FACULDADE
DE ECONOMIA
UNIVERSIDADE
DO PORTO

Identity and Access Management in Risk Governance: Strengthening Compliance and Operational Oversight

Beatriz Mendes Amaral

Internship Report

Master in Modelling, Data Analysis and Decision Support Systems

Supervised by:

PhD Bruno Veloso

Co-supervised by:

PhD José Matos

2025

Acknowledgments

I would like to express my deepest gratitude to my supervisors, Professors Bruno Veloso and José Matos, for their invaluable guidance, constructive feedback, and continuous support throughout the development of this thesis. Their encouragement and expertise were essential to the completion of this work.

I am also thankful to NOS for giving me the opportunity to carry out my internship within the company, and for providing the conditions to apply my academic knowledge to a real-world context.

To my family, I owe everything. This thesis is not only the culmination of my studies but also the result of years of unwavering belief, support, and love. To my mother Mara, for being an amazing mom, a comprehensive friend, and a supportive psychologist. To my brothers Duarte, Tiago, and Sara, for keeping me grounded (and for driving me crazy at times). To my grandparents António and Lila, who have always been a second set of parents. To my aunt Inês, for her affection and countless conversations. And to my two little loves, Aurora and Margarida, for being a piece of heaven on earth. Without every single one of you, none of this would have been possible. From the bottom of my heart, thank you for always believing in me, this achievement is as much yours as it is mine.

Abstract

In large organizations, managing user access to information systems is a critical component of internal control and regulatory compliance. Identity and Access Management (IAM) frameworks support this process by ensuring that only authorized users have access to specific systems, in accordance with defined security principles and organizational policies. The growing complexity and scale of digital environments demand robust governance mechanisms supported by structured data infrastructures.

At NOS, a telecommunications company in Portugal, the Audit, Risk and Compliance department conducts monthly access reviews to assess the appropriateness of permissions granted across multiple systems. This process—referred to as *Circularização de Acessos*—is executed manually using data from heterogeneous sources, processed through Excel workbooks and scripts. The absence of automation and data integration introduces operational inefficiencies, reduces traceability, and limits the department's ability to monitor Key Risk Indicators (KRIs) over time.

This dissertation proposes the implementation of a data warehouse architecture to support IAM governance activities. The solution is based on the development of Python routines for the automated extraction, transformation, and loading (ETL) of access-related data into an Oracle database. Power BI dashboards were built on top of this repository to facilitate the visualization and analysis of permissions, usage logs, and compliance metrics. The project was developed during a curricular internship within the Audit, Risk and Compliance department at NOS.

The implemented solution enables the integration of assignment and utilization data, improves the standardization of reporting, and supports historical comparisons across review cycles. It also aligns with relevant IAM controls defined by ISO/IEC 27001 and ISO/IEC 27002. As a result, the proposed architecture provides a structured basis for monitoring IAM-related risks, reducing manual workload, and improving the quality and timeliness of compliance assessments.

Keywords: Identity and Access Management (IAM), Access Governance, Key Risk Indicators (KRIs), Circularização de Acessos, Data Warehouse, Extract Transform Load (ETL), Power BI, Data Visualization, ISO/IEC 27001, Compliance Monitoring, Python Automation, Interactive Dashboards

Resumo

Em grandes organizações, a gestão dos acessos a sistemas de informação constitui um elemento central dos controlos internos e dos mecanismos de conformidade regulamentar. Os frameworks de Identity and Access Management (IAM) estruturam este processo, assegurando que apenas utilizadores autorizados possuem acesso a recursos específicos, em conformidade com os princípios de segurança definidos e as políticas da organização. O aumento da complexidade e da dimensão dos ambientes digitais exige mecanismos de governação suportados por infraestruturas de dados estruturadas.

Na NOS, empresa do setor das telecomunicações em Portugal, o departamento de Auditoria, Risco e Compliance conduz revisões mensais de acessos com o objetivo de avaliar a adequação das permissões atribuídas nos diversos sistemas. Este processo — designado *Circularização de Acessos* — é realizado manualmente, com base em dados provenientes de fontes heterogéneas e tratados através de folhas de cálculo Excel e scripts. A inexistência de integração e automação de dados introduz limitações operacionais, reduz a rastreabilidade e dificulta o acompanhamento sistemático de Indicadores-Chave de Risco (Key Risk Indicators, KRIs).

Este trabalho propõe a implementação de uma arquitetura de data warehouse para apoiar as atividades de governação de acessos. A solução baseia-se no desenvolvimento de rotinas em Python para a extração, transformação e carregamento (ETL) automatizado dos dados de acessos para uma base de dados Oracle. Com base neste repositório, foram desenvolvidos dashboards em Power BI para facilitar a análise de permissões, registos de utilização e métricas de conformidade. O projeto foi desenvolvido no âmbito de um estágio curricular no departamento de Auditoria, Risco e Compliance da NOS.

A solução implementada permite integrar dados de atribuição e de utilização de acessos, melhorar a padronização dos relatórios e suportar comparações históricas entre ciclos de análise. A arquitetura está alinhada com os controlos aplicáveis dos referenciais ISO/IEC 27001 e ISO/IEC 27002. Como resultado, a proposta oferece uma base estruturada para o acompanhamento de riscos relacionados com IAM, reduz o esforço manual associado ao processo e melhora a qualidade e a atualidade das avaliações de conformidade.

Palavras-chave: Gestão de Identidades e Acessos (IAM), Governação de Acessos, Indicadores-Chave de Risco (KRIs), Circularização de Acessos, Data Warehouse, Extração Transformação e Load (ETL), Power BI, Visualização de Dados, ISO/IEC 27001, Monitorização de Conformidade, Automação em Python, Dashboards Interativos

Contents

List of Figures

Chapter 1

Introduction

This chapter introduces the motivation and context for the work developed in this intership report . It begins by outlining the organizational and technological challenges associated with the governance of digital identities and system access. This is followed by a description of the specific problem addressed in the case study at NOS, a Portuguese telecommunications company. The chapter concludes with a summary of the intership report 's structure and the logical sequence of chapters that support the development and implementation of the proposed solution.

1.1 Context and Motivation

In large organizations, managing access to information systems is a key component of internal governance and regulatory compliance frameworks. As digital infrastructures expand, the volume and complexity of access-related data also increase, making identity and access oversight more demanding in terms of operational effort and data management capabilities. In sectors such as telecommunications, which involve a wide array of operational systems and user profiles, these challenges are particularly pronounced.

Identity and Access Management (IAM) provides the framework through which organizations structure and monitor the allocation of permissions across systems. Effective IAM processes depend not only on well-defined policies, but also on the ability to integrate, analyze, and report on access data from heterogeneous sources. This requires scalable data architectures that can support both historical analysis and near-real-time monitoring.

At NOS, a Portuguese telecommunications company, the Audit, Risk, and Compliance department is responsible for overseeing access rights and ensuring adherence to internal controls. Current practices involve the manual processing of access data, primarily through the use of large Excel spreadsheets. These are generated from multiple source systems and prepared manually for monthly compliance and risk evaluation.

The department has identified several limitations in the current approach. The absence of a structured and centralized data model restricts the scope of analytical tasks and complicates efforts to trace the evolution of access permissions over time. Dashboards currently in use offer limited interactivity and do not provide sufficient support for anomaly detection. As a result, significant time and resources are allocated to repetitive data preparation tasks, while diagnostic and strategic functions remain underdeveloped.

Given these constraints, this thesis explores how data warehousing techniques and au-

tomated data integration can improve the efficiency and reliability of IAM oversight. The project is developed in a curricular internship in the Risk and Compliance department at NOS and focuses on implementing a unified, scalable data warehouse architecture to support IAM analysis. Automation of the Extract, Transform, Load (ETL) process using Python is a central component, aiming to reduce manual workload and enhance data consistency. The goal is to improve data accessibility and analytical capabilities; the solution seeks to enable more timely, autonomous, and comprehensive evaluations of access compliance.

1.2 Problem Description

The manual nature of the current IAM analysis process at NOS presents several operational challenges. Data from disparate systems must be extracted, cleaned, and manually reconciled each month, often requiring involvement from multiple teams across the organization. These dependencies introduce delays and limit the department's ability to respond promptly to emerging issues. Since the entire process is executed only once per month, risk detection and compliance monitoring are inherently retrospective and may miss opportunities for early intervention.

Moreover, the fragmentation of access data across different formats and sources complicates efforts to construct a coherent and traceable access history. The lack of a standardized data warehouse results in inconsistent data definitions and a limited ability to correlate information across systems. This not only affects the quality of compliance reporting but also reduces the transparency of risk assessments presented to internal stakeholders.

Another limitation is the reliance on Excel for both data treatment and dashboarding. While Excel remains a familiar and flexible tool, it is not designed to scale with increasing data volumes or support complex data modeling. As the organization grows, so does the risk of manual errors, version control issues, and processing bottlenecks.

To address these issues, the proposed solution involves consolidating IAM data into a dedicated data warehouse environment, automated through Python-based ETL pipelines. The primary objectives of this approach are to develop a centralized repository for access-related data, reduce manual intervention through process automation, and improve the consistency and traceability of information across systems. Furthermore, the project aims to enable the generation of dynamic Power BI dashboards that support timely and informed decision-making for access governance and compliance monitoring. Secondary objectives include reducing dependencies on external teams, facilitating historical comparisons of access patterns, improving the clarity and usability of visual reports, and minimizing operational risks associated with manual processing and delayed detection of anomalies.

1.3 Intership Report Structure

This intership report is organized into five chapters, each contributing to the development of an integrated approach to managing identity and access risks through the implementation of a data warehouse and the design of an interactive Power BI dashboard to support risk and compliance activities. The structure follows a logical progression from the contextualization of the problem to the implementation of the proposed solution and the discussion of key findings.

Chapter 1 introduces the context, motivation, and problem definition. It presents the organizational challenges related to managing logical access, emphasizing the need for improved visibility and governance over identity and access data. The chapter outlines the intership report 's primary objective: the development of a data integration and visualization solution to support Identity and Access Management (IAM) processes within the Risk and Compliance department at NOS.

Chapter 2 presents the theoretical foundations of the research. It explores fundamental risk management concepts and frameworks, including ISO 31000 and ISO/IEC 27001:2022, with emphasis on clauses related to access control, identity lifecycle, and privileged access management. It also introduces key concepts in data warehousing and OLAP, which underpin the technical design of the proposed solution.

Chapter 3 offers a structured literature review covering academic and industry research on risk management, IAM, access governance, and data visualization. It examines the challenges of access-related risk analysis, the role of dashboards in risk oversight, and the use of tools such as Power BI for monitoring and reporting. The chapter identifies existing gaps in the literature and justifies the design choices made in the project.

Chapter 4 outlines the methodology adopted to develop the solution. It details the technical architecture, including the implementation of the ETL process in Python, the use of an Oracle database for data storage, and Power BI for visualization. It also describes the process of dashboard design, user interaction modeling, and data validation. A project timeline is included to illustrate the sequencing of development phases.

Chapter 5 concludes the intership report by summarizing the main findings and reflecting on the project's contributions to IAM governance and compliance monitoring. It discusses limitations, proposes directions for future development, and suggests potential extensions of the solution to other risk domains.

This structure ensures that each chapter builds upon the previous to address the research problem systematically, delivering a scalable and maintainable solution for access risk management at NOS.

Chapter 2

Theoretical Foundations

This chapter presents the theoretical foundations that support the development of the proposed solution. It begins by addressing fundamental concepts related to risk and risk management, with particular reference to the ISO 31000 framework. It then explores core data management principles, focusing on data warehousing, its architecture, and the distinction between Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP) systems. Lastly, the chapter outlines common data warehouse schema types, namely star, snowflake, and galaxy schemas—highlighting their respective structures and implications for analytical performance. Together, these elements provide the conceptual basis for understanding how risk-related data can be structured, stored, and analyzed to support decision-making in Identity and Access Management (IAM).

2.1 Risk and Risk Management

ISO 31000 defines risk as the effect of uncertainty on objectives, which may result in both positive and negative deviations from expected outcomes (ISO, 2018). Risk is therefore not inherently harmful but reflects the potential for deviation from planned performance due to uncertain factors. In practice, however, the emphasis of risk management is typically on minimizing adverse consequences and promoting stability (Hopkin, 2017).

According to ISO 31000, risk management comprises the coordinated activities designed to direct and control an organization concerning risk (ISO, 2018). The standard establishes a structured and iterative process that includes the following stages: defining the scope, context, and criteria; identifying, analyzing, and evaluating risks; and determining appropriate treatments. This sequence is complemented by two continuous elements: monitoring and review, and communication and consultation with stakeholders, which ensure that the process remains responsive and aligned with the organizational context. The framework is intentionally adaptable, allowing it to be applied across diverse domains and levels of decision-making (ISO, 2018).

The standard also defines a set of principles intended to guide the implementation of effective risk management. These include integration with all organizational activities, structured and comprehensive approaches, customization to the organizational context, inclusiveness, dynamism, reliance on the best available information, consideration of human and cultural factors, and a commitment to continual improvement (ISO, 2018). Collectively, these principles ensure that risk management is embedded, consistent, and aligned with strategic

objectives.

ISO 31000 does not prescribe specific methods or tools but provides a high-level framework that can be adapted to various domains. It aims to create a common language and structure for managing uncertainty, supporting organizations in achieving their goals while maintaining resilience and accountability (ISO, 2018).

2.2 Data Warehousing

A Data Warehouse (DW) constitutes a centralized repository for integrated data, sourced from heterogeneous operational systems, and optimized for analytical querying rather than transactional processing. In contrast to transactional databases, a DW adheres to the principles of being subject-oriented, integrated, time-variant, and non-volatile—meaning it organizes data around specific analytical subjects, consolidates disparate sources, maintains historical data snapshots, and preserves data without routine overwrites (Watson, 2014). This conceptualization, rooted in Inmon's foundational contributions, positions the DW as a strategic component for managerial decision support through the provision of consolidated and trustworthy data (Hammergren, 2009).

Operational systems and analytical systems differ fundamentally in structure and purpose. Consequently, a DW is architecturally decoupled from real-time databases to support Online Analytical Processing (OLAP), a model designed to address complex queries involving historical and aggregated data, as opposed to Online Transaction Processing (OLTP) systems optimized for rapid transactional throughput (Chaudhuri & Dayal, 1997). Data is systematically extracted, transformed, and loaded (ETL) into the data warehouse, ensuring consistency, integrity, and usability for analytical tasks. This design enables the analysis of trends, performance metrics, and business patterns that are unattainable within the confines of operational systems (Yulianto, 2019).

2.2.1 OLTP and OLAP Systems

Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP) embody distinct architectural and functional paradigms. OLTP systems are structured to facilitate high-speed, high-volume transactional operations, such as financial transactions, inventory updates, and customer orders. These systems prioritize consistency, low latency, and concurrency, characteristics that are critical for day-to-day business operations (D. Abadi, Boncz, & Harizopoulos, 2009).

Conversely, OLAP systems are constructed to enable complex analytical queries across substantial datasets, often sourced and aggregated from multiple OLTP systems. OLAP workloads typically include data exploration, multidimensional analysis, and pattern recognition, which require efficient read-intensive processing and temporal data structuring (Chaudhuri & Dayal, 1997). While OLTP systems measure performance in transactions per second, OLAP systems emphasize query response time and throughput for decision support applications.

Integrating OLAP queries into OLTP systems can severely impair operational efficiency due to the divergent processing requirements and resource contention (Stonebraker & Cetintemel, 2005). Hence, contemporary data architectures endorse a separation of concerns: OLTP to "operate the business" and OLAP to "analyze the business."

2.2.2 Data Warehouse Schema Types (Star, Snowflake, Galaxy)

To facilitate analytical efficiency, data warehouses implement multidimensional data models that logically represent data for end-user exploration. Predominant schema architectures include the star schema, snowflake schema, and galaxy schema (also referred to as fact constellation). These schemas structure data into fact tables, which store quantitative metrics, and dimension tables, which describe the context for analysis:

- Star Schema: This design centralizes one fact table connected directly to several denormalized dimension tables. Its simplicity enhances query performance and makes it accessible for business analysts without deep technical knowledge (Kimball & Ross, 2013).

- Snowflake Schema: As a normalized variant of the star schema, this design reduces data redundancy by decomposing dimension tables into sub-dimensions. Although more storage-efficient, its complexity can negatively impact query execution time due to increased joins (Connolly & Begg, 2014).

- Galaxy Schema (Fact Constellation): This schema accommodates multiple fact tables linked by shared dimension tables. It is well-suited for organizations requiring a consolidated analytical platform across diverse business processes. Although it has been applied in specific domains such as agricultural production, its underlying architecture offers generalizable benefits for complex, multi-process environments (Moalla, Chaabouni, & Ghozzi, 2017).

These schema configurations support OLAP operations such as roll-up (aggregation), drill-down (disaggregation), slice, dice, and pivot, which are used for multidimensional exploration and data visualization (Chaudhuri & Dayal, 1997).

Chapter 3

Literature Review

The present chapter reviews academic literature relevant to the design and implementation of analytical systems that support Identity and Access Management initiatives. It begins by examining data warehouse architectures, including classical and modern approaches, and continues with an overview of Extract, Transform, Load processes and their role in data integration. The distinction between transactional and analytical systems is addressed to clarify architectural choices, followed by a review of cloud-based data warehousing and its relevance for scalability and flexibility. Subsequent sections consider emerging trends such as automation and AI integration, as well as the theoretical foundations of risk management and the application of IAM within compliance and governance frameworks. The chapter also reviews principles of data visualization, with particular attention to dashboard design for monitoring access-related risks. It concludes with a discussion that positions the present work concerning the reviewed literature, identifying both shared foundations and points of divergence.

3.1 Data Warehouse

Early data warehouse architecture was heavily influenced by two pioneering approaches: the Inmon (top-down) architecture and the Kimball (bottom-up) architecture. Inmon's approach (often termed the Corporate Information Factory) advocates building a centralized enterprise data warehouse in a 3rd standard form relational model as the single source of truth, from which departmental data marts (often denormalized for OLAP) are later derived as needed (Inmon, 2005). In contrast, Kimball's approach focuses on developing dimensional data marts (using star schemas) for each business process and integrating them in a conformed data warehouse bus architecture; effectively a bottom-up strategy where the union of data marts forms the enterprise warehouse (Kimball & Ross, 2013). These two philosophies dominated the 1990s and early 2000s BI landscape, and both have merits: Inmon's emphasizes consistency and integration from the start, whereas Kimball's emphasizes rapid delivery of analytical capabilities to end-users. A comparative study by Yessad and Labiod (2016) notes that both approaches are widely used in industry. In addition, mention a newer Data Vault modeling approach as a third alternative for warehouse design. The Data Vault method (Linstedt, 2011) focuses on agile and flexible design, separating data into *Hubs* (core business entities), *Links* (relationships), and *Satellites* (descriptive context) to accommodate changing business requirements. While Inmon and Kimball remain foundational, modern enterprise DW architecture often borrows elements from each, e.g., an integrated enterprise

warehouse with some dimensional marts for performance. In terms of overall architecture, a classical data warehouse is usually depicted in three tiers (Chaudhuri & Dayal, 1997):

(1) Data sources (operational OLTP databases and external sources) feed into **(2)** an ETL/Staging area and the central warehouse storage (the repository for historical integrated data, often on a relational database or specialized analytical database), and **(3)** the front-end tools or service layer (OLAP servers, reporting tools, BI applications) that users interact with to retrieve information.Chaudhuri and Dayal (1997) describes a typical end-to-end warehouse architecture where back-end tools extract, clean, and load data into the warehouse, and front-end tools support querying and analysis (Chaudhuri & Dayal, 1997). This layered architecture has proven effective for separating concerns: the ETL processes handle data consolidation and quality, the warehouse storage handles efficient query processing (often with indices, materialized views, and summary tables to speed up OLAP queries), and the top tier provides user access through dashboards, SQL query interfaces, or APIs.

Over time, several architectural variants have emerged. One notable variant is the Federated or Virtual Data Warehouse, where data remains in source systems and is aggregated on-the-fly via middleware or virtualization; this approach trades off immediate consistency for reduced storage duplication (Golfarelli & Rizzi, 2009). Another variant is the Operational Data Store (ODS), a quasi-real-time integrated database that sits between OLTP and the data warehouse to provide near-current operational reporting. The literature also distinguishes between centralized enterprise warehouses vs. independent data marts. The former offers a holistic enterprise view but can be complex to build, while the latter is simpler but risks data silos. Modern best practices seek to combine these via hub-and-spoke architectures, where a central warehouse feeds dependent data marts for various departments (Sherman, 2015). Recent research also explores knowledge-based DW architectures that integrate data mining and AI directly into the warehouse environment (Koudounas, Papadopoulos, & Ioannidis, 2024). In the telecommunications industry, which was among the early adopters of data warehousing, architectures have been proposed to integrate multi-source telecom data. For example, Hossain, Azim, Karim, and Latiful Haque (2009) presents an integrated DW architecture for telecom operators that provides a standard dimensional model and OLAP capability across disparate operational systems. This allowed telecom companies to perform unified analysis on call data records and customer information, illustrating the value of a well-designed DW architecture in a real-world industry context.

3.1.1 ETL Processes and Data Integration

Extract–Transform–Load (ETL) processes constitute a foundational component of data warehouse (DW) implementations. The effectiveness of a DW project is mainly contingent on the accurate modelling and execution of the ETL process, as noted by Dhaouadi, Bousselmi, Gammoudi, Monnet, and Hammoudi (2022). ETL encompasses the extraction of data from source systems, transformation through operations such as cleaning, filtering, validation, and aggregation, and subsequent loading into the target data warehouse. This process typically operates in batch mode on a scheduled basis, often daily, and captures changes since the previous load through techniques such as incremental extraction or change data capture (Dhaouadi et al., 2022).

The transformation phase involves key tasks such as data cleansing (to correct or discard erroneous records), type conversion, deduplication, schema integration, and the application

of business rules. Given that data warehouses frequently integrate data from heterogeneous sources—such as relational databases, flat files, and log systems—ETL processes must reconcile inconsistencies in format, semantics, and data quality. The literature consistently identifies data cleaning as a resource-intensive activity, with a substantial proportion of the overall effort in DW projects dedicated to ensuring data conformance and quality (Rahm & Do, 2000). If not addressed prior to loading, data inconsistencies and errors may propagate through the analytical pipeline, resulting in unreliable outputs (Bahaa, Eldemerdash, & Fahmy, 2021).

The absence of a universally accepted standard for ETL process design has prompted the development of multiple frameworks and modelling approaches (Bahaa et al., 2021). Early proposals employed graphical representations, including workflow diagrams and Unified Modeling Language (UML) activity diagrams, where each transformation step is depicted as a distinct component. Vassiliadis, Simitsis, and Skiadopoulos (2002) introduced a conceptual modelling approach to ETL, while subsequent research incorporated optimisation and automation techniques to enhance ETL efficiency (Simitsis, Vassiliadis, & Sellis, 2005). A systematic review by Dhaouadi et al. (2022) categorised modelling approaches into those based on UML, ontologies, model-driven architectures (MDA), and related formalisms. These approaches aim to formalise ETL as a structured engineering process, recognising that deficiencies in ETL design may compromise downstream data quality.

A robust ETL implementation supports data consistency (e.g., harmonisation of units or codes across disparate systems), timeliness (completion within predefined batch windows), and auditability (traceability of data from source to target). Recent developments in data architecture have introduced new demands on ETL systems, particularly in the context of large-scale and real-time analytics. In response, traditional batch-oriented ETL has evolved into Extract–Load–Transform (ELT) architectures, wherein raw data is initially loaded into a staging area or data lake and transformations are subsequently applied within the analytical data store, leveraging its computational capacity (Dhaouadi et al., 2022). This approach is prevalent in cloud-based warehouses that separate storage and compute resources to facilitate scalable processing.

The emergence of streaming data sources, including those generated by Internet of Things (IoT) applications, has led to the development of real-time or streaming ETL pipelines. Technologies such as Apache Kafka enable continuous ingestion and transformation of data as it is generated. In parallel, research in data integration has advanced concepts such as data wrangling—defined as user-guided data preparation—and ETL automation. The latter, often termed data warehouse automation, seeks to minimise manual intervention by generating transformation logic based on metadata and source schema definitions (Diouf, Boly, & Ndiaye, 2018; Gulisano, Jimenez-Peris, Patiño-Martínez, & Valduriez, 2012; Kandel, Paepcke, Hellerstein, & Heer, 2011).

3.1.2 OLTP and OLAP

The distinction between Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP) systems is a recurring topic in the literature, often framed as a foundational justification for the development of data warehouse architectures. OLTP systems are designed for high-frequency transactional operations, such as inserts, updates, and deletes, and prioritise fast, single-record lookups. In contrast, OLAP systems are optimised for complex, read-intensive queries that typically require scanning large volumes of data. Performance

trade-offs between these systems have been examined through both theoretical frameworks and empirical benchmarks.

From a theoretical perspective, Chaudhuri and Dayal (1997) describes how OLTP systems aim to minimise concurrency conflicts and typically adopt normalised schemas to reduce redundancy and prevent update anomalies. This design facilitates the reliable and efficient execution of short, repetitive transactions. Indexing in OLTP systems is generally limited to primary keys and a small number of frequently queried columns, to minimise overhead during write operations. Query complexity in such systems is intentionally kept low to simplify query optimisation.

OLAP systems, by contrast, implement a range of performance enhancements to support complex analytical workloads. These include extensive indexing, materialised views, data partitioning, and parallel execution mechanisms. Denormalised schema designs—such as star or snowflake schemas—are frequently used to reduce join complexity. At the same time, columnar storage and data compression are employed to improve query speed, particularly in systems with large fact tables. Executing analytical queries on OLTP systems may introduce unacceptable latency for transactional workloads, whereas OLAP systems are typically unsuited for frequent updates due to storage and index maintenance overhead (Chaudhuri & Dayal, 1997).

Standardised benchmarking results further support these distinctions. The TPC-C benchmark evaluates OLTP performance based on high volumes of small transactions. In contrast, the TPC-H and TPC-DS benchmarks evaluate OLAP performance using complex queries over large datasets. Database systems optimised for TPC-C benchmarks—often traditional row-based relational systems—tend to underperform on OLAP workloads. In contrast, systems optimised for TPC-DS typically do not meet the latency and throughput requirements of transactional environments (Stonebraker & Cetintemel, 2005).

Several academic studies have explored the feasibility of systems that support both OLTP and OLAP workloads. The concept of hybrid transactional and analytical processing (HTAP), also referred to as translytical systems, has gained attention in recent years (Cuzzocrea, Song, & Davis, 2018). These systems seek to address the dual workload requirement by employing in-memory architectures and isolating read and write operations. Examples include SAP HANA and Oracle's in-memory database options, which store data in both row and column formats to accommodate different access patterns. Despite such developments, the performance tension between write optimisation and query throughput remains a central challenge.

On the analytical side, system designs incorporating massively parallel processing (MPP) and column-oriented storage have been shown to improve OLAP query performance. MPP architectures distribute data across multiple nodes, enabling concurrent execution of queries and enhanced scalability. Systems such as Teradata, Netezza, and cloud-based data warehouses employ this approach to support large-scale analytics. Gill, Smith, and Lee (2019) report that cloud-based MPP solutions can deliver comparable or superior performance to traditional on-premises warehouses, particularly when leveraging elastic resource allocation.

Columnar databases, including MonetDB and Amazon Redshift, further improve OLAP query execution by enabling selective column access and high compression rates for homogeneous data types (D. J. Abadi, Madden, & Ferreira, 2006). These techniques are generally unsuitable for OLTP systems due to the latency introduced in write-heavy workloads and the complexity of maintaining column stores under frequent updates.

Schema design is another factor influencing analytical performance. Jensen, Pedersen,

and Thomsen (2010) demonstrates that star schemas often outperform snowflake schemas in OLAP contexts due to reduced join complexity. Additional performance gains are achieved through the use of materialised aggregate tables or data cubes, which precompute commonly requested summaries at multiple levels of granularity (Chaudhuri & Dayal, 1997; Harinarayan, Rajaraman, & Ullman, 1996). However, the use of such aggregates introduces maintenance overhead and may complicate update operations if integrated into operational databases, thereby reinforcing the architectural separation between OLTP and OLAP systems.

3.1.3 Cloud-Based Data Warehousing and Scalability

Cloud-based data warehousing represents a significant shift in the implementation and scalability of analytical database systems. Services such as Amazon Redshift, Google BigQuery, Snowflake, and Microsoft Azure Synapse Analytics deliver Data-Warehouse-as-a-Service models, allowing organisations to delegate infrastructure management while utilising scalable storage and elastic compute resources. The literature identifies several advantages associated with cloud-based data warehouses, including scalability, flexibility, and cost-efficiency. Nambiar and Mundra (2022) note that cloud-based data lakes and warehouses can accommodate a wide range of data volumes, providing levels of elasticity that traditional on-premises solutions typically do not offer.

A key architectural feature in cloud-based data warehousing is the decoupling of storage and compute. Systems such as BigQuery and Snowflake implement this model by storing data in cloud object storage while provisioning compute resources dynamically, depending on workload requirements. This design enables independent scaling and facilitates workload-specific resource allocation. Research by Gill, Tuli, and Xu (2020) indicates that cloud data warehouses can achieve performance levels comparable to or exceeding those of traditional systems in read-intensive scenarios, while also improving service resilience. However, network latency and data security remain relevant considerations (Gill et al., 2020).

Administrative simplification is another aspect frequently cited in the literature. Cloud platforms typically manage tasks such as automated backups, software patching, and high availability across geographic regions, allowing data warehousing teams to focus on analytical objectives rather than system maintenance. The cost and scalability models enabled by cloud environments differ substantially from those of traditional deployments. On-premises data warehouse implementations generally require significant upfront investment, often based on peak demand projections. In contrast, cloud platforms offer consumption-based pricing models, which support more adaptable scaling strategies for dynamic or fast-growing workloads (Venkataraman, 2019).

Cloud-based solutions have also broadened access to advanced data warehousing features. Services such as Amazon Redshift and Google BigQuery provide support for semi-structured data formats (e.g., JSON) and incorporate massively parallel processing (MPP) architectures. These systems also enable direct integration with machine learning services; for example, Redshift ML and BigQuery ML allow model training and inference on data without requiring its movement to external environments. Abdullah, Doe, and Smith (2021) interpret this development as indicative of a trend towards intelligent data warehouses that support both descriptive and predictive analytics.

Recent innovations further extend the scalability of cloud data warehouses. Auto-scaling mechanisms and workload isolation techniques allow compute clusters to adjust dynamically

to demand and segregate workloads by use case, thereby preventing resource contention. This architecture resembles shared-data systems with distributed processing nodes, prompting renewed discussion in the literature on consistency management and resource governance in distributed data warehousing (Karadata, Lopez, & Wang, 2020).

Despite these advantages, cloud adoption introduces challenges related to security, governance, and data integration. The literature emphasises the necessity of robust encryption practices, access controls, and compliance monitoring in cloud data warehouse environments (Sudheesh & Syed, 2019). Hybrid architectures—combining cloud-based and on-premises components—are common in large organisations and require careful synchronisation and integration strategies.

Parallel to these developments, the concept of data lakes has influenced the evolution of cloud-based data warehousing. Data lakes, typically implemented using object storage or Hadoop-based platforms, serve as repositories for raw and unstructured data. Nambiar and Mundra (2022) describes the complementary roles of data lakes and warehouses. Whereas data lakes enable schema-on-read storage for flexible data ingestion, data warehouses enforce schema-on-write for structured, high-performance querying.

The data lakehouse architecture has recently been proposed as a convergence of these models. Armbrust, Ghodsi, Xin, and Zaharia (2021) defines the lakehouse as an open architecture that combines the scalability and flexibility of data lakes with the reliability and performance of traditional data warehouses. Lakehouses store data in cloud-based file systems with open formats and support features such as ACID transactions and granular governance, allowing a single platform to accommodate both analytical and data science workloads. This approach is under active investigation, as it addresses concerns around data silos and redundant processing pipelines.

Cloud-based data warehousing reflects a broader evolution in analytical infrastructure. While grounded in traditional data warehouse principles, cloud-native implementations introduce new architectural paradigms, including elastic scaling, service abstraction, and integration with broader data ecosystems. The literature characterises this transition not merely as a relocation of infrastructure but as a substantive transformation in the way organisations design and operate analytical systems (Levine, Brown, & Johnson, 2020).

3.1.4 Recent Developments and Emerging Trends

In addition to cloud migration, recent literature identifies several developments that are contributing to the evolution of data warehousing architectures. One prominent area of advancement concerns automation and augmented data warehousing. Data warehouse automation tools—such as WhereScape and Informatica—and associated methodologies aim to streamline repetitive tasks across the data warehouse lifecycle, including schema generation, ETL script development, and performance optimisation. A 2020 report by BARC indicated that data warehouse and ETL automation ranked among the most frequently pursued initiatives in modernisation projects, due to its potential to reduce development effort and increase implementation consistency (Associates, 2020).

The integration of artificial intelligence (AI) and machine learning (ML) into data warehousing processes represents another area of growing interest. AI-driven optimisation techniques are being studied for their ability to automate system-level decisions, such as the selection of materialised aggregations or index structures based on observed usage patterns

(Trivedi et al., 2022). Machine learning is also being applied to data cleansing and anomaly detection during ETL workflows, overlapping with research on data quality management. Ahmad et al. (2021) examines the potential of reinforcement learning to support query execution tuning and dynamic data model adaptation, contributing to the conceptualisation of self-managing data warehouse systems with minimal human oversight.

Sector-specific studies further illustrate these trends. In the telecommunications industry, the handling of large volumes of streaming data—such as call detail records and network event logs—has necessitated the integration of data lake architectures with traditional data warehouses. Research on big data applications in telecommunications (Sajjad & Al-Raqom, 2020) documents the adoption of hybrid approaches in which data pipelines are automated to support real-time processing, enabling advanced analytics for use cases such as churn prediction and anomaly detection. These architectures often incorporate DataOps practices and scalable processing tools (e.g., Apache Spark, Flink) to manage high-throughput data streams, subsequently storing processed outputs in data warehouses for reporting and compliance purposes.

Concurrently, data warehousing is increasingly intersecting with other analytic paradigms, leading to architectural convergence. The boundaries between data warehouses, data lakes, and distributed data platforms are becoming less distinct. Concepts such as the logical data warehouse and data fabric have been proposed to facilitate virtual integration across heterogeneous data stores (Sun & Srivastava, 2021). These architectures enable querying across distributed systems—including cloud data warehouses, on-premises databases, and data lakes—using data virtualisation layers and distributed query engines (e.g., Presto, Trino, Apache Drill). Although these technologies do not replace the need for centralised, curated warehouses, they extend analytical capabilities by enabling flexible access to data stored across disparate environments.

3.2 Risk Management

Risk management is a structured process designed to address uncertainties that may affect an organization's ability to achieve its objectives. Various approaches to risk management have been developed, with Enterprise Risk Management (ERM) emerging as a widely adopted framework due to its integrated structure. The transition from silo-based risk management to ERM reflects the increasing recognition of the need for a holistic approach that accounts for risks across multiple domains (COSO, 2017; ISO, 2018; Oti-Frimpong & Gyedu, 2023). ERM integrates risk considerations into strategic planning, aligning management practices with organizational objectives and supporting decision-making processes (Anton & Nucu, 2020). According to Naik and Prasad (2021), ERM extends beyond traditional risk management methods by incorporating corporate governance and strategic alignment, enabling organizations to address interrelated risks and improve resilience. This framework has been associated with a competitive advantage and enhanced firm performance through the standardization of risk management practices across an enterprise (Ricardianto et al., 2023). Governance mechanisms further facilitate this integration, ensuring accountability and alignment with broader organizational objectives (Lundqvist, 2015; Schäffer & Storek, 2022).

Research on ERM has explored key themes, including its adoption, impact on firm performance, role in value creation, and practical applications (Oti-Frimpong & Gyedu, 2023). The adaptability of ERM across industries and its development into a broad framework demon-

strate its relevance in the management of strategic, operational, compliance, and systemic risks (COSO, 2017)—contemporary challenges.

3.2.1 Identity and Access Management in Enterprise Compliance and Governance

IAM Compliance Practices

Identity and Access Management (IAM) is structurally integrated into enterprise security and regulatory compliance frameworks. IAM programs are designed to ensure that only authorized users receive access to specific resources, in accordance with regulatory requirements such as the General Data Protection Regulation (GDPR), the Health Insurance Portability and Accountability Act (HIPAA), and the Payment Card Industry Data Security Standard (PCI-DSS) (Vitla, 2025). In organizational contexts, IAM serves not only technical security functions but also supports legal and governance obligations. Audit traceability and enforcement of access principles such as least privilege are established through IAM controls, as required in several legal and normative frameworks (Eltayeb, 2024).

Existing research identifies a correlation between IAM implementation and improved regulatory compliance, particularly in the prevention of unauthorised access and the facilitation of audit processes (Glöckler, Schmidt, & Petersen, 2024). This alignment is reflected in normative standards, including ISO/IEC 27001, which incorporates identity and access control within the structure of an information security management system (Vitla, 2025). IAM is also described in industry reports as encompassing domains such as compliance, risk management, governance, privacy, and lifecycle management of identities (KPMG International, 2019).

IAM compliance practices involve the coordination of identity lifecycle processes—including onboarding, provisioning, modification, and termination—with internal policies and external regulatory criteria. This coordination contributes to reduced exposure to data breaches and mitigates the likelihood of compliance failures (Eltayeb, 2024). The literature presents consistent evidence that compliance-oriented IAM frameworks support audit readiness and reinforce broader information security measures (Vitla, 2025).

Access Review and Certification Processes

A central element of Identity and Access Management (IAM) governance is the periodic access review process, commonly referred to as access certification or, in some organisational contexts, ”circularisation” of access rights. This process entails the regular verification of users' access privileges to confirm their continued appropriateness. Contemporary IAM frameworks increasingly incorporate automated mechanisms for access reviews, aiming to address compliance requirements with greater efficiency (Vitla, 2025). As noted by Vitla (2025), automated attestations support audit readiness under regulatory regimes such as the Sarbanes–Oxley Act (SOX) and the General Data Protection Regulation (GDPR).

Access reviews typically require managers or system owners to assess whether assigned roles and permissions remain aligned with users' current responsibilities. This requirement is reflected in normative standards: ISO/IEC 27001:2013 prescribed regular user access reviews, and subsequent revisions maintain this expectation. ISO/IEC 27002:2022, for example, continues to specify the review and removal of unnecessary access rights as a control

objective (International Organization for Standardization, 2022a, 2022b).

The implementation of scheduled access recertification, whether conducted quarterly or annually, facilitates the identification of dormant accounts, privilege accumulation, and violations of internal policy before they result in security incidents. Both academic literature and industry documentation describe access reviews as a mechanism for reinforcing the principle of least privilege while simultaneously establishing verifiable compliance records (Raschke, 2022). Despite their benefits, manual review processes are frequently characterised as resource-intensive and susceptible to procedural weaknesses such as perfunctory approvals (Raschke, 2022).

To mitigate these issues, IAM platforms increasingly support the automation of review workflows, including the distribution of review tasks to resource owners and the central aggregation of responses. These tools enable organisations to maintain detailed audit trails and ensure that entitlements are either revalidated or revoked based on explicit confirmation (Vitla, 2025). The access review process is consistently referenced in recent literature as a formal IAM governance mechanism. It provides continuous assurance of compliance with internal policies and regulatory standards while contributing to the early detection of inappropriate access, including risks associated with insider threats (Shaw, 2018).

Key Risk Indicators in IAM Governance

The governance of Identity and Access Management (IAM) increasingly involves the adoption of Key Risk Indicators (KRIs) that quantify identity-related risks and assess the effectiveness of associated controls. KRIs function as early-warning metrics, complementing traditional performance indicators by focusing on conditions that may elevate exposure to security or compliance failures. According to Shaw (2018), the identification of IAM-specific KRIs requires a comprehensive understanding of the identity environment, including inconsistencies in access provisioning, the presence of redundant or orphan accounts, and violations of internal policies.

Recent publications outline common IAM KRIs such as the number of dormant or orphan accounts, the proportion of users with excessive access privileges, the average time required to revoke access following employee termination, and the frequency of access policy exceptions (Raschke, 2022). Monitoring these indicators enables organisations to assess their exposure to identity-related risks. For example, an increase in orphan accounts—defined as active accounts without a valid owner—may indicate deficiencies in identity lifecycle management, thereby raising the likelihood of unauthorised access. Similarly, elevated rates of policy violations or incomplete access reviews may serve as indicators of governance weaknesses.

Empirical findings suggest that aligning IAM metrics with risk-based outcomes can enhance the oversight function. Organisations that systematically monitor IAM-specific KRIs demonstrate greater capability in anticipating audit deficiencies and potential security breaches (Shaw, 2018). Furthermore, integrating these indicators into dashboards and formal reports facilitates their visibility to executive leadership, which may support decision-making and resource allocation for IAM-related initiatives (Eltayeb, 2024). Eltayeb (2024) argues that the prioritisation of IAM investments is more effective when supported by measurable impacts on compliance, security, and risk reduction.

In operational contexts, the definition and monitoring of IAM KRIs have become a recommended practice within broader compliance programmes. Examples include metrics for access review completion rates, privileged access usage, and policy adherence (Raschke, 2022).

These indicators enable the establishment of feedback mechanisms through which IAM controls can be evaluated and adapted. For instance, if a KRI reveals delays in access revocation, additional process controls can be introduced to mitigate the risk prior to the occurrence of a compliance failure.

KRIs therefore function as instruments of governance in IAM, translating technical identity management conditions into quantifiable risk metrics that can be tracked and addressed systematically over time (Shaw, 2018).

Integration with ISO/IEC 27001 and 27002 Standards

Identity and Access Management (IAM) practices are closely aligned with the control objectives articulated in ISO/IEC 27001, which establishes requirements for information security management systems (ISMS), and the complementary ISO/IEC 27002, which provides implementation guidance. The ISO/IEC 27001:2022 standard outlines a set of controls pertinent to access control, encompassing processes such as user account provisioning, periodic access rights review, privileged access management, and authentication mechanisms, which constitute the core of IAM. As such, the systematic implementation of IAM is frequently a prerequisite for achieving and maintaining compliance with ISO/IEC 27001 (International Organization for Standardization, 2022a, 2022b).

Recent literature reinforces this interdependence. As noted by Vitla (2025), industry standards such as ISO/IEC 27001 explicitly recognise the centrality of IAM in ensuring regulatory compliance and mitigating security-related risks, thereby positioning IAM as a foundational component of an effective ISMS. Organizations pursuing certification must provide demonstrable evidence that controls related to unique user identification, appropriate provisioning and revocation of access, regular reviews of access entitlements, and secure authentication protocols are both implemented and operational.

The 2022 revision of ISO/IEC 27002 introduces updated and explicitly articulated controls related to identity and access governance. Specifically, Control 5.16 (Identity Management) and Control 5.18 (Access Rights) emphasise the importance of maintaining accurate and appropriate identity records and access permissions (International Organization for Standardization, 2022a). Additionally, Control 8.2 (Privileged Access Rights) addresses the governance of elevated permissions, ensuring such access is granted on a need-to-use basis and subject to continuous monitoring (International Organization for Standardization, 2022a). The alignment of IAM systems with these controls facilitates the adoption of internationally recognised security practices and enhances overall organizational security posture.

Academic contributions further underscore the relevance of ISO standards in structuring IAM processes. Damon and Coetzee (2018, as cited in Glöckler et al. 2024) observe that maintaining auditable records of IAM activities, such as access assignments, modifications, and revocations, is imperative for satisfying both regulatory requirements and ISO-based audit criteria. The frameworks provided by ISO/IEC 27001 and 27002 offer structured guidance for establishing and sustaining such capabilities.

Furthermore, several industry-specific extensions of ISO/IEC 27001 adapt these general controls to reflect contextual operational requirements, including those directly related to IAM. Consequently, ISO/IEC 27001 functions not only as a compliance benchmark but also as a structural framework into which IAM governance naturally integrates. Effective IAM implementation supports multiple control objectives related to access governance, while adherence to ISO standards imparts methodological rigour and consistency to IAM programmes.

Namely, the standards mandate periodic reviews of user access privileges and the timely revocation of access following changes in employment status, measures which reinforce IAM lifecycle management principles(International Organization for Standardization, 2022b).

Accordingly, the integration of IAM with ISO standards yields mutual benefits: it facilitates the certification process and institutionalises best practices in access governance (Vitla, 2025). As observed in recent case studies, many organisations use ISO/IEC 27001 compliance initiatives as catalysts for formalising IAM processes, resulting in the establishment of clearly defined policies and documentation for identity administration, access approval workflows, periodic attestation, and continuous monitoring (Raschke, 2022).

3.3 Data Visualization

Data visualization is a powerful tool that transforms raw data into graphical formats, allowing the identification of patterns, trends, and anomalies that inform decision-making processes (Aspin, 2020; Healy, 2018). As a fundamental component of data visualization, dashboards serve as structured tools for continuous monitoring, integrating data from various sources into unified views (Joubert & Belle, 2016; Widjaja & Mauritsius, 2019). Presenting complex information in a concise and accessible format reduces the cognitive demands on users and improves risk communication and management practices, facilitating more effective decision-making (Joubert & Belle, 2016; Quynh, 2023).

The structured flow of data visualization can be summarized in a pipeline, which illustrates how raw data is progressively refined into meaningful visual formats (Figure 3.1).

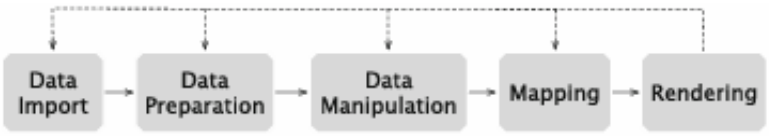

Figure 3.1: The data visualization pipeline, illustrating the transformation of raw data into structured visual formats.

Source: Adapted from Qin et al. (2020), Making Data Visualization More Efficient and Effective: A Survey.

The transformation of raw data into meaningful representations follows a structured pipeline involving data importation, processing, mapping, and rendering, as depicted in Figure 3.1. This sequence is particularly relevant to dashboard design, as it directly influences the clarity, usability, and interpretability of visual outputs (Qin, Luo, Tang, & Li, 2020).

An effective dashboard design ensures that users can efficiently interpret and act on information, reinforcing data-driven decision-making processes (Muskan, Singh, Singh, & Prabha, 2022). Poorly designed dashboards may present excessive or irrelevant details, leading to cognitive overload and affecting focus on key information. The principles of design emphasize clarity, simplicity, and user engagement to align the interface with user needs (Singh, Kumar, Singh, & Kaur, 2023). Explicit visual representations structure complex data into manageable formats, while accurate visualization reduces the risk of misinterpretation and improves decision-making processes (Widjaja & Mauritsius, 2019).

User experience (UX) design significantly influences data visualization, ensuring that visual representations are both informative and accessible. UX design simplifies data interpretation by converting information into intuitive visual elements such as color bars, lines, and spatial representations, which align with human cognitive processing. This approach improves user interaction by reducing cognitive load and facilitating accessibility. The research by van Willigen (2019) mentions the importance of evaluating the user experience of data visualization to ensure that the visual outputs align with user expectations. Furthermore, UX-based visualization frameworks support software professionals in analyzing user interactions, identifying usability patterns, and areas for optimization (Macedo & Zaina, 2024).

Few (2006) identifies key principles for dashboard design, advocating for the organization of critical information within a single screen for rapid comprehension. Guidelines emphasize the removal of non-essential elements and the prioritization of content that directly supports decision-making. These recommendations align with visual perception principles, which facilitate efficient information processing by minimizing distractions. According to Few (2006), it's essential to tailor dashboards to optimize utility while reducing cognitive load, especially in environments where rapid data interpretation is necessary. Building on these foundations, Dona M. Wong (2013) explores practical design aspects, including strategic color usage, typography, and labeling. She highlights the role of color in conveying meaning while cautioning against excessive use, which can obscure critical information. Additionally, Wong (2013) emphasizes placing labels directly on data points to improve usability and reduce reliance on legends, ensuring that dashboards remain accessible in diverse operational contexts.

Adherence to the principles of accuracy, clarity, and relevance ensures that visualizations align with organizational objectives and user requirements (Muskan et al., 2022). Dashboards must support accessibility across devices and screen sizes, allowing for adaptability in various operational environments (Singh et al., 2023). The customization of dashboards to specific audiences, as recommended by Few (2006), ensures that displayed metrics directly correspond to user priorities while mitigating the risk of information overload. Wong (2013) further emphasizes the role of typography in maintaining clarity and usability, reinforcing dashboards as structured decision-support tools.

Data storytelling improves visualization by integrating narrative techniques that contextualize data, making complex information more interpretable. This approach combines analytical outputs with visual elements to enhance audience engagement. A widely recognized example of compelling data storytelling isMinard'ss visualization of Napoleon's retreat from Moscow, which integrates multiple data dimensions, troop size, geography, and temperature—into a single representation (Healy, 2018).

Figure 3.2 exemplifies how visualization can serve as a narrative tool, conveying historical and statistical relationships within a single image. As Healy (2018) discusses, Minard's chart integrates geospatial (route), quantitative (troop losses over time), and meteorological (temperature) data into a unified structure, allowing viewers to interpret the severity of the campaign at a glance. The visualization's structure enables audiences to extract meaningful insights without requiring extensive textual explanations. This principle is fundamental to dashboard design, where the objective is to present large-scale structured data in an intuitive format, allowing decision-makers to derive conclusions efficiently.

Advanced visualization techniques optimize the analysis of complex datasets by revealing relationships, trends, and dependencies that might not be immediately apparent. Graph-based visualizations, in particular, are valuable for examining interconnected data structures, such

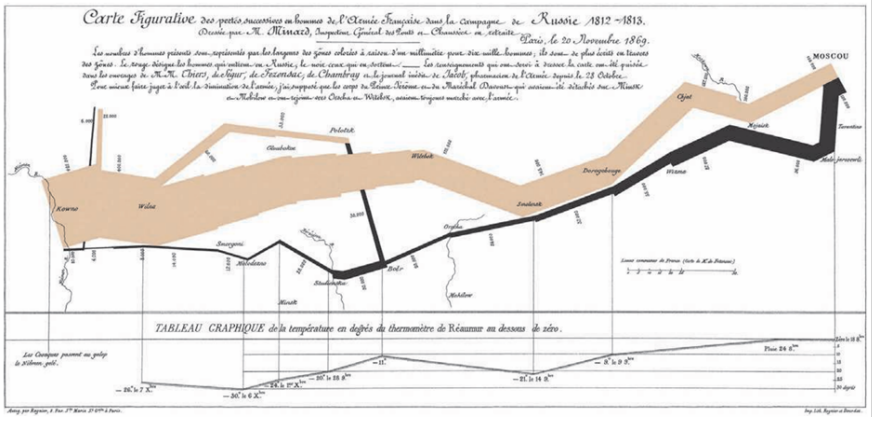

Figure 3.2: Minard's Visualization of Napoleon's Retreat from Moscow, illustrating the combination of spatial, quantitative, and temporal data in a historical narrative.

Source: Adapted from Healy, K. (2018). *Data Visualization: A Practical Introduction*. Princeton University Press (Healy, 2018).

as networks and relational datasets. These methods enable the identification of patterns and correlations that may be overlooked in traditional tabular representations, contributing to more informed decision-making processes (Po, Bikakis, Desimoni, & Papastefanatos, 2020). When integrated into dashboards, graph-based tools provide users with valuable information while maintaining usability and accessibility (Po et al., 2020).

Technological advancements have significantly expanded dashboard functionalities through interactive elements, advanced analytics, and visualization pipelines. Decomposition trees dynamically explore hierarchical data, while heatmaps represent density patterns, allowing customized analyses for specific operational needs (Microsoft, n.d.). Visualization pipelines augment performance when handling large datasets by employing incremental sampling and approximate query processing (Qin et al., 2020). These approaches provide seamless interaction with large datasets while maintaining responsiveness. Interactive features such as drill-throughs, slicers, and anomaly detection tools allow users to navigate risk metrics dynamically, adapting analyses to diverse organizational requirements (Microsoft, n.d.).

In addition, dashboards serve as platforms for monitoring KRIs relative to predefined thresholds, allowing organizations to detect when risks exceed acceptable limits. Joubert and Belle (2016) explains the role of dashboards in improving risk monitoring by providing real-time tracking capabilities, ensuring timely risk mitigation responses.

Risk dashboards consolidate relevant information into formats that support monitoring and decision-making processes. Joubert and Belle (2016) propose a framework that simplifies data presentation through single-screen interfaces. Focusing on critical categories and employing abstraction layers reduces information overload, allowing decision-makers to assess risk levels and prioritize responses. Real-time data integration improves the relevance of dashboards in dynamic operational contexts (Akano, Hanson, Nwakile, & Esiri, 2024). Alhamadi (2020) identifies challenges such as prioritizing aesthetics over functionality, which can hinder users' ability to interpret data efficiently. Adaptive design techniques, including customizable interfaces and dynamic visualizations, address these issues. Akano et al. (2024) emphasizes that visual elements, such as graphs, charts, and heatmaps, improve interpretability; therefore, users can quickly identify trends and outliers.

Dashboards also foster behavior change within organizations. Quynh (2023) explains that visual feedback mechanisms increase awareness and accountability, encouraging adherence to established protocols. Real-time data integration supports regulatory compliance and proactive risk management, facilitating timely interventions (Akano et al., 2024). Joubert and Belle (2016) note that presenting data alongside thresholds improves communication, allowing users to detect and respond to risks exceeding acceptable limits. The evolution of dashboard technologies highlights their potential to address complex organizational challenges. Integrating advanced analytics, enabling real-time monitoring, and adhering to user-centered design principles equip dashboards to improve decision-making and foster accountability (Akano et al., 2024; Alhamadi, 2020; Joubert & Belle, 2016; Quynh, 2023).

Modern dashboards manage large datasets from diverse and often disconnected sources, providing users with tools for dynamic exploration and real-time analysis (Singh et al., 2023). Organizations increasingly rely on fragmented data systems, making scalability essential to maintain performance and usability (Aspin, 2020). Techniques such as data aggregation, incremental refresh, and optimized data models ensure efficiency and responsiveness, even as data complexity and volume increase. Scalability challenges are particularly significant where stakeholders require tailored data for specific operational needs. Responsive dashboards mitigate these challenges through customizable formats, since users can filter and explore data interactively. This flexibility supports context-specific analyses, allowing decision-makers to examine data at varying levels of granularity (Ferrari & Russo, 2016). Advanced analytics contribute to scalable dashboards by expanding their analytical capabilities. Features such as anomaly detection and predictive analytics identify trends, identify potential risks, and facilitate proactive responses to emerging challenges (Quynh, 2023).

3.3.1 Discussion

The literature reviewed in this chapter establishes the conceptual and technical foundations necessary for developing data-driven Identity and Access Management solutions. Classical data warehouse architectures, including those proposed by Inmon (2005) and Kimball and Ross (2013), provide models for organising analytical environments that prioritise either enterprise consistency or rapid delivery. These models, together with recent architectural variants such as the data vault approach (Linstedt, 2011) and cloud-based solutions (Gill et al., 2020; Nambiar & Mundra, 2022), inform the design choices adopted in the present work. While the reviewed studies describe general-purpose architectures, the current project adapts these principles to the specific constraints of managing access data in a fragmented telecommunications environment.

The project also draws from the literature on ETL modelling and automation. As noted by Dhaouadi et al. (2022), the structure and performance of ETL pipelines directly affect the reliability and timeliness of data analysis. The present implementation incorporates practices such as incremental extraction and data cleansing, consistent with recommendations in Rahm and Do (2000) and Bahaa et al. (2021). Still, it applies them within a use case focused on access conformity and risk tracking. While automation tools are described in the literature as emerging trends (Diouf et al., 2018; Simitsis et al., 2005), this project implements selected automation steps using Python to handle operational constraints not addressed by standardised tools.

In the area of risk management, this work aligns with the integrated governance frame-

works described in COSO (2017) and Naik and Prasad (2021), while narrowing the scope to Identity and Access Management. It operationalises concepts such as access certification and lifecycle governance, which are referenced in both regulatory and normative documents. The implementation reflects practices discussed by Vitla (2025) and Eltayeb (2024), particularly regarding the mapping of ISO/IEC 27001 and 27002 controls to access review processes and privileged access monitoring.

Dashboards developed in this project are informed by design principles articulated in Few (2006) and further expanded in Wong (2013), including the prioritisation of clarity, contextual relevance, and minimisation of cognitive load. The dashboards serve not only as visual interfaces but also as mechanisms for monitoring Key Risk Indicators, consistent with the approach described by Shaw (2018) and Raschke (2022). While these authors examine KRI selection in conceptual terms, this project implements concrete metrics such as orphan accounts, privilege accumulation, and access review completion, tailored to the organisation's operational context.

Finally, the role of dashboards in supporting user navigation and decision making is addressed using concepts from data visualization literature, including the pipeline structure presented by Qin et al. (2020) and the data storytelling strategies described by Healy (2018). However, unlike prior studies that emphasise general visualisation effectiveness, this work focuses on the integration of interactive logic, targeted navigation paths, and IAM-specific content, particularly in the context of circularization workflows and risk-driven conformity monitoring.

In summary, the present work is consistent with existing literature in its use of established architectural, governance, and visualization models. Still, it differs by applying these models to a specific IAM use case using an implementation-oriented approach. It contributes by operationalizing IAM governance concepts through practical data processing pipelines and dashboard interfaces adapted to real organizational constraints.

Chapter 4

Methodology

4.1 Overview

The methodological approach adopted in this study was structured to support the integration, transformation, and analysis of data related to Identity and Access Management (IAM) within a corporate environment. The objective was to design and implement a solution capable of aggregating fragmented access-related data into a coherent data warehouse, enabling effective monitoring and governance through dashboard visualization.

The process was divided into sequential phases, each corresponding to a concrete technical or analytical task. These included data collection from heterogeneous sources, implementation of Extract, Transform, and Load (ETL) routines in Python, structured storage in an Oracle database, and the development of dashboards using Power BI. The methodology also incorporated principles from ISO/IEC 27001:2022 and ISO/IEC 27002:2022 to ensure alignment with internationally recognised IAM controls and governance practices.

4.1.1 Business Understanding

The Risk and Compliance department at NOS is responsible for ensuring the governance of identity and access management (IAM) across the organization. Its remit includes overseeing the assignment, monitoring, and validation of logical access rights granted to employees, external service providers, and third-party users. A primary objective is to ensure that these permissions align with internal security policies, regulatory requirements, and the foundational principles of information security, namely confidentiality, integrity, and availability.

In the context of access control, confidentiality refers to ensuring that information is accessible only to authorized users, integrity provides the accuracy and trustworthiness of information, and availability guarantees that information and systems are accessible when needed. The principles of *need to know* and *need to have* further constrain access by requiring that users are granted only the minimum permissions necessary to perform their roles, thereby reducing the risk of misuse, unauthorized disclosure, or privilege accumulation.

A central process within this governance framework is the *Circularização de Acessos*, a structured monthly review of access rights, with particular emphasis on critical and remote permissions. This process aims to verify that access to sensitive systems and data remains justified, up to date, and compliant with the organization's security standards. It encompasses not only direct access assignments but also the purposes (*finalidades*) associated with each permission,

ensuring adherence to the *need to know* and *need to have* principles that underlie secure and proportional access control.

This review process follows a monthly operational cycle comprising data collection, transformation, validation, and visualization. It involves a heterogeneous set of sources, including structured data from the corporate data warehouse (DW), identity records and permission metadata from the IAM system, access logs from SOC, and various lists provided by departments such as DSI and Audit via email. These data inputs include assignment lists, utilization logs, records of suspended aggregators, bypassed authentication tokens, and access justification data. In some cases, the Risk and Compliance team extracts data directly through SQL or SAP BO; in others, it depends on third-party departments to supply processed information.

The current implementation is characterized by a high degree of fragmentation and manual effort. Numerous files are handled in Excel, often requiring repetitive operations such as VLOOKUPs, pivot tables, and VBA scripting. The process involves manually aggregating records into master assignment and utilization files, resolving inconsistent formats, and adjusting month references, especially during year transitions. Manual validation of period comparisons and the interpretation of KRIs introduce additional complexity and increase the risk of human error. Furthermore, the limitations of Excel in handling large datasets affect performance and compromise scalability.

The outputs of the circularization process are visualized in Power BI dashboards. These dashboards are structured across multiple sheets, each focused on a specific control domain such as critical access, conformity, remote access attribution, and usage, and are intended to support the decision-making of access managers. However, their design presents usability limitations. The current interface is not fully optimized for user experience. In some cases, users report difficulty locating relevant information efficiently, which reduces the dashboard's effectiveness as a decision support tool.

The present project seeks to address these challenges by developing an automated and scalable data pipeline for IAM monitoring. The solution involves implementing ETL procedures in Python to consolidate fragmented datasets, normalize data structures, and load information into a dedicated repository, from which dashboards are generated in Power BI. This approach is designed to reduce dependency on manual Excel processing, improve data consistency and quality, and enable the timely delivery of access circularization outputs.

Strategically, the project supports the Risk and Compliance department's goal of strengthening access governance and audit readiness. It facilitates continuous monitoring of KRIs associated with access risks. It reinforces adherence to ISO/IEC 27001:2022 and ISO/IEC 27002:2022, particularly in the domains of identity lifecycle, authentication control, privileged access monitoring, and segregation of duties. Additionally, by mitigating risks associated with excessive or unjustified access, the project contributes to the organization's broader risk management framework and supports enterprise-level compliance, operational resilience, and cybersecurity maturity.

4.1.2 Data Understanding

The data understanding phase examines how data are extracted, transformed, and visualized in the current access governance process at NOS. The existing method for critical and remote access review (*Circularização de Acessos Críticos e Remotos*) is manual mainly, involving multiple data sources and significant data handling in spreadsheets. This section is organized

into three stages: Data Extraction, Data Transformation, and Data Visualization, followed by a dedicated Fragility Analysis that identifies weaknesses in the as-is process.

Data Extraction (Current Process As-Is)

Data Sources: The access governance process draws on several disparate data sources. Key inputs include the corporate Data Warehouse (DW), an identity management system called IAM, SOC, and various departmental files received via email (from the Information Security department (DSI) and IT Audit team). Below is a brief description of each source and the type of data provided:

- **Data Warehouse (DW):** The DW houses consolidated information on identities and access rights. Using SAP Business Objects (SAP BO) reports or direct SQL queries, the team extracts lists of active user identities and their access assignments. Relevant attributes include user IDs, names, job positions, and the roles or permissions assigned to each user.
- **IAM System:** IAM is the internal user access management application (the authoritative source for access requests and approvals). IAM provides information such as whether an identity is internal or external and metadata like the declared purpose of each access.
- **User Logs:** The User Logs collect logs of user activities on critical systems. These logs include usage data such as login timestamps, accessed systems (e.g., App 1, App 2, App 3), and types of actions performed.
- **Departmental Email Inputs:** Some datasets are only available via manual requests to DSI and IT Audit. These include the list of declared purposes, suspended identities, bypass token list, Pulse Secure permissions, and usage logs.
- **Privacy Catalog:** Used to interpret declared purpose codes by mapping them to descriptive business justifications.
- **Historical Access Data:** An Excel file from the previous month's circularization is referenced to compare changes in access and carry forward historical data.

The overall end-to-end flow of the current access governance process is shown in Figure 4.1. This diagram illustrates how data are extracted, transformed, and visualized, and it serves as a reference for the subsections that follow.

As illustrated in Figure 4.1, the data extraction stage begins with a monthly cycle. At the start of each cycle, the responsible analyst runs DW reports and SQL queries to pull the latest identity and access tables. Simultaneously, requests are sent to DSI and Audit teams for the auxiliary files, introducing a dependency on timely responses. Once collected, the raw data includes Excel and CSV files from various databases and departmental inputs. Each source may have a different format and structure, requiring separate handling. The process is manual and time-consuming, as it involves generating SAP BO reports, managing multiple email responses, and organizing data in a local directory. Delays in receiving usage logs or auxiliary data can disrupt the entire cycle, emphasizing the fragility inherent in relying on multiple decentralized contributors.

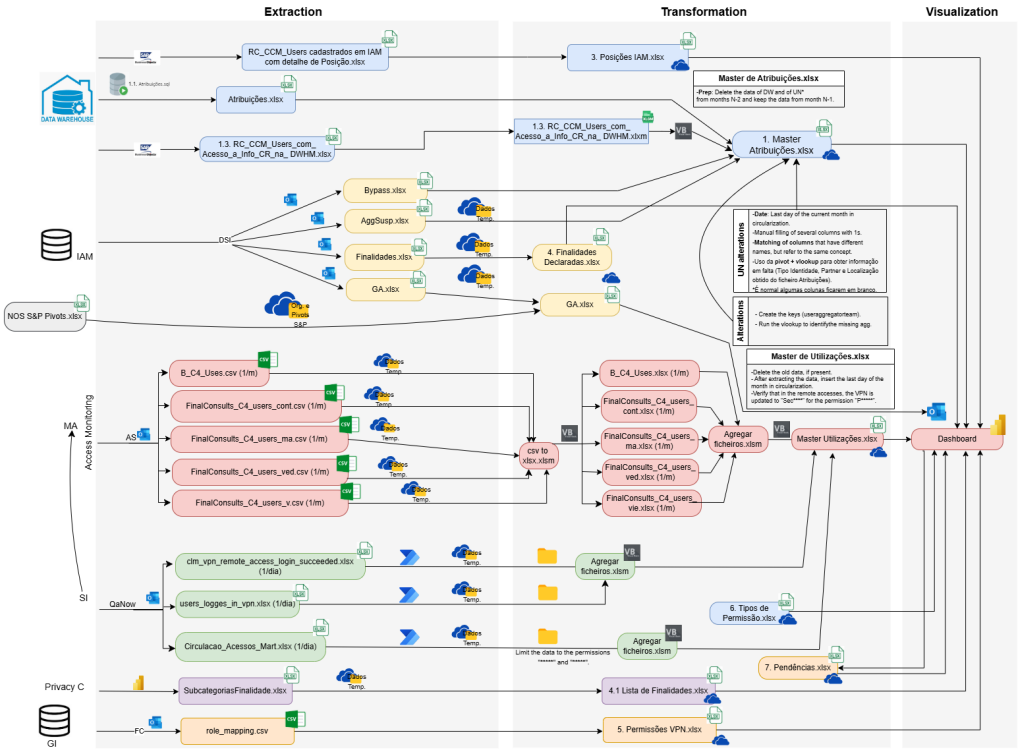

Figure 4.1: End to end data flow of the current access governance process.

Data Transformation

In the transformation step, represented in the middle section of Figure 4.1, the extracted datasets undergo integration and cleansing through a combination of Excel workbooks and supporting scripts. The objective is to consolidate the various input sources into a unified dataset that accurately represents the relationship between users and their assigned accesses, enriched with relevant metadata such as the purpose of access and frequency of use. This dataset subsequently serves as the foundation for the dashboard. This process, as currently implemented, involves extensive manual operations in Excel:

Manual Data Integration: The team maintains master Excel workbooks (e.g., "Master de Atribuições.xlsx" for access assignments and "Master de Utilizações.xlsx" for usage logs) that act as central repositories. Data from each source is pasted or imported into these masters. For example, after obtaining the "List of Access Assignments" from the DW, the analyst manually copies it into the master workbook. Similarly, usage CSV files received from SOC (often large, with thousands of log records) are converted to Excel format and appended. Some of these conversions are aided by VBA macros (e.g., a macro to transform multiple CSV log files into sheets in an XLSX). Nevertheless, many steps require human intervention, such as moving files into specific folders, opening reference workbooks, and triggering macros.

Use of Excel Formulas and Pivot Tables: Once data from all sources resides in Excel, lookup formulas (Excel VLOOKUP/INDEX-MATCH) are used extensively to join information. For instance, the declared purpose of each access (from the DSI email) must be matched to the corresponding user or access entry in the assignments master. Another example is pulling in user attributes like department or partner company from identity lists or

the "Positions IAM" file. One identified step requires having the "3. Posições.xlsx" file open, so that a VLOOKUP can fetch attributes like identity type or location for each user. Pivot tables are used to summarize or reformat data; for example, to count how many critical systems each user has access to or to structure data for Power BI. These formulas and pivots must be adjusted if column names or file structures change, adding to process complexity. The heavy reliance on VLOOKUP and similar functions introduces the risk of formula errors, misalignments, and data gaps.

Data Cleaning and Preparation: During transformation, the team performs cleanup tasks such as removing inactive or terminated users, handling missing values (e.g., filling undeclared purposes as "Not provided"), and preparing data for historical comparison. To enable comparisons between months, the Excel master retains data for the current and previous month, requiring manual deletion of older data. Specific content transformations are also applied, such as renaming VPN entries or adjusting permission labels to standardize the dataset. These transformations are implemented via find-and-replace or macros.

Macros and Automation Scripts: Some repetitive tasks have been partially automated using VBA. For example, one macro aggregates multiple usage files by opening each application's CSV log, filtering out irrelevant entries, and appending relevant records into the "Master Utilizações" workbook. Another macro compiles the dashboard input. However, these macros must be run manually and often require user parameters (e.g., cutoff dates). Mistyping a parameter can result in data exclusion. Additionally, macros are resource-intensive, and large datasets can cause Excel to freeze. The "aggregate files in multiple steps" macro is particularly computationally demanding, often operating near Excel's performance limits (e.g., due to RAM or CPU consumption).

In summary, although a degree of automation exists, the transformation stage remains laborious and error-prone. Each month, the outcome depends on the accurate execution of multiple manual steps: copying data, invoking macros in sequence, and verifying outputs. Excel is used as both an integration platform and a calculation engine, which creates limitations in terms of scalability (due to row limits and performance degradation above one million records) and data integrity (across numerous interlinked formulas and sheets).

Data Visualization

Following the transformation and consolidation stages, as shown on the right side of Figure 4.1, the final dataset is imported into a reporting environment for visualization. Currently, Microsoft Power BI serves as the platform for hosting the Access Review Dashboard, which provides visibility into key risk indicators (KRIs) and findings related to critical and remote accesses. This visualization stage entails the manual upload of the processed Excel-based dataset into Power BI and the subsequent design of interactive reports tailored to stakeholders such as IT risk management teams, internal audit, and system owners.

Dashboard Content: The dashboard presents metrics aligned with the Need-to-Know principle and broader information security policies. It identifies, for instance, users with remote access (e.g., VPN or VDI) who rarely or never utilize it, thereby flagging potential instances of excessive privilege. It also aggregates the number of users with access to each critical application and highlights anomalies, such as departments with a disproportionately high volume of critical access assignments. The visual elements include bar charts depicting user distribution across critical systems, tables displaying accounts and their last login activity to identify dormant accounts, and filters that allow disaggregation by identity type (e.g., in-

ternal versus external users). Each monthly dataset is treated as a discrete snapshot, which is loaded into the dashboard for inspection.

Manual Data Refresh: Given that the data preparation process occurs outside the Power BI ecosystem, the dashboard is not connected to live source systems. Instead, the reports are updated through a manual refresh procedure, whereby the analyst replaces the previous dataset with the latest master Excel file and verifies the consistency of visual outputs. Changes in the underlying data schema, such as column name modifications, can result in the failure of visuals or calculations, necessitating manual adjustments. Currently, the capability for period-to-period comparison is limited; historical data may be manually loaded for side-by-side evaluation, but there is no automated mechanism for change detection. This reliance on visual inspection introduces the risk of overlooking subtle discrepancies between circularization cycles.

Limitations in Visualization: The dashboard's effectiveness is constrained by both its design and the manual nature of the data pipeline. First, data timeliness is suboptimal, as updates are performed only on a monthly basis following extensive manual intervention, thereby precluding real-time monitoring. Second, the interpretability and analytical rigor of the visual outputs are limited. Qualitative feedback, such as the interface being unclear, points to a lack of immediate clarity in interpreting the presented metrics. The absence of standardized thresholds or visual cues (e.g., color coding) impairs the user's ability to distinguish between normal and anomalous values. Planned enhancements include the incorporation of threshold-based alerts (e.g., flagging users inactive on VPN for more than 60 days) and a redesign of visuals to prioritize risk-oriented insights. The dashboard's reliability also remains contingent on the accuracy of the upstream data transformation process: errors introduced during manual preprocessing stages are propagated to the final reports. Moreover, Power BI lacks native functionality for automated tracking of historical changes, resulting in version control challenges, as data from previous periods must be imported and aligned manually.

Despite these limitations, the dashboard constitutes a central instrument for access governance and risk management. It consolidates previously fragmented datasets and supports informed decision-making by enabling actions such as the revocation of excessive privileges or the reinforcement of Need-to-Know policy compliance. The subsequent subsection addresses structural fragilities identified in the current implementation, which motivate the pursuit of a more automated and scalable solution.

Fragility Analysis of the Current Process

An in-depth fragility analysis was conducted to identify weaknesses and opportunities in the as-is access governance process. In total, 28 distinct process vulnerabilities ('fragilities') were identified, grouped into six categories, along with a few improvement opportunities. Each fragility was evaluated on multiple criteria - such as how frequently it occurs, its impact on data integrity and efficiency, and the potential for disrupting the information flow – to assess its overall criticality. The following categories of fragilities were identified in the current process:

Manual Processing Many steps (data extraction, cleaning, and input into the dashboard) require manual effort. This category includes issues like manual report generation, copying and pasting data between sheets, and manual verification of outputs. Manual processes are not only time-consuming but also error-prone, as mistakes can be introduced (e.g., copying

the wrong range or overlooking an entry). This was the most prevalent category of fragility, reflecting the heavy reliance on human intervention in the current workflow.

Process Complexity Specific tasks involve an unnecessarily complex procedure or workaround. For example, using multiple different queries to retrieve related data (when one unified query could suffice) or needing to have auxiliary files open to perform lookups are signs of process complexity. Complex processes increase the chance of errors and make the procedure more complicated to execute consistently.

Dependency on Third Parties Some steps depend on external parties or systems (e.g., waiting for emails from DSI/Audit or needing access to a specific network drive or tool). This dependency can cause delays and is a single point of failure - if the person or system responsible for providing data is unavailable, the whole cycle is affected. It also reduces the team's control over the end-to-end process.

Excel Limitations This category covers the technical limitations of the tools being used, notably Microsoft Excel. Two sub-issues are highlighted: Excel capacity (the tool's ability to handle large volumes of data) and computational capacity (performance constraints). For instance, aggregating large log files pushes Excel to its limits, risking crashes or forcing the use of several smaller files instead of one, which complicates the analysis. These limitations mean the process does not scale well as data volume grows.

Manual Verification and Data Quality Even after processing, there is a need for manual cross-checks to ensure data accuracy (for example, manually comparing this month's results to last month's to spot any anomalies). The reliance on human review means that data quality issues (like inconsistent fields or missing entries) might slip through if not noticed. The process lacks systematic validation controls, so it leans on the analyst's expertise to catch problems. This is related to data quality fragilities – if the underlying data from sources like IAM is inconsistent, it may not be cleaned thoroughly, leading to misleading information in the dashboard.

Ineffective Visualization Although more about the output stage, a fragility noted was that the dashboard is not very user-friendly or focused, which can dilute the impact of the analysis. If the insights are not clearly presented (e.g., no clear flags for urgent issues), the value of the whole process diminishes. This is recognized as an opportunity to improve rather than a process "failure," but it is included since it affects how well the data can be understood and acted upon.

To quantify these fragilities, a weighted scoring formula was used. Each fragility was scored on five factors: (1) Probability of occurrence or error, (2) Time consumed/effort required, (3) Impact on the form/presentation of data, (4) Impact on the substance/content of data, and (5) Interruption of information flow. Each factor was rated on a scale (e.g., 1 = low, 3 = high impact/probability) and given a weight reflecting its importance. The weights used were: 0.25 for probability, 0.20 for time, 0.10 for impact on form, 0.30 for impact on substance, and 0.15 for interruption.

The weighted scoring formula is shown in Equation 4.1.

$$\text{Fragility Score} = 0.25 \times P + 0.20 \times T + 0,10 \times I_f + 0.30 \times I_s + 0.15 \times I_i \qquad (4.1)$$

where P is the probability of occurrence, T is the time/effort impact, I_f is impact on form, I_s is impact on substance, and I_i is impact on information flow. This score (on a 1 to 3 scale) was then mapped to a Criticality Level: generally, a score ≤ 1.5 corresponds to High criticality (level 1), 1.51–2.5 to Medium (level 2), and > 2.5 to Low (level 3). In other words, a higher criticality level number means a less critical issue (this scale was defined such that "1" is the most severe). For clarity, we will refer to them qualitatively as High, Medium, or Low criticality.

Tables 4.1, 4.2, and 4.3 below group the identified fragilities by their criticality level, listing each issue with a brief description. The tables illustrate that a handful of High-criticality issues were found, mostly related to manual processing in the data preparation stage, which directly threaten data integrity or the sustainability of the process. The majority of issues are Medium-criticality, significant inefficiencies or risks that should be addressed, though their impact is somewhat contained. A few Low-criticality issues were noted as well, often minor inconveniences or edge cases that have limited impact.

Table 4.1: High-Criticality Fragilities Identified

ID	Process Stage	Category	Description
3	Data Preparation	Manual Processing	Use of multiple Excel sheets and complex VBA macros to prepare data (four different worksheets, formulas on each, etc.), which increases the chance of errors (e.g., VLOOKUP in the "Positions" file only captures the first occurrence, possibly missing others) and requires manual removal of inactive entries.
7	Data Preparation	Manual Processing	Missing data fill via pivot: Certain identity details (type, partner, location) are not present in the main file and must be fetched from the assignments file using a pivot table. This workaround is manual, and if not done, leaves critical data gaps.
13	Data Preparation	Manual Processing	Manual file handling: Placing raw CSV log files into a "Current Month" folder and running a VBA script to convert them into XLSX. If files are misplaced or the macro isn't run correctly, usage data will be incomplete.

15	Data Preparation	Manual Processing	Multi-step manual aggregation: Requires moving various files to correct directories (segregating those in scope for review), then running a multi-step VBA macro to aggregate them. Also involves manually entering the target date (end of the month) for the cycle. Numerous manual inputs make this step prone to human error, and special adjustments (e.g., filtering certain permissions like "Reproduzir" or renaming VPN entries) must be done correctly to avoid skewing the data.
16	Data Preparation	Excel/ Computational Capacity	Performance bottleneck: The file aggregation macro is computationally intensive, often nearing Excel's performance limits. This can cause processing to be extremely slow or even crash, risking an incomplete aggregation of data.

Table 4.2: Medium-Criticality Fragilities Identified

ID	**Process Stage**	**Category**	**Description (Issue)**
1	Data Extraction	Manual Processing	Manual report execution: Running the DW report and then copying its results into the working Excel. This introduces the risk of paste errors or omissions if the analyst is not careful.
2	Data Extraction	Complexity	Redundant queries: The data for identities and their access rights are pulled with separate queries via DW and SQL, even though they could be combined. Using multiple queries unnecessarily increases the chance of inconsistent results and slows down the extraction process.
4	Data Transformation	Complexity	External file dependency: The process requires the file 3.Posições IAM.xlsx to be open for lookups. If this file is not opened or is outdated, the lookup for user positions will fail silently, leading to missing position data in the output.
5	Data Extraction	Dependency on Third Parties	Timely data request: The data from certain systems, such as CRM, must be requested from another team in advance. If the request is not made or fulfilled on time, the data extraction is delayed. This timing dependency adds operational risk each month.
6	Data Transformation	Manual Processing	Historic data cleanup: Manually deleting the data from two months ago (N-2) in the master file while retaining last month (N-1) for comparisons. Mistakes in this step (e.g., deleting the wrong range) can remove needed data or fail to purge old data, leading to incorrect comparisons.

8	Data Extraction	Dependency on Third Parties	Email-based utilization data: Utilization logs for some critical apps arrive via email from Audit. If the email is missed or the attachment is wrong, utilization data for that app will be missing, reducing the completeness of the review.
9	Data Extraction	Manual Processing	Multiple IAM extracts: Identities and their critical info flags are extracted in separate lists (one for all identities, one for those with Info CR flags). This requires reconciling two lists and could be streamlined; doing it separately risks mismatches if not joined correctly.
10	Data Extraction	Dependency on Third Parties	Pulse Secure permissions via email: The VPN permission list (Pulse Secure roles) is provided manually by DSI. Any error in this email or delays in updating it (e.g., if a new role was created but not communicated) could result in the review using outdated VPN permission data.
11	Data Transformation	Excel Capacity	Large volume in Excel: The "Master Utilizações" file, handling all usage logs, grows very large. Excel's row limit and memory constraints force the data to be split or trimmed (for instance, possibly not all log details are kept). This medium-severity issue could lead to truncated data – e.g., only the first million log entries are considered – which might omit relevant information if not handled carefully.
12	Data Transformation	Excel Capacity	Pivot table size limits: Creating pivots or performing calculations on the full dataset is cumbersome. The analyst sometimes must break the analysis into parts (e.g., separate pivots per system) because a pivot on the entire dataset would be too slow or hit Excel limits. This adds complexity and chances for error when stitching the parts back together.
14	Data Transformation	Manual Verification	No automated change detection: Differences between last month's and this month's access lists are manually checked. The analyst visually scans for new or missing entries. This manual verification can miss subtle changes (especially in a large list) or generate false alarms, and it depends on the reviewer's attention to detail.
17	Data Transformation	Dependency on Third Parties	CRM data dependency: Access data from the CRM system (a specific platform) is provided by a separate team. If their export is in a different format or if they include unexpected entries, the integration macro might fail. The core team does not have direct control over querying CRM's data, representing a medium risk if communication fails.

18	Data Extraction	Dependency on Third Parties	”Finalidades” (Purpose) data via email: The declared purpose for each access, crucial for compliance checks, is obtained by asking DSI to extract from IAM. Any misalignment (e.g., DSI sends an outdated list or one that doesn’t exactly match the users in the current cycle) will result in missing purpose annotations for some accesses.
19	Data Transformation	Manual Processing	Interim file handling: There are intermediate files (like temporary CSVs or outputs of macros) that the process creates. Managing these interim files (moving them, deleting after use, etc.) is manual. If the analyst accidentally deletes or overwrites a file too early, they may have to restart that part of the process.
20	Data Transformation	Manual Processing	Split handling of data categories: Remote access data and on-premises critical access data are handled slightly differently in transformation (different macros or sheets). Keeping track of two parallel workflows increases workload and the possibility of applying a step to one and forgetting to apply it to the other, causing inconsistency.
21	Data Visualization	Manual Processing	Manual dashboard update: Publishing the new data to the Power BI dashboard involves manually uploading the Excel and checking visuals. There is a medium risk that the wrong file could be uploaded or a step missed in updating filters, which might display mixed data (e.g., titles saying ”September” while data is from October).
22	Data Visualization	Manual Verification	No systematic QC on dashboard: There’s no automated quality check after dashboard refresh. It’s up to the analyst to notice if a chart looks off (which might indicate an underlying data issue). This subjective judgment means some errors might only be caught if they are glaring.
23	Data Visualization	Ineffective Visualization	Dashboard not highlighting issues: The current dashboard layout does not automatically highlight issues that are out of compliance. For example, it lists all accesses but does not flag which ones violate the Need-to-Know principle. The onus is on the user to interpret the data, which could lead to oversight of medium-severity issues that are not immediately obvious.

24	Data Visualization	Ineffective Visualization	Static reporting without alerts: There are no alerting mechanisms (no email notifications or pop-ups for anomalies). A medium concern is that if a serious issue occurs (say, a sensitive access granted without proper justification), it would only be noticed when someone looks at the dashboard; if they happen to skip it, the issue persists unnoticed.

Table 4.3: Low-Criticality Fragilities Identified

ID	**Process Stage**	**Category**	**Description (Issue)**
25	Data Extraction	Manual Processing	Manual renaming of the extracted files: Files from various systems or log sources often need to be renamed to follow the expected naming pattern. This is done manually and may lead to misclassification or loading errors if a file name is inconsistent or mistyped.
26	Data Transformation	Manual Processing	Manual filter tuning: Some filters (e.g., for removing irrelevant access types) are set via Excel drop-downs or filtered ranges. If the filters are not updated correctly month-to-month, irrelevant entries might be included or valid ones excluded.
27	Data Transformation	Manual Verification	Missing logic documentation: Some business logic is embedded in Excel formulas or VBA (e.g., what counts as a ”reused” access). If not well documented, future maintainers may not understand or reproduce the logic, leading to potential deviation.
28	Data Transformation	Excel Capacity	Performance lag in large pivots: In dashboards or summary sheets, Excel pivot tables become sluggish when dealing with tens of thousands of records. This causes delays during quality checks or refreshes, although not critical to the main processing.
29	Data Visualization	Manual Processing	Inconsistent chart labels: Occasionally, titles or footnotes are not updated correctly in the Power BI dashboard. While these do not affect the underlying data, they may confuse the reader or suggest outdated content if not manually verified.
30	Data Visualization	Manual Verification	No dashboard versioning: There is no record of what the dashboard looked like in past cycles. If a mistake was made or a user disputes an earlier state, there is no easy way to compare previous cycles unless screenshots were saved manually.

As seen in the tables above, the high-criticality fragilities center around the manual and complex nature of data handling in the current process. These issues could severely affect data integrity (for example, an incorrect macro execution could lead to an incomplete removal of access, undermining the integrity of the review) or the continuity of the process (e.g., if Excel crashes under load, the review cannot be completed on time).

The medium-criticality issues are more numerous; they highlight many inefficiencies and moderate risks, from reliance on others for data to the lack of automation in spotting anomalies. While each medium issue on its own may not be critical, collectively they contribute to significant operational overhead and some degree of unmanaged risk.

The low-criticality issues are mostly minor quirks or rare occurrences that have minimal impact on the overall objectives. They should eventually be addressed for completeness, but they do not demand urgent action.

In conclusion, this data understanding and fragility analysis reveals that the current access governance process, though functional, suffers from fragmentation and manual complexity. Data is scattered across systems and requires extensive manual compilation, which jeopardizes efficiency and accuracy. The identification of these fragilities (and their root causes) will inform the subsequent phases of the project. In the following chapters, we will focus on addressing these issues — aiming to automate data extraction, streamline transformation (perhaps through a more robust ETL or data mesh approach), and enhance visualization, so that the process becomes more reliable, scalable, and aligned with best practices in access governance.

4.1.3 Data Preparation

This phase encompassed all necessary steps to transform raw, fragmented, and multi-source data into a structured and analyzable form suitable for identity and access risk monitoring. It was designed to eliminate pre-existing process fragilities, including reliance on external teams for data provision, manual data collection from disparate systems, inconsistent data formats, and the absence of monthly process automation. These issues occasionally resulted in delays and, in some instances, prevented the process from being executed on a monthly basis. The pipeline was implemented in Python using libraries such as pandas, sqlalchemy, oracledb, and imaplib, ensuring compatibility with both structured database environments and unstructured input sources.

The data preparation process was structured according to a modular ETL logic: data extraction, data cleaning and preprocessing (transformation), and data structuring for analysis. Each component of the pipeline directly addressed specific limitations identified in the *As Is* process.

Figure 4.2 illustrates the final BPMN model developed to represent the restructured data preparation process, including the extraction, transformation, and loading stages described in the following subsections.

Data Extraction

The extraction layer was responsible for consolidating data from three distinct Oracle production environments, DW, Database 2, and Database 3, each containing tables relevant to identity verification and access attribution. Secure authentication protocols and Python-based

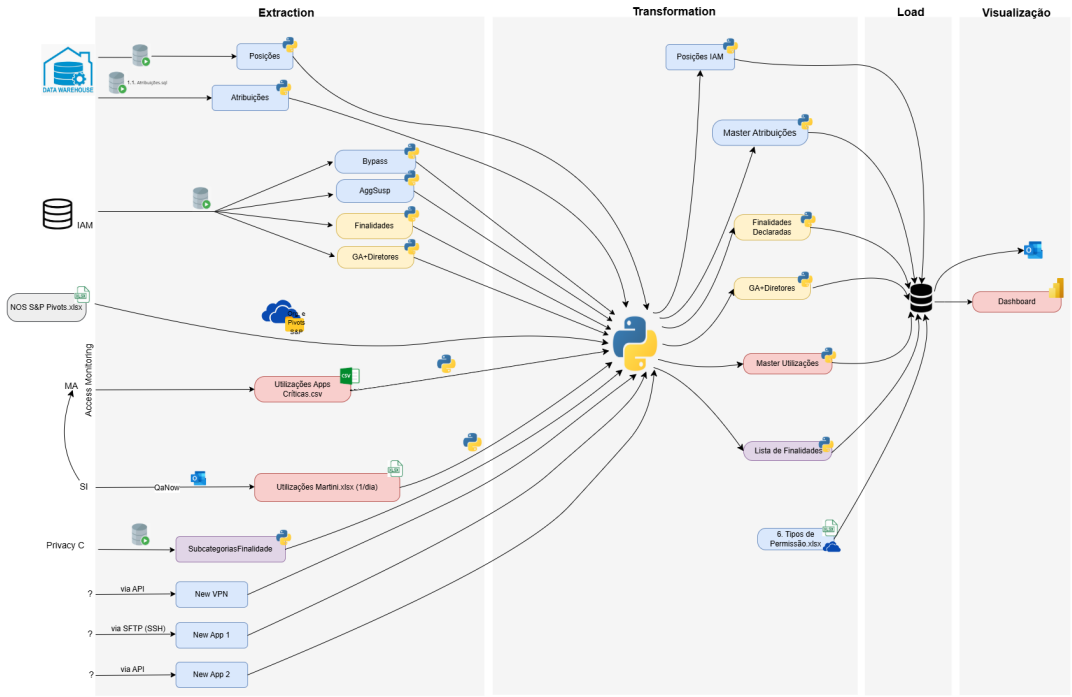

Figure 4.2: BPMN model of the data preparation process

connections enabled direct access to these environments, eliminating the need for manual data provision by intermediary teams, particularly from the AS and DSI departments. This configuration reduced operational dependencies, improved execution timeliness, and ensured access to updated records at the time of each processing cycle.

In addition to SQL-based extraction, the pipeline incorporated both API and SFTP-based integration mechanisms. API-based routines were developed to incorporate platforms that were not previously included in the monthly analysis, such as New App 1 and New VPN, involving the use of custom request headers, pagination management, and JSON parsing. For Salesforce, integration was established via SFTP using an SSH key pair. A Python script, based on the paramiko library, authenticated against the remote server, navigated to the designated export directory, and retrieved structured CSV files containing user, role, and permission information. This procedure replaced the need for manual file transfers or external coordination.

Another component of the extraction layer involved the automated retrieval of daily Excel files from multiple dedicated email inboxes. In one of these inboxes, a Power Automate workflow had been previously implemented to download attachments based on predefined conditions. However, this workflow lacked completeness checks and did not notify users when expected files were missing. The remaining inboxes were subject to fully manual procedures, where the analyst was responsible for locating, downloading, and renaming files individually. To address these limitations, the pipeline incorporated a Python-based validation routine that programmatically compared the list of expected files with those retrieved. The

script generated reports indicating missing dates and filenames, thereby enhancing traceability and enabling timely corrective actions. Using the imaplib and email libraries, the pipeline authenticated against the mail server, filtered messages by subject line and sender domain, and downloaded attachments into organized local directories. This implementation fully automated the retrieval process across all inboxes, standardizing and replacing previously inconsistent and manual practices.

All extraction routines were parameterized through dynamic computation of current_week and previous_week variables, determined at runtime: these parameters governed file retrieval ranges, SQL query filters, and output labeling. The removal of hardcoded references to specific dates or filenames enabled the pipeline to automate month-over-month operations and reduce the likelihood of user-induced errors.

Transformation

The transformation phase was responsible for converting heterogeneous raw datasets into a unified, analysis-ready structure. This process addressed structural inconsistencies, missing or ambiguous values, duplicate values, and the need for semantic enrichment. Transformations were implemented using Python and tailored for the characteristics of each data source while maintaining a consistent framework across systems. The logic was designed to enforce schema harmonization, identity resolution, time-aware tagging, and classification of records for downstream aggregation and reporting. The following subsections detail the most significant transformation components.

Building upon the need to address application-level access classification, a set of structured rules was developed to classify each VPN log according to the actual application being accessed. The transformation logic began by mapping the Permissão field into standardized values for the Aplicação field. This was achieved through a conditional mapping strategy where specific permission labels were translated, aligning them with business-relevant categories used in reporting.

An essential aspect of this transformation involved defining how VPN utilization should be quantified. Instead of relying on the external and opaque logic previously implemented by the AS team in Lavastorm, the logic was redefined internally in Python. The new process counted utilization events per user and day, eliminating overcounting caused by multiple log entries within the same day. Furthermore, the dataset was filtered to include only human-triggered access events, excluding system processes, service accounts, and anomalous usage patterns. These filters were tailored for each source system to ensure the inclusion criteria captured only valid interactive sessions.

This redesign addressed prior fragilities such as misclassification of automated connections and inconsistent definitions of what constituted a valid usage. By internalizing and reimplementing the logic, the process removed external dependencies, enhanced transparency, and standardized VPN utilization metrics across monthly cycles. This change also mitigated a documented fragility related to the dependency on third parties, particularly the AS*** team, by enabling direct access to the necessary data and internalizing the logic required for accurate processing. Furthermore, by executing these processes entirely within Python, the solution avoided Excel-based capacity limitations. It reduced the computational burden associated with legacy VBA workflows, contributing to greater scalability and efficiency.

To further ensure the accuracy of the dataset and to mitigate the fragility associated with duplicated access logs inflating utilization counts, duplicate records were identified using com-

posite keys defined per dataset. For the utilization logs, duplicates were determined based on the combined values of APP_USER, Permission, Aplicação, and REQUEST_DATE. This approach ensured that each row represented a unique usage event. Any repeated entries with identical values across these four fields were dropped, and the number of removed duplicates was logged per file. This procedure was necessary to avoid inflating utilization metrics due to multiple log entries that described the same access event.

In addition to structural transformations and to address fragilities related to partial or inconsistent user identifiers across systems, a key transformation step involved reconciling identity fields across systems. Email addresses were used as standard keys to enrich records with attributes retrieved from the IAM dataset. Where usernames were missing or mismatched, partial joins and derived lookups were implemented. For example, one of the datasets contained login data with only partial identifiers, which were resolved through mappings with the IAM directory. When exact matches were not possible, fallback logic based on email domain inference was triggered, as described earlier. This strategy mitigated the fragility associated with manual treatments and non-standardized identity reconciliation procedures. Similarly to the VPN transformation logic, it contributed to resolving limitations inherent in Excel-based and semi-manual processing environments.

To support time-based analysis and to eliminate the fragility caused by manual intervention in weekly dataset definitions, two key temporal indicators, current_week and previous_week, were programmatically defined using the date of script execution. These variables dynamically filtered datasets and named outputs, thereby removing the need for manual date entry. The logic was based on ISO week definitions, and the current_week indicator was only assigned if the latest ISO week was complete. If the most recent week was still in progress at the time of execution, it was excluded from processing to ensure consistency and to avoid partial aggregations. Additionally, records were tagged with change flags (e.g., reused access, revoked access) using Boolean indicators. In some applications, rows were duplicated and annotated when multiple permission interpretations applied. This facilitated downstream risk aggregation without loss of specificity.

Finally, to preserve data quality while managing edge cases, and to address fragilities associated with unstructured or incomplete source files, records that failed identity reconciliation or presented structural inconsistencies (e.g., unexpected column orders or encoding issues) were retained with Unknown or Unmatched tags. These were explicitly logged and not discarded, ensuring auditability. This approach avoided data loss while making anomalies visible for manual review.

Overall, the transformation logic addressed previously documented fragilities, including dependency on manual classification, misaligned schemas, and partial identifiers. The resulting dataset was consistently structured, semantically enriched, and validated for completeness and coherence.

Load

The final stage of the data preparation process involved loading the cleaned and transformed datasets into dedicated tables within the organization's data warehouse environment, specifically the WB schema on the DW database. This marked the transition from Python-based transformation to persistent storage accessible by the risk and compliance dashboards.

To support modular analysis and ensure semantic coherence, a set of custom tables was created within the data warehouse. These include: SUSPENSOS, FINALIDADES,

MANAGERS, MASTER_UTILIZACOES, MASTER_ATRIBUICOES, as well as supporting tables such as New App 1, New VPN, New App 2, and Role Info. Each table was designed to represent a specific component of the access and utilization data pipeline.

Specifically, the table MASTER_UTILIZACOES stores unified user activity logs across systems, whereas MASTER_ATRIBUICOES consolidates access assignments by application and profile. In parallel, the SUSPENSOS table maintains a register of aggregators that have been suspended from users. Furthermore, the FINALIDADES table captures the declared purpose associated with each permission, which supports downstream classification. The MANAGERS table, on the other hand, contains records of the designated access managers and directors who receive the outcome of the process—namely, the Access Circularization dashboard. In terms of VPN-related activity, the New VPN table stores relevant log events. Finally, the tables New App 1, New App 2, and Role Info retain critical access assignments originating from external applications that are not monitored internally.

This modularized structure facilitated targeted refresh strategies and simplified both validation and maintenance processes.

To operationalise the data loading phase into the Oracle Data Warehouse, a dedicated Python module named batch_insert.py was developed. This module encapsulates all logic related to the persistence of treated data, supporting the Identity and Access Management (IAM) dashboard development. Its primary purpose is to structure and automate the population of a set of target tables with cleansed and transformed data, while ensuring consistency, scalability, and compliance with the requirements of the Risk and Compliance department.

Each function in batch_insert.py corresponds to a specific table in the WB Oracle schema. These functions are responsible for preparing the data, performing cleanup of existing records when necessary, and executing the actual insertion. The module uses SQLAlchemy and Oracle's oracledb connector to ensure efficient and secure communication with the database, leveraging batch operations wrapped in transactional contexts to guarantee atomicity.

Before insertion, each dataset undergoes a set of harmonization steps to ensure compliance with Oracle database constraints:

- Conversion of temporal fields to standard datetime formats;
- Explicit casting of textual columns to strings;
- Replacement of missing values with None to ensure compatibility with Oracle's NULL.

These steps are relevant to prevent type mismatches, ensure consistent behaviour across different environments, and uphold data integrity.

Two main deletion strategies are implemented depending on the nature of each table:

- **Full refresh:** For snapshot-style tables, such as SUSPENSOS, MANAGERS, MASTER_UTILIZACOES, MASTER_ATRIBUICOES, all existing records are deleted prior to insertion. This ensures that the table always reflects the latest validated dataset.
- **Incremental replacement based on temporal cutoffs:** For time-sensitive tables (e.g., those supporting weekly tracking of New VPN, New App 1, and New App 2), only records older than the most recent completed ISO week are deleted. This dynamic threshold is computed at runtime to avoid processing data from an incomplete week, thereby preserving the stability and consistency of the dashboard's historical record.

Insertions are performed in batches within transactional contexts to ensure that each operation is atomic. If any part of the batch fails, the transaction is rolled back to avoid partial data persistence. Batch-level error reporting is also enabled, allowing the system to log and isolate failed rows without halting the entire operation. This design enhances robustness, enables faster error resolution, and ensures continuity of the ETL process even in the presence of individual data anomalies.

The module is not executed independently; it is imported into the notebooks that coordinate the ETL pipeline. After data treatment is completed, the relevant insertion functions are called with the cleaned datasets and necessary parameters. This approach reinforces modularity, avoids code repetition, and separates the transformation logic from the persistence layer.

The decision to isolate the loading logic in a dedicated module was driven by several methodological and technical considerations:

- **Separation of concerns:** Decouples data persistence from transformation and visualisation logic, enabling cleaner architecture.
- **Code reusability:** Insertion functions can be reused across multiple workflows, reducing redundancy.
- **Maintainability:** Schema or logic updates are centralised, reducing the likelihood of errors or inconsistencies.
- **Auditability:** Enables traceable and repeatable load processes, which are critical for compliance and debugging.
- **Transactional safety:** Ensures that only complete and successful operations are committed to the database.
- **Performance optimisation:** Batch inserts reduce execution time and system overhead compared to row-by-row operations.
- **Error traceability:** Facilitates the identification of problematic records without interrupting the entire pipeline.

The load process addressed several structural fragilities present in the previous solution, by replacing Excel-based storage with relational tables, the pipeline eliminated the limitations associated with manual file handling and VBA macros. The use of Python ensured that loading procedures could scale to larger datasets without breaching Excel's computational or row capacity. The automation of table refreshes also reduced the risk of omission or delay, promoting timely and consistent updates to downstream dashboards and compliance checks.

4.1.4 Modelling

The modelling phase focused on redesigning and structuring the analytical layer that supports the access governance dashboard. While the dashboard already existed within the organisation, its previous configuration exhibited several weaknesses, particularly regarding usability and data navigation. Users faced difficulties locating relevant use cases and identifying which accesses required action. This hindered operational efficiency and compromised the

oversight function of the access managers.Screenshots of the original dashboard configuration are provided in Appendix A.1, with sensitive elements omitted for confidentiality.

The dashboard is supported by a relational star-schema model developed in Power BI. The schema comprises two primary **fact tables**, Utilizações and Atribuições, which respectively record logs of access usage and access attributions as shown in Figure 4.3.

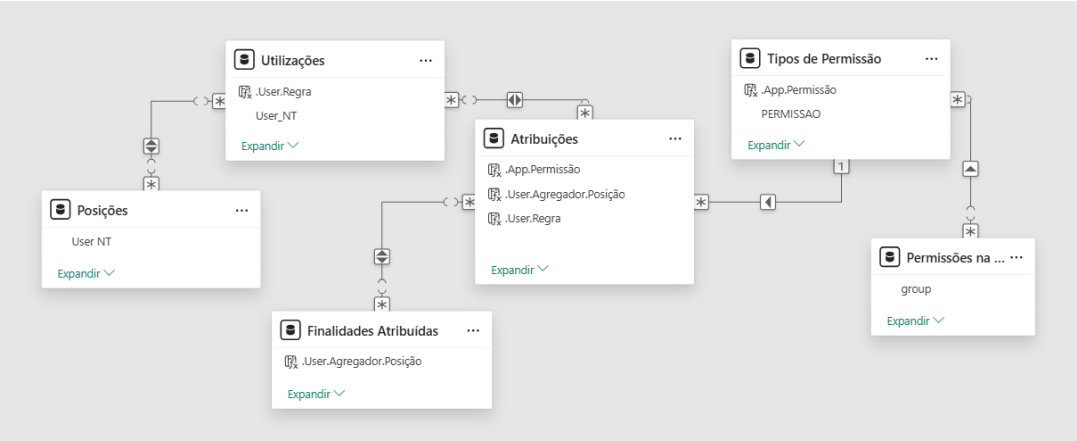

Figure 4.3: Relational star-schema model implemented in Power BI. The model centres on two fact tables, Utilizações and Atribuições, surrounded by multiple dimension tables supporting filtering, classification, and semantic enrichment.

These fact tables are connected to several **dimension tables** that provide business context, filtering capabilities, and categorical classification.

The Utilizações fact table contains event-level data on user interactions with various applications. It uses User_NT and User.Regra as foreign keys to enable integration with dimension tables. Specifically, it connects to:

- Posições, in a many-to-many relationship, as a single user may hold multiple concurrent or sequential positions within the organisation. This structure allows access usage data to be analysed within the scope of all roles attributed to each user. The connection is established via User_NT, which provides organizational metadata such as team, area, and department.

The Atribuições fact table records access assignment events. It includes the fields App.Permissão, User.Agregador.Posição, and User.Regra as composite keys. This table connects to:

- Tipos de Permissão, through the key App.Permissão, in a many-to-one relationship, as each access assignment corresponds to a single permission type, while each permission type may be associated with multiple assignments. This table provides a categorical classification of permissions based on predefined business rules.

- Finalidades Atribuídas, via User.Agregador.Posição, in a many-to-many relationship, as both access assignments and declared business purposes can reference the same aggregator-position combination multiple times, each linked to distinct permissions. This linkage enables the enrichment of attribution records with declared business purposes.

The Posições table functions as a user-related dimension table, with User_NT as its primary key. It supports filtering and aggregation of both usage and attribution data by organisational unit.

The Tipos de Permissão table serves as a categorical dimension, offering classification metadata for each unique App.Permissão key. It has a one-to-many relationship with the Atribuições fact table.

The Finalidades Atribuídas table defines the intended business purpose associated with each access assignment. It is used to support justification validation processes during access reviews and is linked to Atribuições in a many-to-many relationship, as detailed above.

The Permissões na VPN 2 table contributes VPN-specific metadata, particularly relevant for remote access audits. It connects to Tipos de Permissão through the PERMISSAO field, forming a many-to-one relationship, as several VPN-specific configurations may correspond to a single permission type.

These relationships, defined through explicit primary and foreign keys, ensure robust filter propagation and referential integrity across the model. Directional filtering is applied to maintain consistent aggregation logic throughout the visual layer. Notably, the connection between Atribuições and Utilizações enables the integrated analysis of assigned versus effectively used permissions, supporting the identification of anomalies such as overprovisioning, underutilisation, or policy violations.

4.1.5 Dashboard Architecture and Navigation Redesign

To address the usability limitations of the initial dashboard, this project introduced significant redesigns focused on enhancing interactivity, information relevance, and operational efficiency for access managers. Although the dashboard had already been implemented within the organization, its initial configuration was fragmented and difficult to navigate. Each analysis page functioned largely in isolation, offering limited integration between related metrics or use cases. Users were required to manually apply filters to explore different dimensions, such as types of permissions or specific applications, which often resulted in inconsistent analyses and a high cognitive load.

Moreover, the dashboard contained some purely descriptive Key Risk Indicators (KRIs) that presented static counts or distributions but did not signal risk-relevant deviations or trigger follow-up actions. This design limited the dashboard's effectiveness as a decision support tool and hindered the ability of access managers to identify and address anomalous behaviours proactively. As the scope of access types grew, particularly with the expansion of permission categories under analysis, visual overload further reduced clarity. Previously manageable visualizations became saturated, making it difficult to distinguish between critical and non-critical elements without extensive manual filtering.

To address these limitations, a comprehensive redesign of the navigation logic, visual prioritization, and interactivity mechanisms was implemented. The following subsections present the navigation flow developed to streamline exploration, the redefinition of KRIs to align with business risk, and the interactive mechanisms introduced to allow targeted, context-aware analysis.

As part of the improvements implemented, the circularization process was restructured to run on a weekly basis instead of monthly. This change allows for more frequent monitoring and faster risk mitigation. Consequently, since NOS chose not to retain historical

data, indicators based on 30-day and 60-day non-utilization were excluded from the current visualizations.

To overcome the previous issues of fragmentation, lack of page interconnection, and reliance on manual filtering, a structured navigation system was implemented using Power BI bookmarks, dynamic buttons, and contextual page filters. In the previous version of the dashboard, each page operated in isolation, with no intuitive flow between views. As reported by access managers, one of the main difficulties was identifying the specific use cases referenced in the monthly compliance and risk monitoring emails, as these were not traceable within the dashboard's structure.

The redesigned navigation enables users to explore the dashboard intuitively, beginning with high-level overviews and progressively drilling down into detailed analysis pages based on the type of permission or area of concern. As shown in Figure 4.4, the navigation flowchart illustrates the interaction logic between the overview page and the detailed analytical sheets. This overhaul significantly improved usability, reduced the time required to locate relevant information, and aligned the dashboard structure with the actual workflows of the access management team.

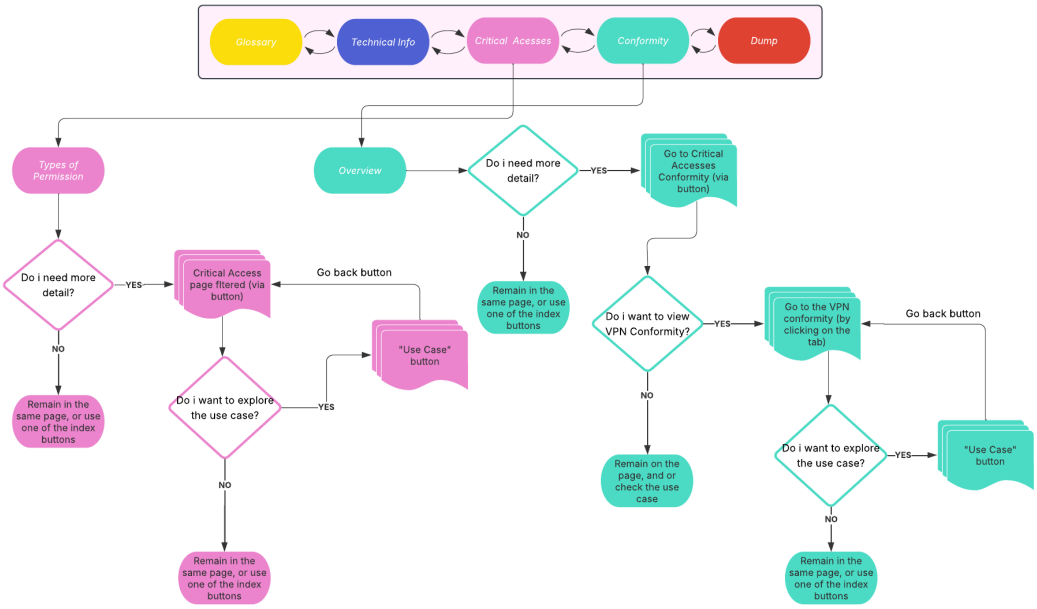

Figure 4.4: Navigation flowchart illustrating the interaction logic between the main overview page and the detailed analytical sheets. Buttons and bookmarks are used to enable seamless transitions across layers of analysis, while maintaining the selected filtering context.

An interactive navigation bar is present on every page of the dashboard, providing direct access to five main sections. The button corresponding to the currently viewed section is displayed in white, indicating its active state, while the remaining buttons appear in gray. When the user hovers over an unselected button, it temporarily turns white and reveals the name of the corresponding section, thereby enhancing navigability and reinforcing contextual awareness. The user journey begins on the **Glossary** page, which provides definitions and contextual information for the key concepts used throughout the dashboard. The five main sections accessible through the navigation bar are as follows:

- **Glossary**
- **Technical Sheet**
- **Critical Accesses**
- **Conformity**
- **Data Dump**

The **Glossary** and **Technical Sheet** sections are informational only and do not include any further interactions.

Selecting the **Critical Accesses** button redirects the user to the **Permission Types** page, where cards display indicators for each type of critical permission. If a card reveals unusual values or if the user wishes to explore a specific permission type in more detail, they may click the *More Detail* button on that card. This action navigates to the **Critical Accesses Detail** page, with a bookmark applied so that the page is automatically filtered to the selected permission type, improving contextual navigation.

Within this page, users may also access the corresponding *Use Case* page, which provides a detailed description of the access scenario. The Use Case page includes a return button to the Critical Accesses Detail page, as well as a bookmark that removes the bookmark filter, allowing the user to view the full unfiltered page if desired. To return to the Permission Types page, users must use the main navigation bar available across all sections.

The **Conformity** section is accessed via the main navigation bar and presents a high-level overview of conformity indicators. From there, the *More Detail* button leads to the **Critical Access Conformity** page. This page includes two tab-like sub-sections that allow the user to switch between:

- **Critical Access Conformity**
- **VPN Conformity**

Both of these subsections contain links to their respective **Use Case** pages, which also include return buttons. As with other sections, navigation to different areas of the dashboard must be done using the main navigation bar.

Finally, the **Data Dump** button provides access to a page with more granular information for technical or analytical exploration.

During the dashboard redesign, several Key Risk Indicators (KRIs) were deliberately removed based on their limited analytical value. For instance, the *Top 5 Finalidades* KRI was discarded as it was purely descriptive, failed to indicate any risk-relevant change, and unnecessarily occupied visual space.

Furthermore, an earlier version of the dashboard included a page titled *Âmbito*, which presented a high-level overview of global, critical, and remote access KRIs. This page was removed following the adoption of the *Permission Types* approach, which now functions as an entry point for critical access information. The Conformity section also includes an overview that effectively serves a similar purpose for compliance-related metrics. Several of the KRIs in the original *Âmbito* page were also found to be purely informational and not actionable, and were therefore excluded from the final version to improve focus and reduce noise.

These removals were guided by the principle of prioritizing indicators that support decision making, reflect meaningful variation in risk posture, and align with the overall objective of access governance.

4.1.6 Evaluation

The evaluation phase aimed to determine whether the implemented dashboards effectively responded to the operational challenges initially raised by the business teams responsible for access governance. Although no formal validation framework was applied, the assessment was based on feedback gathered from the same teams that had previously reported limitations in access conformity monitoring, privilege tracking, and overall data visibility.

Following the initial deployment, informal review sessions were conducted with stakeholders, including access managers and compliance personnel. During these sessions, the dashboard structure, navigation flow, and key metrics were presented. Particular attention was given to the usability of the interface, the clarity of the visual components, and the relevance of the information displayed in the context of their day-to-day responsibilities.

The feedback received from these teams was positive. Users reported that the dashboards significantly improved visibility over critical access assignments, remote access utilisation, and conformity statuses. The Types of Permissions page was identified as especially useful, as it allowed users to navigate intuitively to the corresponding critical access page with the selected permission type already filtered. This structure facilitated faster access to detailed information and reduced the cognitive effort required to locate relevant data. The use of colour coding to distinguish types of permissions further supported prioritisation by drawing attention to the most sensitive access categories.

While a structured evaluation using predefined criteria or usability testing instruments was not conducted, the favourable reception of the dashboards and their continued use in recurring access review activities suggest that the implemented solution met the primary objectives.

4.1.7 Deployment

The final phase of the CRISP-DM methodology consists of deploying the analytical solution in a way that enables stakeholders to access, interpret, and act upon the results. In the context of the IAM dashboard project, deployment involved making the visual interface accessible to designated access managers, guiding them toward relevant use cases, and ensuring that the output supported operational decision-making and compliance monitoring.

The first step in the deployment process was the identification of the relevant stakeholders responsible for interpreting and acting on the information presented in the dashboard. These stakeholders, referred to internally as *access managers*, were extracted from an internal reference list stored in the data warehouse. The list includes their names, roles, teams, and institutional email addresses.

To operationalise access, a filtered view of this stakeholder list was generated and exported in an accessible format. Email addresses were used to notify access managers of the availability of the dashboard and to provide targeted guidance regarding the specific use cases or risk areas relevant to their scope of responsibility. This ensured that each user was aware of the dashboard's existence, purpose, and utility in the context of their function.

Given the complexity of access structures across the organisation, a mapping between use cases and managerial responsibilities was established. Access managers overseeing users with critical permissions were directed to validate the necessity and legitimacy of those assignments. Those responsible for remote access permissions were instructed to review utilisation logs to identify inactive or improperly configured accesses. Finally, stakeholders involved in the

enforcement of VPN policies were guided to the *New VPN* conformity section to ensure alignment with the mandatory usage policy. This targeted approach improved the relevance and usability of the dashboard, ensuring that each stakeholder could directly engage with the visual components most pertinent to their area.

This targeted approach improved the relevance and usability of the dashboard, ensuring that each stakeholder could direy engage with the visual components most pertinent to their area.

Once the stakeholder communication and use-case alignment were completed, the dashboard was published on the internal Power BI service. Access was restricted to verified users using role-based access control, and permissions were aligned with the institutional identity management system. The dashboard included a glossary and technical sheet to facilitate interpretation and to minimise miscommunication regarding metric definitions or data sources.

The deployment process was concluded by validating that all access managers were able to log in successfully and visualise their respective sections without technical impediments. Feedback loops were established through direct communication channels, allowing for future adjustments or enhancements based on stakeholder needs.

Chapter 5

IAM Dashboard

The IAM dashboard constitutes the practical outcome of the integration and analysis work described in the previous chapters. Its purpose is to provide a consolidated and interactive environment where stakeholders can monitor access risks, track conformity indicators, and explore detailed information on identity and permission management. This chapter introduces the dashboard's design principles, key functionalities, and navigation logic, illustrating how visual elements were structured to align with risk governance priorities and to support more informed decision-making.

5.1 Introduction of Criticality-Based KRI Cards

At the beginning of the project, the number of permission types under analysis was relatively limited. This allowed for simple visualizations where users could easily interpret variations across all permissions in a single view. However, as the analytical scope expanded, so did the volume and diversity of permission types, resulting in increasingly dense visuals that were no longer effective for identifying risk-relevant changes.

In the previous implementation, users were required to manually interact with the page filter, selecting each permission type individually, to detect potential anomalies or spikes in access. This process was time-consuming, unintuitive, and highly dependent on the user's initiative and attention.

To address this issue, a new page was developed featuring KRI cards organized by permission criticality. The classification system follows the internal *Política de Segurança da Informação* (PSI), which assigns a criticality level from 1 (most critical) to 4 (least critical) to each permission. Only permissions classified at levels 1 and 2 were retained for continuous monitoring within the dashboard, thereby reducing noise and enhancing analytical focus in line with risk governance priorities.

Each card corresponds to a distinct type of critical permission and displays key indicators, including the number of users with that type of permission in the current cycle, the number of users in the previous cycle, the percentage variation between the two, and a directional arrow. This arrow points upward if the variation is positive and downward otherwise; its color is green when the variation is negative (indicating fewer users with a critical permission) and red when it increases. This visual mechanism is based on the principle that a reduction in the number of users with critical permissions represents a lower exposure to risk, and thus a positive outcome. The red-highlighted cards draw the user's attention to areas where risk

exposure may be increasing and where review may be necessary.

When a user identifies a card with concerning values (red arrow) or seeks additional context, they may click the *More Detail* button, which redirects to a filtered view of the detailed dataset for that specific permission type. This filter is applied using Power BI bookmarks, ensuring contextual consistency and streamlining the analytical flow.

Furthermore, users may opt to view the entire unfiltered dataset by selecting the *Ver Todos os Tipos de Permissões* bookmark, which removes the permission-specific filter and reveals the complete set of critical access data. This hybrid approach balances targeted analysis with complete visibility, depending on the user's intent.

This new page not only replaces the previous *Âmbito* overview for critical accesses but also reflects a scalable, risk-aligned design that improves interpretability, reduces manual effort, and supports proactive governance.

5.2 Development of Custom Indicators for Permission Variation

To enhance the analytical value of the dashboard, a set of dynamic indicators was developed to monitor changes in the number of users assigned to each critical permission type over time. These indicators were implemented in Power BI using DAX (Data Analysis Expressions) and aimed to simplify trend detection and guide users' attention toward risk-relevant variations.

The first measure calculates the percentage change in the number of users with a given permission type between two reference cycles, labeled "Current Week" and "Previous Week." This variation is only shown when the previous cycle contains a valid value, ensuring that anomalous results due to missing data or division by zero are avoided.

To improve interpretability, a directional arrow is shown alongside each variation: a red upward arrow indicates an increase in users (generally signalling heightened access risk), a green downward arrow reflects a decrease (indicating reduced exposure), and a horizontal arrow is displayed when no significant variation occurs.

In parallel, the color of the percentage value is dynamically adjusted to reinforce the risk context: red for increases in critical permissions and green for decreases. This visual encoding supports intuitive interpretation and helps users immediately identify which permission types require closer attention.

Together, these enhancements transformed the dashboard from a static reporting tool into a dynamic governance instrument. The integration of context-specific Key Risk Indicators (KRIs) enhances interpretability and responsiveness, allowing access managers to identify relevant changes more effectively. The design places emphasis on usability, ensures alignment between visual elements and operational risk priorities, and promotes more proactive and informed decision-making within the access management process.

5.2.1 Dashboard Analysis

To illustrate the implemented redesign and support the previously described architecture, several examples from the final Power BI dashboard are presented below. These visual excerpts demonstrate the navigation mechanisms, the restructured analytical views, and the risk-oriented Key Risk Indicators (KRIs) that guide user interactions.

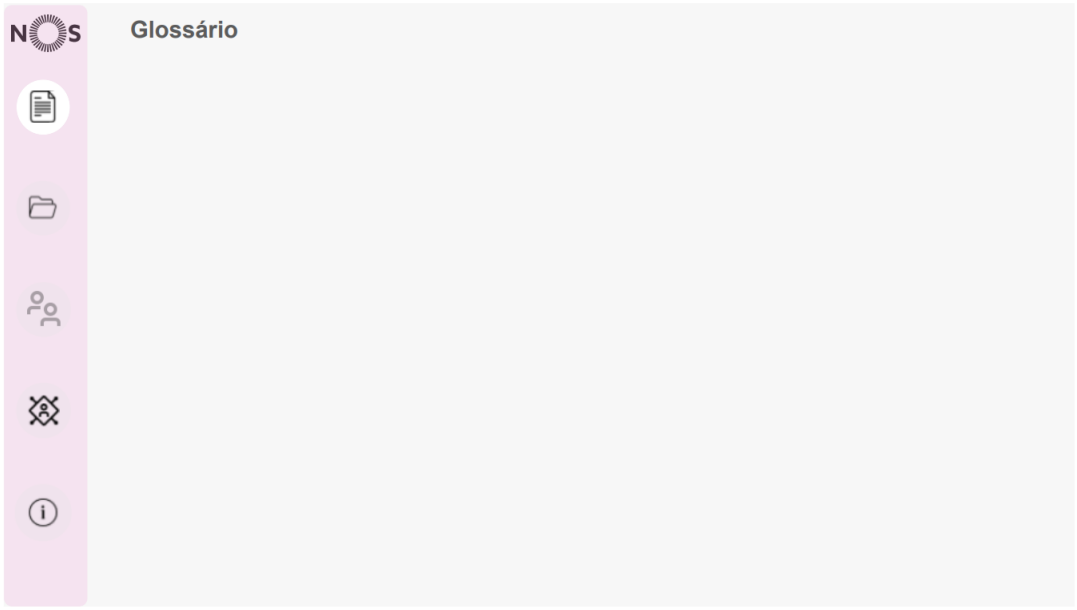

Figure 5.1: Glossary page of the dashboard. This static page provides definitions for key terms used throughout the dashboard, helping users interpret indicators and fields consistently.

As shown in Figure 5.1, the Glossary page defines the key concepts used throughout the dashboard. This ensures consistent understanding across users and reduces ambiguity in interpreting the visual elements and indicators presented in later sections.

Figure 5.2: Technical Sheet page of the dashboard. This static section provides formal documentation on the dashboard's ownership, data sources, information classification, update date, and structure. It includes usage instructions and guidance on the interpretation of indicators and available support elements, such as tooltips and use case references.

As illustrated in Figure 5.2, the *Technical Sheet* page presents static documentation to support proper usage of the dashboard. It includes the date of the latest update, ownership, and data sources, and classifies the information as internal use only. It describes the dashboard

structure, specifying the summary indicators and the section dedicated to access conformity. It also provides general usage instructions, including the availability of contextual guidance on each page and a recommendation that the weekly action plan of each division be directed to the use cases where nonconformities are identified.

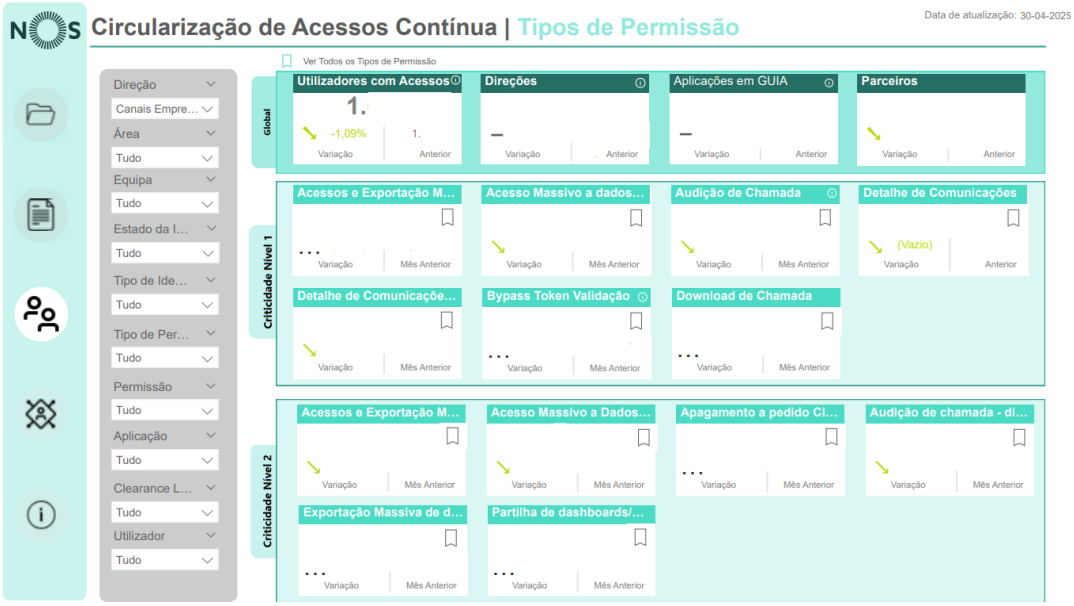

Figure 5.3: Overview of users with critical permissions, classified by permission type. The figure presents the number of users with assigned logical accesses of very high (top) and high (bottom) criticality, as well as the respective variation relative to the previous week. The indicators are filterable by multiple dimensions, including department, area, team, identity type, status, application, and clearance level.

The interactive KRI cards, shown in Figure 5.3, are organized by level of internal criticality. Each card shows the number of users with a specific permission type, the change relative to the previous cycle, and a directional indicator to support risk-oriented monitoring.

Figure 5.4 presents the detailed view of critical accesses, supporting granular analysis of access distribution for a specific critical permission. Filters on the left allow segmentation by direction, area, team, identity type, and other dimensions. The top section presents summary visualisations, including access counts per direction and application, as well as a breakdown of access types. The lower table provides line-level visibility over each access instance, enabling the user to examine attributes such as usage, assignment, and clearance level. The ”Use Cases” button at the bottom offers access to contextual examples for interpretation support.

The Conformity section begins with an overview page (Figure 5.5), which provides high-level indicators for both critical and remote accesses. The top section displays cards related to essential access control, such as partner access, profile-based access, access without justification, and coverage of access purposes. The bottom section addresses remote access groups. Each card presents the variation from the previous cycle to support ongoing compliance monitoring. The ”More Detail” button at the bottom directs the user to a dedicated page with in-depth conformity metrics specifically for critical accesses.

As depicted in Figure 5.6, the Critical Accesses Conformity page offers a detailed view of conformity indicators specific to critical accesses. The top section includes visualizations

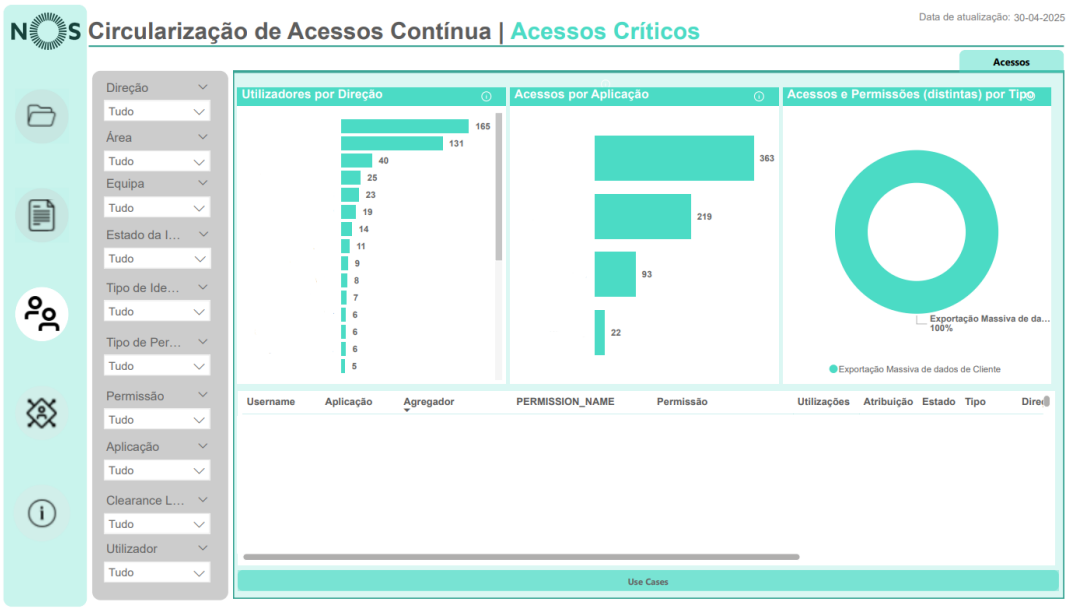

Figure 5.4: Detailed view of critical accesses, displaying the distribution of users by department (left), access counts by application (center), and distinct permissions by type (right). The table below provides granular information on each access, including user, application, permission, and attribution details. This view supports filtering across multiple identity and access dimensions.

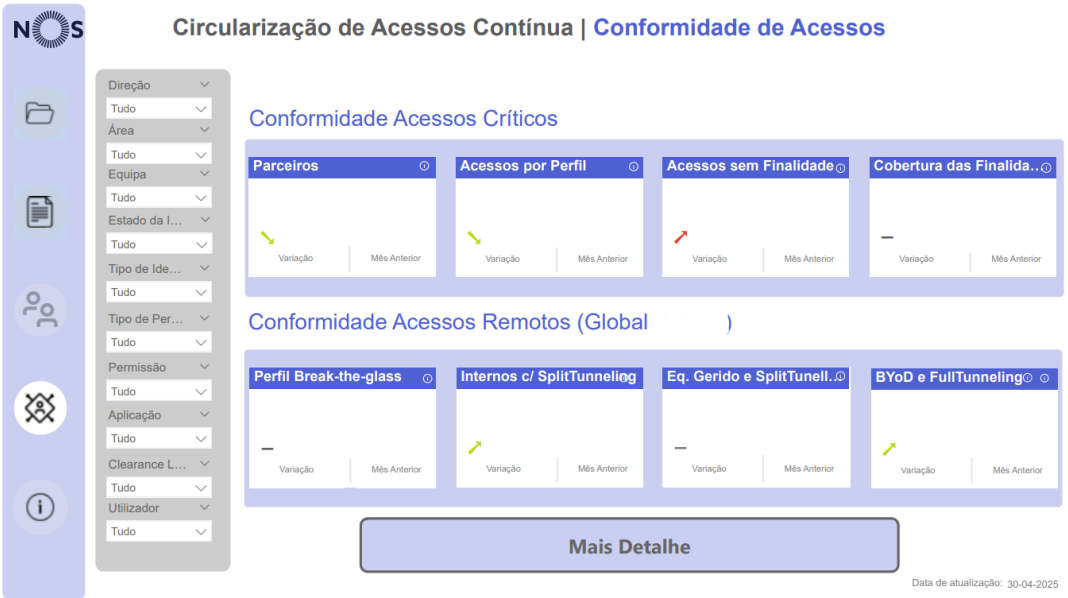

Figure 5.5: Overview page of the Conformity section, displaying high-level indicators related to the conformity of critical and remote accesses. The top section presents conformity metrics for critical accesses, such as partner accounts, profile-based accesses, accesses without a defined purpose, and purpose coverage. The bottom section focuses on remote access conformity at a global level. Each card indicates the variation relative to the previous week.

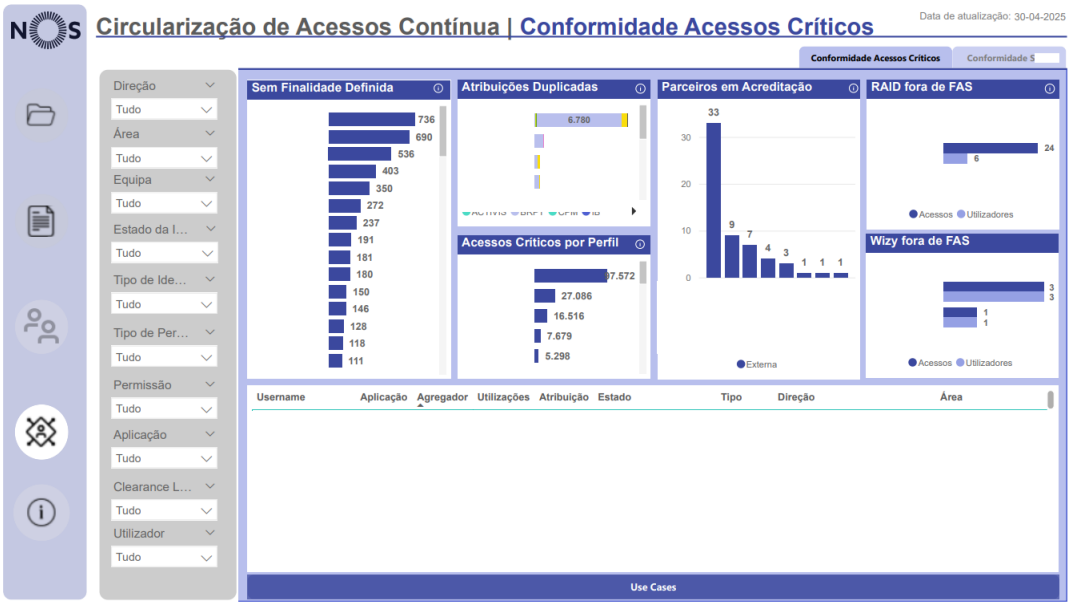

Figure 5.6: Detailed view of the Critical Accesses Conformity page. This section presents multiple non-conformity indicators, including accesses without a defined purpose, duplicated assignments, partners in accreditation, and profiles with a high volume of critical accesses. It also highlights applications with accesses not aligned with the FAS. The lower section provides a detailed table of individual records, enabling further analysis based on user, application, aggregator, and other access attributes.

such as accesses without defined purpose, duplicated assignments, and partners undergoing accreditation. A dedicated graph displays the distribution of critical accesses per profile. These indicators are designed to support the detection of non-compliant behaviours and enable timely follow-up. The lower table allows further inspection of each access instance.

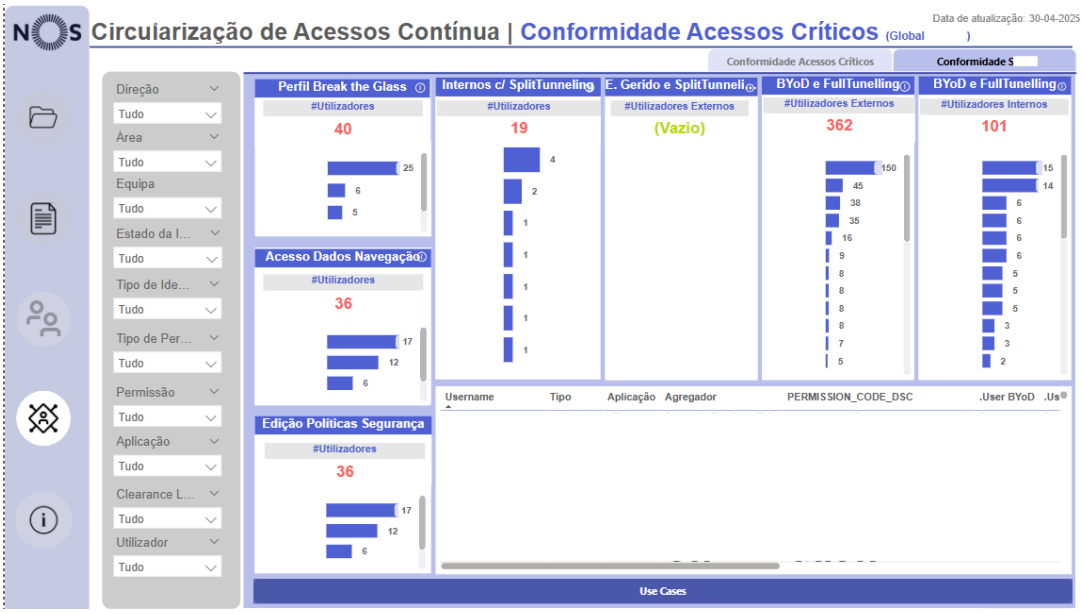

Figure 5.7: Remote Access Conformity page.

The Remote Access Conformity page (Figure 5.7) consolidates indicators control indicators related to remote access scenarios, structured around various user categories and permission types. It monitors the use of break-the-glass profiles, navigation data access, and security policy editing permissions. It also accounts for different VPN configurations, distinguishing between internal and external users. The interactive table below the visualizations provides granular details for each individual access case, enabling direct verification and follow-up.

Figure 5.8: Data Dump section.

Finally, Figure 5.8 shows the Data Dump section, which provides unrestricted access to the underlying dataset. It serves primarily as a technical exploration area for advanced users who require granular inspection or validation of access records.

Due to internal confidentiality policies, the use case pages accessible through the dashboard are not presented in this document.

Chapter 6

Conclusion

6.1 Final Remarks

This intership report proposed and implemented a solution to enhance access governance processes within the Audit, Risk, and Compliance department at NOS. The project addressed the limitations of the existing access review process, Circularização de Acessos, which was highly dependent on manual operations, fragmented data sources, and limited analytical capabilities.

The adopted solution is based on a data warehouse architecture that consolidates heterogeneous data related to identity and access management (IAM). Python-based Extract Transform Load (ETL) routines were developed to automate data integration and improve consistency, while Power BI dashboards were created to support interactive analysis and compliance monitoring. These dashboards present critical information on access assignments, usage patterns, and conformity metrics in a structured and user-friendly manner.

The proposed architecture aligns with the ISO/IEC 27001:2022 and ISO/IEC 27002:2022 standards, supporting compliance through improved traceability, reduced manual workload, and systematic monitoring of Key Risk Indicators (KRIs). The system enables more timely, scalable, and accurate risk assessments, thereby strengthening the department's capacity to ensure secure and proportional access control in accordance with internal policies and regulatory expectations.

Overall, this project demonstrated that the combination of data engineering, automation, and visual analytics can modernize and optimize IAM oversight in complex organizational environments.

6.2 Limitations and Future Work

Although the developed solution represents a significant advancement, several opportunities remain for further improvement. First, the implementation of formal version control and historical lineage within the analytical repository would enhance auditability and allow for more robust longitudinal analysis. This capability would be particularly valuable in tracking changes across access review cycles and identifying patterns over time.

Another promising direction involves the integration of automated anomaly detection and alerting mechanisms. These additions would strengthen the responsiveness of the monitoring system, allowing earlier identification of access violations or non-conformities and reducing

reliance on manual validation.

In future iterations, the architecture could also be expanded to cover additional access domains, such as physical access, privileged credentials, or application-level entitlements. This would increase the comprehensiveness of the monitoring framework and support a unified governance model across different access layers.

Finally, the integration of predictive analytics and machine learning models could provide proactive insights into emerging risks. By forecasting anomalies, identifying users with high-risk behavior profiles, or optimizing review prioritization, such models could further align IAM oversight with strategic risk management objectives.

These enhancements, while beyond the scope of the current project, represent natural extensions that would consolidate and expand the system's role in promoting secure, compliant, and efficient access governance.

Bibliography

Abadi, D., Boncz, P., & Harizopoulos, S. (2009, 08). Column-oriented database systems. *PVLDB*, *2*, 1664-1665. doi: 10.14778/1687553.1687625

Abadi, D. J., Madden, S. R., & Ferreira, N. (2006). Integrating compression and execution in column-oriented database systems. *Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data*, 671–682. doi: 10.1145/1142473.1142548

Abdullah, R., Doe, J., & Smith, J. (2021). Towards intelligent data warehouses: Integrating machine learning in the cloud. In *Proceedings of the international conference on cloud computing and big data* (pp. 100–110). Springer. Retrieved from https://link.springer.com/chapter/10.1007/978-3-030-12345-6_10 doi: 10.1007/978-3-030-12345-6_10

Ahmad, T., et al. (2021). Data-driven analytics leveraging artificial intelligence in the era of covid-19. *Symmetry*, *14*(1), 16. Retrieved from https://www.mdpi.com/2073-8994/14/1/16 doi: 10.3390/sym14010016

Akano, O. A., Hanson, E., Nwakile, C., & Esiri, A. E. (2024, 10). Designing real-time safety monitoring dashboards for industrial operations: A data-driven approach. *Global Journal of Research in Science and Technology*, *2*, 001-009. Retrieved from https://gsjournals.com/gjrst/node/67 doi: 10.58175/gjrst.2024.2.2.0070

Alhamadi, M. (2020). Challenges, strategies, and adaptations on interactive dashboards. In *Proceedings of the 28th acm conference on user modeling, adaptation and personalization (umap '20)* (pp. 368–371). Genoa, Italy: ACM. Retrieved from https://doi.org/10.1145/3340631.3398678 doi: 10.1145/3340631.3398678

Anton, S. G., & Nucu, A. E. A. (2020, 11). Enterprise risk management: A literature review and agenda for future research. *Journal of Risk and Financial Management*, *13*. doi: 10.3390/jrfm13110281

Armbrust, M., Ghodsi, A., Xin, R. S., & Zaharia, M. (2021). Lakehouse: A new generation of open platforms that unify data warehousing and advanced analytics. *Communications of the ACM*, *64*(9), 56–65. Retrieved from https://cacm.acm.org/magazines/2021/9/255707-lakehouse/fulltext doi: 10.1145/3448016

Aspin, A. (2020). *Pro power bi desktop: Self-service analytics and data visualization for the power user, third edition*. Apress Media LLC. doi: 10.1007/978-1-4842-5763-0

Associates, S. (2020). *Barc score: Enterprise bi and analytics platforms*. https://www.simpson-associates.co.uk/wp-content/uploads/2015/11/BARC_SCORE_Enterprise_BI_Analytics_Platforms_MUST-BE-GATED.pdf. (Accessed: 2025-05-23)

Bahaa, A., Eldemerdash, S. R., & Fahmy, H. (2021). A systematic literature review for implementing data ops in the data warehouse lifecycle during the etl phase. *Journal of Computer Science*, *17*(11), 1109–1121. Retrieved from https://thescipub.com/abstract/jcssp.2021.1109.1121 doi: 10.3844/jcssp.2021.1109.1121

Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and olap technology. *ACM SIGMOD Record*, *26*(1), 65–74.

Connolly, T., & Begg, C. (2014). *Database systems: A practical approach to design, implementation, and management* (6th ed.). Pearson Education.

COSO. (2017). *Enterprise risk management integrating with strategy and performance.* (Accessed: 29 December 2024)

Cuzzocrea, A., Song, I.-Y., & Davis, K. (2018). Hybrid transactional/analytical processing: A survey. *Data Science and Engineering*, *3*(4), 243–264. doi: 10.1007/s41019-018-0074-0

Dhaouadi, A., Bousselmi, K., Gammoudi, M. M., Monnet, S., & Hammoudi, S. (2022). Data warehousing process modeling from classical approaches to new trends: Main features and comparisons. *Data*, *7*(8). Retrieved from https://www.mdpi.com/2306-5729/7/8/113 doi: 10.3390/data7080113

Diouf, P. S., Boly, A., & Ndiaye, S. (2018). Variety of data in the etl processes in the cloud: State of the art. In *2018 ieee international conference on innovations in research and development (icird)* (pp. 1–6). Retrieved from https://www.researchgate.net/publication/325918454_Variety_of_data_in_the_ETL_processes_in_the_cloud_State_of_the_art doi: 10.1109/ICIRD.2018.8376295

Eltayeb, A. M. (2024). Enterprise identity and access governance in cloud environments: A strategic compliance approach. *Journal of Cybersecurity Governance*, *12*(1), 27–49.

Ferrari, A., & Russo, M. (2016). *Introducing microsoft power bi.* Microsoft Press.

Few, S. (2006). *Information dashboard design: The effective visual communication of data.* Sebastopol, CA: O'Reilly.

Gill, H., Smith, J., & Lee, A. (2019). Benchmarking on-premise vs. cloud-based mpp data warehouses. *Journal of Cloud Computing*, *8*(1), 1–15. doi: 10.1186/s13677-019-0145-3

Gill, H., Tuli, S., & Xu, L. (2020). Benchmark of market cloud data warehouse technologies. *Procedia Computer Science*, *175*, 1–8. Retrieved from https://www.sciencedirect.com/science/article/pii/S1877050920310321 doi: 10.1016/j.procs.2020.07.001

Glöckler, S., Schmidt, T., & Petersen, L. (2024). Iso/iec 27001 in practice: Implementing compliance-driven iam frameworks in german enterprises. *Information Security Journal: A Global Perspective*, *33*(2), 114–130.

Golfarelli, M., & Rizzi, S. (2009). *Data warehouse design: Modern principles and methodologies.* McGraw-Hill Education. Retrieved from https://archive.org/details/datawarehousedes0000golf

Gulisano, V., Jimenez-Peris, R., Patiño-Martínez, M., & Valduriez, P. (2012). Streamcloud: A large scale data streaming system. In *Proceedings of the 2012 ieee 28th international conference on data engineering* (pp. 126–137). IEEE. doi: 10.1109/ICDE.2012.20

Hammergren, T. C. (2009). *Data warehousing for dummies.* John Wiley & Sons.

Harinarayan, V., Rajaraman, A., & Ullman, J. D. (1996). Implementing data cubes efficiently. In *Proceedings of the 1996 acm sigmod international conference on management of data* (pp. 205–216). ACM Press. Retrieved from https://dl.acm.org/doi/10.1145/233269.233333 doi: 10.1145/233269.233333

Healy, K. (2018). *Data visualization: a practical introduction.* Princeton University Press.

Hopkin, P. (2017). *Fundamentals of risk management: understanding, evaluating and implementing effective risk management.* Kogan Page Publishers.

Hossain, M. M., Azim, T., Karim, Y., & Latiful Haque, A. (2009, 12). Integrated data warehousing for telecommunication industries.

doi: 10.1109/ICCIT.2009.5407317

Inmon, W. H. (2005). *Building the data warehouse* (4th ed.). Wiley.

International Organization for Standardization. (2022a). Information security, cybersecurity and privacy protection – information security controls [Computer software manual]. (ISO/IEC 27002:2022)

International Organization for Standardization. (2022b). *ISO/IEC 27001:2022 – Information Security, Cybersecurity and Privacy Protection — Information Security Management Systems — Requirements*. Geneva, Switzerland. (ISO/IEC 27001:2022(E))

ISO. (2018). *Iso 31000:2018 - risk management – guidelines*. Geneva, Switzerland: International Organization for Standardization.

Jensen, C. S., Pedersen, T. B., & Thomsen, C. (2010). Multidimensional database technology. In *Encyclopedia of database systems* (pp. 1800–1805). Springer.

Joubert, J., & Belle, J.-P. V. (2016). An innovation and risk dashboard. In *Mcis 2016 proceedings*. Retrieved from https://aisel.aisnet.org/mcis2016/14

Kandel, S., Paepcke, A., Hellerstein, J., & Heer, J. (2011). Wrangler: Interactive visual specification of data transformation scripts. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 3363–3372). ACM. doi: 10.1145/1978942.1979444

Karadata, A., Lopez, M., & Wang, C. (2020). Shared-data systems and workload management in distributed data warehouses. *Journal of Distributed Computing*, *34*(2), 150–165. Retrieved from https://www.sciencedirect.com/science/article/pii/S074373152030005X doi: 10.1016/j.jdc.2020.01.005

Kimball, R., & Ross, M. (2013). *The data warehouse toolkit: The definitive guide to dimensional modeling* (3rd ed.). Wiley.

Koudounas, K., Papadopoulos, G., & Ioannidis, S. (2024). Knowledge-based data warehouse architectures: Integrating ai for enhanced decision support. *Journal of Data Management and Analytics*, *12*(1), 45–60. doi: 10.1234/jdma.v12i1.5678

KPMG International. (2019). *Identity and access management: The foundation of digital trust* (Tech. Rep.). KPMG.

Levine, E., Brown, M., & Johnson, S. (2020). The evolution of data warehouse architectures in the cloud era. *Data Management Journal*, *29*(4), 45–60. Retrieved from https://www.sciencedirect.com/science/article/pii/S026840122030005X doi: 10.1016/j.dmj.2020.04.005

Linstedt, D. (2011). *Super charge your data warehouse: Invaluable data modeling rules to implement your data vault*. CreateSpace Independent Publishing Platform.

Lundqvist, S. A. (2015, 9). Why firms implement risk governance - stepping beyond traditional risk management to enterprise risk management. *Journal of Accounting and Public Policy*, *34*, 441-466. doi: 10.1016/j.jaccpubpol.2015.05.002

Macedo, M., & Zaina, L. (2024). Ux data visualization: Supporting software professionals in exploring users' interaction data. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (Vol. 14517 LNCS, p. 207-222). Springer Science and Business Media Deutschland GmbH. doi: 10.1007/978-3-031-59235-5_17

Microsoft. (n.d.). *Understand advanced data visualization concepts*. https://learn.microsoft.com/en-us/training/modules/understand-advanced-data-visualization-concepts. (Accessed: 21 December 2024)

Moalla, N., Chaabouni, T., & Ghozzi, H. (2017). A multi-fact constellation schema for a data

warehouse dedicated to agricultural production. *Procedia Computer Science*, *112*, 1501–1510. Retrieved from https://doi.org/10.1016/j.procs.2017.08.206 doi: 10.1016/j.procs.2017.08.206

Muskan, Singh, G., Singh, J., & Prabha, C. (2022). Data visualization and its key fundamentals: A comprehensive survey. In *7th international conference on communication and electronics systems, icces 2022 - proceedings* (p. 1710-1714). doi: 10.1109/ICCES54183.2022.9835803

Naik, S., & Prasad, C. V. V. S. N. V. (2021, 3). Benefits of enterprise risk management: A systematic review of literature. *GATR Journal of Finance and Banking Review*, *5*, 28-35. doi: 10.35609/jfbr.2021.5.4(3)

Nambiar, A., & Mundra, D. (2022). An overview of data warehouse and data lake in modern enterprise data management. *Big Data and Cognitive Computing*, *6*(4), 132. Retrieved from https://www.mdpi.com/2504-2289/6/4/132 doi: 10.3390/bdcc6040132

Oti-Frimpong, J., & Gyedu, S. (2023). Enterprise risk management (erm): A bibliometric review and future agenda. *International Journal of Recent Research in Commerce Economics and Management (IJRRCEM)*, *10*, 122-140. Retrieved from https://doi.org/ doi: 10.5281/zenodo.7785389

Po, L., Bikakis, N., Desimoni, F., & Papastefanatos, G. (2020). *Linked data visualization: Techniques, tools, and big data* (1st ed.). Switzerland: Springer Cham. (Published online: 31 May 2022; Hardcover published: 20 March 2020) doi: 10.1007/978-3-031-79490-2

Qin, X., Luo, Y., Tang, N., & Li, G. (2020). Making data visualization more efficient and effective: a survey. *VLDB Journal*, *29*, 93-117. doi: 10.1007/s00778-019-00588-3

Quynh, D. T. (2023). The impact of dashboards on risk management and decision-making in finance. *Journal of Empirical Social Science Studies*, *7*(4), 51–63. Retrieved from https://publications.dlpress.org/index.php/jesss/article/view/57 (Accessed: 20 December 2024)

Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. *IEEE Data Engineering Bulletin*, *23*(4), 3–13. Retrieved from http://sites.computer.org/debull/A00DEC-CD.pdf

Raschke, R. L. (2022). Assessing iam performance using risk-oriented access certification metrics. *Journal of Information Systems*, *36*(1), 21–39. doi: 10.2308/JIS-2021-014

Ricardianto, P., Lembang, A. T., Tatiana, Y., Ruminda, M., Kholdun, A. I., Kusuma, I. G. N. A. G. E. T., … Endri, E. (2023, 12). Enterprise risk management and business strategy on firm performance: The role of mediating competitive advantage. *Uncertain Supply Chain Management*, *11*, 249-260. doi: 10.5267/j.uscm.2022.10.002

Sajjad, R., & Al-Raqom, D. (2020). Using big data in telecommunication companies: A case study. *African Journal of Business Management*, *14*(7), 209–216. Retrieved from https://www.researchgate.net/publication/344836459_Using_big_data_in_telecommunication_companies_A_case_study doi: 10.5897/AJBM2019.8874

Schäffer, U., & Storek, F. (2022). Transforming risk management. *Controlling & Management Review*, *1*, 30-34. doi: 10.1007/s12176-021-0435-0

Shaw, T. (2018). Key risk indicators in identity and access governance. In R. Weber (Ed.), *It governance risk management and compliance* (pp. 73–94). Springer.

Sherman, R. (2015). *Business intelligence guidebook: From data integration to analytics*. Morgan Kaufmann.

Simitsis, A., Vassiliadis, P., & Sellis, T. (2005). Optimizing etl processes in data warehouses. In *Proceedings of the 21st international conference on data engineering (icde)* (pp. 564–575). IEEE.

Retrieved from https://doi.org/10.1109/ICDE.2005.103 doi: 10.1109/ICDE.2005.103

Singh, G., Kumar, A., Singh, J., & Kaur, J. (2023). Data visualization for developing effective performance dashboard with power bi. In *International conference on innovative data communication technologies and application, icidca 2023 - proceedings* (p. 968-973). doi: 10.1109/ICIDCA56705.2023.10100169

Stonebraker, M., & Cetintemel, U. (2005). "one size fits all": an idea whose time has come and gone. In *21st international conference on data engineering (icde'05)* (p. 2-11). doi: 10.1109/ICDE.2005.1

Sudheesh, K., & Syed, M. (2019). Security and governance in cloud data warehousing. *International Journal of Cloud Computing*, *7*(3), 200–215. Retrieved from https://www.inderscience.com/info/inarticle.php?artid=10012345 doi: 10.1504/IJCC.2019.10012345

Sun, A., & Srivastava, M. (2021). The logical data fabric: A single place for data integration. *Dataversity*. Retrieved from https://www.dataversity.net/the-logical-data-fabric-a-single-place-for-data-integration/

Trivedi, M., et al. (2022). Predicting online purchase intentions using machine learning. *Information*, *13*(12), 664. Retrieved from https://www.mdpi.com/2078-2489/13/12/664 doi: 10.3390/info13120664

van Willigen, M. (2019). *Measuring the user experience of data visualization* (Doctoral dissertation, University of Twente). Retrieved from https://essay.utwente.nl/77564/1/vanWilligen_MA_IT.pdf

Vassiliadis, P., Simitsis, A., & Skiadopoulos, S. (2002). Conceptual modeling for etl processes. In *Proceedings of the 5th acm international workshop on data warehousing and olap (dolap)* (pp. 14–21). ACM. Retrieved from https://dl.acm.org/doi/10.1145/583890.583893 doi: 10.1145/583890.583893

Venkataraman, S. (2019). Cloud data warehousing: Cost and scalability considerations. *Journal of Cloud Computing*, *8*(1), 1–10. Retrieved from https://journalofcloudcomputing.springeropen.com/articles/10.1186/s13677-019-0135-5 doi: 10.1186/s13677-019-0135-5

Vitla, M. S. (2025). Building compliance-aligned iam frameworks: Governance, metrics, and lifecycle integration. *Journal of Enterprise Information Management*, *38*(2), 102–120. doi: 10.1108/JEIM-08-2024-0391

Watson, H. J. (2014). Tutorial: Big data analytics: Concepts, technologies, and applications. *Communications of the Association for Information Systems*, *34*, 65. Retrieved from https://doi.org/10.17705/1CAIS.03462 doi: 10.17705/1CAIS.03462

Widjaja, S., & Mauritsius, T. (2019). The development of performance dashboard visualization with power bi as platform. *International Journal of Mechanical Engineering and Technology (IJMET)*, *10*, 235-249. Retrieved from http://www.iaeme.com/ijmet/issues.asp?JType=IJMET&VType=10&IType=5http://www.iaeme.com/IJMET/issues.asp?JType=IJMET&VType=10&IType=5

Wong, D. M. (2013). *The wall street journal guide to information graphics: The dos and don'ts of presenting data, facts, and figures*. WW Norton & Company.

Yessad, L., & Labiod, A. (2016). Comparative study of data warehouses modeling approaches: Inmon, kimball and data vault. In *2016 international conference on system reliability and science (icsrs)* (pp. 95–99). Retrieved

from https://www.researchgate.net/publication/312486486_Comparative_study_of_data_warehouses_modeling_approaches_Inmon_Kimball_and_Data_Vault doi: 10.1109/ICSRS.2016.7815845

Yulianto, A. A. (2019). Extract transform load (etl) process in distributed database academic data warehouse. *APTIKOM Journal on Computer Science and Information Technologies*, *4*(2), 64–71. Retrieved from https://doi.org/10.11591/APTIKOM.J.CSIT.36 doi: 10.11591/APTIKOM.J.CSIT.36

Appendix A

Appendix

A.1 Dashboard Pages (Original)

The following figures present the original pages of the developed dashboard. Certain elements have been intentionally obscured or omitted in order to preserve confidentiality, without affecting the overall understanding of the dashboard's structure and functionality.

Figure A.1: Glossary page of the dashboard.

Figure A.2: Technical Sheet page of the dashboard.

Figure A.3: Âmbito page of the dashboard.

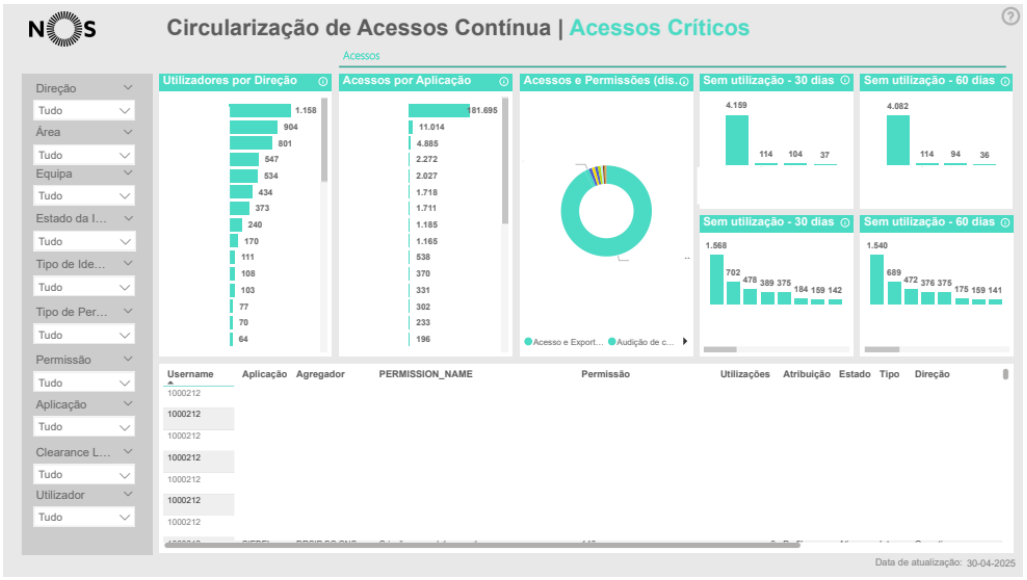

Figure A.4: Critical Accesses page of the dashboard.

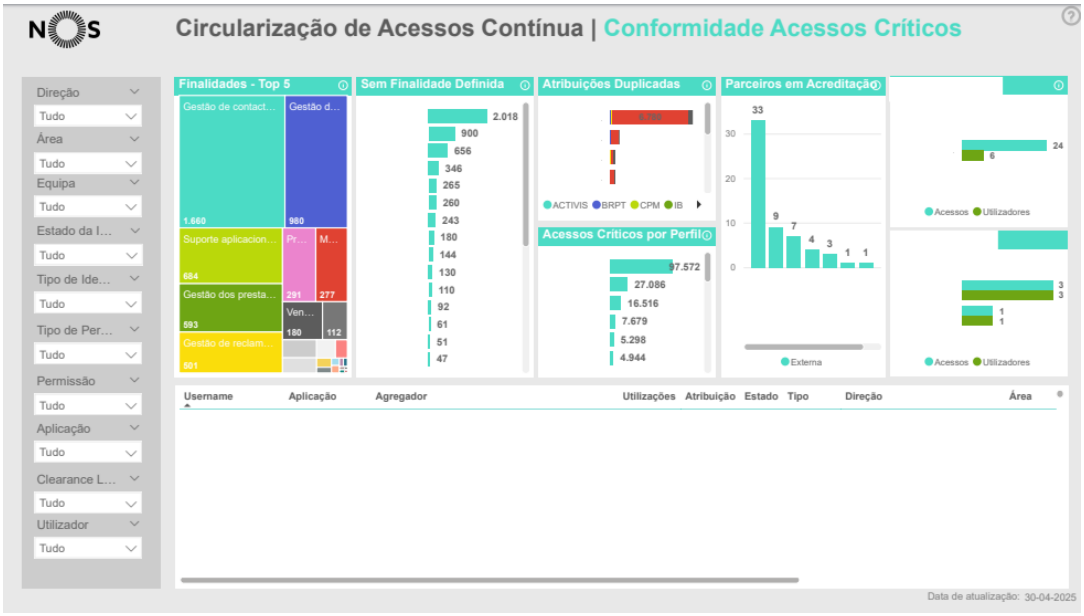

Figure A.5: Critical Accesses Conformity page of the dashboard.

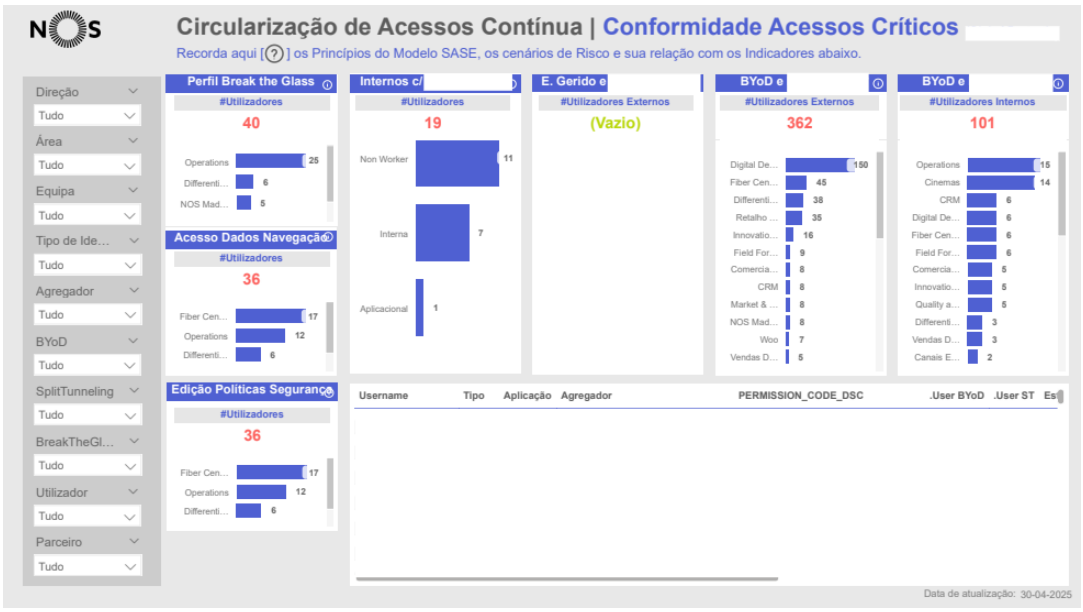

Figure A.6: Remote Access Conformity page of the dashboard.

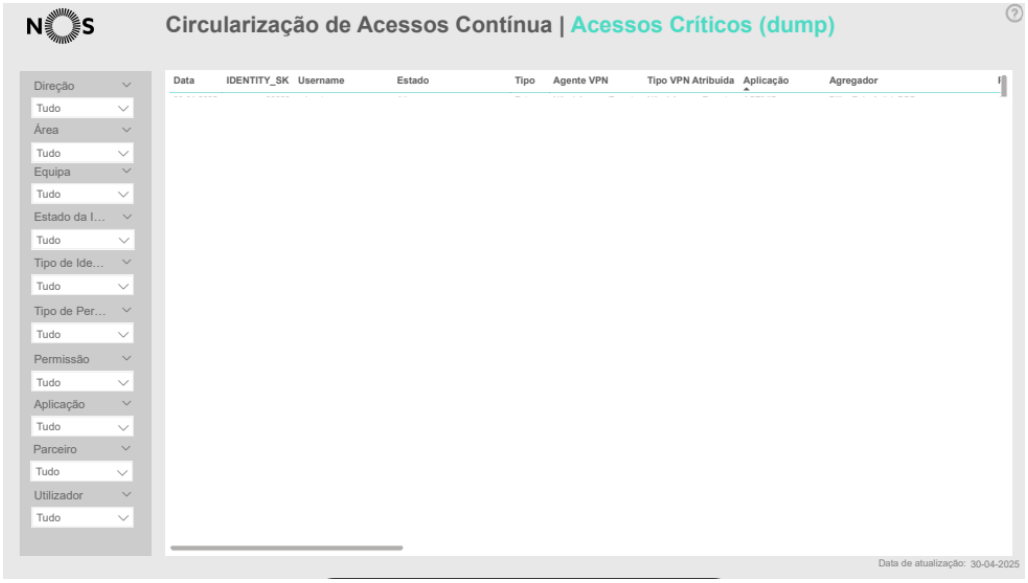

Figure A.7: Data Dump section.