Big Data and Analytics
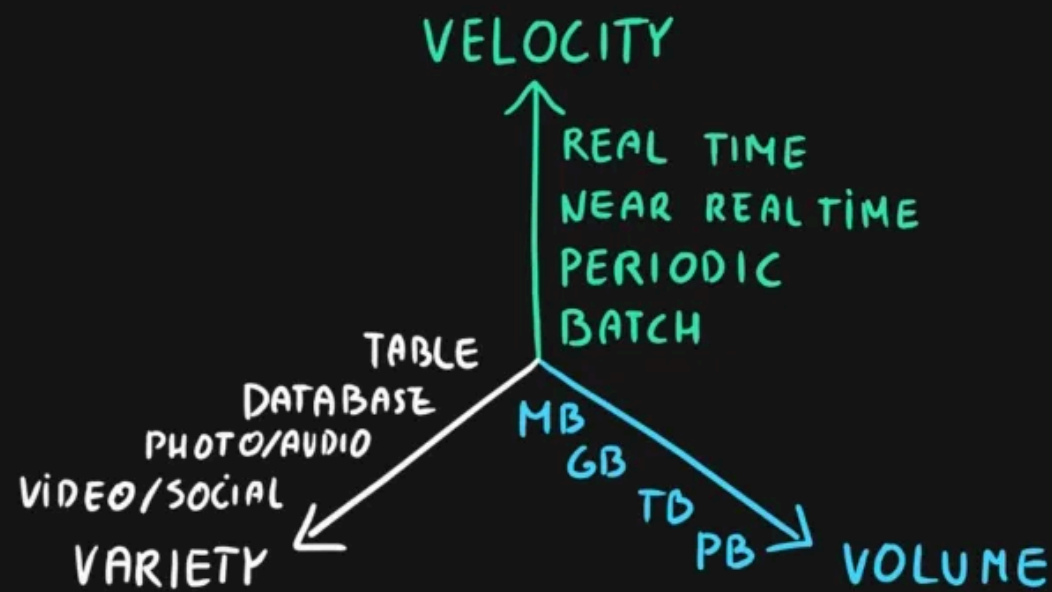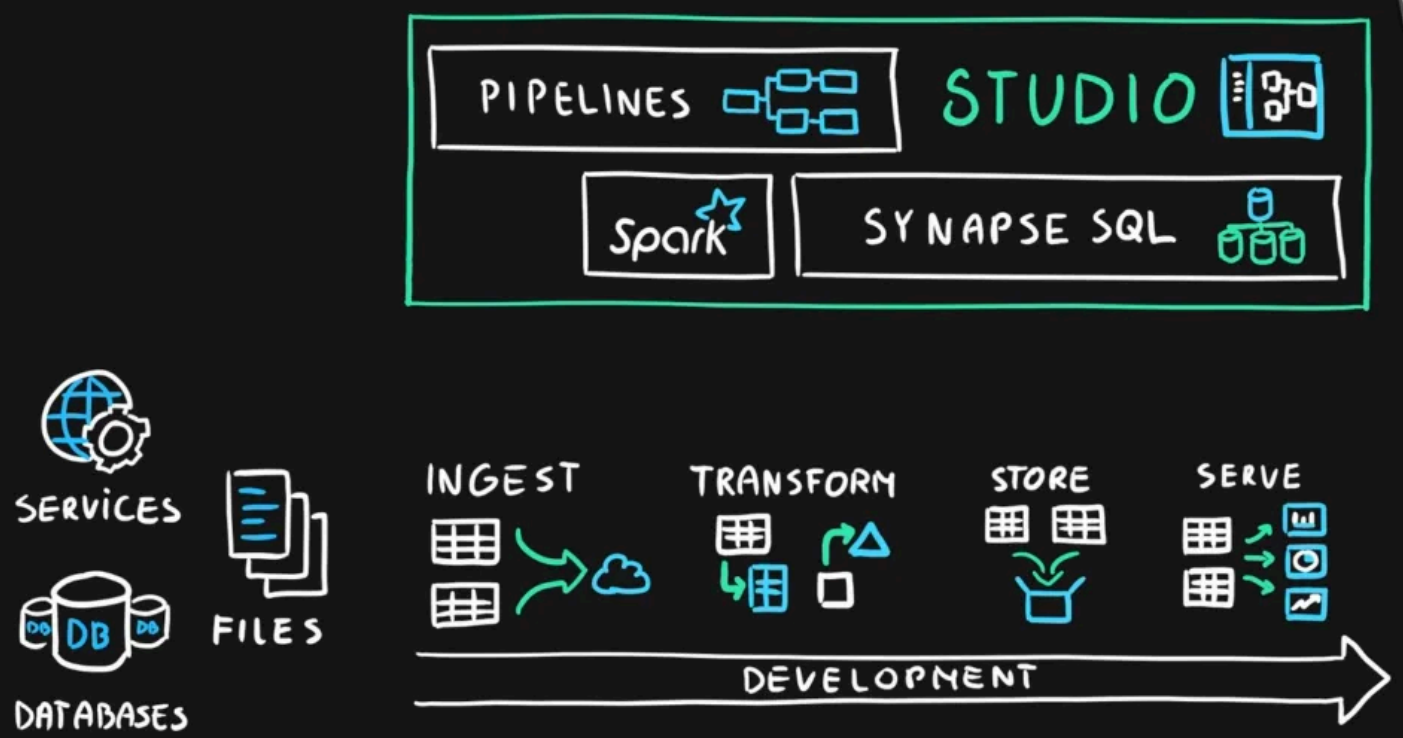
- Describe Big Data and Analytics products such as
  - Azure Synapse Analytics
  - Azure HDInsight
  - Azure Databricks

Big Data is a field that of technology that helps with extraction, processing and analysis of information that is too large or complex to deal with traditional software.

VELOCITY

REAL TIME
NEAR REAL TIME
PERIODIC
BATCH

TABLE
DATABASE
PHOTO/AUDIO
VIDEO/SOCIAL
VARIETY

MB
GB
TB
PB
VOLUME

Azure Synapse Analytics
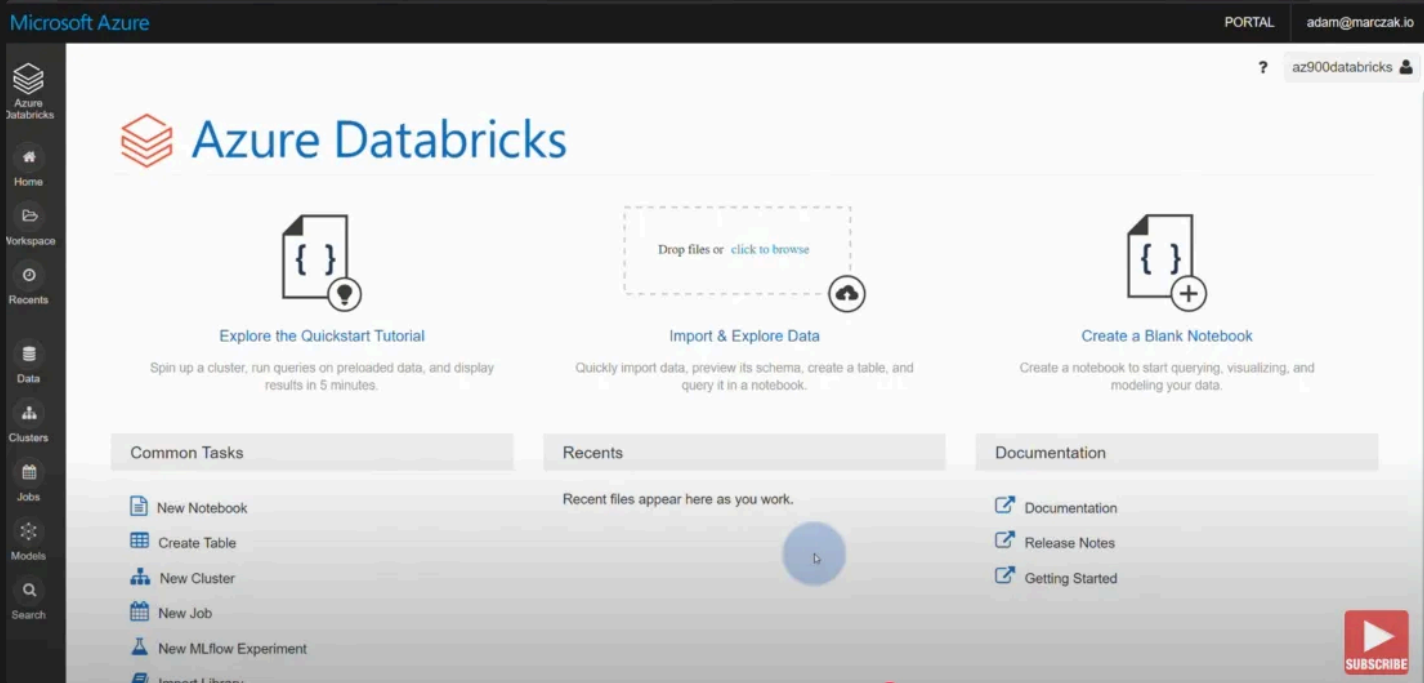
Azure Synapse Analytics

Key Characteristics

- Big data analytics platform (PaaS)

- Multiple components
  - Spark
  - Synapse SQL
    - SQL pools (dedicated – pay for provisioned performance)
    - SQL on-demand (ad-hoc – pay for TB processed)
  - Synapse Pipelines (Data Factory – ETL)
  - Studio (unified experience)

Azure HDInsight

Azure Databricks

Key Characteristics

- **Big data collaboration platform** (PaaS)

- **Unified workspace** for notebook, cluster, data, access management and collaboration

- Based on **Spark**

- **Integrates** very well with common Azure data services

Summary

- **Azure Synapse Analytics** – Modern workspace for end-2-end enterprise data warehousing & analytics with a lot of integrated tools like Data Factory, Spark, SQL, etc.

- **Azure HDInsight** – Fully-managed open source analytics service with a lot of supported frameworks & tools like Hadoop, Spark, Kafka, Hive, etc.

- **Azure Databricks** – Apache Spark based analytics platform for data transformation and collaborative analytics