L20-1116 Faiqa Adnan & L20-0921 Aisha M Nawaz
PDC BSCS 6E PROJECT

Azure Synapse Analytics, HDInsight, and Azure Databricks
Summary

INTRODUCTION:

In today's digital era where data is being generated at unprecedented rates, businesses find it hard to manage & integrate data in the data-driven environment. Microsoft Azure services provide a solution to just this. Focusing on Big data & analytics, these services provide companies with a robust, cloud-based data platform to manage, integrate and analyze their data. Microsoft Azure's services, Azure Synapse, HDInsight & Azure Databricks, each satisfy different needs and together provide complete & comprehensive tools for building & managing big data solutions. These services can be used by large companies that have multiple systems gathering data from nation-wide to help manage & analyze their data quickly & efficiently that was otherwise inconsistent & incomplete. It helps companies integrate data sources into one platform giving a unified view and aiding in identification of trends & patterns in data which can be useful for making wise business decisions.

AZURE SYNAPSE ANALYTICS:

Azure synapse is an integrated analytics service that uses massively parallel processing. It combines big data and data warehousing into a single service. By using serverless or provisioned resources, businesses are allowed to query & analyze large amounts of data. It can be flawlessly integrated with other Microsoft products like Power BI & Azure Machine Learning to provide holistic data analytics & integration solutions. It is cloud-based & supports structured, semi-structured & unstructured data. Azure Synapse Analytics service is quick, efficient & may be viewed as a virtual assistance for your data as it provides intelligent insights & recommendations to aid in making informed business decisions.

To drive businesses forward, several unique features & capabilities of Azure Synapse can be leveraged like: Serverless or Provisioned Resources, Seamless integration & Security features. Using Serverless Resources, businesses can query & analyze large volumes of data without additional stress of underlying infrastructure. This is especially useful for businesses that have unpredictable workloads or wish to minimize costs. By using Provisioned Resources, businesses can allocate dedicated computing and storage resources to handle different workloads. This is helpful for businesses with predictable or high-volume workloads. Due to seamless integrations with other Microsoft products like Power BI, it is easy to create dashboards, reports & predictive models using data stored in Azure Synapse. Lastly, robust security features are provided to ensure the confidentiality, integrity and availability of data. Azure Synapse Analytics provides support for numerous Authorization & Authentication mechanisms while also providing encryption of data in transit and at rest. It also provides the features of data ingestion, preparation, Data Warehousing, Big data analytics & Machine Learning.

Industries that have successfully made use of Azure Synapse include: Healthcare, Financial Services & Retail. Moreover, Business outcomes of using Azure Synapse include: Improved Decision Making, Increased Efficiency, Enhanced Security, Improved customer Experience & Faster time-to-market

However, the service has its disadvantages that includes cost, it has a steep learning curve so complexity is high for new users, has limited compatibility & customization.

AZURE HDINSIGHT:

Using a simple yet cost-effective way to analyze large amounts of data on the cloud, HDInsight, much like Azure Synapse Analytics, is a cloud-based big data analytics service. It provides a platform for running well-known big data processing frameworks like: Apache Hadoop, Spark Hive, HBase & Storm. It simplifies the process of deploying, managing & scaling Apache Hadoop clusters in the cloud and allows running of big data workloads like: Batch processing, Interactive Querying & Real-time stream processing. It supports structured & unstructured data and offers integration with Azure services like: Azure Blob Storage & Azure Data Lake Storage.

You may be wondering what benefits are there of using Azure HDInsight? Well, First of all, it is a highly scalable platform for big data processing, is a cost-effective way to process & analyze big data without the need to manage complex infrastructure or invest in hardware, it seamlessly integrates with other Azure services making it easier to store & access big data, it consists of built in security features that help provide a secure environment. Lastly, it has a user-friendly interface making it easy to use for developers & data scientists to process & analyze big data. Additionally, it has features of Data ingestion, processing, Analysis & visualization. However, there are some downsides as well like high complexity for new users, dependency on cloud infrastructure, limited control over underlying infrastructure & additional data transfer cost for transferring large volumes of data to & from Azure HDInsight.

AZURE DATABRICKS:

Azure Databricks is a very fast, easy & collaborative Apache Spark-based Analytics allowing organizations to quickly build & deploy big data solutions. It provides a unified analytics platform for Data engineering, Machine Learning & Analytic. Azure Databricks are a combination of Apache Spark, Databricks & Enterprise Cloud. It enables users to collaborate & share insights in real-time. Azure Databricks, like other services discussed above, has several features that include: Data ingestion, preparation, Analysis & Machine Learning. It supports structured & unstructured data and offers integration with Azure services like: Azure Data factory, Azure Blob storage & Azure Data Lake Storage.

Advantages of Azure Databricks include the fact that clusters are easy to setup and configure here, it has an Azure Synapse Analytics connector & can connect to Azure DB, it is integrated with Active Directory & supports multiple languages. Disadvantages are that it does not integrate with Git or any other versioning tool. It currently only supports HDInsight and not Azure Batch or Azure Distributed Data Engineering Toolkit (AZTK).

CONCLUSION:

All in all, we discussed 3 powerful big data analytics & services that together provide a comprehensive suite of tools for building & managing big data solutions. We also discussed several advantages & disadvantages of using each service. They all cater different needs & use cases to help unlock the true value of your data and together provide a complete solution for organizations looking to leverage the power of big data to improve businesses.