Migration from Legacy System to Azure Synapse Analytics: Challenges and Outcomes

Hari Prasad Bomma

Data Engineer, USA
haribomma2007@gmail.com

Abstract

Legacy systems have persistently been a major obstacle for organizations, preventing them from aligning with evolving business needs and technological advances. These outdated systems often lead to inefficiencies, increased maintenance costs, and difficulty in integrating with newer technologies. Consequently, organizations find it challenging to remain competitive and responsive to market changes, ultimately impacting their growth and innovation potential. This research paper delves into the intricacies of migrating from a legacy system to Azure Synapse Analytics, a comprehensive analytics service that offers a unified platform for data integration, data warehousing, and big data analytics.

Keywords: Legacy system, Migration, Azure Synapse, Azure Analytics

Introduction:

Legacy systems, while once the backbone of many organizations, often become problematic as technology evolves. These older systems, typically built on outdated hardware and software, can lead to several challenges that hinder an organization's growth and efficiency. Moreover, the rigid and inflexible nature of legacy systems can make it challenging to meet evolving business needs and adapt to new technological advancements. [8]

High Maintenance Costs: Legacy systems require constant upkeep, and finding skilled personnel to maintain and troubleshoot these outdated systems can be costly and challenging. The expenses associated with hardware parts and specialized support can add up quickly. Often, these systems come with no documentation**,** making problem solving even more time consuming.

Integration Difficulties: As organizations adopt new technologies, integrating these with legacy systems can be complex and resource intensive. Incompatibility issues might arise, making seamless data flow and communication between systems difficult. Additionally, the presence of data duplicity and redundancy can lead to inconsistencies and inefficiencies.

Limited Scalability: Older systems are not designed with the flexibility to scale according to modern business needs. This limitation can restrict an organization's ability to grow and adapt to increasing demands, especially when dealing with huge volumes of data and transactions. Data duplicity, redundant information, and the complexity of legacy systems often hinder organizations from effectively leveraging their data assets.

Security Vulnerabilities: Legacy systems often lack the advanced security features found in newer technologies. This absence of modern security measures makes them vulnerable to cyber attacks and data breaches, posing significant risks to the organization. The lack of governance and lineage further exacerbates these vulnerabilities.

Reduced Efficiency: Over time, legacy systems become sluggish and less reliable, impacting the overall efficiency and productivity of the organization. Outdated interfaces and workflows, coupled with unused reports, can slow down operations and frustrate users.

Compliance Challenges: Keeping legacy systems compliant with modern regulatory standards can be difficult. These systems may not support necessary updates and changes, potentially leading to legal and financial risks. The absence of proper documentation and governance can make this even more challenging.

Inflexibility: Legacy systems are often rigid and difficult to modify or upgrade. This inflexibility can prevent organizations from implementing new features, improving processes, and adopting innovative solutions. The lack of clear data lineage and documentation only adds to the complexity of making changes or improvements.

Transitioning to Azure Synapse Analytics:

Azure Synapse Analytics offers a comprehensive solution to address the challenges faced by organizations with legacy systems. This cloud based platform provides a unified data analytics service that integrates data integration, data warehousing, and big data analytics capabilities [3]. By migrating to Azure Synapse Analytics, organizations can leverage the power of a scalable, flexible, and modern data platform that can handle the growing volume, velocity, and variety of data [4]. Azure Synapse Analytics offers a range of benefits that can help organizations overcome the limitations of legacy systems. By migrating from legacy systems to Azure Synapse Analytics, organizations can unlock the full potential of their data, driving innovation, improving decision making, and gaining a competitive advantage in their respective industries. [1][6][7]

Improved Data Integration and Governance: Azure Synapse Analytics provides a centralized platform for data integration, enabling organizations to consolidate and transform their data from various sources, reducing data silos and improving data governance.

Enhanced Performance and Scalability: The platform's scalable architecture and in memory processing capabilities can handle large scale data processing and analytics, delivering faster insights and improved performance compared to legacy systems. Reusable templates and automation capabilities in Azure Synapse Analytics can streamline the migration process, reducing the time and effort required to transition from legacy systems.

Cost Optimization: By migrating to a cloud based platform, organizations can leverage Azure's pay as you go pricing model, reducing the upfront and ongoing costs associated with maintaining on premises legacy infrastructure.

Increased Agility and Flexibility: Azure Synapse Analytics allows organizations to rapidly adapt to changing business requirements and take advantage of new technologies, enabling them to stay competitive in the rapidly evolving digital landscape.

Improved Scalability and Flexibility: The cloud based nature of Azure Synapse Analytics allows organizations to easily scale their data analytics infrastructure up or down, depending on their evolving business needs and data processing requirements.

Improved Security and Compliance: The platform offers strong security features, including advanced data encryption, access controls, and compliance with industry standards, ensuring the protection of sensitive data and meeting regulatory requirements.

Streamlined Data Analytics: Azure Synapse Analytics provides a unified platform for data warehousing, data lakes, and advanced analytics, allowing organizations to leverage a range of analytical tools and techniques to gain deeper insights and drive data driven decision making.

Notifications and Alert Mechanisms: Azure Synapse Analytics can help organizations proactively identify and address potential issues, reducing the risk of system failures and data loss.

Performance tuning and Optimization features: Azure Synapse Analytics can help organizations maximize the efficiency and performance of their data analytics workloads, leading to faster insights and increased productivity.

Methodology:

The migration from legacy systems to Azure Synapse Analytics involves several crucial steps, ensuring an easy transition and optimized performance. This process encompasses a comprehensive assessment and planning phase, where the existing systems' data sources, models, and workflows are thoroughly analyzed

Assessment and Planning: Start with a thorough evaluation of the current legacy systems, including data sources, models, workflows, and IT infrastructure. This step helps in identifying potential challenges and defining the scope of the migration. It's essential to create a detailed migration plan that outlines timelines, resources, and risk mitigation strategies. *Example:* If a company relies on an old SQL database for their data storage, the assessment phase would involve analyzing the database's structure, the data it holds, and how it interacts with other systems. This comprehensive analysis helps guide decisions on how to map and transfer the data into Azure Synapse Analytics, ensuring compatibility and optimal performance.

Data Analysis: Data analysis involves identifying variables from various source tables to gain insights. The data from these source tables is processed and cleaned to ensure it's in a suitable format for analysis. The target table is defined, aligning with specific business requirements and goals. The data schema is designed to improve readability and maintainability. Updating and optimizing of the data systems may be necessary to optimize performance and scalability for future analyses. *Example:* A retail company analyzes data from source tables like customer information, sales transactions, and product catalog. They process and clean the data, define a target table with relevant fields, and design a data schema for better readability.

Data Migration: After assessing, migrate data from legacy systems to Azure Synapse Analytics. This process can involve several techniques, such as ETL (Extract, Transform, Load), to ensure data is correctly formatted and transferred. During migration, it's crucial to maintain data integrity and security. *Example:* Using Synapse Pipelines in Azure Synapse Analytics, a company can automate the process of extracting data from their legacy systems, transforming it to fit the structure of their new analytics environment. Additionally, Data flows within Synapse allow for cleansing, enriching, and transforming data before loading it into Synapse Analytics.

Pipeline Orchestration Using Synapse: Azure Synapse Analytics provides advanced capabilities for orchestrating data pipelines, enabling integration, transformation, and loading of data from various sources to a target destination. Synapse Pipelines, a key feature of Synapse Analytics, allows you to create and manage complex workflows with ease. This includes orchestrating data movement and transformation activities like running Spark jobs, calling stored procedures, and leveraging data flow activities.

Orchestrating Pipelines:

1. **Data Ingestion:** Synapse Pipelines can ingest data from multiple sources, such as on premises databases, cloud storage, and APIs. This is achieved through various linked services that connect to the source systems.
2. **Data Transformation:** Using Data Flows within Synapse Pipelines, you can perform complex transformations on the ingested data. This includes tasks like filtering, aggregating, joining, and deriving new columns from existing data.
3. **Data Loading:** The transformed data can be loaded into a target destination such as a Synapse SQL Pool, Azure Data Lake Storage, or an Azure SQL Database. This ensures that the data is ready for further analysis and reporting.
4. **Scheduling and Monitoring:** Synapse Pipelines offer scheduling capabilities to automate the execution of data workflows. You can define triggers based on time or events to ensure data is processed at the right intervals. Monitoring features provide insights into pipeline performance, helping you identify and resolve issues promptly.

Application Integration: Ensure existing applications and workflows are integrated with Azure Synapse Analytics for smooth data flow and interoperability. This integration often requires updating API connections, adapting data pipelines, and testing interactions between systems to minimize disruptions. *Example:* A retail company using a legacy ERP system can integrate it with Azure Synapse Analytics, allowing real time data from sales, inventory, and customer databases to be analyzed in a unified platform.

Optimization and Tuning: Post migration, focus on optimizing the performance and efficiency of Azure Synapse Analytics using features like query optimization, workload management, and resource scaling. This step helps in enhancing the system's responsiveness and cost effectiveness by fine tuning its operations. *Example:* By analyzing query performance, a financial institution can identify bottlenecks and optimize SQL queries, leading to faster data processing times and more efficient resource utilization.

Testing and Validation: Conduct extensive testing and validation to ensure the integrity, accuracy, and reliability of the data and system performance before completing the migration. This stage includes various testing methods, such as unit testing, integration testing, and user acceptance testing, to confirm the system's readiness. *Example:* A healthcare provider can use test data to validate that patient information is accurately migrated and matched across systems, ensuring there are no discrepancies or data loss.

Notifications and Alert Mechanisms: Email notifications are essential for monitoring jobs and informing stakeholders about their status. In case of job failures, immediate email alerts are sent to administrators to prompt quick troubleshooting. Success notifications can be sent to users to inform when tasks are completed, ensuring transparency and efficient workflow. *Example:* A retail company can use automated notifications if a data loading job fails in Synapse, an immediate email alert is sent to administrators with details about the error, for quick troubleshooting. Once the data processing job is successful, a "Report Ready" notification is sent to users, informing them that the latest sales report is now available for review.

Change Management and Training: A successful migration requires a well planned change management strategy, including training employees and adopting new processes and tools. Communicating the benefits of the new system and providing hands on training sessions can help ensure a smooth transition. *Example:* Organizing workshops and training sessions for the IT staff and end users of a manufacturing company can help them become familiar with Azure Synapse Analytics, making them more comfortable with the new system and its features.

Following this comprehensive approach, organizations can effectively navigate the challenges of migrating from legacy systems to Azure Synapse Analytics and unlock the full potential of their data analytics capabilities.

Considerations:

The migration from legacy systems to Azure Synapse Analytics is not without its challenges. Some of the key challenges and considerations include:

Data Complexity and Heterogeneity: Legacy systems often contain large amounts of data stored in various formats, structures, and locations. Consolidating this data presents a significant challenge. It requires extensive data mapping, transformation, and cleansing efforts. These steps are essential to standardize and harmonize the data effectively.

Legacy System Integration: Integrating legacy systems with the Azure Synapse Analytics platform can be complex, as organizations may need to address compatibility issues, data migration, and the integration of existing workflows and applications.

Organizational Change Management: Migrating to a new data analytics platform can disrupt existing business processes and workflows, requiring a well planned change management strategy to ensure smooth adoption and minimize resistance to change.

Skill Gap and Talent Acquisition: The migration to Azure Synapse Analytics may require organizations to upskill their existing workforce or

acquire new talent with the necessary expertise in cloud based data analytics, data engineering, and Azure platform management.

Security and Compliance: Ensuring the security and compliance of data during the migration process is crucial, as organizations must adhere to industry regulations, data privacy laws, and best practices for data protection.

Conclusion:

The migration from legacy systems to Azure Synapse Analytics is a complex but necessary journey for organizations seeking to harness the power of data driven decision making and advanced analytics. By following a comprehensive approach that includes thorough assessment, data migration, application integration, optimization, testing, and change management, organizations can successfully navigate the challenges and unlock the benefits of this transformation.

Azure Synapse Analytics provides a unified platform that combines data warehousing, data lakes, and advanced analytics capabilities, enabling organizations to gain deeper insights, improve scalability, optimize costs, and increase efficiency. Although the migration process may present various challenges, such as data complexity, legacy system integration, organizational change, and skill gaps, a well planned and executed strategy can help organizations overcome these obstacles and realize the full potential of their

data analytics capabilities. By embracing the migration to Azure Synapse Analytics, organizations can position themselves for long term success, staying ahead of the competition, and driving innovation in their respective industries.

References:

1. Bohlouli, M., Schulz, F., Angelis, L., Pahor, D., Brandić, I., Atlan, D., & Tate, A. R. (2020).*" Towards an Integrated Platform for Big Data Analysis".* In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2004.13021
2. Chawla, H., & Khattar, P. (2020).*" Data Preparation and Training Part I."* In Apress eBooks (p. 99). https://doi.org/10.1007/978-1-4842-6252-8_6
3. Chessell, M., Wolfson, D., & Vincent, T. (2017). *"Architecting to Deliver Value From a Big Data and Hybrid Cloud Architecture".* In Elsevier eBooks (p. 33). Elsevier BV. https://doi.org/10.1016/b978-0-12-805467-3.00003-x
4. Henke, N., Bughin, J., Chui, M., Manyika, J., Saleh, T., & Wiseman, B. (2016). *"The age of analytics: competing in a data-driven world".* https://apo.org.au/node/241286
5. Kishnani, N. (2019). *"Talent Analytics for Organizations of 21st Century".* In SSRN Electronic Journal. RELX Group (Netherlands). https://doi.org/10.2139/ssrn.3307505
6. Mao, Y., Fu, Y., Gu, S., Vhaduri, S., Cheng, L., & Liu, Q. (2020). *"Resource Management Schemes for Cloud-Native Platforms with Computing Containers of Docker and Kubernetes."* In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2010.10350
7. More, S., Chowdhury, S., Kumar, R., Bohare, S., & Khairnar, Prof. D. (2022). *"Research and Analysis of Different Cloud Storage Databases. In International Journal of Advanced Research in Science Communication and Technology"* (p. 296). Shivkrupa Publication's. https://doi.org/10.48175/ijarsct-3922
8. Strobl, S., Bernhart, M., & Grechenig, T. (2020).*" Towards a Topology for Legacy System Migration "* (p. 586). https://doi.org/10.1145/3387940.3391476