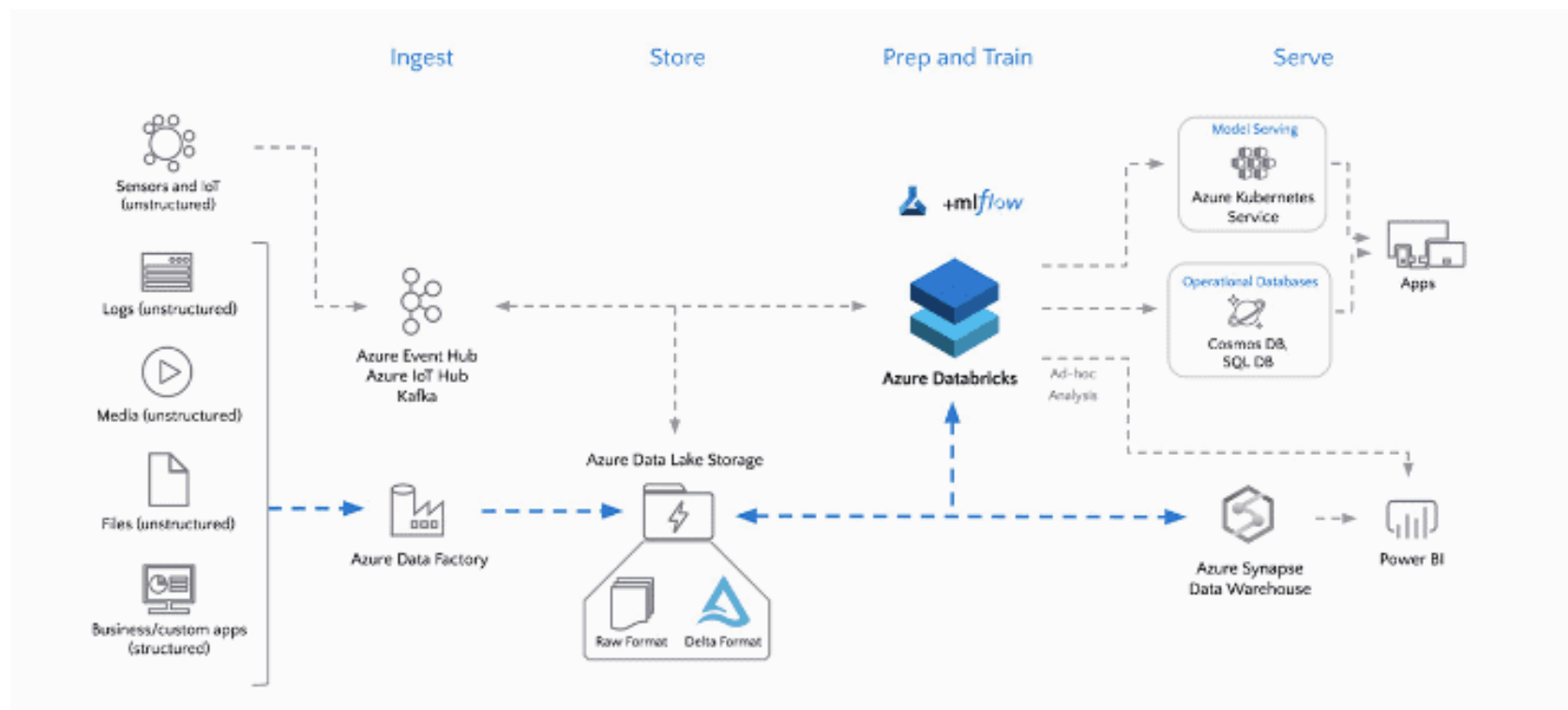# What Is Azure Data Factory (ADF)?

- Azure Data Factory is a fully managed, cloud-based data integration service that allows you to create, schedule, and monitor data pipelines.
- Can also be defined as a data integration service.
- The aim of Azure Data Factory is to fetch data from one or more data sources and convert them into a format that we process.
- The data sources might contain noise that we need to filter out. Azure Data Factory enables us to pull the interesting data and remove the rest.
- Azure Data Factory to ingest data and load the data from a variety of sources into Azure Data Lake Storage.
- It is the cloud-based ETL service that allows us to create data-driven pipelines for orchestrating data movement and transforming data at scale.

Read more about Azure Data Factory on this blogpost!

What Is Data Integration Service?

- Data integration involves the collection of data from one or more sources.
- Then includes a process where the data may be transformed and cleansed or may be augmented with additional data and prepared.
- Finally, the combined data is stored in a data platform service that deals with the type of analytics that we want to perform.
- This process can be automated by ADF in an arrangement known as Extract, Transform, and Load (ETL).

How does Azure Data Factory (ADF) work?

Connect and Collect

- Enterprises have data of various types such as structured, unstructured, and semi-structured.
- The first step collects all the data from a different source and then move the data to a centralized location for subsequent processing.
- We can use the Copy Activity in a data pipeline to move data from both cloud source and on-premises data stores to a centralized data store in the cloud.

Transform and Enrich

- After data is available in a centralized data store in the cloud, transform, or process the collected data by using ADF mapping data flows.
- ADF supports external activities for executing our transformations on compute services such as Spark, HDInsight Hadoop, Machine Learning, Data Lake Analytics.

CI/CD and Publish

- ADF offers full support for CI/CD of our data pipelines using GitHub and Azure DevOps.

More on CI/CD on this blogpost!

- After the raw data has been refined, ad the data into Azure SQL Database, Azure Data Warehouse, Azure CosmosDB.

K21Academy
Learn Cloud from Experts

Monitor

- ADF has built-in support for pipeline monitoring via Azure Monitor, PowerShell, API, Azure Monitor logs, and health panels on the Azure portal.

Pipeline

- A pipeline is a logical grouping of activities that execute a unit of work. Together, the activities in a pipeline execute a task.