and trends. Predictive Analytics is characterized by searching patterns and historic variables relationships to generate a forecast, using past events to predict future behaviors, anticipate optimal outcomes or decision-making. Recent technology that supports this analysis are NoSQL, Cloud Data Bases, scalable Processing power and, Data Lakes.

4. **Prescriptive Analytics** is an additional advanced Analytics that supports decision-making processes and suggest potential courses of action for taking advantage of future opportunities or reduce risks probabilities. In other words, Prescriptive Analytics helps organizations consider the best "course of action" in light of available data and information obtained through Descriptive and Predictive Analytics. It answers to the question "What can we do to make XYZ happen?".

5. **Cognitive Analytics** uses Artificial Intelligence (AI), high-performance data analysis technology, and real-time data to increase decision-making efficiencies by using human-like intelligence for specific tasks. Some of the current technologies that support Prescriptive Analytics are Artificial Intelligence Algorithms, Deep Learning, and Machine Learning.

In recent years organizations have matured in terms of digital maturity and increased their experiences in the use of Descriptive/diagnostic Analytics, many of which are preparing to advance in their continuum Analytics path. For instance, by following the multiple steps towards higher levels of Data Analytics, as Predictive Analytics or envisioning the future state-of-the art Cognitive Analytics [31]. Nonetheless, to continue progressing in the Advance Analytics Maturity Path, organizations will require a different set of Infrastructures and Maturity Capabilities to extract disruptive strategic insights and create incremental value from the different Advance Analytics Products (refer to Figure 4.3).

The current business analytics processes within a company will determine the state of its Analytics capability "as-it-is" state based on the level of analytics maturity and the corresponding digital maturity of its big data (BD) resources. As organizations progress and enhance their BD resources to achieve the desired future state of big data analytics (BDA) capability, various digital transformations involving people, processes, and technology become necessary. For example, different BD resources and levels of BDA capability maturity are needed for descriptive analysis compared to predictive analysis. Therefore, it is crucial to evaluate the maturity levels of BD and the dimensions of capability required for the specific analytics approaches within a company's business analytics processes.

BDA resources challenges

Deploying a Big Data Analytics project is a complex process, that requires not only a enormous cross functional team, but inherit a set of common resource challenges that are

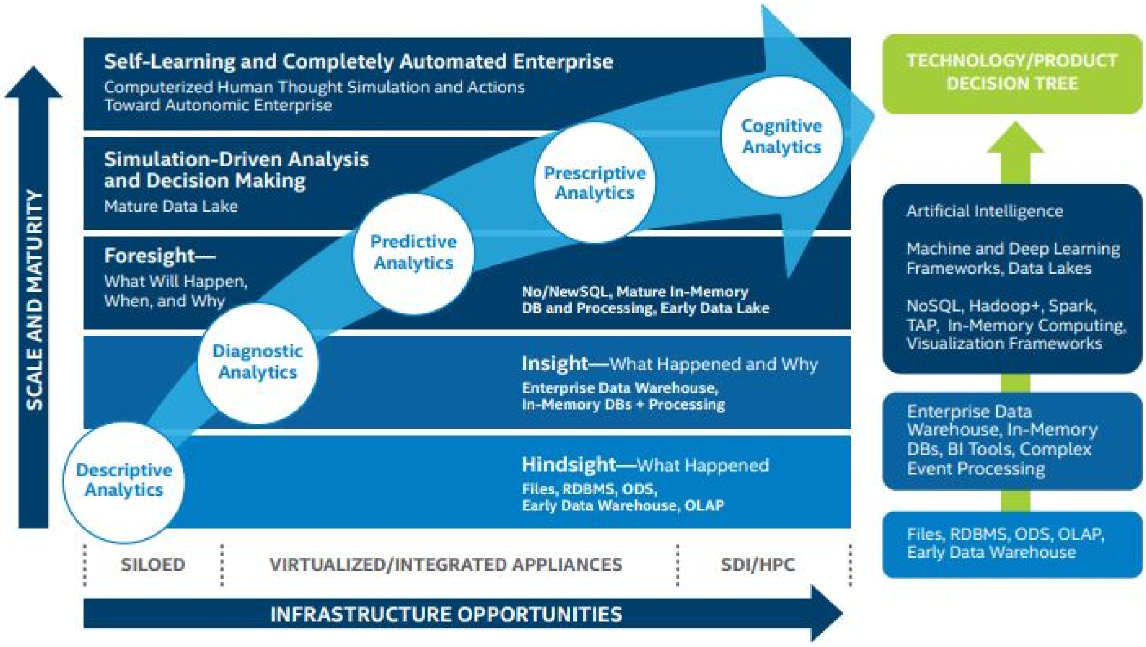


Figure 4.3: Advanced Analytics Maturity Path [31]

part of the cross-functional use and integration of DBA resources. The following are the most common BDA resources challenges with a Big Data Analytics project initiative:

Resources	Challenges
People	Data driven culture, limited skills to extract meaningful information from data, ML adoption training and upgrading of analytical skills and competences
Process	The data generated along the network systems is complex, lack of shared data standard because of its sensitivity within organisations, Data abnormalities due to human or business complexity, legacy systems conditions or by human interference (e.g., Financial compliance, GDPR law), Data leakage.
Technology	Fragmented and siloed data and applications, difficulty integrating BDA into the current IT landscape, complex data integration, limited data standards, data security & privacy.

Table 4.1: BDA Resources Challenges, based subsection: 3.4.2

Once the organization evaluates and clearly defines the required maturity levels of various analytics resources, along with the associated implicit challenges, the next step is to understand the scope of the analytical business. This entails gaining a deeper understanding of the multiple stakeholders, processes, systems, applications, and dimensions of capabilities needed to implement big data analytics (BDA) within an existing business process. To accomplish this objective, enterprise architecture serves as a strategic roadmap blueprint, enabling the integration and operation of resources and capabilities within the given business context.

4.3.6 Architecture Vision Phase:

This phase describes the initial step of the Architecture Development Method (ADM). It includes information about defining the scope, identifying the stakeholders, and creating the Architecture goals. The Architecture Vision provides the sponsor with a key tool to sell the benefits of the proposed capability to stakeholders and decision-makers within the enterprise. Architecture Vision describes how the new capability will meet the business goals and strategic objectives and address the stakeholder concerns when implemented. Clarifying and agreeing the purpose of the architecture effort is one of the key parts of this activity, and the purpose needs to be clearly reflected in the vision that is created.

Architecture projects are often undertaken with a specific purpose in mind - a specific set of business drivers that represent the return on investment for the stakeholders in the architecture development. In the case of Bid Data Analytics projects, the purpose of the present architecture process were presented in section: 4.1. The Architecture capability vision provides a first-cut, high-level description of the Baseline and Target Architectures, covering the business, data, application, and technology domains. These outline descriptions are developed in subsequent phases.

BDA Capabilities dimensions

The business capability serves as an abstraction representation view of the business operation reality that facilitate the different stakeholders roles common communication. These capabilities permit the creation of a common contextual framework that integrates the main supporting dimensions/ components (People, Processes, Technology) and facilitate the stakeholders communication in order to achieve a particular business goal and make solid business decisions [1]. Within the current research context, the chapter 3, subsection 3.4.1 have establish a robust common theoretic background through the Integrated Big Data Analytics Capabilities (BDAC) framework that propose how these BDAC dimensions leverage big data firms business value (Please refer to figure 3.5). In this integrated framework, the different capabilities dimensions are divided in two groups (Please refer to table 3.2): Functional and **Evolutionary**.

Particularly, this second group dimensions is understood as capabilities that allow firms to reconfigure their functional resources, capabilities and information assets to respond to changes in business environments. For this reason, the framework propose that these dimensions are to be examined different, as they directly or indirectly support the BDA investments success and drive performance effects in innovation and **agility**. In order to assess the the BDAC deployment architecture capabilities, the Evolutionary dimensions have been grouped based on the relation of their current construct nature and goals in three main capabilities, group by the definition integration in deployment process in: 1) What to change?, 2) How to change? , and 3) What are the required data transformation processes?:

1. **What to change? Business-Infrastructure alignment:** Ability to optimize

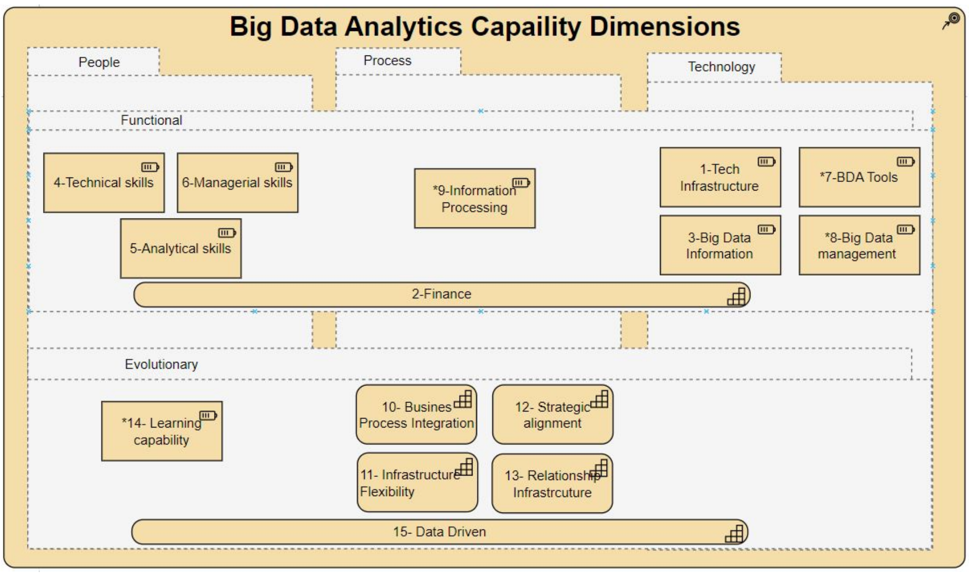


Figure 4.4: BDAC Deployment Reference Architecture BDA Capabilities and Resources dimensions

existing/new business processes information using internal and external opportunities (i.e., functional operational data, Nielsen, world-bank). Integrated BDAC evolutionary dimensions:

(a) **Business Process Integration**

(b) **Relationship Infrastructure**

2. **How to change? Seizing & Reconfiguration:** Ability to make optimal investment decision to quickly develop, deploy and support the firms' resources (People, Process, and Technology) and target performance gaps. Integrated BDAC evolutionary dimensions:

   (a) **Strategic alignment**

   (b) **Infrastructure Flexibility**

   (c) **BDA Driven**

3. **Which data transformations? Information Transformation:** Ability to extract, transform, store and load high quality information from different functional sources for better decision making.

   (a) **BDA learning ability**

   (b) **Infrastructure flexibility**

These three evolutionary capabilities: business process integration, strategic alignment, and infrastructure Flexibility will be utilized to subsequently support a firm ´s BD resources and BDA capabilities dimension reconfiguration and deployment architecture levels following the TOGAF ADM phase Vision - Opportunities and Solutions (Figure: 2.3),

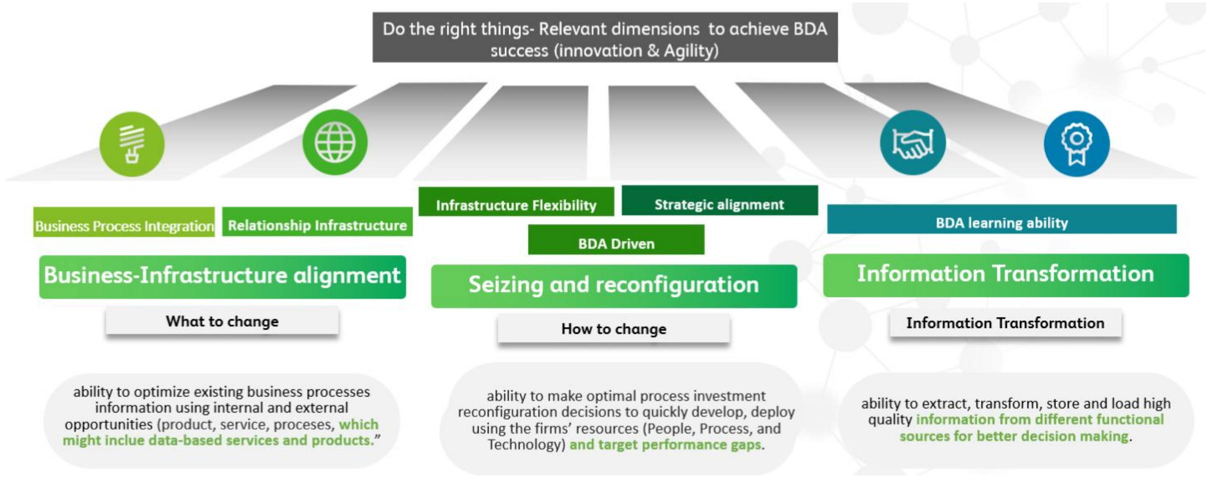


Figure 4.5: BDAC Deployment Reference Architecture capabilities

employing the Multi-Criteria and Model-Based Analysis Capability Model Methodology (Figure: 2.4). However, for the current research scope and contextual limitations, only the Prerequisites and Capability Analysis will be utilized, along with the Analytics Maturity assessment as a capability measured instead of the AHP method.

Integrated Multi-Criteria and Model-Based Project Selection- BDA capabilities dimensions- Analytics Maturity Assessment

To establish the architectural vision for specific Big Data analytics projects, the Multi-Criteria and Model-Based Project Selection Method [6] is employed, which involves the following steps: Pre-requisites analysis and the first three steps of Capabilities analysis.(Please refer to Figure: 2.0.3 & Figure:4.6):

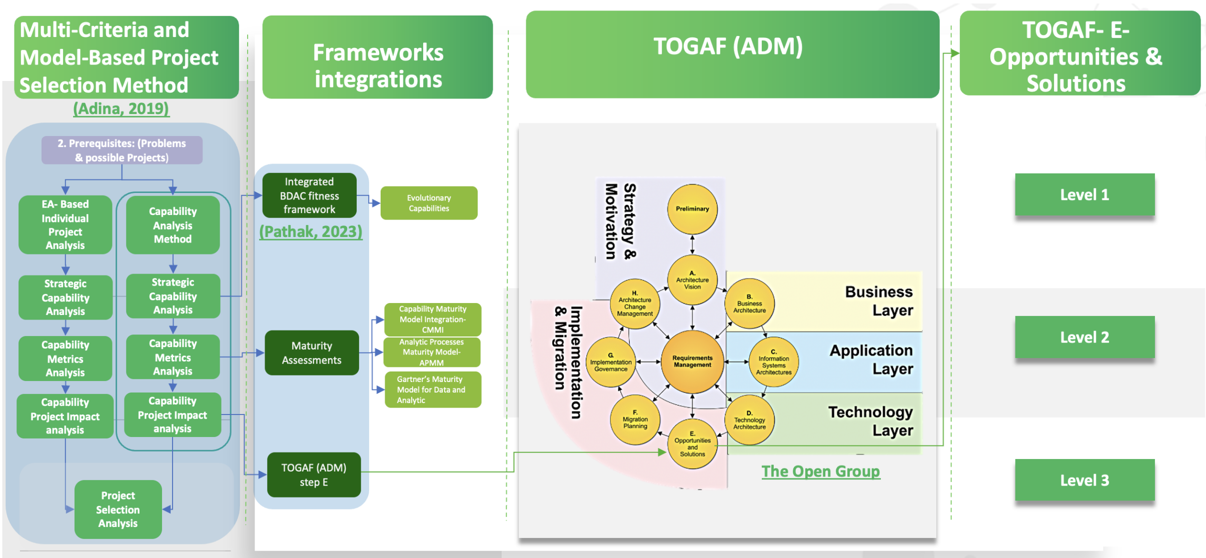


Figure 4.6: Method Multi-Criteria and Model-Based Project Selection- BDA capabilities dimensions

1. **Pre-requisite- Determined the specific concern or business problem:** validates the Big Data project initiatives implementation problems, such as the BDA project implementation challenges ones found in the SRL subsection: 3.4.2.

2. **Pre-requisite- Determined possible projects:** assess the adequate big data analytical project depending on the business goal and Analytics Maturity path (e.g., Diagnostic, Predictive, Prescriptive) (Please refer to subsection: 4.3.5).

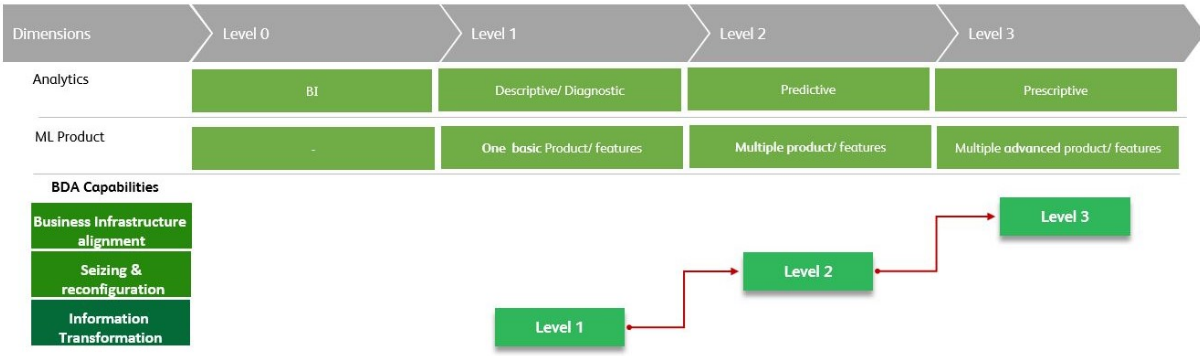


Figure 4.7: Multi-Criteria and Model-Based Project Selection- BDA capabilities dimensions

3. **Capability Analysis-Strategic Capability Analysis:** validates the BDA Capabilities levels (Section: 4.3.6) required for the Big Data Analytics projects.

4. **Capability Analysis-Capability Metrics Analysis:** This step aims to define the criteria to evaluate the selected BDA capabilities as the ability to use the BDA resources maturity (People, Process, Technology) to deploy the BDA project initiative (Please refer to subsection: 4.3.4). To evaluate this process, it will be required to map the current BDA capabilities dimensions to the Analytics Maturity assessment questions and use them as an architecture guiding architecture patterns. (Please refer to the Figure: 4.8).

5. **Capability Analysis- Project Impact Analysis:** select projects that can contribute to the improvement of a capability based on the metrics identified in the previous step. This step will be demonstrated in the TOGAF- ADM- phase E: Opportunities and solutions.

BDA Capability deployment Architecture goals

The architecture goals are as follow, based on the EA framework, SLR and BDA goals is to create a BDA capability reference Architecture. This process involves the use of insights gained from the SLR to identify the requirements and design patterns that should be considered in the artifact design. For this step, multiple BDA dimensions are gathered for the BDAC reference architecture and EA frameworks. The following represent the integrated Architecture goals (Please refer to Figure:4.2):

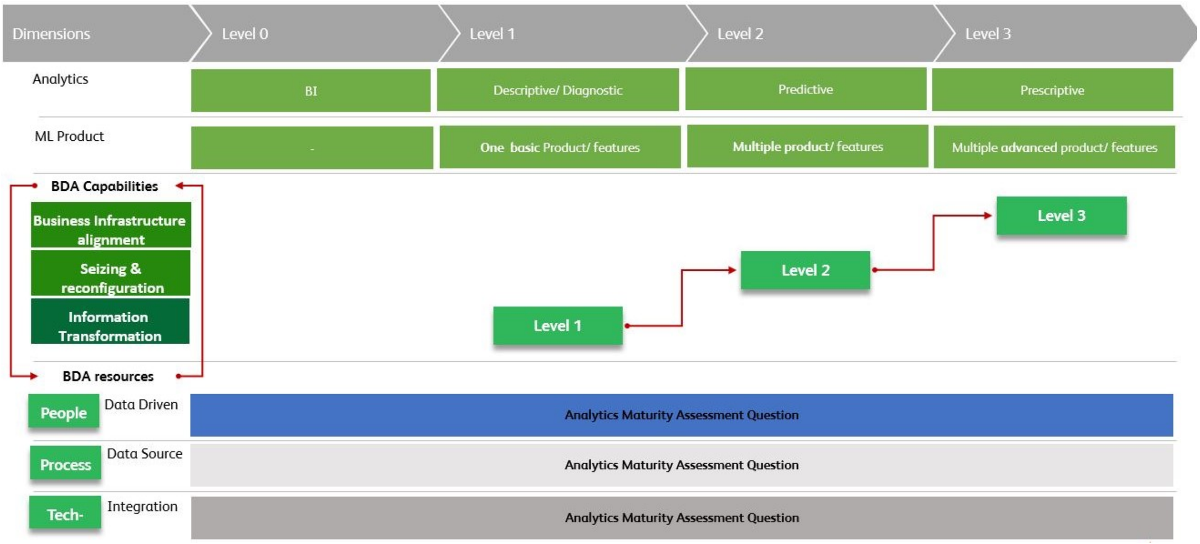


Figure 4.8: Strategic Capability Analysis- Integrated BD Analytics Maturity assessment and capabilities dimensions

EA Benefits	Architecture goals	BDA Capabilities
Resource Complementary	Demonstrate empirically the firm's BDA capability building blocks of a firm's BDA capability	Business Process Integration: ability to optimize existing business processes information using internal and external opportunities.
Resource Portfolio Optimisation	Identify the BDA project deployment interdepencies, mediating processes and mechanisms in different local resources and capabilities.	Size & Reconfiguration: Flexibility to quickly develop, deploy and support their firms' resources (People, Process, and Technology).
Organisational Alignment	Portrait the different architecture layers and transformations process to deploy Analytical products.	Data Transformation: ability to extract, transform, store and load high quality information from different functional sources for better decision making.

Table 4.2: BDA Capability deployment Architecture goals

BD Analytics project initiative Deployment reference architecture roles

Once the business process/service analytical needs within its Analytics operating model and their current BDA resources maturity levels, and define the stakeholders roles necessary in the DBA Capability Deployment reference architecture. To achieve this, an MLOps model [38] is an interdisciplinary group process, and the interplay of different roles is crucial to design, manage, automate, and operate an ML system in production and finally impact the final user operations and value creation. In the following, every role, its purpose, and related tasks are briefly described, together with their correspondent roles in the Integrated business analysis process:

1. **Product Owner:** (Project Manager) is responsible for setting the Machine Learning (ML) project objectives and managing the communication aspects of the business(i.e., Churn initiative or return on investment (ROI)) generated by the ML product.

2. **Function Stakeholder:** Local/regional/global function management role that directly or indirectly will benefit from the analytical product. Generally, for the product deployment this person is the local business function that would serve as a point of contact.

3. **Data Engineer:** involves constructing and supervising pipelines for data and feature engineering. Additionally, this position guarantees the correct ingestion of data into the databases of the feature store system.

4. **Data Scientist:** responsible for converting the business problem into a machine learning (ML) problem and managing the model development process, which involves selecting the most effective algorithm and hyper-parameters.

5. **DevOps Engineer:** connects development and operations team to guarantee efficient automation of the CI/CD process, orchestration of the ML workflow, deployment of the model to the production environment, and monitoring of the entire ML system.

6. **ML Engineer/MLOps Engineer.** Cross-functional role from various domains, including data scientists, data engineers, and DevOps engineers. Its main responsibility is to operate within this cross-domain capacity and establish and maintain the machine learning (ML) infrastructure, overseeing the automated ML workflow pipelines, and facilitating the deployment of models into production.

7. **Solution Architect:** create and defines the appropriate technologies to be utilized after a comprehensive assessment.

4.3.7 Business Architecture phase

Function Business Analytics Process

TOGAF's [1] Architecture Development Method (ADM)- step B, facilitates the continuous advancement of the target Business Architecture (BA) to meet evolving business objectives and address the strategic requirements changes. In the case of Big Data analytical initiatives, the process starts within the local function operations systems, which collect and stores the different digital operations decisions taken in the day to day. For instance, the number of customers that were contacted last month, the most sold SKU by different segments, the production time taken for a particular product. As these process are measured and stored, the business actors will be able to take a more data driven decisions. Within these business process, the following is a common representation of a Business Analytics process (Please refer to Figure: 4.9):

1. **Information systems:** represents the actual digital system and data sources required to support the daily operational processes. Different data types are find in this systems, such as: structured data, semi-structured data, or unstructured data.
2. **Data Extraction:** is the data extraction possess from the data sources.
3. **Data Transformation:** processes the extracted data.
4. **Data Analysis:** responsible for all data processing and performing the data analyses.

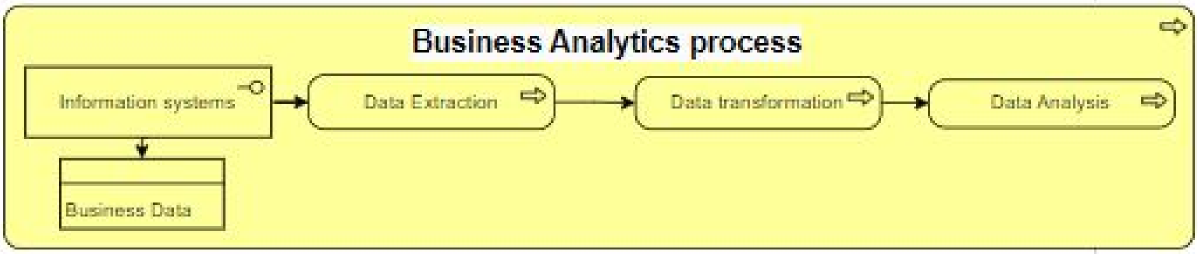


Figure 4.9: Business Analytics Process

Project Initiation

The project Initiation represents the common starting point in the big data analytics project, which triggers the activation and coordination throughout the different actor roles to five architecture layers: Business, Technology, Data Engineering, Continuous Improvement/Deployment, and Machine Learning (Please refer to figure: 4.10). First, the **Business** represents the process Function Business Analytics process, **Technology** presents the hardware, software and local IT elements, arranged in an specific configuration to serve and support the business data operations, **Data Engineering** describe the designing and building of large-scale systems to collect, store, and transform the operational data, **Continuous Improvement/Deployment** portrait the DevOps method to deliver code updates in a frequent, reliable, and quick to build, test, delivery, and deploy steps to provides rapid feedback and improve productivity (Both, for the previous and the next layer). Finally, **Machine Learning** represents the process to build a model based on data coming from the CI/CD component, to be use in different advanced analytics (i.e., predictions or decision making) automatically. The following are the basic processes in this layer:

- The business stakeholder analyzes the business and identifies a potential business problem to be addressed by ML techniques.
- The solution architect articulate the ML system architecture design and, the required technologies.
- The data scientist assess the proper ML problem to be solve the business goal.

- The data engineer and data scientist locate the raw data sources for the initial data analysis. For instance, verifying the data source labeling.
- Finally, the labeled data will be passed through the different ML models, and be presented to the business stakeholder.

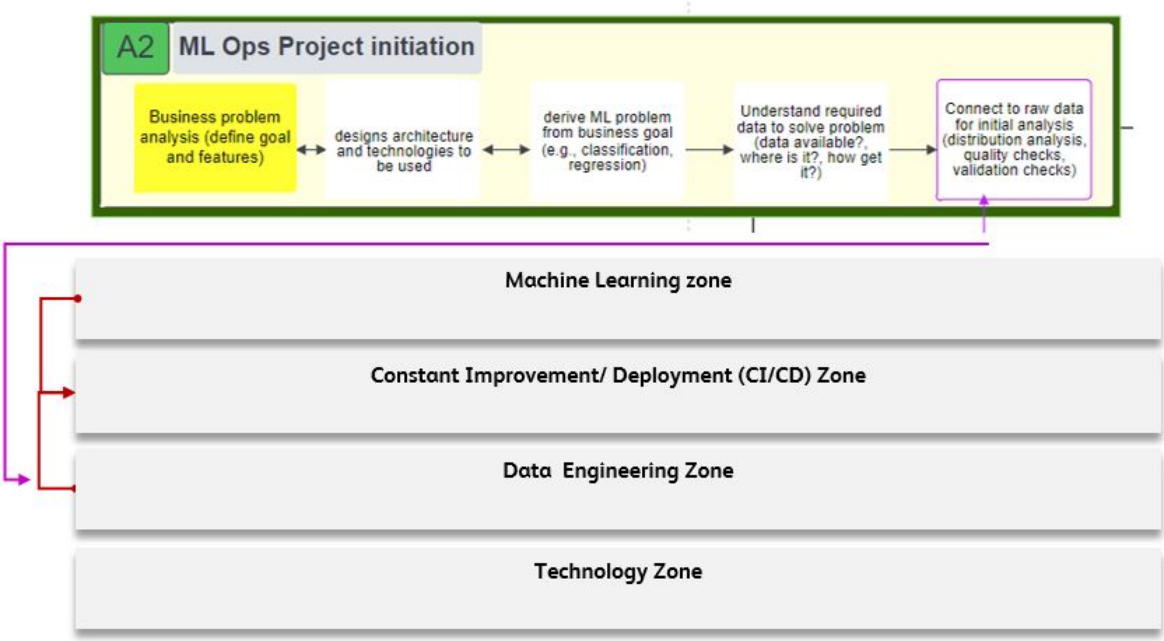


Figure 4.10: Dig Data Analytics Project Initiation zones [38]

4.3.8 Information Systems Architecture phase

TOGAF's [1] ADM-step C outlines the procedure and principles for the Application architecture and Data Architecture base architecture ("as-it-is"), aiming to facilitate the development of target ("to-be") Data Architecture and Application Architecture. Both target architectures goals is to align with the target Business Architecture needs, and the maturity of BDA capabilities and resources (People, process, and Technology). For instance, within a business process (Procurement, Production, Sales or Marketing) a new BDA Analytics initiative project is required, for this purpose the data and application target architecture will be used to identify, map and represent the data and the applications required to deploy the new business analytic process.

The 'business-information' diagrams generated during the target Business Architecture (BA) development in ADM-step B (figure: 5.1.7), guided definition of the different data relationships across different business applications based on the local resources maturity (Please refer to figure: 4.11). The definitions and deployment architecture levels will be further explain in chapter: 5. Nonetheless, in the following are the corresponding reference architecture for the current Integrated Business Analytics process zones: **Data Engineering**, **Continuous Improvement/Deployment**, **Machine Learning**.

Data Engineering pipeline Zone:

1. **Requirements for Feature engineering pipeline:** involve the identification of the essential attributes necessary for model training and model features. Once a preliminary understanding of the raw data and initial data analysis is obtained, the following fundamental requirements for the feature engineering pipeline are established to extract, transform, store and load the required model data/ metadata (Please refer to the figure 4.11- Zone B):
   (a) The data engineer sets the data transformation guidelines, such as normalization and aggregations, in order to extract the required data format.
   (b) The data scientist and data engineer to establish feature engineering rules. These rules entail the calculation of new and more sophisticated features based on the current ones.
   (c) The initially defined rules undergo iterative adjustments by the data scientist. These adjustments are informed by feedback obtained from the experimental model engineering phase or assessed through the model performance.
2. **Feature engineering pipeline:** represents the data feature pipelines that: A) build between the data sources storage to be used in the BDA project the advanced analytical model, and B) connects the extract model outcome to the User interface systems. The following are the most relevant steps on this process (Refer to Figure: 4.11- zone B1 Section-B):
   (a) The feature engineering pipeline starts with the initial data requirements, which serve as a foundation for the data engineer and software engineer to develop a pipeline prototype. The requirements and rules are subject to updates based on the feedback received from the observed model's performance in production.
   (b) The data engineer set the code required for the CI/CD and orchestration component to ensure the task orchestration of the feature engineering pipeline.
   (c) In the feature pipeline, the initial step involves the establishing a connection with the data source raw data, which can be streaming data, static batch data, or data stored in any cloud storage.
   (d) The data is extracted from the multiple data sources.
   (e) The data pre-prossessing begins with data transformation and cleaning tasks to make the source data into a usable format.

Machine Learning Production/ Implementtion Zone

This zone represents the BDA project initiative models and its multiple features that depends on the Analytics business process needs. The ML zone is triggered by the acquisition of data gathered in the previous section, which can encompass numerical values, images, or textual information (i.e., financial transactions, SKU items, maintenance records, time series data from sensors, the brewery, or sales reports). The following are the most important steps by steps on this zone (Please refer to the Figure 4.11- zone C):

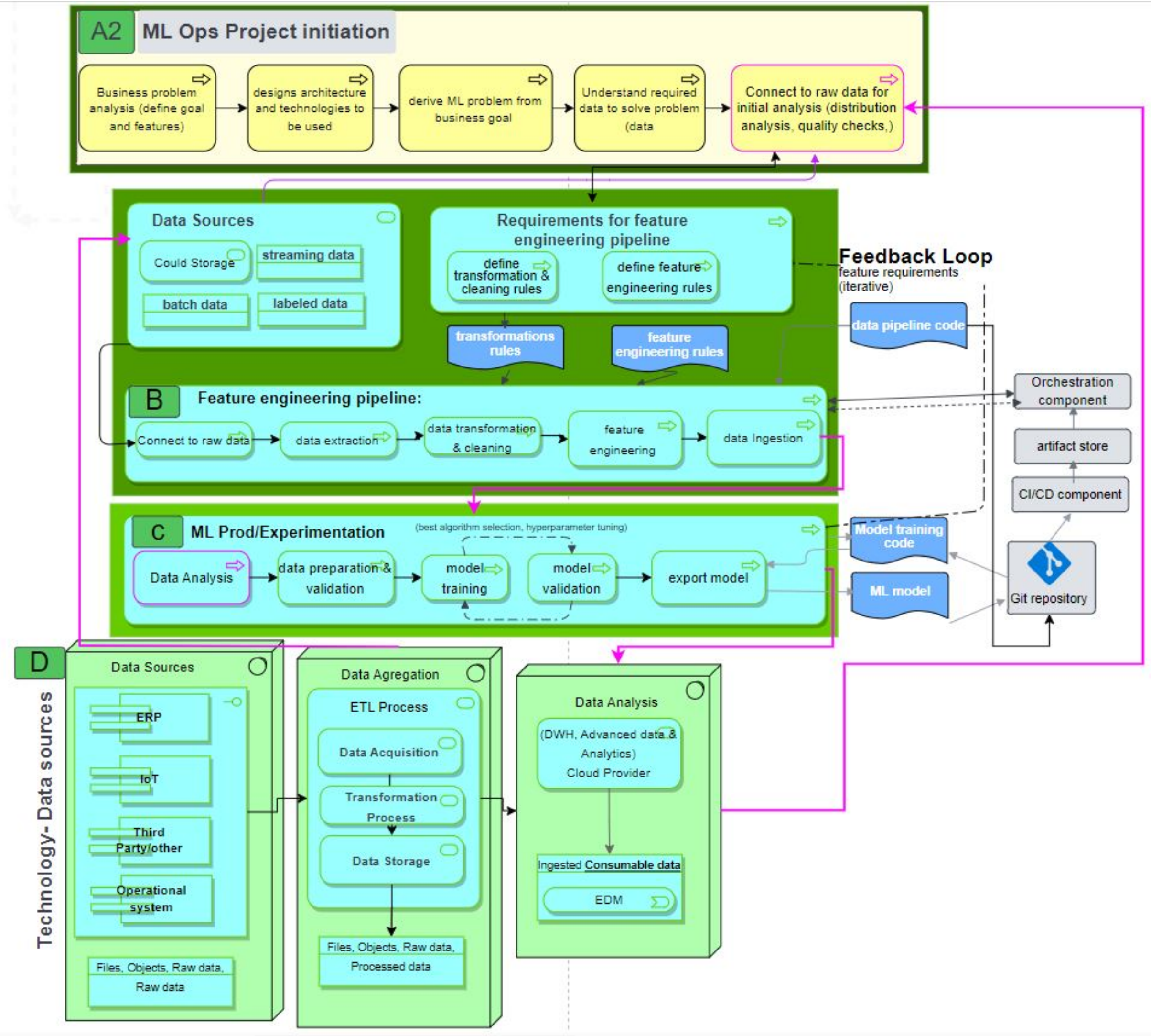


Figure 4.11: BDA Capability Reference Architecture System Layer

1. The data scientist connects to the feature store system for the data analysis, connects directly to the raw data for initial analysis, and communicates the required data changes back to the data engineering zone.

2. Involves data preparation and validation obtained from the feature store system, including the creation of train and test split datasets.

3. The data scientist estimates the best-performing algorithm and hyper-parameters by triggering the model training process using the training data.

4. The optimal model is exported and its code committed to the repository. Additionally, depending of the systems interaction and IT landscape, the DevOps engineer/ ML engineer establishes an automated ML workflow pipeline so that a new ML model or ML workflow pipeline code is committed to the repository.

4.3.9 Technology Architecture phase

The Technology Architecture phase (ADM-step D) supports the development and alignment of based and target architecture to support the vision and the target business, application, and Data Architecture. The outcomes of ADM-step D includes: technology standards and portfolio, application or technology, and platform, processing, networked, and hardware that align the target Business Architecture (BA) and Advanced Analytical product resources and BDA capabilities required. The target Technology Architecture Supports the alignment between specific local technology infrastructure, investments, and financial resources for Big Data Analytics (BDA) context.

The main representation outputs of this phase catalogues (technology standards and technology portfolio), matrices (application or technology matrix), and diagrams, such as platform, processing, networked and hardware diagrams in alignment with the target Business Architecture. The target TA helps align the BDA-specific technology infrastructure, BDA investments, and financial resources. TOGAF suggests that technology becomes a driver and a strategic resource rather than a recipient of BA's change requests. As a result, the Technology Architecture in parallel drive business capabilities and respond to technology requirements [49]. The following are the key Technology data sources and processes, based on the Systematic literature review academic sources: [59], [61], [14], [34]:

1. **Data Sources layer:** This layer assumes responsibility for managing data originating from various data sources. The data undergoes three essential steps within this layer: acquisition, transformation, and storage. The data acquisition goal is to retrieve data from diverse data sources, sizes, and formats. This step often poses a significant challenge during the initial stages of implementing big data analytics since the incoming data can exhibit substantial variations. Consequently, the associated requirement costs may exceed the current company's budget in terms of DBA People, Process, and Technology Resources, such as: Hiring Big data specialists, Analytics training, data warehouses, and incremental cloud services consumption.

   Next, the transformation step possesses the capability to execute operations such as data extraction, cleaning, translation, merging, and data validation. For instance, structured data from an ERP system that must adapt to global/local standards or unstructured IoT device measurements from production lines that must be converted into a specific standardized format. The "digested"/ transformed data is then stored accordingly to specified criteria such as local Enterprise Data Management, to be validated against predefined data quality rules.

   Finally, the transformed data is loaded into target databases, such as a Hadoop cloud environment, for further analysis and processing. Data storage within this layer adheres to principles dictated by organization compliance regulations, data governance policies, and access controls. The methods employed for data storage can be implemented through batch processes or in real-time, depending on the requirements and constraints of the system.

2. **Data Ingestion and aggregation:** This phase processes and analyze the diverse data types. To achieve this process, for instance, this phase realize multiple data analysis depending on the data structure and the analysis goal. For instance, [61] proposed the following three components: Hadoop Map/Reduce, stream computing, and in-database analytics. The choice of component depends on the nature of the data and the objective of the analysis.

   (a) **Hadoop Map/Reduce** enables the cost-effective processing of large data volumes in batch form, facilitating the analysis of both structured and unstructured data within a massively parallel processing environment. Currently, this big data data analysis is considered the predominant programming model employed.

   (b) **Stream computing performs high-performance** processing of semi/ real-time data stream. Real-time analysis allows users to monitor data flows, respond promptly to unforeseen events, and determine optimal actions. For instance, in financial transaction fraud detection or alarming systems.

   (c) **Database analytics** refers to a data mining process that runs within different analytic systems that are integrated within the data warehouse. This analytical process allows a fast, scalable, parallel, and feature tailored optimization for big data analytics. Additionally, it provides a optimal and secure environment for confidential data. Nevertheless, the analytics results are not real-time, and its main outcomes are by nature static predictions. As mentioned previously, this analytics might be appropriate to the different Analytics Maturity path (Please refer to subsection: 4.3.5).

3. **Analytics Data Layer:** The third and final layer, produces outputs including various visualization reports, real-time information monitoring, and valuable business insights derived from the analytics layer, which are then delivered to users within the organization. Similar to conventional business intelligence platforms, reporting plays a crucial role in big data analytics by presenting data in a meaningful manner to support users' daily operations and aid managers in making faster, more data-driven decisions.

4.3.10 Opportunities and solutions

The objective of the Opportunities and Solutions phase (ADM-step E) is to assess and categorize the various levels of Big Data Analytics projects or initiatives in order to achieve the desired target architecture. Level one is defined as the baseline architecture, while level two represents the target architecture. Similarly, the relationship between level two and level three serves as the final target Deployment Reference Architecture. These different architectures are determined through aevaluation of gaps and the candidate architecture roadmap developed in Phases B, C, and D. At this point, various reference architecture layers have been established for common advanced analytical products. Considering the variations in business operational processes based on Data Analytics, BDA capabilities (Business-Infrastructure alignment, Seizing Reconfiguration, and Data Transformation), BDA resources (People, Process, and Technology), and the maturity assessment, the following steps will guide the architecture modeling process:

Level	**Data Analytics**	**BDA Capabilities**	**MA Evaluation**
1	Descriptive	Low levels of BDA capabilities	Low level on the Analytics MA assessment questions
2	Predictive	Medium levels of BDA capabilities	Medium level on the Analytics MA assessment questions
3	Prescriptive	High levels of BDA capabilities	High level on the Analytics MA assessment questions

Table 4.3: Big Data Analytics Projects Opportunity & Solution

Chapter 5

Design & Development- BDAC Deployment reference architecture

5.1 TOGAF ADM Methodology

5.1.1 Preliminary phase

Beer Industry

Beer is the oldest and most widely consumed alcoholic beverage globally, has been an integral aspect of human culture for multiple millennials and only third after water and tea consumption. Belonging to the Food and Beverage Industry, which is consider the 10th biggest industry in the world with a worth of $6,383 billion and employs over 10 million people worldwide. Specifically within this industry and the global Alcoholic Drinks market, beer is the most significant segment in terms of both volume and value.The projected revenue for 2023 is about $610 billion [56] (almost 9% of the total industry size). The global market is expected to growth annually by 10.34% (CAGR 2022-2025 together with a constant growth of 12% to 13% for the Non- Alcoholic Beer category. (see below Figure 5.1)

If divided, the four biggest global beer market regions are : Americas, Africa, Asia and Europe. From the 757.5 billion dollars total beer revenue market of 757.5 billion dollars from 2025, the following will be the biggest markets: First, Americas with 260 billions that represents almost 34% of the total market value, follow by Asia with 199 billion (26%), Asia with 239 billion (32%), Europe with 199 billion (26%) and Africa with 39 billion (6%). (Please refer to Figure 5.2).

It is projected that Americas will continue leading the Average revenue per capita and Asia will lead the total volume.(see Figure 5.3). Nonetheless, the markets will maintain their positions in the Average volume per capita with minimum growth (see Figure 5.5) and an overall significant increase of 24% in the price per unit(Please refer to Figure 5.4).

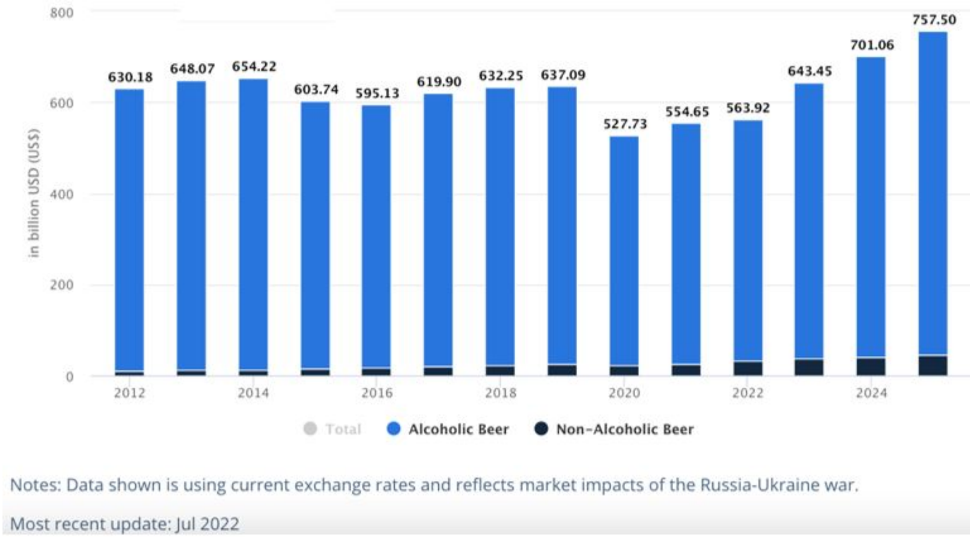
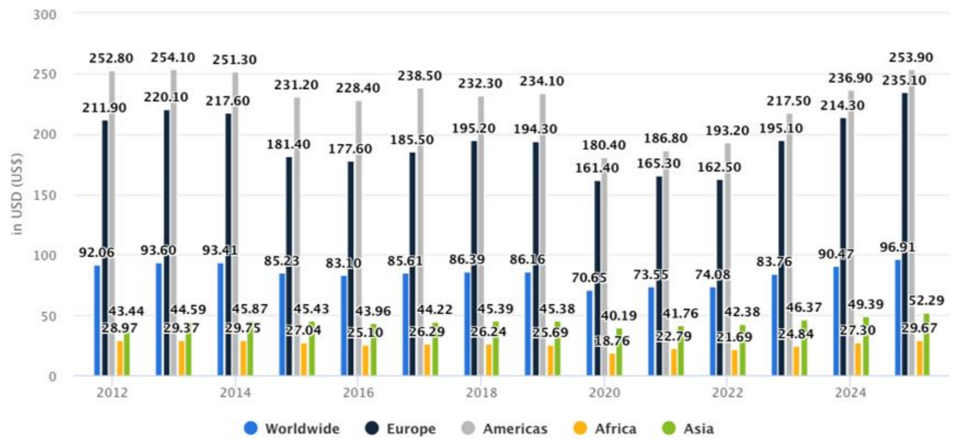


Figure 5.1: Worldwide Beer market Revenue[56]

Figure 5.2: Worldwide Beer market Revenue compare by four Markets[56]

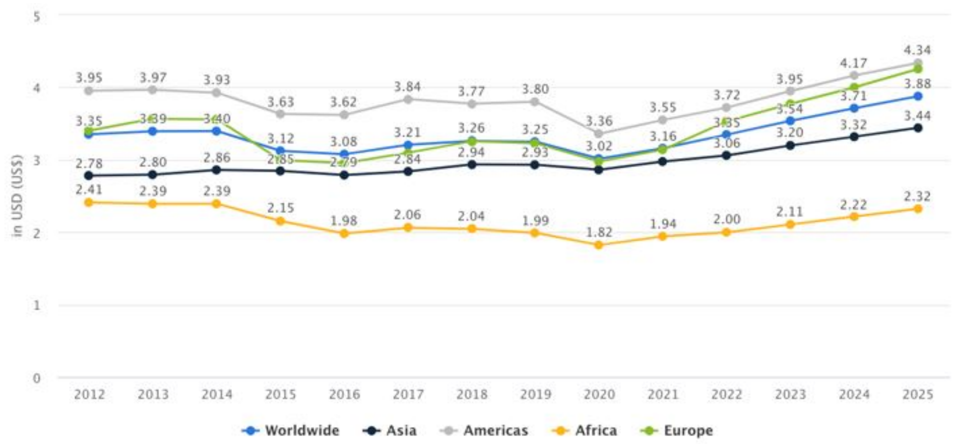
In terms of volume per Capita, the average growth is projected to be stable in the following years. However, Europe and Americas have started to present a decrease on their consumption from 2019 caused due Covid-19. However, both markets are projected to recover by 2025 with a constant growth 3% to 8%. Globally, Anheuser-Busch InBev, Heineken, China Resources, Molson Coors, and Carlsberg are the most important companies in term of volume. Just AB InBev represents more than 30% of the worldwide volume and altogether these five organizations share 60% of the global beer production [56], concentrate most of the brewery industry.

The market for alcoholic and non-alcoholic drinks is divided into two main segments:



Source: Statista

Figure 5.3: Average revenue per capita by four Markets[56]



Source: Statista

Figure 5.4: Average Price per unit by four Markets[56]

the retail market, also known as the off-trade market, which is focused on in-home consumption and includes outlets such as supermarkets, stores, and online stores, and the on-trade market, which caters to out-of-home consumption and includes establishments such as bars, restaurants, nightclubs, and hotels. The prices and product offerings (SKUs) of these markets can vary based on factors such as consumer demand, competition, revenue, and volume. By 2025, it is projected that the volume of out-of-home consumption will account for 33% of the overall volume and 52% of the market, a fundamental part of the market revenue segment. The global beer market presents an expected 5.44% (CAGR 2023-2027) and an average of 22.67 Liters of beer for 2023 [27]. Notably, some of the increase in revenue for multiple beer companies is explained by the price adjustments as the volume maintains stability; this trend is directly perceived in the average volume per capita increase of two or three percent. Nonetheless, the current inflation is caused by consequences of Covid-19, the price increase in raw material and energy prices Russian-Ukrainian war.

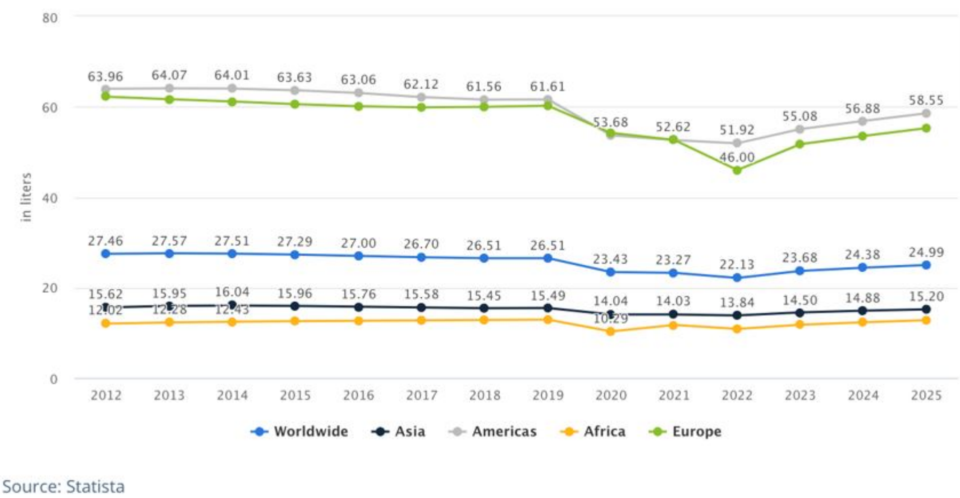


Figure 5.5: Average Volume per capita by four Markets[56]

Responding effectively and agile to these extraneous and local circumstances is increasingly becoming an invaluable competitive advantage for every company; as diverse and raging as these Operational challenges might affect "doing business" worldwide, the agility and proper use of data and Analytics capabilities and resources will play. For better or worse, technological advances are making it possible to produce a quantum leap in multiple industries. Analytics has transformed from an afterthought to a requirement in a digital age powered by data and automation. Data analytics systems are considered a crucial strategic investment for multiple organizations, as they have the potential to enhance firm performance significantly.

5.1.2 Architecture Vision phase

Company Description: Heineken Global

Heineken is a global beer company founded in 1864 by Gerard Adriaan Heineken and currently the second world-biggest beer company in term of volume[27]. It brews and distributes over 300 international, regional, local, and specialty beers and ciders that are available in 190 countries around the globe [27]. By 2021, the company employs approximately over 82,257 employees within its 165 breweries around the world [27]. In terms of its worldwide organization, the enterprise is divided into the following departments: D&T, Supply Chain, Procurement, Commerce, Finance, Audit, Corporate Affairs, Central Transformation Office, HR & Facilities, and Legal.

Globally, Heineken focus on growth, customer centricity, productivity, conscious culture, sustainability and responsibility, talent and capabilities, and becoming the best-connected brewer by digitising their business end-to-end [27]. For instance, some recent growth examples includes robusting its growth and position in its new and current markets by opening new breweries in Mexico in 2018 and Vietnam on 2022, acquisition of South Africa- Brewery Distell in 2023, the launch of Heineken Silver in 2022, Heineken 0.0 and the launch of AIDDA, the AI Data Driven Advisor to maximize customer value interaction. However, as discussed in the previous section, Heineken is encountering several challenges in the industry, including rising prices and declining organic volume growth [28]. These challenges require implementing actions such as cost reduction, creating new revenue streams to boost growth, and the development of an agile deployment capability.

5.1.3 Global Analytics

The Data and Technology department (D&T) at Heineken depends heavily on the Heineken Global Analytics (GA) team to enable the company to derive value from analytics across the entire value chain. The GA team is responsible for streamlining internal stakeholders' business processes and helping Heineken achieve its goal of becoming the most connected brewer. To achieve this objective, the GA team is committed to advancing the "Cost & Value" initiatives and the "Digital & Technology" pillars outlined in Heineken's EverGreen Strategy [27]. The GA team aims to make Heineken a data-driven organization that can create long-term incremental value globally by delivering Advanced Analytics solutions and developing company-wide capabilities. To this end, Global Analytics focuses on digitally transforming business functions (i.e., Commerce or Supply Chain), through the use of scalable analytics use cases, data pipelines harmonization, the establishment of Data & Analytics Foundations, implementation of Data Governance, and development of Management Information frameworks.

In order to process and execute its mission, the Global Analytics team have developed the **Operating and Governance model** to create and deliver scalable high value customer-centric advanced analytics products, through the following fours phases:

1. **Intake:** The business problems are reviewed with GA stakeholders to determine the solution's analytics scalability and value impact. If the assessment is beneficial, the initial stage of developing an analytics product will start.

2. **Plan:** The GA team strives to transform the initial ideas into a refined **product design** by engaging with end-users, comprehending their challenges and requirements, and creating an appropriate requirements design and corresponding Wireframe if necessary. In addition, the GA team collaborates with the Product Owner to project an adequate benefit estimation and potential unlock value.

3. **Prove:** Entails the Change Management and Communication prioritization by maintaining regular communication with key stakeholders regarding progress, sharing initial prototypes, developing programs to enhance skills, and other related activities. Additionally, usage and end-user satisfaction are measured through an established framework once the analytics products are deployed in the country. Outcome: a clear product design and understanding of how the solution can be validated.

4. **Produce:** The analytic product is Deploy locally within a particular business function (Commerce- Supply chain) team and final users, measuring its generated value and support the local countries with different activities such as training's, value

frameworks or support agreements. Outcome: A proven value advanced analytical product and is incorporated through the business processes in an interface (Power BI, analytical tools or B2B platform Back-end) so the final user can interact and take different advance analytical data driven decisions.

(a) **Data backbone validation (ETL):** Data transformation extraction, storage and transformation to be used in the advanced analytical models.

(b) **Local & Regional Pitch:** Analytical product presentation to the local business and technology team (D&T).

(c) **Deployment Process:** Represents all the different steps to be taken in order to effectively implement an advanced analytical products.

(d) **User training:** is the final users product training's and required skills preparation.

(e) **Review usage KPIs & feedback:** are the comments and feedback from the users that actively uses and interact with the platform in terms of model accuracy, front-end "easy to use" perception and further improvements (Always on, Quarterly).

(f) **Value measurement:** represent the measurement of the value created from the models suggestions A/B testing, Regressions, Difference in Difference, etc.

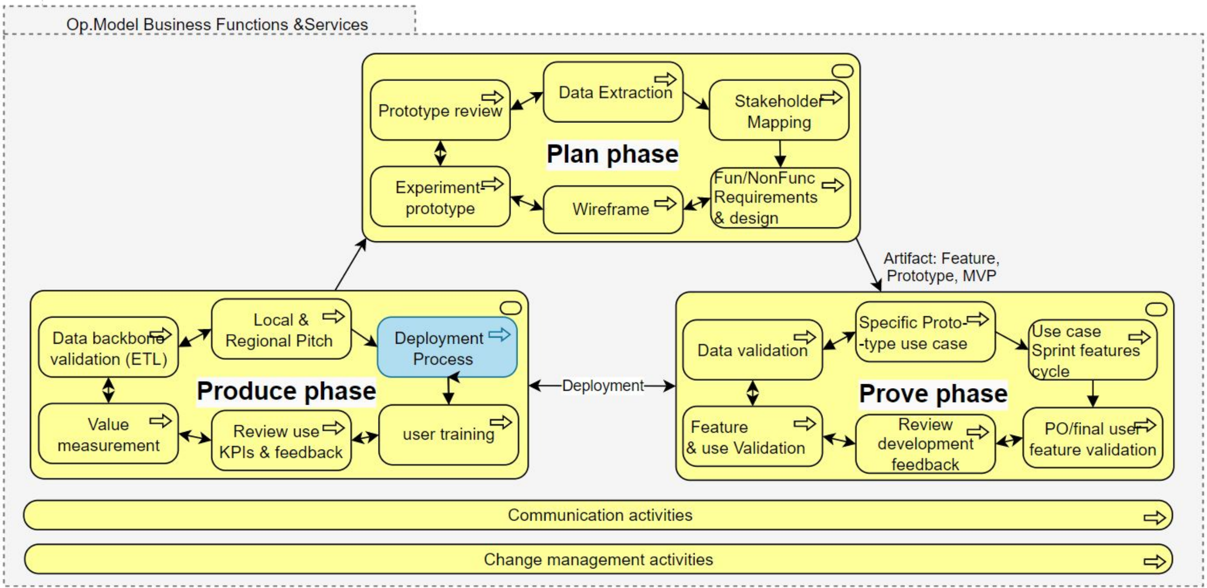


Figure 5.6: BDAC Deployment Reference Architecture Business Layer

As products are analytical products produce and validate in Plan and Prove phase, the next step is to deploy the products in the different being deploy in the different successfully this model, the following are the three interdisciplinary groups that are required:

1. **Global Analytics team:** The team organizationally resides under the umbrella of the Global D&T within Heineken Headquarters in Amsterdam, Netherlands. Their

main goal to to produce and deploy advanced analytical products around the world and its teams is integrated from most of the roles R:1,2,3,4, and zone A2, B1,B2,C, and partially D through data harmonization (Please refer to section 1).

2. **Regional Hubs:** Regional team that produced, deploy, and prepare the analytical maturity levels required in the advanced analytical products. The regional works with the same resources and process as the global team, in a smaller scale.

3. **Local Heineken (OpCo):** (Operating Companies) Are mainly represented by the local functional teams that possesses the product remand on the Business Information processes. The main roles from the business roles: R1.2, R1.3, R1.4 (Global, regional and local) and technical teams (R.2.3, R2.4, R2.5), and additionally in mature markets (R.3.1).

These three groups of interdisciplinary internal and external stakeholders are required in the definition of what business process information will be improve thought the analytical product, which process will need to be reconfigured based on the local resources and the data extraction, transformation and storage that will be required to meet the final user requirements. (Please refer to Figure: 5.1.7). At the center of the operating model, the Analytics translator (R1.1) are in charge to lead the product development and stakeholders interaction, and its business counterpart Product Owners (R1.2) extract and group the business analytics process requirements and define advanced analytical product features. Both of these roles share two fundamental goals: A) Produce an advanced anayltical product that meet the final users product requirements and incremental value (**Produce**), and B) Successfully to deploy in the different local countries through the Produce (**Deploy**) (Please review the Prove and Produce phases in Figure 5.6.

5.1.4 Stakeholders Deployment challenges

Deploying an advanced analytical product is a complicated process that requires the involvement of various interdisciplinary roles in different locations and established procedures. Moreover, the maturity level of local resources (i.e., people, processes, and technology) and capabilities differs in different organizational local context, meaning multiple deployment challenges. As the current research developed within the Heineken Global Analytics team, a set of seven interviews with the different business roles (1) and R3, which includes interviews with (2) product owners, (3) Translators, (1) Harmonization Engagement analyst and (1) Data Harmonization Engineer. From these interviews, a group of insights well catalogued by role, using the Global Analytics Operational model phase, resource type and challenge details (Please refer to tables 5.1 and 5.2). After further analysis, the following represent the main deployment challenges:

- **Business and infrastructure integration:** Aligning the business- IS infrastructure by selecting and optimally leverage business process information required in the local context to the analytical product deployment to create value. In other words, define the proper integration of business process information.

- **Process and resources reconfiguration:** Adequately select and restructure local process that matches local resources and target required performance gaps. As such, defining clear ways reconfigure critical capabilities,process/resources to deploy optimally analytical products.

- **Data availability:** Find efficient ways to support the internal and external data quality, volume, value, variety, velocity, and veracity required for the multiple analytical products.

Stakeholder,Operational Model Phase, Resource type	**Challenge**
Translator, Intake, People	I'll prioritisation of needs & inadequate definition of unique value proposition
Translator, Produce, People	Complex organisational structure (local WoW autonomous)
Translator, Produce, Technology	Siloed data
Translator, Produce, Process	Unclear roles & Responsibilities, things fall through the gaps
Translator, Produce, Process	Changes in data dependencies
Translator, Produce, Technology	Limited data access (makes designing hard)
Translator, Produce, Technology	Low Data Quality
Translator, Produce, People	Limited Engagement Product owners /SME's
Translator, Produce, Process	OpCo (Local) low Engagement
Translator, Produce, People	clear roles boundaries (Translator-PO)
Translator, Produce, People	Product not easy to use - difficult to scale
Translator, Produce, Process	The global Harmonization Data Pipeline is an local enabler/ blocker
Translator, Produce, Process	Change Management (Openness to new process Commerce vs Supply functions)
PO,-, Process	Difficult to project value*, how to prove it, projection are based on the past and validated in the future
PO,-, People	Change Management(Trust) in terms of starting communication, guiding and supporting change
PO,-, Process	Unclear road-map in advanced
PO,-, People	Governance complexity (Deployment requires top-down/ down-top approach)
PO,-, People	Use of internal or external resources (speed, trust, communication and expertise)
PO,- ,Process	The Analytical products might affect current local team economic incentives (Sales incentives)

Table 5.1: Translator and PO Resources Challenges Operational Model phases

Role	Resource	Challenge
R2.1	People	PO and Translator support on local engagement
R2.1.	People	Limited communication with local technical teams
R2.1	Technology	complex and manual local data transformations
R2.1	Process	Analytical products might affect the local incentives
R2.1	Process	Change Management (historic business data rules- i.e., finance)
R2.1	Process	Limited technical knowledge (Consultants)
R2.1	Process	Manual business data rules, transformations and data models
R2.1	Process	Manual business data rules
R2.1	Technology	Low Data Quality

Table 5.2: Data Challenges Operational Model

5.1.5 SWOT analysis

Based on the current context of the deployment of advanced analytics products within the global analytics products and interviews from the previous sections the following Deployment SWOT analytics have been developed (Please refer to figure 5.7) for which the following are the three main analysis conclusions:

- **Weakness (GA-Internal):** Given the immense amount of local systems, IT landscapes, providers, business strategy/ way of working (WoW) is it challenging to the global team to effectively assess, align and leverage business process information to the local resources to create business value (i.e., Data silos, limited data sources, low data quality, complex IT landscape). In other words, identifying efficiently what to change given the ability to leverage current business processes.

- **Threats (Local context-External):** As business processes information are selected to change in order to increment is value creation, the process to change the local resources and capabilities are not clear. For instance, there are currently systems that partially grant the analytical product features, the local resources only requires one/ a few product features or the local team already posses local business strategies that might be affected by analytical product. For which, it is recommended to clearly define ways on how the analytical product might require process changes or reconfigurations through a deployment road-map that portrait the interdependence's blocks in different local context to match the current local resources and target required performance gaps.

- **Opportunity (External):** The main Deployment opportunity is to leverage from one side the multiple global umbrella of projects such as Core, HDP, EDM; from the other side, is to orchestrate the current regional projects (Hubs), the different regional and local resources/ experiences (A.I. business cases, Ml expertise, people talent) to efficiently access, transform and use data to deploy advanced analytical products.

- **Strengths:** The main strengths of the Global Analytical team is the success and value measure of multiple analytical products, which have open the eyes from the C-suite leadership and local team, creating a positive demand to acquire and create incremental value through their products.

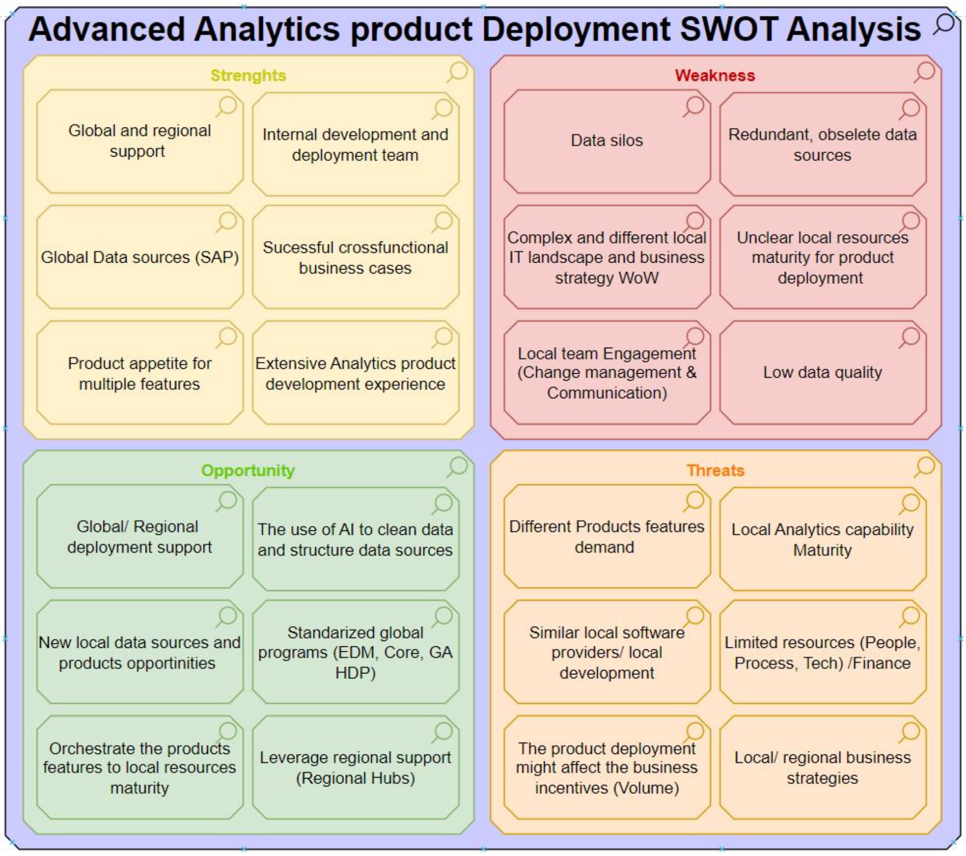


Figure 5.7: Advance Analytics product deployment SWOT analysis

5.1.6 Multi-Criteria and Model-Based Project Selection- BDA capabilities dimensions

The following are the Multi-Criteria and Model-Based Project Selection Method steps applied in the Global Analytics context, as proposed in the previous chapter, subsection 4.3.6:

1. **Pre-requisite- Determined the specific concern or business problem:** validates the Big Data project initiatives implementation problems, as the multiple challenges gathered through the stakeholders interviews in subsection: 5.1.4.

2. **Pre-requisite- Determined possible projects:** assess the adequate big data analytical project depending on the business goal and Analytics Maturity path (e.g., Diagnostic, Predictive, Prescriptive) (Please refer to subsection: 4.3.5). In the Global Analytics context, the analytical products present multiple Analytics Maturity paths.

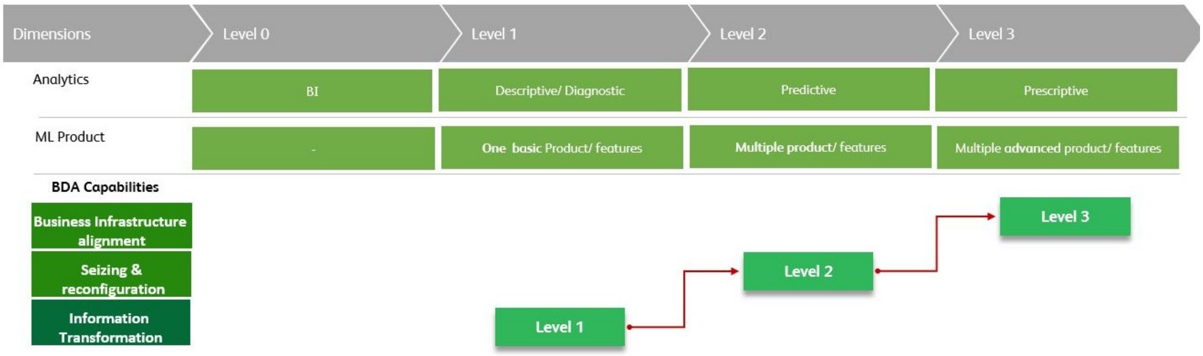


Figure 5.8: Multi-Criteria and Model-Based Project Selection- BDA capabilities dimensions

3. **Capability Analysis-Strategic Capability Analysis:** validates the BDA Capabilities levels (Section: 4.3.6) required for the Big Data Analytics projects.

4. **Capability Analysis-Capability Metrics Analysis:** This step aims to define the criteria to evaluate the selected BDA capabilities will be mapped and evaluated using the current Global Analytics Maturity assessment (Please refer to Appendix: A.2.1). To evaluate this process, it will be required to map the current BDA capabilities dimensions to the Analytics Maturity assessment questions and use them as an architecture guiding architecture patterns. To assess this process, the following three Global Analytics Maturity assessment questions were map to the BDA capabilities (Please refer to the table: 5.3, and Figure: 5.9):

Maturity Assessment	**DBA Resource, Question**	**BDA Capabilities**
GA Analytics Maturity Assessment	People, How data-driven are your decisions?	Business- Infrastructure Technology
GA Analytics Maturity Assessment	Process, Where does the majority of the data that you use in your daily tasks comes from?	Seizing and reconfiguration
GA Analytics Maturity Assessment	Technology, How do you connect two (or more) different data sources?	Infrastructure Flexibility

Table 5.3: Capability Metrics Analysis-Global Analytics Maturity assessment

5. **Capability Analysis- Project Impact Analysis:** select projects that can contribute to the improvement of a capability based on the metrics identified in the previous step. This step will be demonstrated in the TOGAF- ADM- phase E: Opportunities and solutions (Please refer to section: 5.2).

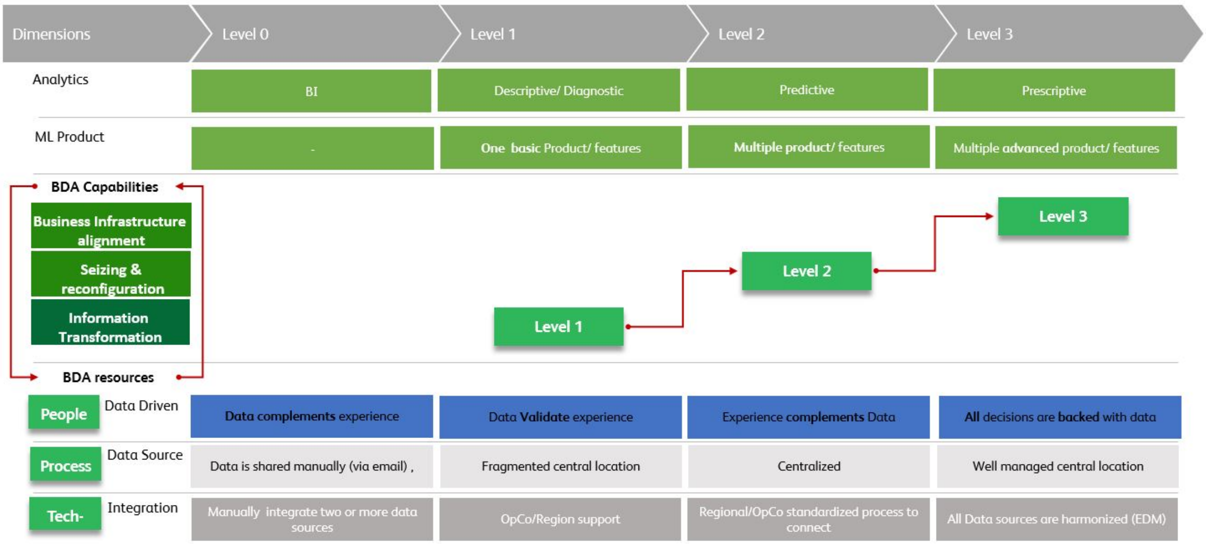


Figure 5.9: Global Analytics- Integrated BD Analytics Maturity assessment and capabilities dimensions

5.1.7 Business Architecture phase

Function Business Analytics Process

In the case of analytical products within Heineken, the process starts within a Country Operating Company- OpCo (i.e., Brazil, Rwanda, Mexico) and a department/ function (Commerce, Procurement, Supply Chain) that within their day to day or periodically operations realized analytical process and take in more or less degree data driven decision to optimize the function Business Process.

Its within these business process that the Business Analytics process starts (Please refer to Figure: 5.10) and continuous within the following process:

1. **Information systems:** represents the actual digital system (i.e., ERP, CRM, SAP modules) that serves as a front-end to the End-user (R1.4) and digitally creates the business data process generated (Sales orders, customer, Financial, Marketing costs).
2. **Data Extraction:** is the extraction of data from homogeneous or heterogeneous sources.
3. **Data Transformation:** processes the extracted data by cleaning and transforming it to a proper storage, structure or format for the analysis or querying purposes.
4. **Data Analysis:** is process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making.
5. **Advanced Analytical Product:** refers to the range of sophisticated techniques and tools to analyze and interpret complex data sets, patterns, predictions or in-

sights to make data-driven decisions. It commonly involve the application of statistical, mathematical and machine learning algorithms to extract and discover insights from the sourced data (For the different Analytics Maturity path type, please refer to Section: 4.3.5).

6. **User interaction systems:** provides the interface between the user and the system (i.e., Power BI, B2B platforms).

7. **Business Decision:** Refers to the business decisions that are meant to be leverage by the advanced analytics products insights as the outcome that is presented through the User iteration systems.

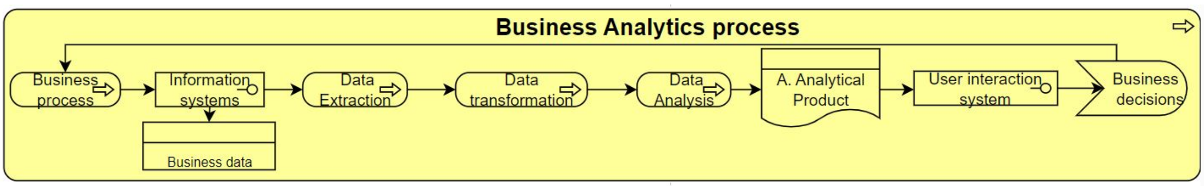


Figure 5.10: Function Business Analytics Process

Integrated Local, Regional and Global Analytics business process Deployment view

The next step in the deployment process is to place the Analytics products on top of the Analytics business processes mentioned in the previous subsection, by integrating the local functional team and the Global/Regional teams in a common integrated Business Advanced analytical Deployment process. (Refer to Figure: 5.10). As integration business process, the local team is responsible for the data extraction from the different data sources and the utilization of the analytical outcome. Next, the Global/ regional Team extract the operational data, standardized ("harmonized"), and analyzed it through the different Data science techniques, to finally push the data outcome through the different User Interaction systems as follows:

1. **Local Business Process:** are the functions business process that generate, process and stores the metadata and transactional data.

2. **Data Engineering:** presents the process of (Extract, Transform, and Load) ETL process and the standardized data integration within a common Enterprise Data Model (EDM).

3. **Analytical Product:** represents the different advanced data analysis methodologies and models that are use to analyzed and take actions in the different business processes.

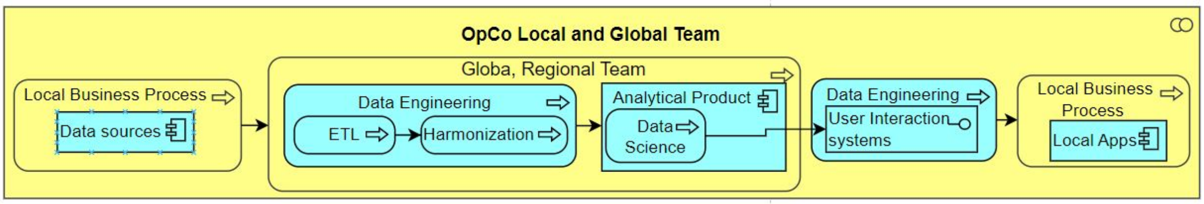


Figure 5.11: Integrated business Analytics process

BDA Capabilities Deployment reference architecture roles

After describing the Global Analytics Operating Model and their resulting steps, the roles list will be presented by the main stakeholders groups that together are necessary roles in order to realize DBA Deployment Capability reference architecture. MLOps is an interdisciplinary group process, and the interplay of different roles is crucial to design, manage, automate, and operate an ML system in production and finally impact the final user operations and value creation. In the following list, every role, its purpose, and related tasks are briefly described, together with their correspondent roles in the Integrated business analysis process (Please refer to the Figure: 5.12):

1. **Business Roles (R1)**

   - **R1.1 Translator:** orchestrate the deployment of high quality, customer-centric advanced analytics products by building close stakeholder relationships (end users, PO, Technology Teams) and leveraging our broad skill set that ranges from product management, change management to communication and upskilling.
   - **R1.2 Product Owner:** (Project Manager) is responsible for setting the Machine Learning (ML) project objectives and managing the communication aspects of the business(i.e., Churn initiative or return on investment (ROI)) generated by the ML product.
   - **R1.3 Function Stakeholder:** Local/regional/global function management role that directly or indirectly will benefit from the analytical product. Generally, for the product deployment this person is the local business function that would serve as a point of contact.
   - **R1.4 End-User:** An individual or an organization that utilizes the Analytical products or services.

2. **Data roles(R2)**

   - **R2.1 Data Harmonization Engagement analyst:** is technical analytics professional specialized in engaging, assessing and guiding the local data IT team, to Extract and Load the function transactional and meta data.
   - **R2.2 Data Harmonization Engineer:** is responsible for setting the Machine Learning (ML) project objectives and managing the communication aspects of the business(i.e., Churn initiative or return on investment (ROI)) generated by the ML product.

- **R2.3 Data Engineer:** involves constructing and supervising pipelines for data and feature engineering. Additionally, this position guarantees the correct ingestion of data into the databases of the feature store system.
- **R2.5 Data Owner:**Responsible for responsible metadata and transactional data storage and/or the quality of a defined data set on day-to-day basis.

3. **Machine Learning (ML) roles (R3)**

- **R3.1 Data Scientist:** responsible for converting the business problem into a machine learning (ML) problem and managing the model development process, which involves selecting the most effective algorithm and hyper-parameters.
- **R3.2 DevOps Engineer:**connects development and operations team to guarantee efficient automation of the CI/CD process, orchestration of the ML workflow, deployment of the model to the production environment, and monitoring of the entire ML system.
- **R3.3 ML Engineer/MLOps Engineer.** Local function management that directly or indirectly will benefit from the analytical product. Generally, this person is the local business function point of contact.

4. **Cross Functional roles (R4)**

- **R4.1 Solution Architect:** create and defines the appropriate technologies to be utilized after a comprehensive assessment.
- **R4.2 UX design/Front end developer:** ensures that products, services, and technology are user-friendly, engaging, and available to all the stakeholders and final users.

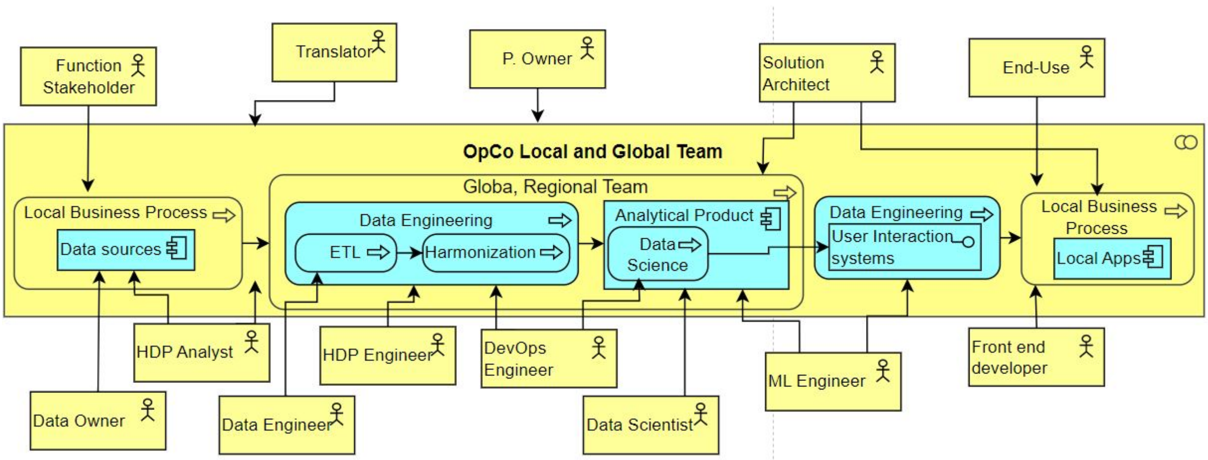


Figure 5.12: Integrated business Analytics process and roles

5.1.8 Project Initiation

The following process represent the Global Analytics Project Initiation for this layer:

- The Function stakeholder (R1.3), Translator (R1.1), and Product Onwer (R1.2) conducts a Function and Product analysis and identifies a potential business challenge and use cases that can be addressed using machine learning (ML). This process involve multiple cross-functional/ cross project initiatives communication and change management that given the research time limitation, model complexity and research goal will not be elaborated, and instead be presented as a future work.
- The solution architect (R4.1) formulates the architectural design for the ML system as a whole and makes informed decisions about the technologies to be employed following a comprehensive evaluation, including the roles and cross functional responsible team/ projects involve.
- The solution architect (R4.1) formulates the architectural design for the ML system as a whole and makes informed decisions about the technologies to be employed following a comprehensive evaluation. Additionally, different Data Roles (R2) will prepare the different data sources, ETL and Harmonziation process in preparation to the data features required for the Analytical product.
- The data scientist (R3.1) formulates an ML problem based on the business objective, considering factors such as whether regression, classification or any other technique that will be used—from the business goal to solve the End-user (R1.4).
- The data engineer (R2.3) and the data scientist (R3.1) collaborate to gain a mutual understanding of the necessary data extraction, transformation, storage and loading to solve the business problem.
- Once the requirements are clarified, the data engineer (R2.3) and the data scientist (R3.1) work together to identify the original data sources for initial data analysis.
- They examine the distribution and quality of the data and perform validation checks, ensuring that the incoming data from the sources is labeled, meaning that a target attribute is known, which is essential for supervised ML. In this particular example, the data sources already had labeled data available as the labeling step was completed during a previous process.
- Finally, on the Outcome-end zone, the model output is transferred to the User Interaction systems (i.e., Power BI/ B2B platform) for the use on the Final user (R1.4) business operation and feedback and value measurement.

5.1.9 Information Systems Architecture phase

Data Engineering Zone

1. **Requirements for feature engineering pipeline:** involve the identification of the essential attributes necessary for model training and model features. Once a

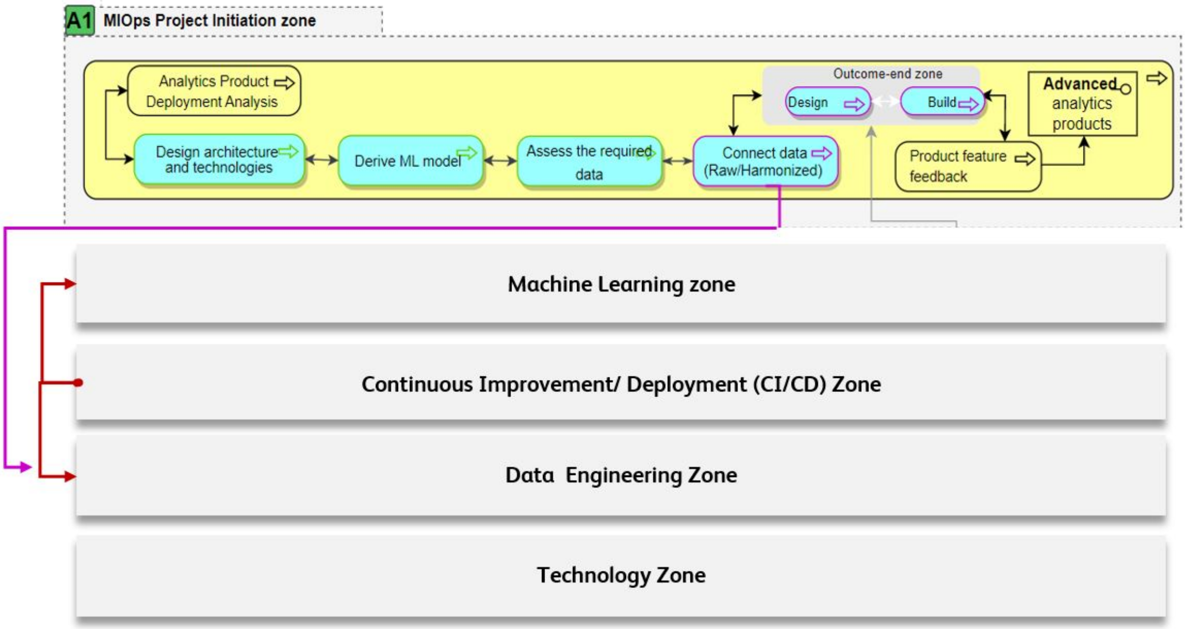


Figure 5.13: Project Initiation Pillars

preliminary understanding of the raw data and initial data analysis is obtained, the following fundamental requirements for the feature engineering pipeline are established to extract, transform, store and load the required model data/ metadata (Please refer to the figure 5.14- Zone C):

(a) The Translator (R1.1) orchestrate a join meeting to define the required data ETL and harmonization process for the Analytical process (i.e., Service Level Agreements). Based on the business data sources and Analytical products two different group are required: Global/regional (Solution Architect (R4.1), Data Harmonization Engagement analyst (R2.1), Product Owner (R1.2), Function Stakeholder (R1.3) and Local teams roles: Function Stakeholder(R1.3), Data Engineer (R2.3), and Data Owner (R2.5).

(b) The data engineer (R2.3) defines rules for data transformation (e.g., normalization, aggregations) and cleaning to ensure data usability, including the data quality and corresponding Enterprise Data Management (EDM) compliance. These process include both Global and Local data sources

(c) The data scientist (R3.3) and data engineer (R2.3) collaborate to define feature engineering rules, including the creation of new and advanced features based on existing ones and the product feature required from the Product Owner (R1.2).

(d) The initial rules will required to be iteratively adjusted by the data scientist (R3.1), either based on feedback from the experimental model engineering stage or from the monitoring component that assesses the model performance. Particularly, in this view this method is semi- automatic.

2. **Feature engineering pipeline:** represents the data feature pipelines that: A)

build between the data sources or Global systems and the local feature store that will feed the advanced analytical model, and B) connects the extract model outcome to the User interface systems. The following are the most relevant steps on this process (Refer to Figure: 5.14- zone B1 Section-C):

(a) The data engineer (R2.3) and Data engineer (R2.3) utilize the initially specified requirements as a foundation to develop a prototype of requirements and rules are continuously updated based on feedback received from the experimental model engineering stage or the monitoring component that assesses the model's performance in production.

(b) Parallel to the previous process, the data engineer (R4) defines the code required for the CI/CD and orchestration component (Refer to zone B2) to ensure the task orchestration of the feature engineering pipeline. This role also defines the pipeline underlying infrastructure resource configuration based on the local or Global data source.

(c) In the pipeline, the initial step involves the establishing a connection with the raw data, which can be streaming data, static batch data, or data stored in any cloud storage. In particular, the data might come from Global sources (i.e., SAP HANA, Oracle ERP, Nielsen) or local internal systems (ERP, CRM, HR) or local external systems (Route planing, Social Media, Weather, World Bank). Additionally, the data pipeline will be push back to the corresponding local system. The data extraction from the multiple data sources will be presented in the Technology phase (subsection: 5.1.10).

(d) The data is extracted from the data sources (Global Extraction systems, the local sources or the external providers).

(e) The data preprossessing begins with data transformation and cleaning tasks, to later on be Harmonized to comply with the EDM requirements and model features.The transformation rule artifact defined in the requirement gathering stage serves as input for this task, and the main aim of this task is to bring the data into a usable format. These transformation rules are continuously improved based on the feedback and the Analytical model features and local BDA capabilities and resources maturity.

Machine Learning Zone

This zone represents the features of the advanced analytics products and its multiple features depending on the local needs. The ML zone is triggered by the acquisition of Harmonized data gathered in the previous section, which can encompass numerical values, images, or textual information (i.e., financial transactions, SKU items, maintenance records, time series data from sensors, the brewery, or sales reports). This data is analyzed, collected, and prepared for utilization as training data, serving as the input for training the machine learning model. The quantity/quality of data plays a pivotal role in enhancing the effectiveness of the model and is critical for the quality of the ML product feature. In the following steps, programmers select a suitable machine learning model,

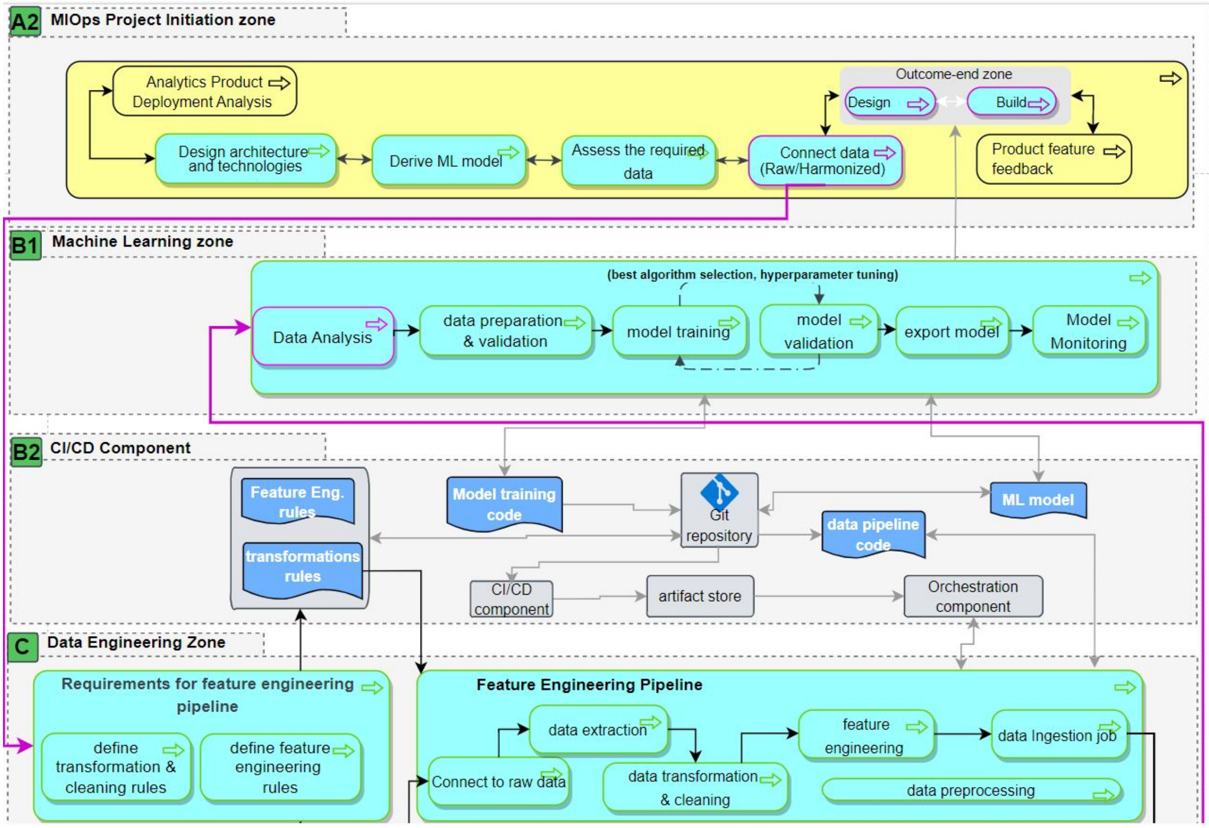


Figure 5.14: Global Analytics BDA Capability Reference Architecture System Layer

input the prepared data, and enable the computer model to undergo self-training, identifying patterns and making predictions. Over time, human programmers have the ability to refine the model, including adjusting its parameters, to facilitate its progression towards achieving more precise outcomes and close to day-to-day business operations complexities. The following are the most important steps by steps on this zone (Please refer to the Figure 5.14- zone B1):

1. The data scientist (R3.1) connects to the feature store system for the data analysis. The accessed data for this process requires to be harmonized accordingly to the entity's definitions and EDM data compliance. The R3.1 cannot connect to the raw data for an initial analysis unless the data is manually harmonized, depending on the local BDA capabilities and Resources. However, the current architecture presented the different data pipelines in the previous subsection, which permits the data scientist (R3.1) to report the required changes back to the data engineering zone (in a semi-automatic feedback loop).

2. The incoming data is repared and validate from the feature store system (Both the training and testing data set creation and splitting).

3. The best-performing algorithm and hyperparameters, and the model training is then triggered using the training data.

4. Multiple iterations of model training involve the interactive testing and validation

of various model parameters. The process continues until the performance metrics (accuracy) indicate satisfactory results, at which the iterative training ceases. To achieve this, parameter tuning is conducted to identify the most effective model parameters. The tasks of model training and model validation are subsequently repeated iteratively, and referred to as "model engineering." The objective of model engineering is to determine the optimal algorithm and hyperparameters that yield the highest performance for the model.

5. The Data Scientist (R3.1) exports the model and push the code to the repository. As a fundamental prerequisite, either the DevOps engineer (R3.2) or the ML engineer (R3.3) formulates the code for the ML workflow pipeline (C) and adds it to the repository. Once the (R3.1) commits a new ML model or the DevOps engineer (R3.2) and the ML engineer (R3.3) add new code for the ML workflow pipeline to the repository.

6. Finally, the export model data outcome is connected to the particular local context through an "ML outcome pipeline" to the "Outcome-end-zone", in zone A2 that represents the "user interaction system".

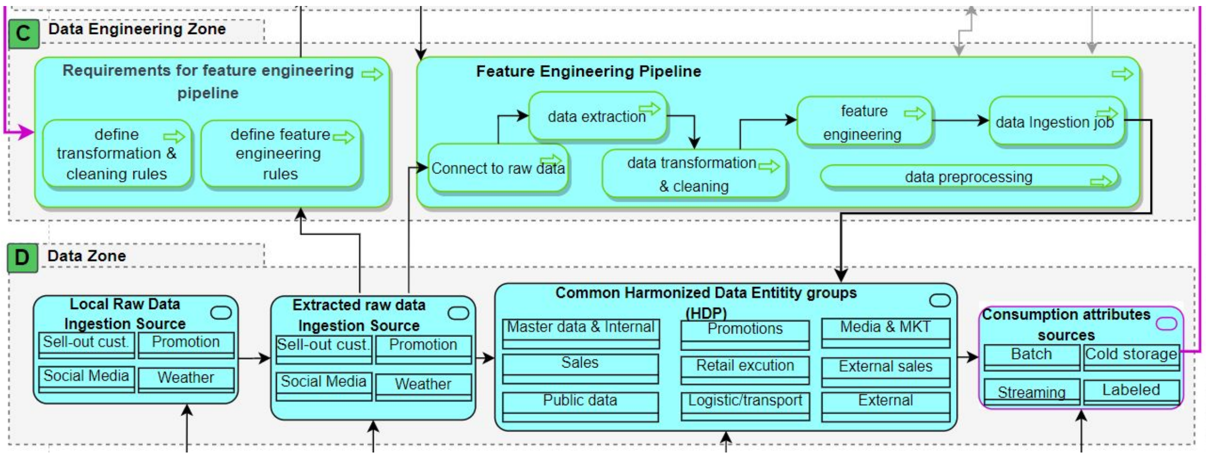


Figure 5.15: BDAC Deployment Reference Architecture Data Layer

5.1.10 Technology Architecture phase

On this architecture layer, the focus will be on the core data sources, processes, and IT landscapes that represent some of most relevant technology required to produce and deploy advanced analytics products. Typically, data is generated and stored across various IT systems like Enterprise Resource Planning (ERP), production/ manufacturing systems, or Customer Relationship Management (CRM). While this approach is suitable for operational processes, it can become a challenge when analytical products require utilizing the data for different purposes. For instance, a significant challenge arises when combining data from various IT systems, such as merging route-to-customer data with ERP and CRM data or marketing initiatives, or when aiming to leverage the data for analysis and reports. To address some of these interoperability issues, multiple data platforms and technology processes offer an enhanced solution by enabling access, integration, storage, analysis, and data reporting from diverse data sources.

The primary objective of a data platform is to gather, store, transform, and analyze data, providing accessibility to business users or other systems. It serves as a foundation for business intelligence, advanced analytics (including Machine Learning), or as a central data hub.This platform consists of distinct components organized into various layers, each serving a specific function. These layers encompass Data Sources, Integration Layer, Processing Layer, Storage Layer, Analytics Layer, Visualization Layer for the different Heineken's local and global technology systems. The following are the three technology process representations of the most relevant data sources and technology process for the Advanced Analytical Product Deployment (Please refer to Figure: 5.17- zone E and D, for Data representation).

1. **Local Data Sources:** involve the multiple data sources and systems that are daily present in the business operations(i.e., ERP, CRM, sales orders). These data sources represents the raw data sources that later, after the ETL and data harmonization process will become in the the attributes necessary in model training and model features (Zone B1), and five main types are of data are identified: Streaming (IoT, Big Data streams), Unstructured (images, videos), Semi-structured (logs, json, xml. logs), structured (Relational data base) and Cloud Data services (Azure data services; Cosmos DB, Data verse, Azure SQL Database). The following are some example of these local data sources:

   (a) Enterprise Resource Planning (ERP) serves as the core system for operational companies, encompassing and supporting various essential processes. It acts as a centralized platform enabling various operations, such as registering an Stock Keeping Unit(SKU) and specifying the necessary materials and processes required to manufacture a final product (including brewing, packaging, and logistics). Additionally, ERP facilitates the seamless flow of sales orders, tracking their progress from order placement to invoice generation and shipment while also providing comprehensive stock-level management capabilities. Essentially, ERP acts as the backbone of an organization, ensuring smooth and efficient operations across multiple functions. Some examples are (SAP B1, JDE, Oracle Fusion, or internal corporate ERP).

   (b) Local promotional systems, functional systems (i.e., Marketing, logistics, finance).

   (c) As the next step. the different ingestions systems are presented to ingest the data throughout different sources technology: (Azure Events/IoT Hubs, Azure synapse, Azure Databricks (Scheduled/ event- triggered data ingestion pipelines, and Pipelines orchestrators). Nonetheless, multiple ingestion technologies are present depending on the local BDA capabilities and resources, such as (i.e., DB file objects, Mesh Broker, API, JSON).

2. **Analytics Enablement Platforms (AEP):** represents the extraction of the raw data from global systems such as Global ERP. Generally speaking, data is pushed by the source to the landing zone into a cloud storage service. From there, the data is ingested to a logical storage in the Data Lake, to later on be incorporated, transformed and orchestrated to be consume by Operational Companies (OpCo's, usually countries), Global Functions (like Global Procurement or Global HR), and Harmonized Data Pipelines (HDP). An example of technology that could be use in this platform are common cloud services use are Data Lake, Delta Lake, Synapse server-less. (Please refer to Figure: 5.17-Section-E and D):

   (a) The data pre-prossessing begins with data transformation and cleaning tasks, to later on be Harmonized to comply with the EDM requirements and model features.The transformation rule artifact defined in the requirement gathering stage serves as input for this task, and the main aim of this task is to bring the data into a usable format. These transformation rules are continuously improved based on the feedback and the Analytical model features and local BDA capabilities and resources maturity.

3. **Harmonizing Data Pipelines (HDP):** is a standardized data integration solution designed to establish an automated data flow for every HEINEKEN Operating Company (OpCo). Its objective is to seamlessly merge global and local, internal and external data sources. By doing so, HDP facilitates efficient access to pertinent, trustworthy, and well-organized data. The main different with (AEP) is that HDP create digested data pipelines with Global and Local systems, particularly on transactional data, instead AEP only Extract and orchestrate raw data. To achieve this goal, HDP solution uses a framework Data Build Tool (dbt) to support the creation and maintenance of the data warehouse pipelines.dbt supports the transformation step of a data pipeline. DBT consists of several core components such as models, seeds, sources and macros. Models are the building blocks of the dbt ETL pipeline through four main layers: Data Source Layer, Staging Source Layer, Core Layer and Advanced Analytics Product Layer. The principal outcome are the specific data marts that represent the Entities and attributes required in the features for the different Advanced Analytics Products (i.e., Sales , Retail Media marketing, Logistics, Public data, promotions) (Please refer to Figure 5.17, Zone D- Common Harmonized Data Entity groups).

5.2 Opportunities and solutions

The Opportunities and solutions (ADM-step E) has the goal to assess the different project or initiatives are categorized into work packages in order to achieve the desired architecture. The different architecture are determined through a comprehensive evaluation of gaps and the candidate architecture road-map developed in the preceding Phases B, C, and D. So far, the different reference architecture layers for the common advanced analytical products. Nonetheless, the business operational process, BDA capabilities(Business-Infrastructure alignment, Seizing & Reconfiguration, and Data Transformation), and re-

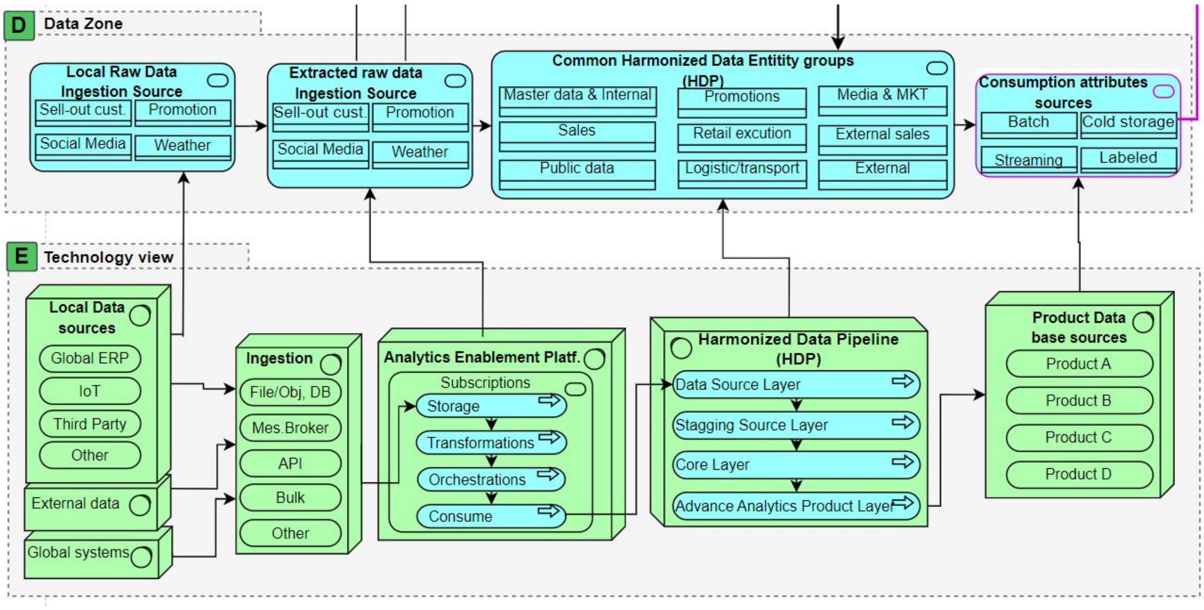


Figure 5.16: BDAC Deployment Reference Architecture Technology Layer

sources (People, Process, and Technology) varies for each of the different Operational Countries.

5.2.1 BDA Capability Deployment Architecture levels

The central Heineken- EverGreen goal of "becoming the most interconnected brewery" requires the interoperability across multiple platforms and IT landscapes. However, its precise implications in the day to day operations and local data and Analytics maturity context is not particularly precise. While this objective is commendable and holds considerable value, achieving it within Heineken organization is a significant challenge ahead.

Traditionally, Heineken has pursued an audacious approach when entering new markets, typically accomplished through acquiring a controlling interest in existing companies. Consequently, it is common to observe operating companies and brands bearing names distinct from Heineken or incorporating a combination of Heineken and local brands. In essence, each of these subsidiaries functions as an autonomous operating company. Heineken's prevailing strategy of gaining market entry through acquisitions rather than organic establishment shapes the operational dynamics of these diverse entities. Throughout these "soft-landing" approach in local markets, Heineken leverages the established operational practices of the acquired breweries within their respective regions. Nonetheless, as this approach possesses its benefits, enabling rapid profitability, and rapid ROI, it also perpetuates the persistence of multiple challenges in terms of systems interoperability, data quality, and legacy operating models.

In the numerous operating companies, multiple approaches are made, such as: assimilating their systems and databases, possesses their unique (ERP) systems implementation or a hybrid local/ local version. Nonetheless, each operating company is independent in selecting independent IT systems. Given the current total amount of ERPs, applications, and IT landscapes, it is difficult for the Global Analytics teams to assess BDA capabilities levels on the use of the BD resources. For this reason, the Global Analytics team has designed a Analytics Maturity assessment to measure the local BDA capabilities use levels based on the BDA resources maturity levels (please refer to appendix:A.2.1. By integrating the GA Maturity assessment within Integrated Multi-Criteria and Model-Based Project Selection- BDA capabilities dimensions (please refer to subsection 4.3.6), is it possible to define the following BDA Capability Deployment Architecture levels represented in the table: 7.1. In the following subsections, the three BDAC deployment reference architecture levels will represent the corresponding Big Data Analytics Capabilities dimensions architecture viewpoints: The business process integration is represented in figure 5.17, the seizing and reconfiguration in the zones B1, B2, and C of each architecture level, and infrastructure Flexibility in zone D and E correspondingly.

BDA Capability	**GA Maturity Assessment Question**	**Level 1 Descriptive**	**Level 2 Predictive**	**Level 3 Prescriptive**
Business Infrastructure alignment (What to change?)	To what extend are you data driven?	Data Validate experience	Experience complements Data	All decisions are backed with data
Seizing and reconfiguration	Manually integrate two or more data sources (How to change?)	Manually integrate two or more data source	Regional/local standardized processes	All Data sources are harmonized
Infrastructure Flexibility	How you integrate two or more data sources?	(email), Fragmented central location	Centralized	Well managed central location

Table 5.4: Global Analytics BDA Capability Deployment Architecture levels

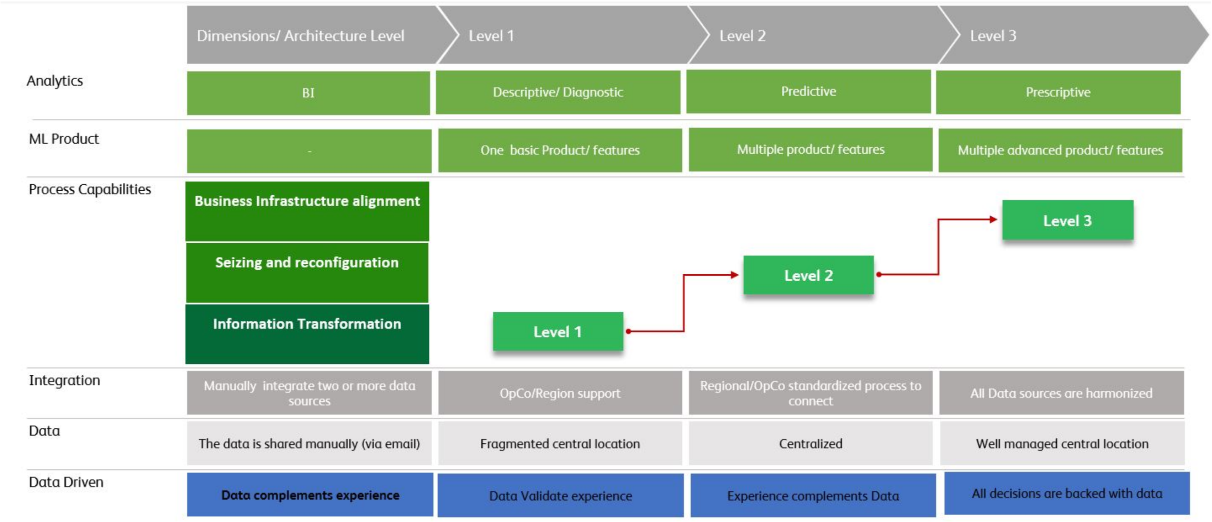


Figure 5.17: BDAC Deployment Reference Architecture levels

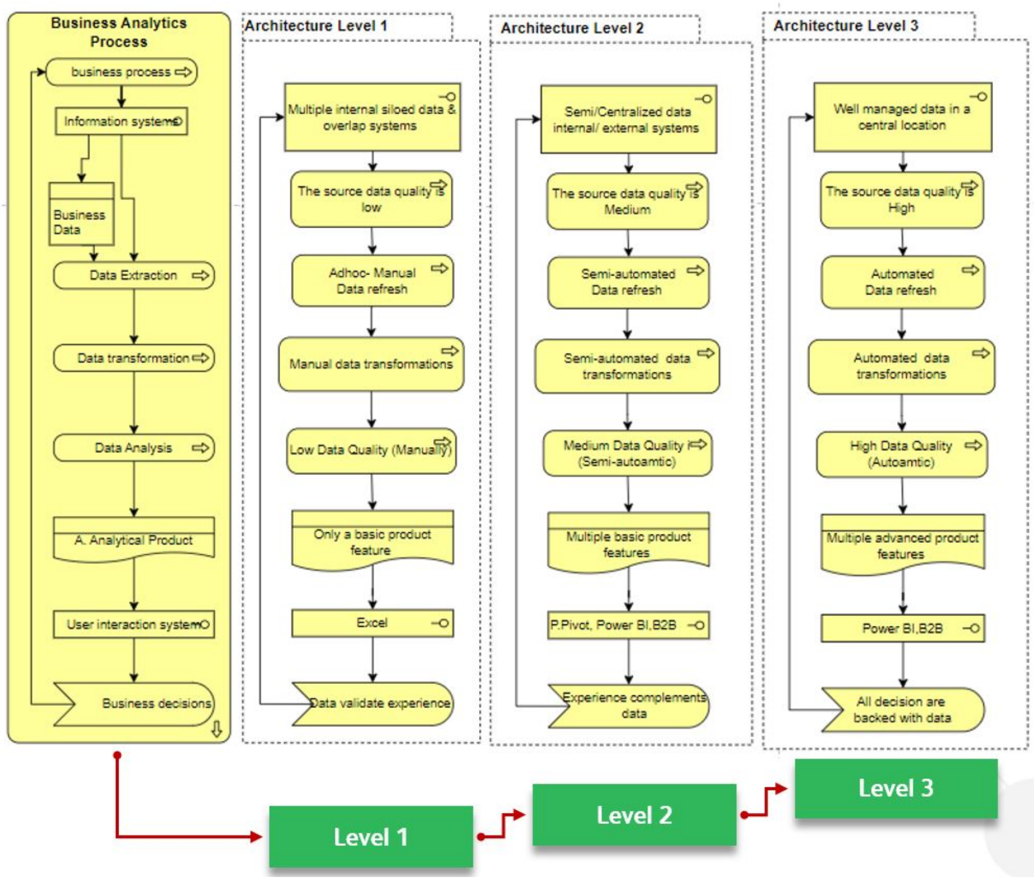


Figure 5.18: BDAC Deployment Architecture Advanced Analytics Business Layer Levels

5.2.2 BDA Capability Deployment Architecture- Level 1

This levels represents a local analytics function skills/ culture that prioritize experience over data ("Data validate experience"), it possesses several siloed data sources that required manual data extraction and integration through CSV/ excel files that possesses low data quality and only permits limited analytics product features. In a short description, is a level of multiple manual process and is supported by Global and Regional teams.The following are main architectural zones characteristics presented in chapter 4 ,with the following differences: (Please refer to Figure: 5.19):

1. **Machine Leaning Zone:** In terms people, this zone is supported by regional and local teams that support and guide the different Advanced analytical process. Particularly, on this levels the technology and Process resources are limited and are supported by regional hubs for this process. For instance, utilizing a common set of Data scientist and Engineers and processes. Additionally, the models KPI's are mostly focus on past business Analytical metrics and limited to just a few feature and models.

2. **CI/CD component Zone:** is limited, as the local data extraction is done manually and no data data pipeline is created from the data sources to global data source to feed the models.

3. **Data Engineering Zone:** works manually, with the local team manually extracting from local internal data and creating manual data transformation from ad-hoc rule based. No Particular semi-automatic ingestion data pipelines or process is put in places, and instead the file are shared using CSV, XML or corporate email.

4. **Technology Zone:** presents a siloed data sources landscape, with mostly internal data sources that are being stored in a non/centralized location and present low Infrastructure flexibility.

5.2.3 BDA Capability Deployment Architecture- Level 2

Presents a more local function context with mature resources and BDA capabilities with local analytics skills/ culture that balance experience and data on a ("Experience complements Data"), it possesses a more centralized (Internal, external) data sources with a semi-automatic data extraction and integration through Data pipeline's that possesses data quality and permits multiple analytics product features and outcomes through a front end such as Power BI or a B2B systems. In other words, this level presents a higher automation level, with the different steps by zones and process presented in chapter 4, figure: **??**, and its technology resource instantiation level as follows:

1. **Technology Zone:**

   (a) **Local Data Sources:** involve the multiple data sources and systems that are daily present in the business operations(i.e., ERP, CRM, sales orders). These data sources represents the raw data sources that later, after the ETL and data harmonization process will become in the the attributes necessary in model training and model features (Zone B1), and five main types are of data are identified: Streaming (IoT, Big Data streams), Unstructured (images, videos), Semi-structured (logs, json, xml. logs), structured (Relational data

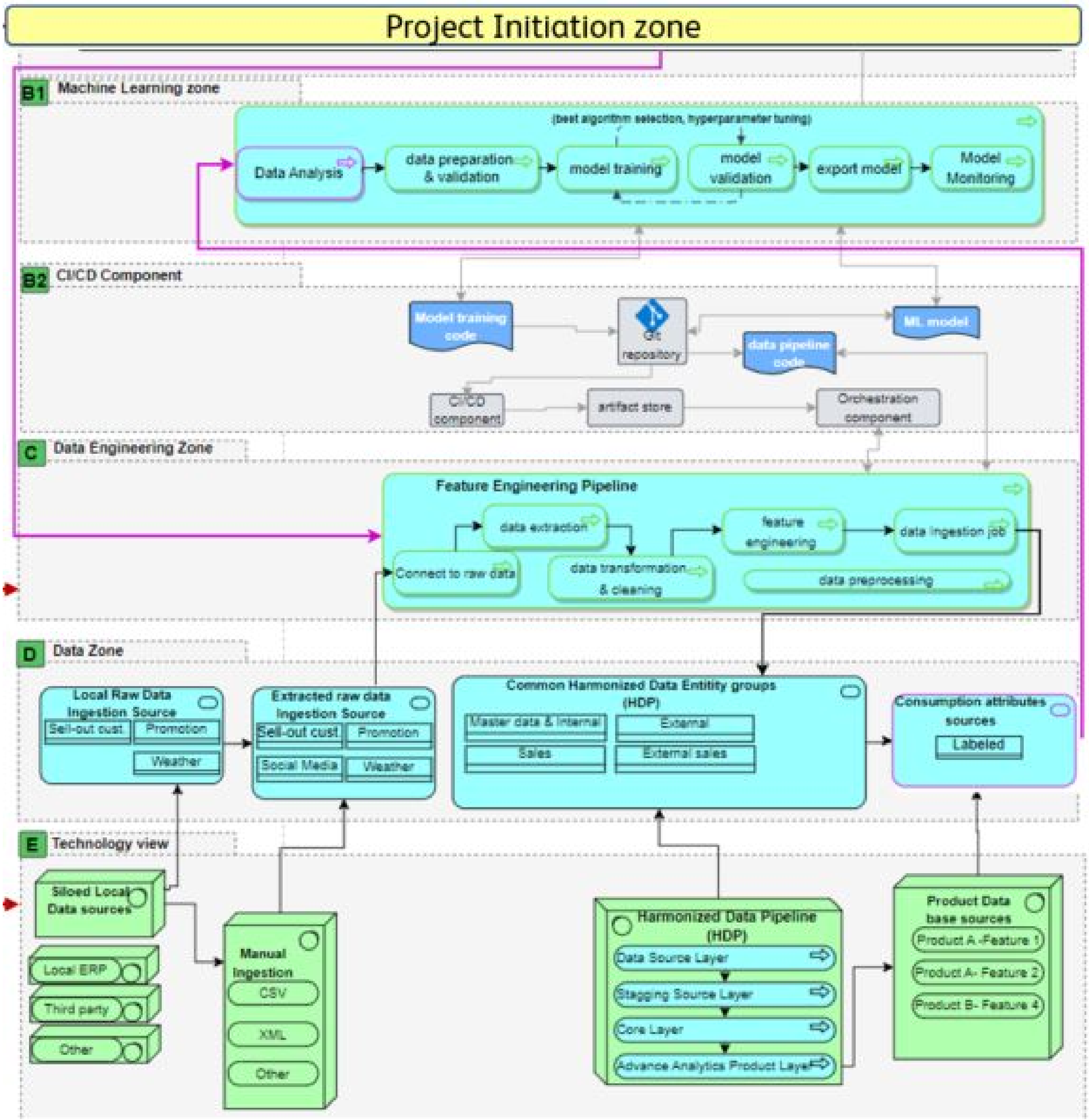


Figure 5.19: BDAC Deployment Reference Architecture level 1

base) and Cloud Data services (Azure data services; Cosmos DB, Data verse, Azure SQL Database). The following are some example of these local data sources:

i. Enterprise Resource Planning (ERP) serves as the core system for operational companies, encompassing and supporting various essential processes. It acts as a centralized platform enabling various operations, such as registering an Stock Keeping Unit(SKU) and specifying the necessary materials and processes required to manufacture a final product (including brewing, packaging, and logistics). Additionally, ERP facilitates the seamless flow of sales orders, tracking their progress from order placement to invoice generation and shipment while also providing comprehensive stock-level management capabilities. Essentially, ERP acts as the backbone of an organization, ensuring smooth and efficient operations across multiple functions. Some examples are (SAP B1, JDE, Oracle Fusion, or internal corporate ERP).

ii. Local promotional systems, functional systems (i.e., Marketing, logistics, finance).

iii. As the next step. the different ingestions systems are presented to ingest the data throughout different sources technology: (Azure Events/IoT Hubs, Azure synapse, Azure Databricks (Scheduled/ event- triggered data ingestion pipelines, and Pipelines orchestrators). Nonetheless, multiple ingestion technologies are present depending on the local BDA capabilities and resources, such as (i.e., DB file objects, Mesh Broker, API, JSON).

(b) **Analytics Enablement Platforms (AEP):** represents the extraction of the raw data from global systems such as Global ERP. Generally speaking, data is pushed by the source to the landing zone into a cloud storage service. From there, the data is ingested to a logical storage in the Data Lake, to later on be incorporated, transformed and orchestrated to be consume by Operational Companies (OpCo's, usually countries), Global Functions (like Global Procurement or Global HR), and Harmonized Data Pipelines (HDP). An example of technology that could be use in this platform are common cloud services use are Data Lake, Delta Lake, Synapse server-less. (Please refer to Figure: 5.17-Section-E and D):

i. The data pre-prossessing begins with data transformation and cleaning tasks, to later on be Harmonized to comply with the EDM requirements and model features.The transformation rule artifact defined in the requirement gathering stage serves as input for this task, and the main aim of this task is to bring the data into a usable format. These transformation rules are continuously improved based on the feedback and the Analytical model features and local BDA capabilities and resources maturity.

(c) **Harmonizing Data Pipelines (HDP):** is a standardized data integration solution designed to establish an automated data flow for every HEINEKEN Operating Company (OpCo). Its objective is to seamlessly merge global and local, internal and external data sources. By doing so, HDP facilitates efficient access to pertinent, trustworthy, and well-organized data. The main different with (AEP) is that HDP create digested data pipelines with Global and Local systems, particularly on transactional data, instead AEP only Extract and orchestrate raw data. To achieve this goal, HDP solution uses a framework called Data Build Tool (dbt) to support the creation and maintenance of the data warehouse pipelines.dbt supports the transformation step of a data pipeline. DBT consists of several core components such as models, seeds, sources and macros. Models are the building blocks of the dbt ETL pipeline through four main layers: Data Source Layer, Staging Source Layer, Core Layer and Advanced Analytics Product Layer. The principal outcome are the specific data marts that represent the Entities and attributes required in the features for the different Advanced Analytics Products (i.e., Sales , Retail Media marketing, Logistics, Public data, promotions) (Please refer to Figure 5.17, Zone D-Common Harmonized Data Entity groups).

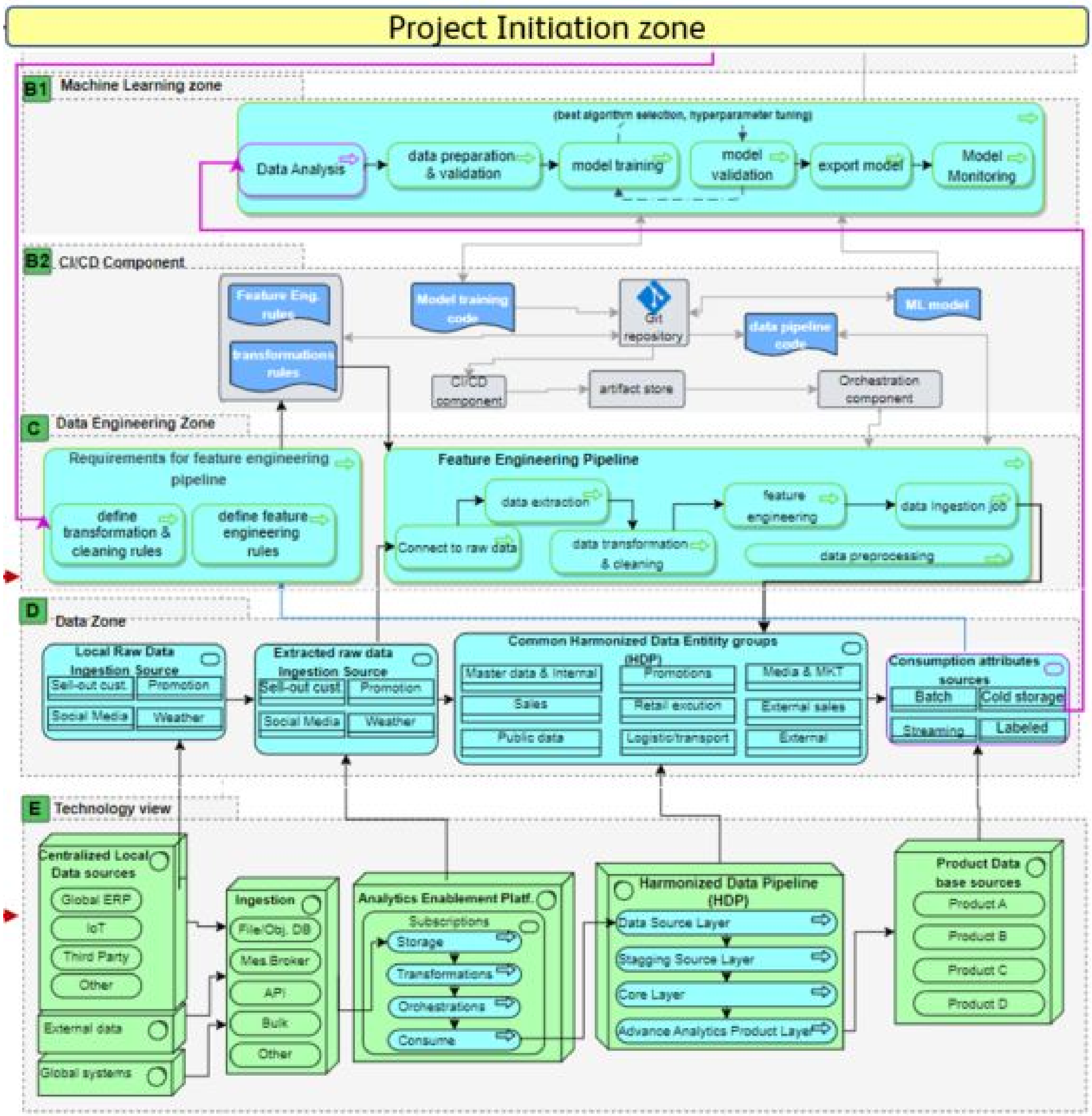


Figure 5.20: BDAC Deployment Reference Architecture level 2

5.2.4 BDA Capability Deployment Architecture- Level 3

This level present the "target architecture" of the reference architecture for level 2, with a high level of automation on the different process (Data Engineering Zone, ML Zone) that supports advanced Analytics features that support a highly mature skills/culture function driven by a "All decisions are backed with data approach". The following are the main difference from architecture level 2 (Please refer to Figure: 5.21):

1. **Machine Leaning & Data Engineering Zone:** automatized and becomes the **Automated ML workflow pipeline** flow data from the multiple data sources in the technology layer to the Machine Learning Zone, following the next steps:
   (a) The DevOps engineer (R6) and the ML engineer (R7) will oversee infrastructure construction for model training, which includes hardware resources and computation frameworks (e.g., Kubernetes).
   (b) The workflow orchestration component coordinates the tasks on the automated ML workflow pipeline by retrieving the necessary task artifacts from the artifact store (e.g., image registry) to be executed in an isolated environment (e.g., containers).
2. Throughout the ML automated ML workflow the following task are deployed automatically:
   (a) Data preparation, validation, and the division between the train and test data.
   (b) Refresh os updated model training on new data (versioned features) using predefined algorithms and hyper-parameters from previous experimentation. The main outcome is the model retraining and refinement.
   (c) Executing automated model evaluation and making iterative adjustments to hyper-parameters if necessary. The automated iterative training continues until satisfactory performance metrics are achieved.
   (d) Export the trained model and store (model registry).
   (e) Once a optimal performing model moves from staging to production, it is shared to the DevOps engineer or ML engineer for deployment. This triggers the continuous deployment pipeline, initiated by the CI/CD component (C1). The production-ready ML model and serving code are retrieved, prepared initially by the software engineer (R5). The continuous deployment pipeline builds, tests, and deploys the model for production serving.
3. Feedback-loop:
   (a) The ML engineer (R7) manage the infrastructure related to model serving.The continuous monitoring component verifies the model service system performance. As a low model prediction accuracy is detected the information is promptly relayed through the feedback loop. This feedback loop, permits a direct feedback, thereby fostering more robust predictions, by facilitating a constant training, retraining, and enhancement process.
   (b) The feedback-loop support the model monitoring component status to be transmitted to various upstream receiver points, including the ML Zone, data engineering zone, and the scheduler (trigger), fostering a collaborative cross-functional environment.

(c) The feedback received in the data engineering zone (Data sources or final user feedback) enables adjustments to be made to the features prepared for the feature store system. Moreover, the detection of concept drifts serves as a feedback mechanism, facilitating continuous training. For instance, if the model monitoring component identify a data anomaly for the scheduler to triggers the automated ML workflow pipeline for retraining purposes.

(d) Finally, the automated automated ML workflow pipeline outcome is push once again to the en-user interaction system (e.g., B2B platform or PowerBI dashboards)

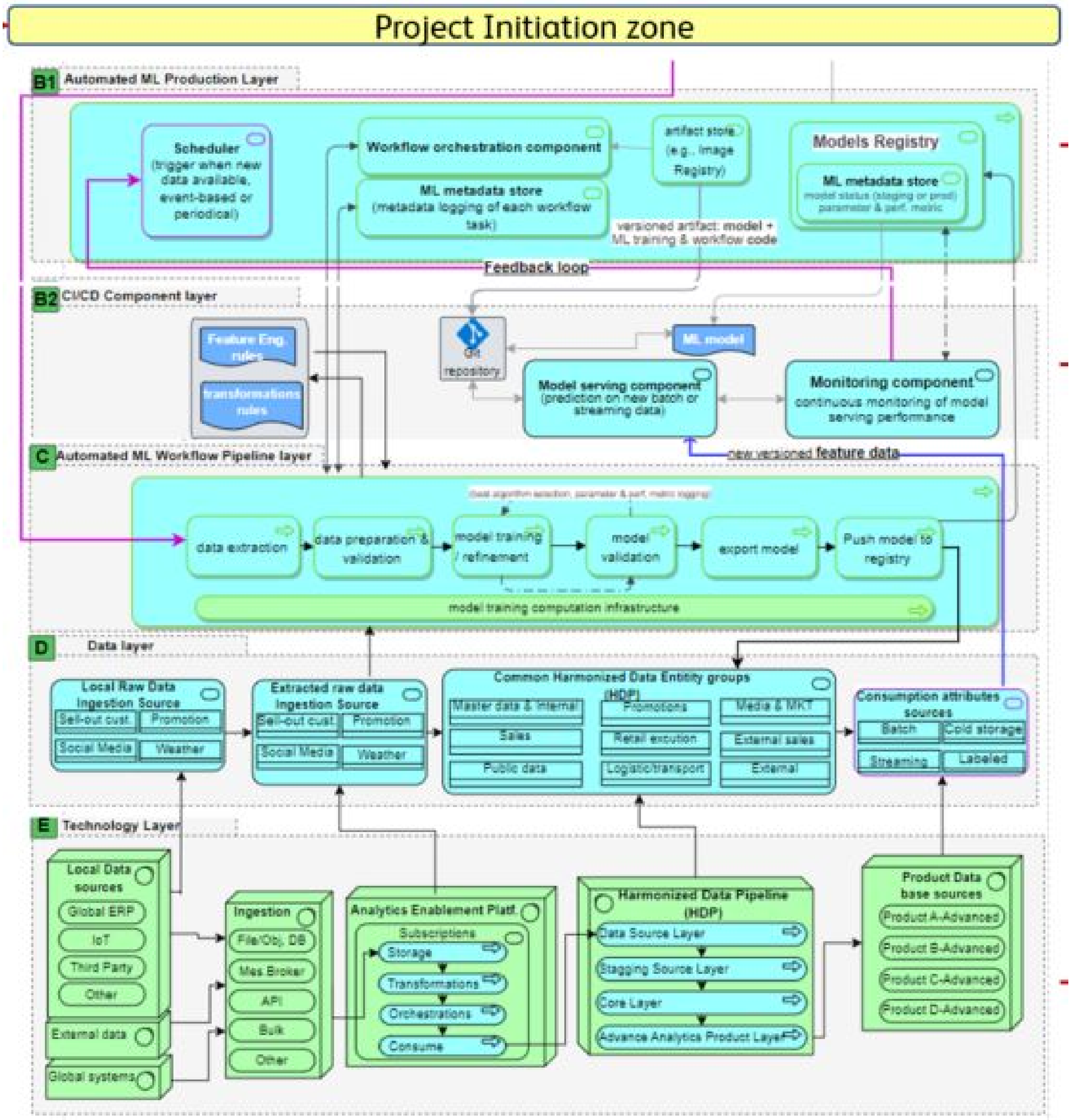


Figure 5.21: BDAC Deployment Reference Architecture level 3

Chapter 6

Validation Global Analytics BDAC Deployment Reference Architecture

The current chapter represents the fifth step in the DSRM methodology ([51]), which corresponds to the Evaluation step of the Global Analytics BDAC Deployment Reference Architecture, as stated in SQ: 4. To conduct this evaluation, the reference architecture was introduced to key stakeholders in a real-world scenario involving the deployment of advanced analytical products within Heineken's Global Analytics team. The evaluation process consisted of four semi-structured interviews, wherein the methodology and complete architecture development cycle were presented.

The interviews also included open-ended questions aimed at addressing specific architectural objectives. The validation artifact for this evaluation comprised the architecture prototype developed in the previous chapter. This prototype was applied to a real product deployment in a specific Operational country and was compared the architectural levels with the "base" and "target architectures" from the past, present, and future of the Advanced Analytical product within Heineken.

The evaluation of the Global Analytics BDAC Deployment Reference Architecture followed an expert opinion approach, which required conducting interviews with four team members of the Global Analytical team. These interviews aimed to validate the extent to which the designed architecture artifact effectively addresses the main research question, considering the outcomes derived from the Technical Action Research case [52]. The expert opinion validation method involved capturing the domain experts' positive and negative opinions based on their expertise in the fields of Enterprise Architecture, Data Engineering, Data Science, Product Development, and Deployment.
The interviews aimed to assess the domain's expert understanding of the presented artifact and the results effects of implementing the BDAC deployment architecture in the advanced Analytics project initiative. The validation process played a critical role in

identifying and eliminating architecture design choices that may be inadequate or ineffective. Additionally, the cross-functional experts insights from Heineken's real-world implementation added multiple valuable design improvements decision to the artifact.

6.1 Domains Interviews

To ensure a thorough artifact validation, a set of three interviews with key stakeholders from diverse domains and disciplines develop. These stakeholders had direct involvement in various real-world deployment process of an advanced analytics product milestones within the OpCo context. Their perspectives provided valuable insights into the BDAC Deployment architecture, considering the historical product deployment process in the past, present, and future. These stakeholders were selected based on their expertise and process integration relevance with different deployment global and local domains. The following four domain experts were chosen for their significant contributions:

Stakeholder	Domain	Relevance
Lead Data Scientist	Data science and Engineering	Lead Data science responsible to develop and implement the advanced analytical products features, and supporting the OpCo data transformations adjustments.
Chapter Lead Data Engineer	Data Engineering	Lead the global analytic ETL and data harmonization process between Global Analytics and the OpCos.
Analytics Translator	Data Engineering	Responsible to engage,prepare and guide the local data extraction from the OpCos to the global analytical storage.

Table 6.1: Domains experts selection

6.1.1 Interviews questions

The practical and deployment relevance questions were formulated to capture the expert's predictions regarding the implementation and adoption of the BDAC deployment architecture within Heineken's D&T Global Analytics team. These questions were carefully selected to evaluate the effects and perceptions of the deployment architecture levels and their alignment with the evolving context and solutions of the analytics products over time. The following are the proposed questions:

6.1.2 Interviewers answers

This subsection compiles the most significant insights obtained from the individual sessions with three Global Analytics team interviewees. The results include discoveries related to the three architectural goals, the effects of its instantiation, and its contextual effects.

#	Interview Questions	**Architecture goals, subsection 4.1.1**
1	Does the BDAC Deployment architecture levels represent the most important building blocks, layers and processes of the deployed Advanced analytics product?	Demonstrate empirically the firm's BDA capabilities building blocks of a firm's BDA capability
2	How are the BDAC (Business Infrastructure alignment Capability,Seizing & Configuration Capability, and Information Transformation Capability) present in the deployment architecture levels?	Portrait the different deployment architecture layers and transformations process to deploy Analytical products
3	What would be the effects if the DBAC Deployment architecture would have been instantiated back in the past for the analytical product deployment?	Identify the BDA deployment interdepencies, mediating processes and mechanisms in different local resources and capabilities
4	To what extent is the deployment architecture expected to contribute in the cross- domains teams collaboration efforts to deploy analytical product (i.e., Final user, Product Owner, Translator, Local Data team, Data Scientist, and engineers)?	Identify the BDA deployment interdepencies, mediating processes and mechanisms in different local resources and capabilities.

Table 6.2: Interview questions

Interviewer: Lead Data Scientist

1. Yes, it does represent the most important building blocks, layers, and processes. However, the analytical product is in between levels being proposed in the solution architecture, i.e. between level 2 and 3. It is important to consider that the architecture should represent the global and local integrated process. Particularly, is clearly that B1,B2, C and particularly D are supported by the global team. Finally, the Analytical product, the data sources and data extraction process are in constant evolution, this process complexity are partially represented in the architecture. However, the level 3 the model registry is relevant to be represent partially in the architecture as it will feed the model data to the different ML models.

2. BDAC deployment also will help with the mapping of right resources to their respective processes, and how these different resources fit in and rely to each other in different processes and used in the proper process/ job. For instance it will be particularly useful for the resources in terms of hiring to hire the correct resource for the job. Additionally, transformation capability is evident through the three levels that are proposed in the solution.

3. The main benefit of the BADC deployment is going to be cost saving in terms of finance and time consumption since it is automation processes are going to

make most process more effective performance (model and data engineering reliability).Additionally, for both Global and local OpCo Heineken will transform in a cost reduction (working hours), by defining a common communication tool to optimally assess the local resources and process orchestrations. However, this value need to be communicated clearly through the deployment process and not only the business outcome based approach.

4. The deployment of the proposed artifact is going to be mostly helpful for technical side of the organization, while the business side will not be entirely concerned aside from the fact that how this deployment will benefit them in terms of outcome. In this sense, the architecture, even in a high level still represent a highly complex and limited to business "related processes".

Interviewer: Chapter Lead Data Engineer

1. Yes, the architecture presents the most relevant building blocks, data process evolution and product evolution through time. The architecture level 1 and 2 represents the product and data evolution and feedback look of level 3 (Automatio- feedback loops processes). However, this might differ as the product matures and is deployed in another OpCo. Additionally, the expert recognized that the model is missing the change management process, which at the moment represents a significant challenge on the different OpCo Deployment processes.

2. The main effect of the architecture would have been the representation of product and deployment processes, local resources, and capabilities architecture levels. Additionally, it would allow the local teams to assess the detailed product that will be deployed, the resources, and the related cost. The architecture views/ layers would also guide the deployment communication and describe the process's complexities (e.g., The effect of particular changes on the final deployment time)

3. First, it would have been beneficial for Opco local team, as they would understand the detailed product to be deployed and its dependencies. More than a concrete simplification, the levels allow the OpCos to assess their deployment plan based on various maturity levels of each deployment zone. Second, the architecture portrays the complexity of the cross-functional integration process and the fundamental reasons for multiple teams' involvement.

4. This architecture help to achieve the cross-domain teams collaboration and the involvement of individuals belong to different projects. Additionally, the architecture helps in the collaboration of multiple teams and particularly in the technical realm is architecture levels portrait the multiple technical dependencies.

Interviewer: Analytics Translator

1. The architecture is clear in representing the most critical aspects. However, in the engagement aspect, it is concerning that the "product initiation zone" is fully complied with before the project deployment. It is fundamental for the deployment process to understand the Local Data layer and align with the local teams regarding

the EDM process. Particularly, the local OpCo function team might use local data rules/ transformations specific to their operations and data processing, and the architecture portrays and communicates these differences, for instance, with a new team member. Subsequently, the expert mentions that given the IT landscape complexity, is important in the model to add an specific timeline for the different process, such as the data preparations and transformations, as they are required to be ready ahead of the analytical product deployment.

2. From the Information transformation and Seizing & Configuration, the architecture process recreates the most significant changes base on the OpCo maturity levels. From a manual process to a more semi-automatic and later or automated pipelines and the architecture assess the local process and the evolution on these process on time. For instance, in the changes on a new ERP.

3. The three main effects identified in the target architecture are: 1) Establish a standard automation process to verify the work done by the multiple teams (e.g., consultants) and review the standard framework for features requirements. 2) It will benefit multiple stakeholders in the local minimum resources and data requirements. 3) The elaboration of a "faster, sharper, data-driven strategy", with a common language that eases the communication between the different local D&T leads.

4. The deployment architecture supports cross-domain collaboration as it clearly defines the different processes and the roles/teams' responsibility for the different roles (Local, global, Consultant). By doing this, the communications improve as the different teams relate to the common process and the reasons why specific data requirements are made on each deployment process steps.

6.2 Validation Results and Analysis

6.2.1 Validation Results

The three domain experts considered the BDAC Deployment architecture levels as representative of the most important building blocks, layers, and processes of the deployed Advanced analytics product. The perception of the three architecture levels were related but different based on the current background, their deployment experience and their day to day responsibilities. For this reason, having the domain experts that work in the BDA deployment process complements the insights and feedback about the proposed artifact. In the case of the Data Science lead, his review brackground relates to multiple years of experience from multi-industry projects and the gradual growth of the current advanced analytical project. The combined role both from strategic and technical encompasses his feedback in critical aspects such as the product evolution, local analytics capabilities, and data processing evolution, at levels 1 and 2. It also acknowledges the relevance of level 3 as a realistic target architecture, which includes the model registry that feeds data to different ML models through the automatic feedback loops as a core technical process. The domain expert also recognizes the importance of global and local integrated processes, with certain elements (B1, B2, C, and D) supported by the global and local team. However, the analytical product and the data sources and extraction process are in constant evolution, and their complexity is only partially represented in the architecture. For instance, multiple OpCos might not be even in level 1 or might be between levels. Based on this high level approach and the historic deployment process, the Data science expert recognized the value in terms in cost and time saving of the deployment artifact (i.e., data pipelines and technical teams alignments). However, the technical alignment perceived that this benefit might.

The chapter Lead Data Engineer, presented a more specific approach regarding its field and expertise dealing with a wide variety of local IT landscapes and data transformation processes. He acknowledge the local architecture evolution and different as the Analytical product matures and is deployed in different operating companies for different uses. For instance, a basic product feature (i.e., Customer churn) in one country might represent a highly advanced product in another one less matured in terms of analytics. Additionally, giving the current experience, he considered change management process as one of the most important problems in the data transformation, for which it should be reflected in the in architecture. Nevertheless, the multiple levels, zones, and buildings blocks permits the technical teams to improve the collaboration and review the roles and process interdependencies (i.e., the effect of changes in the overall deployment process).

Finally, the Analytics translator in its role to engage and review some of the data challenges recognized the importance of the engagement processes. for example, She found the architecture "product initiation zone" a corner stone in the deployment process, but to be effective, it required deep understanding of the Local Data layer and align with local teams regarding the EDM (Enterprise Data Management) process. For her, each OpCo team D&T and business stakeholders may use specific local data rules and transformations, and the architecture should portray and communicate these differences to new team members. For this reason, the domain expert recommendations approach related to the alignment of data/process and responsibilities standards to address the multiple IT landscape complexity. For instance, proposing to add on top of the architecture levels an specific timelines for different processes and zones, such as data preparations and transformations, to ensure they are organized accordingly and ready before the analytical product deployment. In confirmation to the previous two domains experts, the Analytics Translator perceived the architecture as "highly technical" but valuable cross- domains communication and alignment tool.

6.2.2 Validation Analysis

The validation results encompass diverse perspectives from three distinct domain experts involved in deploying advanced analytical products at Heineken. Each expert shared their valuable "hands-on experiences," offering a wealth of insights and reflections on various aspects such as architecture levels, dimensions of capabilities, and Analytics maturity levels when applied in the product deployment process. The following three validation analysis outcomes derive from integrating the practitioner-based findings and the Systematic Literature Review (SLR).

1. **The DBAC deployment architecture generalization** Through conducting multiple validation interviews, the domain experts instantiated the artifact within a previous advanced analytical big data project deployment in a specific Operational Company (OpCo) at Heineken. Surprisingly, despite the diverse array of internal projects these experts had encountered in their past or present experiences with big data product deployments at Heineken, the architecture's main blocks and processes remained relevant across different products and architecture levels. While the architecture does not precisely define the status of each OpCo deployment, the experts found the architecture's levels useful as a standard "blueprint BDA Capability deployment roadmap." For example, the data scientists' input indicated that the architecture represents two deployment processes, namely global and local, and certain OpCos have yet to reach level one regarding deployment analytics maturity.

   These comments illustrate the architecture's applicability as a common framework for various projects and landscapes, as corroborated by the Analytics translator, who highlighted the complexity of the company's IT landscape. Moreover, a thorough analysis of the interview results from Heineken's Global Analytics domain experts revealed that they employ theoretical models and methodologies to define and compartmentalize the deployment process. This approach facilitates the analysis, comparison, and guidance of current OpCo deployment status in terms of processes and levels while also defining communication tools, standards, and best practices.

   In essence, the BDAC architecture goes beyond mere maturity assessment by defining deployment levels based on local capabilities and resources and outlining the processes and structures required for successful BDA Capability deployment. For instance, if an OpCo is assessed at a particular deployment level, the architecture offers a means to position it between deployment levels and provides a road-map for achieving the subsequent development steps. Finally, from the academic standpoint, the theoretical foundations for the present BDAC deployment architecture (levels), such as Big Data Capability Dimensions, MlOps model, TOGAF (ADM), and the Multi-Criteria and Model-Based Project Selection Integrated, are considered to be Big Data Analytics projects "agnostic," which means that the first and the former in a high degree applicable to a wide variety of industries, organizations, and data analytics initiatives.

2. **DBAC deployment architecture feedback & improvements** As part of the interview process, the experts express the following set of architecture improvement opportunities to develop in future architectural views (Please review Table 6.3).

3. **BDAC architecture validation limitations** The present Qualitative research leverage the exper domains experience in the deployment of advanced analytical products to validate the artifact in the "real- context" in Heineken. Nonetheless, the validation process followed in the present subsection is subject to a series of validity threads, including: Transferability, Confirmability, the domains experts.

   (a) **Transferability:** refers to the extent in which the results of a qualitative study can be extrapolate and applicable beyond the specific research participants and

Feedback improvements	Architecture impacts
First, both the Data Science lead and Data Engineering- chapter lead expressed the need to further include in the architecture the communication and change management process. As different local teams would continuously evolve and matures in their deployment cycles, a clear communication process and change management are fundamental to speed-up the deployment process.	The architecture view in the business layer (refer to Figure: 5.6) three phases (Plan, Prove, and Produce) included the cross-phases communication and change management activities, which some of them includes: BD Deployment status, Review of the deployment architecture levels, Mapping of the Maturity assessment resources to each level, graded in terms of colors, time and cost the corresponding resources (People, Process, and Technology) plan versus the executed (i.e., Traffic light views).
Second, add process and more specifically data preparation (i.e., Data Backbone) timelines given the current IT landscape complexities, local data transformations, and engagement challenges.	The architecture impact will be the adjustments on the business layer view and generate the corresponding timelines and ETL/ data processes required. For instance, depending on the Capabilities dimensions maturity levels, the Technology zone could be place in different order and adjust to the local processes including timelines requirements for those steps.
Third, in overall the domains experts express the satisfaction for the deployment architecture levels a common language communication tool. However, unanimously they assess that the architecture levels communication nature is still highly technical and significantly away from the business "outcome approach" stakeholders.	The overall impact of this feedback limits significantly the target group and the "blue print road map" approach to multiple teams. One option to assess add different business KPIs related to investment, ROI, time range, and planing on simplified but most relevant process (i.e., ETL, data harmonization or ML production zone accuracy or scalability). Additionally, add visual forms to establish the "healthy" levels of the core zones and layers, including the previous two feed-backs improvements.

Table 6.3: Interviewers feedback and responses

circumstances, allowing for potential generalizations or comparisons in different situations or contexts. In the research case, the theoretic methodologies and the nature of the data and cloud project initiatives permits the extrapolation of the significant part of the architecture levels, zones and views, including the technology. However, is important to clarify that in some degree, the architecture levels are driven by the local Maturity levels (i.e., Analytics) to divide the capability metrics and analysis. For instance, selecting different questions to validate the maturity levels (People, Process, and Technology) resources or selecting different BDA Capabilities dimensions.

(b) **Confirmability:** refers to the importance of researchers being aware of their own beliefs, biases, and assumptions throughout the research process and taking steps to minimize their influence on data collection, analysis, and interpretation. In other words, ensure that the study's outcomes are genuinely representative of the domains participants' experiences and perspectives. In the present section, this validity threat is present in the analysis and translation of the domains interviews artifact´s feedbacks and improvements. Additionally, some of the insights and example cases used for the domain experts are related to specific cases and products that are not public. For instance, the maturity and example of maturity cases or challenges faced during the implementation processes of current advanced analytics products in Heineken (i.e., AIDDA deployment process).

(c) **Credibility:** Lastly, credibility is a crucial aspect that ensures the reliability and validity of findings, conclusions, and interpretations. A study or piece of work that is deemed credible is considered to be based on sound evidence. The research validation and domain experts selection from the three most essential deployment domains(Translator, Data, and Data Science), which is included in the global team with significant experience and the insights of the thesis company supervisor as a Solution Architect. However, given the deployment nature on multiple local contexts (OpCos), the thesis validation possesses limited local stakeholders to validate the possible challenges, advantages, practicality, and best practices to communicate the BDAC deployment "blueprint". As such, it represents a cornerstone to add to future research and architectural design impacts.

Chapter 7

Conclusions

The current research has introduced an empirical deployment architecture for Big Data Analytics Capabilities, defined by essential foundational elements or zones. To achieve this goal, this architecture has been developed using state-of-the-art MLOps design patterns and guided by integrated methodologies from the field of Enterprise architecture. The research methodology followed the Design Science Research Methodology as discussed in subsection: 1.4.1, starting with the identification of problems and motivation in the chapter: 3. Next, the formulation of the solution objectives was defined in subsection 4.1. Based on the reference architecture goals and theoretical design patterns, chapter:4 resulted in developing a theoretical reference architecture artifact for deploying big data analytics projects. This architecture incorporates the TOGAF methodology, including various architecture views and layers tailored to specific levels of Analytics Maturity. In chapter: 5, the proposed theoretical deployment architecture was operationalized within the context of deploying Global Analytics -Advanced analytics products in a specific country of Heineken's operational company (OpCo). Finally, in the present chapter: 6, the real-world effects of the artifact in the Heineken context were evaluated through interviews with cross-functional domain experts within Global Analytics, aiming to gather insights and feedback regarding the architecture goals (subsection: 4.2) and challenges encountered during real-world deployment context.

7.1 Research Questions

1. **How can Enterprise Architecture be used to identify big data analytics capabilities (BDAC) to improve business value?**

   In order to discover the impact of EA to identify BDAC to improve business value, chapter 3 conducted a SLR to assess the companies ability to analyze and leverage business value from large operational data volume. On this research, the three most relevant obstacles to develop a big data analytic capability resides in the use of BD resources: People (i.e., Knowledge, skills, communication), Process (i.e., Change management, process complexity, roles & responsibilities), technology (i.e., Data: quality, access, standardization, Transformations and Integration). The ability to use BD resources is defined as Big Data Analytics capabilities and these are found to be relevant sources to increase business value and generate competitive advantage.

Extant literature presents multiple definitions of BDA dimensions, however the following six Evolutionary BDA dimensions support the firm's growth and value creation: (1)Business process integration, (2) Infrastructure flexibility, (3) strategic alignment, (4) Learning capability, (5) Relationship infrastructure, (6) BDA-driven decision making. These are found to be the BDA dimensions that support a firm´s resources and capabilities reconfiguration. The (1) and (5) dimensions allow firm´s to sense **What** to change to leverage existing business process information. The (2),(3), and (6) allow firms to define **How to change** seize and reconfigure (i.e., Investment, management commitment, tools or Analytics training's). Finally, (6) and (2) dimensions support the organization´s overall **Information transformations**. Multiple studies and business cases have demonstrated different processes and applicability of big data Analytics in a wide range of industries (i.e., Health care, Telecommunications, Retail, Finance, or Water management). Nonetheless, little is known about the process and structures necessary to orchestrate BD resources and capabilities dimensions to develop firm-wide BD Analytical capability.

Extant literature presented multiple definitions of BDAC dimensions, frameworks, and methodologies of project selection and capability analysis to represent the Big Data project processes. For instance, Project Portfolio Management (PPM) or Capability-Based Planning (CBP) are known framework methodologies used in the industry. Nonetheless, these approaches are limited to represent the project interdependences, local resources maturity, and project deployment impact process. Another approach, based on current literature, suggests Enterprise Architecture as an approach that allows to explicitly define the project's Big Data interdependencies, analyze and represent the most important structural building blocks processes necessary to deploy a BD Analytical capability. For this reason, the present research has selected EA and frameworks (i.e., TOGAF-ADM) as the discipline to demonstrate the Big Data Analytics capability to help firm´s to achieve reconfiguration and orchestrate its core BD resources and BDA capabilities.

2. **How can Big Data Analytics capabilities and Enterprise Architecture integrate into a BD Analytic capability deployment reference architecture?** Different firms across multiple industries are embarking on a gradual digital transformation of their business processes, aiming to integrate business analytics and leverage operational data to create business value. For this purpose, this research is to empirically integrate the core dimensions of big data analytics (BDA) capabilities, big data (BD) resources, and the required analytics types. The goal is to establish an empirically-based BD Analytics capability for analytics project initiatives. As organizations progressively develop their strategic analytics capabilities, the TOGAF ADM method will guide the Enterprise Architecture (EA) to integrate and model the BD capabilities dimensions corresponding to analytics types (e.g., Descriptive,

Diagnostic, Prescriptive), as well as the required maturity of BD resources. Multiple cross-functional stakeholders (e.g., Data Engineering, Data Scientist, Final user, Product Owner), processes, requirements, and data-flow transformations will be represented in various architecture viewpoints.

The initial step integrates the three Analytics types (Descriptive, Predictive, and Prescriptive) to the corresponding BD resources and capability maturity levels required in figure: 5.8. For this reason, the architectural views will portrait the integrated three core evolutionary BDA Capabilities dimensions (business process integration, strategic alignment, and infrastructure Flexibility) using the the corresponding BD resource maturity level. The first capability dimension refers to the ability to optimized existing business processes using IT systems (i,e., Analytical products). BDA strategic alignment Flexibility refers to the ability to quickly develop, deploy and support the firms' BDA resources (People, Process, and Technology). Finally, Infrastructure Flexibility is the ability to manage and exploit data to generate value. The level of usage of the analytic capability will measured through an analytics maturity assessment (i.e., CMMI) to assess the BD resources maturity (e.i., Analytics knowledge, integration process, Data sources) following the Multi-Criteria and Model-Based Project Selection method [6]. This step allowed to TOGAF Architecture Vision Phase to defined the conceptual base and target architecture. As the next step, the MLOps architecture is used an architectural design pattern to portrait the multiple integrated processes viewpoints.

First, the business architecture, represented by the Business Analytics process in the figure: 4.9 possesses four processes steps: (1)Information systems, (2)Data Extraction, (3)Data transformations, and (4) Data Analysis. As a new Big Data Analytics project (ML) is initiated, on top of the business analytics process, a new the MLOps Project Initiation (A2) starts, by the defining the business problem, Data requirements and Data connection as shown in figure 4.11. The next step in the process is to establish the data-flow between the Technology, Data Engineering and Machine Learning zones, as shown in figure: 4.10. Once the analytical problem is identified, the technology data view points (Section D) represent data sources such as ERPs, IoT, and different operational systems that are selected to extract the operational structured or semi-structured data. The different data sources are aggregated in the Data aggregation service, where they are extracted, transform, and storage. Follow by this process, the data Engineering view (Section B) create the feature engineer pipeline that extracts the data and transforms it in a common standard form to be used as a feature for the different ML models. Finally, the Ml Production/experimentation zone (Section C) ingests the processed and standardized data in different ML models to test and develop different analytics test and analysis (i.e., Descriptive, predictive) to solve a business problems (i.e., churn prediction, cross-selling offers). From these viewpoints, it is possible to demonstrate the requirements to develop a DBA capability deployment architecture by connecting the different BDA processes through the architecture viewpoints. As the final step, TOGAF step Opportunities and solutions three architectural levels (1,2,3) are propose to represent the BDA Capabilities and measure by the use level of BD resources through an analytics maturity assessment level).

3. **How to operationalize the BDA Capability Deployment architecture to support the deployment of advanced Analytical products in Heineken's Global Analytics?** Heineken's strategy known as EverGreen defines the global aim to become "the most connected-brewery". Within this goal, the global analytics team, as part of the D&T, has the goal to create incremental value through the development and deployment of Advanced Analytics products. Nonetheless, the deployment in more than 90 countries that possesses multiple ERP systems, IT landscapes and business process, implying variability and challenges in the deployment process. Based on interviews with Product owners, Translators, Data Engagement analyst reported three main business context problems: Deployment speed (Data Quality,Integration, extraction, skills & knowledge), Communication (Engagement, Change management (incentives)), and complexity (IT landscape, roles, organizational structure). For this reason, the operationalized architecture possesses three levels (1, 2, 3) to represent the base and target architecture view for each evolutionary capability based on the maturity use of the BD resources measured by three GA Analytics maturity assessment questions as follows:

BDA Capability	**GA Maturity Assessment Question**	**Level 1 Descriptive**	**Level 2 Predictive**	**Level 3 Prescriptive**
Business Infrastructure alignment (What to change?)	To what extend are you data driven?	Data Validate experience	Experience complements Data	All decisions are backed with data
Seizing and reconfiguration	Manually integrate two or more data sources (How to change?)	Manually integrate two or more data source	Regional/local standardized processes	All Data sources are harmonized
Infrastructure Flexibility	How you integrate two or more data sources?	(email), Fragmented central location	Centralized	Well managed central location

Table 7.1: Global Analytics BDA Capability Deployment Architecture levels

The three BDAC deployment reference architecture levels will represent the corresponding Big Data Analytics Capabilities dimensions architecture viewpoints: The business process integration is represented in figure 5.17, the seizing and reconfiguration in the zones B1, B2, and C of each architecture level, and infrastructure Flexibility in zone D and E correspondingly. In other words, each architecture levels has the same zones and capabilities dimensions , what it different one to another is the ability to use the DB resources maturity measure the Maturity assessment levels.

In level 1, which represents the base architecture, the local analytics function skills/ culture are low and prioritizes experience over data, in other words ("Data validate experience"). Additionally, it possesses several siloed data sources that require manual data extraction and integration that generate low data quality and only limited analytics product features (Please refer to figure: 5.19. Level 2, is the next level in the analytics maturity of the BDA resources with mature resources and BDA capabilities with local analytics skills/ culture that balance experience and data on a ("Experience complements Data"). It possesses more centralized (Internal, external) data sources with a semi-automatic data extraction and integration that generates a higher data quality and multiple analytics product features (Please refer to figure: 5.20. Finally, level 3 represent the target architecture from the previous levels and portrait a high level of automation on the different process (Data Engineering Zone, ML Zone) that supports advanced Analytics features that support a highly mature skills/culture function driven by an "All decisions are backed with data" in the figure: 5.21).

4. **What are the effects of the implementation of the BDA Capability Deployment architecture in Heineken's local Operational Company Advanced Analytics product deployment context?**

   The BDAC deployment reference architecture validation presented the deployment architecture levels as a concrete evolutionary architecture. This architecture serves as the base and target architecture for past, present, and future advanced analytics product deployments. These specific architecture levels feature allowed the domain experts to easily review the differences in architecture viewpoints and assess the application of real-world analytical products.

   To evaluate the main architecture differences and applied to the deployment context, domain experts are presented with each view(section: 6.1.2). In the evaluation, the designed artifact contributes to the research problem and the practical implementation of a DBAC deployment architecture to guide the advanced analytics products deployment. As assessed by the domain experts in the interviews, the artifact demonstrates to represent the most relevant deployment building blocks process through the representation of the business initiation, technology, data Engineering, Machine Learning, and initiation zones. Furthermore, the architecture levels 1, 2, and 3 represent significantly the most relevant "based/target" architecture of the advanced analytic product evolution and future product development steps, such as the manual data storage and processing to a semi/automated storage, data pipeline, and ML experimentation. In that sense, the product and the architecture are still evolving, but the most relevant product deployment milestones are present through the different architecture viewpoints (including the Integrated business Analytics process and roles in figure 5.12). The target architecture in level 3 is found as a realist next step in the analytical deployment product journey and in combination, the three levels permits establishing a "common language" to elaborate an agile strategy with the different stakeholders, establish synergies in the cross-functional and cross-domains team, and a blueprint representation of an advanced analytics product deployment.

However, the high-level representation through the Archimate modeling language limits the representation of some of the most relevant deployment challenges, such as Change management, demonstrating the difference between the global and local teams' processes, the process scheduling timing of each zone, and its representation contributes mostly to the cross-domains efforts of technical teams. Finally, the representation of the deployment of the advanced analytical process is a complex challenge. However, the artifact permits answering the main research question and elaborates a "blueprint"/roadmap steps to represent the most important process in the deployment of advanced analytical products and improve the overall cross-functional integration by assessing the local BDAC resources and capabilities. The answers and details collected from these interviews are compiled in subsection 6.1.2 and will serve as inputs for future research in section: 7.4.

7.2 Contributions

Former research has considered EA's relevance and suggested the need for more empirical evidence on using EA frameworks in coordinating BDAC within the industry to develop a competitive advantage and create business value. Considering the overall impact of this research from the practitioner and scientific relevance, and the evaluation results presented in section 6.1.2, the following are the four main contributions:

1. From a theoretical standpoint, this thesis introduces an innovative approach to developing Enterprise Architecture for Big data Analytics Capability. The artifact combines essential building blocks, zones, and levels that facilitate modular and resource-driven architectures. This is achieved by incorporating insights from a systematic literature review and the instantiation of a multinational-beverage organization. Additionally, Technical Action Research is conducted to provide valuable insights into how the proposed artifact contributes to Advanced analytics product deployment by integrating new architecture patterns and methods into a novel Enterprise Reference Architecture.

2. Researchers and practitioners can employ the proposed artifact to analyze the business impact of aligning people, processes, and technological elements into critical functional capabilities for successful Digital product implementations in organizations—disciplines such as DevOps. MlOps highlight the significance of adopting new empirical approaches to foster strong organizational communication and change management. In that sense, the reference architecture integrates the benefits of designing a cross-functional and interdisciplinary organization as part of the architecture development, considering the complexity of business analytics and moving away from the standard and more static traditional Enterprise Architecture approaches.

3. From a practical perspective, this study validates an empirical DBAC approach to articulate complex processes and a variety of contexts in real-world Advanced Analytical product deployment. The Evergreen strategy and the digitalization across different functions in Heineken provide a suitable context to assess the usefulness of concepts absorbed from other architecture frameworks or methods, which may have been presented solely at a theoretical level or were limited to specific real-world implementation cases. Thus, this study serves as a reference for validating different theoretical and niche business context approaches, elucidating their specific advantages and disadvantages in practical implementation. Overall, experts interviewed during the evaluation stage of this research have positively evaluated the framework, particularly for offering a well-grounded Enterprise Architecture approach that addresses the unique aspects of Analytical deployment implementation capabilities and resources.

4. Finally, the architecture methodologies employed in developing the DBAC deployment architecture, along with the integration of Analytical Maturity assessment and design patterns, hold potential for application in diverse domains beyond the beer industry or analytical products. For instance, these methods can be employed to assess the global or local deployment of big data/ advanced analytics products or services by utilizing multiple IT or cloud assessments. These can also serve as a methodological foundation for the integration of Analytical maturity assessments into different reference architectures views. By leveraging these methodologies, organizations can benefit from a comprehensive approach to designing and deploying their products or services, ensuring alignment with best practices and optimizing overall business value using Enterprise Architecture.

7.3 Limitations

Throughout the research process, despite successfully addressing the main research questions and sub-questions and making contributions to both the academic and practitioner realms, several limitations have been identified.

The first limitation pertains to the restricted access to empirical BDAC frameworks or big data business cases. This constraint restricts the technology solutions and business cases considered in the architectural views and data applications. Secondly, the proposed Multi-Criteria and Model-Based Analysis for Project Selection, as discussed in subsection 2.0.3, introduces the AHP method but is not utilized in the research. Instead, the novel method relies on the company's Analytics Maturity assessment to measure the maturity of resources and capabilities. As a result, some steps of the method are overlooked, and the Global reference architecture patterns for Heineken's context are not fully explored. Additionally, due to time limitations and a focus on domain-specific problems, certain steps of the TOGAF ADM are not described. Consequently, the reference architecture artifact is well articulated in terms of integrating cross-technical functional views. Still, the solution is limited in the comprehensive coverage for business stakeholders and the change management approach necessary for product deployment.

Thirdly, although the model partially represents the significant processes and views related to the Systematic Literature Review, complexities that emerged during real-context validation (such as stakeholder roles) should have been considered, as incorporating them would have increased the complexity of the views. Moreover, while the DBAC deployment architecture incorporates core technical building blocks, data transformation, and data engineering processes, there is a need for further validation in terms of validating the artifact with business stakeholders, end-users, or local operational countries. Lastly, limited validation was conducted with Heineken's enterprise architecture professionals, who play a vital role in real-world analytical deployment processes. This limitation arose due to time constraints and the specific domain expertise required. However, the present thesis supervisor was guided by the solution architect of Heineken's advanced analytics product that provided an extensive context of the deployment and architecture process.

7.4 Discussion & Future work

Future research is proposed in the BDAC deployment architecture to improve the different views and levels in order to improve the deployment of analytical products. Further research initiatives are proposed towards:

1. To ensure effective management and communication of architecture changes in a business analytical process deployment, it is recommended to implement additional steps in the TOGAF ADM and implement Multi-Criteria and Model-Based Analysis for Project Selection. These steps will facilitate the proper articulation of change management and communication initiatives, leading to a more impactful deployment.

2. As discussed in the previous subsection, the inclusion of Analytics Maturity Assessment (MA) or the MLOps model are a unique feature of Heineken's MA. However, the questions and mapping process used to measure the evolutionary dimension of Big data Analytical Capability through the MA is innovative. In particular, the selection of questions to measure analytical maturity levels and map the ability to utilize BD analytics resources in future academic work. (refer to Subsection: 3.5).

3. The deployment architecture presented in this study is based on SRL and open MlOps and DevOps model design patterns that can be applied to any function, organization, or business analytical process. As such, the architecture views are assessed as "industry agnostic" characteristic of the present deployment architecture levels. As such, further research is recommended to explore applying the methods, frameworks, and architectural views presented in this study to different fields and research and industry domains.

7.5 Lessons Learned

Throughout the thesis development, multiple leanings have been made, particularly combined with the opportunity to work in Global Analytics teams in Heineken for almost a year. The following are the most important:

1. **A number says everything and nothing at the same time:** As the different operating companies were evaluated through the common GA Analytics maturity assessment, the results revealed that less mature countries tended to have higher maturity scores compared to more mature countries. A more insightful review on the reason for these indicated that as the most mature countries gets more mature in terms of analytics, it was found that "The more you know, the more you realize you don't know."(Aristotle). A possible hypothesis is that as the more mature countries matures, the stricter they were on their self evaluation versus the less mature but higher "maturity" analyitics countries. A further comparative analysis is need it beyond the direct score results and articulate architectures or artifacts to guide the digital transformations processes .

2. **The opportunity is on the People resource:** Within the Heineken global team context, the case evidence the enormous amount of data is present through the different countries, with a talented team and state-of-the-art technology. However, in my experience within the global team, the main challenges in creating value from data are communication and change management. The former is present in the structure/ communication complexity and processes within the different countries, global and cross-functional teams. The latest is present in the nature of humans to change, and difficult to accept new processes that affect their status quo. What is the best way to communicate or at least define and understand the essential processes with a clear blueprint deployment or architecture The artifact evaluations results indicates a future opportunity to integrate the technical and business process through a common communication artifact/tool.

3. **More is not better:** This lesson, as in life, was found quite surprising from a set of interviews and related to the previous lesson. The advanced analytics solutions have proven their value to the different organizational functions/ departments. Nonetheless, as the same analytical products and solutions might create an overall benefit to the organization but at the same time affect its final users incentives. For instance, the final user of an analytical product might affect its historic KPIs results for which they are currently being measured. As the KPIs might not be met or affect the overall results, the user is disincentive to use or test the new Analytical product. The question is, how to speed up the digital transformations processes and incentives accordingly to support the growth in the use of advanced Analytical products in our people? and how to incorporate these insights in the architecture processes and KPIs as non- technical requirements?

Bibliography

[1] Togaf standard, version 9.2. 2017. URL https://pubs.opengroup.org/architecture/togaf9-doc/arch/.

[2] Analysis of digital business transformation tools. 2019.

[3] Definition of advanced analtyics. 2022. URL https://www.gartner.com/en/information-technology/glossary/advanced-analytics.

[4] B. Adhytia and R. Dilnutt. Investigating enterprise architecture practice for big data analytics implementation. 2022. URL https://aisel.aisnet.org/pacis2022/192/.

[5] B. Adhytia, R. Dilnutt, and A. N. Hidayanto. Investigating enterprise architecture practice for big data analytics implementation. *AIS*, 2022. URL https://aisel.aisnet.org/pacis2022/192/.

[6] A. Adina, M.-E. Iacob, M. Daneva, and L. Helmy Masyhur. Multi-criteria and model-based analysis for project selection: An integration of capability-based planning, project portfolio management and enterprise architecture. *IEEE*, 2019. doi: DOI:10.1109/EDOCW.2019.00032. URL https://ieeexplore.ieee.org/document/8907326.

[7] P. Ahlawat, J. Borgman, S. Eden, S. Huels, J. Iandiorio, A. Kumar, and P. Zakahi. A new architecture to manage data costs and complexity. *Boston Consulting*

Group (BCG), pages 1–12, 2023. URL https://www.bcg.com/publications/2023/new-data-architectures-can-help-manage-data-costs-and-complexity.

[8] S. Akter, S. Fosso Wamba, A. Gunasekaran, R. Dubey, C. Rameshwar, and Stephen. How to improve firm performance using big data analytics capability and business strategy alignment? 2016. doi: http://dx.doi.org/10.1016/j.ijpe.2016.08.018. URL https://www.sciencedirect.com/science/article/pii/S0925527316302110.

[9] A. Aldea. Enterprise strategic alignment method: a cross-disciplinary capability-driven approach. 2017. doi: 10.3990/1.9789036543309. URL https://research.utwente.nl/en/publications/enterprise-strategic-alignment-method-a-cross-disciplinary-capabi.

[10] D. Arora and P. Malik. Analytics: Key to go from generating big data to deriving business value. *IEEE*, page 7, 2015. doi: 10.1109/BigDataService.2015.62. URL https://ieeexplore.ieee.org/document/7184914.

[11] C. Azevedo, J. P. Almeida, M. van Sinderen, and L. Ferreira Pires. Towards capturing strategic planning in ea. *IEEE*, 2015. doi: DOI10.1109/EDOC.2015.31. URL https://ieeexplore.ieee.org/document/7321168/.

[12] C. Azevedo, M.-E. Iacob, J. Almeida, M. Hvan Sinderen, L. Ferreira Pires, and G. Guizzardi. Modeling resources and capabilities in enterprise architecture: A well-founded ontology-based proposal for archimate. *ELSEVIER*, 2015. doi: http://dx.doi.org/10.1016/j.is.2015.04.008. URL https://www.sciencedirect.com/science/article/pii/S0306437915000745.

[13] M. Bilal, L. O. Oyedele, H. Kusimo, H. Owolabi, , L. Akanbi, A. Ajayi, O. Akinade, and J. M. Davila Delgado. Investigating profitability performance of construction projects using big data: A project analytics approach. 2019. doi: https://doi.org/10.1016/j.jobe.2019.100850. URL https://www.sciencedirect.com/science/article/pii/S2352710219302463?via%3Dihub.

[14] A. Cakir, E. Ozkaya, F. Akkus, E. Kucukbas, and O. Yilmaz2. Real time big data analytics for tool wear protection with deep learning in manufacturing industry. *Springer*, pages 148–155, 2022. doi: https://doi.org/10.1007/978-3-031-09176-6_18. URL https://link.springer.com/chapter/10.1007/978-3-031-09176-6_18.

[15] M. . Company. The age of analytics: Competing in a data-driven world. *McKinsey Global Institute*, page 136, 2016. URL https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-age-of-analytics-competing-in-a-data-driven-world.

[16] B. Darvish Rouhani, M. Naz'ri Mahrin, F. Nikpay, R. Binti Ahmad, and P. Nikfard. A systematic literature review on enterprise architecture implementation methodologies. *ELSEVIER*, 2015. doi: https://doi.org/10.1016/j.infsof.2015.01.012. URL https://www.sciencedirect.com/science/article/pii/S0950584915000282.

[17] Y. Demchenko, C. de Laat, and P. Membrey. Defining architecture components of the big data ecosystem. 2014. doi: DOI:10.1109/CTS.2014.6867550. URL https://pure.uva.nl/ws/files/49579691/bddac2014_bd_ecosystem_archi_v05.pdf.

[18] A. Elsheikh. Blockchain analytics reference architecture for fintech. 2022. doi: https://doi.org/10.1145/3531056.3531068. URL https://dl.acm.org/doi/fullHtml/10.1145/3531056.3531068.

[19] F. N. . S. L. Erevelles, S. Big data consumer analytics and the transformation of marketing. page 7, Feb 2016. doi: https://doi.org/10.1016/j.jbusres.2015.07.001. URL https://www.sciencedirect.com/science/article/abs/pii/S0148296315002842.

[20] I. Gartner. Gartner survey shows organizations are slow to advance in data and analytics. Feb 2018. URL https://www.gartner.com/en/newsroom/press-releases/2018-02-05-gartner-survey-shows-organizations-are-slow-to-advance-in-data-and-ana

[21] E. GERLEIN1, M. Mcginnity, A. Belatreche, and S. Coleman. Evaluating machine learning classification for financial trading: An empirical approach. 2016. doi: DOI:10.1016/j.eswa.2016.01.018. URL https://www.sciencedirect.com/science/article/abs/pii/S0957417416000282?via%3Dihub.

[22] Y. Gong and J. Marijn. Roles and capabilities of enterprise architecture in big data analytics technology adoption and implementation. *Journal of Theoretical and Applied Electronic Commerce Research*, 2021. doi: DOI:10.4067/S0718-18762021000100104. URL https://www.mdpi.com/0718-1876/16/1/4.

[23] R. Grossman. A framework for evaluating the analytic maturity of an organization. *ELSERVIER*, 2018. doi: https://doi.org/10.1016/j.ijinfomgt.2017.08.005. URL https://reader.elsevier.com/reader/sd/pii/S0268401217300026?token=AFABE506DCA1BBF058CD69B2106153A8893E4FA8211B1225E3579F02E17697850B5E59449A37EE87D1 originRegion=eu-west-1&originCreation=20230109212759.

[24] V. Grover, R. Chiang, T.-P. Liang, and D. Zhang. Creating strategic business value from big data analytics: A research framework. 2018. doi: https://doi.org/10.1080/07421222.2018.1451951. URL https://www.tandfonline.com/doi/abs/10.1080/07421222.2018.1451951?journalCode=mmis20.

[25] M. Gupta and J. F. George. Toward the development of a big data analytics capability. *ELSERVIER*, pages 1049–1064, 2016. doi: https://doi.org/10.1016/j.comcom.2021.05.008. URL https://www.sciencedirect.com/science/article/pii/S0378720616300787.

[26] D. Halper, F.; Stodder. Tdwi analytics maturity model guide. interpreting your assessment score. page 23, 2020. URL https://go.tdwi.org/rs/626-EMC-557/images/TDWI_Analytics-Maturity-Model-Assessment-Guide_2020.pdf.

[27] heineken. Heineken n.v. annual report 2021. 2020. URL https://www.

theheinekencompany.com/sites/theheinekencompany/files/Downloads/PDF/AR-2021/heineken-nv-annual-report-2021-25-02-2022.pdf.

[28] heineken. Heineken n.v. reports on 2023 first-quarter trading. 2023. URL https://www.theheinekencompany.com/newsroom/heineken-nv-reports-on-2023-first-quarter-trading/.

[29] R. Hribar. Overview of business intelligence maturity models. 2010.

[30] I. Ilin, S. Maydanova, W. Kersten, C. Jahn, and J. Weigell. Digitalization of global container shipping lines. *Springer*, 2022. URL https://link.springer.com/chapter/10.1007/978-3-030-92291-7_8.

[31] Intel. Getting started with advanced analytics: How to move forward with a successful deployment. intel:planning guide. 2017. URL https://www.intel.com/content/www/us/en/artificial-intelligence/getting-started-advanced-analytics-planning-guide.html.

[32] H. Issa, R. Jabbouri, and M. Palmer. An artificial intelligence (ai)-readiness and adoption framework for agritech firms. *ELSERVIER*, page 17, 2022. doi: https://doi.org/10.1016/j.techfore.2022.121874. URL https://www.sciencedirect.com/science/article/pii/S0040162522003985.

[33] C. Kearny, A. Gerber, and A. v. d. Merwe. Data-driven enterprise architecture and the togaf adm phases. 2016. doi: DOI:10.1109/SMC.2016.7844957. URL https://ieeexplore.ieee.org/document/7844957.

[34] K. Kenda, N. Mellios, M. Senožetnik, and P. Pergar. Computer architectures for incremental learning in water management. *MDPI*, page 18, 2022. doi: https://doi.org/10.3390/su14052886. URL https://ieeexplore.ieee.org/document/7184914.

[35] B. Kitchenham and S. Charters. Guidelines for performing systematic literature

reviews in software engineering. 2007. URL https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf.

[36] T. Koehler and S. Alter. Using enterprise architecture to attain full benefits from corporate big data while refurbishing legacy work systems. *Proceedings of CBI*, 2016. URL https://ceur-ws.org/Vol-1753/paper1.pdf.

[37] M. Koot, M.-E. Iacob, and M. K. Mes. A reference architecture for iot-enabled dynamic planning in smart logistics. 2021. URL https://research.utwente.nl/en/publications/a-reference-architecture-for-iot-enabled-dynamic-planning-in-smar.

[38] D. Kreuzberger, N. Kühl, and S. Hirschl. Machine learning operations (mlops): Overview, definition, and architecture. 2022. doi: https://doi.org/10.48550/arXiv.2205.02302. URL https://arxiv.org/pdf/2205.02302.pdf.

[39] Z. D. Krol, K. Analytics maturity models: An overview. page 19, 2020. doi: https://doi.org/10.3390/info11030142. URL https://www.mdpi.com/2078-2489/11/3/142/htm.

[40] K. Król and D. Zdonek. Analytics maturity models: An overview. *MDPI AG*, 11, 2020. doi: 10.3390/info11030142. URL https://www.mdpi.com/2078-2489/11/3/142.

[41] I. Laguir, S. Gupta, I. Bose, R. Stekelorum, and L. Laguir. Analytics capabilities and organizational competitiveness: Unveiling the impact of management control systems and environmental uncertainty. *ELSEVIER*, 2022. doi: https://doi.org/10.1016/j.dss.2022.113744. URL https://www.sciencedirect.com/science/article/pii/S016792362200015X.

[42] M. Lankhorst. Enterprise architecture at work. 2013. doi: https://doi.org/

10.1007/978-3-642-29651-2. URL https://link.springer.com/book/10.1007/978-3-642-29651-2.

[43] S. LaValle, E. Lesser, R. Shockley, M. Hopkins, and N. Kruschwitz. Big data, analytics and the path from insights. 2011. URL https://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to-value/.

[44] P. Maltaa and R. D. Sousab. Process oriented approaches in enterprise architecture for business-it alignment. 2016. doi: doi:10.1016/j.procs.2016.09.239. URL https://www.sciencedirect.com/science/article/pii/S1877050916324085.

[45] I. Merete Enholm, E. Papagiannidis, P. Mikalef, and J. Krogstie. Artificial intelligence and business value: a literature review. 2018. URL https://www.tandfonline.com/doi/abs/10.1080/07421222.2018.1451951?journalCode=mmis20.

[46] P. Mikalef, I. Pappas, J. Krogstie, and M. Giannakos. Big data analytics capabilities: a systematic literature review and research agenda. *Springer*, 2017. URL https://link.springer.com/article/10.1007/s10257-017-0362-y.

[47] G. Muan Sang, L. X. Xu, and P. de Vrieze. Simplifying big data analytics systems with a reference architecture. 2017. URL https://core.ac.uk/download/pdf/83943361.pdf.

[48] I. Parra and G. Arroyo. Electric utility enterprise architecture to support the smart grid - enterprise architecture for the smart grid. 2014. doi: DOI:10.5220/0005014006730679. URL https://www.scitepress.org/Link.aspx?doi=10.5220/0005014006730679.

[49] S. Pathak, V. Krishnaswamy, and M. Sharma. Big data analytics capabilities: a novel integrated fitness framework based on a tool-based content analysis. *Taylor Francis*, 2021. doi: https://doi.org/10.1080/17517575.2021.1939427. URL https://www.tandfonline.com/doi/full/10.1080/17517575.2021.1939427.

[50] T. Pearson and W. Rasmus. Big data: The organizational challenge. 2013. URL https://media.bain.com/Images/BAIN_BRIEF_Big_Data_The_organizational_challenge.pdf.

[51] K. Peffers, T. Tuunanen, and S. Rothenberger, Marcus Chatterjee. A design science research methodology for information systems research. 2014. doi: https://doi.org/10.2753/MIS0742-1222240302. URL https://www.tandfonline.com/doi/epdf/10.2753/MIS0742-1222240302?needAccess=true.

[52] W. Roel J. Design science methodology. 2014. doi: https://doi.org/10.1007/978-3-662-43839-8. URL https://link.springer.com/book/10.1007/978-3-662-43839-8.

[53] J. Ross and C. Beath. Why you — yes, you — need enterprise architecture. 2020. URL https://sloanreview.mit.edu/article/why-you-yes-you-need-enterprise-architecture/.

[54] J. Ross, C. Beath, and A. Quaadgras. You may not need big data after all. *Harvard Business Review*, page 15. URL https://hbr.org/2013/12/you-may-not-need-big-data-after-all.

[55] K. Shadrack. Enterprise architecture – a value proposition for records professionals. *IEEE*, 2019. doi: DOI:10.1109/BigData47090.2019.9006603. URL https://ieeexplore.ieee.org/document/9006603.

[56] Statista. Beer - worldwide. 2022. URL https://www.statista.com/outlook/cmo/alcoholic-drinks/beer/worldwide.

[57] S. Suoniemi, L. Meyer-Waarden, A. Munzel, and A. R. Zablah. Big data and firm performance: The roles of market-directed capabilities and business strategy. *ELSERVIER*, 2020. doi: https://doi.org/10.1016/j.im.2020.103365. URL https://www.sciencedirect.com/science/article/pii/S0378720620303037.

[58] T. Tamm, P. Seddon, G. Shanks, and P. Reynolds. How does enterprise architecture add value to organisations? 2011. doi: https://doi.org/10.17705/1CAIS.02810. URL https://aisel.aisnet.org/cais/vol28/iss1/10/.

[59] L. Tamym, L. Benyoucef, A. Nait Sidi Moh, and M. Driss El Ouadghiri. A big data based architecture for collaborative networks: Supply chains mixed-network. *ELSERVIER*, pages 201–111, 2021. doi: https://doi.org/10.1016/j.comcom.2021.05.008. URL https://www.sciencedirect.com/science/article/pii/S0140366421001924.

[60] R. Teichert. Digital transformation maturity: a systematic review of literature. 2019. doi: https://doi.org/10.11118/actaun201967061673. URL https://www.researchgate.net/publication/338110138_Digital_Transformation_Maturity_A_Systematic_Review_of_Literature.

[61] Y. Wang, L. Kung, and T. A. Byrd. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *ELSERVIER*, page 11, 2018. doi: https://doi.org/10.1016/j.techfore.2015.12.019. URL https://www.sciencedirect.com/science/article/pii/S0040162516000500.

[62] T. M. B. D. F. P. M. A. Westerman, G. The digital advantage: how digital leaders outperform their peers in every industry. 2011. URL https://www.capgemini.com/wp-content/uploads/2017/07/The_Digital_Advantage__How_Digital_Leaders_Outperform_their_Peers_in_Every_Industry.pdf.

[63] S. A. K. V. Zaychenko, I. Application of digital technologies in human resources management at the enterprises of fuel and energy complex in the far north. 2019. doi: https://doi.org/10.1007/978-3-030-19868-8_33. URL https://link.springer.com/chapter/10.1007/978-3-030-19868-8_33.

[64] B. Zheng Zhang, A. Ashta, and M. E. Barton. Analytics: Key to go from generating big data to deriving business value. *Wiley*, page 12, 2021. doi: https://doi.org/10.

1002/jsc.2405. URL https://onlinelibrary.wiley.com/doi/full/10.1002/jsc.2405.

[65] A. Zimmermann, M. Pretz, G. Zimmermann, D. G. Firesmith, I. Petrov, and E. El-Sheikh. Towards service-oriented enterprise architectures for big data applications in the cloud. 2013. URL https://ieeexplore.ieee.org/document/6690543.

[66] A. Zimmermann, R. Schmidt, K. Sandkuhl, K. Jugel, M. Möhring, and M. Wißotzki. Enterprise architecture management for the internet of things. 2015. URL https://subs.emis.de/LNI/Proceedings/Proceedings244/139.pdf.

Appendix A

Appendix

A.1 Maturity Assessments Frameworks

A.1.1 Capability Maturity Model Integration-CMMI

The **CMMI model** first version was published in 2010, by the CMMI institute and developed by Carnegie Mellon University (CMU). The CMMI serves as a framework for evaluating the process maturity of the organization, as well as an standard structure with a sequence of improvements. Additionally, CMMI offers guidance for creating, comparing and enhancing processes that align with the business objectives. Currently the model maturity levels are: (1) Initial, (2) Managed, (3) Defined, (4) Managed, and (5) Optimized.

1. **Ad hoc or Unknown**: tasks may or may not get completed.
2. **Initial**: Processes are reactive and might be completed, but they are poorly controlled or delayed.
3. **Managed**: Projects are proactive, planned, deploy, measured and controlled.
4. **Defined**: Organization provide guided standards provide across projects, programs, and portfolios.
5. **Qualitative Managed**: Data drive Organizations with quantitative performance improvement objectives that are predictable and aligned to meet internal and external stakeholders' needs.
6. **Optimizing**: The organization is focused on continuous improvement and building and responding to change and opportunities and change. The organization's stability and maturity provide a platform for agility and innovation improvements.

A.1.2 Analytic Processes Maturity Model-APMM

The Analytic Processes Maturity Model is an organizational analytics maturity framework based on the following analytic concepts: models, infrastructure and operations. This model focus on measuring six (6) maturity process for software Analytics developments: (1) building models; (2) deploying models; (3) managing and operating infrastructure; (4) protecting assets through appropriate policies and procedures; (5) operating governance structure; and (6) identifying opportunities, making decisions, and allocating resources based upon an analytic strategies. The following are the five APMM organization maturity levels:

1. **Building Reports**
2. **Building & Deploying Models**
3. **Building & Deploying analytics**
4. **Consistent Analytics enterprise-wide processes**
5. **Enterprise Analytics driven Strategy**

A.1.3 Analytics Processes Maturity Model-APMM

The Analytic Processes Maturity Model is an organizational analytics maturity framework based on the following analytic concepts: models, infrastructure and operations. This model focus on measuring six (6) maturity process for software Analytics developments: (1) building models; (2) deploying models; (3) managing and operating infrastructure; (4) protecting assets through appropriate policies and procedures; (5) operating governance structure; and (6) identifying opportunities, making decisions, and allocating resources based upon an analytic strategies [23]. The following are the five APMM organization maturity levels:

1. **Building Reports**
2. **Building and Deploying Models**
3. **Building and Deploying analytics**
4. **Consistent Analytics enterprise-wide processes**
5. **Enterprise Analytics driven Strategy**

A.1.4 Blast Analytics Maturity Assessment Framework

The Blast analyitics matury assessment measures six key process areas and success factors variables: resources, data management, governance, strategy, insights and evolution. Each variable is evaluated from 1 to six levels and the organization is placed in one of the five development stage: Laggard, Follower, Competitor, Leader and Innovator. Additionally, the assessment propose a quarterly survey to assess and measure the Analytics implementation(Benchmark) progress and the current condition (Strategic Roadmap).

Base on the Strategy, an Action Plan is develop to ensure the objectives fulfillment. [39],

Figure A.1: Stages of analytical maturity (Blast Model)[39]

,

A.1.5 Gartner's Maturity Model for Data and Analytic

Gartner's Maturity Model for Data and Analytics is based on the following five levels [20] (Please refer to Figure A.2 :

1. **Level 1:Basic** data are not exploited, D&A (Data Analytics) is managers in silos, people arguing about whose data are correct, analysis is ad hoc, spreadsheet and information firefighting.

2. **Level 2:Opportunistic** IT attempts to formalize information availability requirements, inconsistent incentives. Organizational requirements, Organizational barriers and lack of leadership, not business-relevant, data quality and insight efforts, but still siloed to one to another.

3. **Level 3: Systematic** different content types are still treated differently, strategy and vision formed (five pages), agile emerges, exogenous data sources are readily integrated, business executives become D&A champions.

4. **Level 4: Differentiating** executives champions and communicate best practices; business-led/driven, with chief data office (CDO); D&A is an indispensable fuel for performance and innovations, and linked across programs; link to outcome and data used for ROI (return on investment).

5. **Transformational** D&A is central to business strategy, data value influences investments, strategy and execution aligned and continually improved, Outside-in perspective, CDO sits on board.

A.2 MLOps Architecture

A.2.1 Global Analytics: Maturity Assessment

The Global Maturity Assessment methodology is a systematic series of steps to assess, measure, and represent local Data and Analytics perceptions regarding opportunities, expectations, and commitments toward future analytics infrastructures and initiatives. GA-MA serves as a D&A perception benchmark against strategic Analytics maturity

Figure A.2: Gartner's Maturity Model for Data and Analytics

knowledge and processes. The Methodology process (Refer to figure:**??**) is divided into the following (7) seven phases: (1) **Country Selection**; (2) **Introduction Meeting**; (3) **Take the Assessment**; (4) **Review of the outcomes**; (5) **Workshop** (6) **Action Plan**; and, (7) **Retake the Assessment**. The process currently takes two to three months and will involve the Global Analytics Team, local Data and technology-analytic leader, and multiple participants from every department (Please refer to Figure: A.3).

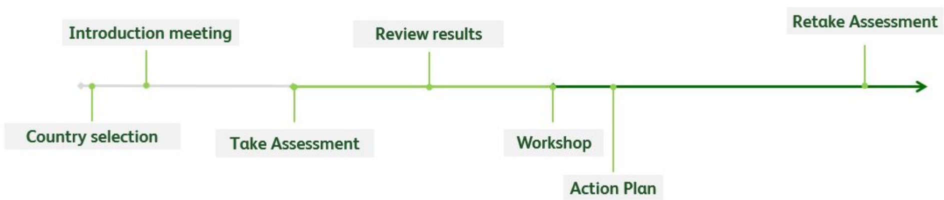


Figure A.3: Data and Analytics Maturity Assessment process

1. **Country Selection** is made for more than 90 countries based on current and future Digital projects and current IT landscape development. Typically, this process is aligned with the different regional leadership that introduces the project to the different local D&T stakeholders and assesses the timing and approach to each country. For instance, particular countries might be deploying multiple projects that require most of the time and talent resources, or local business context might limit access to particular countries.

2. **Introduction Meeting** sets in emotion the Maturity Assessment for each country. In this meeting, the assessment's purpose, goals, process, common questions, and next steps are introduced. Usually, the local D&T representative, together with the global team, is usually present for the thirty-minute meeting. Some of the most common questions are regarding the target assessment's survey and workshop participants, the different types of internal assessment communication, Alignment with current digital projects, the technical survey questions and target audience, and the next steps. Finally, the tentative survey and assessment's date is defined with a target participants list from multiple departments.

3. **Take the Assessment-survey** happens at the global and local team pre-defined launch date, where the participants receive an invitation mail with the steps and link to answer the survey's questions. In this process, the global and the local team maintain fluid communication with the total day-to-day participants. If the participation is different than expected, multiple actions might be recommended, such as extending the survey time range, additional email communication, High management communication support, or using internal communications tools. Additionally, setting a proper survey launch date is crucial to the participant's rate, as typically, the different countries possess multiple Climate-training Surveys, end-of-the-month sales closing, or yearly financial reports. By 2023, the survey's questions will have been translated into four languages: Portuguese (Brazil), Korean, Spanish, and Mandarin. (Please refer to Figure: A.4)

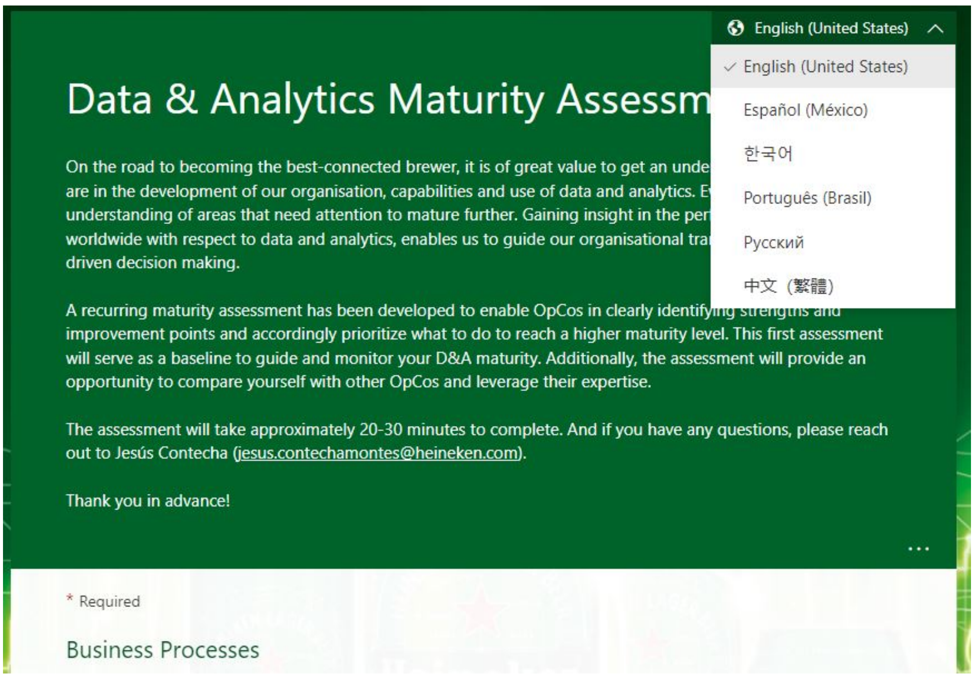


Figure A.4: Sanitized Data and Analytics Maturity Assessment survey

4. **Review the outcomes** occurs a couple of weeks after the participation survey is closed. In this step, the Global and Local D&T team reviews the survey results(Dashboard), analyzes the general overview results report, and defines the

Workshop's strategic priority topics, participants, and tentative dates. For example, horizontal analysis is based on comparing the different answers to questions from different sections by department insights and opportunities. The primary outcome of the meeting is the priority topics definition that will be presented and improved in the following phases (Please refer to Figure: A.5).

Figure A.5: Sanitized D&A Maturity Assessment dashboard

5. **Workshop** represents the qualitative assessment phase, where the results and priorities defined in the previous quantitative phase are used as insights to brainstorm improvement ideas and actions. Regularly local participants or representatives from multiple departments, the Global Analytics team, and local D&T teams participate in the activity. In other words, the Workshop works as a joint interdepartmental space for professionals to "spark" the conversation about actions that leads toward higher D&A maturity levels (Please refer to Figure: A.6).

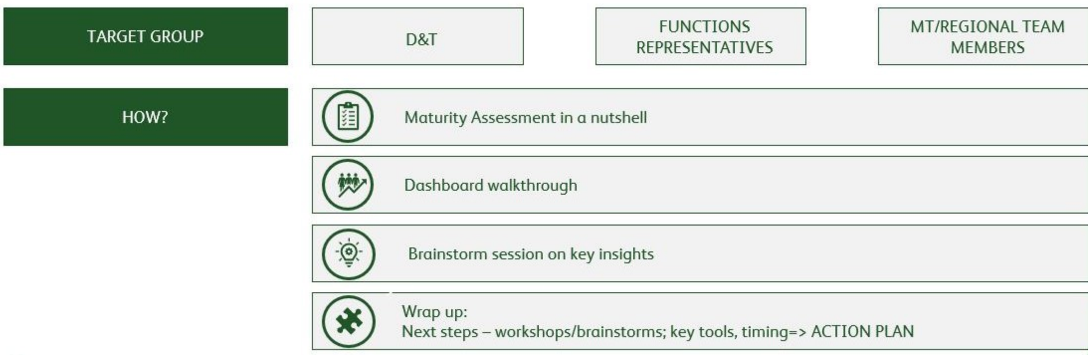


Figure A.6: Sanitized D&A Workshop

6. **Action Plan** is the last methodology phase, the local D&A leader constructs an actionable list of measures to improve the D&A Maturity Levels. Typically, both the

assessment's quantitative (Survey) and qualitative(Workshop) insights are used to fill the list together with follow up meetings and high management support. Figure A.7 represents an optimal framework for the local countries to define the different actions reasons, actions, goals, KPIs, owners, and dates.

Figure A.7: Sanitized Action Plan

7. **Retake the Assessment** yearly, to measure the impact of the different action plan measures and the current impact in terms of higher Data and Analytics maturity levels for multiple departments.

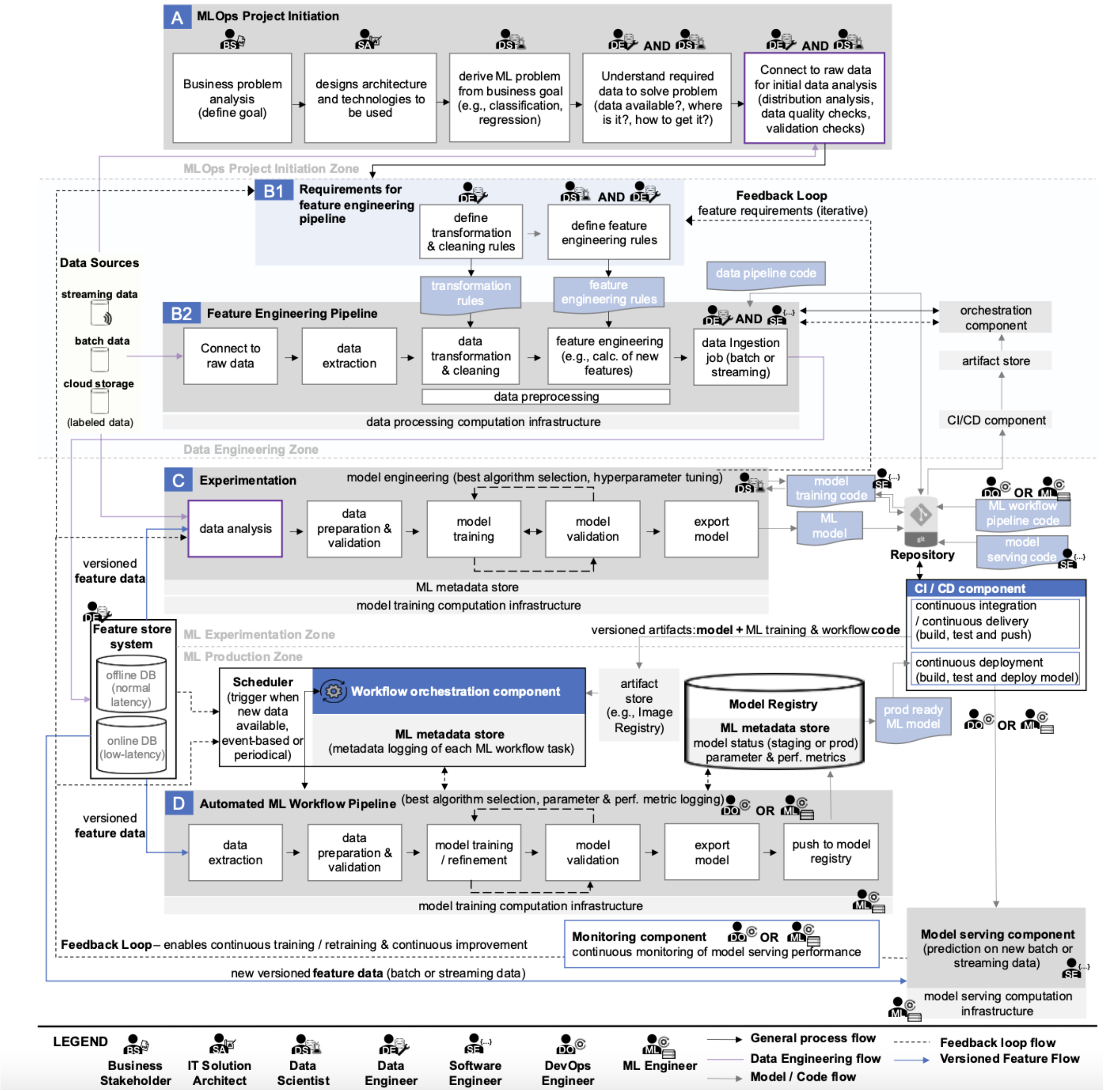


Figure A.8: End-to-end MLOps architecture and workflow with functional components and roles [38]