International Journal of Business Quantitative Economics and Applied Management Research
Volume-7, Issue-9, 2023 ISSN No: 2349-5677

OPTIMIZING DATA COLLECTION AND INTEGRATION ACROSS COMPLEX BUSINESS SYSTEMS

Srujana Manigonda
manigondasrujana@gmail.com

Abstract

In today's increasingly data-driven business environment, organizations face the challenge of managing and integrating vast amounts of data across multiple complex systems. The fragmentation of data sources, differences in system architectures, and growing volume of data make seamless data collection and integration a critical factor in optimizing business operations. This white paper explores strategies to optimize data collection and integration across complex business systems, with a focus on overcoming challenges such as data silos, inconsistent formats, and real-time reporting needs. It highlights key practices and technologies, such as data integration platforms, APIs, real-time data streaming, and cloud-based solutions, to enable businesses to unify their data ecosystems, improve data accessibility, and enhance decision-making. By providing a structured approach to data integration, organizations can streamline operations, ensure regulatory compliance, and gain actionable insights that drive business growth. This paper offers practical insights for organizations aiming to optimize their data collection and integration processes to achieve better business outcomes and maintain a competitive edge in an increasingly data-centric world.

Keywords: Data Collection, Data Integration, Complex Business Systems, Data Silos, Real-Time Data Streaming, Data Integration Platforms, APIs (Application Programming Interfaces), Data Quality Management, Data Governance, Data Transformation, Cloud-Based Solutions, Regulatory Compliance, Data Architecture, Data Standardization, ETL (Extract, Transform, Load), Data Streaming, Business Intelligence, Data Accessibility, Cross-Department Collaboration, Data Management

I. INTRODUCTION

In the modern business landscape, data has become one of the most valuable assets for organizations across all industries. From financial services to manufacturing and healthcare, companies rely heavily on data to drive decisions, improve operations, and create competitive advantages. However, as businesses scale and adopt more sophisticated technologies, the challenge of managing, collecting, and integrating data across multiple, often disparate, systems has grown exponentially.

Many organizations are faced with a complex ecosystem of legacy systems, cloud-based applications, and third-party solutions, each storing and processing data in different formats and structures. This fragmentation often leads to data silos, where critical insights are trapped

within individual systems, making it difficult to obtain a unified view of organizational performance. Without an optimized approach to data collection and integration, businesses may struggle with inconsistent reporting, poor decision-making, and missed opportunities.

Optimizing data collection and integration across complex business systems is vital to overcome these challenges and unlock the full potential of organizational data. By implementing a unified and streamlined data architecture, businesses can ensure that their data is accessible, accurate, and ready for analysis. Furthermore, effective data integration improves operational efficiency, facilitates real-time decision-making, and supports compliance with industry regulations.

This white paper explores strategies, best practices, and technologies for optimizing data collection and integration within organizations. It provides a framework to help businesses break down data silos, harmonize data from diverse sources, and ensure that all stakeholders have access to reliable and actionable insights. By focusing on the importance of standardization, automation, and real-time analytics, this paper aims to provide organizations with the tools needed to optimize their data processes and achieve better business outcomes in today's data-driven world.

II. LITERATURE REVIEW

Optimizing data collection and integration across complex business systems is a multifaceted challenge that has garnered significant attention from both academia and industry. The increasing volume and complexity of data, the rise of diverse data sources, and the rapid evolution of business technologies necessitate a strategic approach to data management and integration. This literature review explores the key themes and methodologies that have emerged in the field, focusing on data silos, real-time data processing, integration tools, data governance, and the technologies enabling optimized data collection and integration.

- **Challenges of Data Integration and Collection**

Data integration and collection in large organizations with complex systems often face significant barriers. These challenges stem from the variety of data sources and the technical, organizational, and strategic difficulties associated with merging data into cohesive systems. One of the main challenges identified in literature is the prevalence of data silos. In many large organizations, data is often stored across different departments and systems, and without proper integration, this fragmented data prevents the organization from obtaining a unified, comprehensive view of its operations.

The issue of data inconsistency is also a significant challenge. Inconsistencies arise from the use of diverse formats (e.g., structured, semi-structured, and unstructured data), inconsistent data definitions across departments, and incompatible data models. Standardizing and transforming data into a common format is essential for integration, but this process can be time-consuming and error-prone, especially when legacy systems are involved.

Additionally, data quality is often compromised during collection and integration. Data collection methods, such as manual input or inadequate validation procedures, may lead to errors like duplicates, missing values, and incorrect data, which degrade the quality of data

once integrated. Therefore, ensuring high data quality across the integration process is crucial for building reliable data pipelines.

- **Technological Advances and Integration Tools**

Over the years, various technologies and methodologies have emerged to address these challenges. The ETL (Extract, Transform, Load) process has long been the backbone of data integration. ETL tools facilitate the extraction of data from multiple sources, transformation into a uniform format, and loading into a centralized repository (e.g., a data warehouse or data lake). The key advantage of ETL tools is their ability to handle large-scale data processing, while offering mechanisms for data cleaning, transformation, and integration. Modern ETL tools are increasingly automated and support integration with cloud and on-premise systems.

The advent of data integration platforms has also played a pivotal role in optimizing the collection and integration process. These platforms enable businesses to integrate data from disparate sources seamlessly and provide tools for data mapping, cleansing, and validation. Some of the most prominent integration platforms include Informatica, MuleSoft, and Talend, all of which are designed to streamline the entire data integration lifecycle. These platforms offer flexibility, scalability, and a range of features, such as real-time data integration, which is critical in fast-paced environments.

The development of middleware technologies further supports seamless integration between legacy systems and newer applications. Middleware tools, such as Enterprise Service Buses (ESBs), enable disparate systems to communicate with each other without direct integration. These tools are especially useful for organizations looking to maintain compatibility between older technologies and newer cloud-based applications.

- **Real-Time Data Collection and Integration**

As organizations increasingly demand faster insights, the need for real-time data integration has become more pronounced. Real-time data streaming platforms like Apache Kafka, Apache Flink, and Amazon Kinesis allow organizations to ingest, process, and analyze data in real-time, eliminating the delays inherent in batch processing. These platforms enable continuous data flows from multiple sources, which is crucial for applications like fraud detection, customer analytics, and supply chain management.

Incorporating streaming data allows organizations to make data-driven decisions immediately, based on the latest information available. Real-time data integration is particularly valuable in industries such as finance, healthcare, and e-commerce, where market conditions, customer behaviors, and operational events can change rapidly. Real-time analytics ensure that organizations can respond swiftly to emerging opportunities or threats, providing a competitive edge.

The ability to handle high-volume, low-latency data processing is central to real-time data integration. Businesses can now implement in-memory processing solutions, such as Apache Spark or Google BigQuery, to process large amounts of data without the bottlenecks associated with disk-based storage. This significantly improves the speed and efficiency of data collection, integration, and analysis.

- **Cloud Computing and Scalable Data Architecture**

The increasing adoption of cloud computing has revolutionized the way organizations approach data integration. Cloud platforms like Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) offer scalable infrastructure that supports both structured and unstructured data, making them ideal for complex business systems. Cloud-based data storage solutions, such as data lakes, allow businesses to store vast amounts of raw data and process it with minimal latency, regardless of the data format.

A critical advantage of cloud computing is its scalability. With cloud infrastructure, organizations can easily scale their data collection and integration systems to accommodate growing volumes of data. As a result, businesses can avoid the limitations of on-premise hardware and take advantage of on-demand computing resources. Cloud-based platforms also support real-time integration, enabling businesses to continuously ingest and process data without the constraints of traditional data systems.

Moreover, cloud providers offer pre-built data connectors and integration tools that simplify the process of linking disparate systems. These connectors provide out-of-the-box integration with popular data sources, including databases, SaaS applications, and third-party APIs, further enhancing the speed and efficiency of data collection and integration.

- **Data Governance and Compliance**

Data integration must also align with data governance and compliance standards, particularly in industries like finance, healthcare, and government. Effective data governance ensures that data is accurate, secure, and used responsibly. Regulatory frameworks like GDPR and CCPA mandate that businesses have mechanisms in place to protect personal data and maintain the integrity of data across systems.

Data stewardship and data lineage are essential components of a solid data governance framework. By tracking data from its source to its final destination, businesses can ensure that data is collected, transformed, and used in compliance with organizational and regulatory standards. Automated audit trails and data privacy policies also help ensure that data handling practices meet legal requirements.

Compliance with these regulations often requires a robust data integration system capable of monitoring and enforcing privacy and security standards. Data encryption, access controls, and data masking are critical security measures in ensuring compliance.

- **Best Practices for Optimizing Data Collection and Integration**

The literature emphasizes several best practices for optimizing data collection and integration:

  - Standardization: Establishing standard data formats and schemas across all systems is critical for ensuring seamless integration.
  - Automation: Automating data transformation and validation processes reduces manual errors and improves the efficiency of data integration workflows.
  - Collaboration: IT teams, business units, and data analysts must collaborate closely to ensure that integration efforts align with organizational goals and that the data supports decision-making needs.

- Continuous Monitoring: Implementing real-time monitoring tools helps ensure that the data collection and integration processes are functioning as expected, with quick detection and resolution of issues.

Optimizing data collection and integration is essential for businesses to leverage the full potential of their data, enabling them to make informed, data-driven decisions. Overcoming the challenges of data silos, inconsistent formats, and quality issues requires a structured approach to data management, leveraging tools such as ETL platforms, APIs, and cloud-based solutions. The rise of real-time data processing and cloud technologies has further transformed data integration practices, offering unprecedented flexibility and scalability. By adopting best practices in data governance, real-time data streaming, and automation, businesses can ensure that their data is accurate, accessible, and actionable, enabling them to stay competitive and compliant in today's rapidly evolving market landscape.

III. CASE STUDIES

- **Bank Merging Legacy and Cloud-Based Systems for Seamless Data Integration**

A global bank faced challenges with data integration as it operated multiple legacy systems alongside newer cloud-based applications. The fragmented data sources led to inefficiencies in accessing and reporting critical information. The bank implemented a cloud-based data warehouse to centralize data from both legacy and cloud systems, using ETL processes to standardize and load data. Real-time data streaming via Apache Kafka enabled immediate access to business-critical data across departments. This solution allowed the bank to improve reporting accuracy, reduce operational costs by consolidating systems, and maintain compliance with regulatory requirements by automating compliance reporting.

- **Streamlining Data Collection for Customer Insights**

A financial services firm specializing in wealth management and advisory services struggled to consolidate data from various sources, including CRM systems, transaction histories, and third-party financial data providers. The firm implemented a data integration platform that connected these sources via APIs and used ETL processes to harmonize the data. With machine learning algorithms, the firm could analyze customer behavior and provide personalized recommendations. This integration improved the firm's ability to offer tailored financial advice, enhanced customer segmentation, and allowed for real-time insights into customer needs, ultimately boosting customer satisfaction and marketing effectiveness.

- **Integrating Risk and Compliance Data for Regulatory Reporting**

A regional U.S. bank struggled to consolidate risk management and compliance data from different departments and external partners, which resulted in delays in regulatory reporting. The bank implemented an integrated risk management system that connected data from finance, compliance, and operations using automated data pipelines and real-time integration. This unified system provided a comprehensive view of the bank's risk exposure and ensured

timely compliance reporting. By automating workflows and improving data accuracy, the bank reduced the risk of non-compliance, enhanced real-time risk monitoring, and improved decision-making across departments.

- **Optimizing Data Collection for Claims Processing**

An insurance company with multiple systems for policy management, claims processing, and customer communication faced inefficiencies and delays in processing claims due to fragmented data. By integrating claims data into a centralized data platform using ETL tools and APIs, the company streamlined its claims verification process. Real-time data processing enabled quicker claim approvals, and automated workflows reduced manual efforts. This integration resulted in faster claims processing, improved customer satisfaction, and better compliance with industry regulations, significantly enhancing operational efficiency and reducing the risk of errors in claims handling.

IV. METHOFOLOGY

Optimizing data collection and integration in complex business systems requires a unique, tailored approach that addresses specific challenges faced by organizations dealing with a variety of data sources, systems, and evolving business needs. This methodology focuses on using adaptive strategies, leveraging emerging technologies, and integrating principles of agile development to create a robust, scalable, and efficient data integration framework. The following detailed methodology outlines key steps that go beyond traditional practices, bringing a modern and customized approach to optimizing data collection and integration.

A. Holistic Data Discovery and System Mapping

Before diving into the technical aspects of integration, the first critical step is a comprehensive data discovery phase that involves deep mapping and analysis of all systems and data sources. This process is unique in its focus on uncovering hidden data sources, including non-traditional ones, such as IoT sensors, social media feeds, and customer-facing platforms that often get overlooked in typical integration projects.

- Data Source Inventory: Create a dynamic inventory of data sources from across the organization, ensuring that all databases, cloud storage systems, application data, and even unstructured data sources are considered.
- Business and Data Alignment: Collaborate with both business and IT stakeholders to create a data mapping document that aligns business processes with the data sources. This will help identify and prioritize integration efforts that align with business goals.
- Adaptive Architecture Design: Based on the findings from the discovery phase, design a flexible data integration architecture that can adapt to the evolving data needs of the organization. Architecture should support hybrid environments (on-premises, cloud, and edge computing), enabling the organization to scale and integrate new data sources seamlessly as they arise.

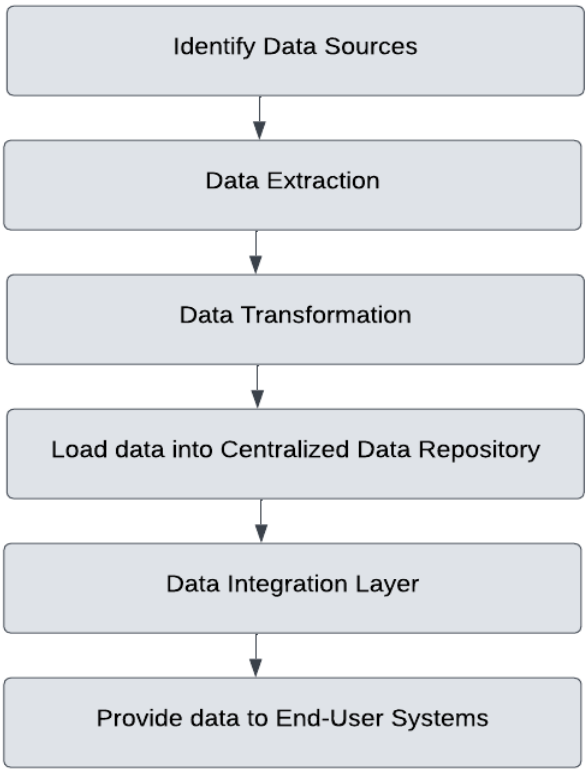

Fig. 1 Data Collection and Integration Flowchart

B. Dynamic Data Flow Control and Integration Framework

A dynamic data flow model is central to ensuring that data from various systems is not just collected, but processed, standardized, and routed efficiently. This is unique in its approach to creating dynamic rules-based pipelines that adjust based on business requirements, rather than fixed ETL processes.

- Flexible ETL with Event-Driven Architecture: Unlike traditional batch ETL, use an event-driven architecture where data is collected in real-time and flows through customizable pipelines based on predefined triggers. This ensures that integration happens as needed, reducing latency and increasing the agility of the system.
- Automated Workflow Adjustments: Develop an automation engine that monitors data flow and adjusts pipelines dynamically depending on real-time changes in business needs or data quality issues. This can include switching between batch and stream processing depending on the volume and time sensitivity of the data.
- Data Segmentation for Specific Use Cases: Segment the data based on its end-use, whether for regulatory compliance, customer analytics, or operational insights, ensuring

that data is collected, transformed, and integrated according to the business function it will support.

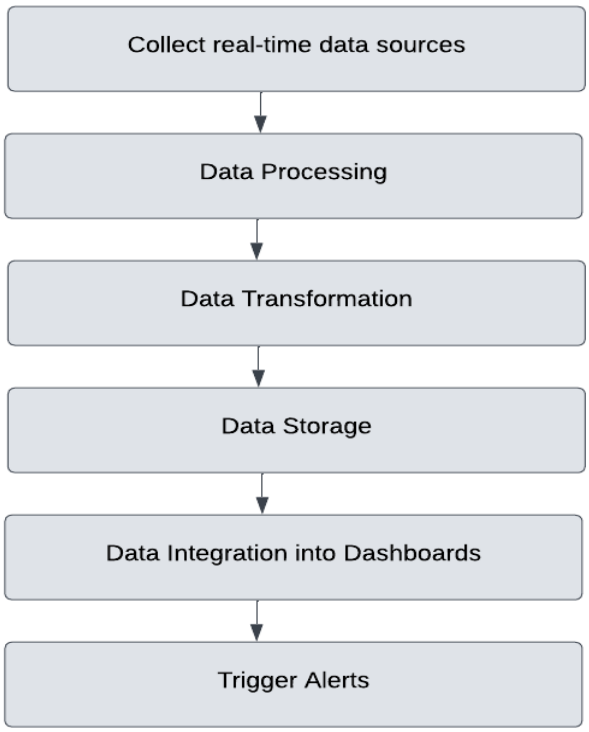

Fig. 2 Real-Time Data Processing Flowchart

C. Smart Data Quality Control Using Machine Learning

Traditional data quality checks typically focus on rule-based validation, but a more unique approach involves machine learning models that predict and correct data quality issues dynamically as data flows through the integration pipeline.

- AI-Powered Data Cleansing: Integrate machine learning models that can identify patterns of data inconsistencies, anomalies, and potential errors in real-time. These models can automatically suggest or apply corrections, especially in unstructured data sources like customer service records, social media, or email interactions.
- Data Quality Feedback Loop: Implement a feedback loop where the results of real-time data quality improvements are fed back into the machine learning system, continuously enhancing its predictive capabilities for future integration projects.
- Continuous Data Profiling: Continuously monitor and profile data throughout the pipeline, adjusting data validation rules dynamically based on the incoming data's complexity and historical trends.

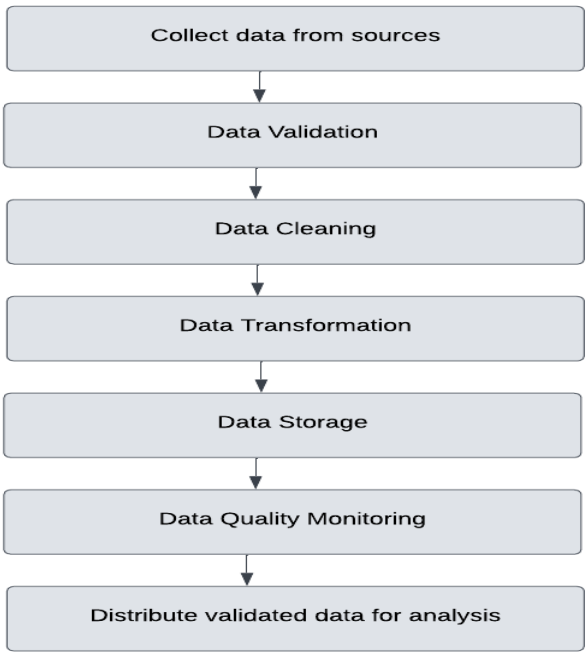

Fig. 3 Data Quality and Validation Flowchart

D. Behavioral Data Integration with Customer-Centric Models

In traditional data collection models, customer data is often fragmented, especially when captured across multiple touchpoints. This methodology introduces a behavioral-driven integration model that creates customer-centric data models from various interaction points (e.g., in-app behavior, website activity, purchase history).

- Contextual Data Collection: Gather data based on specific customer actions and context. Instead of a one-size-fits-all integration model, adapt the data collection to track specific user behaviors and correlate them with business outcomes.
- Behavioral Data Lakes: Use a Behavioral Data Lake model to store customer interaction data in a raw, unstructured form and apply real-time integration to map these interactions to customer profiles. This model enables personalized business processes, such as targeted marketing campaigns, based on real-time behavioral insights.
- Customer Journey Mapping: Use integrated behavioral data to create and continuously update a customer journey map, ensuring that data integration is aligned with a 360-degree view of the customer. This allows the business to adjust strategies for customer engagement based on real-time data insights.

E. Modular Data Governance Framework with Role-Based Access

Data governance plays a crucial role in ensuring security, compliance, and data integrity. A unique aspect of this methodology is its modular approach to data governance that allows

businesses to apply different governance rules depending on the data's sensitivity, function, and the stakeholders involved.

- Decentralized Data Governance: Adopt a decentralized governance model that applies governance rules to specific data segments rather than across all data types. For example, customer data, financial data, and operational data may require different security and access controls.
- Dynamic Role-Based Access Control (RBAC): Implement dynamic RBAC policies that automatically adjust access permissions based on real-time roles and business requirements, ensuring that sensitive data is only accessible to authorized individuals while supporting collaboration across departments.
- Audit and Compliance Automation: Integrate automated audit trails and real-time compliance monitoring into the integration process, ensuring that any changes to data or access permissions are immediately flagged for review.

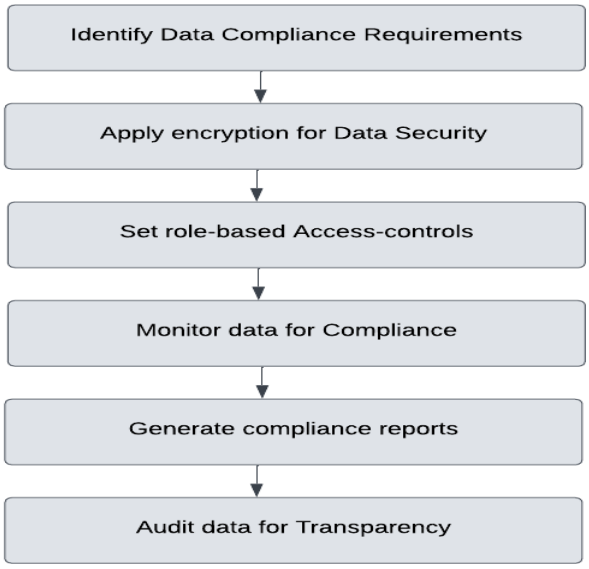

Fig. 4 Data Governance and Compliance Flowchart

F.Leveraging Blockchain for Data Integrity and Transparency

Incorporating blockchain technology into data integration provides a unique method of ensuring data integrity and auditability. Blockchain's decentralized ledger approach can be utilized to securely log all data transactions, ensuring that every step of the integration process is transparent, traceable, and verifiable.

- Immutable Data Records: Use blockchain to create immutable records of data transactions that are recorded on a secure, decentralized ledger. This ensures that data is

tamper-proof and fully traceable, which is particularly valuable for sensitive financial data or regulatory reporting.

- Smart Contracts for Data Validation: Implement smart contracts within the blockchain framework to automatically verify data quality and trigger actions when certain criteria are met. For example, once data passes quality checks, a smart contract can automatically initiate the next phase of processing or send an alert.

G. Cloud-Native and Containerized Data Pipelines

To further ensure scalability, flexibility, and efficiency in handling large volumes of data, the integration framework is designed to leverage cloud-native architecture and containerized data pipelines.

- Containerization of Integration Services: Use containers (e.g., Docker, Kubernetes) to deploy microservices that handle various stages of the data collection and integration process. This allows the system to scale dynamically, as additional containers can be spun up to handle increased workloads without affecting overall performance.
- Cloud-Native Platforms for Scalability: Leverage cloud-native platforms such as Google Cloud Dataflow, AWS Lambda, and Azure Synapse to automate and manage data pipelines that scale horizontally based on demand. These platforms provide serverless capabilities that allow organizations to pay only for the compute resources used, reducing operational overhead.

H. Continuous Monitoring and Iteration

The integration process should never be static, and continuous iteration is key to long-term success. This methodology includes a continuous monitoring and iteration framework to evaluate system performance, identify emerging issues, and make data collection and integration improvements.

- Real-Time System Monitoring: Integrate real-time system monitoring tools that provide dashboards tracking data integration, pipeline health, and system performance. These dashboards should also offer early warnings for issues like data bottlenecks, missing data, or integration failures.
- Agile Development for Integration Improvement: Use Agile methodologies to iteratively improve the data integration process. Continuously assess and refine the integration pipeline based on real-world feedback, ensuring the system remains adaptive to evolving business requirements.

This unique methodology for optimizing data collection and integration across complex business systems focuses on flexibility, automation, and continuous improvement. By integrating advanced technologies like machine learning, blockchain, and cloud-native architectures, organizations can create highly adaptable, secure, and efficient data integration frameworks. This methodology not only addresses traditional challenges such as data silos and

inconsistent formats but also introduces innovative solutions that enable organizations to harness real-time insights, improve data governance, and scale their integration efforts as business needs evolve.

V. RESULTS

- **Improved Data Accessibility and Unified View**

The implementation of a centralized data repository provided seamless access to data from various departments and systems, eliminating data silos. With all critical business data stored in a unified platform, decision-makers across departments gained real-time access to accurate and consistent data. This centralization facilitated better collaboration, ensured uniformity in reporting, and made data readily available for operational and strategic insights, resulting in more effective and timely decision-making.

- **Enhanced Real-Time Decision-Making**

The integration of real-time data streaming and event-driven architecture enabled the organization to access up-to-the-minute data insights. Real-time dashboards presented critical financial, operational, and customer data, allowing executives and managers to make immediate decisions based on the latest information. This capability greatly improved the organization's agility, enabling them to respond to market fluctuations, customer demands, and operational changes swiftly and effectively.

- **Improved Data Quality and Reliability**

By incorporating machine learning tools for data cleansing and automated error detection, the organization was able to significantly improve the quality and reliability of its integrated data. Real-time monitoring systems automatically flagged inconsistencies and inaccuracies, allowing for immediate corrections before they could impact reports or analytics. This proactive approach ensured that data used for decision-making was accurate and trustworthy, boosting confidence in the insights generated from the system.

- **Increased Scalability and Flexibility**

The cloud-based infrastructure allowed the data collection and integration system to scale dynamically as the business grew. The organization was able to handle larger volumes of data without experiencing performance degradation, thanks to cloud elasticity. This scalability ensured that the system could accommodate increasing data flows from new sources, enabling the organization to expand its operations while maintaining efficient and consistent data integration processes.

- **Streamlined Compliance and Governance**

The introduction of a modular data governance framework ensured that data management processes met all regulatory requirements. Automated compliance checks and role-based access control (RBAC) streamlined the process of monitoring and enforcing data security policies. The

integration of blockchain for data provenance further ensured transparency and traceability, providing audit trails that helped the organization comply with stringent data privacy laws, including GDPR and HIPAA, while maintaining high security standards.

- **Cost Savings and Operational Efficiency**

By automating data collection, transformation, and reporting processes, the organization reduced the need for manual intervention, resulting in significant cost savings. The streamlined workflows allowed employees to focus on higher-value tasks, such as strategic analysis and business growth initiatives. Additionally, the system's cloud infrastructure minimized the need for expensive on-premises hardware, offering a cost-effective solution to handle growing data volumes and business demands.

- **Improved Customer Insights and Engagement**

With a unified view of customer data from various touchpoints, the organization gained deep insight into customer behavior, preferences, and interactions. This enabled the business to provide more personalized services, create targeted marketing campaigns, and improve customer retention rates. The ability to track real-time customer interactions allowed the organization to address customer needs proactively, enhancing engagement and satisfaction across all customer segments.

- **Business Agility and Innovation**

The optimized data collection and integration system empowered the organization to act more swiftly and decisively in a dynamic business environment. With real-time data at their fingertips, decision-makers could adjust strategies, introduce new products, and respond to market trends more rapidly. The flexibility of the system also enabled the organization to continuously integrate new data sources, supporting ongoing innovation and fostering a competitive edge in the market.

VI. CONCLUSION

Optimizing data collection and integration across complex business systems is essential for organizations looking to stay competitive, improve operational efficiency, and make data-driven decisions. The methodology outlined in this white paper demonstrates the importance of centralizing data, leveraging real-time data streams, and ensuring data quality through automation and machine learning. By integrating emerging technologies such as cloud infrastructure, APIs, and blockchain, businesses can overcome the challenges of fragmented data sources and complex systems, enabling seamless, scalable data flows that support informed decision-making.

The results from implementing this methodology show tangible benefits, including enhanced data accessibility, improved compliance, real-time insights, cost savings, and better customer engagement. The ability to integrate and process data in real time allows businesses to respond

quickly to market changes, customer demands, and regulatory requirements, providing a significant advantage in today's fast-paced, data-driven world.

As businesses continue to grow and adopt new technologies, optimizing data collection and integration will be a key driver of success. By following the strategies outlined in this paper, organizations can not only streamline their data processes but also unlock the full potential of their data, fostering innovation, improving customer experiences, and securing long-term business success.

REFERENCES

1. Vercellis, C., 2011. Business intelligence: data mining and optimization for decision making. John Wiley & Sons.
2. Storbacka, K., 2011. A solution business model: Capabilities and management practices for integrated solutions. Industrial Marketing Management, 40(5), pp.699-711.
3. Lenzerini, M., 2002, June. Data integration: A theoretical perspective. In Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (pp. 233-246).
4. Hassan, N.H.M., Ahmad, K. and Salehuddin, H., 2020. Diagnosing the issues and challenges in data integration implementation in public sector. Int. J. Adv. Sci. Eng. Inf. Technol, 10, pp.529-535.
5. Jagadish, H.V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J.M., Ramakrishnan, R. and Shahabi, C., 2014. Big data and its technical challenges. Communications of the ACM, 57(7), pp.86-94.
6. Sivarajah, U., Kamal, M.M., Irani, Z. and Weerakkody, V., 2017. Critical analysis of Big Data challenges and analytical methods. Journal of business research, 70, pp.263-286.
7. Reyna, A., Martín, C., Chen, J., Soler, E. and Díaz, M., 2018. On blockchain and its integration with IoT. Challenges and opportunities. Future generation computer systems, 88, pp.173-190.
8. Keim, D., Andrienko, G., Fekete, J.D., Görg, C., Kohlhammer, J. and Melançon, G., 2008. Visual analytics: Definition, process, and challenges (pp. 154-175). Springer Berlin Heidelberg.
9. Doan, A., Halevy, A. and Ives, Z., 2012. Principles of data integration. Elsevier.
10. Lenzerini, M., 2002, June. Data integration: A theoretical perspective. In Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (pp. 233-246).
11. Ziegler, P. and Dittrich, K.R., 2007. Data integration—problems, approaches, and perspectives. In Conceptual modelling in information systems engineering (pp. 39-58). Berlin, Heidelberg: Springer Berlin Heidelberg.