NOVA IMS Information Management School

MGI

Mestrado em Gestão de Informação
Master Program in Information Management

COMPARATIVE STUDY OF IMPLEMENTING THE ON-PREMISES AND CLOUD BUSINESS INTELLIGENCE ON BUSINESS PROBLEMS IN A MULTI-NATIONAL SOFTWARE DEVELOPMENT COMPANY

Alexandre Nicolau Jdanok Vassilenko

Internship report
presented as partial requirement for obtaining the
Master's degree in Statistics and Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

2019

Title: COMPARATIVE STUDY OF IMPLEMENTING THE ON-PREMISES AND CLOUD BUSINESS INTELLIGENCE ON BUSINESS PROBLEMS IN A MULTI-NATIONAL SOFTWARE DEVELOPMENT COMPANY

Alexandre Nicolau Jdanok Vassilenko

MGI

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

COMPARATIVE STUDY OF IMPLEMENTING THE ON-PREMISES AND CLOUD BUSINESS INTELLIGENCE ON BUSINESS PROBLEMS IN A MULTI-NATIONAL SOFTWARE DEVELOPMENT COMPANY

by

Alexandre Nicolau Jdanok Vassilenko

Internship report presented as a partial requirement for obtaining the Master's degree in Information Management, with a specialization in Knowledge Management and Business Intelligence

Advisor Flavio Pinheiro, PHD

June 2022

ABSTRACT

Nowadays every enterprise wants to be competitive. In the last decade, the data volumes are increased dramatically. As each year data in the market increases, the ability to extract, analyze and manage the data become the backbone condition for the organization to be competitive.

In this condition, organizations need to adapt their technologies to the new business reality in order to be competitive and provide new solutions that meet new requests. Business Intelligence by the main definition is the ability to extract analyze and manage the data through which an organization gain a competitive advantage. Before using this approach, it's important to decide on which computing system it will base on, considering the volume of data, business context of the organization and technologies requirements of the market.

In the last 10 years, the popularity of cloud computing increased and divided the computing Systems into On-Premises and cloud. The cloud benefits are based on providing scalability, availability and fewer costs. On another hand, traditional On-Premises provides independence of software configuration, control over data and high security. The final decision as to which computing paradigm to follow in the organization it's not an easy task as well as depends on the business context of the organization, and the characteristics of the performance of the current On-Premises systems in business processes. In this case, Business Intelligence functions and requires in-depth analysis in order to understand if cloud computing technologies could better perform in those processes than traditional systems.

The objective of this internship is to conduct a comparative study between 2 computing systems in Business Intelligence routine functions. The study will compare the On-Premises Business Intelligence Based on Oracle Architecture with Cloud Business Intelligence based on Google Cloud Services. A comparative study will be conducted through participation in activities and projects in the Business Intelligence department, of a company that develops software digital solutions to serve the telecommunications market for 12 months, as an internship student in the 2nd year of a master's degree in Information Management, with a specialization in Knowledge Management and Business Intelligence at Nova Information Management School (NOVA IMS).

KEYWORDS

Business Intelligence; On-Premises; Cloud; SaaS; PaaS;IaaS;BI SaaS;BI PaaS; Oracle; Exadata

INDEX

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

BA PaaS	Business Analytic Platform as a Service
BI	Business Intelligence
CPU	Central Processing Unit
DW	Data Warehouse
GCP	Google Cloud Platform
ETL	Extract Transform Load
IaaS	Infrastructure as a Service
IT	Information Technologies
OLAP	Online Analytical Processing
OOP	Object-Oriented Programming
PaaS	Platform as a Service
PL/SQL	Procedural Language/Structured Query Language
SaaS	Software as a Service
SQL	Structured Query Language
VPN	Virtual Private Network

1. INTRODUCTION

In the last years, the data volumes are increased dramatically. As each year data in the market increases, the ability to extract, analyze and manage the data become the backbone condition for the organization to be competitive.

At the current moment, when the system of the world is changing politically, and economically the uncertainty, and risks are highest than ever. The enterprises need to consider it as the new reality. The nature and the structure of the current dynamic world cause that nowadays, in times of uncertainty, risks and incomplete information, the crisis becomes a feature of modern business, not a state of emergency. (Korol, T., Korodi, A, 2010) Organizations in this way, need to adapt and be more agile to external events and have the necessary infrastructure that gives this capability. Today, the world, including the market, is becoming more data-driven and given that, any event that occurs generates information. Information is a strategic resource for companies, and decisions must be taken based on a huge amount of real-time information, from a high variety of internal and external sources, unstructured and structured sources. (Muntean, 2015)

According to Craig le Clair: there are 3 categories of main factors that influence business agility: Organization (knowledge dissemination, digital psychology and change management), marketing (market responsiveness and channel integration) and Information Technologies (IT): (Business Intelligence (BI), infrastructure elasticity, business processes architecture, software innovation and sourcing and supply chain).

IT Technologies in this way play an important role not only in the context of business agility but also in strategic points. As the current economic crisis is so uncertain, at this condition the Information Technology industry is oriented to effectiveness by creating new models of provision, management and IT security. Given this, the enterprises start to re-creating flexibility and strong sensitivity to the environment and the use of strategic systems in real-time. Enterprises need to have a good approach and methodology to use IT technologies to better and more effectively way process and analyse information and Business Intelligence is one of them.

Business Intelligence refers to computer-based techniques used in spotting, digging out, and analysing business data, such as sales revenue by products and/or departments, or by associated costs and incomes. BI technologies provide historical, current, and predictive views of business operations. Ceborarian, E. (2008) Business Intelligence is a process to exploit the available data to generate information and knowledge useful for complex decision-making processes.BI creates value for organizations by using data or facts and increases capabilities of decision-making for managerial processes by providing Online Transaction Processing (OLAP) through interactive dashboards, using the data in the form of facts.

However, traditional Business Intelligence solutions are usually implemented in organizations as enterprise software. In this way BI have enterprise software properties as software solutions that are installed and supported on the premises by its own IT team; acquisition and implementation of the software are also expensive, in addition to the costs with the employees. In the Future, BI solutions could require the implementation and purchase together with the integration of multiple software components, such as software for extracting data, data integration, storage, analysis, and reporting.

(Traditional Enterprise Business Intelligence Software Compared to Software as a Service Business Intelligence).

Given this, with Traditional Business Intelligence, organizations have a risk to fail to create innovative BI solutions and implement the new models of provision, and strategic management that influences business agility. To fulfil the requirements of organizations, on the market appeared new solutions based on Cloud Computing, called Cloud BI. Cloud-based BI refers to BI SaaS (BI software as a service), BI for PaaS (platform as a service), BI for SaaS and BA PaaS (business analytic platform as a service).

According to the NIST (National Institute of Standards and Technology) and CSA (Cloud Security Alliance) definition: Cloud computing is a model that provides on-demand network access to a common reserve of configurable computing resources (for example, networks, servers, storing, applications and services). Cloud BI brings the components and capabilities of Traditional Business Intelligence (Data Warehouse, ETL Tools, Analytical Tools, Reporting Tools, business performance management tools and applications), and makes it accessible via the cloud as a service. Depending on deployment models for cloud-based BI (public, private) and company business objectives any combination of Cloud BI components is possible. Cloud BI are much more flexible than traditional BI solutions, which gives flexibility to the organization to adapt to economic changes. Integration of a Cloud BI solution has a special interest for organizations that desire to improve agility while at the same time reducing IT costs and exploiting the benefits of Cloud Computing. (Bowen, F., 2009)

But at the same time implementing Cloud BI in the organization is a complex task and not all organizations do it. Initially, cloud solution was designed for small and medium organizations to improve business agility at low cost. In the last years, large organizations trying to implement Could BI, but the components remain for small and medium-sized organizations. As a result, the use of a Cloud Computing based solution is not suited for all organizations, especially in the field of BI. (Wayne, W. E. (2009).

The success of implementing Cloud BI depends on the organization's business, current processes and how the main benefits of the Cloud can improve the company, and which risks it will have. Cloud adoption is not a simple task as this process depends on the type of service model (IaaS – Infrastructure as a Service, PaaS - Platform as a Service, SaaS - Software as a Service) and the development models (private cloud, community cloud, public cloud, and hybrid cloud).

The objective of this internship is to conduct a theoretical and practical study comparing the Business Intelligence components implemented in On-Premises and environment BI given the conditions of the Information system Business Intelligence Architecture and business model of the company. The internship will provide a comparative analysis of the performance of components of Current Traditional BI Architecture: how components of the ETL, and OLAP analysis influence the organization's agility and compare this between Cloud Based BI.

The Final output will help the companies with the decision to choose the deployment model for Business Intelligence or combine both.

2. LITERATURE REVIEW

2.1. BUSINESS INTELLIGENCE

2.1.1. Definition

The concept of Business Intelligence originated in the 1950s and has changed over time with the evolution of technology, taking its place in organizations today. First-time Business Intelligence was defined as “The ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal” (Luhn, H.P, 1958)

Originally Business Intelligence had one role with Decision Support System (DSS) that developed actively in the 80s. DSS was defined by Gory and Scott Morton and represents a computer system that deals with a structured decision-making problem (routine, repetitive, well-structured, easily solved) and unstructured (new, novel, ill-structured, difficult to solve) (Shim et al., 2002). From DSS, Data Warehouses, Executive Information Systems, OLAP and Business Intelligence came into focus beginning in the late 80s. In the 1980s the DSS was involved in Enterprise Resource Planning (ERP), and the concept of Relational DMBS appeared. At the beginning of 90´s ERP transformed into Executive Information Systems, appeared the possibility of providing the dashboard and scorecards to the top executives and the Data warehousing concept was created. In 1989 Howard Dresner defined Business Intelligence (BI) as an umbrella term to describe "concepts and methods to improve business decision-making by using fact-based support systems.". But Business Intelligence started to be involved at the end of the 90s at the beginning of the 2000s.

Later when Business Intelligence started to be commonly implemented in organizations this was defined “as a set of mathematical models and analysis methodologies that exploit the available data to generate information and knowledge useful for complex decision-making processes.” (Aslan et al., 2021). As well, the main purpose of BI is to support better and faster business decision-making where a decision is based on analysing the data in form of facts. The fact-based decision-making with technology enables not only top executives, but middle management and operational management to conduct analyse and measure the performance of the organization. The BI objectives are to enable everyone throughout an organization to quickly and easily access any data in the enterprise, possibly in real-time, as well as conduct appropriate manipulation and analysis.

2.1.2. Traditional Business Intelligence Architecture

Business Intelligence Architecture consists of several key components, such as Data source, Data Warehouse & Storage and Presentation layer. Schematically it can be represented in Figure 1.

Components collect data from different sources and extract, transform and load it into a Data Warehouse. Then the data is analysed by using BI tools and analytical models. Lastly, reports, insights, and dashboards are distributed for the decision-making process.

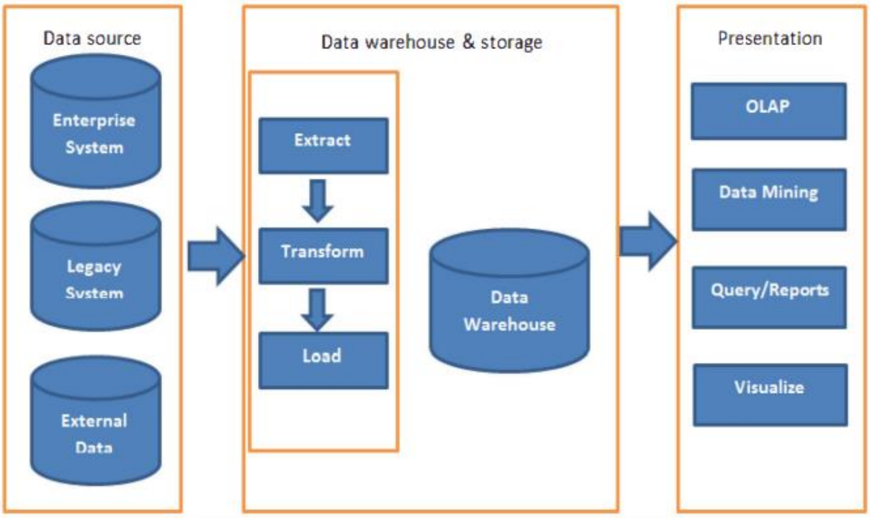

Figure 1. Traditional Business Intelligence Architecture (Turban E, et al., 2011)

2.1.3. Data source

Data Warehouse receives data from various data sources: Enterprise systems, Legacy systems or Data from external souses. Enterprise systems with Legacy systems can contain historical data, and operational data and this data is stored on operational databases. The data sources can be relational databases or any other data structure that supports the line of business applications. Data sources also can reside on many different platforms and can contain structured information, such as tables or spreadsheets, or unstructured information, such as plain text files or pictures and other multimedia information. (Ranjan, J., 2009)

Fabret (1997) argued in his research that "The data sources can be either part of the operational environment of a company or an organization, or external, that is produced by a third entity. They are usually heterogeneous, which means that the same data can be represented differently, for instance through different database schemas, in the sources".

2.1.3.1. ETL Process

Data Warehouse has a special structure known as schema which defines the type of relationship between the tables: for example, a more relational-based schema known as Snowflake schema or focusing on the speed of data querying provided at Star schema., and in the end, the schema defines the datatypes of attributes of each table. The data is received from a variety of sources, that can contain various data types and structures of data. At this condition is important to conduct the data integration process to transform the data.

ETL is a type of data integration that refers to the steps of Extract, Transform and Load used to combine data from multiple data sources.

Extract:
The extract step refers to the process of getting the data from all involved data sources. During this stage, data is obtained and filtered out from various forms of data sources. After, the data is validated by verifying consistency and deleting invalid records. At the end of the process,

validated data stored on a relational database significantly facilitates further data processing at the transformation stage.

Transform:

This step is computationally complex in the overall ETL Process. The filtered data from the previous stage is unified, the necessary aggregated is calculated and duplicate or missing data is identified. At this step, data is combined and transformed into standard formats, and structures which correspond to the Data Warehouse Schema. Transformed data could be temporarily stored on the intermediary database – a staging area where isn't any relationship between the tables which facilitates the load process to the Data Warehouse. In another approach, the transformed data is directly loaded into the Data Warehouse.

Load:

The transformed and cleaned data from the transform stage is loaded into Data Warehouse. This is the last stage of ETL, and the moved data is ready to use for BI Purposes.

Generally, an ETL process is essentially composed of a collection of different tasks, going from getting new data from one source system to another destination system.

2.1.3.2. Data Warehouse

Definition and design

According to Inmon (1996) Data Warehouse (DW) is described as a "collection of integrated, non-volatile, subject-oriented databases designed to support the DSS function, where each unit of data is relevant to some moment in time. It contains atomic data and lightly summarized data." The Purpose of BI is to provide Decision Support capabilities for all members of organizations forming then the "single version of the truth". Providing decision support requires 2 characteristics that traditional Data Warehouses couldn't give: End users' access (simple and intuitive database structure), Read only access (Faster's query performance). (Sharda, Delen & Turban., 2017)

As Data Warehouse environment is different from the operational environment (Operational Databases) and, therefore, requires a different approach in terms of database design to enable Decision Support. In 1996, Kimball created a Dimensional Modelling approach witch specifically conceived for Data Warehouse design, answering the limitation found in the traditional approaches to database design. This design approach was a fundamental development for data warehousing since it provided a viable way to represent data in order to support end users in understanding and being able to query them. According to (Kortnik, M. A. R.,1999) the objectives of dimensional modelling are to produce database structures that are easy for end-users to understand and write queries against, and to maximise the efficiency of queries. It could be achieved by minimising the number of tables and relationships between them. Data Warehouse contains relation data that is organized to provide enterprise-wide, cleansed data in a standardized format which enables the conduct of analytical processing activities (OLAP, Data Mining, querying, reporting, and other decision support applications. (Sharda, Delen & Turban., 2017)

Mohammad Rifaiea (2008) In his research about Data Warehouse common design, lists the characteristics of information warehouse data stores which distinguish them from existing production operational data stores as follows:

1) Non-Volatile: Real-time updates occur to selective Data Warehouse data stores. Most data stores are refreshed in batches, not less than every 24 hours. Time consistent context of data across different sources needs to be maintained.

2) Time Variant: An information warehouse typically stores the data on a 3 to 7- year time horizon. The data is periodic and maintained as a series of snapshots, taken at some moment in time. The key structure of data tables must contain some element of time.

3) Granularized structure: Data is maintained at various levels of granularity and summarization. Frequently accessed data can be re-joined and summarized to enable quick turnaround on queries and reports. Detailed and atomic level data will be maintained alongside summarized and pre-calculated data"

Data organization

The Dimensional Modelling approach uses the next key concepts: Facts, Dimensions and Measures. Fact represents the subject of business which needed to be analyzed. There are 4 types of facts: Event Fact (business event), Fact-less Fact (fact per event), Snapshot fact (a fact that captures current status), and cumulative snapshot facts (captures cumulative status up to now). (Sharda, Delen & Turban., 2017) Dimension could be defined as a collection of data that describe one business dimension (Fact Table). Dimensions determine the contextual background for the facts, they represent the parameters over which OLAP analyze is performed. Measures are a fact property which the end-user wants to study and optimize. It represents the performance or behaviour of the business event relative to the dimensions.

Relational Models

Dimensional modelling is usually based on the next relational models which defined the organization of the tables: "Star" model and "Snowflake" schema. The Star schema model is the basic structure for a dimensional model. In this model central table is a fact table surrounded by dimensions tables and linked via foreign keys and has a one-to-many relationship). (Sharda, Delen & Turban., 2017) This type of relational model allows fasters "read" (SQL queries) but slow "write" operations.

The snowflake schema has a similar structure which difference is that dimensions are de-composed from many-to-one relationships among sets of attributes which forms a hierarchy. The decomposed snowflake structure visualises the hierarchical structure of dimensions very well. (Gutiérrez A, Marotta A, 2000)

Data Warehouse Types

There are several Data Warehouse types which are implemented depending on the elaboration model of the Data Warehouse (Kimball or Inmon), the architecture of BI and the business objective of the enterprise.

In the literature three principal types of Data Warehouse are appointed:

- Enterprise Data Warehouse (EDW);
- Operational Data Source (ODS);
- Data Marts.

Enterprise Data Warehouse

Enterprise Data Warehouse (EDW) is a large-scale Data Warehouse that is used across the enterprise for decision support. EDW is used in analytical tasks as OLAP and it could be a source for other Decision support systems such as Customer Relationship Management (CRM), Supply Chain Management (SCM), Business Process Management (BPM), Product Life Cycle (PLM), Knowledge Management System (KMS), etc. This type of Data Warehouse is implemented on traditional Business Intelligence architecture.

Operational Data Source (ODS)

Inmon defined Operational Data Source (ODS) as a "subject-oriented, integrated, volatile, current-valued, detailed-only collection of data in support of an organization's need for up-to-the-second, operational, integrated, collective information". (Gutierrez A, Marotta A, 2000) ODS receives data that comes from the operational systems environment and this data is used for Decision Support based on the company's specific subject areas. This type of database could be implemented when there is a need to consult reports and make near-real-time analytic decisions based on recent data. In some cases, the ODS can be merely a mirrored copy of an operational database (usually an online transaction processing, or OLTP, database). In this example, the ODS can be used strictly for reporting and data, thus avoiding the need to exhaust the resources of the hardware hosting the OLTP database or of the database itself. (Smith, 2009)

In these cases, during the ETL process, no data integration or summarization during the load phase is needed. However, on the ODS, any data manipulation required for decision support can then be executed without affecting the integrity of the core operational data. But in most cases, the ODS is more complex as receives data from various souses and could have a structure similar to the Data Warehouse.

Data Marts

A Data mart represents a subset or simple form of a Data Warehouse that is focused on a single area (or functional area), such as Sales, Finance, or Marketing. In an organization, data marts are often built and controlled by a single department. Datamart receives data from a few sources because they are implemented to the scope of the specific needs of the department. (Inmon., W.H., 2002)

The sources can be a central Data Warehouse or external operating systems depending on the implemented type of Datamart. An organization can have several Data Mart's that are relevant to one or more business units for which were designed.

Usually, there are two types of Data Marts: Depended Data Mart and Independent Data Mart.

The Dependent Data Mart is a subset that is created directly from a central Data Warehouse where the data is loaded already from the centralized warehouse and then extracted, enhanced and loaded into this data mart. As data stored in the Data Warehouse is already cleaned and have predefined formants, the ETL process is simplified. (Hajmoosaei et al., 2011) Thus, it is partially dependent on the Data Warehouse and independent in terms of clients getting the data from data marts which are specific. The dependent Data Mart follows Inmon's methodology of implementation of a Data Warehouse where Data Marts are created from the existing Data Warehouse.

The dependent Data Mart provides the ability to the end user to get the same version of the data that are accessed by all other Data Warehouse users which supports the concept of the single version of the truth. The disadvantages of a dependent data mart are the costs of implementation as a central Data Warehouse must construct first which requires clear business needs resources and time.

The Independent Data Mart follows the opposite of Kimball – the Inmon methodology of Data Warehousing. In this methodology, BI systems are constructed by implementing the independent Datamart's fist.

Independent data marts are primarily characterized as independent entities which receive the data directly from operational sources which are specifically subject-oriented to the needs of the exact business unit. (Inmon, W.H., 2002) describes the characteristics of the independent Data Mart as follows: "First, each data mart I sourced directly from the operational systems without the structure of a Data Warehouse to supply the architecture that is necessary to maintain and grow the data marts. Second, these data marts are naturally built separately from one another by autonomous teams. Typically, these teams will usually utilize varying tools, software, hardware, and processes". The advantage of Independent Data Mart is the low cost of implementation Data Mart is focused on specific department needs and requires dome data process and data volume. The disadvantage is each department will have a different version of the data as the structure of Data Mart focuses on the business context in which was developed.

Analise Capabilities

The BI architecture provides the end-users with the capability of analyzing such as OLAP Data mining, Querying and reports, and visualization. OLAP provides multidimensional, summarized views of business data and is used for reporting, analysis, modelling and planning for monitoring and making decisions based on the business data. The key features of an OLAP application are multidimensional views of data, calculation-intensive capabilities and time intelligence. (Forsman, S.,1996). Typical applications executed on multidimensional data views may include roll-up (summarizing data into a high-level view), drill-down (moving downward through a data hierarchy), slice and dice (segmenting data into smaller parts), and pivoting (performing cross-tabulation).

Data mining is the nontrivial extraction of previously unknown and potentially useful information from data. This involves various technical approaches, such as clustering, data summarization, learning classification rules, analyzing changes, and detecting anomalies. (William J. et all., 1992). Datamining contains techniques involved in the descriptive analysis of data to answer the question: "What happened? Why it happened?" but could be also involved in prescriptive analyze.

2.1.4. Benefits and Limitations of Business Intelligence

Today the data is generated in high quantity and represents a significant value to organizations and requires the actions that need to be taken. The decision-makers need the right information at the right moment in the right place. Business Intelligence helps organizations gain actionable insights from data in order to achieve business goals. Usually, most departments can use the Information stored in traditional databases for some applications. But Business Intelligence help´s a company not only consult the data but also easily obtain insights by making the analysis and integration of larger amounts of data from multiple areas of the company and allows them to perform complex calculations and distribute compelling reports and dashboards along all enterprises. Thus, the BI simplifies the communication flows in the enterprise while coordinating activities and enable companies to respond quickly to changes in financial conditions, customer preferences, and supply chain operations.

It is important to mention that BI improves the overall performance of the company (Ranjan, Jayanth, 2009). For example, the principal benefits could be summarised as follows (Eckerson, W., 2003):

- Time savings;
- Provide the single version of the truth;
- Enables the improved strategies and plans;
- Improved tactical decisions;
- Make the processes more efficient;
- Cost savings

In the survey Thompson (2004) suggested that BI tools can help to generate faster and more accurate reporting to support decision-making that can lead to improved customer service levels and improved profits. S. William and N. Williams (2003) also appointed out that BI tools help to improve both management and operational processes, because of their ability to support managers in planning, organising, leading and controlling. On another hand, Business Intelligence has the challenges of implementation such as alignment with strategy, costs management, trainee, etc. However, the success of BI in the enterprise depends on its implementation.

2.2. CLOUD BUSINESS INTELLIGENCE

2.2.1. Cloud Computing

2.2.1.1. Definition

Cloud computing is currently used by a large number of organizations and has gained fast adoption for the last few years as corporate sought more efficient and effective ways in utilizing their IT investment. Cloud computing has particularity by enabling to use or implement flexible and scalable services without having the computing resources directly allocated. In this way, companies have powerful computing resources for executing complex tasks without a big investment. (Amanatullah et al., 2013)

The National Institute of Standards and Technology (NIST) defines cloud computing as “a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.” (Mell, P. and Grance, T., 2011). Companies have the advantage to access powerful computing resources fast, without big investments for executing complex tasks.

Cloud has been adopted as an important utility across all aspects of society, from academia, governmental institutions and industry. Amanatullah (2013) and Casavant et al. (1998) concluded that the characteristics of cloud computing such as dynamic, metered access to a shared pool of computing resources have enabled the realization of new technologies and paradigms to fulfil the demands of emerging applications including scientific, healthcare, agriculture, smart city, and traffic management.

2.2.1.2. Computing Paradigms and Evolution

Cloud Computing is one of the computing paradigms that evolved every time. From 1958 computing systems started evolving by creating new paradigms, intending to improve the use of hardware resources in an efficient way. Computing systems have evolved during these decades of computing, there have been various types of computing paradigms and technologies have been developed and invented, which contribute extensively to the current research in the field of computing (Gill et al., 2019). In the research on defining the taxonomy of cloud computing systems.

Rimal et al. (2009) argued that only one system can execute one specific task at a time and execute multiple tasks but at the same time, there is a need for multiple systems to run parallelly to execute multiple tasks concurrently. For this capability, a secure communication network is required to exchange data among different computing systems. The evolution of computational technologies from 1958 to implement the paradigms discussed is represented schematically in Figure 2.

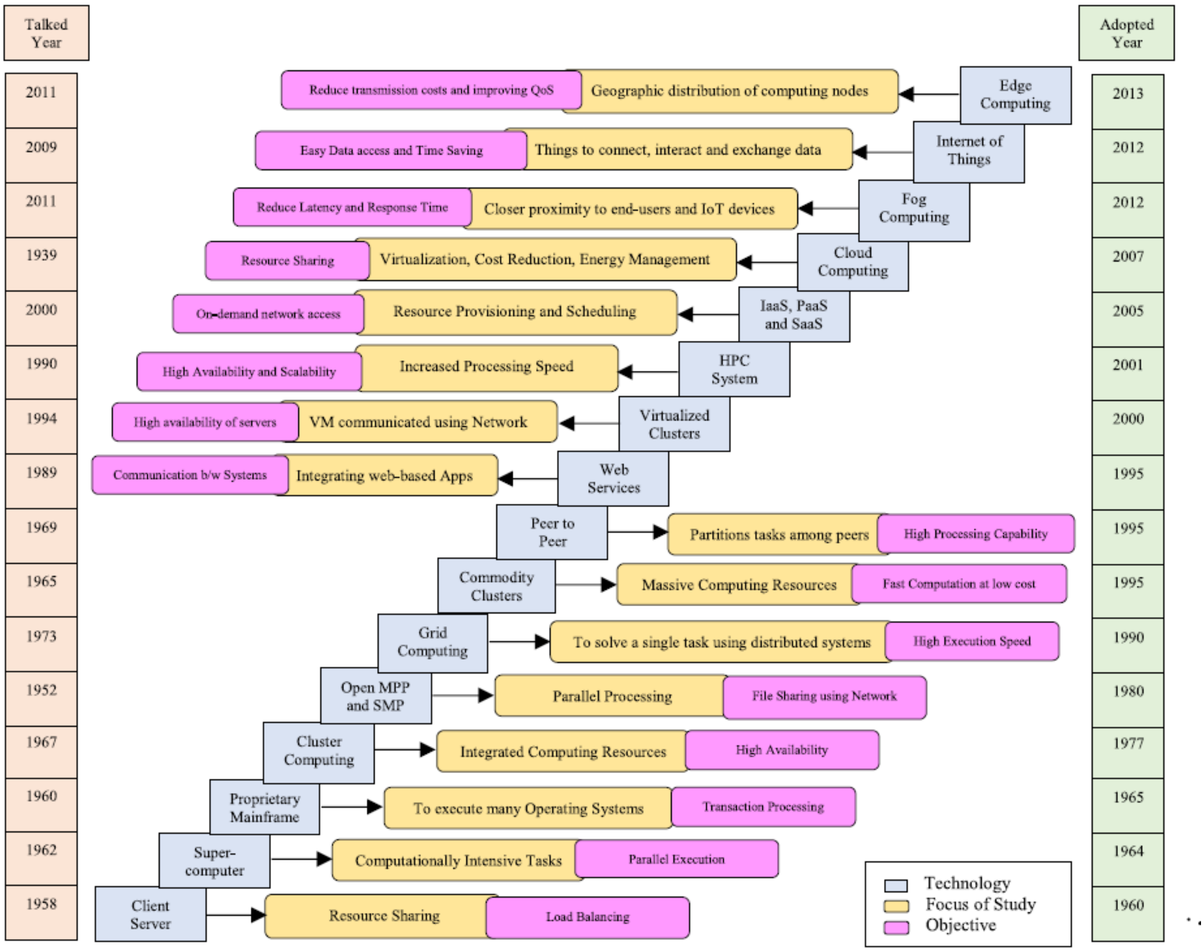

Figure 2. Computing paradigms evolution (Gill S, Tuli S, Garraghan P, 2019)

The technology evolution is composed of several steps as follows:

- Client-server: This is a distribution application or centralized system that was developed in 1960 aiming to divide the workloads or tasks among resource providers (servers) and client service requesters. (Rimal et al., 2009) . In the research, M. J. Flynn (1966) described that "Servers and Clients communicate through the computer network and server shares resources with clients to execute their workloads in a load balancing manner". Email and World Wide Web (WWW) are two examples of a client-server model.

- Supercomputer: In 1962 appeared the need to develop a system with high-performance computing capability to execute computationally intensive tasks in different scientific fields such as molecular modelling, climate research and quantum mechanics (Casavant, T. L. and Kuh, J. G., 1998). Important to mention that after the development of supercomputers the main research challenge of energy usage and heat management remained since 1960. Examples of supercomputers are Multivac, HAL-9000 and The Machine Stops.

- Proprietary mainframe: This is large-high speed computer, which can further support various devices and workstations, are used to process a large amount of data such as transaction processing, consumer statistics and census (Yu, J., Buyya, R., 2005). The latest version of the mainframe IBM z14 was developed by IBM at the year 2017.

- Cluster computing: This technology uses a fast local area network to communicate available computing nodes and clustering middleware is used to make coordination among different computing nodes. (Compton, K., Hauck P., 2002) The main objective of cluster computing is to execute a single task using different interconnected computing nodes to improve the performance of the computing environment.

- Open massively parallel processing (MPP) and symmetric multi-processing (SMP): There are two main types of parallel processing environments: massively parallel processing (MPP) and symmetric multiprocessing (SMP) systems. In an SMP environment, other hardware resources (disk space, memory) are shared by multiple processors but using a single operating system. In the MPP environment, the only file system is shared but there is no sharing of resources during the execution of the job. The scalability can be improved by adding computers and related disk and memory resources. (Gill et al., 2019)

- Grid computing: The grid computing paradigm was developed to achieve a common objective using distributed computing resources and executes non-interactive workloads, which contain a huge number of files. (Compton, K., Hauck P., 2002). The single grid is dedicated to an execution of a specific application. Grid computing is commonly a resource allocation and management service, secure infrastructure and monitoring and discovery service.

- Commodity clusters: It is also called commodity cluster computing, which offers low-cost computation of user workloads by using huge numbers of computing resources concurrently. (Compton, K., Hauck P., 2002), (Flynn, M. J., 1966). Different vendors are using open standards to manufacture commodity computers to reduce the variation among the products of vendors. Presently, off-the-shelf commodity computers are available to fulfil business computing requirements quickly.

- Peer to peer: In the research Yu, J., Buyya, R., (2005) Peer to peer is described as distributed architecture which divides the workload or task among different peers or computing nodes and peers can communicate with each other directly at the application layer. In Peer-to-peer architecture, peers can access different resources such as processing power, disk storage or network bandwidth without the requirement of a central coordinator. Today this type of paradigm started implemented in the 90s on TCP/IP networks and is used to exchange data among peers. The main applications of peer-to-peer architecture are multimedia, file-sharing networks and content delivery.

- Web services: This technology enables communication among different electronic devices through World Wide Web using different types of machine-readable file formats such as JavaScript Object Notation (JSON) and Extensible Markup Language (XML) (Rimal et al., 2009). Web service provides the user interface to the end user for interaction with the database server.

- Virtualized clusters: These clusters are implemented as a real computing system to perform similar functions using virtualized environment. Virtualized cluster enables the sharing of resources among different Virtual Machines (VM) to execute workloads or tasks. (Gill et al., 2019)

- HPC system: This is the tool which is used to solve large tasks (which require high computing power) of business, engineering and science. High-Performance Computing (HPC) system contains different types of computing resources to solve different types of problems and access to these resources are controlled by a batch system or scheduler. (K. Compton, S. Hauck,2002). HPC systems are sharing resources and they can access different resources remotely and execute workloads or tasks using the scheduling of parallel resources.

- IaaS, PaaS, SaaS: As the 200s internet started to expand, the necessity of expanding the service of the Internet appeared. There are different types of web services which can be accessed via the Internet such as SaaS (Software as a Service), PaaS (Platform as a Service) and IaaS (Infrastructure as a Service). SaaS offers software functionality as a service without any maintenance and initial cost with high quality and an example of SaaS is Gmail. PaaS offers a framework, where user can deploy their application with the required scalability and an example of PaaS is Microsoft. IaaS offers infrastructure resources such as network, memory, storage and processor to execute workloads or tasks in a cost and time-optimized manner and an example of IaaS is Amazon. This concept started to implement in 2005.
- Cloud computing: The cloud services are generally denoted by – XaaS where X consists of I (IaaS), P (PaaS), S (SaaS), and is defined as the practice of using remote resources to execute user tasks (processing, management and storage of data) through the Internet. (Singh, S., Chana, I., 2016) Cloud computing enables sharing of resources to reduce execution costs and increase the availability of services. This paradigm was discussed in 1939 as "Sharing Recourses" but was implemented only after 89 years. Cloud contains the four deployment models of cloud computing models: public, private, and hybrid.
- Fog computing: This is the latest architecture which performs a significant amount of storage and computation using end devices or fog nodes and the Internet is used to establish communication among these devices(Rao & Clarke, 2020). Fog computing consists of a data plane and a control plane. The Data plane provides services at the edge of the network to reduce latency and increase QoS, while the control plane is part of the router and decides network topology. (Gerla et al., 2012.).
- Internet of Things: IoT can be described as the network of devices such as actuators, software, home appliances and sensors and Internet connectivity which is used to exchange data among these network devices (Gerla et al., 2012.).
- Edge computing: It is a distributed computing paradigm, which performs computation on distributed edge devices, and it enables data collection and communication over a network. (Singh, S., Chana, I., 2016) Further, edge computing moves the large volume of data by processing at edge devices instead of a cloud server, which improves the quality of service and reduces latency and transmission costs.

2.2.1.3. Cloud Computing Architecture

The official article Final Version of NIST Cloud Computing Definition, (Mell, P., Grence, T., 2011) defined three fundamental models for Cloud Computing services: Infrastructure as a Service, (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). Figure 3 represents the layered Architecture of Cloud Computing that consists of various layers that are the core of each service model. It is important to note that each software model is oriented to specific types of customers.

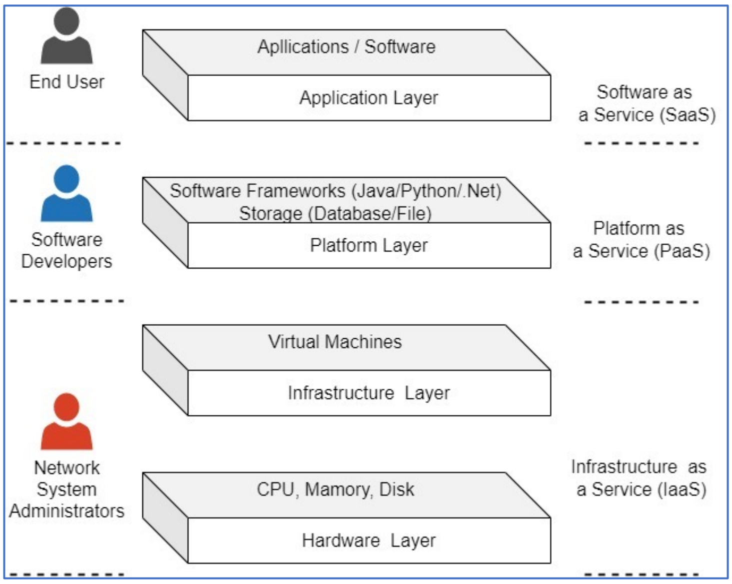

Figure 3. Cloud Applicational Architecture. Adapted from (Alqarni, 2021)

Infrastructure-as-a-Service (IaaS) is defined as the service model which allows the use of computing resources such as servers, storage devices and network infrastructure (addresses, load balancers and firewalls), often delivered as virtual machines hosted by the service provider and managed remotely (Burtescu et al., 2014).

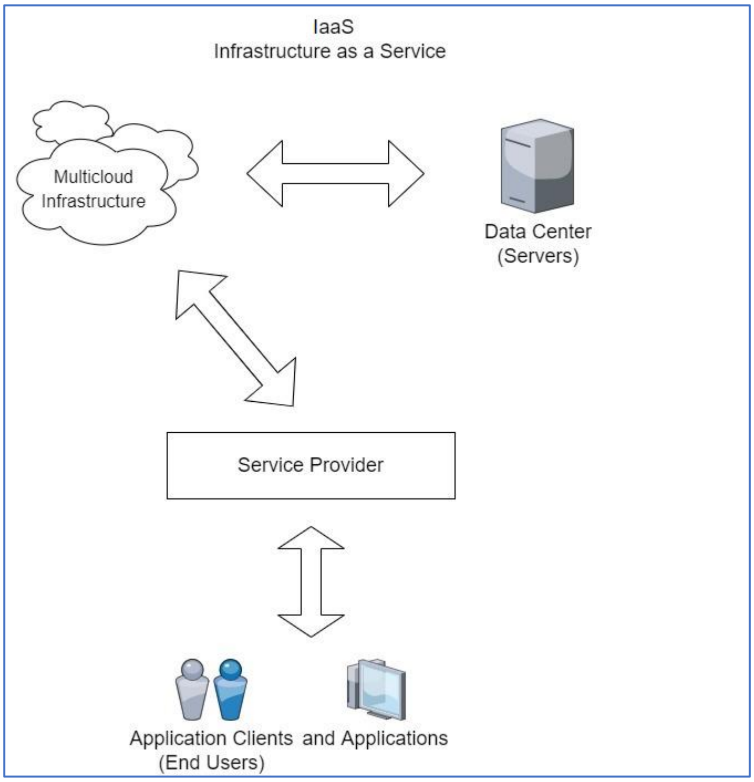

Figure 4. Infrastructure-as-a-Service Applicational Architecture. Adapted from (Alqarni, 2021)

A key benefit of IaaS is the usage-based payment scheme which allows customers to pay as they grow. Another important advantage is that it always is using the latest technology, so, the customers can achieve much faster service delivery and time to market (Rimal et al., 2009). IaaS also have benefit for small organizations as enables them to emphasize their core business activities instead of focusing more on IT infrastructure. Key examples are GoGrid, Flexiscale, Layered, Technologies, Joyent and Mosso/Rackspace etc.

Platform-as-a-Service (PaaS) is the type of cloud that provide developers with a platform including all the systems and environments comprising the end-to-end life cycle of developing, testing, deploying and hosting of sophisticated web applications as a service delivered by a cloud-based, a Platform-as-a-Service (PaaS). PaaS is completely managed by the provider, but end users can customize their environment settings and deploy everything they need.

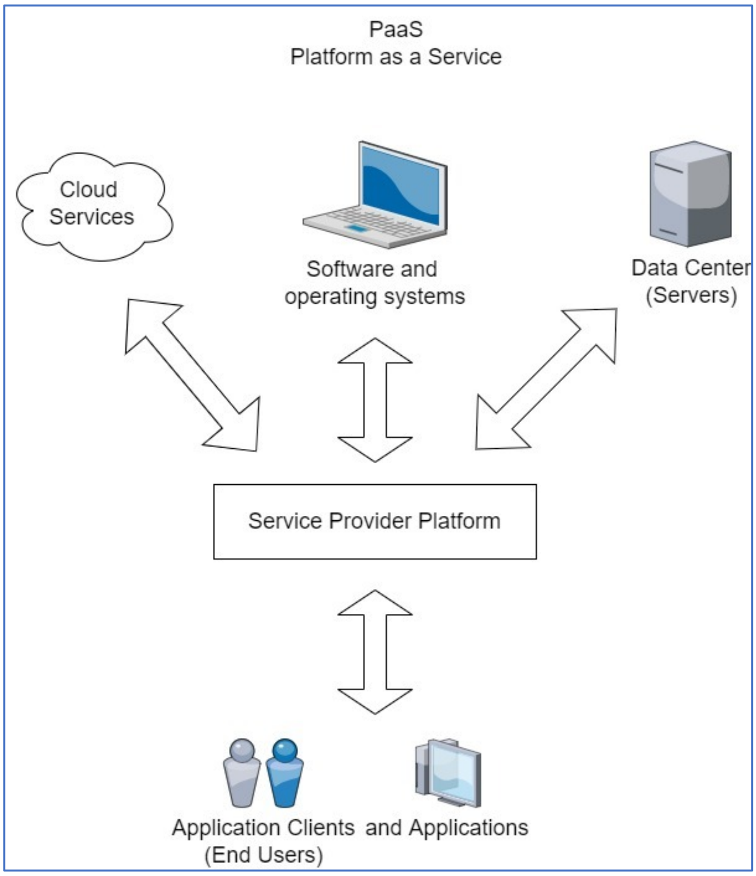

Figura 5. Platform-as-a-Service Applicational Architecture. Adapted from (Alqarni, 2021)

Key examples are the companies GAE, Microsoft's Azure etc. Compared with conventional application development, this strategy can slash development time, offer hundreds of readily available tools and services, and quickly scale.

Software-as-a-Service is a multi-tenant and the most popular model platform. It uses common resources and a single instance of both the object code of an application as well as the underlying database to support multiple customers simultaneously (Rimal et al., 2009). In other words, SaaS requires that cloud providers install and operate the complete application stack on their platform and users access these applications remotely through specific client software. In this manner, the customers do not manage the cloud infrastructure on which their application is running and are not responsible for maintenance and support.

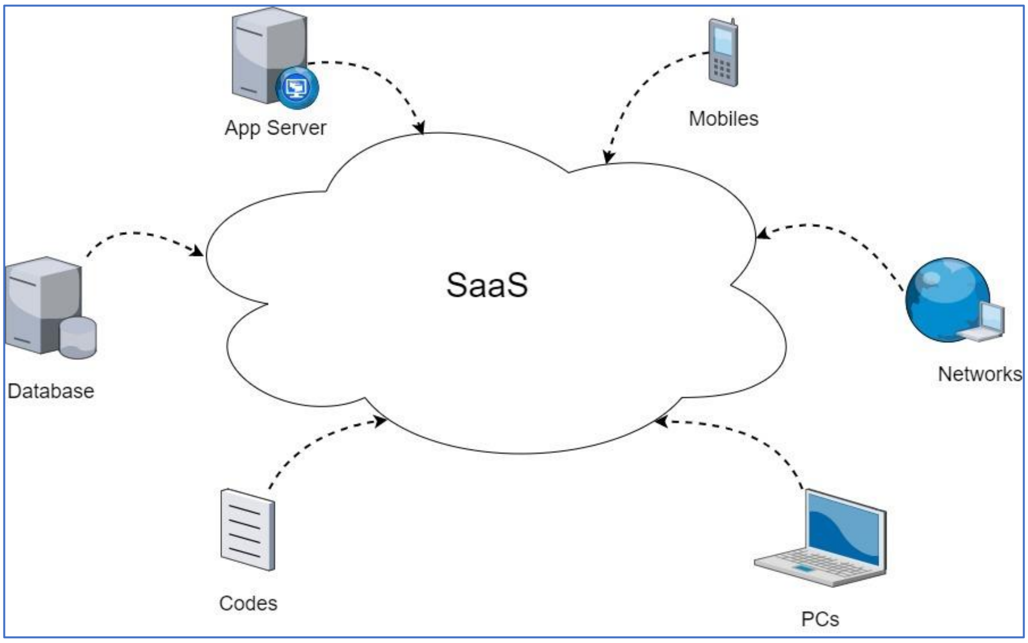

Figure 6. Software-as-a-Service Applicational Architecture. Adapted from (Alqarni, 2021)

2.2.1.4. Benefits and Limitations of Could Computing

The Could Computing paradigm in all service forms provides significant benefits compared with the traditional On-Premises approach. Griffith (2015) argues that cloud computing provides benefits such as data access anytime anywhere benefits, cost saving, mobility, scalability, agility, efficiency, environmental benefits and flexible security. Scalability is characterized by enabling the use of resources for a large pool of users who have different needs. Agility gives short response time on operations and it is important in critical situations. The multi-sharing is a benefit that characterized cloud computing itself. Aslan (2021) explained the multi sharing is “ the main characteristic of the cloud that enables users to share resources”

Cloud-based solutions have been designed initially for small and medium-sized organizations and this is suited initially for them increased agility at low cost. Given this, cloud computing has benefits, especially for these organizations as allows to them focus more on business needs without having concerns about hardware maintenance and software configuration. Cloud computing can be used as a service rather than hardware-based computing, it is cost-effective and can be used as a utility (Miller, 2009). Another attractive benefit of cloud computing is that the application processing load shift from local computing to cloud computing. For example, using a client, the application can be processed on a remote server (Strickland, 2008).

The important question is if the computing paradigm has issues and concerns that the organization needs to consider before implementing it. Indeed, the first very important concern of organizations is information security as there are several security challenges and opportunities in cloud computing (Krishnamurthi R, 2019). Usually, the data is stored in the cloud by a third party and can be accessed through the Internet and has the property of giving limited access to control and visibility on data. This makes concerns organizations on how to ensure data can be suitably secured with low risks at the same time.

The questions of risk became critical even before the cloud computing implementation as several risks in cloud computing is associated with the security of cloud data. These security risks include the lack of visibility towards data and failure to control data from theft or breaches. Most business organizations have encountered security issues in SaaS, IaaS, and private clouds (Raja JB, Rabinson KV, 2016). Also, one of the critical issues in cloud computing is suiting cloud systems to the big companies. As was mentioned could computing was initially created for small-medium organizations. As a result, the use of a Cloud Computing based solution is not suited for all organizations, especially in the field of BI. (Gill S, Tuli S, Garraghan P, 2019). Success in using Cloud Computing opportunities depends on conditions existing within the organization and on the understanding of the main benefits and risks of Cloud Computing.

2.2.2. Cloud Business Intelligence

According to Dresner H. and Ericson J. (2015) cloud-based, BI refers to the BI technologies and solutions that employ one or more cloud deployment models. Cloud Business Intelligence use its own infrastructure and delivers its components through the cloud storing the data. On this type of cloud, the components of traditional BI, such as Data Warehouse, ETL process and instruments, and analytical OLAP tools) are implemented on the cloud platform and could be delivered as a service.

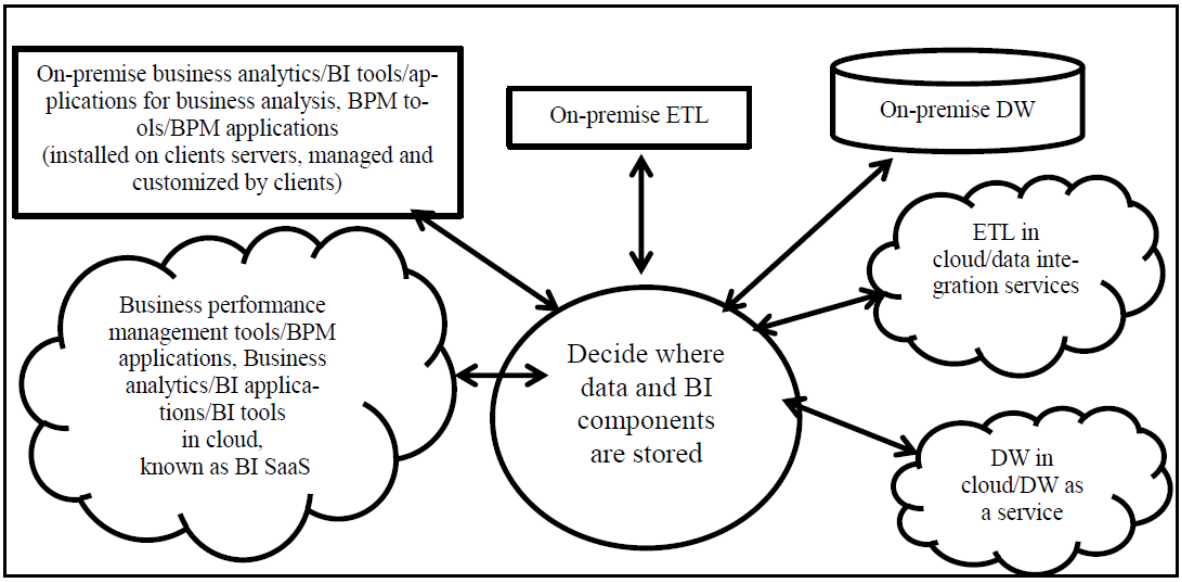

Figure 7. Cloud Business Intelligence Architecture (Muntean, M., 2015)

The several combinations of Business Intelligence components that could be used in the cloud are shown in Figure 7. The decision of how data will be stored, and where and in which form will be accessed depends on the following cloud deployment model.

In the literature the principal Cloud BI services models fall into the following three categories: i) BI SaaS (BI software as a service); ii) BI PaaS (platform as a service), and iii) BA PaaS (business analytic platform as a service).

BI SaaS (BI software as a service):

BI SaaS include the functionality of BI in SaaS Application:

BI SaaS could be defined as on-demand BI (Muntean, 2015), which has the following components:

- BI SaaS tools that can be used to develop BI applications for deployment in the cloud;
- Packaged BI SaaS applications that can be deployed in a cloud environment (for example, applications for business analysis or business performance management applications) data integration services for BI;
- Developing/ testing services for BI.

BI for PaaS (Platform as a Service) and BA PaaS (Business Analytic Platform as a Service):

BI for PaaS usually includes IaaS layers and could be used as a part of analytical and information delivery services that are integrated into the platform which focuses on Business Intelligence tasks. Muntean, M., (2015),

On the other side, the Business Analytic PaaS (BA PaaS) represents a shared and integrated analytic platform in the cloud and delivers several services such as Data Warehouse services, Data Integration services and Infrastructure services. The summarized description of the capabilities of the Business Analytic Platform as a Service is represented in Figure 8.

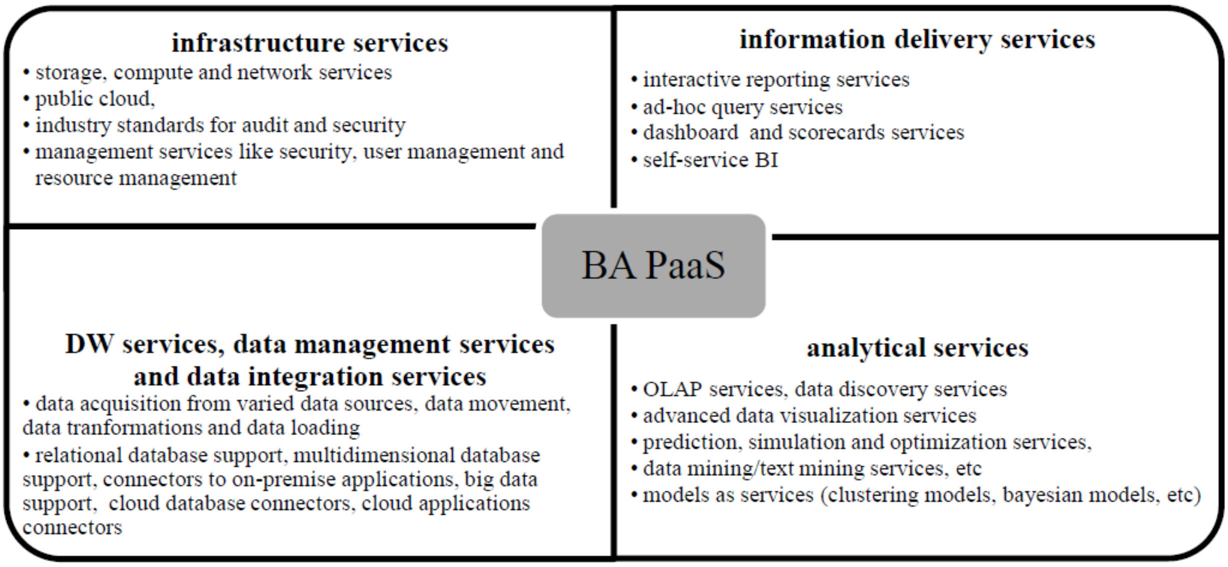

Figure 8. Capabilities of Business Analytic Platform as a Service (Muntean, M., 2015)

3. COMPARATIVE ANALYSIS OF ON-PREMISES AND CLOUD BI

In the previous section, the definitions of the traditional BI components, such as Data Warehouse, ETL process and instruments, and analytical OLAP tools, were described, as well as their delivery capabilities via the cloud. Each component could be deployed in a public or hybrid cloud deployment model depending on the enterprise's business needs.

The BI components implemented in On-Premises or in the cloud have their own application features. At the same time, each of them has advantages and disadvantages, depending on in which business context they are used.

For the successful use of certain On-Premises and BI components in different contexts, a more complete analysis of their capabilities and comparison is necessary. In this section, a comparative analysis between BI components (ETL Process, Data Storage, OLAP Operations) in On-Premises and in the cloud is presented and discussed.

3.1. COMPONENTS ANALYSIS

3.1.1. ETL Process

In the On-Premises environment, traditionally the ETL is more representative of the scope of the process and relies only on the meaning that the data is loaded and transformed from one On-Premise system to another. The overall process is started with the decision of which business scope of data is needed based on the structure of the Data Warehouse. The identified souses of data are quite limited: typically, this can be legacy systems, databases, transactional systems, Excel sheets, and CSV files.

After defining data sources, the entire pipeline of the ETL process is configured manually by IT and Business Intelligence teams via On-Premises tools. This process is repeated for each BI initiative within the organization. The main limitation is that a new data source is difficult to add since it must strongly rely on the DW components which is a very difficult change. If the company change completely ETL sources this will impact the Data Warehouse changes according to the business needs. Usually, it is costly in the On-Premises scenario and requires a lot of time.

In the cloud environment, the ETL process has various scopes of implementation. Compared to the traditional implementations, the ETL process could be partly represented as one of the components deployed in the architecture of cloud service which received a big amount of data in real-time from various data sources. ETL processes have more instruments for data transformations with powerful computational capacity. In the BI context, companies can start the ETL process with the available data without determining the exact data sources. Then, the additional data sources can be added continuously increasing the scope of business. But at the same time there some questions appear that need to be answered:

- What kind of data is entered into the system?
- What data types can be used by BI solutions to contain information useful for a business user?

The BI solution in a cloud should allow companies or groups to get the data in files of large size with each application that can export data for a variety of sources. Optimally, the BI solution should have an automated data updating system, otherwise, the data must be uploaded and updated manually.

Could Business Intelligence vendors be able to offer the development of an integration plan directly or through a partner?

The ETL process on the cloud is flexible and enables obtaining data from several sources by continually adding. This provides benefits, especially for small companies where projects are implemented with a large scope and then several data souses could be added, and this will impact the continuous change of business needs and context. At the same time when the quantity of data is higher, the ETL process turns more computationally complex. SaaS Cloud implements a "pay-what-you-use model at which the costs might be increased due to the expansion of the resources to complete computation tasks. Besides the several techniques implemented in order to reduce the costs of the ETL process, on the Cloud, the problem remains, and organisations needed to have attention to it and continuously monitor the overall ETL process thought tools. Security during ETL processing is a critical issue as the transformed data on architectures could be transferred via the internet and there is a risk of data hacking Cloud providers and client companies must use a strong encryption algorithm.

On the other hand, the traditional On-Premises ETL solutions give more data control to the users during all pipelines: monitoring all processes of transformation and computation when cloud tools provided for this initiative gives limited control. On complex ETL processes, there isn't a risk of increased costs of processing compared to the cloud solution. At the same time, enterprises need attention to not overload the computational resources on ETL tasks as there are risks of global down of the DW which is critical for the entire organization. The On-Premise BI ETL process also has a limitation on the data sources as typically the data is received from initially defined sources. Those data sources need to be defined initially with the elaboration of the business needs during the implementation of the Data Warehouse. The security of the ETL process has low risks competing with the cloud ETL Processes as the data is transferred only On-Premises environment.

3.1.2. Comparison of the traditional and Cloud Data Warehouse in BI

Traditional Data Warehouse is the main component of BI which stores the data used for analytical purposes. The implementation strategy in the On-Premise DW shall be first defined and followed up according to the business needs. For example, Kimball scaling-up methodology starts from the creation of Data Marts with several ETL processes which reduce the scope of an application (Sharda, Delen & Turban, 2017). On the other side, Immon approach role-up an Enterprise Data Warehouse (EDW) with scaling but has the limitation of providing new analytical purposes (Inmon, W.H., 2002). Generally, the implementation of strategy for the creation of DW defines the architecture of all Business Intelligence in an organization. Thus, wrong implementation or changing the structure is a costly and time-consuming task.

On-Premises Data Warehouse (DW) requires configuration of the hardware, software and infrastructure which needs constantly implemented, monitored, and resolve challenges in the case of problems. The DW is costly and time-consuming to scale as requires manually buying additional memory storage configuration. Those limitations result in high costs and focus on data management instead of getting the analytical capabilities of the DW. The security On-Premises DW, on the one hand, has no risk as the data again is not transferred via the internet and stored in internal systems but for another hand, there is another risk of those system's security policies and capability of these systems ensure disaster recovery and backups functionality.

Cloud Data Warehouse provides the service model of "Storage as a Service". The first strong advantage is that it is enabled to connect to various data sources and is designed to take a larger number of users and applications. On the cloud BI, Data Warehouses are created for each Business Intelligence project or application independently and there is no dependence on a single EDW approach as On-Premises. This gives the enterprise high flexibility on the implementation of BI on different business initiatives, especially for slow-medium organizations without the additional time and costs of creating a new Data Warehouse.

Another important component of storage is scaling up which is a tedious and resource-exhaustive task. The key benefit of Cloud DW over On-Premises DW is that scaling can be achieved effectively and easily by acquiring more space with immediate implementation. Using the cloud data can be scaled up or scaled down quickly. On the other hand, a traditional Data Warehouse cannot be scaled up and down instantly as this requires reviewing the business context in which BI was implemented before. Scaling is costly and time-consuming. For analytics purposes, Cloud DW is efficient as uses columnar storage and massively parallel processing (MPP) which allows better performance for running complex queries. At the same time, in the ETL process, the companies need to pay attention to the costs of the analytical queries as complex tasks in the cloud, which require more computational resources and can result in high unexpected costs.

It is important to note that the security in On-Premise DW is not the same as in the cloud as the main risk of data transfer via the internet is avoided. However, the cloud provides and enables data availability, as well as disaster recovery and backups, which are passed to the cloud provider side, and it is much more flexible than traditional On-Premises.

Cloud Data Warehouse allows enterprises to move their focus from system management to business analysis whereas traditional Data Warehouse focuses on data management. Cloud DW allows cost benefits by eradicating hefty upfront costs.

Cloud Data Warehouse technology avoids the need for dedicated hardware and server rooms, IT-related issues, and operational expenses to maintain the DW. By reducing this complexity of managing in comparison to the On-Premises systems, organizations can focus on extracting value from their data rather than on maintaining hardware and software infrastructure.

One of the biggest advantages of Cloud Data Warehouse technology consists in the automatisation of many manual tasks, such as:

- Managing encryption;
- Automatic determination of data distribution;
- Delivering automatic performance optimizations;
- Deployment;
- Limitation of users;
- Affordability;
- Volume and pace of data;
- Data managing.

It could be concluded, that cloud technology provides cheap storage and on-demand computation with the lowered cost of complexity and lengthy time-to-value, which has limited the adoption and successful use of traditional Data Warehouse technology. Moreover, a cloud Data Warehouse offers fully independent scaling for computing, storage, and services.

The modern Data Warehouse changes Business Intelligence by providing effective and powerful ways to achieve new trends and adapt quickly to external changes with performance increasing through time. Consequently, the cloud Data Warehouse replaces the traditional warehouse as provides business agility to organizations by automatizing maintenance tasks and giving the flexibility to change according to the business needs in the rapidly changing world.

3.1.3. Online Analytical Processing

Online Analytical Processing (OLAP) is the capability a manipulate the data for performing multidimensional analysis at high speeds on large volumes of data from a Data Warehouse, Data Mart, or another data store. In the literature, OLAP is defined as a computational approach to answering multi-dimensional analytical queries and is part of Business Intelligence, which also encompasses relational databases, report writing and data mining (Deepak P., 2007). Typical applications of OLAP include business reporting for sales, marketing, management reporting, business process management, financial reporting and similar areas (Queiroz-Sousa, P.O. and Salgado, A.C., 2019).

In this section, the differences in OLAP implementing in On-Premises and cloud environments will be shortly explained.

In On-Premise, the data for OLAP purposes must be firstly properly prepared. The subsets of the data from the Data Warehouse are extracted, organized and loaded in the OLAP cubes or custom cubes. This cube is defined by several dimensions and measures. Cubes usually require significant technical resources to building and re-building the cube, as well as some times manual configuration in case of new data (Queiroz-Sousa, P.O. and Salgado, A.C., 2019). On some cubes is possible to handle business complex models but there is another limitation such as the impossibility to scale large data sets with high cardinality.

Cloud SaaS BI solutions offer an automated approach to the analysis and move the data online with an integrated On-Line Analytical Processing (OLAP) engine. Cloud OLAP provides advantages such as performing ad-hoc queries in a second with important components of analysis perspective, scaling large datasets and quick response to new data. However, on the market, there are few cloud providers for OLAP, the most known are Microsoft Cloud Azure and AWS Cloud Servers of Amazon.

Table 1 represents a resumed comparison of the main requirements and characteristics of the implementation of Cloud OLAP and traditional On-Premises OLAP and On-Premises Data Marts.

Table 1. Comparison of OLAP Implementation

Requirement	Cloud OLAP	On-Premises OLAP	On-premises Data Marts
Ad hoc query performance under 1 second	Yes (Automatic)	Yes (Automatic)	No. Manual engineering is required and sub-second response time is impossible.
Universal Semantic consistency & governance	Yes	Only for Tools of MDX	No
Ability to handle complex business models	Yes	Yes	No, SQL is not suitable
Scale to large datasets with high cardinality	Yes	No (pre-computation /memory limited).	Yes. Manual data engineering is required.
Predictable cloud costs	Yes	Yes for Azure but N/A for other clouds because they don't offer OLAP services.	No, require control input queries.
Quick response time to introduce new data	Yes	No. Must rebuild cube and change ETL.	No. Requires manual data engineering &ETI modifications.
Hybrid cloud support	Yes	No. Only Azure supports OLAP & other clouds don't.	No, except for Snowflake. Manual data engineering & ETL are cloud-specific.
Unified security & authentication	Yes	No. A separate data store requires separate security.	No. Each BI tool requires different security settings.
Heterogeneous BI tool support w/ live connection	MDX & SOL	Only MDX.	Only SQL.
Avoids redundant data	Yes	No. Requires Cube build.	It depends. Usually requires different tools.

3.2. Critical Questions

The Business Intelligence components in the On-Premises and Cloud perspectives have several benefits, as well as limitations, but the choice of paradigm depends on the company's business needs. However, in any case, two critical issues must be analyzed carefully when the company chooses the path to follow: Security and Cost Management.

Each of the two options, On-Premises or Cloud, provides completely different approaches to managing these critical components, and companies need to pay attention to them as the decisions made will affect long-term prospects.

In this section, the critical questions arising from an On-Premises and Cloud perspective will be analyzed.

3.2.1. Security

Information security is a part of information risk management and involves preventing or reducing the probability of unauthorized access, use, disclosure, disruption, deletion, corruption, modification, or recording. The implementation of information security is different from the cloud and On-Premises paradigms.

Cloud provides multiple services where each microservice runs in every point of the world and is linked with another environment through the network. So, Cloud computing information security is organized in a different way from On-Premise since its components require more security concepts involved. Those concepts need to be well implemented to avoid the protentional risk of unauthorized data access.

The three main levels of security are:

1) Overall cloud security. At this level, there is a need to ensure that the environment from which the cloud is accessed is unbreachable and secure. Also, at this level, there is a need to ensure that each microservice on the cloud is protected as those microservices have accessed from an external entity and received external data. An example is user identification security and encryption.

2) Level is the application or microservice security. On this level is ensured that the application code and microservice are secured. Examples are API Security and application code security.

3) Level is the security of the architecture of microservice. There are 2 security components: Cloud Security Posture Management (CSPM), and Cloud Workload Protection. Cloud Security Posture Management (CSPM) is a security tool that is designed to identify misconfiguration issues and compliance risks in the cloud. An important purpose of CSPM programming is to continuously monitor cloud infrastructure for gaps in security policy enforcement. These security tools are usually used by organizations that have adopted a cloud-first strategy and want to extend their security best practices to hybrid cloud and multi-cloud environments. Cloud Workload Protection is the process of keeping workloads that move across different cloud environments secure. In the cloud, the workload includes the application, data generated by or entered into an application, and the network resources that support a connection between the user and the application. A cloud-based application will not function properly if any part of the workload is compromised. Workload security is especially complicated in hybrid data centre architectures that employ everything from physical, On-Premises machines to multiple public cloud infrastructures as a service (IaaS) environments to container-based application architectures. Cloud workload security is particularly complex because as workloads pass among multiple vendors and hosts, the responsibility for protecting the workload must be shared.

Cloud: Could provide the category of services depending on the service model. Starting from IaaS to SaaS where at each next level of the service model the cloud provider has more control and responsibility for the components provided and at the same time user control and responsibility decrease. On the cloud security implemented the shared responsibility concept which defines that each participant: cloud provider or user has the security responsibility on those components with full control, depending on the cloud service model. For example, on PaaS, the cloud responsibility of security is ensuring that the workloads run on the side of the platform for developing the application are secured but for other workloads, which ran out of the platform on the user or another vendor's side the information security responsibility of those workloads is also them.

Another concern that exists on the cloud is data security as several applications are involved in transferring decision support data between the cloud and networks, especially on Cloud BI. To ensure security the data is encrypted using services such as Keep Your Own Key (KYOK) and Bring Your Own Key (BYOK). The key management from the user and cloud provider is very important. KYOK is a security service which can be used to encrypt data encryption keys (DEKs) that are used to encrypt data in services. BYOK enable public cloud users to maintain control of the cryptographic keys used in the cloud to keep their data secured. KYOK enables public cloud users to generate their high-quality master key locally On-Premises, and securely transfer the key to their cloud service provider (CSP) to protect their data across multi-cloud deployments. To generate and manage high-quality keys, BYOK uses FIPS and Common Criteria-certified hardware security modules (HSMs) that the cloud user maintains On-Premises or leases as a service.

However, there are other problems which still important and need to be considered such as:

- Virtualization security problems
- Internet and services-related security issues
- Network security problems
- Access control problems
- Software security problems
- Trust management problems

On-Premises: On another side, On-Premises information security is different as there isn't any external access to the internet and all processes such as data transfer, the workload runs in a private zone inside the organization. The focus of information security here is to protect the organizational system to not authorized use. This is accomplished by firewalls, VPN, User Authentication, etc. On On-Premises data is not passed through the internet meaning that there is no need for encryption security. The information security risk in these conditions is lowest compared to the cloud. In the On-Premises, the user has overall control of the data and processes. Users can monitor in detail what's happening on the system and in case of any threat stop all the processes immediately. Also, the On-Premises architecture allows properly configure security and accomplishing the regular standards while on the could security is configured depending on the provider.

However, there are On-Premises security risks such as the physical threat of hardware systems in the case of disaster or human factors when employees transfer sensitive information to competitive organizations.

3.2.2. Cost management

Costs management is another critical issue with enterprises have in the decision between On-Premises and cloud.

Cloud solutions provide that users, on the one hand, don't spend the cost of building, and maintaining the availability of computing resources by providing those resources. Public User only has the costs of system implementation and supports costs based only on the number of resources used accordantly the "pay-what-you-use" model or through a monthly basis subscription depending on the type of service model used. Below represented a table of comparison Active Project Costs with Hidden/Outgoing costs.

Table 2. On-Premises and Cloud BI Costs

On-Premises		Cloud	
Active Project Activities Costs	**Ongoing Support Activities**	**Active Project Activities Costs**	**Ongoing Support Activities**
Hardware Software Licenses System Implementation	Hardware Upgrade & Refresh Software License Maintenance Technical System Administration Patch Management Performance tuning Training	System Implemention	Subscription Fee Configuration Changes Training

However, in the On-Premises approach, the main costs of building the solution are at the first phase when the computing storage needs to be constructed first. The maintenance costs exist but, in some cases, the total costs are low compared to the cloud annual payments. Also, important to consider that the final costs from short and long perspectives depend on the cloud service model, cloud deployment model and the business needs. To have reasonable expenses organizations need from a cloud perspective to perform an estimation of costs based on business needs and manage by scaling or down the resources and constantly monitoring the expanses of use of resources by cloud monitoring tools which cloud providers give. On another side on On-Premises, the support activities costs are stable and do not require special control and this provides the possibility to perform higher calculations with fewer costs compared to the cloud. At the same time, the costs can increase due to the scalability process.

4. TOOLS

This section represents the tools and features used during the internship on cloud and On-Premises environments to fulfil common routine Business Intelligence processes such as ETL, reporting, and analysis. The functionality of each toll will be described in the context of utility.

Oracle Database:

Oracle is a Relational Database Management System (RDBMS) created in 1970. This database is highly scalable and provides main features such as partitioning, logical data structure, and Memory caching which makes it very popular and used widely by many organizations. In Oracle are used SQL language is in data manipulation and properly created language PL/SQL which is used on the stored procedures and during the transactional processing.

Oracle Exadata:

Exadata is a Database Machine created in 2008. Consists of database and storage servers with connected by active Remote Direct Memory Access over Converged Ethernet (RoCE) internal network fabric. Exadata runs the main workloads of databases such as OLTP, Data warehousing. The processing capability and speed depend on the configuration selected.

PG Admin:

PG Admin is a web-based Graphical User Interface (GUI) management application used to communicate with Postgres and derivative relational databases on both local and remote servers. PG Admin uses the PostgreSQL language and help PostgreSQL users get the most out of their database. The purpose is to provide a graphical administration tool to make it easier to manipulate schema and data in PostgreSQL (or even multiple installations thereof). Also, PG Admin provides several features such as working across operating systems, compatibility with all versions of PostgreSQL and EDB Postgres Advanced Server, a variety of deploying and query tools for features for faster data entry, debugging, and more. This GUI is used to connect to the cloud database PostgreSQL instance and enable full control of execution processes, the status of operation and automatizing the jobs. This client helps manage the Postgres database processes such as triggers, stored procedures, jobs executions, monitoring the database processes and querying.

Visual Studio:

Microsoft Visual Studio is an IDE made by Microsoft in 1997. Used for different types of software development such as computer programs, websites, web apps, web services, and mobile apps. It contains completion tools, compilers, and other features to facilitate the software development process. For BI purposes Visual Studio is used as a reporting tool to provide analytical reports and ETL tools. During the Internship, the Visual Studio will only use for Reporting building.

Toad:

This is a client application which enables connection to the On-Premise DW and Transactional databases. By Toad users can execute complex queries over DW, Create scheduled Jobs on running specific tasks, and create stored procedures. This tool is important in the context of the internship as this will be used to construct complete queries to the reports, and create and schedule the ETL Process.

Structured Query Language

SQL:

According to ANSI (American National Standards Institute), it is the standard language for relational database management systems. SQL statements are used to perform tasks such as updating data on a database or retrieving data from a database. Some common relational database management systems that use SQL are Oracle, Sybase, Microsoft SQL Server, Access, Ingres, etc.

PL/SQL:

PL/SQL or Procedural Language for SQL is an extended version of SQL programming, designed specifically for Oracle relational databases, to be programmed alongside SQL and Java. As a programming language, it is derived from SQL and incorporates object-oriented programming (OOP) concepts such as procedures, functions, loops, conditional statements, etc. Also, this programming language consents to variables & constants declaration, similar to Java and other OOP-based programming languages. PL/SQL will be used on Oracle to affect the ETL through stored procedures.

5. PROJECTS

This section contains a description of the BI projects which were implemented on On-Premises and cloud environments. The projects will elaborate on the main tasks of BI in those environments to measure and compare the uselessness of BI in the context of cloud and On-Premise.
The core tasks will be conducted as follows: Configuring the ETL process from an external (Production Data source), and Creation analytical reports.

It will be presented a short introduction to Business Intelligence from the view of the enterprise: how Business Intelligence works and communicates with all organization departments, the value provided to the company and how BI helps to gain business agility. After the introduction, the hardware which will support the execution of the projects will be described.

Hardware infrastructure

During the internship were used the computer DELL L-3420 Laptop with Operating System Windows 10 and 11 Gen Intel(R) Core(TM) i7 processor.

The DW, notification platform and information systems of Company **A** were assessed as follows:

- Through the corporate network in the office, where the PC is connected to the network provided;
- Through a Virtual Private Network (VPN) where is possible to connect to the private network of Company **A** out of the office. This approach is used in "hybrid work", where the working days are rotated (2 days work from home, 3 days work from the office).

5.1. ON-PREMISES BUSINESS INTELLIGENCE PROJECT

Company **A** is an international software development company which operates in the telecommunication market serving as the intermediary between the telecommunication operator and the client. Besides providing technical architecture, software, and marketing services, the enterprise provides analytical services to promotional campaigns, such as analyzing customer trends, offering efficiency, estimating costs, etc.

The promotional campaigns are the main Business process of company **A** which involves all critical departments and systems. The On-Premises Business Intelligence plays an important role in promotional projects as started evolving in the company from the beginning over the years have stabilized the necessary data source for the enterprise and provided the analytical capacity to the top and middle management.

The central Data Wharehouse of Company **A** is based on Oracle Environment Database and receives the data from various Oracle Environment production databases located in different countries and conciliates it into a single version of truth containing relevant data of all business scopes of the company. The DW for warehousing workloads contains the specialized computing platform Exadata. Exadata Database Machine is a database architecture which consists of scale-out industry-standard database servers, and scale-out intelligent storage servers which help runs multiple complex analytical queries with faster speed. Production databases are also based on the single Oracle Environment and linked with each other and with DW through a single corporation network and accessed through Client based Software "Toad".

The Oracle Database Environment enables databases with corporation networks direct access from one database to another without directly logging through a database link. The ETL Process uses the Client based Software “Toad” as intermediary software where each production table is directly accessed, data transformed and loaded to the DW.

All departments and top executives access the relevant business data stored on DW through Business Intelligence analytical features such as data analytical reports, OLAP On-Premises Cubes and Alarms. The data analytical reports are stored on the Microsoft Reporting server where each report queries the data from DW and is accessed by all enterprises.

The departments establish communication with the BI department through a notification platform, where new requests are created. The Business Intelligence team are notified and responds to new requests in less than 24 h.

This project is a part of the main Business Intelligence projects involved in analytical campaigns where the functionality of BI On-Premises is strong.

Use case: BI department received the request from the marketing department: there is a need to analyze the promotional campaign to track the client's flow of receiving and following the game steps of the campaign of Uruguay, the correct answer to questions user wins more points. There is a need to receive on daily basis the statistic of users' steps in aggregational form by region and date and directly consult the report at any time and download full data.

5.1.1. ETL Process

Company **A** implements ETL through the stored procedures inside DW with direct access to the production database through database links.

Oracle database link is a pointer that defines a one-way communication path from an Oracle Database server to another database server. This allows to users access a link to a remote database without having to be a user on the remote database.

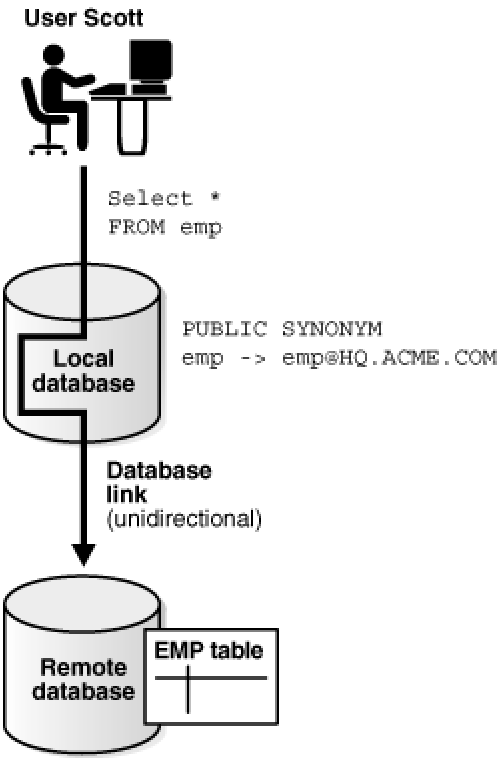

Figure 9. User access through a database link

The Procedure is created via Toad on thatch accessed the production databased and extracts the data at a specific time by special scheduled jobs within a certain time executes the stored procedures

Company **A** has pre-defined main tables containing the main information of all business processes but sometimes appear a new project where new tables are required to be added or procedures created. This project is a promotional campaign business process which contains a special structure with is not defined in the DW schema. For this reason, new tables need to be created and procedures need to be created.

As company **A** has distributed database system, the link must contain the global database name. Global database name consists of prefixing the database network domain, with the individual database name.

The tables for this campaign are defined as follows:

step_table(customer_id, step_type,step_date,questions)

sms_billings(customer_id, date,total_charged)

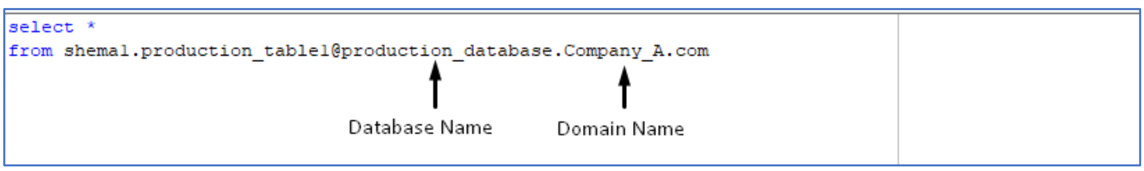

Figure 10. External tables accessing

The main stage is to create an ETL process, which will be accomplished by creating the stored procedure on the side of DW. The stored procedure is a series of SQL statements compiled and saved on the database. Stores procedures contain conditional programming elements such as if-else programming conditions and “for”, and “while” loops, and accept the parameters. In Oracle Stored Procedures is used PL/SQL language.

```
create  or replace procedure schema_DWH.UPDATE_STEP(vdate  IN DATE)
IS

reference_date VARCHAR2(10);

BEGIN


EXECUTE IMMEDIATE ('alter session set time_zone= ''America/Montevideo''');


/*-----DELETE------*/
DELETE
FROM schema_DWH.step_date
WHERE step_date>=TO_DATE(reference_date,'ddmmyyyy')-1
      AND step_date<to_date(reference_date,'ddmmyyyy');

      COMMIT;


/*----INSERT-----*/
INSERT INTO schemal.step_table
SELECT customer_id,step_type,step_date,questions_name,Correct_Answer,points_winned
FROM schemal.step_table@production_database.Company_A.com
WHERE
           step_date>=
           TO_DATE(reference_date,'ddmmyyyy')-1
           AND step_date<TO_DATE(reference_date,'ddmmyyyy');

           COMMIT;


           END;
```

Figure 11. On-Premises stored procedure

This procedure defines the ETL which gets the data from the production table to the DW on the daily basis. The PL/SQL code of the stored procedure changes the database time to the local time zone of the promotional service country in order to avoid data silos and inconsistency. In the case, that the procedure was executed more than 1 time during 1 day, the previously inserted data is deleted using the variable *reference_date* with stores the Oracle date/time function - *current_date* . After the data is accessed to the production database table through the database link and inserted directly into the destination table.

To execute those stored procedures automatically every day, the scheduled job needs to be created and configured. The Oracle Scheduled jobs is a script which receives the specified function/stored procedure and executes it automatically at a specified time.

The program type will be PL/SQL Block in which the stored procedure is created before it consists of. The store procedure is called by passing the 1 parameter defined which is *current_date*. After the time interval of the job, execution needs to be defined as follows: *FREQ=DAILY*; *INTERVAL*=1.

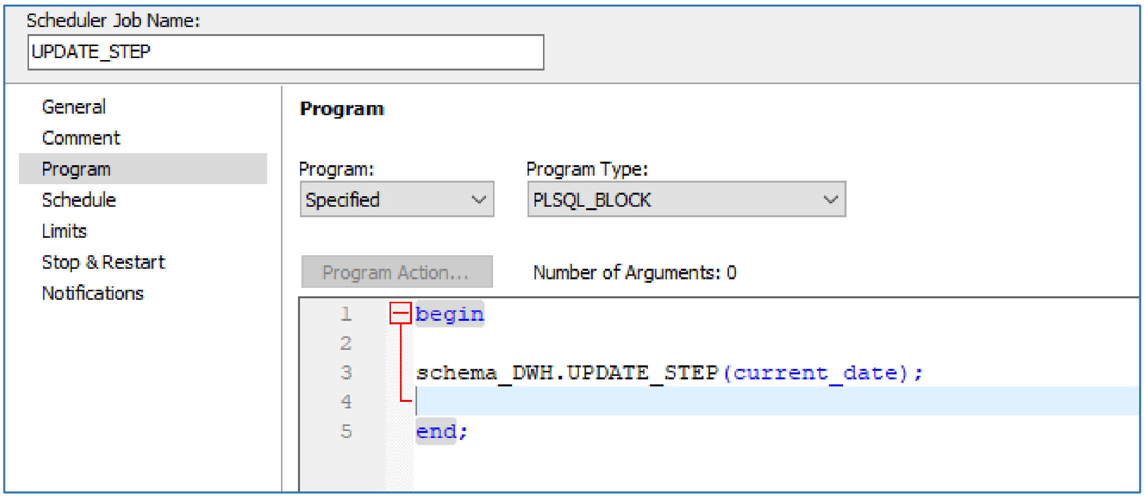

Figure 12. Oracle job stored procedure call

5.1.2. Data Warehouse

The On-Premises Data Warehouse is based on Oracle Environment which retrieves business data from production tables. The DW business data is consulted through 2 forms: Reports and direct SQL queries executed on the DW environment through client software Toad. The DW execute the complex queries through the database computing engine Oracle Exadata. Oracle Exadata is a database machine and consists of database and storage servers which enable to execution of complex calculations with a fast speed of data retrieval. The number of database and storage servers depends on the configuration chosen. There are 4 configurations: eight racks, quarter pack, half rack and full rack. Company **A** has the recent version of Exadata x9-m with quarter pack configuration with Hight Capacity Storage Server.

Exadata Rack Configurations [1, 2]

RACK SIZE	DATABASE SERVERS AND CORES	STORAGE SERVER AND CORES	HIGH CAPACITY STORAGE SERVER CAPACITY (RAW)		EXTREME FLASH STORAGE SERVER CAPACITY (RAW)
Eighth Rack [3]	2 x servers, 64 cores	3 x servers, 48 cores for SQL offload	324 TB disk, 38.4 TB Flash, 2.3 TB Persistent Memory	or	n/a
Quarter Rack	2 x servers, 128 cores	3 x servers, 96 cores for SQL offload	648 TB disk, 76.8 TB Flash, 4.5 TB Persistent Memory		153.6 TB Flash, 4.5 TB Persistent Memory
Half Rack [4]	4 x servers, 256 cores	7 x servers, 224 cores for SQL offload	1512 TB disk, 179.2 TB Flash, 10.5 TB Persistent Memory		358.4 TB Flash, 10.5 TB Persistent Memory
Full Rack [4]	8 x servers, 512 cores	14 x servers, 448 cores for SQL offload	3024 TB disk, 358.4 TB Flash, 21 TB Persistent Memory		716.8 TB Flash, 21 TB Persistent Memory
+Database Servers	Up to 19x servers[5], 1216 cores max per rack	n/a	n/a		n/a
+Storage Servers	n/a	Up to 18x servers[5], 576 cores max per rack	3888 TB disk, 460.8 TB Flash, 27.6 TB Persistent Memory		921.6 TB Flash, 27.6 TB Persistent Memory maximum per rack

Figure 13. Oracle Exadata configurations

The Exadata server is activated on SQL Statement through Oracle hint. Oracle hint is an optimizer code embedded into an SQL statement before defining columns and indicates to Oracle Engine how the statement should be executed. On the BI requests when the data retrieve operation is relevant from DW the /*+ PARALLEL */ Oracle hint is used. The parallel hint defines the number of execution processes attributed to each database server of Exadata. The parallel Oracle hint enables fast data retrieval from SQL queries. On the internship, will be used Oracle /*+ PARALLEL (8)*/ hint where 8 defines the number of execution processes executed on each Exadata database server.

5.1.3. Reporting Component

The presentation layer of business architecture plays an important role for Company **A** in all business processes which includes the capabilities such as OLAP, Reporting, direct querying through direct access to the DW and Data Mining

The data stored on DW is aggregated into the analytical cube with enables users access to the analytical information and business insights. The DW enables quieting capability where technical users from IT and departments can immediately extract the necessary business information. The reporting capability represents by business alarms with occurred inside the DW and reports. The business alarms collect information from tables on recall time and monitor the critical business units and projects. The visualization and analytical reports are stored on a Microsoft Power BI report server. Microsoft On-Premises Power BI report server is defined as a platform for publishing, sharing, and consuming Business Intelligence contained within the organizational network. Analytical visualization reports play one of the highest business roles of BI. The report published is accessed by more than 500 users those reports represent any kind of information in a variety of forms. For example, getting information insight into the recent promotional campaigns with special rules is unavailable on the OLAP cube. In this project will be created 2 paginated reports and deployed to the report server.

The first report request for the promotional campaign was a create report which counts the users aggerated by step_type date and region. The report query created uses the table migrated and the table of regions and is represented as follows:

```
select /*+ PARALLEL (8)*/ r.region_name,trunc(step_date) as step_date,
count(customer_id) as steps_quantity,
count(correct_questions_ansvered) as number_correct_questions_ansvered,
sum(points_winned) as total_points
from shemal.step_table st
join shema_region.region r
on substr(st.customer_id,0,3)=r.id
group by r.region_Name,trunc(step_date);
```

Figure 14. Matrix report query

The Visual Studio report can be taken a variety of forms: Tabular, Matrix, or Waterfall. Matrix report provides a scale-out vision with enables the facts (quantity of steps, answers) seen from other visions and business grains. In this report, the regional department of company **A** will consult the performance of the campaign in the regions while the marketing department will use the information on the overall performance through dates.

Matrix Promotional Report

Region	01/09/2021			02/09/2021			03/09/2021		
	Quantity Steps	Correct Answers	Points Winned	Quantity Steps	Correct Answers	Points Winned	Quantity Steps	Correct Answers	Points Winned
Artigas	8493	3098	32509	9894	9361	36738	579	8568	61135
Barros Blancos	4407	1987	50391	7105	6084	4276	2701	165	7764
Canelones	125	551	65472	153	2125	47980	6981	6573	87074
Ciudad de la Costa	7011	6057	38057	8040	1632	50119	9439	7959	43917
Colonia del Sacramento	903	1207	26446	1190	3557	67546	4532	2654	1
Durazno	7156	6743	63309	6351	3490	8513	6512	4403	67842
Florida	3048	2013	50230	930	7519	79404	4597	6948	20087
Fray Bentos	2448	3654	47443	7794	546	1652	918	4426	24559
Las Piedras	1937	9940	31481	4265	5448	47289	3427	9163	34357
Maldonado	1309	3911	63917	8097	4744	91702	5608	1683	43111
Melo	9178	1699	11011	6979	9118	68641	1286	7608	40039
Mercedes	8633	9971	28625	1623	7146	79041	5744	1329	60366
Minas	2244	1594	3432	816	2134	80336	8573	1097	60960
Montevideo	2857	2610	59337	2074	8011	9420	3587	2128	47759
PaysandÃº	8121	6668	20170	3480	9729	48625	1440	1304	35554
Rivera	1097	4560	78521	3262	8877	87596	7359	995	30147
Rocha	3606	1458	6533	2471	9098	23889	711	8262	22493
Salto	5252	173	76545	7273	1152	28788	8146	6081	23900
San JosÃ©	8445	8354	73821	6272	2892	39499	886	2403	99804
TacuarembÃ³	252	4298	50559	7457	843	4897	2850	7279	37315
Treinta y Tres	1838	5881	92569	3839	2876	11912	7971	2055	30101

Figure 15. Matrix promotional report

Tabular reports on the other hand have another benefit providing information which can't be represented in aggregated form. This kind of information can't be represented in the OLAP analysis.

The tabular report was implemented on the request too where there is a business need to monitor the statistics of the game of each user per date and region. The second report will be represented the statistics of the progress of the user: the quantity step completed, and the challenge questions answered by a user on region and date.

```
 select /*+ PARALLEL (8)*/  r.region_name,
 trunc(step_date) as step_date,
 customer_id,
 count(*) as steps_quantity,
  sum(correct_questions_answered) correct_questions_answered,
  sum(points_winned)points_winned
from shemal.step_table st
join shema_region.region r
on substr(st.customer_id,0,3)=r.id
group by r.region_name,trunc(step_date),customer_id;
```

Figure 16. Tabular report query

Tabular Promotional Report

Region	Step Date	Customer ID	Steps Quantity	Correct Questions Answered	Point Winned
Artigas	01/09/2021	9555	7	2	190
Barros Blancos	01/09/2021	6179	2	2	2128
Canelones	01/09/2021	7805	4	5	2090
Ciudad de la Costa	01/09/2021	6868	1	3	2096
Colonia del Sacramento	01/09/2021	3410	3	1	761
Durazno	01/09/2021	9093	10	6	4373
Florida	01/09/2021	6167	2	10	3690
Fray Bentos	01/09/2021	9255	6	1	2644
Las Piedras	01/09/2021	5620	6	5	3368
Maldonado	01/09/2021	3014	1	2	2200
Melo	01/09/2021	7707	6	3	91
Mercedes	01/09/2021	9437	8	2	2531
Minas	01/09/2021	9648	7	3	754
Montevideo	01/09/2021	7351	9	4	3414
PaysandÃº	01/09/2021	2036	8	4	3033
Rivera	01/09/2021	2400	3	1	408
Rocha	01/09/2021	6707	1	1	2235
Salto	01/09/2021	5013	8	4	4629
San JosÃ©	01/09/2021	6317	8	9	2402
TacuarembÃ³	01/09/2021	1513	10	10	4291
Treinta y Tres	01/09/2021	3922	1	4	1252
Artigas	02/09/2021	9418	5	5	4633
Barros Blancos	02/09/2021	1638	9	6	565
Canelones	02/09/2021	1541	10	3	3468
Ciudad de la Costa	02/09/2021	3860	2	10	3831
Colonia del Sacramento	02/09/2021	9356	6	2	1106

Figure 17. Tabular promotional report

The statistic of each user isn't seen on the OLAP cube or aggregated Power BI reports witch the report is directly provide

The benefits of the 2 reports and the On-Premises here are that the user at any time cloud consults the data but visualises it not only on the report server but downloading at any time.

5.2. Google Cloud Business Intelligence Project

Company **A** recently started using components of Google Cloud Platform (GCP) and GCP Power BI visualization Reporting Platform and elaborating new projects of analyzing short-term campaigns. Those projects are pilot projects where only basic features of the reports and DW are implemented.

The cloud architecture of Company A consists of central Google Stored Databased which receives the data from other cloud production tables. The DW is based on the Google SQL Database part of GCP which is defined as a fully manageable SQL database. This cloud database cloud based on PostgresSQL and the instance is accessed and manipulated through the *PG Admin* client software. The data is processed through stored procedures implemented in *Postgres* environment without using cloud tools.

The reporting platform is based on the power BI report server. This platform is specially dedicated to the power BI reports where the reports are based on Power BI Desktop Reports (ptcx) with published into the Power BI workspace. In the Power BI workspace, there are defined datasets and analytical dashboards and analytical Power BI Reports.

Use case: 2-month promotional services started in Mexico and there is a request for implementing the analytical power bi reports to an external group of clients. The objective is a have full control over the billings affected by promoted services through aggregated and interactive information and visual capabilities of Power BI. The project is implemented on the cloud to have to access the data with a low risk of DW unsuspected failure.

5.2.1. ETL Process

The ETL process similar to On-Premises will be conducted inside the PG Admin through a stored procedure. Therefore, proceeding to the ETL process complaint to the Oracle On-Premises DW, some configurations on Postgres are required.

The data souses of this project are cloud SQL Postgres google production databases with a limited capacity of storage.

The external table is defined as follows:

Service_Billings(service_id, customer identifier, region, billing_quantity,date_of_billling)

On Postgres SQL the external tables are referenced through foreign tables. The foreign tables are accessed through the *Data wrapper*. A foreign data wrapper is defined as an extension which enables it to connect to the external database and retrieve the data.

The data wrapper is created on the Postgres database with connects to the database:

```
CREATE FOREIGN DATA WRAPPER postgres_fdw
    VALIDATOR public.postgres_fdw_validator
    HANDLER public.postgres_fdw_handler;

ALTER FOREIGN DATA WRAPPER postgres_fdw
    OWNER TO admin;
```

Figure 18. Foreign data wrapper creation

```sql
    CREATE SERVER cloud_prod_server
    FOREIGN DATA WRAPPER postgres_fdw
    OPTIONS (host '10.10.10.200', port '5000', dbname 'cloud_prod_db');

ALTER SERVER cloud_prod_server
    OWNER TO admin;
GRANT USAGE ON FOREIGN SERVER cloud_prod_server TO user1;
GRANT USAGE ON FOREIGN SERVER cloud_prod_server TO user2;
```

Figure 19. Connection to the external Postgres SQL Server

After the data wrapper creation, the connection to the server is defined by the host, post number and database name. These connection parameters are stored in the *pg_catalog.pg_user_mapping*. In the end, the server owners, grant usage of the server.

```sql
CREATE USER MAPPING FOR admin SERVER cloud_prod_server;
```

Figure 20. User mapping creation

Figure 20 represents the step of mapping creation which define a user connected to the database.

The final step is creating the following foreign table where this table is stored on table *pg_foreign_server* catalogue.

```sql
CREATE FOREIGN TABLE IF NOT EXISTS foreign_schema.Service_Billings(
    service_id bigint NOT NULL,
    region character varying(255) NOT NULL COLLATE pg_catalog."default",
    customer_identifier bigint NOT NULL,
    date_of_billling timestamp without time zone NOT NULL,
    billing_quantity        bigint
) (service_id, customer_identifier, region, billing_quantity,date_of_billling
    SERVER prod_db
    OPTIONS (schema_name 'mep', table_name 'bundle');

ALTER FOREIGN TABLE foreign_schema.Service_Billings
    OWNER TO admin;

GRANT SELECT ON TABLE foreign_schema.Service_Billings TO user1;

GRANT ALL ON TABLE foreign_schema.Service_Billings TO admin;
```

Figure 21. Foreign table creation

The ETL process is conducted through the stored procedures inside the PG Agent client aiming to extract the data from the external table to the Data Warehouse environment.

```sql
CREATE OR REPLACE PROCEDURE schema1.Service_Billings(
    )
LANGUAGE 'plpgsql'
AS $BODY$
DECLARE
      vnow        timestamp;
      v_after     timestamp;

   BEGIN
        set timezone = 'america/costa_rica';

      SELECT current_date - interval '1 day'
        INTO vnow;

        SELECT current_date + interval '1 day'
        INTO v_after;


        truncate table schema1.Service_Billings_Stage;


        insert into schema1.Service_Billings_Stage(service_id,region, customer_identifier,date_of_billling, billing_quantity)
        select service_id,region,customer_identifier,date_of_billling,billing_quantity
        from foreign_schema.Service_Billings
        where  date_of_billling>=vnow and date_of_billling<v_after;


        insert into  schema1.Service_Billings (service_id,region, customer_identifier,date_of_billling, billing_quantity)
        select service_id,region,customer_identifier,date_of_billling,billing_quantity
        from schema1.Service_Billings_Stage;
```

Figure 22. Cloud ETL stored procedure

On the stored procedure the language is defined as '*plpgsql'* to enable declaring the variables in the body of the procedure where variables *vnow* and v_*after* receive the 1-day interval. The procedure extracts external data by consulting the public table and the data into the staging table by extracting only for time intervals specified at stored procedure variables.

5.2.2. Data Warehouse

The operational databases and business data storage service are provided by Google Cloud as a Google SQL Database. According to Google's information, the Google SQL Database is defined as a "Fully managed relational database service for MySQL, PostgresSQL, and SQL Server with extension collections, configuration flags, and developer ecosystems the instance is configured on google cloud". The main benefit is reducing maintenance costs with a fully saved and operational database, continuous service backups, automated database provisioning, storage capacity management and integration with Cloud Google apps such as Big Query.

Company **A** has elaborated on the service Google SQL Database instance on PostgresSQL which is configured and accessed from the PG Admin application. PGAdmin is a web-based Graphical User Interface (GUI) management application used to communicate with Postgres and derivative relational databases on both local and remote servers. This client helps configure the computational queries such as triggers, stored procedures, and jobs executions, monitoring the database processes and enabling the querying ability.

5.2.3. Reporting Component

Reporting and the analytic component of the Cloud BI of Company **A** are provided by a different vendor - Microsoft. Microsoft provides cloud-based BI services, known as "Power BI Services," which includes a desktop-based interface "Power BI Desktop". The Platform contains a workspace which is dedicated only to Power BI Dashboards.

The primary focus of "Power BI Desktop" is a Business Intelligence service by enabling the creation of interactive dashboards and reports with monitoring alarms. The tool enables connection to various data sources such as Oracle, SQL, MYSQL, POSTGRES SQL, and analytical services.

One of the main benefits is that the Power BI report has interactivity. The user on the dashboards can interact with data, applying and choosing what is needed to visualize without waiting and at a very fast speed.

The necessary information from 2 tables will be uploaded on the single dashboard consisting of the Bar chart, Pie chart list, and Card. On the report building the connection to the PostgresSQL Database is stabilized by the import method entering the server IP, and the database name.

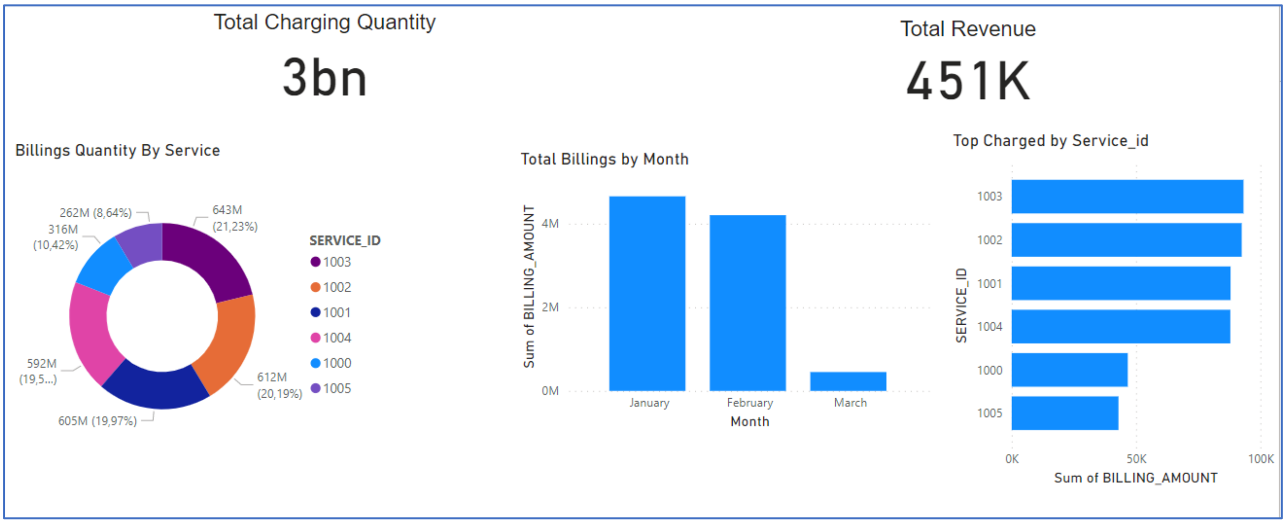

Figure 23. Power BI Promotional Dashboard

The report is published on the report Power BI server on the defined workspace where this workspace is created, and the respective access is configurated.

6. RESULTS AND DISCUSSION

The two Business Intelligence projects were executed to provide the Business Intelligence functions to the main Business process such as Analysis of the Promotional campaigns. The On-Premises and cloud components involved here are different and have positive and negative impacts on the company. This section provides a Post comparative analysis of each component involved in those projects. At the same time, complete completion of the components is not possible as the cloud solution of company **A** not used on all BI components.

The ETL process of company **A** is very specific as on the On-Premises and in the cloud, the ETL process is executed through stored procedures which run under the DW environment. It originates that the characteristics of the DW have a strong influence. However, the first main benefit of the On-Premises ETL process of Company **A** is high processing speed. On-Premises DW during the ETL loads a high quantity of data with fast speed enabled by Oracle Exadata by using the Exadata Oracle Hints on PL/SQL stored procedure queries. On the opposite, in the cloud environment, the ETL is executed through a storage procedure which runs completely slowly.

Google storage PostgresSQL, on which data injection procedures occur, is designed for storage purposes and not designated for high and computationally complex computational expenses as compared to the Oracle Exadata. The Google Postgres Database is used as a part of Google Cloud tools which serves as the storage database of analyzed data used on complex data processing tools such as Big Query and as a data source where the data is read for processing. Also, the Google SQL Database stores the data from a variety of data sources: any On-Premises Data Warehouse or transactional processing databases such as Oracle, or Microsoft SQL. At the same time, the main limitation in the On-Premises system of Company **A** is data sources. The On-Premises DW based on Oracle at the ETL process can get only the datasets from other production databases which belong to the organizational network.

The On-Premises Data Warehouse have strong Business Intelligence relevance since it contains the business-relevant data for all processes from On-Premises production with data used on BI reports. On the other side, the Google Cloud Database contains only one promotional campaign Project. Besides this is not a direct technical comparison, the database's importance is measured by the relevance of the data and which data is received.

In terms of comparing the speed of querying, and complex PL/SQL query execution the On-Premises Database DW has better performance and higher speed than Google PostgresSQL Database.

The resumed comparison of main characteristics for On-Premises and Cloud BI components implemented in Company **A** is resumed in Table 3.

On-Premises DW contains an Oracle Exadata engine with consists of multiple database servers while at the same time, the google database is designed only for storage purposes and uses the memory and CPU (Central Processing Unit) for formal and medium queries. The On-Premises DW is central and receives the data from production databases containing the relevant business data. The Google SQL

Table 3. On-Premises and Cloud BI components characteristics

On-Premises BI	Cloud BI
ETL	**ETL**
Fast Data injection Transform and Load Using Fast Exadata.	Slow speed of injection Transform and Load through Google Cloud SL Postgres Instance
Available data sources are only Oracle production databases accessible through a single organizational network.	Google Cloud SQL Postgres database cloud receives a variety of data sources: Postgres databases, SQL management databases, and On-Premises transactional processing databases On cloud environment could receive data from cloud souses and can be accessible by other Google cloud tools for example Google cloud big query.
Strongly Dependence on DW configurations	Strongly Dependence on DW configurations
On-Premises ETL Process despite being implemented through On-Premises ETL tools is implemented on the Oracle environment which maximizes the On-Premises ETL by using Oracle Exadata.	ETL Process implemented without using cloud ETL tools at the same time implemented on the environment on which is not dedicated for purposes: Google Cloud PostgresQL Database.
DW and Exadata	**Google Cloud Postgres Database**
Hight relevant: Contains the relevant business data for all areas of Business of Company **A** and Receives the business necessary data from Production Oracle Databases.	At the moment low relevance on the business processes. The cloud DW receives the business data only from a specific project.
Power engine Oracle Exadata quarter Pack which allows executing complex Analytical queries with fast speed.	Contain only standard Database space, memory and CPU to process the analytical data. Not suitable for complex queries.
Database flails or Unavailability due to updates or failures solving.	Availability upgrades are fully managed by Google. Stable data source.
Exadata Required Updates every 3 years.	
Database Management Required	Cost Management Required
Reporting	**Reporting**
Fully managed Report server, anal updates, and risks of unavailability which is critical for Business Units.	Fully managed updates by Microsoft Vendor, Low risk of unavailability for critical Business Units
The unlimited extraction of the reports without cost management	Provide only Power BI reports. Advanced platforms may require the limitation of extracting due to cost management
On-Premises Report server supports a variety of reports: Aggregated reports, Tabular, Matrix, Waterfall, and Power BI reports, enabling the high scope.	Power BI Platform Supports only Power BI reports.
OLAP Availability is the strong influence of On-Premises DW.	Google Cloud doesn't provide OLAP Availability. On-Premises dependence on OLAP.

Database receives the data from only one production database. The speed of high processing with fast time is the main factor of booth DW where the ETL process is conducted in thought-stored procedures.

The main disadvantage of On-Premise DW with Exadata is its low availability. The DW on Company **A** is accessed not only through reports but also directly by multiple technical departments IT, DBA, and BI. This causes the risks of DW to go down. To avoid this problem, database management is conducted. Hight risk of database latency causes the unavailability of the whole department to access critical business data and the reporting capability and monitoring which has a strong business impact. Additionally, the business users will not always have updated business data due to the migration delays of Data Warehouse where the data need to be recovered from production databases. During the internship, this problem occurred once every 3 months, which means that in Company **A** the database management issues could have a frequency of such occurrences, approximately, 4 times per year.

Another problem of On-Premises Oracle DW is space allowance. 95 % of the space of DW is occupied which requires additional data management as data cleaning and moving the data to archive databases every two years. So, at these points, the Google Could PostgresSQL Database has the advantage due to its availability, and database management since it is fully managed by a cloud vendor.

The space could be scaled at any time but with an additional disadvantage: space could be only increased, not decreased or optimized, so the space needs to be calculated well to avoid unnecessary costs. Database availability and maintenance are the main disadvantage point of On-Premise DW**,** and the cloud DW solution will help reduce its risks by a minimum.

Additionally, the main limitation of On-Premises is data sources. The On-Premises DW based on Oracle gets only the datasets from other production databases which belong to the same organizational network. On the opposite, the Cloud SQL Database allows a variety of resources from On-Premises databases such as Oracle, Microsoft SQL, Postgres, Maria DB. In the cloud, the database receives data sources from the Google Cloud Platform such as Google *Big Query*, etc.

Reporting

For comparison of reporting capability of both systems, the reporting platform plays an important role in the reports consulting.

At Company **A** the On-Premises report server is managed, and updated, enabling to deploy the variety of reports such as tabular, matrix, waterfall reports and Power BI reports, as well as providing its assessment by a variety of business users. At the same time, Power BI Microsoft platform enables only deploying Power BI reports which represent limited information. However, the platform is fully managed by the Microsoft party which allows high availability. Availability plays an important role, as the long unavailability of business data is very undesirable, so the Cloud Power BI Platform provides low risk.

7. CONCLUSIONS

During the internship, the Business Intelligence capabilities were executed on On-Premises and Cloud environments.

Company **A** started recently implements the Cloud BI solution by using only some tools of Business Intelligence Cloud SQL Database using only on the small business scope. Cloud BI of Company **A** at the moment is not capable to process big amounts of data and the report cloud service doesn't provide the necessary business reports.

At the same time, On-Premises Business Intelligence components have a strong impact on Company **A** Data Warehouse. DW receive all business data from production data sources and processes high amounts of data on ETL and executes complex queries during reports. The report server requires management and monitoring while the risk of unavailability for business users is present.

The Oracle Exadata DW on Company **A** requires strong Database management and has a high risk of unavailability which causes a critical situation for business processes. During the Internship, the unavailability of DW occurred due to the database error and the Exadata Upgrade.

During the internship was verified that implemented On-Premises and BI Architecture have a high impact on Company **A** and provide high Business Intelligence capability despite the problems of low availability. On-Premises BI **A**. However, the Cloud BI represents a strong advantage of high availability which is relevant for Company **A**. To maximize the BI capabilities cloud architecture needs to be completed.

8. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE WORKS

During the internship, the main limitations were also identified.

Firstly, the Cloud Business Intelligence architecture implemented in Company **A** is not on full scale and only with small business scope in comparison with the On-Premises complex BI architecture.

The main limitation was found during the comparison of the DW BI Component and ETL Process. The Oracle On-Premises DW Based on Exadata has the scope by storing and processing Analytical data, while the Google SQL Database is based only on storage processes and is not dedicated to a high amount of processing. Due to these limitations, only descriptive and functional comparative analyses were possible.

In the future, Company **A** may use this internship report as input for a pilot study to improve the Cloud architecture components on critical pipelines: DW and ETL. The Cloud architecture has a big potential to transform the BI Structure of Company **A**, however, the decisions of migration from On-Premises to Cloud systems need to be taken. The study of migration from the On-Premises systems from SQL to the Cloud needs to be conducted on Company A considering the costs and evaluating which components on current on-remises BI need to be migrated.

9. BIBLIOGRAPHY

Amanatullah Y. et al., (2013), Toward cloud computing reference architecture: Cloud service management perspective, Proceedings - International Conference on ICT for Smart Society 2013: "Think Ecosystem Act Convergence", ICISS 2013

Aslan, I., et al. (2021) "Collaboration of Business Intelligence and Cloud Computing and Selecting the Best Cloud Business Intelligence Solution", Online Academic Journal of Information Technology 12(46) 29-45

(Alqarni, 2021) "A Secure Approach for Data Integration in Cloud using Paillier Homomorphic Encryption"

Ballard, C. Data Modeling Techniques for Data Warehousing, (1998) SG24-2238-00. IBM Red Book. ISBN number 0738402451.

Banica, L., Stefan, C., (2013), From Grid Computing to Cloud Infrastructures, International Journal of Computers & Technology, Vol 12, No.1, pp. 3187-3194.

Berkowitz, J. (2009), Cloud Computing (Part 1): Advantages, Types and Challenges, CRM Mastery Weblog

Birst (2010), Why Cloud BI? The 9 Substantial Benefits of Software-as-a-Service Business Intelligence, Birst, Inc.

Bonomi, F., Milito, R., Zhu, J., Addepalli, S. (2012) Fog computing and its role in the internet of things, in Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing, ACM, pp. 13–16.

Bowen, F., (2009), How SOA can ease your move to cloud computing, IBM.

Burtescu et al., (2014), Advanced Security Models for Cloud Infrastructures Journal of Emerging Trends in Computing and Information Sciences Advanced Security Models for Cloud Infrastructures. CIS Jornal

Casavant, T.L. and Kuhl, J.G., (1988) "A taxonomy of scheduling in general-purpose distributed computing systems," in IEEE Transactions on Software Engineering, vol. 14, no. 2, pp. 141-154, doi: 10.1109/32.4634.

Catteddu, D., Hogben, G. (2009), Cloud computing: benefits, risks and recommendations for information security, European Network and Information Security Agency

Ceborarian, E. (2008), "Jornal of Knowledge Management, Economics and Information Technology"

Compton, K., Hauck, S. (2002), Reconfigurable computing: a survey of systems and software, ACM Comput. Surv. (CSUR) 34 (2), 171–210.

Deepak P. (2007). Business Intelligence for Telecommunications. CRC Press. pp. 294 pp. ISBN 978-0-8493-8792-0.

Eckerson, W. (2003), Smart Companies in the 21st Century: The Secrets of Creating Successful Business Intelligent Solutions. The Data Warehousing Institute: Seattle, WA

Fabret F, Matulovic M, Simon E, (1997) DWQ Foundations of Data Warehouse Quality (DWQ) ,State of the Art : Data Warehouse Refreshment

Final Version of NIST Cloud Computing Definition, (2011), NIST Special Publication 800-145

Flynn, M.J. (1966), Very high-speed computing systems, in Proceedings of the IEEE, vol. 54, no. 12, pp. 1901-1909, Dec., doi: 10.1109/PROC.1966.5273.

Forsman, S., (1996) OLAP Council White Paper, OLAP Council, San Rafael, CAwww. olapcouncil.orglresearch/whtpaply.htm.

Gill S, Tuli S, Garraghan P, (2019), Transformative effects of IoT, Blockchain and Artificial Intelligence on cloud computing: Evolution, vision, trends and open challenges,Internet of Things

Griffith, E. (2015), What Is Cloud Computing?

Gutiérrez A, Marotta A, (2000), An Overview of Data Warehouse Design Approaches and Techniques

H. Dresner, J. Ericson (2015), “Cloud Computing and Business Intelligence market study”, Wisdom of crowds series, Dresner Advisory Services

Hajmoosaei A, Kashfi M, Kailasam P, (2011), Comparison plan for Data Warehouse system architectures, The 3rd International Conference on Data Mining and Intelligent Information Technology Applications, 290-293.

Inmon, W. H., Imhoff, C., and Battas, G. (1996) Building the Operational Data Store. New York: John Wiley and Sons, Inc.

Inmon, W.H. (2002), “Building the Data Warehouse “ , Third Edition , WILEY

ISACA (2009), Cloud Computing: Business Benefits, With Security, Governance and Assurance Perspectives, Emerging Technology White Paper, 1-10

Korol, T., Korodi, A. (2010), Predicting Bankruptcy with the Use of Macroeconomic Variables, Journal of Economic Computation and Economic Cybernetics Studies and Research, volume 44, number 1, 201-220.

Kortnik, M.A.R., Moody, D.L. (1999) From Entities to Stars, Snowflakes, Clusters, Constellations and Galaxies: A Methodology for Data Warehouse Design. 18 th. International Conference on Conceptual Modelling. Industrial Track Proceedings.

Krishnamurthi R, Goyal M. (2019) Enabling Technologies for IoT: Issues, Challenges, and Opportunities. Handbook of Research on Cloud Computing and Big Data Applications in IoT: IGI Global. p. 243-270.

Li, H., Sedayao, J., Hahn-Steichen, J., Jimison, E., Spence, C., Chahal, S. (2009), Developing an Enterprise Cloud Computing Strategy, Intel Corporation, 1-16

Luhn, H.P. (1958). "A Business Intelligence System" (PDF). IBM Journal of Research and Development. 2 (4): 314–319.

McKendrick, J, (2007), BI, Delivered from the Cloud, Ebizq Net, The Insider's Guide to Business and IT Agility

Mell, P. and Grance, T. (2011), The NIST Definition of Cloud Computing Recommendations of the National Institute of Standards and Technology

Miller, M., (2009), Cloud Computing Pros and Cons for End Users

Misra, S.C., Mondal, A., (2010), Identification of a company's suitability for the adoption of cloud computing and modelling its corresponding return on investment, Mathematical and Computer Modelling, doi:10.1016/j.mcm.2010.03.037

Mohammad Rifaiea, Keivan Kianmehrb, Reda Alhajjb, (2008), Data Warehouse Architecture and Design, IEEE International Conference on Information Reuse and Integration

Muntean, M., (2015), Considerations Regarding Business Intelligence in Cloud Context, Informatica Economica, Vol 19, 55-97

Queiroz-Sousa, P.O. and Salgado, A.C. (2019). A Review on OLAP Technologies Applied to Information Networks. ACM Trans. Knowl. Discov. Data 14, 1, Article 8 (February 2020), doi.org/10.1145/3370912

Rimal et al., (2009), A taxonomy and survey of cloud computing systems NCM 2009 - 5th International Joint Conference on INC,44-51

Raja JB, Rabinson KV, (2016), Iaas for private and public cloud using Openstack. International Journal of Engineering.;5

Ranjan, J. (2009) "Business Intelligence: Concepts, components, techniques and benefits " Journal of Theoretical and Applied Information Technology VoI. 9(1), 60-70

Rao, A.R., Clarke, D. (2019), Perspectives on emerging directions in using IoT devices in blockchain applications, Internet Things, doi:10.1016/j.iot.2019. 100079.

Sharda, Delen & Turban, (2017). Business Intelligence, Analytics, and Data Science: A Managerial Approach, Global Edition

Sheelvant, R. (2009). 10 Things to Know about Cloud Computing Strategy. IT Strategy.

Shim J, Warkentin M, […] Carlsson C (2002) "Past, present, and future of decision support technology", Decision Support Systems, 33(2), 111-126

Singh, S., Chana, I., (2016), QoS-aware autonomic resource management in cloud computing: a systematic review, ACM Comput. Surv. (CSUR) 48 (3) 42.

Strickland, J., (2008), How cloud computing work

Smith (2009), Building An Operational Data Store for a direct marketing application system

Thompson, O., (2004), Business Intelligence Success, Lessons Learned.

Turban E., Sharda R., Dursun D., Aronson JE.,(2011), Decision Support and Business Intelligence Systems9th Ed , Prentice Hall, ISBN 978-0-13-610729-3

Wayne, W. E. (2009), Implementing BI in the Cloud, The Data Warehousing Institute.

William J. F., Gregory, P. & Matheus, C J, (1992) "Knowledge Discovery in Databases: An overview", AI Magazine, Volume 13, Issue 3

William, S. and N. Williams, N., (2003), The Business Value of Business Intelligence. Business Intelligence Journal

Yu, J., Buyya, R, (2005), Taxonomy of workflow management systems for grid computing, J. Grid Comput. 3 (3–4) 171–200