9-30-2024 10:30 AM

Security and Integration in Business Intelligence Tools: A Comprehensive Study

Lakchchayam Divya Khare Mr, *Western University*

Supervisor: Dr. Milani Mostafa, *The University of Western Ontario*
: Dr. Iftekher Chowdhury, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Computer Science

Abstract

This thesis examines business intelligence (BI) tools, which are essential for processing and analyzing business data to support informed decision-making and enhance operational efficiency in modern organizations. Our research includes a comprehensive survey focusing on small to medium enterprises to assess BI tools' adoption, usage, and satisfaction levels. Unlike previous studies, our survey uniquely begins by identifying the evaluation criteria users consider important for BI tools rather than assuming predefined criteria. This user-centered approach led to a two-part survey: Section A collects insights on the general criteria relevant to BI tool evaluation. In contrast, Section B assesses specific BI tools based on the criteria established in Section A.

Our research was motivated by a collaboration with Instacart, a leading online grocery platform. We identified significant security concerns related to using BI tools to manage sensitive centralized data. In particular, we observed that user management and access control pose substantial challenges in environments where BI tools are shared among multiple users with varying access levels, such as managers and data analysts generating reports from sensitive client data. These challenges are exacerbated by frequent changes in data and users accessing it, which necessitates constant updates to access controls.

Through our survey and collaboration with Industry Collabration, we highlight the critical role of security in evaluating and selecting BI tools. We also propose a prototype for a central system designed to address the identified security challenges, ensuring that BI tools remain secure and effective in dynamic environments.

This thesis offers practical implications for companies in selecting BI tools, emphasizing security concerns. By providing a user-centered evaluation framework and addressing common security issues, this research aims to enhance BI tools' overall effectiveness and security, reinforcing their value as indispensable assets for organizations.

Keywords: BI Tools, Data Integration, Data Visualization

Summary for a Lay Audience

One key aspect of my study is the evaluation of BI tools from the user's perspective, emphasizing what criteria they consider most important. This user-centered approach allows companies to select BI tools that best fit their specific needs. Security emerged as a significant concern, especially when managing sensitive data shared among different users within a company. My research, which included collaboration with the online grocery platform Instacart, highlighted challenges such as managing access for users with different roles, like managers and data analysts. These challenges are particularly tricky in dynamic environments where data and user access requirements frequently change. To address these security issues, my study proposes a central system to better manage user access and ensure data security, making BI tools safer and more effective for companies. This research not only helps organizations choose the right BI tools but also guides them in securing their data, ultimately making BI tools more valuable assets in their operations. Through this work, I aim to improve how businesses use BI tools, focusing on security and user needs, which are crucial for making data-driven decisions that can boost their performance and competitiveness.

Acknowledgments

The completion of this thesis would not have been possible without the guidance, support, and encouragement of numerous individuals to whom I am deeply grateful.

First and foremost, I would like to express my profound gratitude to my co-supervisors, Dr. Mostafa Milani and Dr. Iftekher Chowdhury. Their unwavering support, insightful feedback, and deep expertise have guided me through every stage of this research journey. Their patience, intellectual rigor, and encouragement allowed me to navigate through the complexities of my research with confidence. I consider myself incredibly fortunate to have had the opportunity to work under their supervision and to learn from their wealth of knowledge and experience. Their mentorship has not only shaped this thesis but has also significantly contributed to my personal and academic growth.

I would also like to extend my heartfelt thanks to Western University for providing an excellent academic environment and for the continued institutional support throughout my studies. The resources, facilities, and academic opportunities offered by the university have been instrumental in the development of my research. I am grateful to the faculty, staff, and fellow students in the Department of Computer Science who have contributed in various ways to my journey.

I owe a special debt of gratitude to my parents, whose love, sacrifices, and unwavering belief in me have been the foundation upon which all my achievements stand. To my mother and father, I am deeply thankful for their constant emotional support, which has been a source of strength during the challenging phases of my academic pursuit. Their encouragement has always been a motivating force, and without their unconditional love, none of this would have been possible. Words cannot adequately express the depth of my appreciation for everything they have done for me.

Furthermore, I would like to acknowledge the support of my friends and colleagues who, in various ways, have been by my side throughout this academic endeavor. Their encouragement and shared experiences have made this journey more enjoyable and fulfilling.

Lastly, I would like to acknowledge the many authors and researchers whose work has laid the foundation for my thesis. The literature and ideas that preceded my own have been a vital part of my research journey, helping me to understand the broader context and significance of my work.

To everyone who has contributed to this thesis and supported me in one way or another, I offer my sincerest thanks.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In today's business environment, organizations are flooded with vast amounts of data from various sources. Effectively managing and utilizing this data to gain actionable insights is crucial for maintaining a competitive edge. BI tools transform raw data into meaningful information, enabling businesses to make informed decisions by providing historical, current, and predictive views of business operations. These tools facilitate the analysis of complex data sets, helping organizations identify market trends, understand customer behavior, and enhance operational efficiency. By converting raw data into actionable insights, BI tools empower companies to make data-driven decisions, improving their competitive position.

Key components of BI tools include data integration, data visualization, analytics, and user-friendliness. Data integration involves combining data from various sources into a unified view for comprehensive analysis. Data visualization, through charts, graphs, and dashboards, makes complex information accessible, aiding quick decision-making. Analytics, including predictive and prescriptive techniques, extract meaningful patterns from data, while user-friendly interfaces ensure that business users can efficiently utilize these tools without extensive training. Over the years, the concept of Business Intelligence has evolved significantly, from the early decision support systems of the 1950s to the advanced AI and machine learning-driven tools of today. Modern BI tools now offer sophisticated analyses, enabling businesses to stay ahead in an increasingly competitive market.

Our research is motivated by the practical needs identified through collaboration with industry partners[1]. There is a clear need for a comprehensive framework to evaluate BI tools, particularly in understanding the most relevant criteria for selecting the right tool for specific business needs. Furthermore, security concerns, particularly around user access management, have emerged as critical challenges. BI tools are often used by multiple users with varying levels of access, leading to significant security challenges in managing user roles and protecting sensitive data.

[1]The industry supervisor for this thesis is working at a management position at Instacart. His experience includes evaluating multiple publicly available BI/Analytics tools for Instacart usage as well as leading the development of an in-house Business Intelligence Analytics tool for Instacart. He provided his valuable feedback throughout this research project to implement the artifacts for this research (e.g., survey questionnaire). Additionally, we surveyed multiple Instacart employees with experience in Business Intelligence Analytics tools. We will also present our findings to Instacart and explore opportunities for relevant research.

1.1 Problem Statement

In modern business environments, BI tools are indispensable for leveraging data to gain critical insights that drive strategic planning and operational efficiency. However, selecting and integrating these tools into existing business systems poses significant challenges. These challenges are twofold: the need for a comprehensive evaluation framework that captures the most relevant criteria from a user perspective and the pressing security concerns related to data access and user management.

First, BI tools' growing diversity and complexity necessitate a detailed framework for their evaluation. Organizations require clear criteria to assess tools based on their effectiveness in data integration, visualization, analytics, and user-friendliness. Effective data integration ensures that all relevant data is accurately analyzed, providing a complete picture of business operations. Data visualization techniques make complex information accessible and actionable, facilitating quicker decision-making. Advanced analytics, including predictive and prescriptive techniques, help extract meaningful patterns and insights from data, while user-friendly interfaces ensure that business users can efficiently utilize these tools without extensive training. Despite these needs, existing evaluations often overlook the nuances that matter most to users, leading to suboptimal tool selection and underutilization.

Second, current BI tools exhibit significant shortcomings in terms of security, particularly in the context of dynamic business environments where data and user roles frequently change. These security concerns are amplified in scenarios where multiple users interact with sensitive data across shared BI tools, often with varying access levels. The specific security challenges include:

1. **Limited Support for Custom Roles and Access Control**: Most existing BI tools, such as Tableau, Domo, and PowerBI, come with predefined user roles and access controls. These tools often lack the flexibility required to tailor access based on organizations' specific and evolving needs. As a result, the inability to adapt to organizational policies and changes in user roles increases the risk of data misuse and inefficiencies in user management.

2. **Frequent Changes in Data and User Access**: In many organizations, especially Industry Collabration, where multiple users use BI tools to generate reports for different clients, the data and user access requirements constantly change. This fluidity makes it challenging to maintain up-to-date access controls, increasing the potential for unauthorized data access and creating significant compliance and security risks.

3. **Disconnect in Data Filtering Based on User Roles**: Current BI tools often lack the functionality to dynamically filter and display data based on user roles and permissions defined within enterprise applications. This disconnect can lead to unauthorized data access, presenting major compliance and security challenges. The inability to align data visibility with user permissions exposes organizations to risks such as data breaches and non-compliance with data protection regulations.

4. **Inadequate Integration Between Identity and Access Management Tools and BI Systems**: While identity and access management (IAM) tools like Okta and Entra ID

offer basic functionalities such as Single Sign-On (SSO), Multi-Factor Authentication (MFA), and Role-Based Access Control (RBAC), they often lack deep, customizable integration with BI systems. This limitation restricts the synchronization of user identities and dynamic data access controls, leading to a fragmented strategy for managing users and securing data. Such fragmentation increases administrative workload and the risk of security breaches due to errors or outdated access permissions.

5. **Need for Enhanced Integration Between BI Tools and IAM Systems**: To address these challenges, advanced integration solutions are needed to define and implement unique roles and precisely calibrate access controls within BI tools. Enhanced integration would improve data security and boost productivity and compliance by streamlining user processes. Better scalability and flexibility would enable organizations to adapt to evolving business needs and maintain secure, efficient BI tool usage.

1.2 Research Objectives

The challenges identified in integrating and evaluating BI tools require a thorough reevaluation of how these tools are assessed and secured within corporate environments. To address these challenges, this research is guided by several key objectives aimed at enhancing the functionality, security, and overall integration of BI tools:

- Develop a comprehensive evaluative framework for assessing BI tools. This framework is structured to cover key dimensions such as data integration, visualization, analytics, and user-friendliness. Each dimension is further broken down into detailed criteria, ensuring a thorough analysis of the selected BI tools. Additionally, the framework includes criteria that address specific needs and challenges unique to different organizations, ensuring its applicability across various contexts. This objective seeks to fill the gap in existing evaluation methods, providing organizations with a robust basis for informed decision-making when selecting the most suitable BI tools.

- Validate the evaluative framework through a survey of BI tool experts and data analysts. The survey is designed to gather insights and feedback on the relevance and effectiveness of the identified criteria. Participants will rate the importance of each criterion and suggest any necessary refinements or additional considerations. This validation process ensures that the framework reflects industry professionals' actual needs and priorities, making it a practical tool for BI tool evaluation.

- Investigate and address security challenges in integrating BI tools, particularly user management, access control, and dynamic access control. This objective involves a detailed analysis of existing BI tools, such as Tableau, Domo, and PowerBI, and their security features. By identifying gaps in these features, the research aims to develop and propose solutions that enhance the security integration of BI tools. Collaboration with industry partners like Industry Collabration provides real-world validation of these challenges and the proposed solutions, ensuring they are both practical and effective.

Realizing these objectives will significantly contribute to the field by addressing immediate integration and security concerns and paving the way for future research and development in BI tool deployment and optimization.

This study is significant for its dual focus on developing a comprehensive evaluation framework for BI tools and addressing critical security challenges during their integration. It fills gaps in existing evaluation methods by introducing a robust, user-centered framework that covers key dimensions like data integration, visualization, analytics, and user-friendliness. Additionally, the collaboration with Industry Collabration provided valuable insights into real-world challenges, such as dynamic data filtering and user access management, leading to practical solutions that enhance security and operational efficiency in BI tool deployment. These contributions are expected to have a lasting impact on how organizations select, integrate, and optimize BI tools, fostering continuous improvement and innovation.

While this research initially focuses on the challenges faced by small to mid-sized enterprises (SMEs), the solutions proposed are equally relevant to larger enterprises. Companies of all sizes, including industry giants like Amazon, encounter similar issues related to managing access control, securing sensitive metadata, and ensuring scalability across multiple BI tools. *The complexities of handling dynamic user permissions, securing data across platforms, and scaling BI operations* are challenges that affect organizations regardless of their size. Therefore, the framework and solutions presented in this thesis are designed to be versatile and scalable, addressing the needs of businesses ranging from SMEs to large enterprises, all of which require effective and secure BI tool management.

1.3 Methodology, Contributions, and Thesis Structure

Our research methodology combines a survey-based approach with an in-depth investigation of security features in popular BI tools. We first developed a comprehensive evaluative framework, focusing on key dimensions such as data integration, visualization, analytics, and user-friendliness. This framework was refined through interviews with BI tool experts and validated by a survey of industry professionals, ensuring its relevance and practical applicability.

To address security concerns, we conducted a detailed analysis of BI tools like Tableau, Domo, and PowerBI, focusing on their handling of custom roles, access control, and dynamic data filtering. This analysis, along with a literature review on data and software security, informed the design of a prototype solution. Developed in collaboration with Industry Collabration, the prototype incorporates advanced integration solutions to enhance user authentication, authorization, and data access control, ensuring its effectiveness in real-world applications.

The primary contribution of this thesis is the development of a validated evaluative framework that addresses both the user experience and security concerns of BI tools, surpassing existing methodologies. Additionally, the thesis provides a significant contribution through the identification of security vulnerabilities in popular BI tools and the proposal of a practical prototype solution. The collaboration with Industry Collabration not only validates these findings but also underscores the practical impact of the research, offering valuable insights and solutions for organizations seeking to optimize their BI tool deployment.

The structure of this thesis is organized to address the research questions and objectives systematically. Chapter 2 presents a comprehensive literature review, covering existing frame-

works for BI tool evaluation and highlighting the security challenges associated with their integration. Chapter 3 focuses on developing the evaluative framework, describing the criteria selection process, the validation survey design, and the survey results analysis. Chapter 4 investigates the security analysis of popular BI tools, discussing the findings and design of the proposed security solution. The chapter includes a case study involving Industry Collabration, where the developed framework and security solutions are applied and tested in a real-world context, demonstrating their effectiveness and practicality. Finally, Chapter 5 concludes the thesis by summarizing the key contributions and discussing the potential for continued advancements in BI tools.

Chapter 2

Evaluation Framework for BI Tools

In this chapter, we present an evaluation framework for BI tools. We begin by briefly discussing the role of BI in modern corporate operations, emphasizing its importance in data-driven decision-making. We then outline the evaluation criteria for BI solutions and explore the development of BI tools in supporting organizations across various contexts. We reference key industry analysts like Gartner, Forrester, and BARC, who shape industry perspectives on BI technologies. For evaluation, we selected five prominent BI platforms—Power BI, Tableau, Qlik Sense, Sisense, and Apache Superset—based on their industry standing. Our goal is to provide organizations with actionable insights for selecting the most suitable BI tools by systematically assessing these platforms against expert-defined standards.

BI tools are integral to corporate success, transforming raw data into actionable insights that guide strategic decisions, improve operational efficiency, and maintain competitive advantage. These tools enable organizations to analyze data from diverse sources, predict future outcomes, and identify trends. By understanding the core principles of BI and the pivotal role of these tools, we can better evaluate and select the best BI solutions to drive organizational success.

2.1 Background on BI Tools

Software programs or platforms known as BI solutions are made to gather, process, analyze, and present data in order to help firms make well-informed decisions. These technologies are essential for turning unprocessed data into insights that can be used to enhance competitive positioning, operational efficiency, and strategic goals. The goal of BI solutions is to enable users at all organizational levels to investigate data, find trends, and obtain insightful knowledge to aid in decision-making. Throughout history, BI technologies have progressed from simple reporting systems to complex platforms with advanced analytics, data visualization, and predictive modeling features. BI tools have developed into vital resources for companies looking to prosper in a time of abundant data and cutting-edge technology. A fundamental paradigm shift in how businesses use data to make decisions is represented by BI. The notion of BI has become an indispensable tool for companies looking to obtain a competitive advantage in the fast-paced and fiercely competitive corporate world of today. BI is a collection of tools, procedures, and technology that turn unstructured data into insightful knowledge and help with strategic decision-making. BI is becoming increasingly important as firms struggle to extract

meaningful insights from the ever-growing volumes of data.

2.1.1 Existing Studies for Evaluation of BI Tools

An essential element of the data-driven decision-making environment is the evaluation of BI tools. Prominent companies like Gartner, Forrester, and the Business Application Research Center (BARC) are important sources of assessments that have a big impact on how the industry views and uses these products. The modern company is characterized by a demand for rapid, accurate, and collaborative decision-making this is where BI comes in. The core of BI is the interaction of information technology IT and business processes, with an emphasis on converting unprocessed data into insightful knowledge. Many components, including databases, applications, techniques, tools, and architectures, are included in this field. Introduced by the Gartner Group in the middle of the 1990s, BI integrates fact-based decision-making, analysis, and business information [28, 39].

- **Gartner**: Gartner has a long list of evaluation criteria for BI and analytics technologies. Essential aspects including security, management, cloud analytics, connectivity to data sources, and data preparation are all included in the requirements. Their assessment also heavily relies on elements such as the construction of catalogs, automated insights, data visualization, data storytelling, natural language query and generation, and reporting. This extensive set of standards guarantees that the BI tools under evaluation are not only highly developed technically but also well-suited to the changing business and data environments The comprehensive visual portrayal of BI vendors included in Gartner's Magic Quadrant reports is a source of great respect. These studies plot vendors' performance against their completeness of vision on a two-dimensional graph. This model offers a thorough analysis of the market dynamics [11, 12].

- **Forrester:** The Wave studies from Forrester employ a nuanced strategy, highlighting the advantages and disadvantages of different BI tool providers. These papers highlight the vendors' market position, current offers, and strategic direction. Businesses can better match their choice of BI solutions with their strategic objectives by using Forrester's rigorous analysis, which ensures a synergy between long-term business goals and technology capabilities as well as practical usefulness in a corporate context. Forrester advocates for a broad assessment strategy in their full evaluation of BI technologies. Beyond the bare minimum, factors like data connectivity, user interface design, sophisticated analytics capabilities, scalability, strong security measures, and dependable vendor support are taken into account. This methodology guarantees a comprehensive assessment, considering not only the BI tools' technical capabilities but also their usefulness in a business setting [9, 8]. .

- **BARC** distinguishes itself with its user-centric approach to BI tool evaluation. BARC's reports offer an authentic portrayal of the user experience by thoroughly examining user input and satisfaction indicators. This includes highlighting aspects like usability, performance, and the useful value that the BI tools bring in real-world scenarios [2].

While the assessments provided by Gartner, Forrester, and BARC offer valuable insights, they also have certain limitations that necessitated our independent survey. One key shortcoming is that these evaluations are often tied to the specific features and strengths of the tools currently dominating the market, potentially overlooking the broader, independent criteria that are essential for building a robust, tool-agnostic evaluation framework. Additionally, these surveys may not always focus on the most up-to-date and state-of-the-art BI tools, leaving a gap in understanding how emerging technologies stack up against established solutions. Our survey uniquely addresses these gaps by developing evaluation criteria that are not bound to any particular tool, allowing for a more flexible and comprehensive comparison across the latest BI platforms.

2.1.2 Selected BI Tools for In-Depth Evaluation

In this section, we explain our selected BI tools for our survey study and the rational behind our choice. We have chosen five well-known systems for our BI tool analysis: Power BI, Tableau, Qlik Sense, Sisense, and Apache Superset. In the market, each of these technologies has established a solid reputation for its capacity to convert unstructured data into insightful understandings, empowering businesses to make informed decisions based on data. Big organizations like Apple, LinkedIn, and Amazon use these tools for their business [32, 33]. By assessing each platform's strengths, weaknesses, and unique selling points, we aim to provide a comprehensive comparison to help businesses identify the most suitable tool for their specific requirements. In selecting the appropriate tools for our analysis, we considered various factors that underscore their dominance in the industry.

- **PowerBI:** Because of its self-service BI features, which let end users generate reports and dashboards without the help of database administrators or IT personnel, Microsoft Power BI, our top option, stands out. Its leading position in the market further motivated our choice. Notable enterprises, including Hewlett-Packard (HP) Inc., Vodafone, and Ernst & Young (EY), utilize Power BI for their analytics and BI needs, attesting to its reliability and efficiency. Also, Gartner recognized Microsoft's Power BI as a leader in the 2020 report for Analytics and BI Platforms.

- **Tableau:** We selected Tableau as our next product after Power BI because it is well-known for enabling large-scale governance, security, and compliance while maintaining cost-effectiveness. According to Google Trends, its popularity is demonstrated by the fact that over the previous five years, there have been more searches for it than for Power BI. Its popularity is evidenced by a higher search frequency compared to Power BI over the past five years, according to Google Trends. The acquisition of Tableau by Salesforce in 2019 and its impressive Q4 revenue of $636 million, a 5.6% year-over-year increase, further substantiates its strong market presence [10, 34, 21, 38, 13, 33].

- **Qlik:** Our analysis also encompasses Qlik, a tool that experienced a remarkable growth of 46.6% in 2020, reaching $153 million in data integration revenue, the highest among the top 10 vendors by market share according to Gartner [24].

- **Sisense:** Sisense is another pivotal tool in our evaluation, chosen primarily due to its endorsement by our industry collaborator, Industry Collaboration. The platform has proven

to be particularly suitable for businesses with 50-200 employees and an annual revenue ranging between $1 million and $10 million [31].

- **Apache Superset:** Lastly, we included Apache Superset in our study, an open-source solution offering customization and extensibility, which is vital for organizations seeking to delve deeper into tool architectures and expand functionalities through novel plugins [1]. Its vibrant community support is reflected in over 10,000 commits, 8,400 forks, and approximately 43,000 ratings on its GitHub repository.

This selection of tools allows us to conduct a comprehensive and detailed analysis, leveraging the strengths of each to enhance our understanding and application of data visualization tools. Other tools with important features not mentioned above and are popular and widely used are briefly discussed below. These are mainly based on Gartner's mentions of leaders and players in the BI domain, each with unique features and capabilities [10]:

- **ThoughtSpot:** Recognized as the leader in BI tools, ThoughtSpot was listed among the "Top 100 North American Companies" by Red Herring [25]. It connects with Google Cloud Platform and Amazon Web Services and can analyze data from sources like Snowflake and Databricks
- **Looker:** Acquired by Google, Looker uses LookML, enabling data teams to define database relationships so business users can work with data easily [14].
- **MicroStrategy:** Known for its multi-source option addition to Intelligence Server, it allows combining data from multiple databases into a single project [22].
- **IBM Cognos:** Offers tools to combine various data sources and an AI Assistant tool for quick, understandable recommendations [18].
- **Dundas BI:** Catering to organizations needing extensive customization and flexibility [7].
- **Domo:** Includes a Magic ETL feature and offers an HTML5 interface with embedded and extended analytics capabilities [20].
- **Pentaho:** An open-source BI suite focusing on big data and data integration, consisting of particular Java classes to build any BI solution [23].

The list of tools above hold niche positions in the diverse landscape of BI tools. They cater to unique requirements and offer distinct functionalities, complementing the offerings of other prominent leaders in the sector. Our analysis deliberately concentrates on our curated set of five tools to enable a more detailed and focused evaluation study. Fig 2.1 illustrates the timeline of the introduction of various BI tools. The tools are plotted in chronological order, with the year of introduction on the x-axis and the tool names on the y-axis. The most recent tools appear at the top of the chart to provide an intuitive view of the evolution of BI tools over time.

This visualization highlights key milestones in the development of BI tools, showing how the industry has progressed from early players like IBM Cognos (introduced in 1984) to more modern tools like Power BI (introduced in 2014).

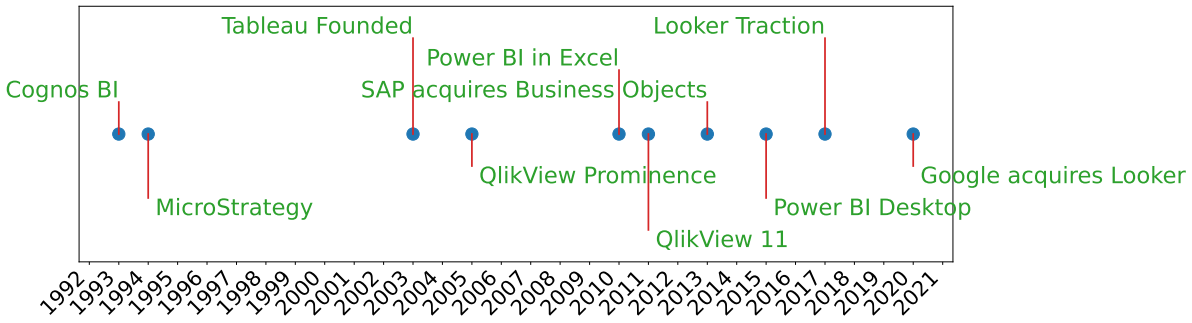


Figure 2.1: Timeline of Business Intelligence Tools Introduction. This figure shows the chronological introduction of various BI tools, with the most recent tools listed at the top.

2.1.3 Main Choice of BI Tools: Tableau and Power BI, and Their Core Benefits

Companies across various industries consistently choose *Tableau* and *Power BI* as their primary BI tools due to several core factors, including *usability*, *scalability*, *AI capabilities*, and *ease of learning*. Tableau's drag-and-drop interface is renowned for its *ease of use*, enabling users with minimal technical expertise to create sophisticated data visualizations and uncover insights quickly. This is particularly valuable in industries like finance, healthcare, and retail, where business users require rapid, data-driven decision-making without the need for extensive IT involvement.

Power BI is equally popular due to its seamless *integration with the Microsoft ecosystem*, allowing businesses already using tools like Excel and Azure to easily connect their data flows and extend their analytical capabilities. Both Tableau and Power BI excel in *scalability*, accommodating the needs of small businesses with basic reporting requirements, as well as large enterprises handling massive datasets across multiple departments. The flexibility of these tools makes them ideal for companies looking to scale their data operations over time.

In addition to usability and scalability, both tools offer advanced *AI-driven features*. Power BI, in particular, integrates Microsoft's sophisticated AI models, enabling businesses to perform predictive analytics, automate data analysis, and uncover hidden patterns within datasets. Tableau also incorporates AI-powered augmented analytics, guiding users toward relevant insights without requiring deep technical knowledge, making it a powerful tool for both technical and non-technical users.

Finally, the *learning curve* for both tools is minimized through extensive support systems, including user communities, online tutorials, and certification programs. This ensures rapid adoption within organizations, allowing employees to become proficient quickly and start leveraging the full potential of these tools. The impact of adopting Tableau and Power BI has been transformative for the industry, enabling companies to democratize data analytics, enhance operational efficiency, and maintain a competitive advantage by making data-driven decisions at every level of the organization.

2.2 Criteria-Based Assessment of BI Tools

After selecting the particular BI tools, we can now compare and thoroughly assess them. The first step in this process is to examine the basic categories, each of which includes essential functions like analytics, data integration, and visualization. We then divide these core areas into more specific subcategories to establish a rigorous platform for comparison, laying the groundwork for a thorough examination of the five chosen BI solutions.

In addition to these basic categories and their thorough analysis, we will also examine other significant factors that don't align perfectly with the main categories but play a crucial role in evaluating the efficiency and suitability of the tools. These requirements were carefully selected to reflect the many demands of the contemporary workplace, and the hierarchical structure ensures a methodical and systematic approach to the evaluation. Through a comprehensive analysis of each tool within this layered criteria, we highlight the unique strengths and potential weaknesses of each, providing a robust basis for informed and sound decision-making.

The category "Data Integration and Connectivity" are essential to businesses because they involve combining data from many sources and making sure it flows smoothly across them. Organizations can create a single view of information by combining data from many systems, including databases, applications, and files, through this method. Businesses can increase performance and make well-informed decisions by gaining a thorough insight of their customers, operations, and market trends through effective data integration. We reiterate this requirement here with a brief focus on the various kinds of data sources that the data integration solution can link to.

- **Files (Excel, CSV, TXT, JSON, XML, PDF)**: This sub-criterion assesses the solution's ability to connect to various file formats commonly used for data storage. It ranges from not being able to connect to any format to being able to connect to all mentioned formats.
- **Databases (Relational and Non-relational)**: This assesses how well the solution can connect to various kinds of databases. The process commences with no database connectivity and advances to establish connections with relational and non-relational (NoSQL) databases, encompassing stored procedure scenarios.
- **Cloud Storage**: This measures integrating cloud-based storage solutions. It begins with no integration, moves to read-only access, then read/write capability, synchronization abilities, and ultimately full integration with versioning support.
- **APIs**: This assesses the solution's integration with APIs. It starts from no API connectivity and progresses to limited integration with standard API endpoints.

The "Data Preprocessing, Cleaning, and ETL" (Extract, Transform, Load) categorization is vital for evaluating BI technologies and enabling accurate data analysis. With these technologies, data extraction from a variety of sources frequently initiates the ETL process and provides the framework for further analytical stages. Data transformation comes next, which entails combining datasets and removing unnecessary information to achieve data homogeneity. Loading the reduced data into the BI tool for analysis and visualization is the next step.

.

- **Data Import and Export**: This feature supports both bulk and incremental data loading, as indicated by the checkmarks in the respective columns.
- **Data Transformation**: This feature supports data transformation using SQL, JavaScript, and Python, as denoted by the checkmarks.
- **Data Quality Management**: This feature involves data cleaning, validation, and enrichment. It seems to be supported to varying degrees, with checkmarks indicating support in certain areas.
- **Stream Processing**: Real-time data analysis is supported, with some checkmarks indicating varying levels of support.
- **Caching and Performance:**: In-memory processing and caching mechanisms are provided, again with differing levels of support indicated by the checkmarks.

The "Integration and API Support" category encompasses the platform's capabilities regarding interoperability with external systems and the tools provided to facilitate this integration process. Essentially, it focuses on enabling seamless connectivity and data exchange between the platform and other software applications or services.

- **Standard API**: The platform supports both REST and GraphQL APIs, as indicated by the checkmarks.
- **SDKs and Developer Tools**: The platform provides SDKs for custom development, showing support for developers to extend the platform's functionality according to their needs, as indicated by the checkmarks.
- **Third-party Integrations**: The platform offers extensive plugin and integration support for third-party services, allowing users to connect with various external tools or systems, as shown by the checkmarks.

The evaluation criteria for the "Data Visualization and Reporting" category centers on the proficiency of BI tools in presenting data in an accessible and insightful manner. Refer to Table 2.2 for a comprehensive rubric outlining this category and its subcategories. Within the "Data Visualization Options". We examine the variety and efficacy of techniques for visual representation, such as graphs, charts, and other graphical components. Another important feature is dashboard customization, which lets customers adapt visualizations to their reporting requirements. We also look at the interactive elements that improve the user's ability to engage dynamically with the data, like drill-down and filtering functions. This rubric helps businesses choose tools that meet their needs for presenting data by providing an organized framework for evaluating the reporting and data visualization capabilities of different BI solutions. We shall examine each subcategory in more detail in the subsequent sections. Please note that the terms "category" and "criteria" are used interchangeably to denote the focal points of our evaluation.

- **Basic Charts and Graphs**: The tools (Tableau, Power BI, Qlik, Superset, and Sisense) all offer a wide range of basic chart types, including line charts, bar charts, pie charts, area charts, scatter plots, and histograms.

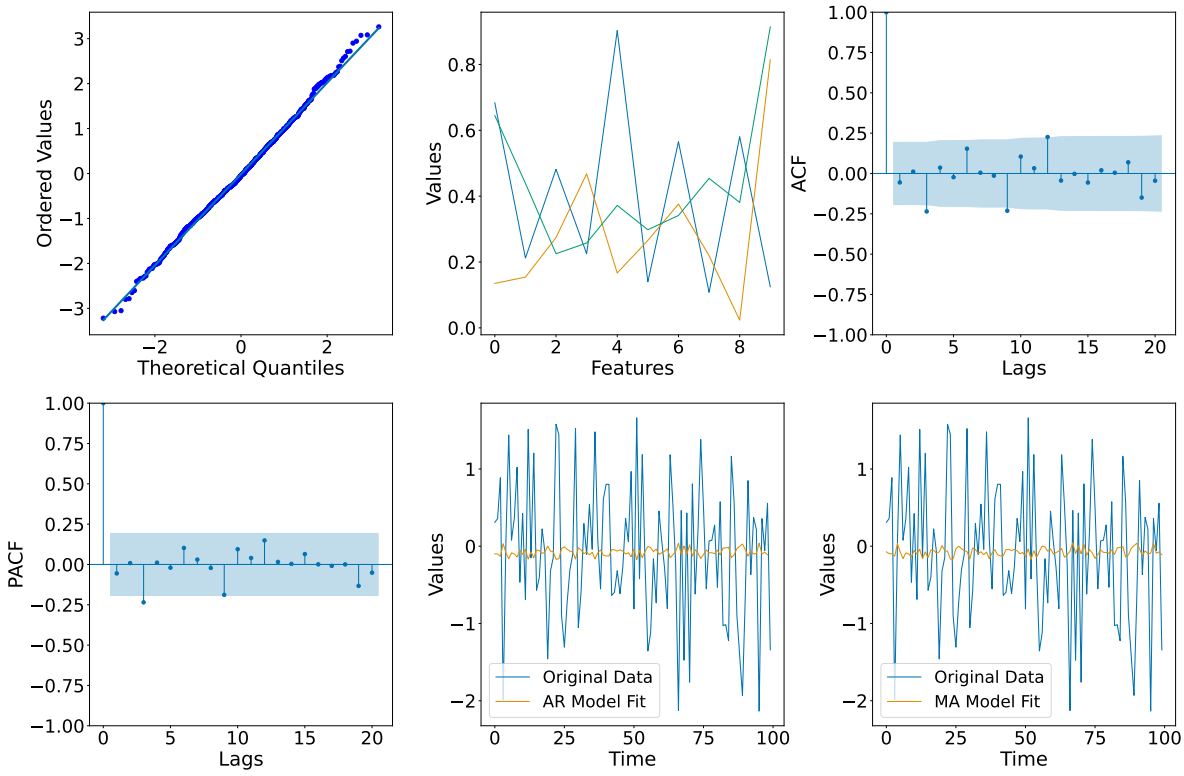


Figure 2.2: (From left to right) The QQ plot compares data quantiles against a standard normal distribution. 2. The Parallel Coordinates plot visualizes relationships among multiple variables. 3, 4. The ACF and PACF plots examine autocorrelations in the time series. 4, 6. The AR model plot shows the AutoRegressive model fit, while the MA model plot displays the Moving Average model fit.

Criteria	Sub-criteria	Item	Tableau	Power BI	Qlik	Superset	Sisense
Source Connectivity	Files	Excel, CSV, TXT, JSON, XML, PDF	✓	✓	✓	✓	✓
	Databases	Relational & NoSQL	✓	✓	✓	✓	✓
	Cloud Storage	AWS, Google Cloud, Azure	✓	✓	✓	✓	✓
	APIs	RESTful, SOAP	✓	✓	✓	×	✓
	Real-Time Sources	Kafka, RabbitMQ, IoT devices	✓	×	✓	×	✓
Data Processing and Management	Data Import and Export	Supports bulk and incremental load	✓	✓	✓	×	✓
	Data Transformation	SQL, JavaScript, Python support	✓	✓	✓	✓	✓
	Data Quality Management	Data cleaning, validation, and enrichment	✓	✓	✓	×	✓
	Stream Processing	Real-time data analysis	✓	✓	×	×	✓
	Caching and Performance	In-memory processing, caching mechanisms	✓	✓	✓	×	✓
Integration and API Support	Standard APIs	REST, GraphQL	✓	✓	✓	×	✓
	SDKs and Developer Tools	Provides SDKs for custom development	✓	✓	✓	×	✓
	Third-party Integrations	Extensive plugin and integration support	✓	✓	✓	×	✓

Table 2.1: Evaluation Rubric for Data Integration and Connectivity Capabilities of BI Tools

- **Advanced Visualizations**: The evaluated tools support advanced visualizations like heatmaps, geospatial maps, waterfall charts, parallel coordinates, sunburst charts, radar charts, and bubble charts.
- **Statistical Plots**: All tools support statistical plots, including box plots, and violin plots.
- **Real-time Visualizations**: All tools except Sisense support live dashboards. Sisense supports streaming charts with some limitations.

This category assesses how customizable the dashboards are in each BI tool:

- **Theming and Styling**: All tools allow the use of predefined themes and custom colors for styling dashboards.
- **Dashboard Configuration**: All tools support drag-and-drop design and resizable widgets for easy dashboard configuration.
- **Report Customization**: All tools allow customization of headers and conditional formatting for reports.

This "Interactivity Features" category evaluates the interactive capabilities of each BI tool:

- **Data Exploration**: All tools support drill-down and filtering options for exploring data in visualizations.

- **User Interaction**: All tools provide hover tooltips and clickable elements for user interaction with visualizations.

- **Collaborative Features**: All tools allow for shared dashboards and real-time collaboration on data visualizations.

This rubric can be used as a thorough evaluation tool to compare the reporting and data visualization features of different BI solutions in a number of ways. It supports users in making well-informed choices on which tool would be most appropriate for their particular requirements.

The evaluation criteria within the "Data Analysis Capabilities" category focus on the effectiveness and depth of analytical functionalities offered by BI tools. Refer to Table 2.3 for an extensive rubric describes this category and its subcategories. Under "Data Preparation and Manipulation" we assess the tools' proficiency in refining and organizing data for meaningful analysis, encompassing data cleaning and transformation tasks. In "Statistical Analysis" We assess how much BI tools help with statistical processes that extract meaningful information from data distributions. Moreover, the area includes "Time-Series Analysis" which investigates how instruments may manage time-varying data for forecasting and trend analysis.

The "Data Preparation and Manipulation" is assessing the tool's ability to prepare and manipulate data in advance of analysis. It looks at how well the tool can handle operations like data aggregation which involves calculating average and sum and data joining, which involves combining datasets based on shared properties. In essence, this subcategory evaluates the tool's capacity to arrange and handle unprocessed data into an analysis-ready structure, which is necessary to glean insightful conclusions and make defensible decisions.

- **Data Joining:** This entails assessing the tool's capacity to combine datasets using shared keys or properties. For thorough analysis, it is essential to combine data from multiple sources into a single dataset.

- **Data Aggregation:** In this case, we evaluate the tool's capacity to process data using aggregation functions like sum, average, count, etc. Large datasets are summarized in aggregated views, which facilitates decision-making.

The "Statistical Analysis" subcategory assesses the tool's capacity to calculate key statistical measures, such as mean, median, and mode, offering perceptions into the central tendencies of the data. In order to provide insights on data spread, it also evaluates the tool's ability to compute measures of dispersion including variance and standard deviation. It also looks at how well the tool can create histograms and calculate percentiles to see data distribution and comprehend data positionality, respectively.

- **Mean, Median, Mode:** We assess the tool's ability to calculate basic statistical measures including the mode (value that occurs most frequently), median (middle value), and mean (average). These metrics shed light on the dataset's central patterns.

Criteria	Sub-criteria	Item	Tableau	Power BI	Qlik	Superset	Sisense
Data Visualization Options	Basic Charts and Graphs	Line Chart	✓	✓	✓	✓	✓
		Bar Charts	✓	✓	✓	✓	✓
		Pie Charts	✓	✓	✓	✓	✓
		Area Charts	✓	✓	✓	✓	✓
		Scatter Plots	✓	✓	✓	✓	✓
		Histograms	✓	✓	✓	✓	✓
	Advanced Visualizations	Heatmaps	✓	✓	✓	✓	✓
		Geospatial Maps	✓	✓	✓	✓	✓
		Waterfall Charts	✓	✓	✓	✓	✓
		Parallel Coordinates	✓	✓	✓	✓	✓
		Sunburst Charts	✓	✓	✓	✓	✓
		Radar Charts	✓	✓	✓	✓	✓
		Bubble Charts	✓	✓	✓	✓	✓
	Statistical Plots	Box Plots	✓	✓	✓	✓	✓
		Violin Plots	✓	✓	✓	×	✓
		QQ Plots	✓	✓	✓	×	✓
		Correlation Matrices	✓	✓	×	✓	✓
		Trend Lines	✓	✓	✓	✓	✓
	Real-time Visualizations	Live Dashboards	✓	✓	✓	✓	✓
		Streaming Charts	✓	✓	✓	×	×
Customizability of Dashboards	Theming and Styling	Predefined Themes	✓	✓	✓	✓	✓
		Custom Colors	✓	✓	✓	✓	✓
	Dashboard Configuration	Drag-and-Drop Design	✓	✓	✓	✓	✓
		Resizable Widgets	✓	✓	✓	✓	✓
	Report Customization	Custom Headers	✓	✓	✓	✓	✓
		Conditional Formatting	✓	✓	✓	✓	✓
Interactivity Features	Data Exploration	Drill-Down	✓	✓	✓	✓	✓
		Filters	✓	✓	✓	✓	✓
	User Interaction	Hover Tooltips	✓	✓	✓	✓	✓
		Clickable Elements	✓	✓	✓	✓	✓
	Collaborative Features	Shared Dashboards	✓	✓	✓	✓	✓
		Real-time Collaboration	✓	✓	✓	✓	✓

Table 2.2: Evaluation Rubric for Data Visualization and Reporting

- **Variance (std):** This entails evaluating the tool's capacity to compute dispersion metrics such as variance and standard deviation. By measuring the variability or spread of data points around the mean, these metrics shed light on the consistency and variability of the data.

- **Histograms:** Here, we investigate the tool's ability to produce histograms, which are graphical depictions of a data set's frequency distribution. Histograms make data's structure and distribution easier to see, which makes it easier to spot patterns and anomalies.

- **Percentiles:** We assess the tool's ability to generate percentiles, which split a dataset into 100 equal segments. Understanding the relative position of a given value within the dataset and recognizing extreme values are made easier with the help of percentiles.

The "Time-Series Analysis" sub-category explores the features of the tool for trend analysis of temporal data. It evaluates if the instrument can compute moving averages to find trends in time-series data. It also assesses how well the tool can separate time-series data into seasonal, trend, and residual components, allowing for a more in-depth analysis of underlying trends. In addition, it looks at whether the tool can generate plots of autocorrelation and partial autocorrelation, which can help analyze temporal dependencies in the data.

- **Moving Averages:** This entails determining whether the tool can compute moving averages, a method for reducing oscillations in time-series data and spotting underlying trends or patterns.

- **Seasonal Decomposition:** The capacity of the tool to separate time-series data into its seasonal, trend, and residual components is assessed here. This makes it easier to spot erratic oscillations, long-term trends, and seasonal patterns in the data.

- **ACF and PACF Plots:** Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots are essential tools for analyzing the correlation structure within time-series data. These plots help identify lagged relationships between observations, aiding in the identification of potential forecasting models and understanding data dependencies.

- **AR and MA Models**: These are AutoRegressive and Moving Average models used in time-series analysis. None of the mentioned tools support these.

The criterion of "Advanced Analysis and Machine Learning" looks at the availability of cutting-edge machine learning and analytical methods. By providing a clear framework for evaluating the data analysis capabilities of various BI systems, this rubric helps organizations choose tools that meet their analytical needs. Each subcategory will be thoroughly explored in the sections that follow.

- **Predictive Analysis and Forecasting**: Assesses if the tool can perform predictive modeling and forecasting.

- **Hypothesis Testing**: Checks if the tool supports hypothesis testing for making inferences about a population.

Criteria	Sub-criteria	Item	Tableau	Power BI	Qlik	Superset	Sisense
Data Analysis Capabilities	Data Preparation and Manipulation	Data Joining	✓	✓	✓	✓	✓
		Data Aggregation	✓	✓	✓	✓	✓
	Statistical Analysis	Mean, Median, Mode	✓	✓	✓	✓	✓
		Variance (std)	✓	✓	✓	✓	✓
		Histograms	✓	✓	✓	✓	✓
		Percentiles	✓	✓	✓	✓	✓
	Time-Series Analysis	Moving Averages	✓	✓	✓	✓	✓
		Seasonal Decomposition	✓	✓	✓	✓	✓
		ACF and PACF Plots	×	×	×	×	×
		AR and MA Models	×	×	×	×	×
	Advanced Analysis and Machine Learning	Predictive Analysis and Forecasting	✓	✓	✓	✓	✓
		Hypothesis Testing	✓	✓	✓	✓	✓
		Clustering	✓	✓	✓	✓	✓
		Text Analysis and NLP	✓	✓	✓	✓	✓

Table 2.3: Evaluation Rubric for Data Analysis Capabilities

- **Clustering**: Assesses if the tool provides capabilities for clustering or grouping similar data points.

- **Text Analysis and NLP**: Determines if the tool supports text analysis and natural language processing for unstructured data.

The evaluation criteria for "User-Friendliness and Ease of Use" revolve around the accessibility and intuitiveness of BI tools. Refer to Table 2.4 for a detailed rubric outlining this category and its subcategories. Within the User "Interface", we scrutinize the design and layout of the tool, assessing its clarity and user-friendliness. We also evaluate the learning curve associated with each tool, acknowledging the time required for users to become proficient. This category seeks to gauge the ease with which users can navigate and utilize the features of the BI tool. This rubric provides a structured framework for assessing the user-friendliness and ease of use of various BI solutions, aiding organizations in selecting tools that align with their user interface requirements. Further sections will offer an in-depth exploration of each subcategory.

The "User Interface" sub-category examines how user-friendly and visually appealing the BI tools are, ensuring an intuitive experience for users. This evaluation is crucial for organizations aiming to adopt tools that enhance user productivity and engagement.:

- **Intuitiveness**: All of the mentioned BI tools are considered intuitive, making it easy for users to navigate and perform tasks.

- **Customization**: Users can customize the interface of all the mentioned BI tools to suit their preferences or specific needs.

- **Accessibility**: All of the mentioned BI tools are accessible, ensuring that users can interact with the tool effectively.

Criteria	Sub-criteria	Item	Tableau	Power BI	Qlik	Superset	Sisense
User-Friendliness	User Interface	Intuitiveness	✓	✓	✓	✓	✓
		Customization	✓	✓	✓	✓	✓
		Accessibility	✓	✓	✓	✓	✓
	Learning Curve	Ease of Onboarding	✓	✓	✓	✓	✓
		Training Resources	✓	✓	✓	✓	✓
		User Support	✓	✓	✓	✓	✓
Ease of Use	Usability	Data Import	✓	✓	✓	✓	✓
		Drag-and-Drop Functionality	✓	✓	✓	✓	✓
		Report/Dashboard Creation	✓	✓	✓	✓	✓

Table 2.4: Evaluation Rubric for User-Friendliness and Ease of Use

The "Learning Curve" category evaluates the ease with which users can learn to use the BI tools effectively. It considers factors such as user-friendliness, availability of training resources, and intuitive interface design. A steep learning curve may indicate a need for extensive training and support, while a gentle curve suggests that users can quickly adapt to the tools with minimal assistance:

- **Ease of Onboarding**: Onboarding is straightforward for all the mentioned BI tools, allowing users to get started without significant hurdles.
- **Training Resources**: Adequate training resources are available for all the mentioned BI tools, aiding users in learning and using the tool effectively.
- **User Support**: Each BI tool provides robust user support, ensuring that users can get assistance when needed.

The "Ease of Use" focuses on how quickly users can become proficient with the BI tools. It assesses various aspects such as data import, drag-and-drop functionality, and report/dashboard creation. For data import, onboarding is straightforward across all mentioned BI tools, ensuring users can get started without significant hurdles.

- **Data Import**: Onboarding is straightforward for all the mentioned BI tools, allowing users to get started without significant hurdles.
- **Drag and Drop Functionality**: Adequate training resources are available for all the mentioned BI tools, aiding users in learning and using the tool effectively.
- **Report/Dashboard Creation**: Each BI tool provides robust user support, ensuring that users can get assistance when needed.

The evaluation criteria for "Performance and Scalability" center on the responsiveness and scalability of BI tools to handle increasing data volumes. Refer to Table 2.5 for an in-depth rubric outlining this category and its subcategories. In terms of "Performance," we assess the speed and responsiveness of the BI tool, considering its efficiency in generating reports and

Criteria	Sub-criteria	Item	Tableau	Power BI	Qlik	Superset	Sisense
Performance	Query Execution and Optimization	Throughput	✓	✓	✓	✓	✓
		Latency	✓	✓	✓	✓	✓
		Query Parallelization	✓	✓	✓	✓	✓
		Hardware Compatibility	✓	✓	✓	✓	✓
	Responsiveness	Loading Time for Plots	✓	✓	✓	✓	✓
		Loading Time for Datasets	✓	✓	✓	✓	✓
		Smoothness in interactions and navigation	✓	✓	✓	✓	✓
Scalability	Data Volume	Handling Large Datasets	✓	✓	✓	✓	✓
		Integration with Big Data Technologies	✓	✓	✓	✓	✓
		Indexing	✓	✓	✓	✓	✓
	User Scalability	Concurrent Users	✓	✓	✓	✓	✓
		High Availability	✓	✓	✓	✓	✓

Table 2.5: Evaluation Rubric for Performance and Scalability

visualizations. Additionally, we examine the tool's capacity to handle complex calculations and queries without significant lag. Regarding "Scalability" we delve into the tool's ability to accommodate growing data sets and increasing analytical demands. This category seeks to determine how well the tool adapts to the expanding data landscape of organizations. This rubric provides a structured framework for evaluating the performance and scalability of various BI solutions, assisting organizations in selecting tools that align with their processing speed and scalability requirements. Subsequent sections will provide a detailed exploration of each subcategory.

- **Query Execution and Optimization**: Query execution refers to the process of actually running a query against a database. It involves the steps taken by the database management system (DBMS) to retrieve the requested data.
- **Throughput**: All of the mentioned BI tools are capable of handling high throughput for executing queries.
- **Latency**: They also exhibit low latency in query execution.
- **Query Parallelization**: The BI tools can parallelize queries effectively.
- **Hardware Compatibility**: They are compatible with a wide range of hardware configurations.
- **Responsiveness**: Responsiveness refers to the ability of a system, application, or interface to quickly and effectively react to user interactions or external events, providing timely and smooth feedback.

- **Loading Time for Plots**: The tools exhibit efficient loading times for generating plots and visualizations.
- **Loading Time for Datasets**: They also demonstrate good performance in loading large datasets.
- **Smoothness in interactions and navigation**: Interactions and navigation within the tools are smooth and responsive.

"Scalability" is a sub-category that evaluates criteria related to the scalability of BI tools. This assessment focuses on how well the tools can adapt to increasing data volumes and analytical demands. It examines aspects such as the ability to accommodate growing data sets and handle higher levels of analytical complexity. This evaluation helps organizations determine whether a BI tool can effectively scale alongside their expanding data landscape. Subsequent sections will provide a detailed exploration of each criterion within this sub-category, offering insights into the scalability capabilities of various BI solutions.

- **Data Volume**: Data volume refers to the quantity of data generated, collected, and stored by an organization.
- **Handling Large Datasets**: All the mentioned BI tools are capable of effectively handling large datasets.
- **Integration with Big Data Technologies**: They can integrate with big data technologies for processing and analyzing large volumes of data.
- **Indexing**: The tools provide effective indexing mechanisms for optimizing data retrieval.

"User Scalability" refers to the ability of a system, application, or platform to accommodate an increasing number of users while maintaining performance, responsiveness, and user experience.

- **Concurrent Users**: The tools can handle a high number of concurrent users accessing the system simultaneously.
- **High Availability**: They ensure high availability, minimizing downtime and ensuring uninterrupted access.

The evaluation criteria for "Security and Access Control" are centered on the robustness of security measures and the management of user permissions within BI tools. Refer to Table 2.6 for a comprehensive rubric outlining this category and its subcategories. Under "Security," we scrutinize the tool's measures to safeguard sensitive or confidential data, including encryption, access restrictions, and compliance with industry standards. Additionally, we assess the tool's ability to detect and respond to security incidents. Within "Access Control," we evaluate the effectiveness of features that govern user permissions and data access levels, ensuring that information is accessible only to authorized personnel. This category aims to determine how well the BI tool protects data integrity and confidentiality. This rubric provides a structured framework for assessing the security and access control capabilities of various BI solutions, aiding

Criteria	Sub-criteria	Item	Tableau	Power BI	Qlik	Superset	Sisense
Security	Data Security	Data Encryption	✓	✓	✓	✓	✓
		Data Privacy Support[1]	✓	✓	✓	✓	✓
		Recovery & Backup	✓	✓	✓	✓	✓
	Network Security	Firewalls	✓	✓	✓	✓	✓
		Secure Transmission Protocol	✓	✓	✓	✓	✓
Access Control	Authentication and Authorization	Single Sign-On (SSO) Support	✓	✓	✓	✓	✓
		Role-Based Access Control (RBAC)	✓	✓	✓	✓	✓
	User Management	User Provisioning and Management	✓	✓	✓	✓	✓
		Customizable Permission	✓	✓	✓	✓	✓
		User Activity Tracking	✓	✓	✓	✓	✓
		User Behaviour Analysis	✓	✓	✓	✓	✓

Table 2.6: Evaluation Rubric for Security and Access Control

organizations in selecting tools that align with their security requirements. Further sections will offer an in-depth exploration of each subcategory.

The evaluation criteria for "Cost and Licensing" revolve around the financial aspects associated with adopting and using BI tools. Refer to Table 2.7 for a detailed rubric outlining this category and its subcategories. Under "Cost," we scrutinize the overall financial implications of adopting the BI tool, taking into account factors such as initial investment, ongoing maintenance, and potential additional costs. Additionally, we evaluate the perceived value derived from the tool relative to its cost. In terms of "Licensing," we assess the clarity and fairness of the licensing model, ensuring that it aligns with the organization's usage and scalability needs. This category aims to determine the cost-effectiveness and transparency of the financial arrangements associated with each BI tool. This rubric provides a structured framework for evaluating the cost and licensing aspects of various BI solutions, assisting organizations in selecting tools that align with their budgetary constraints and licensing requirements. Subsequent sections will provide a detailed exploration of each subcategory. This sub-category assesses the cost-related criteria:

- **Licensing Model**: A licensing model outlines the terms and conditions for using software or intellectual property, including permissions, restrictions, and pricing.
- **Subscription-based**: Tableau, Power BI, Qlik, and Sisense offer a subscription-based licensing model, providing flexibility in payment options.
- **Perpetual License**: Tableau, Power BI, Qlik, and Sisense also support perpetual licenses, which involve a one-time purchase for ongoing use.
- **Total Cost of Ownership**: Total Cost of Ownership (TCO) encompasses the complete expenses associated with acquiring, operating, and maintaining a product, system, or service over its entire lifespan. This includes initial purchase costs, operational expenses, maintenance fees, and potential.

- **Initial Costs**: Tableau, Power BI, Qlik, and Sisense may involve initial costs such as setup fees or license purchase fees.
- **Ongoing Maintenance**: These tools may also require ongoing maintenance costs for updates, support, and additional services.

This sub-category evaluates criteria related to licensing considerations. It examines aspects such as the flexibility of licensing options, including pricing models, user tiers, and scalability. Additionally, it assesses whether the licensing terms align with the organization's budget and requirements. This evaluation helps organizations make informed decisions about the most suitable licensing arrangements for their needs.

- **User-Based Licensing**: Tableau, Power BI, Qlik, and Sisense offer licensing based on specific named users with assigned permissions.
- **Concurrent Users**: These tools also support licensing based on the number of users accessing the platform concurrently.
- **Subscription-Based Licensing**: Monthly Subscription: Tableau, Power BI, Qlik, and Sisense provide the option for users to subscribe on a monthly basis.
- **Annual Subscription**: These tools also offer the option for users to subscribe on an annual basis, often at a discounted rate compared to monthly subscriptions.

In this chapter, we developed an evaluation framework for BI tools to help organizations navigate the complex landscape of data analytics solutions. We identified key components of effective BI tools, such as data integration, visualization, analytics, and user-friendliness. Our analysis focused on leading platforms—Power BI, Tableau, Qlik Sense, Sisense, and Apache Superset—providing a detailed comparison of their strengths and limitations. Each tool was assessed based on our established criteria, revealing differences in their capabilities. Some tools excelled in data visualization, while others stood out for data integration or ease of use, which is crucial for non-technical users. This evaluation highlighted the importance of aligning tool selection with organizational goals and user needs.

Our analysis also considered the rapidly evolving BI landscape, emphasizing the need for adaptable tools that can meet changing business requirements. By understanding how each tool performs across various dimensions, decision-makers can make informed choices to enhance their data-driven processes. Selecting a BI tool is a strategic decision with significant implications for an organization's data capabilities. The framework developed in this chapter serves as a foundational tool that can be refined as new solutions emerge and organizational needs evolve. The next chapter will introduce a user survey to validate our evaluation criteria, gathering insights from actual users to ensure the framework's relevance and applicability in real-world settings.

Criteria	Sub-criteria	Item	Tableau	Power BI	Qlik	Superset	Sisense
Cost	Licensing Model	Subscription-based	✓	✓	✓	×	✓
		Perpetual License	✓	✓	✓	×	✓
	Total Cost of Ownership	Initial Costs	✓	✓	✓	×	✓
		Ongoing Maintenance	✓	✓	✓	×	✓
Licensing	User-Based Licensing	Named Users	✓	✓	✓	×	✓
		Concurrent Users	✓	✓	✓	×	✓
	Subscription-Based Licensing	Monthly Subscription	✓	✓	✓	×	✓
		Annual Subscription	✓	✓	✓	×	✓

Table 2.7: Evaluation Rubric for Cost and Licensing

Chapter 3

BI Tools Survey

This chapter presents the details of our BI Tools Survey, focusing on designing and implementing a comprehensive survey to understand the usage, preferences, and challenges associated with BI tools among industry professionals. A total of 28 participants, all industry experts, were surveyed to gather relevant insights. The chapter begins by discussing the user survey design methodologies, including the Goal-Question-Metric (GQM) framework, which guides research objectives with measurable outcomes. It then provides an overview of the survey methodology, detailing the systematic approach to crafting, distributing, and analyzing the survey. Through this structured exploration, the chapter aims to validate the evaluation criteria established in earlier sections, providing insights into the effectiveness of BI tools in real-world applications and offering a foundation for the subsequent analysis of survey results.

3.1 Background on User Survey Design Methodologies

Goals define the primary aims and purpose of a study, setting the overall direction for the research. These goals are then translated into specific research questions, which help to focus the investigation on achieving the intended objectives. Each question is associated with measurable metrics, which serve as criteria to assess progress and success. These metrics provide indicators that allow for the evaluation of research outcomes and the extent to which the goals have been achieved. By linking goals, questions, and metrics, the research process becomes more organized and focused, ensuring that the objectives are clear and the results are measurable.

3.1.1 Goal-Question-Metric (GQM) Framework

It is a systematic approach widely used in research to define, measure, and analyze goals, questions, and associated metrics. Introduced by Basili and Weiss in the early 1980s, GQM provides a structured method for aligning research activities with the overarching goals of a study. The framework involves three key components: The GQM approach, introduced by Basili and Rombach in the late 1970s, has become a prominent methodology in software engineering and other fields for structuring and guiding empirical studies. Basili et al. emphasized the importance of aligning software engineering processes with organizational goals to improve quality and productivity. GQM provides a systematic framework for defining goals, de-

riving specific questions to address those goals, and establishing measurable metrics to assess progress and outcomes. The GQM framework has evolved over the years, adapting to various domains and research contexts. In software engineering, GQM has been widely used for process improvement and empirical studies. Kitchenham et al. applied GQM to software metrics, demonstrating its effectiveness in assessing the impact of software engineering technologies. Researchers have explored extensions and variations of the GQM model to enhance its applicability. For instance, GQ models focus on goals and questions without explicitly incorporating metrics. GQ models retain the goal-oriented structure while providing flexibility in metric definition. The GQM+Strategies approach introduced by van Solingen and Berghout extends the original GQM model by incorporating strategies as a crucial element [37, 35, 6, 36]. This extension aims to bridge the gap between organizational goals and concrete actions by explicitly addressing the strategies employed to achieve goals.

The GQM framework was selected for this study due to its effectiveness in aligning research objectives with concrete measurement strategies. In the context of evaluating BI tools, GQM allows for a clear definition of what needs to be achieved (goals), how to achieve it (questions), and how to measure progress (metrics). This structured approach ensures that the research is purposeful, measurable, and directly tied to the overarching objectives.

In this section in addition to understanding the variety of survey questions, it is crucial to recognize the different types of surveys used to collect and analyze data on BI tools. Each type of survey serves a unique purpose, helping researchers to gather and interpret data in ways that best suit their study objectives.

Surveys can be categorized into various types, each serving distinct purposes and providing unique insights. Cross-sectional surveys collect data at a single point in time, offering a snapshot of attitudes, behaviors, or characteristics within a specific timeframe. In contrast, longitudinal surveys gather data from the same respondents over multiple time points, allowing for the tracking of changes, trends, and developments. Descriptive surveys aim to characterize a population or phenomenon, providing insights into the prevalence of specific traits or behaviors. Explanatory surveys, on the other hand, seek to uncover the reasons behind observed phenomena by exploring causal relationships and providing insights into changing factors.

To gather comprehensive data, we employ a variety of question formats: open-ended questions capture detailed thoughts, closed-ended questions offer clear choices, dichotomous questions provide simple yes or no answers, ranking questions help understand preferences, multiple-choice questions allow for structured responses, Likert scale questions measure levels of agreement, and matrix questions explore multiple aspects in one view. Additionally, we use cross-sectional surveys to get a snapshot of attitudes and behaviors at a single point in time, longitudinal surveys to track changes and trends over multiple time points, descriptive surveys to detail the characteristics of a population, and explanatory surveys to uncover reasons behind observed phenomena. This diverse approach helps us gain a thorough understanding of BI tool usage and user experiences.

3.1.2 Survey Questions

This section explores the different types of survey questions used in our study on BI tools. Understanding the variety of survey questions is essential for gathering comprehensive data, as different question types serve different purposes and provide different insights.

Our survey questions are crafted to gather comprehensive data about BI tools using the following types of questions. Open-ended questions allow free-form responses, capturing unstructured data and insights into respondents' thoughts. Closed-ended questions offer predefined response options, such as multiple-choice, yes/no, and Likert scale, facilitating straightforward data analysis. Dichotomous questions present two options, often "yes" or "no" suitable for exploring binary choices. Ranking questions require respondents to prioritize or rank items, which is useful for understanding preferences within a set. Multiple-choice questions provide a set of options, allowing respondents to select one or more choices, making them effective for structured data collection. A *Likert scale* is a commonly used tool in survey research to measure respondents' attitudes, perceptions, or opinions on a set of statements or questions. Named after its creator, psychologist Rensis Likert, the scale typically ranges from strongly agree to strongly disagree, providing respondents with a range of options to express their agreement or disagreement. In the user survey conducted as part of this research, Likert scales were employed to gather qualitative insights into industry partners' perceptions of BI tools. The Likert scale allows participants to express their opinions on various aspects, such as usability, satisfaction, and preferences. This structured format enables the collection of valuable data, which can then be analyzed to understand trends, patterns, and overall sentiments regarding BI tools. By utilizing both the GQM framework and Likert scales, this research aims to combine a structured approach to goal-oriented research with the richness of qualitative insights gathered from user perspectives, providing a comprehensive and balanced evaluation of BI tools

3.2 Survey Methodology Overview

A total of 28 participants were involved in the survey. The target participants were selected based on their professional experience with BI tools, ensuring that the survey captured insights from individuals who actively use these tools in their daily operations. These participants included data analysts, IT managers, and business users from various industries, such as retail, technology, and finance. Participants were recruited through professional networks, collaboration with industry partners like Industry Collaboration, and outreach to small and medium-sized enterprises (SMEs) that rely on BI tools for their business analytics. This diverse group of participants provided a comprehensive perspective on the challenges and requirements associated with using BI tools.

The survey follows a GQM approach serves as the guiding framework for this survey, offering a structured methodology to systematically define and measure the effectiveness of BI tools. Beginning with overarching organizational goals, GQM cascades into specific, measurable questions that inform the identification of relevant metrics. This hierarchical structure ensures a goal-oriented and comprehensive evaluation of BI tools, aligning the survey objectives with organizational aspiration The primary focus of this research is to validate a set of criteria designed for the systematic evaluation of tools within a general framework. The central objective is to ensure an unbiased assessment that transcends specific industry requirements and objectives, thereby providing a broader scope for evaluation. The validation process aims to establish a foundation that can be applied universally, fostering a comprehensive and impartial approach to tool assessment.

Building upon this groundwork, the second phase of the study aligns with existing method-

ologies employed in comparing BI tools. Notably, what sets our approach apart is the integration of the meticulously validated criteria developed in the initial stage. This strategic incorporation enhances the rigor and objectivity of the comparison, elevating the quality of evaluation in a manner that reflects a nuanced understanding of tool performance across diverse contexts.

3.2.1 Section A: Validating Evaluation Criteria for BI Tools

The survey followed a user-centered approach, where the criteria for evaluating the BI tools were first verified by the participants themselves. In the initial phase, users were asked to rank and validate the evaluation criteria based on their experiences and preferences when working with BI tools. This ensured that the criteria used for the evaluation were directly relevant to the needs and priorities of actual BI tool users. *In the second phase of the survey*, the same user-verified criteria were then applied to evaluate selected BI tools, ensuring a comprehensive and user-driven assessment. This approach not only strengthens the relevance of the evaluation but also provides valuable insights into the specific features and functionalities that matter most to end users in real-world scenarios.

In this section, we analyze respondents' demographic data, evaluate their data integration and connectivity methods, and assess their use of BI tools for data visualization, reporting, and analysis. We begin by collecting background information on respondents, such as industry, company size, position, and experience with BI tools. This demographic understanding allows us to tailor the evaluation to each respondent's specific context, ensuring that the BI tools are assessed in a way that is relevant to their particular needs and environments.

We then explore how respondents connect and integrate data within the BI tools ecosystem, focusing on the use of various data sources like spreadsheets, databases, cloud storage, and APIs. This includes evaluating data transformation capabilities and real-time data handling to determine how effectively these tools manage diverse and complex data integration requirements. Following this, we assess the BI tools' data visualization and reporting functionalities, looking at the variety and complexity of available visualization options, the degree of customization, interactivity, and the ability to generate insightful reports that support data-driven decision-making. Additionally, we evaluate the tools' data analysis capabilities, examining their support for various types of analysis—such as diagnostic, prescriptive, predictive, and descriptive analytics—along with advanced features like predictive modeling, natural language processing, and machine learning.

Finally, we assess the overall user-friendliness of the BI tools by examining their interface design, ease of use, and accessibility. This includes evaluating features like drag-and-drop capabilities, guided workflows, and responsive design. We also consider the tools' performance and scalability, focusing on their ability to handle large data volumes and complex queries efficiently, as well as their adaptability to evolving business needs. To ensure the evaluation is thorough and aligned with organizational goals, we implement the Goal-Question-Metric (GQM) approach, which involves setting strategic objectives, translating them into targeted questions, and using the responses to guide iterative improvements in BI tool functionality. The following questions are specifically related to data integration and connectivity.

- **Q: Please rank the following high-level criteria for the evaluation of BI tools based on their importance according to your experience with these tools.** Respondents are

asked to rank criteria such as Data Integration and Connectivity, Data Visualization and Reporting, Data Analysis and Capabilities, Scalability and Performance, User-Friendly Interface, and Cost and Licensing. This prioritization helps to focus on developing features that align most closely with user needs.

- **Q: Which feature related to data connectivity of sources is more important to you?** Choices include Availability of Connections and Data Preprocessing, Cleaning, and ETL. Insights from this question help optimize BI tools to support effective data integration and management, enhancing user experience and tool efficiency.

These are a few sample questions from the survey that ask about data Visualization and reporting, and data analysis capabilities:

- **Q: Which types of visualizations do you primarily use in your reports?** Options such as Bar charts, Line charts, Pie charts, Scatter plots, Heatmaps, Tables, Area charts, Treemaps, and Bubble charts are included. This feedback assists in understanding users' visualization needs and preferences, guiding the development of more effective visualization capabilities in BI tools.

- **Q: How much customization is available for visualizations in your BI tool?** Responses can range from Extensive customization to no customization options. This question assesses the adaptability of BI tools to user-specific needs, which is critical for ensuring that tools are versatile and user-friendly.

- **Q: What are the primary challenges you face in your data analysis process?** Users identify challenges like Data quality issues, Lack of advanced analysis skills, Insufficient computing resources, and Difficulty in integrating diverse data sources. Tool enhancements can significantly improve user satisfaction and analytical accuracy by addressing these challenges.

- **Q: Does your BI tool offer cross-platform compatibility (e.g., desktop, web, mobile)?** Compatibility options are vital for accessibility, with choices ranging from seamless compatibility to no plans for future updates. This ensures that BI tools are accessible and functional across various devices and platforms, enhancing user experience.

There are several questions in the survey that ask about performance and scalability and security, and access control, including the following:

- **Q: How would you rate the performance of your current data evaluation tool?** Performance is rated from Excellent to Poor. This direct feedback provides valuable insights into the tool's capability to handle demanding tasks and large datasets, guiding further optimizations.

- **Q: To what extent does your current data evaluation tool scale with increasing data volumes?** Options from Highly scalable to Not scalable at all are given. Scalability is critical as it determines the tool's ability to grow with the organization's data needs.

- **Q: How satisfied are you with the current user authentication methods (e.g., username/password, multi-factor authentication) in place for accessing BI tools?** Satisfaction levels from Extremely satisfied to Extremely dissatisfied are gauged to assess security effectiveness.

- **Q: Do you believe the current access controls sufficiently limit data access to only authorized personnel?** Responses from Strongly agree to Strongly disagree help evaluate the adequacy of the tool's security measures in protecting sensitive data.

3.3 Section B: Evaluation of Selected Tools

In this section of the survey, we continue applying the GQM approach by focusing on the detailed evaluation of five selected BI tools. These tools were chosen based on their preliminary alignment with the organizational needs identified in Section A. This phase's main goal is to thoroughly assess each tool's performance, user interface, scalability, and security features. Each question has been carefully designed to elicit specific information that will help quantify the strengths and weaknesses of these tools, thereby assisting stakeholders in making informed decisions about which tool best suits their operational needs.

The followings are some of the main questions from the second section of the survey:

- **Q: How well does the BI tool handle real-time data analytics?** Options such as Excellent, Good, Adequate, Poor, and Not applicable are included. This feedback helps in assessing whether the BI tool can provide timely and efficient data processing, which is crucial for organizations that require immediate insights for decision-making.

- **Q: What is the level of support and resources available for new users of this BI tool?** Choices ranging from Comprehensive, Adequate, Limited, to None are provided. This feedback assists in understanding how effectively the tool supports new users, impacting ease of adoption and overall user satisfaction with the tool's learning curve and resource availability.

- **Q: Can you describe the scalability of this BI tool in handling growing data volumes?** Options such as Highly scalable, Moderately scalable, Not scalable, and Uncertain are available. This question is crucial for evaluating whether the BI tool can handle increasing data volumes, which is essential for organizations planning to scale operations or data analytics capabilities.

- **Q: How would you rate your overall satisfaction with the security features of the BI tool?** Feedback options include Very satisfied, Satisfied, Neutral, Dissatisfied, and Very dissatisfied. This question helps determine the effectiveness of the security measures implemented within the BI tool, which is vital for protecting sensitive and confidential data.

- **Q: To what extent does the BI tool integrate with other enterprise applications and data sources?** Integration capabilities are assessed with options like Fully integrated, Mostly integrated, Partially integrated, and Not integrated. This feedback is essential

for understanding how seamlessly the BI tool can be incorporated into the existing IT infrastructure, affecting its usability and efficiency.

- **Q: What are the customization options available in this BI tool, and how do they enhance user experience?** Customization options such as Extensive, Moderate, Limited, and None are included. This feedback aids in evaluating how well the BI tool can be tailored to meet specific organizational needs and user preferences, enhancing overall utility and satisfaction.
- **Q: How effective is the data visualization capability of the BI tool in presenting complex data insights?** Effectiveness of data visualization is rated with options like Very effective, Effective, Adequate, and Ineffective. This feedback is critical for assessing the BI tool's ability to turn complex data sets into understandable, actionable insights through visual representation, impacting decision-making processes.

We now present a summary of our survey results from the two sections.

3.4 Survey Results: Section A

Section A of the survey gathered insights on high-level evaluation criteria for BI tools, focusing on data integration, visualization, analysis, performance, and security. The responses highlight which criteria users find most critical for decision-making, providing a foundation for evaluating BI tools and guiding the analysis in Section B.

3.4.1 BI Tools Usage Patterns

The survey provided valuable insights into the usage of various BI tools across a range of industries. Among the tools evaluated, Tableau and Power BI emerged as the most widely adopted, indicating their strong presence and preference in the market. These tools are known for their advanced data visualization capabilities, user-friendly interfaces, and robust feature sets, which make them suitable for a wide array of business environments, from small enterprises to large corporations. Tableau leads the pack with 38.5% of respondents indicating its use, followed by Power BI at 25.6%. These two tools alone account for most of the market, collectively representing 64.1% of users. Tableau's strength lies in its sophisticated data visualization capabilities, which allow users to create detailed, interactive dashboards that can be easily shared across teams. On the other hand, Power BI is highly valued for its seamless integration with other Microsoft products, particularly Excel and Azure, making it a versatile choice for organizations already invested in the Microsoft ecosystem. The popularity of these tools suggests they are well-suited for handling large datasets and complex data relationships and providing actionable insights in a visually compelling manner. Figure 3.1 shows the popularity of BI tools among survey respondents.

Other tools like QlikView (12.8%), Looker (10.3%), and Google Data Studio (7.7%) also see significant use, though they lag behind Tableau and Power BI. QlikView is noted for its dynamic data exploration capabilities. Looker for its unified data model, and Google Data Studio for its ease of use and integration with Google products. Tableau and Power BI dominate

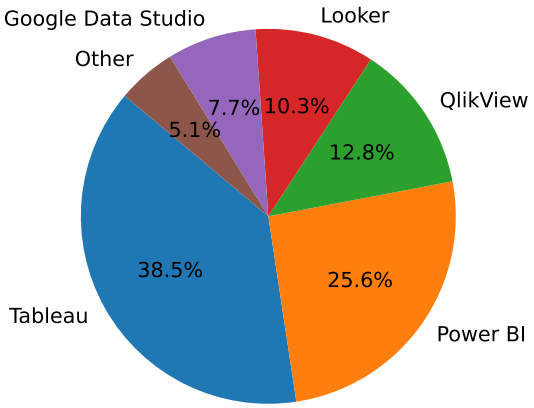


Figure 3.1: Popularity of BI Tools Among Survey Respondents

the market despite their presence, collectively holding over 64% of the share. This dominance underscores their reliability and scalability in meeting diverse data analytics needs. However, the continued use of QlikView, Looker, and Google Data Studio suggests that specific features, cost, or integration preferences may drive organizations to choose these alternatives.

Data Source	**Priority Level (%)**
Relational/NoSQL Databases	85
Cloud-based Databases	85
APIs	80
Flat Files (Excel, CSV)	60

Table 3.1: Priority Levels of Data Sources for BI Tools Integration and Connectivity

3.4.2 Data Integration and Connectivity

The survey results emphasize the critical importance of data integration and connectivity in BI tools. Respondents prioritized seamless access to both relational and NoSQL databases, as well as cloud-based databases, reflecting the need for robust connectivity in today's data-driven business environment. APIs were also highly ranked, underscoring the importance of integrating external data and services. Although flat files like Excel and CSV are considered less critical, they still play a role in many data strategies. Table 3.1 shows the priority levels of various data sources for BI tools integration and connectivity.

The strong focus on cloud connectivity highlights the growing reliance on scalable cloud solutions. The ability to integrate seamlessly with both on-premise and cloud-based environments is increasingly important, and organizations should prioritize BI tools with strong data integration capabilities to manage and analyze data across multiple platforms effectively.

Regarding data visualization and reporting, bar charts (85%), line charts (78%), and pie charts (72%) emerged as the most commonly used visualizations, valued for their clarity in conveying insights. Scatter plots (56%) and heatmaps (48%) are also frequently used for more complex data relationships. Although less common, specialized plots like box plots and violin

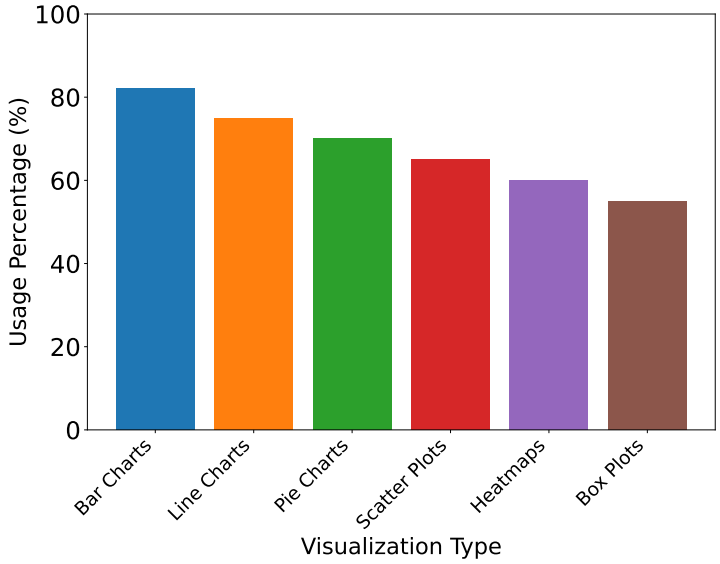


Figure 3.2: Usage Frequency of Various Data Visualization Types in BI Tools

plots are utilized by 35% of respondents. Additionally, 60% of respondents indicated that their BI tools offer extensive customization options, allowing for significant flexibility in modifying visualizations to meet specific needs.

The preference for bar, line, and pie charts highlights their importance in straightforward data reporting, while the availability of customization and interactivity features, such as drill-downs and dynamic filters, is essential for in-depth data exploration and effective decision-making. Figure 3.2 illustrates the usage frequency of various data visualization types in BI tools.

Organizations should choose BI tools that offer a wide range of visualization options, along with robust customization and interactivity features, to support diverse analytical tasks and deliver actionable insights effectively.

Capability	**Priority (%)**
Data Preparation	95
Statistical Analysis	75
Time-Series Analysis	85
Machine Learning	80
Predictive Analytics	80
Data Quality Management	70

Table 3.2: Prioritization of Data Analysis and Capabilities in BI Tools

3.4.3 Data Analysis and Capabilities

The survey results highlight key priorities and challenges in the data analysis capabilities of BI tools. Data preparation and manipulation emerged as the most crucial aspects, underscoring the

need for high data quality before analysis. Statistical and time-series analyses are also highly valued, with time-series features frequently used for tracking trends and forecasting outcomes. Advanced analytics, such as machine learning and predictive analytics, are increasingly important as organizations seek deeper insights from their data. Table 3.2 shows the prioritization of various data analysis capabilities according to the survey respondents.

The emphasis on data preparation reflects its critical role in ensuring effective analysis. The growing demand for advanced analytics indicates a shift toward more predictive insights. However, challenges like data quality issues, insufficient analytical skills, and limited computing resources can impede these efforts.

Organizations should invest in BI tools that provide robust data preparation and advanced analytical features. Addressing challenges related to data quality and resource availability will be essential to maximizing the value of these tools. Collaboration features within BI platforms can also enhance teamwork and the sharing of insights.

3.4.4 User Friendliness and Ease of Use

The survey underscores the importance of user-friendliness in BI tools. A well-designed user interface was the top priority for 65% of respondents, followed by ease of use and a short learning curve. While 60% found it somewhat easy to search for features or reports, 25% noted room for improvement in making these processes more intuitive. Cross-platform compatibility was a key feature for 70% of respondents, with most BI tools offering seamless access across desktop, web, and mobile platforms. Table 3.3 shows the prioritization of various user interface features according to the survey respondents.

The focus on user-friendliness suggests that while analytical capabilities are vital, the usability of a BI tool significantly impacts its adoption and effectiveness. Features like drag-and-drop interfaces, interactive dashboards, and cross-platform compatibility are particularly valued for enhancing user experience and accessibility.

Organizations should prioritize BI tools with intuitive interfaces and easy navigation to maximize adoption and satisfaction. Ensuring seamless access across multiple platforms is essential in today's mobile and remote work environments. User-friendly tools that require minimal training will help organizations leverage BI capabilities more effectively.

Feature	**Priority (%)**
Well-Designed User Interface	65
Short Learning Curve	60
Overall Ease of Use	70
Search for Features	60
Cross-Platform Compatibility	70
Interactive Dashboards	75
Drag and Drop Interfaces	68

Table 3.3: Prioritization of User Interface Features in BI Tools

3.4.5 Performance and Scalability

The data indicates a strong emphasis on the scalability of BI tools, with a significant portion of users rating their tools' scalability as high. This suggests that as data volumes grow, the ability of BI tools to manage and process this data efficiently is becoming increasingly important. Figure 3.3 presents a stacked bar chart that categorizes the performance of BI tools across various functional areas: Performance, Scalability, Datasets, Users, and Processing. The graph segments ratings into three distinct levels—Excellent, Good, and Average—each denoted by a specific color. This visual representation allows for an at-a-glance assessment of where BI tools are excelling and where they might be lacking. Notably, the Performance and Scalability categories have a substantial share of Excellent ratings, indicating robust capabilities in these areas. Conversely, the Processing category shows a more balanced distribution of ratings, suggesting areas where improvements could be beneficial. This chart is instrumental in helping stakeholders identify strengths and areas for development within their BI tools, facilitating informed decision-making based on user feedback.

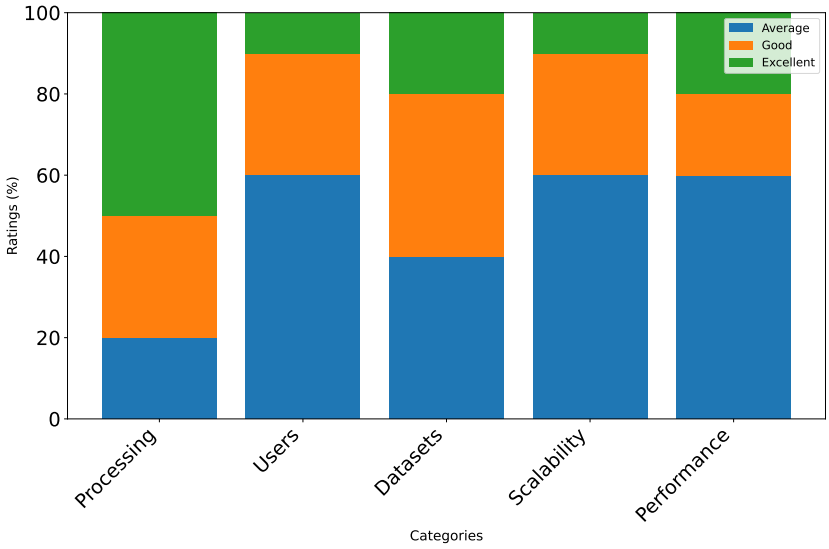


Figure 3.3: Distribution of Ratings Across Categories in Business Intelligence Tools

Despite high ratings in scalability and handling large datasets, real-time processing appears to be a bottleneck, highlighting an area for potential enhancement. The ability to process data in real-time is crucial for making timely decisions, indicating that improvements in this area could enhance the overall utility of BI tools.

Organizations should prioritize BI tools that provide robust scalability and performance capabilities, especially those that can handle large datasets and support multiple users effectively. Ensuring that these tools also offer efficient real-time processing will be key to maintaining competitiveness and operational effectiveness in data-driven environments.

3.4.6 Security and Access Control

The data indicates a generally positive perception towards the security measures implemented in BI tools, with notable satisfaction in user authentication methods. However, there are areas such as data privacy compliance and granularity of permissions where user satisfaction is relatively lower, suggesting room for improvement. Figure 3.4 illustrates the levels of satisfaction with various security features, capturing the areas where users feel improvements are necessary.

Figure 3.4 presents a bar graph that quantitatively assesses satisfaction levels across various security features within business intelligence tools. The specific features analyzed include User Authentication, Access Control, Granularity of Permissions, and Data Privacy Compliance. The varying heights of the bars depict the differing degrees of user satisfaction for each security feature, with User Authentication and Access Control achieving higher satisfaction rates. This visual differentiation underscores the effectiveness of certain security measures and highlights potential areas for improvement in others. Such insights are crucial for developers and administrators aiming to enhance the security and compliance capabilities of their BI tools.

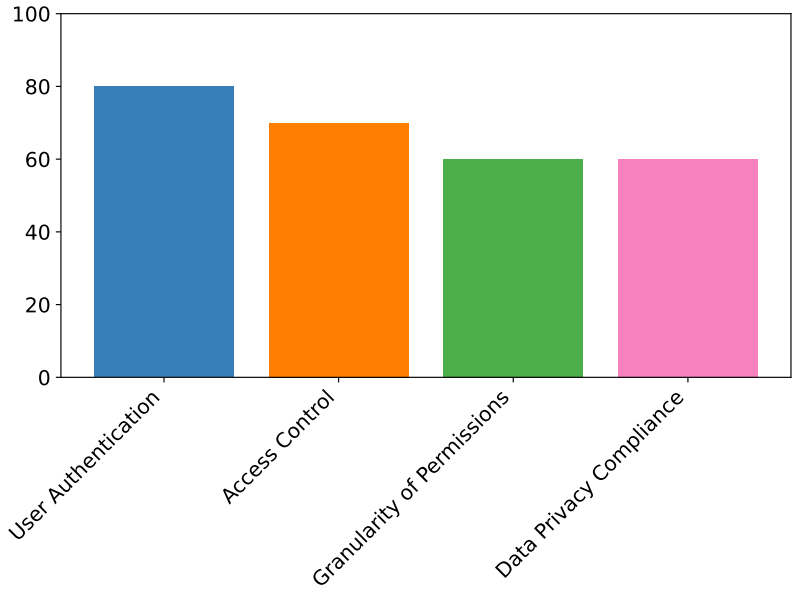


Figure 3.4: Security Features Satisfaction

Security is a critical component of BI tools, particularly in industries handling sensitive data. The survey highlights that while users find authentication measures satisfactory, the management of permissions and compliance with data privacy laws are not as robust. This could potentially affect the adoption of BI tools in sectors with stringent security requirements.

3.5 Survey Results: Section B

Our analysis shows that 70% of respondents rate their BI tool's integration capabilities as either Adequate or Comprehensive, indicating a generally positive perception among users. However,

the 30% who reported Limited Integration highlight a significant area for improvement. Survey comments may further clarify specific integration challenges, such as compatibility issues with legacy systems or difficulties in real-time data synchronization. This structured approach ensures a thorough analysis, providing a comprehensive view of how BI tools meet integration needs in modern enterprises.

The survey results reveal that BI tools like Tableau, Power BI, QlikView, Sisense, and Google Data Studio are widely used, with varying degrees of integration effectiveness. Tableau and Power BI show integration levels ranging from limited to comprehensive, suggesting that integration potential may depend on specific organizational requirements and technical environments. Lesser-known tools like Klipfolio and IBM Cognos also play a role in certain setups, indicating their functional relevance. Overall, the majority of BI tools mentioned in the survey offer at least adequate integration with organizational systems, supporting effective data management and decision-making. This underscores the importance of selecting BI tools that align well with an organization's technological infrastructure to maximize data utilization and operational efficiency.

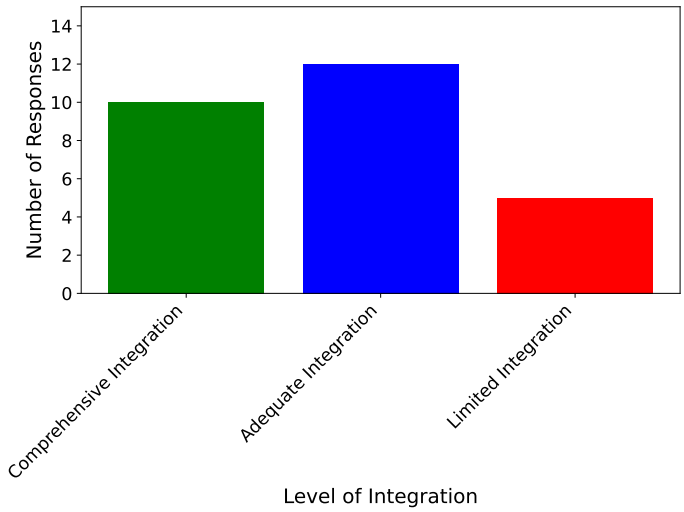


Figure 3.5: Distribution of Responses Across Levels of Integration

The survey results reveal varied perspectives on the effectiveness of BI tools in facilitating collaboration and sharing insights among teams or departments. Tableau shows significant variability, with responses ranging from slightly well to not well at all, indicating its effectiveness in collaboration depends heavily on the organizational context. Qlik exhibits a similar range, from not well at all to extremely well, suggesting its utility also varies by organization. In contrast, Power BI generally receives positive feedback, with many users rating it as very well, highlighting its strong capability to support collaborative efforts. Google Data Studio and BigQuery are moderately effective, as indicated by moderately well and slightly well responses, implying they may not fully meet all collaborative needs. Sisense and Klipfolio show varied effectiveness, with instances of both excellent and less satisfactory performance, emphasizing the need for careful selection of BI tools to align with specific organizational requirements.

The survey also indicates that Power BI and Google BigQuery are highly effective in han-

dling real-time data, consistently receiving very good and excellent ratings. Tableau performs well, with most users rating it as good or very good. In contrast, QlikView and QlikSense show more variability, with responses ranging from fair to excellent, suggesting their effectiveness depends on organizational context. SAP BusinessObjects and Oracle BI receive moderate to good feedback, indicating adequate real-time data capabilities. Overall, the data suggests that while many BI tools are effective in real-time data handling, the choice of tool should align with the specific needs of the organization.

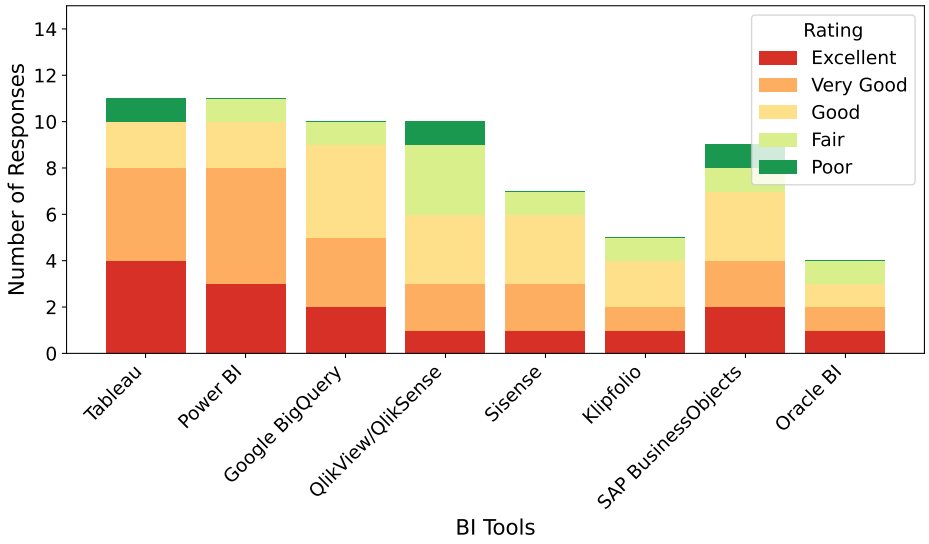


Figure 3.6: Real-Time Data Handling by BI Tools

In evaluating BI tools for customizable dashboards and reports, Power BI and Tableau emerged as top performers. Power BI received the highest ratings, with most users rating it as excellent or good, reflecting its strong capability to meet diverse user needs. Tableau also performed robustly, with similar ratings highlighting its flexibility in customization. Google BigQuery and QlikView received positive evaluations, though some average ratings suggest areas for improvement. Sisense, Klipfolio, and SAP BusinessObjects were generally well-rated, indicating solid performance in specific use cases. Oracle BI, however, had more modest evaluations, with no excellent ratings and a mix of good and average responses, pointing to room for improvement. These results suggest that while many BI tools offer effective customization, Power BI and Tableau are particularly well-regarded for their versatility.

The survey also revealed varying levels of satisfaction with data cleansing and preprocessing capabilities, essential for data accuracy. Tableau received mixed feedback, with a range of satisfaction levels from somewhat satisfied to extremely dissatisfied. Power BI showed a similar split, with some users satisfied and others not. Google BigQuery generally received positive feedback, while QlikView had mixed reviews. Other tools like Sisense, Klipfolio, SAP BusinessObjects, and Oracle BI received lower satisfaction ratings. These findings indicate that while some tools perform well in data cleansing, the choice should align with organizational needs for data accuracy and reliability.

When comparing BI tools for their ability to provide actionable insights, Tableau and Power BI were top performers. Tableau received a high number of superior and above-average ratings,

as did Power BI, indicating their effectiveness in delivering meaningful insights beyond raw data. Google BigQuery and QlikView had mixed reviews, showing variability in effectiveness based on context. Sisense and SAP BusinessObjects generally performed well, though they had some lower ratings. Oracle BI showed a balanced distribution of ratings, with some superior but also average and below-average responses. Overall, the data suggests that while most BI tools can provide actionable insights, Tableau and Power BI are particularly well-regarded for their ability to offer strategic insights that influence business decisions.

Figure 3.7 presents a stacked bar chart evaluating BI tools based on their effectiveness in providing actionable insights. Tableau and Power BI stand out with a significant proportion of superior and above-average ratings, illustrating their strength in supporting strategic decision-making.

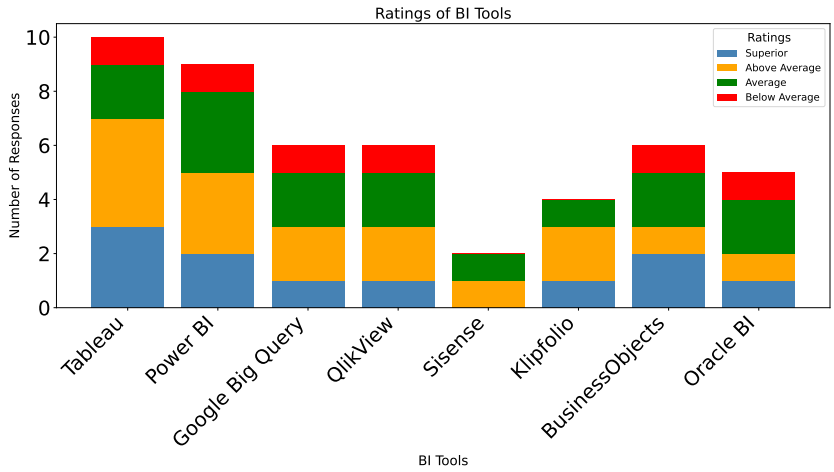


Figure 3.7: Rating Distribution Across BI Tools for Providing Actionable Insights

The survey results indicate that BI tools' adaptation to changing data privacy and compliance regulations varies among different tools. Tableau demonstrated strong adaptability, with most respondents highlighting its effective response to regulatory changes, though a few noted room for improvement. Power BI also performed well, showing generally positive adaptability, with most users feeling it manages regulatory shifts adequately. In contrast, Google BigQuery, QlikView, and Sisense received mixed feedback, suggesting they have strengths in some areas but also face challenges in fully adapting to evolving data privacy and compliance requirements. showing both strengths and areas for improvement.

Regarding the capability to facilitate seamless data sharing and collaboration with external stakeholders, Tableau was rated highly, with many respondents considering it extremely crucial or crucial, reflecting its strong performance in this area. Power BI also received positive feedback, with most users emphasizing its importance for collaboration. In contrast, Google BigQuery and QlikView had more varied responses; while some users rated their collaboration capabilities as highly crucial, others perceived them as less important. Overall, these results suggest that while Tableau and Power BI are well-regarded for their compliance and collaboration capabilities, other tools like Google BigQuery and QlikView may need to enhance these aspects to better meet industry expectations.

Figure 3.8a presents a stacked bar chart that evaluates how well various BI tools adapt to

changing data privacy and compliance regulations. The chart includes responses for tools such as Tableau, Power BI, Google BigQuery, QlikView, Sisense, Klipfolio, SAP BusinessObjects, and Oracle BI. Each bar is segmented to show the different levels of adaptability, ranging from Not well at all to Very well.

Figure 3.8b presents a stacked bar chart that compares the perceived importance of various BI tools for data sharing and collaboration. Each bar in the chart represents a different BI tool, such as Tableau, Power BI, Google BigQuery, QlikView, Sisense, and Klipfolio. The segments within each bar correspond to different levels of importance, ranging from Not Important to Extremely Crucial. The height and distribution of these segments provide insights into how essential respondents consider each tool. Tools with taller bars and a higher proportion of Extremely Crucial or Crucial ratings are deemed more important for data sharing and collaboration. In comparison, those with shorter bars or more Not Important segments may be viewed as less critical.

3.6 Summary of Results, Discussion, and Takeaways

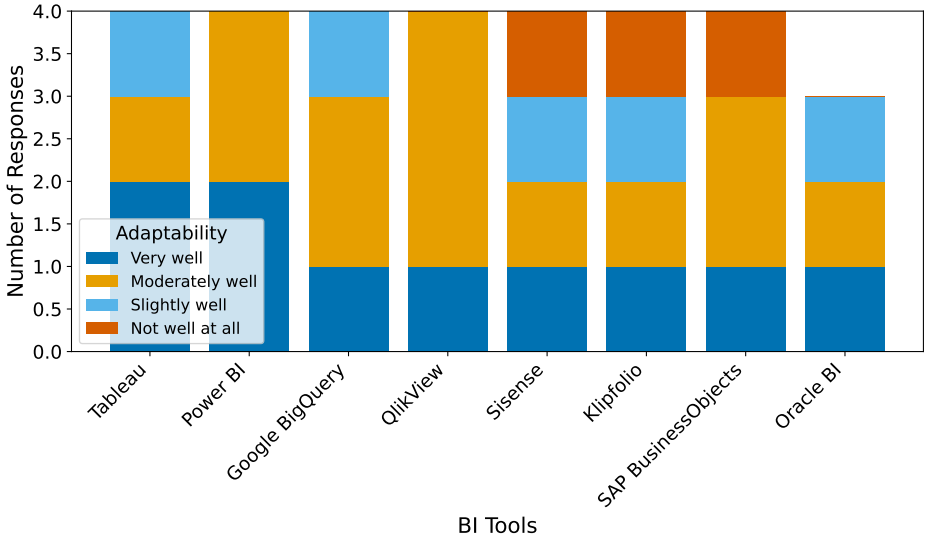
The survey results provided valuable insights into the adoption and utilization of BI tools within the industry. One of the key takeaways is that most users prioritize data integration and connectivity features, followed closely by ease of use and the availability of advanced analytics. This highlights the growing need for BI tools that can seamlessly integrate with various data sources and offer user-friendly interfaces without requiring extensive technical expertise. *Another significant finding is the importance of security and access control*, with many participants indicating that they face challenges in managing access to sensitive data across multiple platforms. These insights are particularly useful for BI tool developers and companies in the industry, as they emphasize the need for enhanced security measures, better integration capabilities, and tools that cater to a diverse user base, from technical users to business analysts.

The evaluation of BI tools based on user feedback and performance criteria reveals key insights into their strengths and areas for improvement. Power BI and Tableau consistently emerge as leaders, particularly in customization, real-time data handling, and providing actionable insights. These tools are highly rated for data visualization and dashboard capabilities, making them suitable for organizations prioritizing flexibility and ease of use.

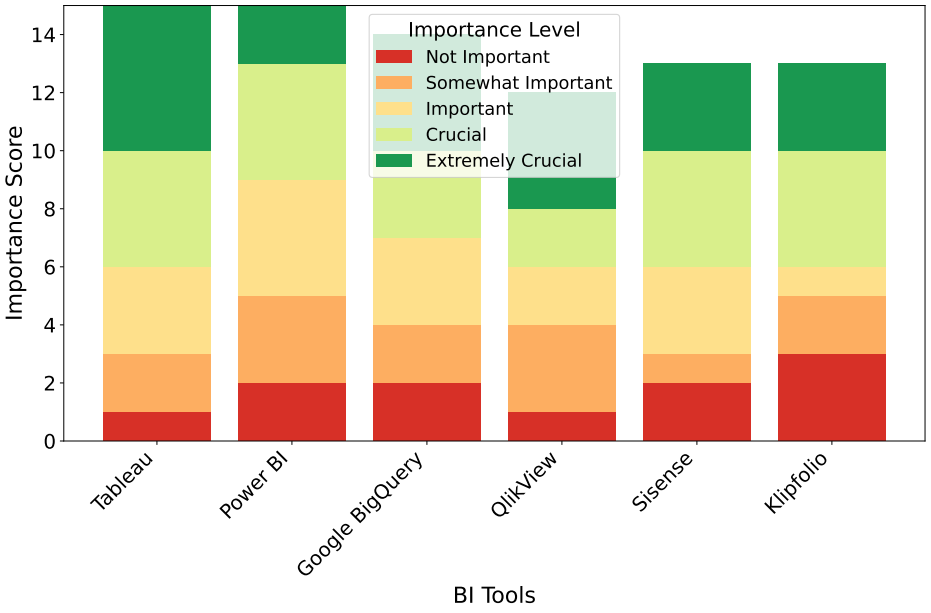
Google BigQuery and QlikView also demonstrate strong capabilities, especially in real-time data processing, though their effectiveness varies across different organizational contexts. Sisense is recognized for its robust data visualization and customization features, particularly in niche environments with specific integration requirements.

SAP BusinessObjects and Oracle BI received moderate feedback, indicating adequate but not outstanding performance, particularly in real-time data handling and customization. These tools may require enhancements to better meet the evolving demands of modern enterprises, especially regarding user satisfaction and adaptability.

The survey highlights the critical importance of security and data integration. While the tools generally provide secure data environments, there is a clear need for improved customization in access controls and data governance. This underscores the necessity for BI tools to evolve continuously, enhancing user-friendliness and ensuring seamless data integration across various platforms.



(a) Adaptability of BI Tools to Data Privacy and Compliance Regulations



(b) Importance of Data Sharing and Collaboration Capabilities Across BI Tools

Figure 3.8: (a) Stacked Bar Chart Showing Adaptability of BI Tools to Data Privacy and Compliance Regulations. (b) Stacked Bar Chart Showing the Importance of Data Sharing and Collaboration Capabilities Across BI Tools.

Overall, Power BI and Tableau are particularly well-regarded for their comprehensive capabilities in addressing diverse business intelligence needs. Organizations are encouraged to select tools that align with their specific operational requirements, considering factors such as scalability, ease of use, and security.

This chapter details the design and implementation of a comprehensive survey assessing the usage, preferences, and challenges associated with BI tools among industry professionals. Utilizing the GQM framework, the survey aligns with research objectives, providing measurable outcomes that validate the evaluation criteria established in earlier chapters. The survey covers dimensions such as data integration, visualization, reporting, analysis capabilities, user-friendliness, and performance. By focusing on real-world user feedback, the survey ensures that the evaluation framework is relevant and applicable to diverse business contexts. The inclusion of demographic data enhances the robustness of the evaluation, allowing for a tailored assessment of BI tools across different user groups. These insights provide a foundation for selecting and implementing BI tools that best fit organizational needs while addressing the critical challenges of data management, security, and user engagement.

Chapter 4

Security in Integrated BI Tools

Security in BI tools is a critical concern, as these tools are often used to process, analyze, and present sensitive proprietary data of businesses. With the growing reliance on BI tools for data-driven decision-making, the risks associated with unauthorized access, data breaches, and non-compliance with data protection regulations have also increased. This chapter explores the security challenges specific to BI tools, focusing on key issues such as user management, access control, and integration with identity and access management systems. By examining these challenges and proposing practical solutions, the chapter aims to provide a comprehensive understanding of how to enhance security within BI environments, ensuring that these tools can be used effectively and safely within organizations.

4.1 Background on Software Security

Software security is essential for protecting systems, applications, and data from unauthorized access, breaches, and other security threats. It involves implementing measures such as secure coding practices, authentication protocols, encryption, and access controls to protect information and maintain software systems' integrity. In the context of BI tools, security extends to protecting sensitive data and ensuring that access is restricted based on user roles and permissions. This section provides an overview of fundamental software security concepts, including common threats and vulnerabilities, and discusses the principles of secure software design. Understanding these foundational aspects of software security is crucial for developing robust BI tools that can withstand evolving security challenges and protect organizational data.

Before discussing security in BI tools, we briefly review key concepts from security literature, focusing on access control and user management, which are essential for securing BI tools and relevant to the next section. Encryption and other security topics are not covered here.

Access control is essential for protecting information, systems, and resources by ensuring that only authorized users can access specific data or functionalities. This involves several key components and mechanisms:

- **Authentication**: The process of verifying user identities, typically through passwords, biometrics (e.g., fingerprints, facial recognition), or other credentials, ensuring that only authorized users access BI tools.

- **Authorization**: Once authenticated, users are granted permissions based on their roles or responsibilities. This determines what actions they can perform within BI tools, such as viewing or editing data, reducing the risk of unauthorized access.

- **Single Sign-On (SSO) and SAML**: SSO allows users to access multiple BI applications with a single set of credentials, improving security and user convenience by minimizing the number of passwords. SAML further enables secure identity management and authentication across various services and domains.

- **Role-Based Access Control (RBAC) and Attribute-Based Access Control (ABAC)**: RBAC assigns permissions based on predefined roles (e.g., Marketing Manager), streamlining access management. ABAC provides more granular control by considering user attributes, environment, and context, allowing dynamic adjustments to access permissions.

- **Multi-factor Authentication (MFA)**: Enhances security by requiring multiple verification factors (e.g., a password and a code sent to a mobile device), adding an extra layer of protection against unauthorized access.

- **LDAP**: A protocol for managing directory information, LDAP supports centralized control over user identities, group memberships, and access permissions, facilitating uniform security management across BI tools.

- **Public Key Infrastructure (PKI)**: Uses digital certificates and encryption to secure communications and data transfers between users and BI tools, ensuring confidentiality and integrity over potentially insecure networks.

- **OAuth**: Allows secure authorization of third-party applications or services to access user data without sharing credentials, commonly used for secure API access in BI environments.

By integrating these access control mechanisms, organizations can ensure that software systems are secure, access is appropriately managed, and sensitive data is protected from unauthorized use, enhancing overall security and operational efficiency.

4.1.1 Security in Software Systems

Security in software systems involves implementing practices and technologies to protect applications, data, and systems from unauthorized access, misuse, and malicious activities. Key security measures include secure coding practices, encryption, access control, authentication, and continuous monitoring. These measures ensure that software functions correctly and can resist attacks.

Access control is a fundamental aspect of software security, ensuring that only authorized users can access and manipulate system resources. This includes enforcing security policies at various levels, from operating systems to database management systems, using models like

Role-Based Access Control (RBAC), Discretionary Access Control (DAC), and Mandatory Access Control (MAC). More advanced models, such as Attribute-Based Access Control (ABAC), provide granular and flexible control over access rights [27, 17].

Authentication is crucial for verifying user identities before granting access to resources. Techniques such as multi-factor authentication enhance security by requiring multiple forms of verification. Authorization follows authentication, determining what resources authenticated users can access based on predefined roles or attributes [16].

To mitigate risks like SQL injection, cross-site scripting, and buffer overflow attacks, secure coding practices and regular security assessments are essential. These practices help developers address vulnerabilities early in the development process, contributing to a resilient architecture that can defend against evolving threats [26].

4.1.2 Unique Security Challenges in BI Tools

BI tools present a unique set of security challenges that differ significantly from traditional software systems and DBMSs. One of the most critical concerns is the need for fine-grained, dynamic access control mechanisms. BI tools are utilized by multiple stakeholders, including data analysts, executives, and managers, all requiring varying levels of access to sensitive data. *The challenge lies in ensuring that each user has access only to the data relevant to their role while preventing unauthorized access to other parts of the system.*

Moreover, BI tools manage not only the underlying data but also metadata, such as report configurations, dashboard layouts, and data visualizations. This metadata can contain proprietary or confidential information that must be protected. Securing this metadata is just as important as securing the data itself, a challenge that is often overlooked in traditional software systems.

Another unique challenge in BI tools is the requirement for real-time access adjustments. As organizational roles and responsibilities evolve, traditional access control models often struggle to accommodate the dynamic nature of permissions required in BI environments. BI tools frequently operate across multiple platforms, which complicates security management. For example, users may need consistent permission levels across different BI tools (such as Tableau and Power BI), but each tool might handle user permissions differently, leading to inconsistencies and potential security gaps.

Addressing these unique challenges requires the development of more sophisticated access control models, such as dynamic role-based or attribute-based models, that are tailored specifically for BI environments.

4.1.3 Security in Database Management Systems

Security in Database Management Systems (DBMSs) is essential for protecting sensitive data and ensuring the integrity, availability, and confidentiality of database resources. Key security measures include access control, authentication, encryption, and auditing, all of which contribute to a robust and resilient database environment.

Access control mechanisms in DBMSs regulate who can access specific data and what operations they can perform. Common models include DAC, where users control their own

data, and RBAC, which grants permissions based on user roles within an organization. ABAC allows for granular and dynamic access decisions based on user attributes [27, 17].

DBMSs can also serve as the backbone for user and group management in access control for other software systems, including BI tools. By leveraging the robust access control mechanisms native to DBMSs, organizations can manage user identities and permissions centrally. This approach allows for consistent security policies across multiple applications, simplifying the management of complex user hierarchies and ensuring that access rights are uniformly enforced. For instance, using a DBMS to manage user groups and roles can streamline the integration with BI tools, where access to sensitive data must be carefully controlled to prevent unauthorized access.

Transaction processing and concurrency management are critical aspects of DBMS security. Techniques such as transaction isolation levels and concurrency control schemes ensure that transactions are processed reliably and securely, even in the face of potential attacks or system failures. For instance, hierarchical concurrency control algorithms use global locking tables based on semantic hierarchies to manage data access and preserve system integrity across distributed databases. These methods help serialize transactions and detect possible deadlocks, enhancing system resilience, particularly in mobile and distributed environments [19].

Additionally, security in DBMSs extends to protecting data during transaction processing. This involves managing concurrency and ensuring data consistency across mobile and heterogeneous environments, where challenges like restricted bandwidth and frequent disconnections can pose significant risks. Research in this area focuses on developing robust transaction recovery methods and advanced concurrency management schemes to ensure secure and reliable data access, even under challenging conditions [4].

The SQL authorization model introduces Authorization IDs, which are used to grant specific permissions such as SELECT, INSERT, DELETE, and UPDATE. These privileges are essential for executing SQL statements and are checked against the current authorization ID associated with a session or module. The integration of single sign-on (SSO) and attribute-based access control (ABAC) frameworks within DBMSs further enhances security by streamlining authentication and enabling fine-grained access control based on user attributes [5, 40].

4.1.4 Identity and Access Management (IAM)

Due to its critical role in ensuring the security of software systems, particularly in cloud-based environments, Identity and Access Management (IAM) warrants a dedicated discussion. IAM is fundamental to managing user identities and controlling access to resources, making it a cornerstone of modern cybersecurity strategies.

IAM involves managing user identities, defining and enforcing access control policies, and ensuring that only authorized users can access specific data, applications, or systems. It provides a framework for verifying user identities, determining permissible actions, and managing user accounts and access rights throughout their lifecycle.

In the context of cloud computing, IAM becomes even more crucial as it facilitates secure access to resources across distributed and dynamic environments. IAM solutions integrate authorization and authentication processes, often featuring single sign-on (SSO) capabilities to simplify access management and enhance security. By centralizing and automating user identity management, IAM reduces administrative burdens and ensures compliance with security

policies and regulations [29].

IAM is compatible with various cloud service models, including SaaS, PaaS, and IaaS, enabling secure, anytime access to cloud resources without direct provider involvement. To streamline access control, Role-Based Access Control (RBAC) is often employed, assigning permissions based on user roles, which simplifies administration in large-scale applications and reduces costs [3, 30].

Given the complexity of managing identities in cloud environments—where numerous users and services interact across different platforms—robust IAM is essential for maintaining data security and integrity. A comprehensive taxonomy of IAM features, including advanced authentication methods, access control policies, and encryption techniques, is necessary for evaluating and improving existing systems.

The importance of IAM cannot be overstated in addressing the growing challenges in cloud security. Solutions such as federated identity management allow for the secure exchange of identity information across trusted domains. Additionally, robust encryption and multi-factor authentication are critical in protecting sensitive identity data, ensuring its security both in transit and at rest [15].

4.1.5 Security Risks and Attacks in Business Intelligence Tools

BI tools, due to their central role in processing and visualizing sensitive business data, face a variety of security risks and potential attacks. One significant risk is *unauthorized access*, where attackers exploit weak access control mechanisms to gain access to sensitive data or reports. BI tools often rely on user roles to manage access, but these roles can become overprivileged or misconfigured, leading to data exposure.

Another risk specific to BI tools is *data leakage*. BI dashboards and reports often integrate data from multiple sources, and if one of these sources is compromised, it can lead to the exposure of sensitive information across the entire system. Furthermore, the metadata associated with dashboards—such as report configurations or access logs—can also be targeted in attacks, as this metadata can contain proprietary business information or insights that competitors might seek.

Additionally, BI tools can be vulnerable to *insider threats*, where employees or contractors with legitimate access misuse their privileges to extract, manipulate, or leak sensitive data. These threats are difficult to detect due to the legitimate nature of access, but can lead to significant data breaches.

Finally, BI tools are susceptible to *man-in-the-middle attacks* during data transmission. If communication between BI tools and data sources is not properly encrypted, attackers can intercept or manipulate data in transit, leading to both security breaches and the corruption of business intelligence outputs.

To mitigate these risks, BI tools must implement robust security measures, such as fine-grained access control, data encryption, regular security audits, and monitoring of user behavior. Additionally, integrating BI tools with comprehensive Identity and Access Management (IAM) systems can help prevent unauthorized access and reduce the impact of insider threats.

4.2 Security Challenges in BI Tools: Problem Statement

One major challenge in BI tools is the management of access control for a large and varied user base. In companies like Industry Collaboration, numerous users with different access levels interact with BI tools continuously. These access levels often change as roles and responsibilities evolve, requiring a system that can dynamically update permissions. Furthermore, the use of multiple BI tools within the same organization complicates security management. Ensuring secure data transfer and consistent access control across these tools is critical, especially when the same users need to interact with multiple systems.

Another significant issue is the need to secure not only the analyzed data but also the metadata related to dashboards and reporting elements. *This metadata can be proprietary, requiring restricted access to protect sensitive information.* The challenge is further compounded by the fact that existing BI tools provide only basic support for user and group management and offer limited integration with IAM systems. While IAM solutions can manage user identities and perform basic access control, they often fall short in handling the nuanced, dynamic access requirements specific to BI tools.

For instance, IAM systems may define who the users are and manage their access to the BI system, but they struggle to support scenarios where access to specific data sets or dashboard elements must be adjusted dynamically based on user roles or contexts. For example, a regional manager may need access to sales data specific to their region while being restricted from accessing data from other regions. Current IAM solutions often cannot dynamically adjust these permissions within BI tools, leading to either over-privileged access or insufficient access that hinders decision-making.

Moreover, when enterprises use multiple BI tools (e.g., Sisense and PowerBI in Industry Collaboration) to meet various analytical needs, the challenge of maintaining consistent user access policies across all platforms becomes more pronounced. Each BI tool may handle user access differently, and even if they integrate with the same IAM system, discrepancies in how access controls are interpreted and applied can lead to inconsistencies in user permissions and data access.

These security challenges highlight the need for innovative solutions that can bridge the gap between IAM systems and BI tools. *Our proposed solution addresses these concerns by integrating centralized access control with BI tools, allowing for more flexible and dynamic permission management.* The existing integration methods often lack the flexibility to support the detailed access control required in BI environments, especially when data, user roles, and access needs are constantly changing. Developing middleware that translates IAM policies into the fine-grained access controls required by BI tools or enhancing IAM systems to natively support these needs could provide a path forward.

By addressing these challenges, organizations can ensure that their BI environments are secure, compliant with regulations, and capable of providing the right data to the right people at the right time. This section has outlined the key issues related to security in BI tools, setting the stage for exploring potential solutions in the following sections.

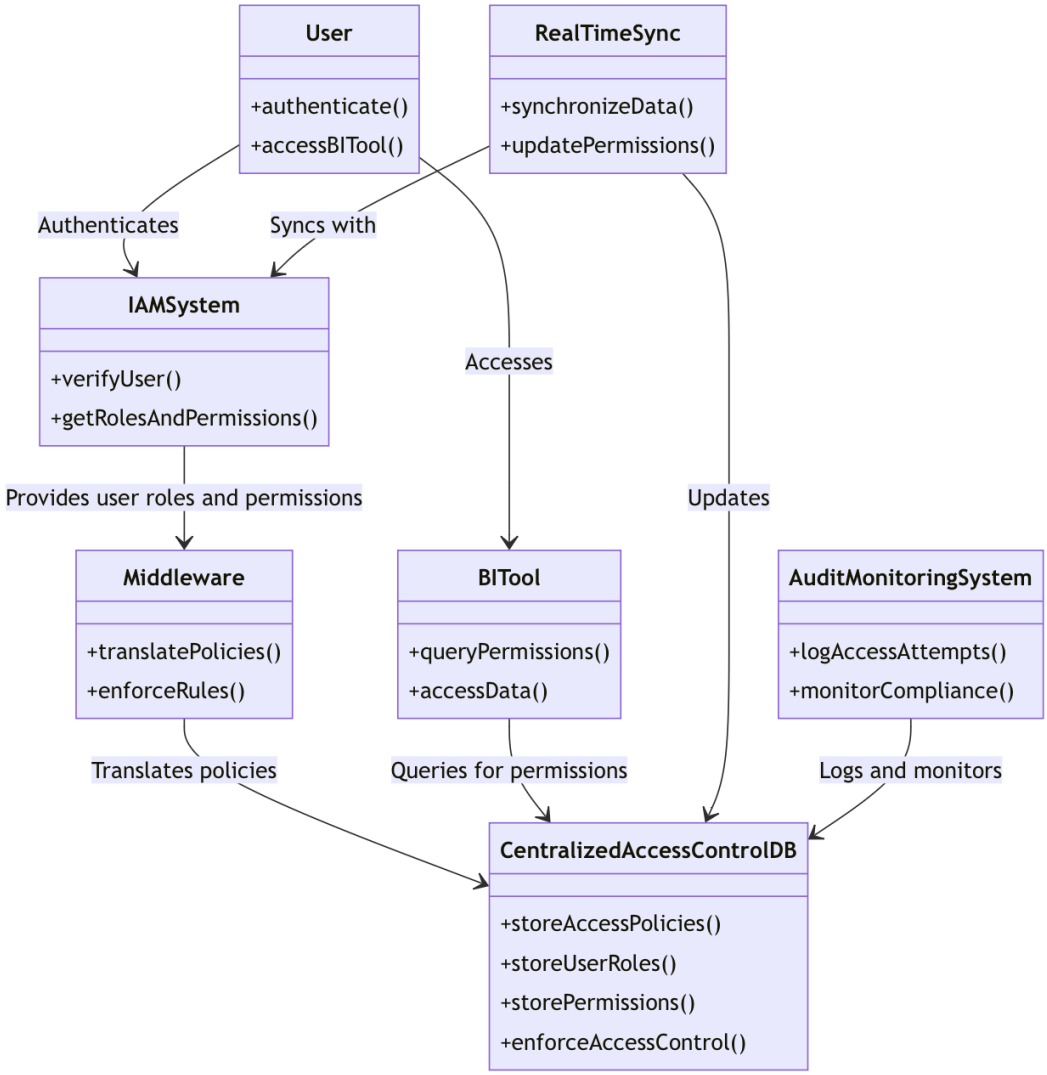


Figure 4.1: Architecture of the Database-Level Access Control Mediator Solution

4.3 Solution Prototype: Centralized Access Control Using DBMS Integration

The proposed solution addresses the security challenges in integrating IAM systems with BI tools by leveraging a DBMS as a central element for access control. This approach ensures consistent and scalable management of user access across multiple BI tools, facilitating seamless synchronization and enforcement of security policies.

At the heart of this solution is a centralized access control database, implemented using a DBMS. This database acts as the single source of truth for access control policies, user roles, and permissions. By centralizing access control management within a DBMS, the solution simplifies the process of updating and maintaining user access permissions as data, users, and roles evolve. This is particularly beneficial in dynamic environments, such as those in companies like Industry Collaboration, where client data changes frequently and user roles are constantly being updated.

The central DBMS serves as the core repository for managing access control across all BI tools within the organization. This centralized approach offers several advantages. First, it simplifies the management of access control by providing a unified interface for updating user roles and permissions, reducing the administrative burden associated with maintaining consistent access controls across multiple BI tools. Second, by using a central DBMS, different BI tools can sync their access control settings, ensuring that user permissions are applied consistently across all platforms. This is critical in environments where the same users interact with multiple BI tools, each handling different types of data. Third, the DBMS can dynamically adjust access control based on real-time changes in user roles and data, allowing for tailored access control that responds to specific contexts, such as restricting a regional manager to data relevant only to their region.

The middleware component of the solution acts as an intermediary between the IAM system and the central DBMS. It translates high-level IAM policies into the fine-grained access controls required by BI tools, ensuring that IAM systems manage user identities and roles, while the DBMS handles the detailed, context-specific access controls necessary within BI environments. The solution includes real-time synchronization capabilities to keep the central DBMS up-to-date with any changes in user roles or permissions as managed by the IAM system. Methods such as webhooks and event-driven updates are employed to achieve this synchronization, guaranteeing that the BI tools always enforce the most current access controls. Additionally, the central DBMS enables context-aware access control, allowing permissions to be adjusted based on the specific context in which the BI tool is being used.

Using a DBMS as the central element for access control provides several key benefits. The robust transaction management features of DBMSs ensure that access control operations are handled efficiently, even under heavy loads, guaranteeing performance and consistency in access control enforcement. DBMSs also offer strong crash recovery mechanisms, which are essential for maintaining the consistency of access control data in the event of a system failure, ensuring that user permissions are always accurate and up-to-date. Moreover, the fast query processing capabilities of DBMSs enable quick verification and enforcement of access control rules, minimizing latency in BI tools and ensuring a seamless user experience. The solution can also be crafted to manage not only traditional data types but also metadata related to dash-

boards, plots, and figures, allowing for comprehensive management of all data types handled by BI tools and ensuring that both data and metadata are securely accessed and managed.

The solution also incorporates audit and monitoring capabilities within the DBMS to track and log all access attempts and permissions checks. These logs provide visibility into the access control processes, enabling the auditing of policies and detection of unauthorized access attempts, which is crucial for maintaining the security and compliance of BI environments. By implementing this DBMS-based mediator solution, organizations can achieve seamless integration between IAM systems and BI tools, ensuring that user access is consistently controlled and aligned with security policies across all platforms. This solution not only enhances data security but also simplifies user access management, providing a flexible, dynamic, and centralized approach to access control. The architecture of the solution ensures that BI tools can efficiently manage the access control of diverse data types while benefiting from the DBMS's strengths in transaction management, crash recovery, and performance optimization.

While the prototype solution demonstrates the feasibility of integrating centralized access control within BI tools, it faces several challenges. *One of the primary concerns is scalability*. As organizations grow and the volume of data and the number of users increase, the system must be able to efficiently handle the load. Without adequate scaling mechanisms, performance may degrade as more users simultaneously access the BI tools. One possible way to address this challenge is through the implementation of distributed systems. By distributing the data and access control management across multiple nodes, the system can balance the load, improving both performance and scalability.

Another major issue is the potential for a *single point of failure*. In a centralized system, if the core access control mechanism fails, the entire system's functionality could be compromised. This risk can be mitigated by introducing redundancy and failover mechanisms or by employing distributed architectures. In a distributed setup, access control responsibilities can be divided among multiple servers or nodes, ensuring that if one component fails, others can take over without disrupting the service. These approaches will not only enhance the system's resilience but also ensure continuous, secure access to BI tools across the organization.

4.3.1 ABAC vs. RBAC

In this thesis, we primarily focus on Role-Based Access Control, where users are assigned specific roles, and each role has predefined access rights. This approach is simple and effective in managing user permissions based on their job responsibilities. On the other hand, Attribute-Based Access Control offers a more dynamic approach. In Attribute-Based Access Control, access decisions are based on multiple attributes, such as user location, time of access, or the sensitivity of the data, allowing for more granular control.

Although Attribute-Based Access Control is theoretically more flexible, it is not used in our current solution for several reasons. First, managing the various attributes required for Attribute-Based Access Control introduces significant complexity. Unlike Role-Based Access Control, which only requires assigning users to roles, Attribute-Based Access Control requires the system to evaluate different attributes for each access request. This added complexity would increase the system's computational load and reduce performance, particularly in environments with many users accessing data simultaneously.

Another challenge is that Attribute-Based Access Control requires real-time evaluation of

attributes, which could lead to performance bottlenecks. As the number of users, data sets, and reports increases, the system may struggle to keep up with the constant evaluation of access permissions based on dynamic attributes. Many Business Intelligence tools are not designed to handle such complex access control mechanisms, meaning extensive customization would be required to implement Attribute-Based Access Control, increasing both development time and costs.

Additionally, implementing Attribute-Based Access Control demands a more advanced identity and access management infrastructure. This would require significant changes to the existing system architecture, which goes beyond the scope of this project. The goal of our current solution is to provide an access control system that is simple, scalable, and efficient. Role-Based Access Control meets these needs without the added complexity and performance issues that would come with Attribute-Based Access Control. While Attribute-Based Access Control could be explored in future work, Role-Based Access Control is the more practical choice for this implementation.

4.4 New Solution Prototype: Proposed Middleware Architecture

To address the challenges of managing access control across multiple BI tools and dynamic user permissions, we propose a middleware-based solution. The architecture is illustrated in Figure **??**. The key components of this architecture include:

- **Users and User Groups:** Different user roles (e.g., data analysts, managers, data engineers) are grouped into roles with specific access rights.
- **IAM System:** Handles authentication and user group management. All user access requests go through this system to verify their credentials.
- **Raw Data Database:** The primary database where analytical data resides. Access to this data is controlled by the middleware.
- **BI Tools:** Various BI tools (e.g., Tableau, Power BI) that interface with the middleware to retrieve data based on user access.
- **Middleware:** Centralized component for access management. It communicates with the IAM system for authentication, checks permissions in the Access Control Database, and processes data queries from the BI tools.
- **Access Control Database:** Stores the access control policies and user group permissions. It dynamically updates when data or user roles change. It can be considered as part of the middleware.

The workflow for data access begins with a user request through a BI tool. The request is authenticated via the IAM system, and access is validated using the Access Control Database. Once validated, the middleware communicates with the Raw Data Database to retrieve the necessary data, which is then passed back to the BI tool.

While our prototype offers an innovative approach to integrating centralized access control within BI tools, it is not without its challenges. One major concern is *scalability*. As the number of users and data points grows, the system must be able to efficiently handle the increased load without compromising performance or responsiveness. To address this, the solution could be extended to leverage distributed systems, where access control and data processing are handled across multiple servers to prevent bottlenecks and ensure smoother operations even at scale.

Another critical issue is the *risk of a single point of failure*. Centralized access control systems often introduce vulnerabilities, where failure in one part of the system could potentially disrupt the entire access management process. To mitigate this, the solution could be enhanced with redundancy measures, such as failover mechanisms or by distributing access control responsibilities across different nodes in a distributed architecture. This would ensure that if one part of the system fails, another can take over without compromising security or system performance.

4.4.1 Discussion and Takeaways

This chapter has explored the security challenges inherent in integrating IAM systems with BI tools, particularly in environments where data is dynamic and user roles frequently change. The sensitive nature of the data handled by BI tools, combined with the need for consistent and secure access control, underscores the critical importance of a robust security framework.

The proposed solution, which leverages a centralized DBMS for managing access control, offers a promising approach to addressing these security challenges. By centralizing access control in a DBMS, the solution ensures that user permissions are consistently enforced across all BI tools, even as data and user roles evolve. The dynamic synchronization of access control policies between IAM systems and the central DBMS, along with the context-aware access control mechanisms, provides a flexible and adaptable security framework that can respond to the specific needs of different users and data contexts.

However, the implementation of this solution is not without its challenges. One significant issue is the need to modify existing BI tools to enable integration with the DBMS for user management. This may require changes to the codebase or the addition of new plugins, particularly in commercial BI systems that are not open source. While many BI tools, including some commercial ones, do support the addition of new plugins, achieving consistent implementation across different BI tools is crucial for the effectiveness of this solution. In our preliminary exploration, we have begun examining the architecture of open-source BI tools like Superset to assess the feasibility of implementing this integrated security solution. Nevertheless, ensuring uniform integration across diverse BI tools remains a challenge.

Another potential drawback is the central point of failure that comes with relying on a single DBMS for access control management. Although this risk can be mitigated through replication and other fault-tolerance techniques inherent to DBMSs, it still represents a vulnerability that must be carefully managed. Additionally, the performance of the DBMS under heavy load, particularly when managing a large number of real-time access control requests, could be a concern. Ensuring that the DBMS can handle such demands without introducing significant latency into the BI tools is critical to the overall success of the solution.

Moreover, while the centralized approach simplifies management and ensures consistency, it also requires rigorous security measures to protect the DBMS itself from unauthorized access

or breaches. The DBMS, being the central repository of all access control policies, becomes a highly sensitive component that needs robust protection, including encryption, access auditing, and continuous monitoring.

In conclusion, the proposed solution presents a robust framework for enhancing security in BI tools by leveraging a centralized DBMS for access control. While it offers significant advantages in terms of consistency, flexibility, and dynamic control, it also introduces new challenges that must be carefully addressed. These include the need for modifications to existing BI tools, the risk of a central point of failure, and the performance demands on the DBMS. The ongoing exploration of these challenges, particularly through the examination of open-source tools like Superset, will be crucial in refining and validating the proposed solution. The collaboration with Industry Collaboration continues to provide valuable insights, demonstrating the potential of this approach in real-world environments while also highlighting areas where further innovation and adaptation may be necessary.

Chapter 5

Conclusion and Future Work

In this chapter, we summarize the contributions of the thesis and discuss future research.

In Chapter 1, we introduced the fundamental role of BI tools in modern business environments, emphasizing their importance in data-driven decision-making and operational efficiency. We outlined the main problem statement and research objectives, focusing on the challenges organizations face in integrating these tools seamlessly into their existing systems. Chapter 2 detailed the development of an evaluative framework for BI tools. We explored foundational concepts in BI and conducted a comprehensive literature review to identify the key criteria for evaluating BI tools. This chapter established the systematic approach we used, highlighting the dimensions of data integration, visualization, analytics, and user-friendliness that guided our evaluation. In Chapter 3, we designed and executed a comprehensive survey to gather insights into BI tool usage patterns and preferences. The methodology, including design, distribution, and data collection processes, was thoroughly detailed. The results provided valuable insights into the real-world applicability and effectiveness of our evaluation criteria. Chapter 4 focused on the critical domain of security within integrated BI tools. We conducted an in-depth literature review on software security and IAM to understand existing challenges. Engaging with Instacart allowed us to validate the practical implications of these challenges and propose actionable solutions, including a DBMS-centric approach to access control. This solution addresses the dynamic nature of user roles and data, ensuring consistent security enforcement across multiple BI tools.

The contributions of this thesis are multi-faceted. First, the evaluative framework provides organizations with a structured approach to selecting BI tools that align with their operational and strategic needs. Second, our focus on security integration addresses the critical need for robust data protection mechanisms in BI tool deployments. Lastly, insights gained from real-world engagements, particularly with Instacart, underscore the practical applicability of our research findings.

Future research should focus on several key areas to further enhance the functionality and integration of BI tools in corporate systems. One important direction is the development of compliance frameworks within BI tools to address the increasing complexity of international data protection laws. Such frameworks could help organizations automatically track and report data usage, ensuring adherence to regulatory standards and reducing the cost and effort of regulatory audits.

Another critical area for future work is the development of industry-wide standards for BI

tools. Establishing these standards would promote interoperability between various platforms and systems, easing integration and reducing incompatibilities. Future efforts could involve leading or participating in initiatives aimed at creating and advancing these standards, which would facilitate the adoption of best practices across the industry.

Additionally, future research should explore the creation of more sophisticated integration frameworks that can seamlessly link BI tools with various enterprise applications and data sources. Developing middleware or APIs that dynamically translate security policies and user permissions across systems will be crucial for ensuring data flows comply with organizational access rules. Improving these integration capabilities can help organizations manage complex data ecosystems more efficiently, reducing latency and errors associated with manual integrations.

There is also significant potential for AI and machine learning to enhance and automate the security and usability of BI tools. Future studies should investigate how these technologies can be leveraged to anticipate user behavior, detect anomalies, and automatically adjust access rules to prevent unauthorized data access. AI could also improve user productivity and experience by optimizing data queries and visualizations based on user preferences and historical usage patterns.

Another important area for future work is improving the real-time data processing and analytics capabilities of BI tools. As enterprises increasingly rely on real-time data for decision-making, developing more efficient streaming data processing algorithms and integrating complex event processing engines will be essential to handle large volumes of real-time data effectively.

Finally, enhancing the user experience remains a priority for future research. This includes developing more intuitive and accessible BI tools, exploring adaptive user interfaces and personalized dashboard configurations, and integrating natural language processing to allow users to interact with BI tools using conversational language.

By addressing these areas, future research can significantly enhance the utility, security, and efficiency of BI tools, making them more adaptable to the needs of modern organizations and capable of handling the growing complexities of data-driven environments.

Bibliography

[1] Apache. What is apache superset? https://superset.apache.org/docs/intro/, Jan 2023.

[2] BARC. Overview of the global enterprise bi & analytics platforms market in 2023. *BARC*, 2023. Link to Source.

[3] Elisa Bertino. Rbac models—concepts and trends. *Computers & Security*, 22(6):511–514, 2003.

[4] Elisa Bertino and Ravi Sandhu. Database security-concepts, approaches, and challenges. *IEEE Transactions on Dependable and secure computing*, 2(1):2–19, 2005.

[5] Jan De Clercq. Single sign-on architectures. In *International Conference on Infrastructure Security*, pages 40–58. Springer, 2002.

[6] Mariayee Doraisamy, Suhaimi bin Ibrahim, and Mohd Naz'ri Mahrin. Formulation of metric based software project performance monitoring model: A roadmap. In *2014 IEEE Conference on Open Systems (ICOS)*, pages 48–53. IEEE, 2014.

[7] DundasBI. overview. https://www.dundas.com/support/videos/dundas-bi-overview, July 2023.

[8] Forrester. The forrester wavetm: Enterprise bi platforms with majority on-premises deployments, q3 2017. *Forrester*, 2017. Link to Source.

[9] Forrester. The forrester wave methodology. *Forrester*, 2024. Link to Source.

[10] Gartener. *Magic Quadrant for Analytics and Business Intelligence Platforms*. Gartener, 2021.

[11] Gartner. Magic quadrant for analytics and business intelligence platforms. *garterner*, 2021. Link to Source.

[12] Gartner. Magic quadrant research methodology. *Gartner*, 2024. Link to Source.

[13] Google. Google trends comparison. https://shorturl.at/D3Mud, July 2023,.

[14] Google. Looker. https://cloud.google.com/looker/docs/what-is-lookml, JULY 2023.

[15] Umme Habiba, Rahat Masood, Muhammad Awais Shibli, and Muaz A Niazi. Cloud identity management security issues & solutions: a taxonomy. *Complex Adaptive Systems Modeling*, 2:1–37, 2014.

[16] Dick Hardt. The oauth 2.0 authorization framework. Technical report, Hardt, 2012.

[17] Vincent C Hu, David Ferraiolo, D Richard Kuhn, et al. *Assessment of access control systems*. US Department of Commerce, National Institute of Standards and Technology . . . , 2006.

[18] IBM. Benefits. https://www.ibm.com/products/cognos-analytics/, July 2023.

[19] James B Lim and Ali R Hurson. Transaction processing in mobile, heterogeneous database systems. *IEEE Transactions on Knowledge and Data Engineering*, 14(6):1330–1346, 2002.

[20] Ashwini Dubbewar Meet Joshi. Review on business intelligence, its tools and techniques, and advantages and disadvantages. *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 10*, 2021.

[21] Microsoft. Ppower bi customer stories. https://powerbi.microsoft.com/en-ca/customer-showcase/, July 2023.

[22] Microstrategy. Accessing multiple data sources in a project. https://shorturl.at/BdVCo, JULY 2023.

[23] pentaho. pentaho. https://www.hitachivantara.com/en-us/products/pentaho-platform/data-integration-analytics.html, July 2023.

[24] Dan Potter. Gartner's magic quadrant or data integration tools is latest signal that qlik data integration continues to set the market standard. https://shorturl.at/ksJ3v, August 2021.

[25] Redherring. 2017 red herring north america: Top 100 winners. https://www.redherring.com/2017-rhna-winners/, jan 2017.

[26] G Dumindu Samaraweera and J Morris Chang. Security and privacy implications on database systems in big data era: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 33(1):239–258, 2019.

[27] Ravi S Sandhu. Role-based access control. In *Advances in computers*, volume 46, pages 237–286. Elsevier, 1998.

[28] R. Sarda, D. Delen, and E. Turban. Business intelligence, analytics and data science: A managerial perspective. *Journal of Software Engineering and Applications*, 2018. Link to Source.

[29] Anuja Sharma, Sarita Sharma, and Meenu Dave. Identity and access management- a comprehensive study. In *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*, pages 1481–1485, 2015.

[30] Anuja Sharma, Sarita Sharma, and Meenu Dave. Identity and access management-a comprehensive study. In *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*, pages 1481–1485. IEEE, 2015.

[31] Sisense. Our investors. https://www.sisense.com/about/investors/, August 2023.

[32] Tableau. Business intelligence: A complete overview. https://www.tableau.com/learn/articles/business-intelligence, July 2023.

[33] Tableau. Our customers. https://www.tableau.com/why-tableau/our-customers, March 2023.

[34] Arun Ulagaratchagan. Microsoft. https://shorturl.at/eJoPG, April February 12, 2020.

[35] Rini Van Solingen, Vic Basili, Gianluigi Caldiera, and H Dieter Rombach. Goal question metric (gqm) approach. *Encyclopedia of software engineering*, 2002.

[36] Rini van Solingen and Egon Berghout. The goal/question/metric method:. In *Proceedings of the [Conference Title]*, 2000.

[37] Rini Van Solingen, Egon Berghout, and Erik Kooiman. Assessing feedback of measurement data: relating schlumberger rps practice to learning theory. In *Proceedings Fourth International Software Metrics Symposium*, pages 152–164. IEEE, 1997.

[38] Rob Watts. Forbes. https://www.forbes.com/advisor/business/software/best-data-visualization-tools/#tableau_section, April Jun 25, 2023,.

[39] Steve Williams and Nancy Williams. *The Profit Impact of Business Intelligence*. Elsevier, 2010. Link to Source.

[40] Eric Yuan and Jin Tong. Attributed based access control (abac) for web services. In *IEEE International Conference on Web Services (ICWS'05)*. IEEE, 2005.