Data Migration from On-Premise Database to SQL Managed Instance on Azure Cloud Using Azure Data Factory – A Working Approach

Chandrika Bhargavi Achanta

School of Information Technology, the University of Cincinnati.

1 OVERVIEW

The selected paper explores the critical aspects of data migration, focusing on the transition from on-premise Oracle databases to cloud-based SQL Managed Instance databases. This paper presents a comprehensive case study employing Azure Data Factory, a cloud-based ETL (Extract, Transform, Load) tool, to facilitate the data migration process for an organization[1]. The objective of the paper is to evaluate the implementation of data migration, emphasizing the use of Azure Data Factory and the associated technologies involved.

The paper begins by highlighting the increasing importance of data migration, particularly in the context of moving applications to the cloud. It emphasizes the need for thorough discussions on data migration essentials and the establishment of proper architecture[2]. The significance of conducting a proof of concept to ensure data integrity and the seamless functioning of existing logic post-migration.

A significant aspect discussed in the paper is the encryption process during data migration, especially when dealing with sensitive user data in Oracle databases. This delves into the analysis of existing encryption algorithms and the implementation of a suitable architecture using Azure Data Factory to maintain data integrity during the migration process. The role of an ETL tool, in this case, Azure Data Factory, is elucidated as a crucial component in the data migration process. ETL, as a process, involves data extraction, transformation based on business rules, and loading into the target system. It emphasizes that ETL tools play a vital role in connecting source and target databases and facilitating the migration process, enabling applications to transition to the cloud with minimal downtime and effective data validation methods.

A critical discussion in the paper revolves around the migration of database objects such as Stored Procedures, Triggers, and Functions. This acknowledges the manual effort involved in rewriting these objects and introduces the SQL Server Managed Assistance (SSMA) tool to expedite the conversion process.

2 ALIGNMENT WITH THE COURSE

This paper aligns with the course content by addressing the crucial topic of data migration, specifically focusing on the use of Azure Data Factory for migrating data from on-premise Oracle/any popular databases to cloud-based SQL Managed Instance databases. The paper delves into the essential considerations for data migration, such as the need for proper architecture, proof of concept, and encryption processes. It emphasizes the significance of evaluating factors like table size, throughput, and network bandwidth during the data migration process. Additionally, the paper explores the role of an ETL (Extract, Transform, Load) tool, highlighting Azure Data

Factory's capabilities in connecting source/target databases and facilitating the migration with minimum downtime. The discussion on creating architecture for data migration, including the use of configuration tables, pipelines, and activities, aligns with the course's emphasis on practical implementation and strategies for efficient data movement. Overall, the paper provides a practical case study that reinforces the concepts learned in the course regarding data migration and the use of relevant tools in a real-world scenario.

3 COMPANIES THAT ARE USING AZURE RESOURCES TO SOLVE BUSINESS PROBLEMS

Various companies have successfully implemented a comparable strategy, utilizing Azure resources for data migration and cloud adoption. For instance, Smithfield Foods, a company with a business worth $15 billion, has effectively employed Azure within a hybrid cloud model. This strategic move resulted in a notable 60% reduction in data center costs, a streamlined application delivery timeline from two months to just one day, and a substantial enhancement in network security.

Similarly, Chevron has opted for a hybrid migration approach, utilizing Azure to facilitate the transition of numerous applications while still making the most of existing infrastructure and technology. The collaboration between Chevron and Microsoft has played a crucial role in comprehending enterprise needs and navigating the intricacies of the migration journey.

J.B. Hunt, another successful example, has embraced a hybrid cloud infrastructure facilitated by Azure. This approach enables J.B. Hunt to strategically migrate workloads to the cloud while continuing to operate critical business functions on-premises. Microsoft's flexibility in adapting to its existing IT landscape has been instrumental in the seamless and successful transition.

These instances highlight the adaptability and effectiveness of Azure in addressing diverse business challenges through strategic data migration and the adoption of hybrid cloud solutions. The success stories of Smithfield Foods, Chevron, and J.B. Hunt underscore Azure's capacity to cater to various organizational needs, bringing about positive outcomes such as cost reduction, operational efficiency, and heightened security.

4 ARCHITECTURE

In the conventional approach, the data migration architecture involves a basic pipeline with a copy activity for each table, directly transferring data from the on-premise database to the SQL Server database in Azure. In the advanced approach, a more sophisticated architecture is implemented[3]. The pipeline includes a lookup activity to fetch all tables, followed by a for each loop to iterate through each table. Within the loop, copy activities are utilized to transfer data from the on-premise database to the SQL Server database in Azure. This advanced approach offers a more scalable and efficient solution for migrating multiple tables seamlessly.

Mater pipeline
- Load master configuration table.

Parallel pipeline
- Number of executes activities configured for data migration matter.
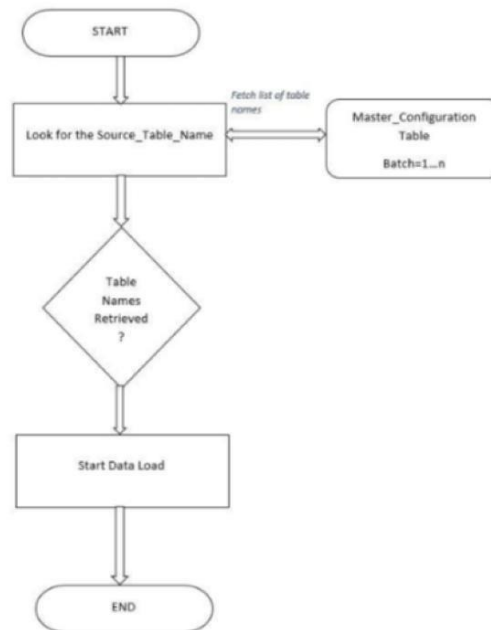
Base pipeline
- Look up and for each activity.

Fig 1 : Flow of Master, Parallel and Base pipeline system in ADF

Key actions can be used to efficiently orchestrate the data migration process from an on-premises Oracle database to a SQL Managed Instance. The primary function of the "Copy Data" action is to facilitate the transfer of data from the source to the destination. To ensure smooth communication, use this activity to define connected services for both Oracle and SQL. Use "ForEach" and "Lookup" operations for more precise control and transformation. Iterative processes are enabled by the "ForEach" activity, which is advantageous in situations such as batch processing. Concurrently, the "Lookup" function enables the querying of external data while the pipeline is running, offering useful contextual data for making decisions or setting up dynamic setups.
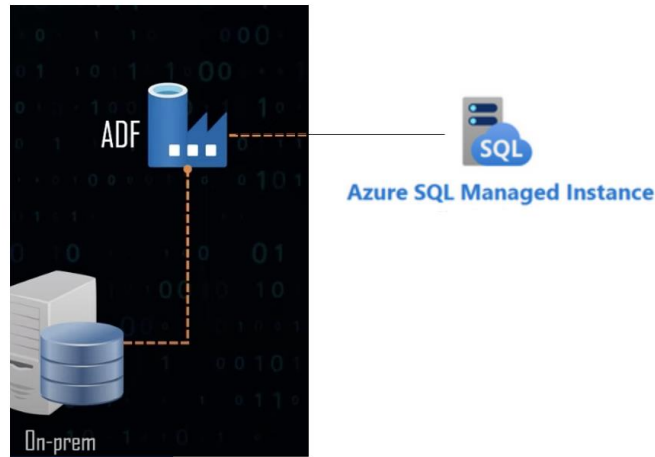
Fig 2 : Architecture of Data migration from on-prem to SQL managed instance

5   CONCLUSION

Cloud computing, especially in the realm of data engineering, serves as a powerful catalyst for data aggregation and learning. The paper highlights the significance of data migration from on-premise databases to cloud-based SQL-managed instances using Azure Data Factory[4]. The adoption of cloud technologies facilitates seamless data movement, enabling organizations to leverage the benefits of scalability, efficiency, and cost-effectiveness. By utilizing tools like Azure Data Factory, the migration process becomes streamlined, allowing for the efficient transfer of data while addressing encryption, architecture, and validation considerations[5]. Cloud computing indeed acts as a catalyst, providing a robust infrastructure for organizations to aggregate and learn from their data effectively.

REFERENCES

[1]   Milind Jadhav, et. al. (2021). Data migration from on-premise Oracle Database to SQL manage instance on Azure cloud using Azure data factory - a working approach. Turkish Journal of Computer and Mathematics Education (TURCOMAT). https://turcomat.org/index.php/turkbilmat/article/view/8093

[2]   Kajáti, E., & Miškuf, M. (2017). *Advanced analysis of manufacturing data in Excel and its add-ins | IEEE …* IEEE Xplore. https://ieeexplore.ieee.org/document/7880359

[3]   Jamshidi, P., & Ahmad, A. (2013). *Cloud migration research: A systematic review | IEEE journals …* IEEE Xplore. https://ieeexplore.ieee.org/document/6624108

[4]   Rawat, S., & Narain, A. (2018). *Understanding Azure Data Factory: Operationalizing Big Data and advanced analytics solutions*. O'Reilly Online Learning. https://www.oreilly.com/library/view/understanding-azure-data/9781484241226/

[5]   Pahl, C. (2013). *A comparison of on-premise to cloud migration approaches*. Research gate. https://www.researchgate.net/publication/286437766_A_Comparison_of_On-Premise_to_Cloud_Migration_Approaches