POLITECNICO DI TORINO

The “Delta” Case: New AWS Data Platform Implementation

Commission of Computer Engineering, Cinema and Mechatronics

Master's Degree Course in Data Science and Engineering

Supervisor:
Daniele Apiletti

Candidate:
Sara Ferrua

Internship tutor:
Alessandro Beninati

ACADEMIC YEAR 2022-2023

A mamma e papà, il mio primo e più grande sostegno.

Ringraziamenti

Desidero esprimere la mia gratitudine a tutte le persone che mi hanno sostenuto durante questi ultimi anni di università.

Ci tengo a ringraziare il prof. Daniele Apiletti e il mio collega Alessandro Beninati per avermi permesso di affrontare questo progetto in azienda e avermi dato la fiducia necessaria per trasformare un progetto lavorativo in una tesi magistrale.

Ringrazio Giuseppe e Sabrina, due colleghi straordinari, per la loro eccezionale collaborazione e supporto lungo l'intero percorso di questo progetto. La loro competenza, la loro dedizione e il loro impegno sono stati fondamentali per il successo di questa sfida che abbiamo affrontato insieme. Si sono dimostrati dei veri punti di riferimento per me. La loro capacità di guidarmi attraverso le complessità del progetto, di fornire preziose risorse e di condividere la loro esperienza sono stati fattori determinanti per il raggiungimento dei risultati. La loro pazienza nel rispondere alle mie domande, nel chiarire i dubbi e nel fornire spiegazioni dettagliate è stata inestimabile.

Desidero dedicare un ringraziamento speciale ai miei genitori, poiché il loro sostegno e la loro dedizione non si sono limitati a darmi l'opportunità di studiare, ma hanno rappresentato un autentico punto di riferimento nella mia vita. Fin dall'inizio, hanno creduto in tutti i miei sogni e hanno lavorato instancabilmente per aiutarmi a realizzarli. La loro presenza costante e il loro amore incondizionato mi hanno dato la sicurezza e la serenità necessarie per affrontare le difficoltà con coraggio. Sono stati sempre pronti ad ascoltare le mie preoccupazioni, a darmi consigli saggi e a offrire un supporto pratico quando ne ho avuto bisogno. La loro saggezza e la loro esperienza di vita sono state un faro nella mia strada, illuminando il cammino e aiutandomi a prendere decisioni importanti.

Ringrazio mia sorella Greta dal profondo del cuore per essere stata una presenza preziosa e irrinunciabile lungo tutto il mio percorso di vita, e non posso che esprimerti la mia gratitudine infinita per il tuo amore sincero, il tuo sostegno incrollabile e la tua ispirazione costante che mi hanno guidato e motivato in ogni passo del mio cammino.

Ringrazio i miei nonni per essere una bellissima certezza. La vostra esperienza di vita e il vostro affetto hanno costantemente illuminato il mio percorso, donandomi la forza e la determinazione necessarie per affrontare le sfide accademiche. Non posso fare a meno di riconoscere il vostro ruolo fondamentale nel plasmare la persona che sono diventata oggi. Il vostro costante interesse per le mie conquiste accademiche ha rappresentato una fonte di motivazione, spingendomi sempre a dare il massimo.

Un pensiero speciale e colmo di gratitudine lo dedico ad Alfonso, la mia metà, il compagno di vita che ha camminato al mio fianco ininterrottamente per questi 5 anni. Grazie per il tuo amore

quotidiano, che si è manifestato in ogni gesto premuroso, in ogni parola di conforto e in ogni sorriso condiviso. La tua dedizione e la tua presenza costante mi hanno dato la sicurezza di poter affrontare qualsiasi sfida, sapendo che non sono mai sola. Non posso che ringraziarti di cuore per esserci stato ogni singolo giorno, ma soprattutto grazie per questa meravigliosa vita che stiamo costruendo insieme.

Desidero manifestare la mia sincera gratitudine nei confronti dei miei zii e delle mie cugine, Rossana e Laura con le quali condivido ogni aspetto della vita quotidiana. La loro presenza non è solo un legame familiare, ma un dono prezioso che arricchisce il mio percorso. L'affetto, la comprensione reciproca e la solidarietà che ci caratterizzano vanno oltre il vincolo di parentela, trasformando la nostra relazione in un legame profondo e duraturo.

Un ringraziamento va a Ursula, Dario, Michela, Angelo, Ernestina per essere una seconda casa per me, per avermi accolta a braccia aperte. La vostra accoglienza calorosa, unita alla costante considerazione nei miei confronti come membro della vostra famiglia, è qualcosa di inestimabile. Vi sono grata per il vostro affetto sincero e il calore che avete sempre dimostrato nei miei confronti.

Ringrazio Alessandra e tutti gli amici e i compagni di corso per aver condiviso con me il percorso accademico di questi anni universitari. La vostra preziosa presenza ha aggiunto un significato speciale a ogni momento di questa esperienza. Vi ringrazio di cuore per aver condiviso passo dopo passo ogni giorno di università. Insieme abbiamo affrontato nuove sfide, superato ostacoli e raggiunto traguardi importanti. Il vostro contributo alla mia esperienza universitaria va oltre l'ambito accademico; avete reso questo periodo un capitolo ricco di relazioni significative.

General Index

List of Figures

Abstract

Nowadays consolidated data insights is an essential competitive advantage for companies.

As Delta[1] embarked on an ambitious growth strategy to establish itself as Italy's premier pharmacy network, consolidated data insights became essential to effectively scale operations across 300 locations without compromising quality of care or service. However, in the current common scenario data collection remained siloed within separate Microsoft Business Central instances, hindering strategic planning efforts.

To overcome these constraints, Delta has decided to enlist PwC consultants to modernize its reporting infrastructure. The project lasted 6 months and it has been divided into different phases.

The first part of the project consisted of an initial assessment. PwC consultants commissioned a thought highlighting the critical need to centralize information to facilitate data-driven management. Consultants contracted a thorough audit of existing systems, tools, and datasets to understand collection points and information needs.

In-deep interviews with the company's leadership illuminated essential metrics and performance indicators required for executive-level reporting and decision-making.

PwC then mapped data flows and linkages between internal departments like procurement, finance, inventory, and customer service to establish an optimal organizational schema.

During the first phase of the project, consultants merged isolated Business Central platforms into an integrated AWS Redshift data lake. This unified environment organized raw operational data into a common business framework.

The data was meticulously modelled, creating a well-defined structure that served as a connection point. This enabled the extraction of valuable insights and information tailored specifically to meet the customer's needs. Through careful analysis and organization, the data became a powerful resource, empowering the customer with actionable knowledge. By harnessing the potential of data modelling, a comprehensive framework was established, facilitating the efficient exploration of information, and yielding valuable outcomes for the customer.

Data quality checks validated accuracy and completeness. This empowered analysts with comprehensive, consistently updated information for advanced analytics. This meticulous approach ensured that the extracted information was not only relevant but also maximized the value derived from the data, resulting in an enhanced experience for the customer.

[1] "Delta" is a fancy name used in compliance with the respect of the privacy of the company.

Subsequent work centered on building an intuitive decision support layer. Power BI visualized insights through tailored reports, interactive dashboards, and data exploration tools. Pharmacists, store managers, and executives now access personalized views directly relevant to their roles from a single portal.

As a result, Delta leadership gains enterprise-wide transparency into all facets of pharmacy operations spanning financial performance to inventory levels. Consolidated indicators replace assumptions, facilitating network-wide optimization to improve patient care and experience uniformly across their pharmacies. Empowered by data, Delta can sustain its rapid growth trajectory while retaining a community-centric focus.

The companies

This chapter outlines PwC, a consulting firm, and their client Delta. Delta name is a fancy name used in compliance with the respect of the privacy of the company. Delta, a holding of pharmacies engaged PwC for assistance in upgrading its reporting systems to extract valuable insights from documents and data to aid strategic decision-making. Currently, Delta gathers information but lacks robust analytics. PwC aims to develop an improved system to process Delta's data. This will allow Delta to better understand performance, decrypt trends and define organizational objectives using data-driven intelligence from its upgraded reporting infrastructure.

PricewaterhouseCoopers

PricewaterhouseCoopers, or PwC, is a reputable professional services company that provides clients in a variety of industries with a wide range of consulting, advisory, and auditing services. PwC's Data Analytics business is a fundamental component of its service offerings, as it plays a pivotal role in harnessing data's ability to yield useful insights, facilitate well-informed decision-making, and open strategic prospects for clients.

The PwC Data Analytics business blends cutting-edge technology with a wealth of market and analytical understanding. The division gives clients a competitive edge in their particular sectors by utilizing the potential of state-of-the-art tools and procedures.

The team's examination of enormous volumes of data allows them to find hidden patterns, spot new trends, and develop correlations that enable clients to take part in data-driven projects and make strategic decisions.

PwC's data analytics team is made up of highly skilled professionals with a variety of backgrounds in business analytics, statistics, mathematics, and data science. These experts collaborate directly with clients to comprehend their distinct objectives, difficulties, and data needs. Gaining a thorough grasp of the client's business environment enables the team to create solutions that are specifically tailored to meet needs and produce measurable business results. The PwC team uses cutting-edge analytics techniques including artificial intelligence, machine learning, predictive modeling, and data visualization to produce insights that are useful for driving organizational success.

Whether the objective is to enhance customer experience, optimize operational processes, or foster innovation, the PwC Data Analytics division is well-equipped to offer comprehensive and

creative solutions. By combining technical expertise with industry-specific knowledge, the division enables clients to unlock the full potential of their data assets. The team's ability to navigate the complexities of different sectors allows them to deliver tailored solutions that address specific industry challenges and capitalize on emerging opportunities.

However, PwC's dedication to data analytics extends beyond technical proficiency. The firm recognizes the critical importance of data security, privacy, and ethical considerations in today's data-driven landscape. The Data Analytics division at PwC upholds strong data governance procedures to protect client data and ensure utmost confidentiality. They adhere to relevant laws and regulations, ensuring compliance and ethical data handling throughout the analytics process. By prioritizing data governance, PwC establishes trust and confidence with clients.

The commitment to data security and privacy is an integral part of PwC's overall approach to data analytics. The firm understands that responsible and ethical data practices are not only essential for maintaining client trust but also for mitigating risks associated with data breaches and regulatory non-compliance. By adhering to rigorous data governance procedures and staying abreast of evolving data protection regulations, PwC's Data Analytics division ensures that client data is handled securely and in compliance with applicable laws. [2]

It is a vital component of the firm's service offerings, enabling clients to harness the power of data for strategic decision-making and competitive advantage. With their blend of technological expertise, market knowledge, and analytical proficiency, the team at PwC delivers customized solutions that provide valuable insights and drive tangible business outcomes. Moreover, the firm's commitment to data security, privacy, and ethical considerations ensures that clients can trust PwC with their sensitive information, knowing that their data is handled responsibly and within the bounds of legal and ethical frameworks. As businesses increasingly recognize the value of data, PwC remains at the forefront of the data analytics landscape, helping clients navigate the complexities of the digital age and achieve sustainable growth.

Figure 1 - Photo of the headquarters of Milan PwC

[2] PUBLISHER, PWC Consulting, Forbes Italia.

Delta company

Delta is a leading pharmacy network operating over 300 locations across central and northern Italy. The company seeks to become the largest socially responsible pharmacy brand in the country.

In 2021, Delta underwent a strategic refinancing process led by professional services firm PwC to strengthen internal reporting functions. The goal was to establish efficient and transparent reporting systems to support the network's rapid growth trajectory.

Currently headquartered in Milan, Delta has expanded significantly since its inception through both new store openings and acquisitions. It employs over 2,000 individuals and generated over €500 million in annual revenues in 2022. The pharmacy network focuses on providing first-rate healthcare services to local communities across key regions like Lombardia, Veneto, and Emilia-Romagna.

To finance further expansion ambitions, Delta negotiated a multi-million-euro financing package in 2022. This landmark transaction represented one of Italy's largest-ever private debt agreements. Fifteen international lenders participated, led by coordinators Intesa Sanpaolo and BNP Paribas. The deal refinanced existing credit lines and injected over €420 million in new capital.

Delta leverages stable regulations and high patient demand for pharmacy services. The brand aims to strengthen its nationwide presence.

Going forward, the new reporting systems implemented with PwC's guidance will provide leadership with clear operational visibility. This allows for optimizing growth strategies while delivering excellent care that emphasizes pharmacies' important role as local healthcare access points. Delta is well-positioned to solidify its position as Italy's premier pharmacy network through continued expansion and superior patient services.

Figure 2 - Photo of a Delta pharmacy

Business Case Introduction

The collaborative project between Delta, a pharmaceutical and parapharmacy holding company, and PricewaterhouseCoopers, a renowned consulting firm specializing in data and analytics services, aimed to revolutionize the customer's reporting practices. Delta had been using a data collection management tool that relied on vast tables, posing significant challenges in conducting comprehensive analyses and extracting meaningful insights.

The primary objective was to consolidate data from various management systems into a unified data platform with a standardized data model. By achieving this, the company's sales reports have become more accessible and actionable, allowing better performance monitoring and facilitating future integrations.

The thesis project focused on conducting a thorough examination of the entire design and implementation process of the data platform, with specific emphasis on three key components: Data Warehouse, Data Integration, and Data Modeling processes. Each of these aspects played a vital role in the overarching goal of enhancing the company's reporting capabilities and leveraging data effectively.

To store and manage the company's data, the first component, the Data Warehouse, involves building a central repository. Data had to be gathered from diverse sources, formatted consistently, and loaded into the data warehouse. The project's goal was to make sure that data would be stored and retrieved for analysis in an effective and scalable manner.

Designing a logical and physical structure for arranging the combined data within the data platform was the task of the final component, data modeling. As part of this process, relationships, hierarchies, and attributes were defined to make efficient querying and reporting possible. The data model acted as a guide for the platform's representation and access of data, assuring correctness and consistency in reporting.

Throughout the thesis project, a meticulous analysis of each stage of the design and implementation process of the data platform was conducted. Various methodologies, best practices, and industry standards were explored to ensure the successful integration of data, optimal performance, and scalability of the platform.

By addressing the challenges posed by the existing data collection and reporting practices, the collaboration between Delta and PricewaterhouseCoopers aimed to streamline the company's reporting processes, enhance decision-making capabilities, and pave the way for future data-driven initiatives. The comprehensive examination of the Data Warehouse, Data Integration, and Data Modeling processes within the thesis project provided valuable insights and

recommendations for implementing an efficient and effective data platform that have empower Delta to leverage data as a strategic asset.

Microsoft Business Central

Figure 3 - Business Central icon

Delta collected all the data in three Business Central environments. It is a management cloud of Microsoft, which allows you to save the accounting data of the company.

In more detail, Microsoft Business Central is a cloud-based ERP and finance management solution designed specifically for small and mid-sized businesses. It offers a comprehensive back-office system to assist businesses in expediting crucial procedures including accounting, sales, purchasing, inventory, manufacturing, and more. It was formerly known as Microsoft Dynamics 365 Business Central.[3]

Financial management, sales and customer management, procurement, supply chain management, project management, reporting, and analytics are a few of the key features offered by Business Central. Organizations are given capabilities for general ledger, accounts receivable/payable, cash flow forecasting, banking, and financial reporting through the financial management module. Additionally, it allows complex dashboards and multilayer budgets.

The sales module has tools that make it easier to do tasks like generating quotes, invoices, and orders. It enables effective management of shipping, commissions, and sales pipelines, among other things. The procurement module in Business Central manages purchase requests, orders, receipts, and vendor invoicing. With tools for inventory valuation, location transfer, quantity tracking, and replenishment, its inventory management technology streamlines supply chains.

Planning resources, and tracking time and expenses for projects are made easier by project management software for organizations like construction companies and consultancy firms. It allows billing, profitability tracking, and earned value analysis at the project level. Production management tools including routings, quality control, production orders, and shop floor management are advantageous to manufacturers.

3 PUBLISHER, Features of Dynamics 365 Business Central, Microsoft.

For mid-sized manufacturers, Business Central provides extensive functionality for configurations, demand planning, MRP, capacity planning, and material requirements planning. Its supply planning module caters to distribution businesses for inventory optimization, available-to-promise, order promising, and deployment planning. Service management assists firms in delivering field service with features like scheduling, dispatching, equipment tracking, and mobile access.

To provide staff like accountants, customer care representatives, warehouse managers, and executives with pertinent information, Business Central offers role-tailored views and KPI-driven dashboards. Reporting, interactive dashboards, and potent data visualization are all provided via its built-in analytical capabilities. The Office 365 platform and the ERP system are tightly integrated to allow for more collaboration utilizing applications like Word, Excel, Outlook, and SharePoint.[4]

Since it is delivered as a service, there are no up-front costs for licenses or hardware investments. Customers just need to subscribe every month to get the most recent features through their browsers. For international operations, Business Central enables multi-company and multi-country/currency configurations. It can meet specific data collection needs thanks to user-defined custom fields and tables.

Business Central offers SMBs an affordable, fully featured cloud ERP alternative to pricey on-premises systems. It is ideal for encouraging growth among smaller businesses by digitally altering business processes thanks to its simplicity of use, role-based experiences, and easy Office 365 integration.

How Business Central was used by Delta

Owing to the large number of pharmacies it managed, Delta had to deal with the difficulty of using three different Business Central settings. This choice was taken to prevent any performance problems that might have developed from having all of the pharmacies run in the same environment. Through the implementation of numerous settings, Delta was able to collect data with effective operations.

The incapacity of this architectural structure to offer the appropriate degree of data analysis was one of its drawbacks. In particular, SelfBI, a term used to characterize users' capacity to conduct their own data analysis, could not be supported. A thorough solution requiring the development of a data model was developed to get around this restriction. This data model facilitated the connection between the front-end tool, Power BI, and the management system through a data

4 PUBLISHER, Microsoft Dynamics 365 Business Central Documentation, Microsoft.

warehouse. By establishing this connection, Delta was able to achieve the desired level of customization and enable seamless data analysis for informed decision-making.

The implementation of separate environments within Business Central allowed Delta to effectively manage a large number of pharmacies under its purview. This approach prevented any potential performance bottlenecks and ensured the smooth functioning of the management system. By distributing the workload across multiple environments, Delta maintained efficiency and optimized resource utilization.

However, the standard capabilities of Business Central were insufficient to meet the specific data analysis requirements of Delta. The absence of SelfBI functionality prompted the development of a tailored solution. This involved the creation of a data model that served as a bridge between Power BI, a powerful data visualization tool, and the management system. Through this integration, Delta gained the ability to perform advanced data analysis.

Architectural Business Intelligence

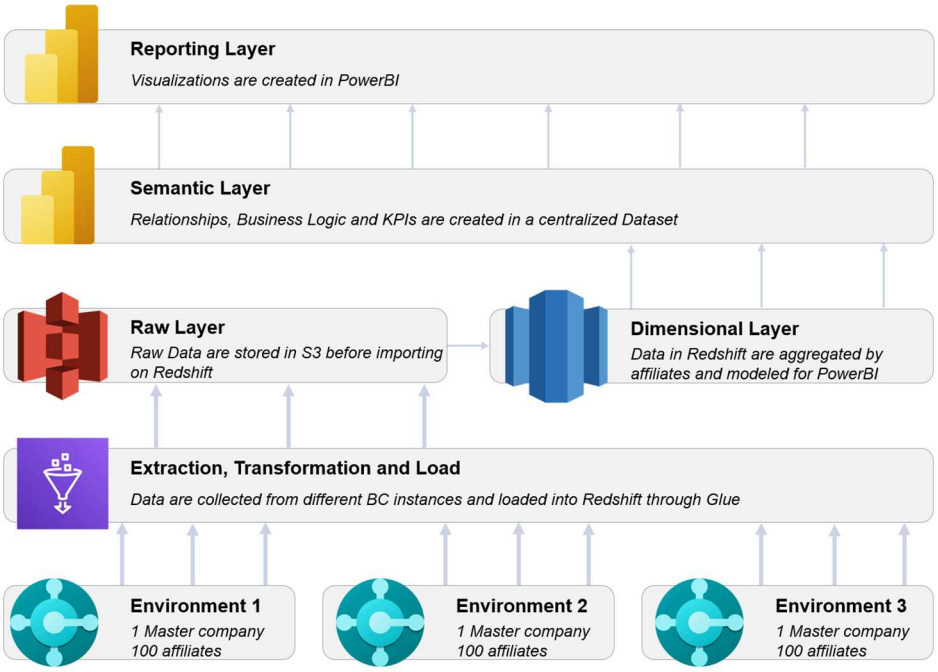


Figure 4 - Graphic designer of the proposed architecture

To enable the utilization and analysis of data from three central business environments using Power BI, a specific architecture has been created. Nowadays, many organizations recognize the significance of business intelligence (BI) in meeting specific business needs and enhancing organizational effectiveness. A robust architecture plays a pivotal role in effectively controlling the implementation process and ensuring the smooth functioning of the entire Business Intelligence environment. This architecture is composed of five distinct levels, namely source. These levels are designed with a focus on the value of data, data quality, and the seamless flow of information within the system.

During the data ingestion phase, AWS Glue, an AWS ETL service, is employed. This service utilizes API libraries to retrieve data from the respective business central environments.

In the raw layer, the data obtained in JSON format is saved in S3 in Parquet format, to optimize storage and retrieval efficiency. This format provides a columnar storage approach, enabling faster queries and minimizing storage costs.

Moving to the dimensional layer, the data undergoes modeling and analysis processes. In this layer, the analyzed data is stored in Redshift, a powerful data warehousing solution. Redshift provides scalability and performance optimization, allowing organizations to handle large volumes of data effectively. The management of Redshift follows a similar approach to the previous architecture, ensuring consistency and ease of maintenance.

The semantic layer is constructed using the AWS Redshift native connector. This layer serves as a Power BI data source. By leveraging the native connector, data replication across multiple reports is avoided, ensuring data consistency and reducing the risk of conflicting information. This semantic layer acts as an intermediary between the raw data and the reporting layer, providing a unified and standardized view of the data for reporting purposes.

Finally, in the reporting layer, data consumption takes place as reports are generated using Power BI. This layer is where the data is transformed into meaningful visualizations and insights. By establishing a Live Connection to the centralized dataset in Power BI, the reports are based on real-time data, providing up-to-date information for decision-making. This approach ensures that the reports are dynamic and reflect the most recent changes in the underlying data.

By implementing this architecture, clients can effectively harness the data from their central business environments and leverage the capabilities of Power BI for data analysis and reporting. The architecture provides a structured and controlled approach to data management, ensuring high data quality, consistent information flow, and accurate reporting throughout the entire Business Intelligence environment.

The choice of provider

The objective was to establish a connection between the data stored in three separate business centers and consolidate it into a unified and centralized database. To accomplish this task, a comprehensive evaluation was conducted to identify the most suitable service providers available in the market.

A server provider refers to a company or service that offers server hosting solutions to external parties. These providers furnish customers with the essential hardware infrastructure, network connectivity, and hosting environments required to effectively host and administer their websites, applications, or online services.

The process of selecting the appropriate server provider involved conducting a meticulous analysis of various factors. This examination encompassed scrutinizing the provider's reputation, reliability, security measures, scalability, pricing structure, and customer support capabilities. Additionally, the compatibility and integration capabilities of the provider's systems with the existing infrastructure were thoroughly evaluated.

By carefully evaluating these criteria, a comprehensive understanding of the available options was obtained. This enabled the identification of the most suitable server provider that could meet the specific requirements of linking the data from the three business centers into a centralized database.

The chosen server provider will facilitate the establishment of a robust and efficient infrastructure that seamlessly integrates the data from disparate business centers. This consolidation will enhance accessibility, streamline data management processes, and enable more effective analysis and reporting.

What is a server provider?

By providing server hosting services to various entities, a server provider, also known as a server hosting firm or service, plays a crucial part in the digital world. These businesses act as middlemen, making it easier for their customers to deploy, administer, and access applications, websites, and online services.

Essentially, a server provider oversees providing the physical infrastructure, network connectivity, and hosting environments needed to support and sustain their clients' online presence. Server providers help companies, organizations, and people create an online presence and manage their digital assets efficiently by utilizing their technology resources, know-how, and infrastructure.

To satisfy their customers' various needs, server providers provide a variety of services. Dedicated servers are one such service, where the supplier provides a physical server that is only allotted to a single client. As a result, the server can be completely customized and controlled by the client, perfectly meeting their unique needs.

Another type of service provided by server providers is virtual servers. In this scenario, the provider creates virtual machines within a shared physical server. Each virtual machine operates independently, providing customers with an isolated and configurable environment. This allows for greater flexibility, scalability, and efficient resource utilization.

Cloud hosting is another popular service offered by server providers. With cloud hosting, customers can leverage virtualized compute resources and storage within a shared cloud infrastructure. This enables businesses to scale their operations seamlessly, adapt to fluctuating demands, and enjoy high availability and reliability.[5]

Some server providers also offer colocation services, which allow customers to house their servers and equipment within the provider's data center facilities. By doing so, customers benefit from high-speed network connectivity, robust security measures, and the advantage of utilizing the provider's infrastructure without the need for extensive in-house resources.

Additionally, many server providers offer managed services to optimize the server hosting experience for their clients. Managed services involve the provider taking on the responsibility of monitoring, maintaining, and managing the servers on behalf of the customer. This includes tasks such as software updates, backups, security measures, and technical support. By offloading these tasks to the provider, customers can focus on their core business objectives while ensuring their server environment remains secure and optimized.

Three main cloud server providers performed a rigorous and in-depth review, specifically looking at the requirements of this project. These were Google Cloud Platform (GCP), Microsoft Azure, and Amazon Web Services (AWS). The team thoroughly considered the capabilities, pricing, features, and services of each alternative to choose the best one based on the project's technical specifications and financial constraints. The team will decide which provider offers the best cloud infrastructure solution for successfully deploying and hosting this project after carefully weighing the benefits, drawbacks, strengths, and weaknesses of AWS, Azure, and GCP through research, testing, cost analysis, and feature assessments.

5 PUBLISHER, What is a cloud service provider?

Amazon Web Services

Figure 5 - Amazon Web Services icon

One of Amazon's branches, Amazon Web Services (AWS), offers a vast cloud computing infrastructure. It provides a broad range of services and tools that enable businesses and organizations to develop and implement various infrastructures and applications in a highly flexible and scalable manner.

The main purpose of AWS is to provide users with on-demand access to computer resources through the Internet, such as virtual servers, storage, databases, and other services. The platform, which is exceptionally reliable, secure, and economical, allows enterprises the flexibility to scale their resources up or down in response to demand.

Amazon Web Services provides virtual infrastructure to meet the computational demands of its clients. This includes the option to start virtual machines on demand using Amazon EC2, which gives you customizable control over processing power. Through AWS Lambda, users can also run code without having to manage or pre-allocate any computers. Instead, they pay by the millisecond as their programs run in reaction to events. Both services offer elastic computing power that may be scaled to meet needs.

Amazon Web Services offers storage options to meet a range of data requirements. Glacier for archives, EBS for persistent disks, and Amazon S3 for files are available as options. Together, they provide scalable on-demand client information protection and availability across databases, files, and long-term backup.[6]

Amazon Web Services provides database technologies to power applications. Amazon RDS, DynamoDB, and Aurora deliver relational and NoSQL data options with managed upkeep of replication, protection, and updates so users can dedicate efforts toward code instead of infrastructure supervision. Customers gain agile database capabilities without related chores.

Amazon Web Services possesses potent networking and digital content conveyance functions including Amazon VPC, AWS Direct Connect, and Amazon CloudFront. VPC facilitates shielded digital circuits and Direct Connect supplies committed interfaces between user platforms and

6 IGNACIO BERMUDEZ, STEFANO TRAVERSO, MARCO MELLIA, MAURIZIO MUNAFO', Exploring the Cloud from Passive Measurements: the Amazon AWS Case.

the cloud. As a content delivery system, CloudFront boosts web tool functionality and extensibility by locally stocking and providing Internet material from areas near end users, aiding speedy access.

Amazon Web Services supplies management utilities such as AWS IAM, CloudTrail, and Systems Manager. IAM allows customization of access controls for AWS resources. CloudTrail maintains API activity logs within accounts. Systems Manager automates operational tasks and large-scale administration through the programming of resources without human execution of individual actions.[7]

Amazon Web Services caters to groups of every magnitude and variety of sectors. Specifically advantageous for new ventures and smaller companies, it supplies functionally extensible foundations affordably without vast preliminary expenses. Larger corporations also apply AWS to energize their digital evolutions, permitting accelerated novelty, compressed periods to commercialization, and streamlined functional mechanisms.

AWS likewise finds applications in numerous situations such as web/mobile app staging, information archiving and investigation, AI/ML, Internet of Things, and serverless processing. Its comprehensive solutions address diverse requirements, allowing companies to productively harness cloud provisions toward accomplishing their objectives.

AWS constitutes an all-encompassing cloud computing environment furnishing an expansive selection of utilities and instruments for devising, releasing, and administering programs and substructures. Its extensibility and plasticity combined with dependability and safety render it a well-liked selection for companies wanting to make the most of the cloud's might to stimulate novelty and expansion.

7 PUBLISHER, Amazon AWS: what it is and why use it in your business.

Azure

Figure 6 - Azure icon

Microsoft Azure consists of cloud infrastructure, programming resources, and prefabricated functions hosted and maintained by Microsoft in their information facilities. It supplies the foundation, instruments, and options essential for businesses to both move present initiatives to the flexibility of the cloud in addition to creating contemporary, cloud-primarily based functions. Azure gives infrastructure as a service element like virtual servers and storage solutions, in addition to a platform as a carrier gear like instrumenting for application construction, trying out, and control. Additionally, it gives prepared software program answers through the software program as a provider type that businesses can easily access and integrate into their operations. Organizations get the pliability to choose the necessary cloud services from Azure to meet their precise needs, regardless of whether modernizing current systems or developing new digital lines of business.[8]

Virtual Machines, which enable Windows and Linux virtual machines that can be swiftly scaled up or down to meet computational demands, is one of the main services that Azure delivers. With virtual machines, developers have a lot of freedom in selecting the language or framework of their choice without thinking about acquiring and maintaining the supporting infrastructure.

App Services, a platform as a service for quickly developing and deploying cloud apps without having to handle infrastructure management, is another important service. Web apps, mobile apps, APIs, and Azure functions are all supported by App Services. For continuous deployment and delivery, it also interfaces easily with well-known DevOps tools like GitHub, Docker, and Visual Studio.

Azure offers several database services, such as its Azure SQL Database, a fully managed relational database service with built-in intelligence, intelligent scaling, and strong security, for storage. Because Azure SQL Database and SQL Server databases are interoperable, conversion is simple. A globally distributed database service called Azure CosmosDB is another option. It supports NoSQL databases such as key-value, document, and graph databases. With guaranteed performance and availability, CosmosDB offers tremendous scalability.

8 DAVID CHAPPELL, May 2009, Introducing the Azure services platform.

To support the Internet of Things, Azure offers Azure IoT Hub, which is a service for ingesting high volumes of telemetry data from connected devices and managing bidirectional communication with devices. IoT Hub enables building complete IoT solutions with machine learning and analytics capabilities.

Data scientists can create, train, deploy, and manage machine learning models using an intuitive drag-and-drop visual interface or Python-based SDKs using Azure Machine Learning, a service that specializes in artificial intelligence. To quickly prepare deployment models, Azure Machine Learning automates the entire machine learning process.

Additionally, Azure offers Kubernetes Service, a fully managed container orchestration service for large-scale deployment and management of containerized applications. It makes Kubernetes cluster deployment and management simpler, enabling the quick development and operation of containerized programs.

There are other additional services available, like as real-time data stream processing, DevOps tool integration, and Azure Active Directory-based access management. The Azure platform's goal is to offer a broad range of cloud services that businesses can use to quickly design, deploy, manage, and scale applications with high availability and business-class security.[9] Azure offers a wide range of use cases, from web apps to machine learning and IoT solutions, thanks to its comprehensive collection of integrated services.

9 MUHAMMAD AYOUB KAMAL, HAFIZ WAHAB RAZA, MUHAMMAD MANSOOR ALAM, MAZLIHAM MOHD SU'UD, January 2020, Highlight the Features of AWS, GCP and Microsoft Azure that Have an Impact when Choosing a Cloud Service Provider.

Google Cloud Platform

Figure 7 - Google Cloud Platform icon

The Google Cloud Platform (GCP) is a broad range of cloud computing services that Google provides. It offers a broad range of infrastructure and tools for creating, deploying, and managing cloud applications and services. GCP gives businesses the freedom to grow their resources, gain access to cutting-edge technology, and make use of Google's wide network for improved performance and dependability.

GCP's primary operating platform is a huge global network of data centers. High-speed fiber-optic cables connect these data centers, establishing a worldwide infrastructure that supports GCP services. The low-latency access to resources and easy data movement between areas are both made possible by this dispersed network.

GCP provides a wide range of services and goods to meet different computing demands. These services can be roughly divided into tools for administration, big data, machine learning, networking, storage, and computing.

It offers a variety of computing solutions to accommodate various workload requirements. Virtual machines (VMs) on Google Compute Engine are configurable with various CPU and memory configurations. It enables businesses to operate their applications in a versatile and scalable setting. On the other side, Google Kubernetes Engine is a managed service that makes it easier to develop and manage containerized applications using Kubernetes.

GCP offers reliable and scalable storage solutions. Google Cloud Storage provides object storage for storing and retrieving data, suitable for a wide range of use cases such as backup, archiving, and content distribution. Google Cloud SQL and Google Cloud Spanner provide managed relational databases for storing structured data.

Building safe and efficient apps is made possible by GCP's networking features. Users of Google Cloud Virtual Network can create custom IP addresses and subnets for their own private networks. It makes it possible for resources inside the virtual network to communicate securely. Additionally, GCP provides load balancing services to improve application performance by

distributing traffic across different instances, such as Google Cloud Load Balancing and Google Cloud CDN.

GCP provides several tools for handling and assessing enormous volumes of data. Google BigQuery is a serverless data warehouse that provides fast and interactive queries on large datasets. Google Cloud Dataflow allows enterprises to manage and examine both batch and streaming data by utilizing Apache Beam. Google Cloud Pub/Sub is used to provide event-driven systems and real-time messaging.

With the range of machine-learning tools it provides, businesses can create intelligent apps. Pre-trained machine learning models, including translation, language, and vision APIs, are available through Google Cloud Brain. Users can create unique machine learning models using Google Cloud AutoML even if they don't have a lot of experience with data science. Machine learning models can be created and implemented on a scalable platform with the help of the Google Cloud AI Platform.

It increases the range of management tools available to streamline cloud operations. Google Cloud Identity and Access Management (IAM) allows organizations to manage user access and permissions. Google Cloud Monitoring provides information about the state and functionality of apps and infrastructure. Furthermore, the infrastructure-as-code and command-line interfaces of Google Cloud Shell and Deployment Manager make resource deployment and management easier.

Users can utilize the Google Cloud Platform Console, a web-based graphical interface that offers a unified view of their resources and facilitates simple management and setup, to interact with GCP services. For programmatic control and automation, GCP also provides API access and a command-line interface (CLI).

GCP's interoperability with other Google services and technologies is one of its main benefits. For instance, GCP easily connects with Google Workspace (previously G Suite), enabling businesses to use Google Docs, Drive, and Gmail as productivity tools in their cloud environment. Additionally, Google Cloud Marketplace, a platform that provides a vast selection of pre-configured software packages and solutions for different use cases, is integrated with GCP.

GCP has a multi-layered approach to security. To protect the confidentiality and integrity of data, it offers encryption both at rest and while in transit. To manage user rights and access policies, GCP additionally provides Identity and Access Management (IAM) controls.[10]

Google Cloud Platform provides a full range of cloud computing services that let businesses create, launch, and expand their apps and services flexibly and safely. With its broad range of

10 RICKY K. P. MOK, HONGYU ZOU, RUI YANG, TOM KOCH, ETHAN KATZ-BASSETT, KC CLAFFY, 2nd November 2021, Measuring the network performance of Google Cloud Platform.

services, world-class infrastructure, and integration with other Google technologies, GCP offers enterprises a strong and trustworthy platform for innovation and cloud success.

The main differences between Azure, AWS, and Google Cloud Platform

Azure, Google Cloud Platform (GCP), and Amazon Web Services (AWS) are three major players in the cloud computing industry, each offering a wide range of services and capabilities to organizations using the power of the cloud. While they share similarities in terms of their core offerings, there are distinct differences that set them apart.

It is important to understand the differences between the cloud computing services mentioned above, to better understand the architecture adopted.

AWS, as the first major player in the cloud computing market, has established a significant market share. It boasts a comprehensive global infrastructure and a broad range of services. AWS is known for its maturity, extensive documentation, and the wide variety of compute, storage, database, networking, analytics, AI/ML, and developer tools it offers. It provides a rich set of services that cater to various use cases and industries.

On the other hand, Microsoft Azure leverages strong integration with Microsoft products and services. It has gained traction in the enterprise market and is recognized for its compatibility with existing Microsoft technologies. Azure offers a comprehensive suite of services, including computing, storage, networking, databases, analytics, AI/ML, and developer tools. Its integration with Microsoft products makes it particularly appealing to organizations already using Microsoft solutions.

Google Cloud Platform (GCP) focuses on its expertise in data analytics, machine learning, and artificial intelligence. It harnesses Google's infrastructure and proficiency in managing large-scale data processing. GCP provides a robust set of services for compute, storage, networking, databases, analytics, and developer tools. It is known for its emphasis on data-driven applications and offers machine learning services such as Google Cloud AI Platform, Google Cloud AutoML, and TensorFlow. GCP's BigQuery is a powerful tool for analyzing large datasets.

In terms of pricing and cost management, AWS, Azure, and GCP offer flexible options. AWS provides a wide range of pricing models, including on-demand, reserved instances, and spot instances. Azure offers pay-as-you-go options, reserved instances, and hybrid benefits for organizations using Microsoft software licenses. GCP utilizes similar models to AWS and Azure, offering on-demand and committed use discounts. All three platforms provide tools and services to help organizations monitor and optimize costs.

Another differentiating factor is their global infrastructure. AWS boasts the most extensive global infrastructure, with data centers in regions worldwide. It allows organizations to deploy resources in multiple geographic locations, ensuring low-latency access and compliance with data residency requirements. Azure also has a global presence with data centers in many regions, providing robust coverage and enabling organizations to build applications with geo-redundancy and high availability. GCP's global infrastructure is continuously expanding, with data centers in multiple regions. It emphasizes Google's high-speed network and global backbone for fast data transfer and low latency.

Each cloud platform offers specific strengths in machine learning and AI capabilities. AWS provides a broad range of machine learning services, including Amazon SageMaker, Amazon Rekognition, and Amazon Comprehend. Azure offers comprehensive AI and machine learning services, such as Azure Machine Learning, Azure Cognitive Services, and Azure Databricks. It integrates well with other Microsoft tools like Power BI and Excel. GCP is renowned for its expertise in data analytics, machine learning, and AI. It offers services like Google Cloud AI Platform, Google Cloud AutoML, and TensorFlow. GCP's BigQuery is a powerful tool for analyzing large datasets.[11]

In conclusion, AWS, Azure, and GCP are dominant players in the cloud computing market, each with its unique strengths and approaches. AWS has a broad service portfolio and market presence, Azure integrates well with Microsoft technologies, and GCP excels in data analytics and machine learning. Organizations should carefully evaluate their specific needs, considering factors such as pricing, integration, specialized services, and desired capabilities, to choose the cloud platform that best aligns with their requirements and business objectives.

The selected provider

AWS was selected as the cloud provider for this project due to its unique advantages over Azure and GCP. AWS offers the broadest set of cloud services, with over 200 in its portfolio, enabling its application across a diverse range of scenarios. The services are also highly integrated and mature compared to alternatives.

AWS has built the largest cloud network globally with the most regions and availability zones. This geographic footprint is crucial for meeting latency requirements, optimizing traffic routing, and ensuring business continuity. Maintaining operations across multiple regions helps minimize downtime from localized disruptions.

11 TABISH MUFTI, POOJA MITTAL, BULBUL GUPTA, 27th February 2020, A Review on Amazon Web Service (AWS), Microsoft Azure & Google Cloud Platform (GCP) Services.

Security, compliance, and data protection are also areas where AWS leads the field based on track record. AWS places a strong emphasis on privacy, procedures to safeguard customer data and infrastructure, as well as compliance with regulations. This gives organizations greater assurance that sensitive workloads and information are handled appropriately in the cloud.

Its support and documentation are often cited as the best in class by users. The depth of technical guidance and responsive customer service helps organizations adopt, operate, and innovate on AWS more seamlessly.

Collectively, these differentiated capabilities position AWS as the best-matched provider to deliver on the requirements of this project at scale.

The connection between Business Central and the Provider

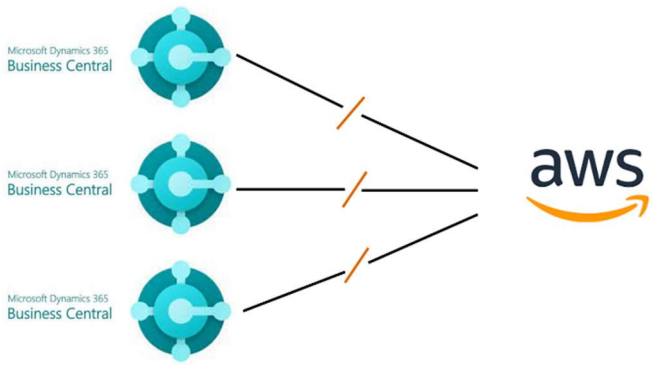


Figure 8 - Representation of the lack of connection between Business Central and AWS

After selecting AWS as our preferred cloud provider, our team set out to establish how we would migrate data from our on-premises ERP system, Business Central, into the centralized data warehouse hosted on AWS. We wanted to leverage native extract, transform, and load ETL (Extraction transformation, and loading) capabilities to automate regular data syncs.

Upon investigating available integration options, we did not find a standardized connector or out-of-the-box ETL tool that could pull data directly from Business Central and load it into our target databases in AWS. Business Central does not natively support direct integration with major cloud platforms through established ETL frameworks. This presented a challenge as we wanted seamless, ongoing data replication.

Extract Transform and Load (ETL)

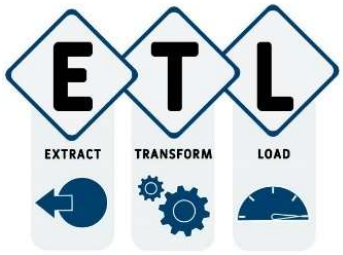


Figure 9 - ETL (Extract Transform and Load) icon

For this project, establishing an efficient ETL process is critical. Our goal is to consolidate operational and transactional data from Business Central into an AWS-hosted data warehouse. This will enable advanced analytics and business insights across the entire organization.

ETL, or extraction, transformation, and loading, is a crucial procedure for integrating, organizing, and analyzing this data. In the extraction stage, raw data is retrieved from a variety of systems, databases, files, APIs, and sources, both structured and unstructured. The objective is to access and gather the data for additional processing. Organizations can create a unified data repository by combining data from sources including sales transactions, customer interactions, and inventory information. The extracted data is transformed in the next steps to make it ready for analysis and reporting by cleaning, validating, and reformatting it. Finally, target databases or data warehouses are loaded with the processed data.

In target databases or data warehouses, the processed data is loaded so that it may be used for analytics, business intelligence, and decision-making. ETL is crucial in gathering dispersed data, processing it for business use, and loading it into reachable systems.

The phase of transformation begins with extraction. The extracted data is cleaned, validated, and restructured in this step. Data is standardized and normalized during transformations to improve consistency and quality. Making calculations, aggregating data, removing irrelevant or incorrect data, and applying business rules are possible. This stage is essential for assuring data accuracy and integrity and for getting the data ready for insightful analysis.[12]

The phase of transformation begins with extraction. The extracted data is cleaned, validated, and restructured in this step. Data is standardized and normalized during transformations to improve consistency and quality. Making calculations, aggregating data, removing irrelevant or

12 PUBLISHER, What is ETL (Extract Transform Load)?, AWS.

incorrect data, and applying business rules are possible. This stage is essential for assuring data accuracy and integrity and for getting the data ready for insightful analysis.

The data is loaded into a target system, such as a data warehouse or analytical database after it has been extracted and transformed. In the loading phase, the converted data are organized and structured in a way that is suitable for analysis and reporting. This stage makes sure the data is easily available for use in decision-making procedures and that it can be rapidly accessed to produce insights.

ETL is essential for helping firms get value out of their data. ETL eliminates data silos by integrating data from several sources, giving a complete picture of the company's information environment. This combined data can be examined to find patterns, correlations, and trends that might otherwise go undetected. By ensuring that decision-makers have access to accurate, current, and reliable information, ETL supports data-driven decision-making.

In addition, ETL facilitates data migration, allowing for the seamless movement of data between systems during system upgrades or transitions. It guarantees accurate data transformation and loading into the intended system, reducing disruptions and maintaining data integrity.

With the aid of specialized tools and technology, ETL procedures are frequently automated. The procedure is made more effective and scalable by these tools' functionalities for data extraction, transformation, and loading. Automation lessens the need for manual labor, decreases human error, and speeds up data processing, all of which are crucial for businesses dealing with enormous amounts of data.

It is an essential duty for businesses trying to integrate, manage, and analyze data properly. By gathering data from many sources, transforming it into a standardized and pertinent format, and then loading it into the target systems, ETL enables businesses to get crucial insights and make intelligent decisions. As the volume and complexity of data continue to expand, driving businesses toward more effective and efficient data management techniques, the value of ETL in streamlining data integration and analysis will only grow.

Omega: the solution to connect Business Central to the provider

As we began our project to migrate operational data from our on-premise ERP system (Business Central) to a centralized cloud data warehouse on AWS, one of the major hurdles we faced was a lack of out-of-the-box integration capabilities. While Business Central provides useful data, no standardized connectivity exists between it and major cloud platforms like AWS. This meant we would need to pursue a custom solution.

After carefully researching our options, we decided the best approach was to engage an experienced integration partner to develop a suite of APIs facilitating direct data extraction from Business Central. This would allow us to avoid complex workarounds like file-based transfers that were prone to errors and difficult to manage at scale over the long run.

We issued a request for proposal to several IT consulting and integration firms. Among the respondents was Omega, a company specializing in hybrid cloud integration projects involving the migration of on-premise systems to AWS and Azure. Omega's proposal demonstrated a thorough understanding of our requirements and constraints. Additionally, their portfolio showed a track record of successfully delivering similar API-driven ETL solutions on time and within budget.

We ultimately chose to contract Omega based on their specialized expertise and competitive pricing. To get started, their engineers met with our data and IT teams to map out the databases, tables, and fields required for extraction. These involved discussions of data structures, relationships, workflows, and more to ensure holistic requirements capture.

Over the following weeks, Omega's development team designed and coded a comprehensive set of RESTful APIs aligned to our specifications. They employed strategies like API versioning, modeling, and documentation to achieve maintainability, security, and scalability. Comprehensive testing was also performed before delivery.

The completed suite of Business Central APIs has now formed the foundation of our ongoing cloud data integration. Automated extraction and loading are seamlessly happening in the background via these APIs, empowering us with unified analytics and improved decision-making across the business.

Tools of AWS

This project utilizes several Amazon Web Services (AWS) tools to enable its functionality. Specifically, Amazon S3 provides object storage for raw data and processed outputs. Amazon Redshift is used to warehouse the structured data and run SQL queries for analysis. AWS Glue handles extract, transform, and load (ETL) operations to prepare and move data between S3 and Redshift. By leveraging these fully managed AWS services together, the backend pipeline is able to reliably and efficiently ingest, process, store, and analyse large datasets.

As a result, this overview will introduce the various tools utilized in this project, along with an explanation of each tool's specific capabilities and functions within the overall architecture.

AWS S3

Figure 10 - AWS S3 icon

Security is to fully comprehend how S3 facilitates our data migration, it is important to understand what S3 is and its key capabilities. Amazon Simple Storage Service (Amazon S3) is an object storage solution that delivers industry-mainstream extensibility, information availability, safety, and effectiveness. Clients of all sizes can safeguard and retain any measure of material for a scope of situations including cloud programs, cell programs, reinforcement and reestablish, libraries, business applications, IoT, and large information investigation.

Some key capabilities of Amazon S3 include virtually unlimited storage scale, high durability of stored objects, security features, integration with AWS services, and management tools. S3 allows storing unlimited objects ranging from bytes to petabytes in size across a global network of data centers.

It utilizes advanced encryption technologies to protect data in transit and at rest. Objects can be encrypted server-side with AWS-managed keys or customer-provided keys for complete control. Robust access management controls include bucket policies, identity management with IAM roles, and signed URLs. Customers can enforce regulatory compliance with audit capabilities to log access to S3 resources.

Amazon S3 delivers high performance for data-intensive workloads. It offers up to 2,000,000 PUT/COPY/POST/DELETE operations per second globally and a large amount of daily data transfer. With a latency of milliseconds for most S3 operations, it supports real-time big data analytics.

If desired, S3 ensures high availability by storing data redundantly across multiple facilities within AWS regions and global regions. This resilient infrastructure provides consistent and fast access to objects worldwide. If a disaster occurs or regions become unavailable, data will still be accessible.[13]

S3 integrates with other AWS services to build robust application architectures. Pre-built connectors allow seamlessly accessing S3 data from databases like Redshift.

Amazon S3's massive scale, proven reliability, security features, and integration abilities make it well-suited as a storage backbone for modern applications and analytics platforms. This helps businesses innovate faster while avoiding upfront costs and ongoing complexities of data infrastructure management.

Within Amazon S3, some buckets serve as containers for storing objects. Specifically, there are two buckets named *Delta-dynamics-integration-test* and *Delta-dynamics-integration-prod*, which are associated with the test and production environments, respectively. These buckets share the same structure.

Firstly, there is a folder called *Log* that is automatically generated by AWS Glue. This folder contains logs related to the operations within the buckets.

Secondly, there is a folder named *GlueScript* which is also automatically generated. This folder is used to store the scripts for AWS Glue jobs.

Next, there is a folder called *Data* which is created by the *dynamics-integration-load-company* job. In this folder, a file with the upload date is saved.

Furthermore, there is a folder named *BusinessCentralData* which is divided into two subfolders. The first one is named *input*, which it is a subfolder used to store Parquet files that are generated by Glue jobs.

The second one is *success*: once the files from the *input* folder are uploaded to Redshift, they are transferred to this subfolder. These files will be retained for approximately one day before being deleted during the next upload. This process ensures that the subfolder contains the most recent files.

13 MATTHIAS BRANTNER, DANIELA FLORESCU, DAVID GRAF, DONALD KOSSMANN, TIM KRASKA, Building a Database on S3.

AWS Redshift

Figure 11 - AWS Redshift icon

Amazon Redshift is a completely handled, hugely parallel information warehouse assistance furnished by Amazon Web Services. It permits companies of any magnitude to investigate gigantic sums of organized and semi-organized information exploitation regular SQL while not needing to be concerned regarding framework administration, inquiry functionality calibration, or high accessibility.

Redshift spontaneously makes any important data warehousing assets finish Amazon's tested and trustworthy engineering substructure. It deals with everything compulsory to build up, run, and proportion a data warehouse empowered by swift, vigorous knots.[14] Customers can start with a single knot and proportion effortlessly up to 128 knots with over a petabyte of information and a throughput of up to 250,000 megabytes per second. This permits Redshift to question and method exabytes of data for progressive analytics applications crosswise over businesses.

Security is deeply embedded in Redshift through features like encryption at rest using AWS KMS keys, IAM database authentication, private connectivity using Amazon VPC, fine-grained access controls based on identities, encryption in transit, and auditing. This enables organizations to load and analyze sensitive data with confidence that it remains protected and meets necessary compliance standards.

In addition to rock-solid security, Redshift provides tremendous value through tight integration with other AWS services and BI tools. It allows querying and combining data residing in storage as S3 and others as part of a single analytic workflow. Pre-built connectors also empower seamless data access for tools like Power BI, but also Tableau, QuickSight, and Looker to reveal compelling business insights.

The on-demand, pay-as-you-go pricing model eliminates upfront hardware costs and charges only for resources consumed. Ongoing maintenance, upgrades, patch management, and all

14 ANURAG GUPTA, DEEPAK AGARWAL, DEREK TAN, JAKUB KULESZA, RAHUL PATHAK, STEFANO STEFANI, VIDHYA SRINIVASAN, 31 May 2015, Amazon Redshift and the Case for Simpler Data Warehouses.

cluster operations are managed by AWS, meaning no spending overhead for ongoing data warehouse support. Overall TCO is significantly lower than traditional on-premises data warehousing infrastructure over time.
In summary, Amazon Redshift enables data-driven decisions at scale by eliminating the heavy lifting of procuring, configuring, and managing a large data lake. Its seamless integration, security, and performance empower businesses to focus resources on deriving actionable intelligence from all customers, products, and operational datasets.

Amazon Redshift is utilized in the project to store data obtained from Business Central. Subsequently, the data is organized into tables, views, and procedures for efficient data management. There are two schemas involved: *hh_dynamics_svil* for the test environment and *hh_dynamics* for the production environment. Both schemas share identical structures, ensuring consistency in data organization and management.

Configuration tables, table, view, and procedures

There are several configuration tables, which allow an orderly, simple and clean management of the data:

• *Cfg_query*: This table contains the list of all APIs affected by the solution. It consists of the following columns:

- *Query*: the name of the API to be included in the perimeter.
- *Deltaloadingfilter*: name of the data field that will be used for loading in delta: this always coincides with the SystemModifiedAt.
- *Fullloadingfilter*: name of the field used for full loading, where possible: usually it is the functional date of the table with the addition of the date from which you should consider loading (e.g. PostingDate ge 2020-01-01).
- *Reloadfield*: name of the date field used for full loading.

•*Cfg_dataframeschema*: In this table the API structure is saved: this is used to force all companies the same table structure and avoid errors during the download. The fields are:

- *Query*: the name of the API to be included in the perimeter.
- *Schemastructure*: list of API fields, this can be obtained using the AWS Glue dynamics-integration-print-schema job.

• *Cfg_logdeltaloading*: Here are the last loading dates of the tables. The columns are as follows:

- *query*: the name of the API to be included in the perimeter.
- *company_id*: company code.

- *maxdeltadate*: date the table was last loaded.
- *environment*: environment to which the company belongs.

•*cfg_reloadcompany*: the table, which must be populated manually if necessary, shows the list of queries and companies that need a full load: if yes, the date used as a filter for downloading the data is also shown. The columns are as follows:

- *query*: the name of the API to be included in the perimeter.
- *company_id*: company code.
- *environment*: environment to which the company belongs.
- *reloaddate*: date used in API filter to download data.
- *isreloaded*: flag indicating if the table has been reloaded in full.
- *loadeddate*: date the table was reloaded in full.

•*cfg_apistring*: This view lists the APIs that will be run to download the data. The rules for defining the API link to be used are defined automatically within the view.
fields are:

- *company*: company description.
- *id*: company code.
- *environment*: name of the environment to which the company belongs.
- *query*: the name of the API to be included in the perimeter.
- *apidelta*: API link used to download data in delta.
- *apifull*: API link used to download data in full.
- *apireload*: API link used to download data in full if the cfg_reaload table has been populated: this is valid for a specific company and query.
- *API*: this returns one of the three API fields above: this is the value that AWS Glue will use when downloading.

Particular importance is given by the *stg_companies* table: this is populated by a job of AWS Glue, and it shows the list of all the companies present on Business Central.

Data-populated tabs enable the implementation of procedures that facilitate the creation of new tables, which can be linked to Power BI for extracting valuable information. These tables serve as a foundation for generating insights and obtaining meaningful data.

Additionally, views offer a robust solution for enhancing query performance within Amazon Redshift. By storing the results of complex queries in memory, views provide an optimized and efficient approach to accessing data. When queries are executed, the system retrieves data

directly from the views rather than performing expensive calculations in real time. This results in significant performance gains, enabling faster and more efficient data retrieval and analysis.

By leveraging data-populated tabs and views, organizations can harness the power of these concepts to unlock valuable insights and improve the overall efficiency of their data processing and analysis workflows.

AWS Glue

Figure 12 - AWS Glue icon

AWS Glue, supplied by Amazon, is a completely handled extract, change, and load solution. It facilitates clients to get ready and load data for analytics effortlessly. With AWS Glue, information can be purified, reshaped, and kept for analytics or machine-gaining knowledge utilizing Spark.

Some key capabilities of AWS Glue include data catalog, ETL jobs, developer-oriented APIs, security controls, and integration with other AWS services. The data catalog in AWS Glue provides a centralized repository that stores all metadata related to data assets. It has a crawler that can automatically detect schema from data sources and populate the catalog. This makes the data easy to find, understand, query, and access.[15]

AWS Glue ETL responsibilities authorize adjusting and transferring information between reservoirs. Jobs can be produced and planned exploitation a visible editor, Python, or API calls. This permit performing ETL undertakings on oversized measures of flowing and group information reliably. AWS Glue moreover facilitates Spark ETL to change and investigate gigantic information at any scale. With built-in Spark capacities, the assistance helps to fabricate machine-gaining knowledge of styles exploitation information pipelines.

The API tools make it simpler to implement data transforms and load data to destinations like Amazon S3, AWS Redshift. AWS Glue integrates closely with other AWS services as well.

For security, AWS Glue enforces encryption, access control, and auditing by default.

15 PUBLISHER, Amazon Glue.

In summary, AWS Glue is a serverless data preparation service that simplifies data integration tasks like classification, extraction, transformation, and load for analytics workflows. It reduces the need to provision and manage ETL infrastructures so that the focus is more on analysis and insights.

Specifically for the project, AWS Glue is responsible for completing all the activities of integration of flows coming from Business Central through the use of customized jobs that, using libraries pyspark, have several functions: manage calls to APIs, download data on parquet files in S3 and the next upload in the Redshift structures.
The main Glue resources that are used for the project are: Crawlers, Databases, Tables, Jobs and Triggers.

Crawlers

The Glues Crawlers are a very useful resource for metadata of the source and destination structures of the data: in particular, to support the developed platform there are two crawlers, *dynamics_integration_crawler* and *dynamics_integration_test_crawler*, respectively for production environment and test.
Through these have been cataloged the structures that make up the db of Redshift to allow Glue Jobs to access information during the run. When a crawler catalogs a new structure it saves the metadata in another Glue object, the Tables.

Databases

Each crawler must historicize the information it collects on the various structures, for this reason, Glue integrates the Databases: the two crawlers previously mentioned are *dynamics_integration* and *dynamics_integration_test*.
Entering the specification of each database is the list of tables of Glue, which we will talk about in the next paragraph.

Tables

A Glue table is the definition of the metadata of the source or destination structure that hosts the data: these are used to retrieve information immediately within Glue jobs.
Below is an extract from the list of tables on Glue to better understand what type of information they contain.

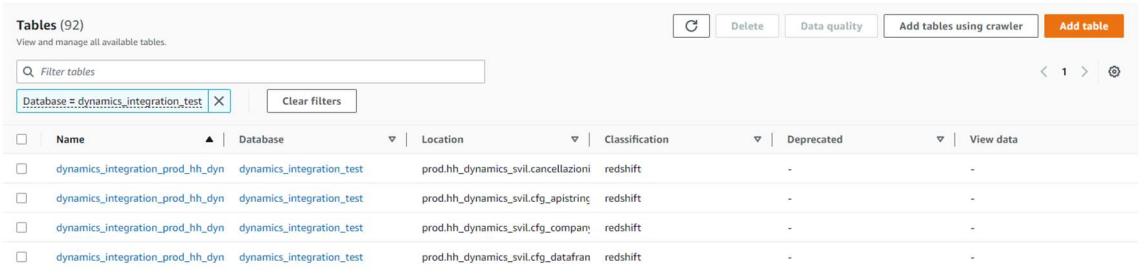


Figure 13 - AWS Glue tables

In detail the fields of the table shown in the previous image:

• *Name*: the name of the structure that the crawler has cataloged (in this case the names in the image refer to some Redshift tables)

• *Database*: The name of the Glue Database in which the information collected by the crawler is redirected.

•*Location*: the path where the structure is located (table, file, etc.). In this case, being Redshift tables, in the location we find the db, the schema and the name of the table.

• *Classification*: The resource hosting the source/destination you are analyzing.

• *Deprecated*: if no longer in use.

Jobs

Jobs serve as the integral component that binds various data sources together, facilitating ETL (Extract, Transform, Load) operations. These jobs can be executed either through a graphical interface or custom scripts. In our specific scenario, the primary function of these jobs is to retrieve data from an API, store it in a historical fashion on the S3 storage system, and subsequently upload it to Redshift. The orchestration of these tasks is managed by another element within Glue known as triggers. All constructed jobs rely on the usage of PySpark for data processing, supplemented by the inclusion of specific libraries tailored to the requirements of each individual job.

Architectural process – TO-BE Solution

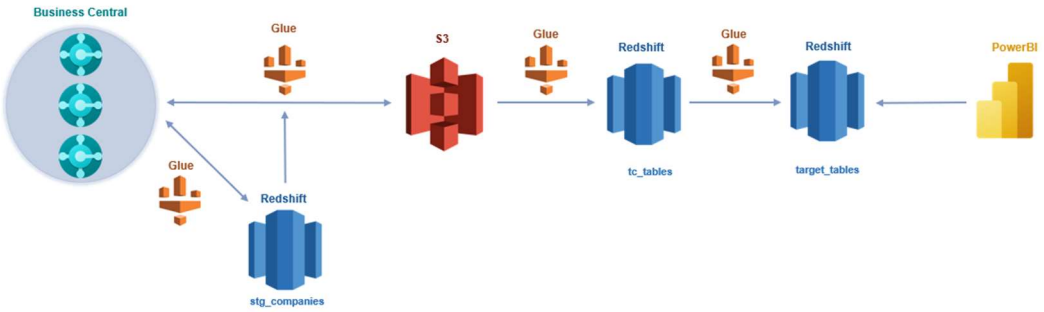


Figure 14 - Representation of architecture

To extract valuable insights for the customer, we have meticulously crafted an architecture that harnesses the data gathered from our three core business environments. This strategic development has successfully propelled us towards our ultimate objective.

The smooth and efficient functioning of the studied system relies on a series of crucial steps that ensure effective data processing and analysis. Let's explore each step in detail:

initially, the system utilizes AWS Glue, a comprehensive data service managed by Amazon Web Services. Through a well-defined Glue job, the system extracts the essential information related to pharmacies from the Business Central (BC) environment. This step ensures that the relevant data is identified and made available for further analysis.

Next, the extracted information undergoes seamless transfer to AWS S3, a robust storage service provided by AWS. This transfer process guarantees secure and efficient storage of the data, ensuring its accessibility for subsequent stages.

Subsequently, the information is continually transferred via Glue to a Redshift table, which serves as a data warehousing system managed by AWS. This transfer operation ensures the data's availability for further analysis and integration within the Redshift environment.

Within the Redshift tables, a series of operations takes place. These operations include querying, integration, and modification, all aimed at refining the data. By executing these operations, the system ensures that the final versions of the tables contain accurate, consolidated, and usable information.

To leverage the refined data, Power BI, a powerful business intelligence tool, establishes a connection with the finalized Redshift tables. This connection enables Power BI to access the data directly from the tables, empowering users to generate comprehensive and visually appealing reports.

By following this meticulously designed process, which involves employing Glue for data extraction, utilizing AWS S3 for secure storage, transferring data to Redshift for analysis and

refinement, and connecting Power BI to the finalized tables for reporting purposes, the system ensures the seamless utilization of the studied data, facilitating informed decision-making and actionable insights.

This overview of the architecture's development will guide the subsequent chapters, providing a step-by-step exploration of the entire process.

1)Glue: list of companies

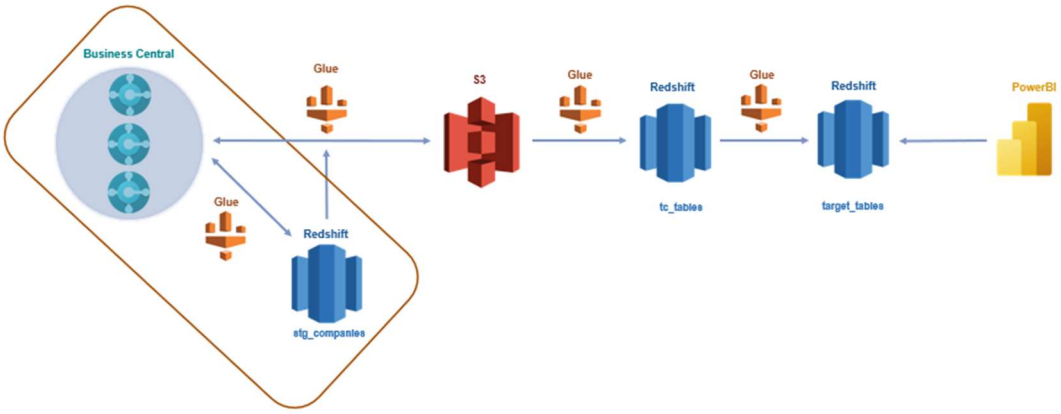


Figure 15 - Representation of architecture – Focus 1

Delta exemplifies an ever-evolving enterprise, consistently broadening its reach in the pharmaceutical sector. Its expansionary approach is characterized by the regular establishment of new centres, reflecting a commitment to meet growing demand and extend its services to a wider audience.

In order to maintain an accurate inventory, the first step of the architectural blueprint necessitates obtaining the latest compilation of pharmacies operating within the system. This crucial task is entrusted to the *dynamics-integration-load-company* Glue job, which dutifully retrieves the list of companies from a dedicated API given by Omega's company.

With a daily execution schedule, this process guarantees that the information remains timely and up-to-date, ensuring the smooth functioning of Delta's architecture future steps.

Glue job *dynamics-integration-load-company* performs several key tasks to process and store data. The PySpark code is organized into distinct sections to import necessary libraries, connect to AWS services, call functions, and write output.

The first part imports core libraries and functions that will be used. A connection is established to AWS S3 to allow future jobs to load and retrieve data from storage: *s3 = boto3.resource('s3')*. For every business central environment, it is sent an API request using the address provided to us by the Omega company. This request aims to gather the necessary information, the list of the

pharmacies, from the respective environment. However, before obtaining the desired data, a client authentication process becomes necessary.

To authenticate and gain access to the information, it is required to get a token generated by the *get_new_token()* function. This token serves as a secure identifier, allowing who want to get information to prove its authenticity and entitlement to access the requested data. By utilizing this token-based authentication mechanism, only authorized clients can retrieve the information from the business central environments. This function defines the rules and procedures for creating a valid token that can be trusted by the system. Once the client successfully authenticates using this token, they gain the necessary privileges to access the information from the business central environments, enabling them to proceed with their intended tasks.

```
environment_list = ['production', 'prod2', 'prod3','prod4']


def get_new_token():
        auth_server_url = '***'
        client_id = '*********'
        client_secret = '*********'
        token_req_payload = {'grant_type': 'client_credentials', 'scope': 'https://api.businesscentral.dynamics.com/',
                             'resource':'https://api.businesscentral.dynamics.com/'}
        token_response = requests.post(auth_server_url, data=token_req_payload, verify=False, allow_redirects=True,
                             auth=(client_id, client_secret))
        if token_response.status_code !=200:
                             print("Failed to obtain token from the OAuth 2.0 server", file=sys.stderr)
                             sys.exit(1)
        print("Successfuly obtained a new token")
        tokens = json.loads(token_response.text)
        return tokens['access_token']

for env in environment_list:
   api_call_company = "https://api.businesscentral.dynamics.com/v2.0/"+env+"/api/v2.0/companies"
   api_call_company_response = requests.get(api_call_company, headers=api_call_headers, verify=False)
   company_df = pd.read_json(StringIO(api_call_company_response.text))
   company_df.info()
   company_df["environment"] = env
   company_json = company_json.append(company_df)
```

Once the pharmacies name is obtained, the information is written to the *stg_companies* table in Amazon Redshift for future analysis steps.

```
Redshift_Companies = glueContext.write_dynamic_frame.from_jdbc_conf(
   frame=company_ddf,
   catalog_connection="dynamics_integration_redshift_connection",
   connection_options={
      "database": "prod",
      "dbtable": "hh_dynamics_svil.stg_companies",
      "preactions": "truncate table hh_dynamics_svil.stg_companies"
   },
   redshift_tmp_dir=args["TempDir"],
)
```

Finally, a new directory is created within the S3 bucket to hold output files from subsequent Glue jobs. The folder name appends the current date to maintain an organized structure over time as more jobs are run. Parquet files from downstream Glue jobs will load results into this shared location for unified data management.

```
NewDate = s3.Object('hippocrates-dynamics-integration-test', prefix + LoadDate)
NewDate.put()
```

By separating functionality into these logical sections and stages, the code demonstrates best practices for readability, reusability and scalability of ETL processes on AWS.

The pharmacy list generated by the AWS Glue job through querying the central business systems is loaded into the *stg_companies* table in Amazon Redshift.

The fields captured and stored within Redshift are illustrated in Figure 16 and Figure 17. By loading the data into a staging area like *stg_companies*, the information can be used in the feature steps to explore, transform and combine this information with other data assets residing in the data warehouse.

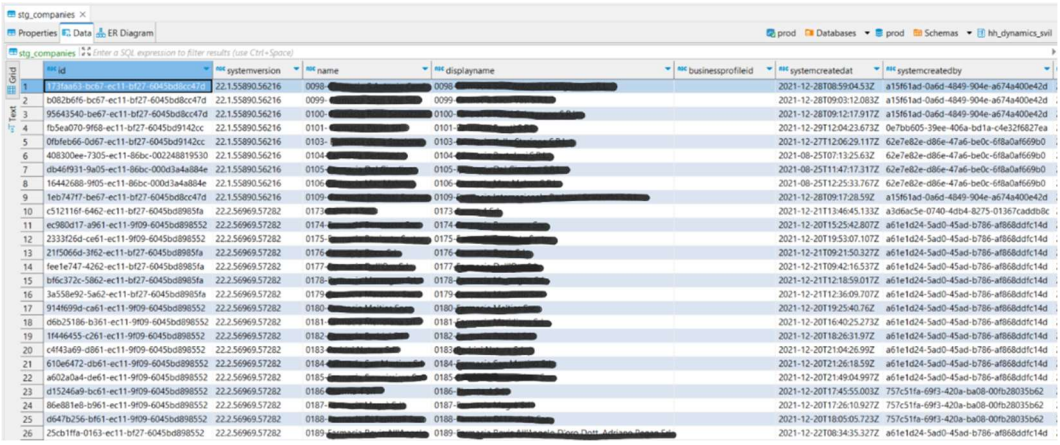


Figure 16 - Redshift stg_companies

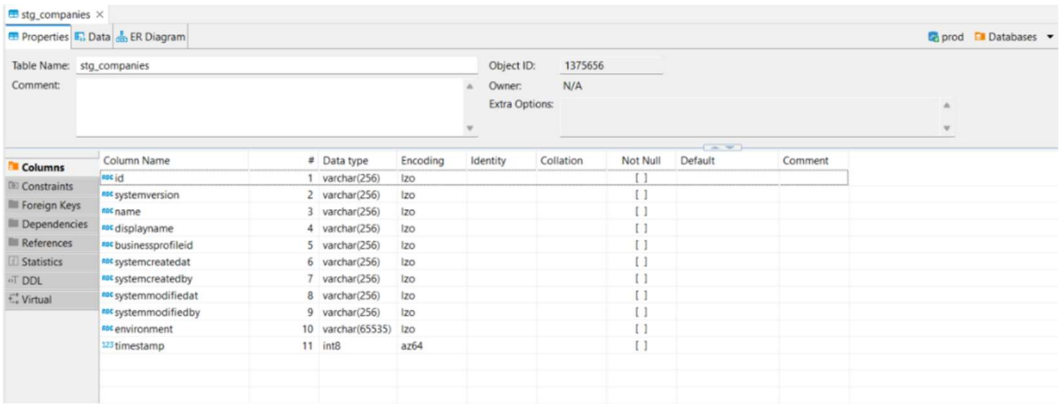


Figure 17 - Redshift stg_companies list columns

2) GLUE: from Business Central to AWS S3

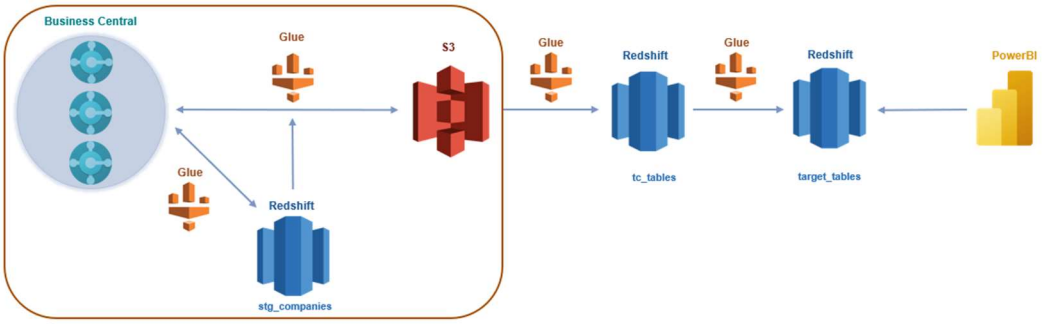


Figure 18 - Representation of architecture – Focus 2

The proposed architecture incorporates three distinct Glue jobs: *dynaimcs-integration-run-api-Prod3*, *dynaimcs-integration-run-api-Production*, and *dynaimcs-integration-run-api-Prod2*. These jobs have been designed to access and retrieve data from different business central environments. The decision to create three separate jobs, despite their similarity, stems from the intention to leverage parallel data extraction, optimizing execution time.

The primary objective of these three jobs is to invoke the APIs of various companies, utilizing environment filters, and subsequently download the data into parquet files on S3. This entire process is managed through a highly flexible and dynamic configuration, allowing for parametric control over the API calls.

Now, let's delve into the execution logic shared by these three jobs, examining the underlying details. In the initial stage, the necessary libraries and features are once again imported, ensuring access to the required resources for seamless execution. This step ensures that the jobs have the essential components readily available to perform their designated tasks effectively.

The subsequent step in the architecture entails the retrieval of API information from Redshift. The Redshift table known as *cfg_apistring*, is depicted in Figure 19 and Figure 20. It serves as a repository for comprehensive information about each company, including its associated link of API.

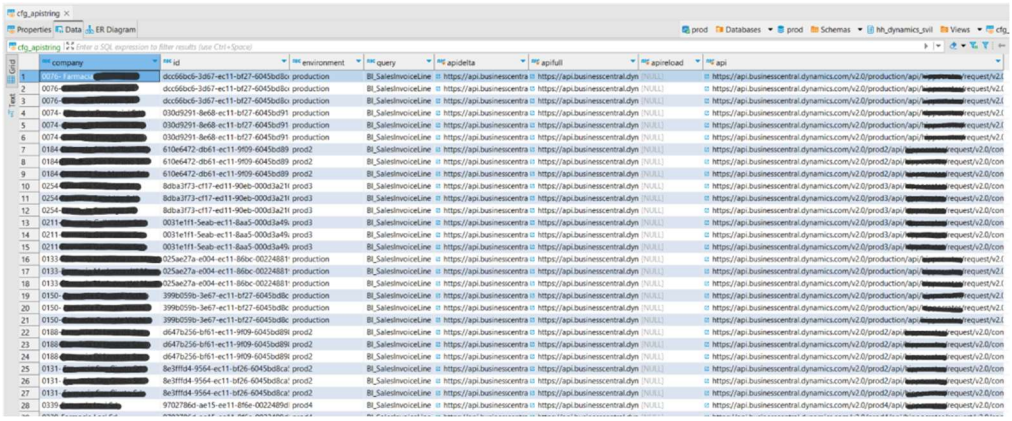


Figure 19 - Redshift cfg_apistring

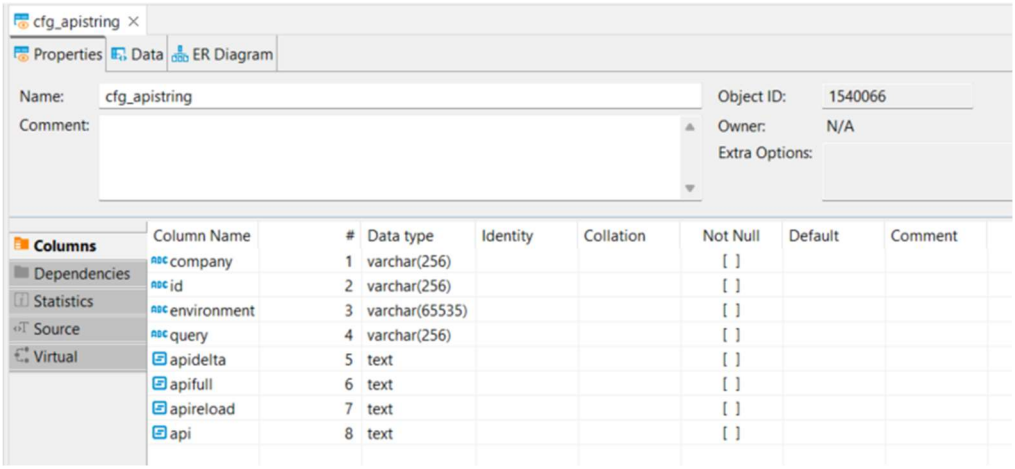


Figure 20 - Redshift cfg_apistring list columns

To ensure specificity for the reference environment, the Redshift *cfg_apistring* table is queried, filtering for the required data related to the URLs of the APIs that the job will invoke for data download. This filtering process results in the extraction of all relevant API URLs, which are then stored in a DataFrame for further processing.

```
# Script generated for node Redshift Cluster
APIToRun_DynFr = glueContext.create_dynamic_frame.from_catalog(
    database="dynamics_integration_test",
    redshift_tmp_dir=args["TempDir"],
    table_name="dynamics_integration_prod_hh_dynamics_svil_cfg_apistring",
)

APIToRun_DF = APIToRun_DynFr.toDF()
APIToRun_DF = APIToRun_DF.filter(APIToRun_DF.environment == "prod2")
APIToRun = APIToRun_DF.collect()
```

Following this, the Token generation process ensues, similar to what was performed in the initial step of the architecture. A dedicated function is employed to establish a connection to the server

and procure the necessary authentication token for making subsequent API calls. This token serves as the key credential required to access the desired data.

```
def get_new_token():
        auth_server_url = "*******"
        client_id = '*******'
        client_secret = '*******'
        token_req_payload = {'grant_type': 'client_credentials', 'scope': 'https://api.businesscentral.dynamics.com/',
'resource': 'https://api.businesscentral.dynamics.com/'}
        token_response = requests.post(auth_server_url, data=token_req_payload, verify=False, allow_redirects=True,
auth=(client_id, client_secret))
        if token_response.status_code !=200:
                print("Failed to obtain token from the OAuth 2.0 server", file=sys.stderr)
                sys.exit(1)
        print("Successfuly obtained a new token")
        tokens = json.loads(token_response.text)
        return tokens['access_token']
```

Subsequently, the extraction of Logs commences. An intelligently designed function takes charge of managing any exceptions that may arise during the process, ensuring robust error handling. Moreover, this function captures and saves the generated logs in a DataFrame, creating a cohesive record of the job's execution. These logs, containing valuable information and insights, are ultimately written to a designated Redshift table for comprehensive logging and auditing purposes.

```
def api_log(company, query, page, api_to_run, api_response):
  company_start_value = [[company, query, api_to_run, page, str(datetime.today()), '','','']]
  company_start_log = spark.createDataFrame(company_start_value,log_schema)
  try:
    api_response.raise_for_status()
  except requests.exceptions.HTTPError as err:
    error_RDD = spark.sparkContext.parallelize([err.response.text])
    error_DF = spark.read.json(error_RDD)
    error_DF = error_DF.select("error.message")
    error_list = error_DF.collect()
    error_message = (error_list[0].__getitem__('message'))
    company_start_log = company_start_log.withColumn('end_datetime', lit(str(datetime.today())))
    company_start_log = company_start_log.withColumn('status_code', lit(str(api_response.status_code)))
    company_start_log = company_start_log.withColumn('error_msg', lit(str(error_message)))
  else:
    company_start_log = company_start_log.withColumn('end_datetime', lit(str(datetime.today())))
    company_start_log = company_start_log.withColumn('status_code', lit(str(api_response.status_code)))
  finally:
    return company_start_log


log_DF = spark.createDataFrame(emptyRDD, schema = log_schema)
```

To provide a comprehensive overview of the job's workflow, it is essential to delve into the intricacies of the process. Firstly, all API schematics, which are stored in another configuration

table known as *cfg_dataframeschema*, are systematically iterated upon. This iterative process involves cycling through each of the previously retrieved APIs and leveraging the schematics stored within the aforementioned DataFrame. These schematics serve as a blueprint for data extraction and transformation, providing vital instructions on how to process the data effectively.

During the execution of the job, each API is called upon, and the corresponding response JSON is read. The job intelligently handles the processing of data spread across multiple pages, seamlessly unifying them into a single cohesive file. This consolidation ensures that the final output adheres to the desired structure and format, which is the parquet format in this case. Once the data has been unified and transformed according to the provided schematics, it is ready to be written to the designated S3 bucket. This storage location serves as a secure and scalable repository for the extracted and processed data.

Overall, this systematic and comprehensive approach to API data extraction and processing ensures optimized execution and effective utilization of resources. By leveraging the power of Redshift, intelligent token authentication, and robust logging mechanisms, the job is able to efficiently retrieve, transform, and store data, facilitating smooth and reliable workflows within the architectural framework.

3)GLUE: from Aws S3 to Redishift TC_tables

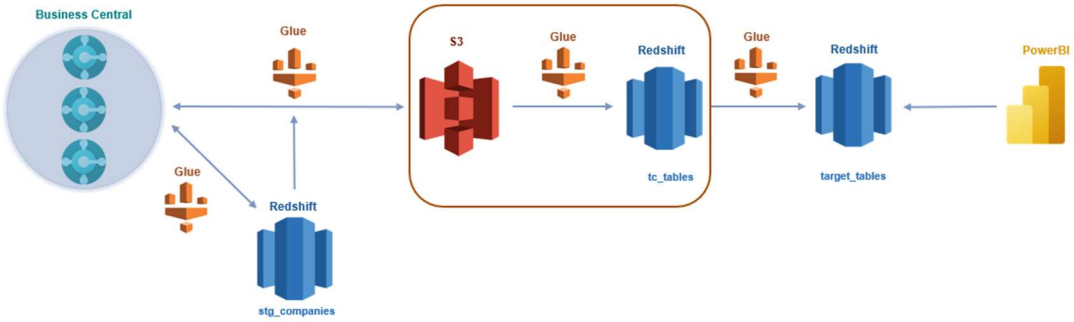


Figure 21 - Representation of architecture – Focus 3

The current data storage mechanism in the process involves the utilization of AWS S3. All data is securely saved within the AWS S3 infrastructure, ensuring its availability and durability.

To facilitate the seamless integration of this data into Redshift, a crucial Glue job called *dynamics-integration-load-tc* comes into play. This job operates loading all the data residing within AWS S3 into the corresponding Redshift tables.

To gain a comprehensive understanding of the Redshift tables' structure, Figure 22 proves invaluable. It provides an exhaustive list of all the tables that currently exist within the Redshift environment. Notably, the job *dynamics-integration-load-tc* primarily focuses on loading data into the initial layer of tables, which can be identified by the presence of the prefix "tc" in their respective names.

- stg_tc_bi_atc_fpas_stagingheaders
- stg_tc_bi_atc_fpas_staginglines
- stg_tc_bi_atc_glentriesec
- stg_tc_bi_atc_wing_itemjnlinvheaders
- stg_tc_bi_atc_wing_itemsjourinvdet
- stg_tc_bi_atc_wing_paymentlines
- stg_tc_bi_atc_wing_purchheaders
- stg_tc_bi_atc_wing_purchlines
- stg_tc_bi_atc_wing_salesheaders
- stg_tc_bi_atc_wing_saleslines
- stg_tc_bi_atclogheaders
- stg_tc_bi_atclogline
- stg_tc_bi_bankaccounts
- stg_tc_bi_cashclosuredetails
- stg_tc_bi_cashclosureflow_wing
- stg_tc_bi_cashclousureheader
- stg_tc_bi_companyinformation
- stg_tc_bi_customer
- stg_tc_bi_customerregisterflow
- stg_tc_bi_customerstaging
- stg_tc_bi_dimensonsetentry
- stg_tc_bi_doclinelinkedtoorders
- stg_tc_bi_filesregistry
- stg_tc_bi_genjournallines
- stg_tc_bi_genjournaltemplate
- stg_tc_bi_glaccounts
- stg_tc_bi_glentries
- stg_tc_bi_itemcategory
- stg_tc_bi_itemjournalflow_wing
- stg_tc_bi_itemledgerentry
- stg_tc_bi_items
- stg_tc_bi_jobqueueentry
- stg_tc_bi_jobqueuelogentry
- stg_tc_bi_location
- stg_tc_bi_purchaseheaders
- stg_tc_bi_purchaselines
- stg_tc_bi_purchcrmemohdrs
- stg_tc_bi_purchcrmemolines
- stg_tc_bi_purchheaderflow_wing
- stg_tc_bi_purchinvheaders
- stg_tc_bi_purchinvlines
- stg_tc_bi_purchrcptheader
- stg_tc_bi_purchrcptline

Figure 22 - Redshift list of tc tables

In Figure 23, a detailed view of one of the tables is presented, focusing specifically on the *stg_tc_bi_cashclosuredetails* table. This image displays a comprehensive list of all the associated columns along with their corresponding data.

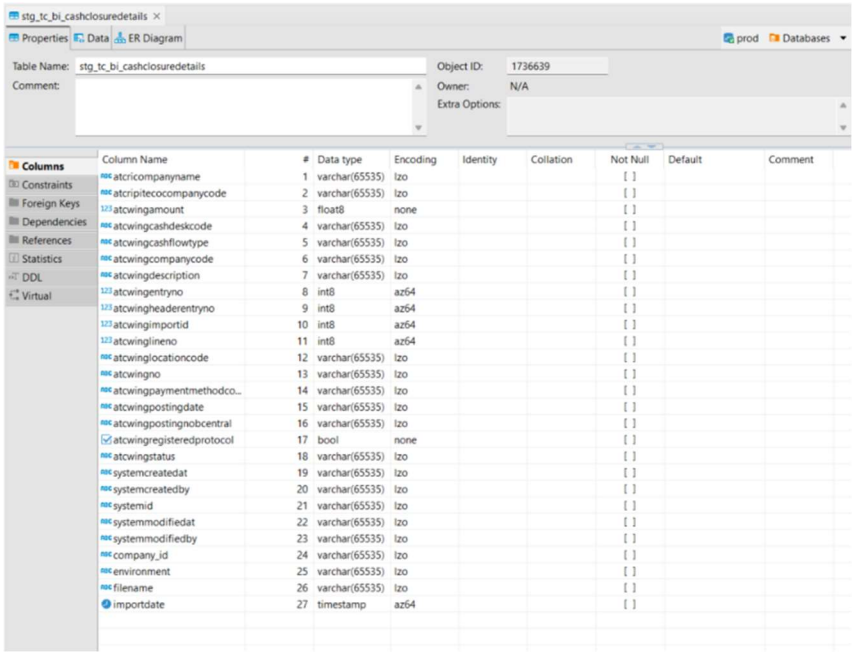


Figure 23 - Redshift stg_tc_bi_cashclosuredetails list of columns

By employing this systematic approach, the integration process ensures that the data is efficiently transferred from AWS S3 to Redshift, enabling further analysis, querying, and manipulation of the information contained within the tables. This seamless flow of data facilitates streamlined operations and enhances the overall effectiveness of data management within the system.

4)GLUE: from Redishift TC_tables to Redishift target_tables

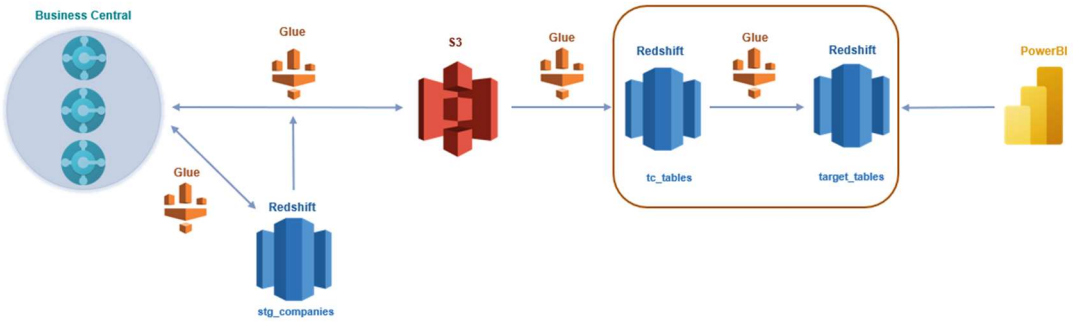


Figure 24 - Representation of architecture – Focus 4

The final phase entails the transmission of a subset of the data contained in the table-tc to the target-tables residing on the Redshift platform. This operation is facilitated by leveraging the capabilities of the Glue job known as *dynamics-integration-load-target*, which serves as the catalyst for executing a series of essential processes for loading data. These processes are executed by invoking specific stored procedures within the job.

To delve into the technical specifics, within the Spark code in the Glue job, subsequent to the initial import of the requisite data, a stored procedure named *hh_dynamic_svil.sp_populate_tables_master()* is defined within the Redshift environment.

```
glue_connection_name = 'dynamics_integration_redshift_connection'
database_name = 'prod'
stored_proc = 'hh_dynamics_svil.sp_populate_tables_master()'
```

This particular procedure, residing within Redshift, boasts a comprehensive implementation, and its underlying code can be examined in detail by referring to Figure 25.

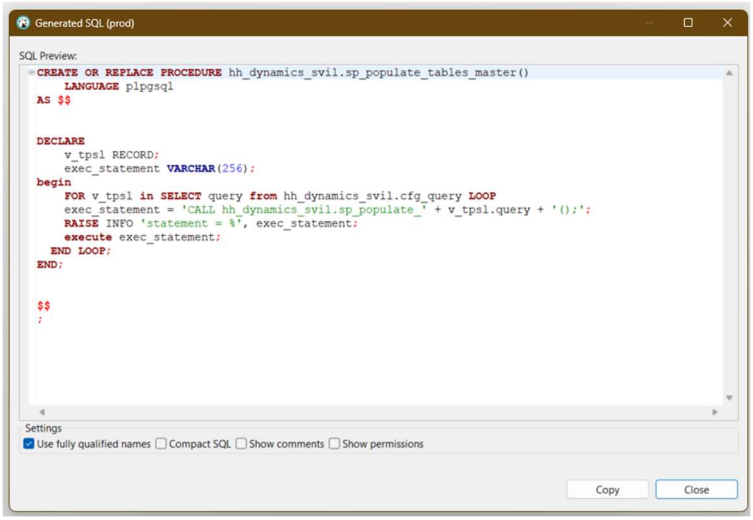


Figure 25 - Redshift sp_populate_tables_master()

Proceeding with the execution, a for loop is employed to discern the individual names of the tc-tables. Each of these tables is then associated with a corresponding function referred to as *sp_populate_bi_*.

It can be seen the list of them in Figure 26.

Figure 26 - Redshift functions

These functions are subsequently launched. A portion of the information housed in the tc-tables is selectively transferred to the target-tables, aligning with the predefined table structure established in Redshift views. Notably, these views are labeled with a distinctive naming convention, commencing with *stg_vw_bi*.

By virtue of this intricate orchestration, the target-tables are populated with great efficacy.

The list of the filled target-tables is visible in Figure 27.

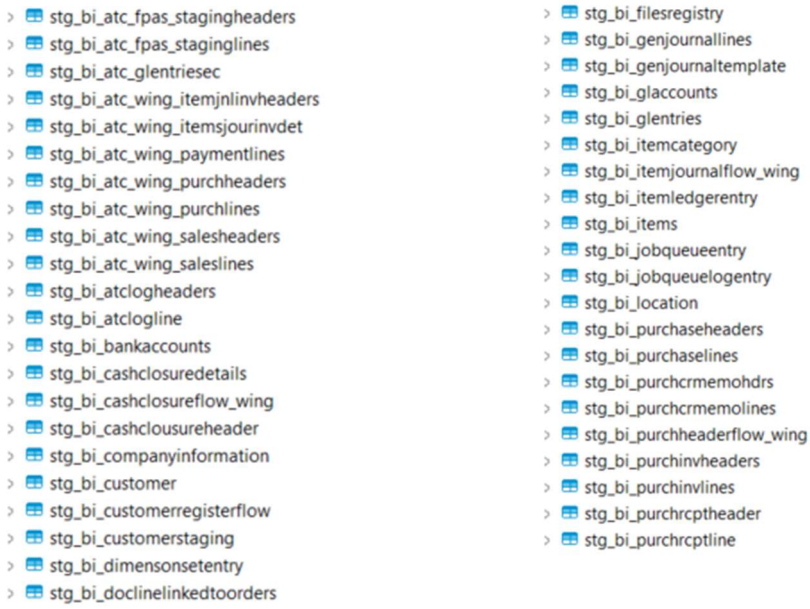


Figure 27 - Redshift target-tables

This transformational process, which entails the transition from tc-tables to target tables, serves as a pivotal means of manipulating and refining the underlying data sourced from the origin. The ultimate objective is to streamline the data by exclusively incorporating the columns that hold significance for future analytical endeavors. In essence, this meticulous optimization ensures that the target tables are enriched with only the pertinent information, thereby facilitating streamlined analysis and expediting insightful decision-making.

5)Power BI form Redishift target_tables

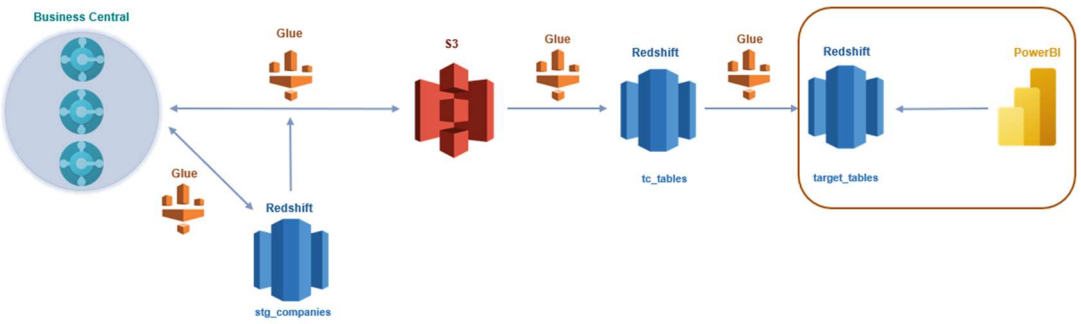


Figure 28 - Representation of architecture – Focus 5

Towards the culmination of the described architecture, the ultimate step encompasses the presentation of the meticulously collected data on Power BI dashboards. This pivotal process is accomplished by establishing a connection between Power BI and the Redshift Views. As expounded earlier, Redshift Views serve as a repository for storing intricate queries pertaining to the target tables, effectively harnessing the power of in-memory data processing.

The Redshift Views in question are manifold, yet they all share a common nomenclature convention, commencing with the prefix *powerbi_vw_* , as discernible from the comprehensive list showcased in Figure 29. Each of these views encapsulates a tailored perspective on the underlying data, strategically designed to align with the specific analytical requirements of the Power BI dashboards.

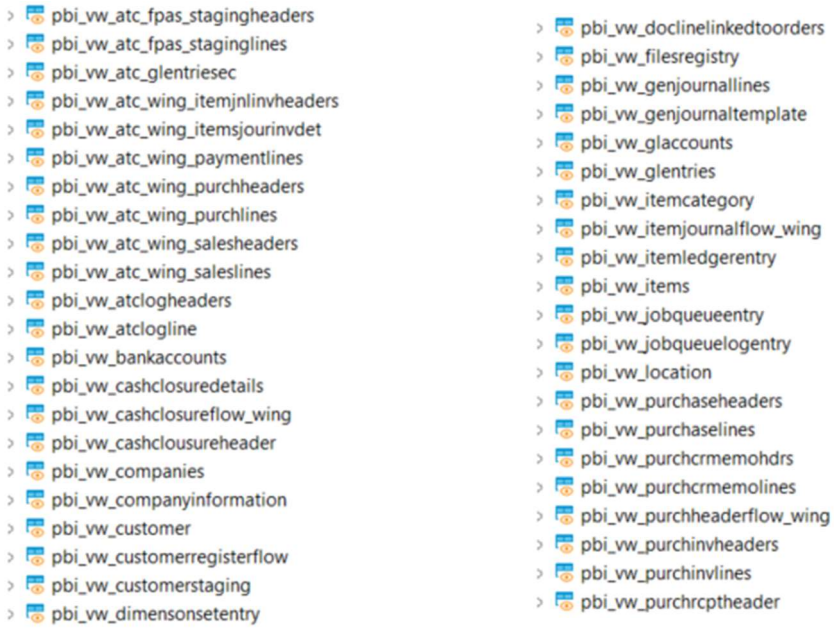


Figure 29 - Redshift views

For illustrative purposes, it is delved into the structure of a particular Redshift View, namely *pbi_vw_filesregistry*, as elucidated in Figure 30.

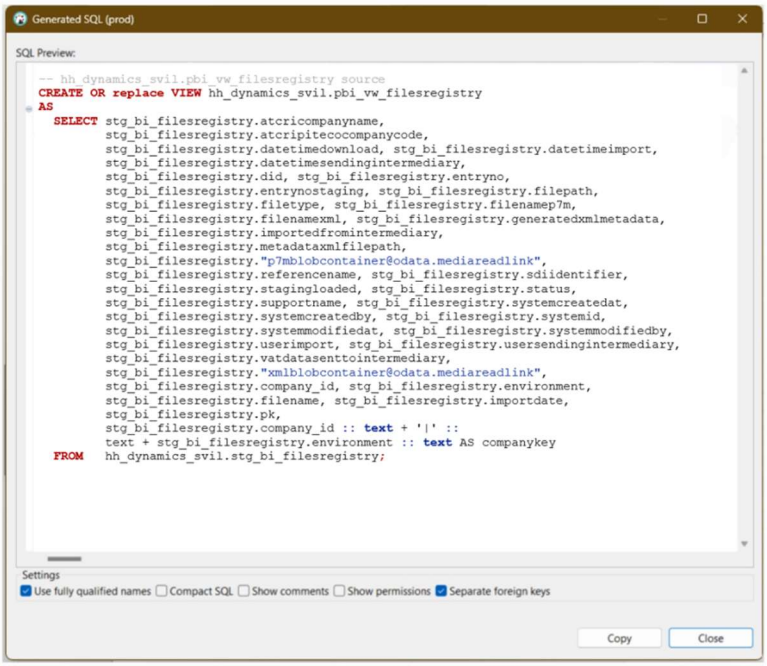


Figure 30 - Redshift pbi_vw_customer view

This particular view exemplifies the intricate interplay between the view and its associated query, which diligently filters the target-table labeled *stg_tc_bi_filesregistry* show in Figure 31. By skillfully applying the filtering conditions stipulated in the query, the view selectively curates the data, ensuring that only the pertinent information pertaining to the *stg_tc_bi_filesregistry* target-table is presented on the Power BI dashboards.

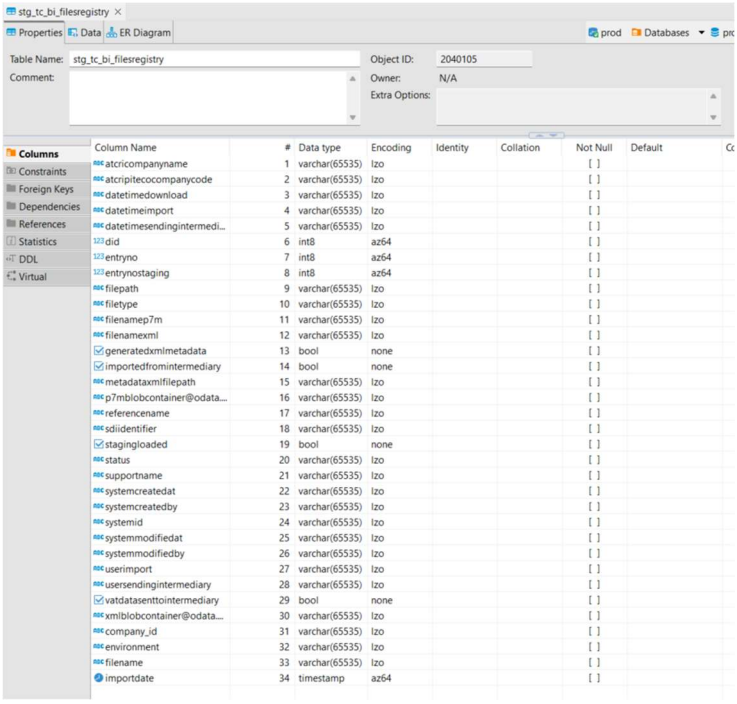


Figure 31 - Redshift stg_tc_bi_filesregistry list of columns

This harmonious integration between Power BI and the Redshift Views empowers users to effortlessly navigate and explore the wealth of data collected, enabling them to glean valuable insights and make informed decisions. The dynamic nature of the Redshift Views, coupled with the robust querying capabilities of the Power BI platform, synergistically converge to deliver a transformative analytical experience.

Power BI showcases four distinct dashboards that provide the users with a platform for conducting in-depth analysis. These dashboards are named as follows: Active Invoice Control, Passive Invoice Control, General Ledger, and Transient Collection. Each dashboard offers a unique perspective and set of insights for analysis. The forthcoming exploration will delve into a comprehensive examination of these dashboards, unraveling their intricacies and unlocking valuable information that can be leveraged for informed decision-making. Through this detailed analysis, users gain a profound understanding of the data presented on each dashboard, enabling them to extract meaningful insights and drive actionable outcomes.

Active invoice control

The Delta company is introduced to its dashboard named *Active invoice control dashboard*, on which the sales data of each pharmacy are collected. The representation of this dashboard as the initial point of data exploration and analysis can be seen in Figure 32.

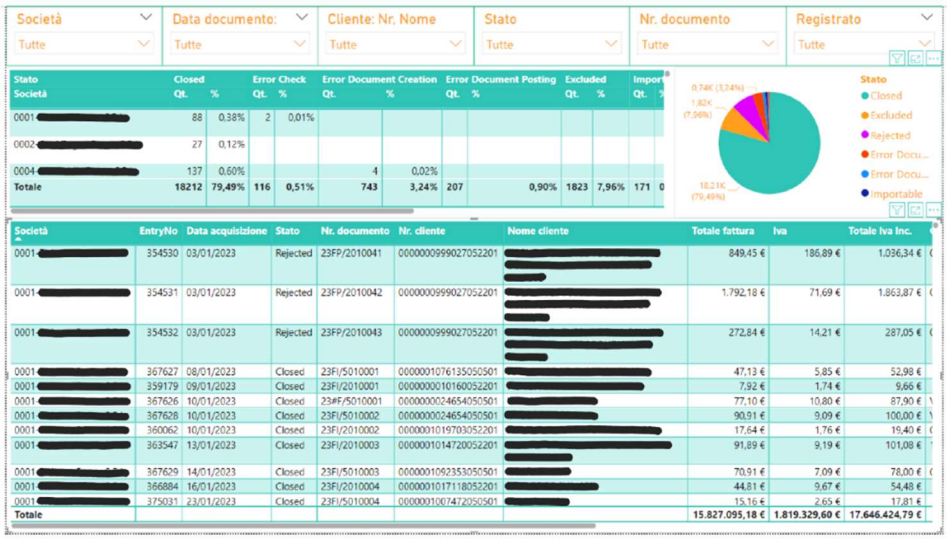


Figure 32 - Power Bi - Active invoice control

The *Active invoice control dashboard* is meticulously crafted by integrating data obtained from four distinct views, which in turn establish connections with specific target tables within the Redshift database. To be specific, these views are referred to as *pbi_vw_atg_wing_salesheaders*, *pbi_v_atcw_wings_saleslines*, *pbi_vw_companies*, and

pbi_v_customer. Each of these views contributes essential information to the construction of the dashboard, enabling a comprehensive assessment of active invoices.

This dynamic dashboard, designed to scrutinize active invoices, harnesses the power of the underlying sales data and meticulously investigates the status and associated details of each invoice. For every invoice, a wealth of information is collected, including the customer code, invoice status, customer name, total invoice cost, VAT and other.

The absence of the pharmacy name in the *header_table* prompted the need for a comprehensive resolution. To address this, a strategic join operation was conducted, seamlessly integrating the *header_table* and the *company_table*. This harmonious union was facilitated through the utilization of the common primary key present in both tables, ensuring a cohesive amalgamation of relevant information.

The significance of this dashboard, rich with customer and pharmacy-related data, becomes apparent when considering its utility for the finance department of the organization. Armed with access to detailed analytics, finance professionals can engage in an in-depth exploration of invoice progress and customer-related insights. This empowers the finance department to conduct granular analyses, monitoring the trajectory of invoices and gaining valuable insights into customer behavior.

By leveraging this dashboard, the finance department can delve into the intricacies of invoice management, track payment statuses, identify potential bottlenecks, and optimize financial operations. Moreover, the availability of customer-specific data equips the finance team to evaluate the financial relationship with individual customers, enabling them to make informed decisions and foster stronger customer partnerships.

Ultimately, through the active invoice control dashboard, the finance department gains a comprehensive toolset that facilitates meticulous analysis, provides visibility into invoice-related metrics, and empowers them to steer financial strategies with precision and confidence.

Passive Invoice Control

Delta's second dashboard offering is the Passive Invoice Control, which serves as a counterpart to the previously described Active Invoice Control dashboard, sharing a similar underlying structure as it can be seen in Figure 33.

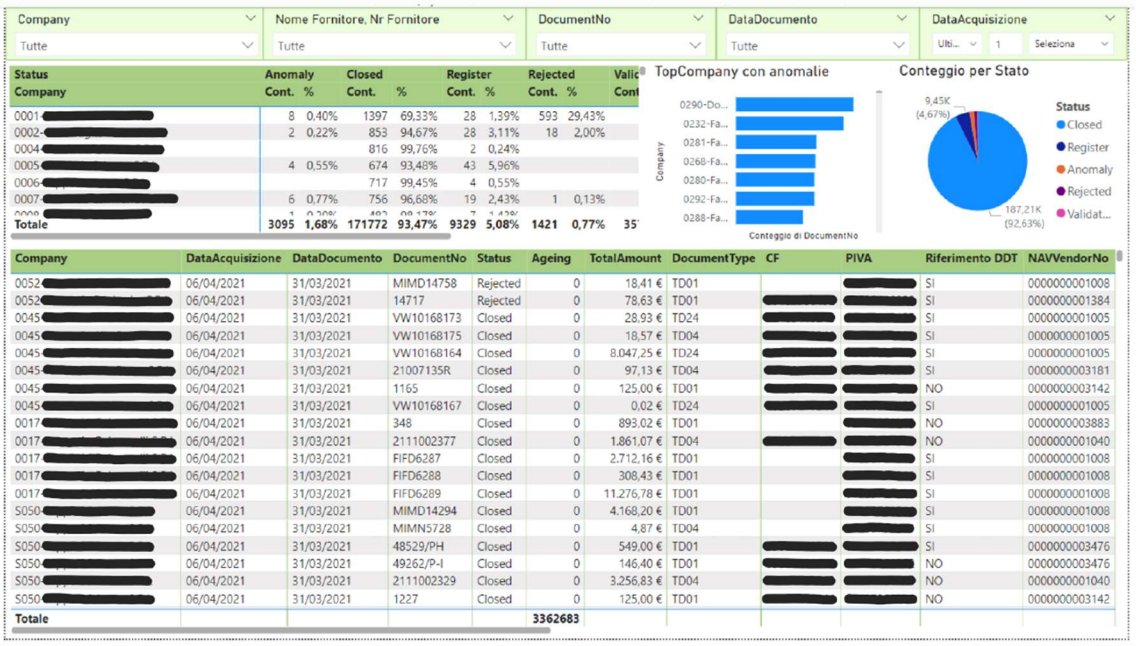


Figure 33 - Power Bi – Passive invoice control

This dashboard establishes connections with new views, namely *pbi_v_atc_fpas_stagingheaders*, *pbi_vw_companies*, and *pbi_v_purchinvheaders*, to facilitate its operations.

The primary objective of the Passive Invoice Control dashboard is to gather and analyze invoices payable, representing the acquisitions made by each pharmacy from various suppliers. It intertwines information from both the pharmacies and the suppliers, creating a comprehensive understanding of the debit transactions.

Within the tables integrated into this dashboard, a wealth of valuable information is available. It encompasses details such as the reference pharmacy, purchase dates of products, invoice amounts, and supplier-related data, including tax codes, VAT, supplier names, payment information, and invoice specifics.

By combining and harmonizing data from pharmacies and suppliers, the Passive Invoice Control dashboard provides a holistic view of financial transactions. It enables insightful analysis of the purchasing activities of individual pharmacies, shedding light on expenses, supplier relationships, and payment trends. This comprehensive perspective empowers Delta's finance

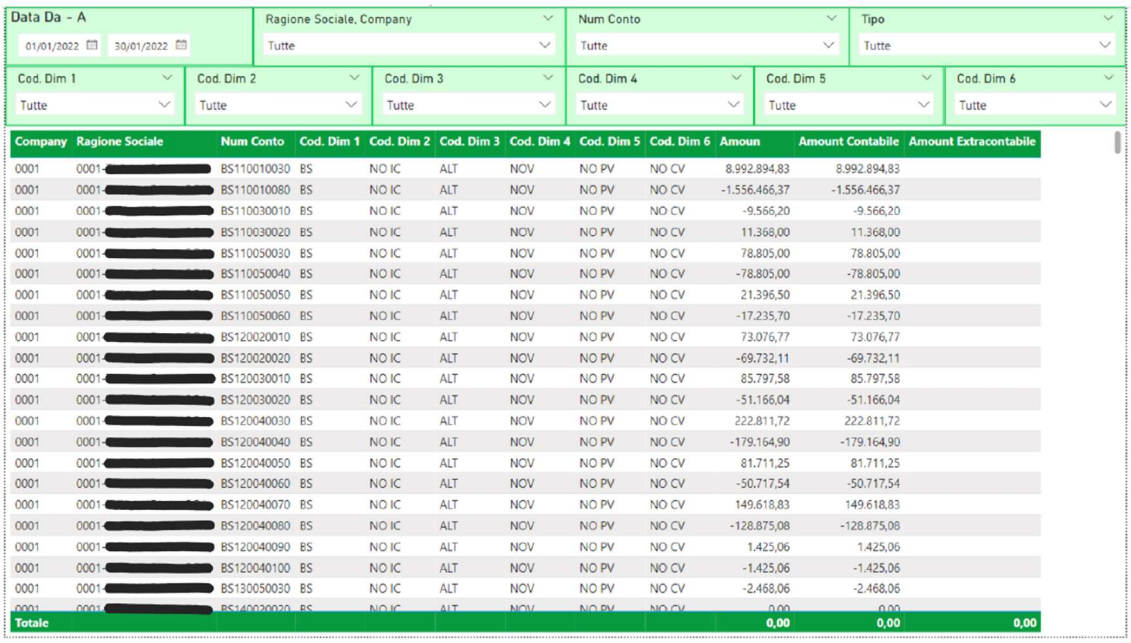
department to optimize cash flow management, negotiate favorable terms with suppliers, and ensure timely payment of invoices.

The seamless integration of pharmacy and supplier data within the Passive Invoice Control dashboard serves as a powerful tool for Delta's financial decision-making processes. By harnessing the detailed information provided, the finance department can identify areas for cost-saving, improve overall financial performance, and foster strong partnerships with suppliers.

The Passive Invoice Control dashboard acts as a bridge between pharmacies and suppliers, facilitating a comprehensive analysis of payable invoices. Its rich dataset empowers Delta's finance department to make informed decisions, streamline financial operations, and forge strategic alliances, ultimately driving the company towards enhanced profitability and financial stability.

General Ledger

The third dashboard in the suite presented to clients, known as the General Ledger, stands as a sophisticated and comprehensive tool that extends far beyond traditional financial reporting within pharmacy management, as it can be seen in Figure 34.



Company	Ragione Sociale	Num Conto	Cod. Dim 1	Cod. Dim 2	Cod. Dim 3	Cod. Dim 4	Cod. Dim 5	Cod. Dim 6	Amoun	Amount Contabile	Amount Extracontabile
0001	0001-[redacted]	BS110010030	BS	NO IC	ALT	NOV	NO PV	NO CV	8.992.894,83	8.992.894,83	
0001	0001-[redacted]	BS110010080	BS	NO IC	ALT	NOV	NO PV	NO CV	-1.556.466,37	-1.556.466,37	
0001	0001-[redacted]	BS110030010	BS	NO IC	ALT	NOV	NO PV	NO CV	-9.566,20	-9.566,20	
0001	0001-[redacted]	BS110030020	BS	NO IC	ALT	NOV	NO PV	NO CV	11.368,00	11.368,00	
0001	0001-[redacted]	BS110050030	BS	NO IC	ALT	NOV	NO PV	NO CV	78.805,00	78.805,00	
0001	0001-[redacted]	BS110050040	BS	NO IC	ALT	NOV	NO PV	NO CV	-78.805,00	-78.805,00	
0001	0001-[redacted]	BS110050050	BS	NO IC	ALT	NOV	NO PV	NO CV	21.396,50	21.396,50	
0001	0001-[redacted]	BS110050060	BS	NO IC	ALT	NOV	NO PV	NO CV	-17.235,70	-17.235,70	
0001	0001-[redacted]	BS120020010	BS	NO IC	ALT	NOV	NO PV	NO CV	73.076,77	73.076,77	
0001	0001-[redacted]	BS120020020	BS	NO IC	ALT	NOV	NO PV	NO CV	-69.732,11	-69.732,11	
0001	0001-[redacted]	BS120030010	BS	NO IC	ALT	NOV	NO PV	NO CV	85.797,58	85.797,58	
0001	0001-[redacted]	BS120030020	BS	NO IC	ALT	NOV	NO PV	NO CV	-51.166,04	-51.166,04	
0001	0001-[redacted]	BS120040030	BS	NO IC	ALT	NOV	NO PV	NO CV	222.811,72	222.811,72	
0001	0001-[redacted]	BS120040040	BS	NO IC	ALT	NOV	NO PV	NO CV	-179.164,90	-179.164,90	
0001	0001-[redacted]	BS120040050	BS	NO IC	ALT	NOV	NO PV	NO CV	81.711,25	81.711,25	
0001	0001-[redacted]	BS120040060	BS	NO IC	ALT	NOV	NO PV	NO CV	-50.717,54	-50.717,54	
0001	0001-[redacted]	BS120040070	BS	NO IC	ALT	NOV	NO PV	NO CV	149.618,83	149.618,83	
0001	0001-[redacted]	BS120040080	BS	NO IC	ALT	NOV	NO PV	NO CV	-128.875,08	-128.875,08	
0001	0001-[redacted]	BS120040090	BS	NO IC	ALT	NOV	NO PV	NO CV	1.425,06	1.425,06	
0001	0001-[redacted]	BS120040100	BS	NO IC	ALT	NOV	NO PV	NO CV	-1.425,06	-1.425,06	
0001	0001-[redacted]	BS130050030	BS	NO IC	ALT	NOV	NO PV	NO CV	-2.468,06	-2.468,06	
0001	0001-[redacted]	BS140020020	BS	NO IC	ALT	NOV	NO PV	NO CV	0,00	0,00	
Totale									0,00	0,00	0,00

Figure 34 - Power Bi – General Ledger

This dashboard provides an all-encompassing perspective on the financial dynamics of pharmacies, shedding light not only on the conventional active and passive invoices but also on the intricate details of any cash flow, including the specifics of employee salaries.

Integral to the functionality of the General Ledger dashboard are the associated Redshift views, which play a pivotal role in offering a nuanced and detailed representation of financial data. These views, including G/L Entries, *pbi_vw_companies*, *pbi_v_glentries*, and *pbi_v_glentriesec*, form the backbone of the dashboard, facilitating a thorough exploration of the intricacies within pharmacy financial transactions.

Tailored to the unique needs of individual reference pharmacies, the dashboard provides a granular breakdown of accounts associated with each pharmacy. The primary table, prominently visualized on the Power BI platform, furnishes key information such as the pharmacy's company name and accounting paragraph numbers. This level of detail proves invaluable, especially when examining transactions like employee salaries, which are systematically linked to distinct account numbers. Delta, leveraging its internal data mapping, can seamlessly trace these

transactions, contributing to a comprehensive understanding of the financial movements within pharmacies.

The General Ledger dashboard, therefore, emerges as a holistic financial compass, offering a panoramic view of the multifaceted dimensions of pharmacy finances. It serves not only as a tool for understanding and analyzing traditional invoicing but also provides a broader context for various financial transactions.

The detailed breakdown of financial data within the dashboard, facilitated by elements like G/L Entries and associated Redshift views, empowers users to navigate through company-specific information. This ensures a comprehensive understanding of transactions tied to individual pharmacies, allowing for informed decision-making, strategic financial planning, and a deeper exploration of the financial health of each entity.

The General Ledger dashboard transcends conventional financial reporting, standing as a sophisticated instrument for clients seeking a profound and comprehensive analysis of pharmacy financials. Its capability to provide detailed insights into financial transactions, coupled with meticulous data mapping for transaction tracing, positions it as an indispensable asset in the realm of strategic financial management and decision support.

Transient Collections

The last report submitted to the client is titled "Transient Collections," and its structural layout is depicted in Figure 35. Within this report, nine tables are intricately linked, namely: *Calendar*, *Cash Closure Details*, *Cassa*, *Companies*, *Verteron*, *GL Entries*, *Sales*, *Store Masterdata*, *WING Sales Header*.

displayname: 0001[redacted] | Month-Year: gen 2023

Company	Posting_Date	TI.Vendite	TI.Scontrino	TI.Documenti	TI.Cassa	TI.DeltaCassa	TI.Resi	NonIncas	CInErrore	Ade.Corrispettivi	Diff = TI-AdE	AdE.Importo	AdE.Resi	AdE.SSN
0001[redacted]	01.1.23	7.160,93	7.160,93		-7.164,96	-4,03	-3,99			€ 7.161,0	€ 0,0	€ 7.164,60	3,63	0,00
0001[redacted]	02.1.23	12.629,88	12.629,88		-11.284,60	1.345,28	-2,30			€ 12.629,2	€ 0,7	€ 12.631,10	1,89	0,00
0001[redacted]	03.1.23	12.035,22	12.035,22		-11.063,34	971,88	-8,37			€ 12.035,2	€ 0,0	€ 12.042,83	7,61	0,00
0001[redacted]	04.1.23	13.508,16	13.508,16		-12.477,86	1.030,30	-41,43			€ 13.508,2	€ 0,0	€ 13.544,92	36,76	0,00
0001[redacted]	05.1.23	14.475,31	14.475,31		-8.456,67	6.018,64	-91,57			€ 61.295,2	-€ 46.819,9	€ 61.376,38	81,16	0,00
0001[redacted]	06.1.23	14.562,46	14.562,46		-14.575,45	-12,99	-12,99			€ 14.562,5	€ 0,0	€ 14.573,11	10,65	0,00
0001[redacted]	07.1.23	18.332,10	18.332,10		-16.403,23	1.928,87	-54,40			€ 18.332,1	€ 0,0	€ 18.380,47	48,34	0,00
0001[redacted]	08.1.23	15.825,39	15.772,41	52,98	-7.576,14	8.249,25	-63,59			€ 7.576,1	€ 8.196,3	€ 7.576,14	0,00	0,00
0001[redacted]	09.1.23	18.352,25	18.342,59	9,66	-16.022,06	2.330,19	-58,81			€ 26.526,7	-€ 8.184,2	€ 65.668,48	104,90	39.036,84
0001[redacted]	10.1.23	17.631,20	17.423,90	207,30	-15.488,61	2.142,59	-36,68			€ 17.424,0	-€ 0,1	€ 17.457,31	33,35	0,00
0001[redacted]	11.1.23	18.501,48	18.501,48		-18.605,87	-104,39	-104,39			€ 18.548,7	-€ 47,2	€ 18.596,57	47,90	0,00
0001[redacted]	12.1.23	17.896,45	17.896,45		-11.079,71	6.816,74	-116,20			€ 17.896,5	-€ 0,1	€ 58.637,02	99,27	40.641,22
0001[redacted]	13.1.23	16.991,75	16.890,67	101,08	-17.067,37	-75,62	-75,60			€ 16.908,3	-€ 17,6	€ 16.958,64	50,35	0,00
0001[redacted]	14.1.23	18.466,71	18.388,71	78,00	-18.541,25	-74,54	-74,54			€ 18.388,7	€ 0,0	€ 18.451,94	63,22	0,00
0001[redacted]	15.1.23	17.210,49	17.210,49		-14.664,21	2.546,28	-60,93			€ 17.221,4	-€ 10,9	€ 17.265,31	43,88	0,00
0001[redacted]	16.1.23	17.451,30	17.396,82	54,48	-16.867,08	584,22	-254,14			€ 16.558,5	€ 838,4	€ 16.781,65	223,19	0,00
0001[redacted]	17.1.23	15.563,41	15.563,41		-13.919,42	1.643,99	-215,49			€ 15.563,4	€ 0,0	€ 15.755,30	191,87	0,00
0001[redacted]	18.1.23	17.695,71	17.695,71		-17.218,71	477,00	-87,91			€ 18.532,7	-€ 837,0	€ 18.612,66	79,92	0,00
0001[redacted]	19.1.23	16.280,12	16.280,12		-16.282,12	-2,00	-2,00			€ 16.280,1	€ 0,0	€ 16.281,94	1,82	0,00
0001[redacted]	20.1.23	18.301,47	18.301,47		-18.575,25	-273,78	-273,76			€ 18.333,6	-€ 32,1	€ 18.548,10	214,49	0,00
0001[redacted]	21.1.23	17.260,62	17.260,62		-17.401,73	-141,11	-141,07			€ 17.260,7	€ 0,0	€ 17.383,28	122,61	0,00
0001[redacted]	22.1.23	16.323,47	16.323,47		-16.323,53	-0,06				€ 16.323,5	-€ 0,1	€ 16.323,53	0,00	0,00
0001[redacted]	23.1.23	17.158,07	17.140,26	17,81	-17.190,01	-31,94	-31,90			€ 17.140,3	€ 0,0	€ 17.168,05	27,75	0,00
0001[redacted]	24.1.23	17.126,96	17.126,96		-14.929,71	2.197,25	-72,30			€ 17.140,6	-€ 13,7	€ 17.206,35	65,72	0,00
Totale		487.093,17	486.267,69	825,48	-448.741,08	38.352,09	-2.465,15			€ 575.357,7	-€ 89.090,0	€ 657.080,69	2.044,94	79.678,06

Figure 35 - Power Bi – Transient Collection

The inherent value of this report lies in its capacity to enable the client to scrutinize and compare cash movements meticulously recorded action by action with the data received from the revenue agency. The process unfolds seamlessly: every transaction initiated by a pharmacy is promptly communicated to the revenue agency, which, in turn, provides the financial department of the holding with a comprehensive record of the corresponding cash flow.

The money flow is characterized by both an incoming stream, reflecting the revenue generated from products sold, and an outgoing stream, accounting for expenditures on purchased products. The columns prefixed with "ti" meticulously document the cash flows recorded by the internal systems, encompassing elements such as *ti.sales*, *ti.cassa*, and *ti.resi*. On the other hand, columns prefixed with "ade" encapsulate all cash flows originating from the revenue agency, featuring entries like *Ade.Imposto* and *Ade.Resi*.

Crucially, the linchpin for accurate data analysis lies in the *Diff = TI-Ade* column, where disparities between the recorded flows are computed. Cells displaying non-zero values, essentially transactions in which there exists a dissonance between the actual cash flows and those declared to the revenue agency, are conspicuously coloured orange.

It is imperative to note that while all the collected data resides in Redshift, there exists a distinction: columns beginning with "ti" are part of the data collection managed directly within this project, whereas the data pertaining to columns referencing the revenue agency were supplied directly by the client, Delta.

This report emerges as an indispensable tool for the financial department, serving as a robust mechanism to scrutinize and identify any disparities between the actual and declared cash flows. The nuanced representation of data, with a keen focus on identifying discrepancies, enhances the financial department's ability to conduct thorough analyses and ensure the accuracy and integrity of financial records.

In essence, the Transient Collections report facilitates a granular examination of the intricacies within cash flows, providing a comprehensive overview that aids in financial transparency and compliance. Through its distinctive visual cues, such as the orange-highlighted cells, it not only streamlines the identification of discrepancies but also empowers the financial department to take swift corrective actions, thereby contributing to the overall financial health and integrity of the organization.

DATA Experience

In the aftermath of the findings garnered throughout the course of this project, it is imperative to underscore the significance of the data experience. The capability to adeptly traverse the data presented on the dashboard is crucial for extracting meaningful information that can serve as a foundation for future decision-making processes. This ability to derive insights from data contributes significantly to elevating the overall efficacy of organizational strategies.

The landscape of contemporary business operations is increasingly becoming data-centric, with companies recognizing the pivotal role of data visualization and the broader data experience. These aspects are now evaluated by numerous enterprises as the linchpin of their decision-making processes. The realization of the importance of a comprehensive data experience is paramount, as it empowers stakeholders to make well-informed decisions grounded in a nuanced understanding of complex datasets.

In essence, the concept of the data experience transcends mere data visualization; it encompasses the holistic interaction with and interpretation of data. Navigating the intricacies of the data landscape requires a nuanced approach that goes beyond surface-level observation. The depth of understanding derived from a rich data experience empowers decision-makers to transcend traditional paradigms and embrace a more conscious, informed, and forward-thinking approach.

Within the contemporary business ecosystem, data visualization has evolved into a central hub, serving as the nexus where raw data transforms into actionable insights. The interactive nature of data dashboards, exemplified by tools such as Power BI, has become a cornerstone of effective decision-making. These tools not only present data in visually engaging formats but also facilitate dynamic exploration, enabling users to delve into the nuances of datasets, uncover patterns, and extract valuable intelligence.

The importance of data visualization extends beyond the realms of aesthetics; it is a strategic imperative that fosters a deeper comprehension of the underlying data. As organizations grapple with increasingly complex datasets, the role of data visualization tools becomes more pronounced in translating information into actionable knowledge. Visual representations, when designed thoughtfully, have the capacity to distill complex datasets into comprehensible visual narratives, aiding decision-makers in discerning patterns and making informed choices.

In the current business landscape, the concept of the data experience is synonymous with cultivating a data-driven culture. This goes beyond the adoption of tools; it entails fostering a mindset that values data as a strategic asset. Organizations that prioritize the data experience recognize that the true power of data lies not just in its abundance but in the ability to harness

it meaningfully. A data-driven culture permeates every facet of an organization, from frontline decision-makers to C-suite executives, fostering a collective consciousness that values data as a potent catalyst for innovation and growth.

The trajectory of data utilization within organizations is underscored by a paradigm shift. Data is no longer relegated to the domain of IT departments; instead, it permeates every layer of the organizational hierarchy. The democratization of data access, facilitated by intuitive visualization tools, ensures that insights are not confined to a select few but are accessible to individuals across departments. This inclusivity in data interaction fosters collaboration and empowers diverse stakeholders to contribute to the decision-making process.

Power BI, as exemplified in the preceding chapter, stands out as a prominent player in the realm of data visualization tools. Its robust features and user-friendly interface make it a preferred choice for organizations seeking to harness the full potential of their data. The tool's ability to seamlessly integrate with diverse data sources, coupled with its dynamic visualization capabilities, positions it as a cornerstone in the architecture of modern data experiences.

The evolution of data visualization tools like Power BI aligns with the growing recognition that effective decision-making requires more than just access to data—it necessitates a comprehensive and immersive data experience. As organizations continue to amass vast datasets, the need to derive actionable insights becomes paramount. The agility offered by tools like Power BI enables users to explore, analyze, and derive meaning from data in real-time, thereby bridging the gap between raw data and strategic decision-making.

The centrality of data visualization in contemporary organizational strategies is a testament to its transformative impact. Beyond being a reporting mechanism, data visualization serves as a catalyst for innovation, driving organizational agility and responsiveness. The iterative nature of data exploration facilitated by visualization tools enables organizations to adapt swiftly to changing landscapes, identify emerging trends, and proactively shape their trajectories.

The paradigm of data experience extends beyond the mere visualization of data; it encapsulates the entire journey of interaction, interpretation, and decision-making fueled by data. The contemporary business landscape demands not only the utilization of powerful tools like Power BI but also a cultural shift that recognizes data as a strategic asset. Embracing a data-driven culture, wherein every stakeholder actively engages with and contributes to the data experience, is fundamental to navigating the complexities of the modern business environment. As organizations continue to harness the potential of data, the cultivation of a rich and immersive data experience emerges as a cornerstone for informed, conscious, and future-oriented decision-making.

A comprehensive exposition follows, delving into intricate details elucidating the operational mechanics of Power BI, expounding on its functionalities, and underscoring the profound importance it holds within the realm of contemporary data analytics and business intelligence.

Power BI

Microsoft developed the business analytics solution Power BI to help organizations transform raw numbers into meaningful insights. Users can structure data, link to many information sources, and construct interactive displays and dashboards using this tool. Power BI is capable of connecting to different data types like Excel sheets, SQL databases, Azure services, and more. Through the Power BI service, reports, dashboards, and data models offering visual perspectives are produced. Subsequently, the online Power BI environment publishes and shares these materials, allowing groups to securely collaborate. Thanks to dedicated Power BI mobile apps, performance can be monitored from tablets and smartphones as well, keeping teams connected regardless of location.

Power BI can be broken down into three core aspects: data connection, data visualization, and data modeling. Chiefly, it emphasizes Power BI's ability to link with over 200 information sources such as SQL databases, XML files, CRM programs, cloud databases, and more, allowing users to then import the data into Power BI. Reports are developed using engaging visual displays like charts, tables, graphs, and maps. Dashboards combine multiple reports and visualizations onto a single screen to provide at-a-glance insight into key metrics and indicators. Moreover, users can integrate, modify, and structure data from different sources by combining datasets, cleansing the information, and defining relationships to construct reports and dashboards. The data connection facility surfaces the information from various programs, the visualization tools permit the construction of insightful reports and dashboards, while data modeling empowers combining, preparing, and linking information from multiple origins for analysis. This three-pronged design allows for extracting meaning from vast pools of disparate information.

Due to its seamless insight distribution across an organization and tight integration with Microsoft items and Office 365, Power BI emerges as a leading business analytical solution. It delivers information refreshes in real-time, guaranteeing customers constantly obtain the latest information promptly in dashboards and reviews, in contrast to different analytical platforms that sporadically encounter facts latency issues. This capacity stems from Power Query's speedy facts modeling function.

Power BI moreover incorporates robust cooperative and circulation instruments, permitting the protected circulation of dashboards and reviews enterprise-wide or with allocated person

classifications. Teams can team up on analytics by exploiting the integrated cooperative choices. Customers value the capacity to effortlessly share penetrating understandings among associates, divisions, and stakeholders. The tight symbiosis with commonplace Microsoft tools streamlines adoption and improves utility. Collectively, these variables positioned Power BI at the vanguard of analytical answers, empowering timely but protected facts sharing and teamwork.

Power BI transforms disorganized data into insightful understandings for decision-making by integrating sophisticated artificial intelligence and cutting-edge analytics capabilities into the platform. Additionally, it provides first-rate data management through detailed security and data policies, ensuring that the proper restrictions are in place while allowing for independent analytical review.

The responsive mobile applications fittingly change reports and visualizations to match any exhibit measurement. Pre-fabricated templates speed up a report and dashboard growth for assorted facts sources and situations. Lastly, Power BI is hugely flexible due to the large selection of third-party expansion and integration choices.

Unstructured playback is mined for insights using cutting-edge AI and advanced analytics methods. While promoting independent exploration, information management capabilities ensure the correct access controls. Apps for mobile devices that work with any device give penetration. Using ready-made templates simplifies creating visualizations. The program's social aspect, which uses API and partner delivery, most dramatically increases its features and cost for every usage scenario. These characteristics maximize the perceived price that Power BI suggests individually.

Power BI is a powerful business intelligence and analytics tool that hastens the analysis of data, the creation of reports, and the disclosure of strategic insight. It allows businesses to continuously maintain a full, real-time view of their most important performance indicators by obtaining and analyzing vast amounts of data from several sources. This enables businesses to gather and analyze crucial data, make smarter decisions, and provide better customer service. The tool provides comprehensive dashboards and analytics that track progress fast. Its comprehensive strategy for gathering and analyzing performance data from various systems promotes well-informed strategic planning and problem-solving. Power BI provides businesses with the deep information needed to increase customer happiness and hone priority areas.

Possible future developments

The implemented project, with its multifaceted implications and strategic foresight, has endowed the customer with a unique vantage point to meticulously evaluate and deliberate upon a myriad of expansion options. This visionary approach not only unlocks the door to immediate growth but also lays the groundwork for a future imbued with innovation and transformative developments. The project, in essence, serves as a catalyst for the emergence of uncharted avenues that beckon exploration and present a canvas upon which the brushstrokes of future developments can be painted.

In the ongoing collaborative discourse with the esteemed company Delta, there exists a dynamic exploration of the potential integration of the existing project structure with an avant-garde phase of generative artificial intelligence (AI). This avant-garde initiative, a testament to the commitment to cutting-edge technological advancements, seeks to infuse the data platform with a transformative layer of intelligence. The focal point of this endeavor is to specifically target and elevate the capabilities of the pharmaceutical domain, envisioning a seamless integration of generative AI through the deployment of an innovative chatbot system.

The conceptualization of an AI-powered chatbot marks a paradigm shift in how customers interact within the pharmaceutical sphere. Beyond its conventional role as a conduit for information, the chatbot emerges as an intelligent interface that transcends transactional engagements. It not only serves as a repository of detailed information on pharmaceutical products but also introduces a dynamic dimension by providing real-time updates on product availability. This nuanced integration is designed to render the customer experience more tangible, intuitive, and holistic, thereby reshaping the narrative of consumer-pharmacy interactions.

The envisioned chatbot functionality is a testament to the commitment to empowerment, offering users a robust platform to make informed decisions regarding pharmaceutical purchases. By facilitating inquiries into the specifics of products and providing real-time updates on stock availability, the AI-driven chatbot transcends the traditional retail paradigm. It transforms the customer experience into a dynamic, information-rich interaction, positioning the pharmacy as a knowledgeable and responsive entity rather than a mere dispenser of products.

As the integration of generative AI unfolds in tandem with the project structure, the potential applications of the data platform expand exponentially. While the immediate focus is on the pharmaceutical sector and the transformative chatbot functionality, the underlying infrastructure remains inherently versatile and open to accommodating a spectrum of diverse

use cases. This deliberate openness is not merely a technical consideration but a strategic decision aimed at future-proofing the technology against the ever-evolving demands and opportunities that lie on the horizon for the company Delta.

In essence, the data platform's receptivity to new use cases serves as a deliberate and strategic decision to future-proof the technology. This forward-thinking approach not only acknowledges the fluidity of technological landscapes but also anticipates the evolving demands of the market. By fostering an environment of continuous innovation, the company Delta positions itself as a trailblazer in leveraging data-driven solutions for not only current challenges but also those that may emerge in the ever-evolving business landscape.

The ongoing collaboration between the project and the esteemed company Delta encapsulates a commitment to remaining at the forefront of technological innovation. The exploration of generative AI integration is not merely a singular initiative; it embodies a broader philosophy of proactively seeking opportunities to enhance and expand operational capacities. This ethos is grounded in the understanding that technology is not a static entity but a dynamic and evolving force, and companies that embrace continuous adaptation are better poised to thrive in the competitive business landscape.

As the integration with generative AI progresses, a symbiotic relationship emerges between technological advancement and customer-centricity. The envisioned chatbot, with its capacity to provide detailed information on pharmaceutical products and real-time stock updates, is emblematic of a customer-centric approach. It transcends the traditional transactional model by placing information and empowerment at the forefront of the customer experience, thus fostering a sense of agency and informed decision-making.

Moreover, the incorporation of generative AI extends beyond immediate customer interactions to encompass a strategic understanding of market dynamics. The data platform, now equipped with enhanced capabilities, becomes a valuable tool for gathering insights into customer preferences, market trends, and product demand. This wealth of information not only informs immediate decision-making but also positions the company Delta as a data-driven entity capable of navigating and shaping the future of the pharmaceutical retail landscape.

The ongoing project, with its visionary integration of generative AI and collaborative partnership with the forward-thinking company Delta, represents a leap into the future of technological integration and business expansion. The deliberate consideration of generative AI integration reflects a commitment to not only meeting current needs but also anticipating and addressing future challenges. The versatility of the data platform, coupled with its openness to new use cases, positions it as a dynamic and adaptive solution that aligns seamlessly with the strategic objectives of the company Delta. As the project unfolds, it serves as a testament to the

company's resilience, innovation, and unwavering commitment to delivering an unparalleled customer experience in the ever-evolving landscape of the pharmaceutical industry.

Conclusion

The project in question has undoubtedly played a pivotal role in enabling Delta, a distinguished and forward-thinking organization, to unlock the true value that had long been concealed within the depths of their expansive and diverse data collection. Through the incisive and meticulous undertaking of a comprehensive analysis of the current state of affairs, the highly skilled and dedicated project team successfully acquired a wealth of invaluable insights pertaining to the most effective intervention strategies and the necessary tools to facilitate the ensuing processes. It is important to emphasize that the project transcended the traditional role of a mere service provider, evolving into a trusted and reliable partner in Delta's ardent pursuit of a data-driven decision-making approach that lies at the very core of their corporate ethos.

The paramount significance of this project becomes even more apparent when we consider the meticulous process of data collection that was undertaken. A vast amount of information was meticulously gathered through this process, serving as the foundation for the subsequent stages of analysis and extraction. This voluminous dataset underwent a thorough and extensive architectural rework, ensuring that it was structured in a manner that would facilitate the extraction of only the most pertinent and meaningful insights. The utilization of query tables played a crucial role in this endeavor, enabling the project team to refine and distill the data to its essence. Ultimately, these invaluable insights were seamlessly transmitted to the Power BI dashboards, which emerged as a powerful instrument wielded by Delta's top management for comprehensive analysis and informed decision-making.

It is worth acknowledging the profound transformation that this project has brought about, particularly when juxtaposed with the limitations imposed by the previous data structure, which hindered the extraction of crucial information by the top management. The newly established architecture ushered in a new era of decipherment and analysis, empowering Delta's decision-makers with a comprehensive understanding of the company's operations. This newfound clarity and insight have undoubtedly propelled Delta towards optimal decision-making, facilitating the formulation and implementation of strategies that align with the organization's overarching goals and objectives.

The Power BI dashboards, intricately linked to the data collection structure expounded upon in earlier chapters, operate on a foundation of real-time data updates. This ensures the accuracy

and currency of the information analyzed within Power BI, providing Delta's top management with a reliable and up-to-date foundation for their decision-making processes. The ability to conduct real-time analyses empowers Delta's leadership to make well-informed and timely decisions, propelling the organization ever forward on its path to success.

Moreover, it is essential to recognize that this project has laid the groundwork for future advancements and possibilities. It has set the stage for the creation of new dashboards, expanding the scope of insights that can be derived from Delta's data. Furthermore, the integration of generative artificial intelligence systems, such as bots, holds immense potential to revolutionize customer interactions within the pharmacy industry. These intelligent bots can serve as intermediaries, facilitating direct and personalized connections between pharmacy customers and the pharmacies themselves. By enhancing customer experiences and fostering stronger relationships, this integration has the capacity to reshape the landscape of customer engagement within the industry.

In conclusion, the project at hand has been nothing short of instrumental in enabling Delta to unlock the true potential concealed within their data. Through the implementation of a robust data collection structure and harnessing the power of Power BI dashboards, Delta's top management has gained access to timely, reliable, and actionable insights. The impact of this project extends far beyond the present, as it has not only empowered Delta's decision-makers to make informed choices but has also laid a strong foundation for future developments and advancements. By embracing a data-driven decision-making approach and prioritizing customer engagement, Delta finds itself at the forefront of innovation and excellence within the pharmacy industry.

Bibliographical references

JULIA KOKINA, DESSISLAVA PACHAMANOVA, ANDREW CORBETT, March 2017, *The role of data visualization and analytics in performance management: Guiding entrepreneurial growth decisions*, Journal of Accounting Education. Available on: https://www.sciencedirect.com/science/article/abs/pii/ S0748575116301221

PUBLISHER, 2016, *PwC in Italy*, Document of presentation. Available on: https://www.ramspa.it/sites/default/files/allegati/ cv_pwc.pdf

PUBLISHER, *PWC Consulting*, Forbes Italia. Available on: https://forbes.it/eccellenza/pwc-consulting/

PUBLISHER, *Features of Dynamics 365 Business Central*, Microsoft. Available on: https://dynamics.microsoft.com/it-it/business-central/capabilities/

PUBLISHER, *Microsoft Dynamics 365 Business Central Documentation*, Microsoft. Available on: https://learn.microsoft.com/it-it/dynamics365/business-central/

MARIJE BRUMMEL, DAVID STUDEBAKER, CHRIS STUDEBAKER, 2019, *Programming Microsoft Dynamics 365 Business Central Six Edition*.

IN LIH ONG, PEI HWA SIEW, SIEW FAN WONG, 2011, *A Five-Layered Business Intelligence Architecture*. Available on: https://ibimapublishing.com/articles/CIBIMA/2011/695619/695619.pdf

MOHAMMAD SHARIAT, ROSCOE HIGHTOWER, JR, *Conceptualizing Business Intelligence Architecture*. Available on: https://www.researchgate.net/profile/Roscoe-Hightower/publication/237013221_Conceptualizing_Business_Intelligence_Architecture/links/0046351ec9d65a42a0000000/Conceptualizing-Business-Intelligence-Architecture.pdf

MATTHEW G. DOYLE, 28th September 2005, *An IP Address Management Solution for a Server Solution Provider*. Available on: https://zaichik.org/thesis/thesis.pdf

IGNACIO BERMUDEZ, STEFANO TRAVERSO, MARCO MELLIA, MAURIZIO MUNAFO', *Exploring the Cloud from Passive Measurements: the Amazon AWS Case*. Available on: https://www.ict-mplane.eu/sites/default/files/public/publications/303amazonfw.pdf

PUBLISHER, *AWS*. Available on: https://www.geekandjob.com/wiki/aws

PUBLISHER, *Amazon AWS: what it is and why use it in your business*. Available on: https://www.ntditalia.com/amazon-aws-cos-e/

PUBLISHER, *Build your future in the AWS Cloud*. Available on: https://www.aws.training/

MUHAMMAD AYOUB KAMAL, HAFIZ WAHAB RAZA, MUHAMMAD MANSOOR ALAM, MAZLIHAM MOHD SU'UD, January 2020, *Highlight the Features of AWS, GCP and Microsoft Azure that Have an Impact when Choosing a Cloud Service Provider*. Available on: https://www.researchgate.net/profile/Muhammad-Ayoub-Kamal/publication/340173446_Highlight_the_Features_of_AWS_GCP_and_Microsoft_Azure_that_Have_an_Impact_when_Choosing_a_Cloud_Service_Provider/links/5e7c397c92851caef49d994f/Highlight-the-Features-of-AWS-GCP-and-Microsoft-Azure-that-Have-an-Impact-when-Choosing-a-Cloud-Service-Provider.pdf

DAVID CHAPPELL, May 2009, *Introducing the Azure services platform*. Available on: http://www.chappellassoc.com/Azure_Services_Platform_v1.1--Chappell.pdf

ROBERT GYŐRÖDI, MARIUS IULIAN PAVEL, CORNELIA GYŐRÖDI, AND DOINA ZMARANDA, 13th December 2018, *Performance of OnPrem Versus Azure SQL Server: A Case Study*. Available on: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8618405

DAVID CHAPPELL, October 2010, *Introducing the Windows Azure platform*. Available on: http://www.davidchappell.com/writing/white_papers/Introducing_the_Windows_Azure_Platform,_v1.4--Chappell.pdf

JOUNI MÄENPÄÄ, 27th April 2009, *Cloud computing with the Azure platform*. Available on: http://www.cse.tkk.fi/en/publications/B/5/papers/Maenpaa_final.pdf

EDITORIAL BOARD, 19th November 2020, *What is Microsoft Azure and what is it for?*, Netinformatica. Available on: https://www.net-informatica.it/microsoft-azure-e-a-cosa-serve/

PUBLISHER, *Endless possibilities, from edge device to the cloud*. Available on: https://azure.microsoft.com/it-it

TABISH MUFTI, POOJA MITTAL, BULBUL GUPTA, 27th February 2020, *A Review on Amazon Web Service (AWS), Microsoft Azure & Google Cloud Platform (GCP) Services*. Available on: https://eudl.eu/pdf/10.4108/eai.27-2-2020.2303255

RICKY K. P. MOK, HONGYU ZOU, RUI YANG, TOM KOCH, ETHAN KATZ-BASSETT, KC CLAFFY, 2nd November 2021, *Measuring the network performance of Google Cloud Platform*. Available on: https://dl.acm.org/doi/pdf/10.1145/3487552.3487862

PUBLISHER, GCP. Available on: https://www.geekandjob.com/wiki/gcp

PUBLISHER, *What is a cloud service provider?*. Available on: https://cloud.google.com/learn/what-is-a-cloud-service-provider

MICHAEL J. DENNEY, DUSTIN M. LONG, MATTHEW G. ARMISTEAD, JAMIE L. ANDERSON, BAQIYYAH N. CONWAY, 29th July 2016, *Validating the Extract, Transform, Load Process Used to Populate a Large Clinical Research Database*. Available on: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5556907/

PUBLISHER, *What is an ETL?*, Talend. Available on: https://www.talend.com/resources/what-is-etl/

PUBLISHER, *ETL (Extract, Transform, Load)*, IBM. Available on: https://www.ibm.com/topics/etl

PUBLISHER, *What is ETL (Extract Transform Load)?*, AWS. Available on: https://aws.amazon.com/what-is/etl/?nc1=h_ls

MATTHIAS BRANTNER, DANIELA FLORESCU, DAVID GRAF, DONALD KOSSMANN, TIM KRASKA, *Building a Database on S3*. Available on: https://d1wqtxts1xzle7.cloudfront.net/43108802/donald-kossman1-libre.pdf?1456516584=&response-content-disposition=inline%3B+filename%3DBuilding_a_database_on_S3.pdf&Expires=1698451514&Signature=B~4B5WgAStwWV7BdhTwjkwahWo1r66t3Ha0jGCnlS7DNaF3bWAMZVCmiilnCF6udJvAzB20axMWOHd0dzPhtOmBm1QfGqRPzfTvxQOqt0CgodgtDixUALWePR2r9JzUw70JPeKQ1apwuUpAGLFdJc4OEc9shuzQJdgmnnKsGOx8oOSzWyY3VIddYRxs3o2VYGokQdoNztbP05M0Y-ROOFcsunwIDiD2NVRODM3apmka6bDDrOYCOm8sAnRYNw-D8NX~b2EhQiMWaEbLvOYs16KtgNvOq~IrpOifPyzqR6NDgwpPe~E4nL-2ujFROySOXekdcyCqfXLqVXjtRC9aJ7w__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

PUBLISHER, *Amazon S3*. Available on: https://aws.amazon.com/s3/?nc1=h_ls

PUBLISHER, *What is Amazon S3?* Available on: https://docs.aws.amazon.com/AmazonS3/latest/userguide/Welcome.html

ANURAG GUPTA, DEEPAK AGARWAL, DEREK TAN, JAKUB KULESZA, RAHUL PATHAK, STEFANO STEFANI, VIDHYA SRINIVASAN, 31 May 2015, *Amazon Redshift and the Case for Simpler Data Warehouses*. Available on: https://dl.acm.org/doi/pdf/10.1145/2723372.2742795

NIKOS ARMENATZOGLOU, SANUJ BASU, NAGA BHANOORI, MENGCHU CAI, NARESH CHAINANI, KIRAN CHINTA, VENKATRAMAN GOVINDARAJU, TODD J. GREEN, MONISH GUPTA, SEBASTIAN HILLIG, ERIC HOTINGER, YAN LESHINSKY, JINTIAN LIANG, MICHAEL MCCREEDY, FABIAN NAGEL, IPPOKRATIS PANDIS, PANOS PARCHAS, RAHUL PATHAK, ORESTIS POLYCHRONIOU, FOYZUR RAHMAN, GAURAV SAXENA, GOKUL SOUNDARARAJAN, SRIRAM SUBRAMANIAN, DOUG TERRY, 12 June 2022, *Amazon Redshift Re-invented*. Available on: https://dl.acm.org/doi/pdf/10.1145/3514221.3526045

PEDRO JOEL FERREIRA, ANA ALMEIDA, JORGE BERNARDINO, *Data Warehousing in the Cloud: Amazon Redshift vs Microsoft Azure SQL*. Available on: https://www.scitepress.org/papers/2017/65871/65871.pdf

PRASHANT DUTTA, January 2019, *Business Analytics using Microsoft Power BI and AWS Redshift*. Available on: https://d1wqtxts1xzle7.cloudfront.net/59350543/187_Business_Analytics_using_Microsoft_Power_BI_and_AWS_Redshift20190521-59265-96mtte-libre.pdf?1558507186=&response-content-disposition=inline%3B+filename%3DBusiness_Analytics_using_Microsoft_Power.pdf&Expires=1698453538&Signature=P6OyrNt15NzNTRB4L5Ai2ewsMDyJQre~H1EAEXpw6ew1p1s2qCNrOvdUDMcuurpoRnxGG0bHNWouArT9t9N~iDzbuSa7kR7BbOYyMpDWuT63i0ajzyXpxFDc0MFVop9TYgAV1X1pU6X6KkXGYFKIH7abWhZwgxh7SKnv18thtxWkqItGIcp8GIj-p16wE-Bfs0za8ncMjjkPkBuD~6veSNbJnZ5UewuCIph74AVFEagpUmspuo~cp-fU3vurnH5vJi8QCSQEllpuaU1TInu4UdNcVS9q37pwPDXRnB4qA~nCwbrsXOrmuYnYcvdd4MWSWLRcyVmdKFWAaufLPVczFQ__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

PUBLISHER, *Amazon Redshift*. Available on: https://aws.amazon.com/redshift/

PUBLISHER, *What is Redshift?* Available on: https://www.snaplogic.com/it/glossary/aws-redshift

MOHIT SAXENA, BENJAMIN SOWELL, DAIYAN ALAMGIR, NITIN BAHADUR, BIJAY BISHT, SANTOSH CHANDRACHOOD, CHITTI KESWANI, G2 KRISHNAMOORTHY, AUSTIN LEE, BOHOU LI, ZACH MITCHELL, VAIBHAV PoRWAL, MAHEEDHAR REDDY CHAPPIDI, BRIAN ROSS, NORITAKA SEKIYAMA, OMER ZAKI, LINCHI ZHANG, MEHUL A. SHAH, *The Story of AWS Glue*. Available on:

https://assets.amazon.science/df/4f/e4acdd234c8296314342e9da4061/the-story-of-aws-glue.pdf

PUBLISHER, *Amazon Glue*. Available on: https://aws.amazon.com/glue/?nc1=h_ls