Akilesh K

k.akilesh123@gmail.com

Data engineering - Batch 1

Date: 27-02-24

DAY 26 – Azure-Synapse

Azure Synapse Analytics, formerly known as Azure SQL Data Warehouse, is a cloud-based analytics service provided by Microsoft Azure. It combines enterprise data warehousing, big data analytics, and data integration to enable organizations to analyse large volumes of data for business insights.

overview of Azure Synapse and its architecture:

Components:

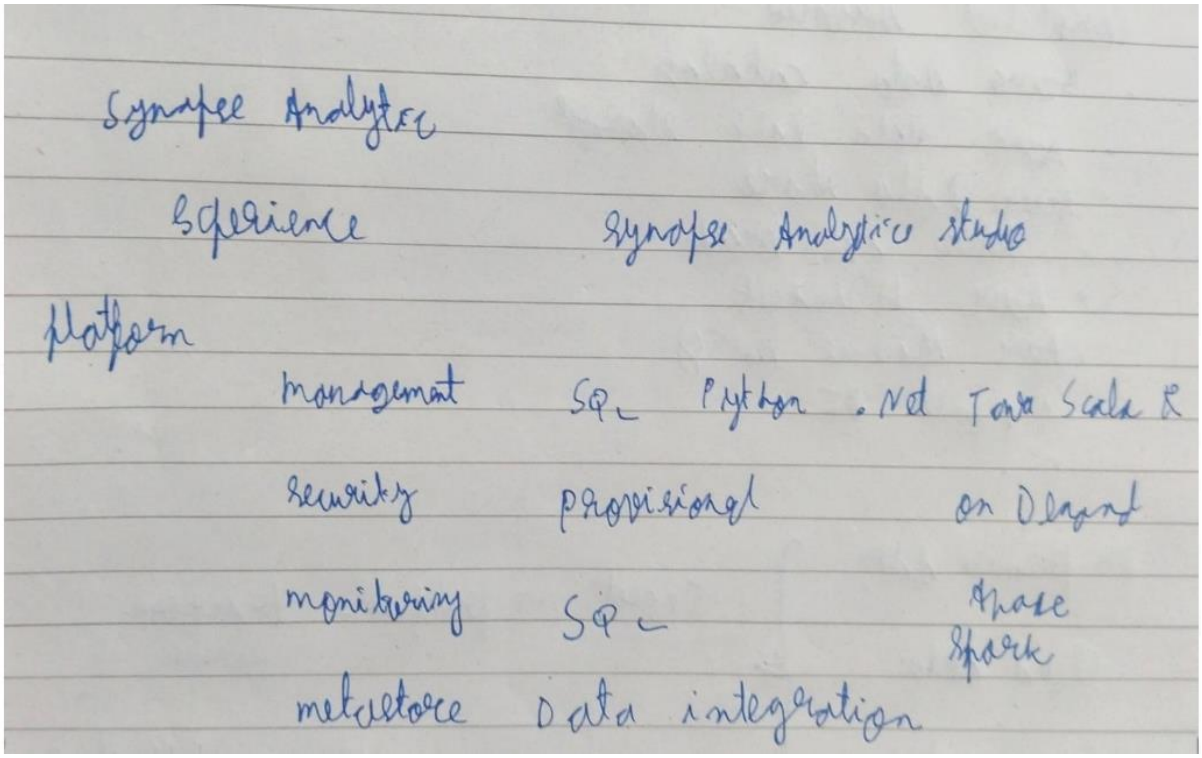
- **SQL Pools**: Azure Synapse provides SQL Pools, which are scalable collections of resources used for data querying and analysis. SQL Pools support both traditional relational data and big data processing.
- **Spark Pools:** Azure Synapse integrates with Apache Spark to provide scalable big data processing capabilities. Spark Pools are used for running Spark jobs on large datasets for data exploration, machine learning, and other big data tasks.
- **Integration with Azure Data Lake Storage (ADLS):** Azure Synapse integrates seamlessly with ADLS, allowing users to store large volumes of structured and unstructured data in a cost-effective manner.
- **Data Integration:** Azure Synapse offers built-in data integration capabilities, allowing users to ingest data from various sources, transform it, and load it into data warehouses or data lakes.

Architecture:

- **Control Plane:** The control plane of Azure Synapse includes management services responsible for provisioning, monitoring, and managing resources. It provides a unified interface for managing SQL and Spark resources, as well as data integration pipelines.
- **Compute:** Azure Synapse leverages distributed compute resources to perform data processing and analysis tasks. These resources can be dynamically scaled up or down based on workload requirements.
- **Storage:** Azure Synapse supports various storage options, including dedicated SQL Pools for structured data storage and ADLS for scalable, cost-effective storage of structured and unstructured data.
- **Query Processing:** Query processing in Azure Synapse involves translating SQL and Spark queries into execution plans that leverage distributed compute and storage resources. The system optimizes query execution for performance and scalability.
- **Security and Compliance:** Azure Synapse provides robust security features, including encryption, authentication, and authorization mechanisms to protect data and comply with regulatory requirements.

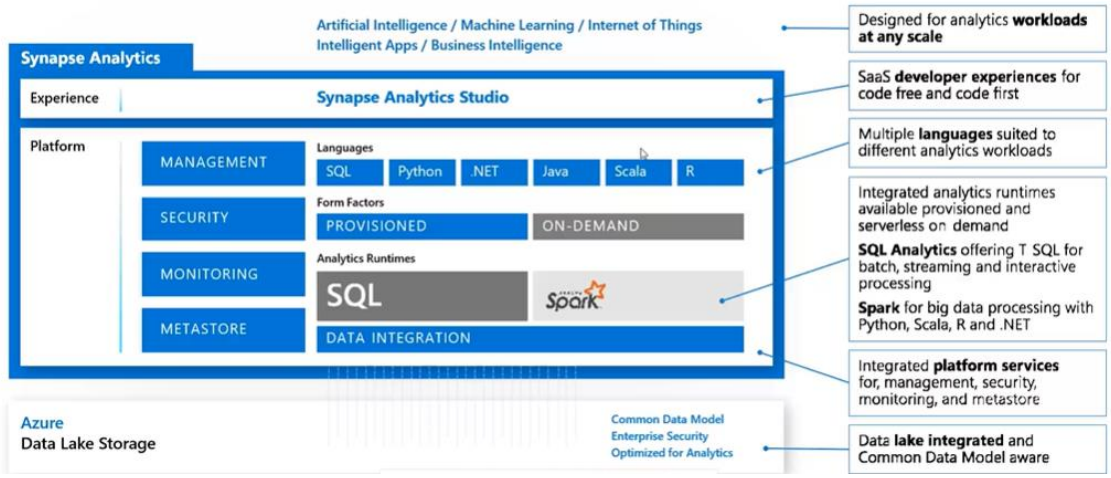
Use Cases:

- **Data Warehousing:** Azure Synapse is commonly used for building and managing data warehouses for business intelligence and analytics.
- **Big Data Analytics:** With its integration with Apache Spark, Azure Synapse enables organizations to perform advanced analytics and machine learning on large datasets.
- **Data Integration**: Azure Synapse simplifies data integration tasks by providing built-in connectors and tools for ingesting, transforming, and loading data from various sources.



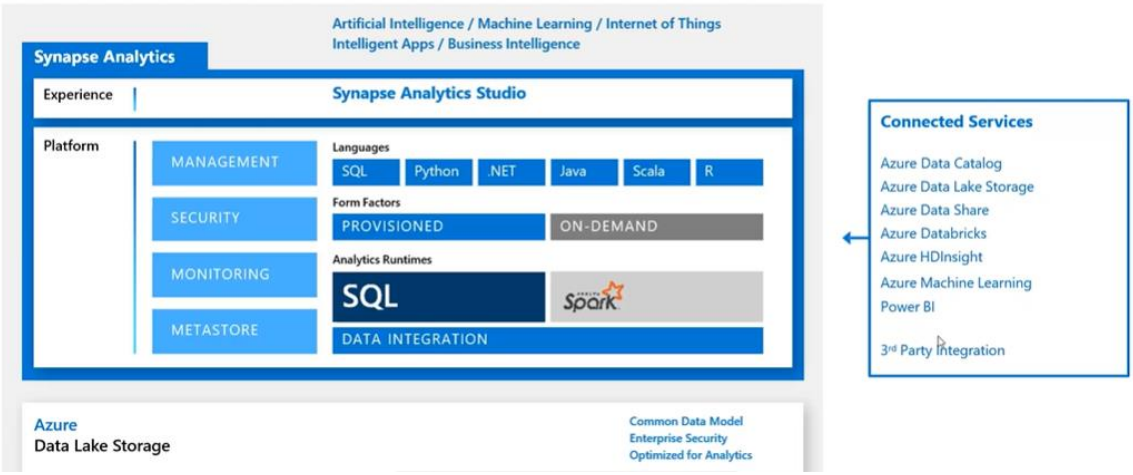
Azure Synapse Analytics

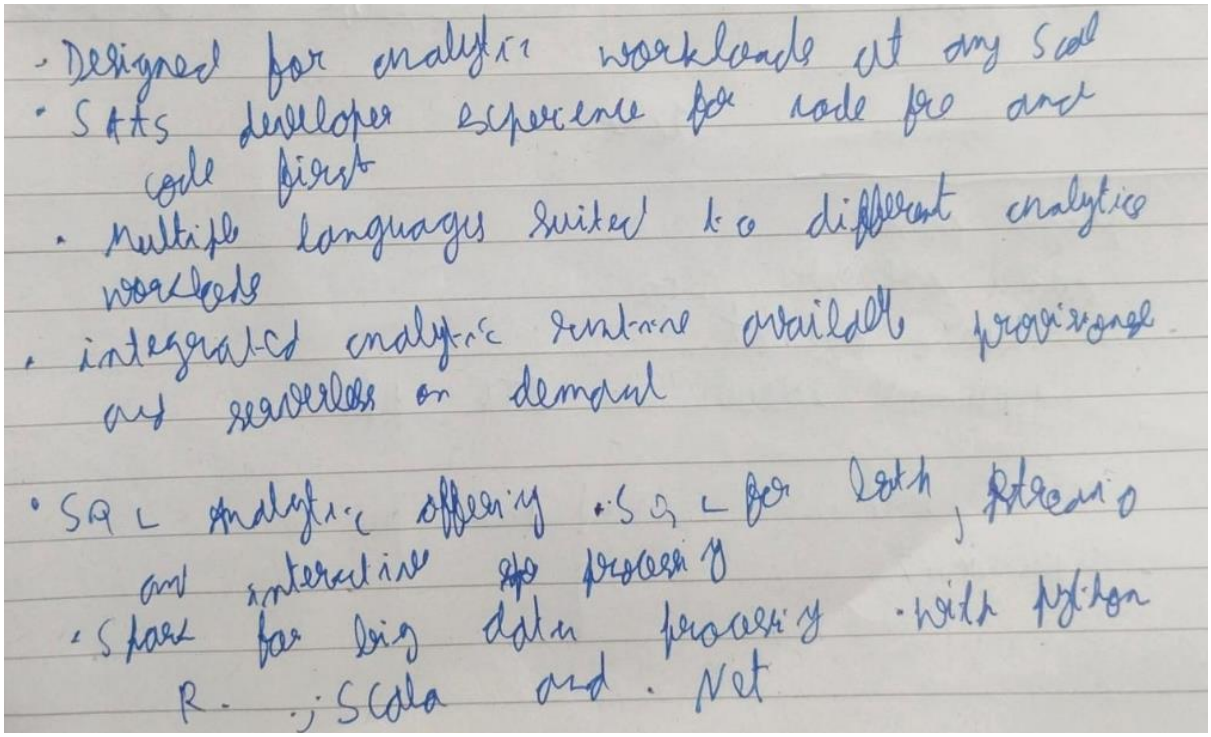
Integrated data platform for BI, AI and continuous intelligence



Azure Synapse Analytics

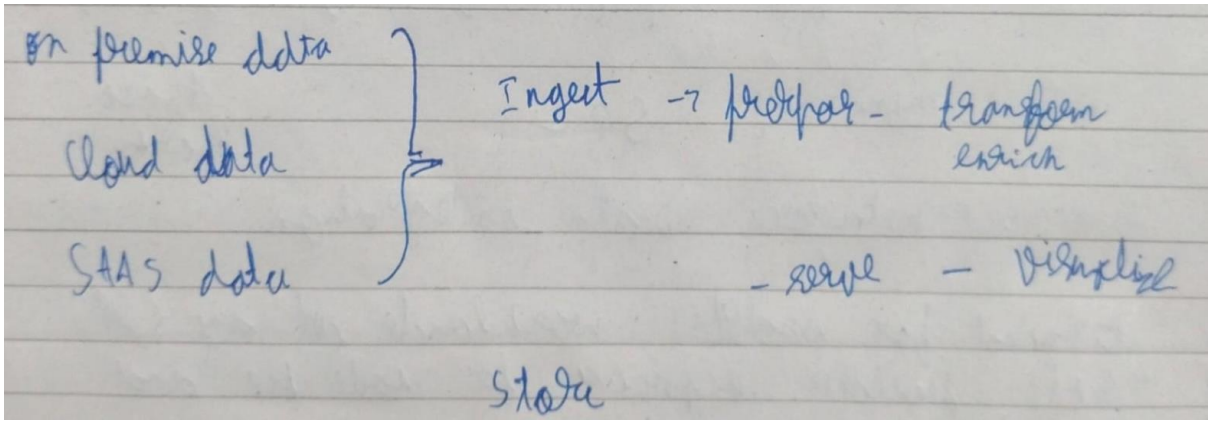
Integrated data platform for BI, AI and continuous intelligence





Connected services

- Azure Data catalog
- Azure Data lake storage
- Azure Data share
- Azure data bricks
- Azure HDInsight
- Azure Machine learning
- Power BI



In summary, Azure Synapse Analytics is a comprehensive analytics service that combines data warehousing, big data processing, and data integration capabilities in a unified platform. Its architecture is designed to provide scalability, performance, and flexibility to meet the diverse needs of modern data-driven organizations.