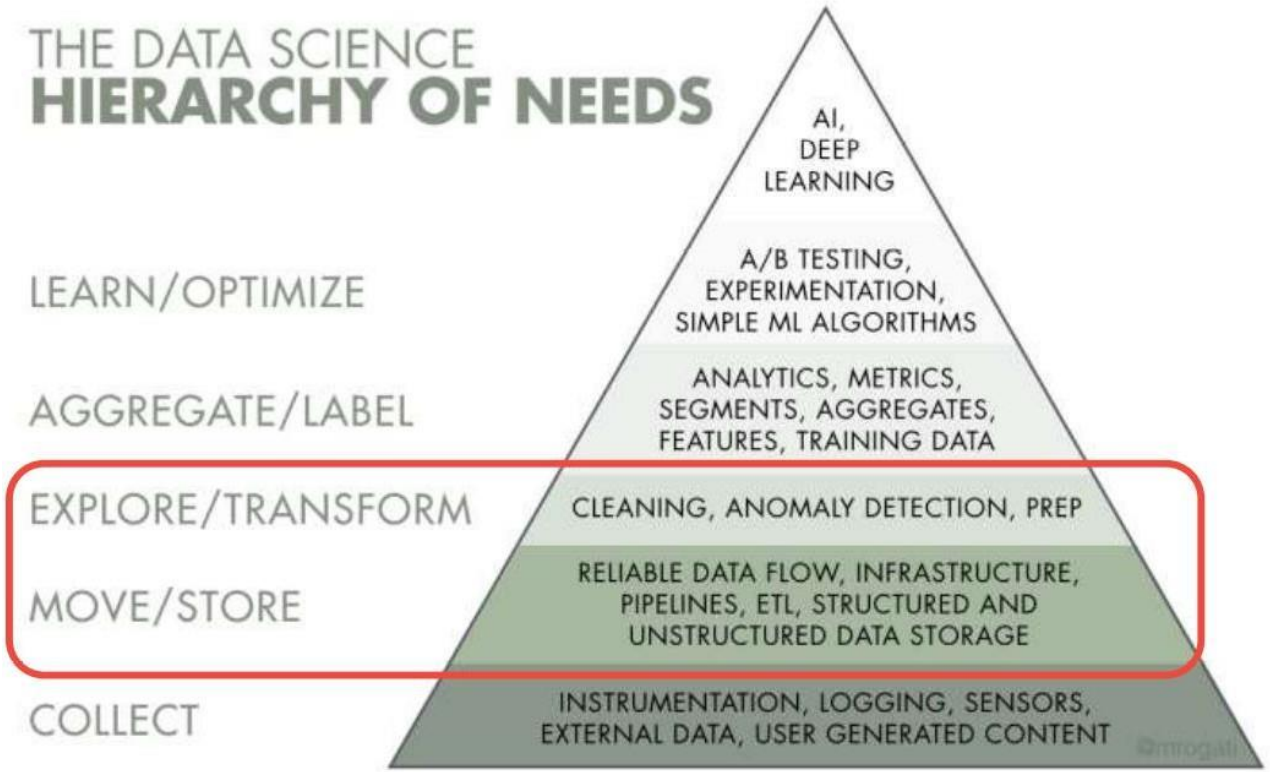
Azure Data Platform

Live Teaching Session

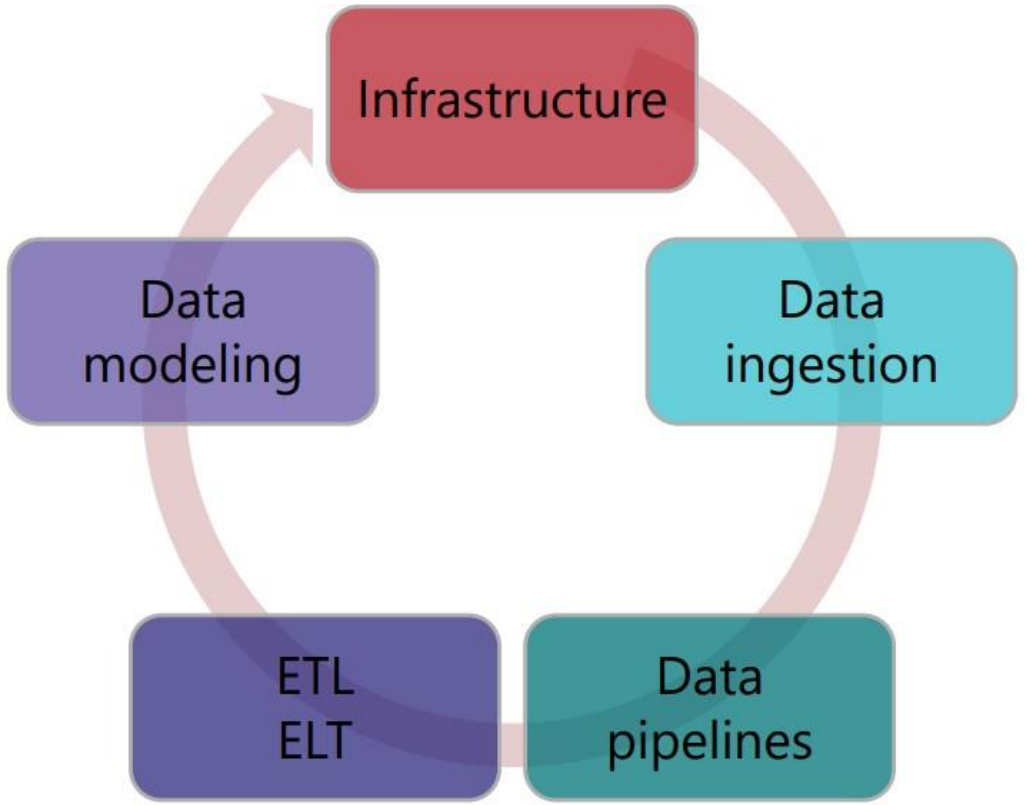
Data Science Hierarchy

Great Learning
POWER AHEAD

Behind every good data scientist is one or more data engineers!



Data Engineer Scope



Networking with AKS

Great Learning
POWER AHEAD

Cosmos DB

- NoSQL
- Document (JSON)
- Key-value
- Graph

SQL Database

- Relational
- Single / elastic pool
- Managed Instance
- Hyperscale

SQL Data Warehouse

- Multi-Parallel Processing (MPP)

Blob storage

- Unstructured data
- Flat files
- Media files

Data Lake Storage

- Unstructured data
- HDFS
- Petabytes of storage

Stream Analytics

- Serverless real-time analytics
- IoT devices
- Event Hubs

Data Factory

- ETL at scale
- 80+ connectors
- SSIS integration runtime

Databricks

- Big Data analytics platform
- Based on Apache Spark
- Notebooks

Case Study

WEB

In the healthcare industry, use Azure Databricks to accelerate big-data analytics and AI solutions. Apply these technologies to genome studies or pharmacy sales forecasting at a petabyte scale. Using Databricks features, you can set up your Spark environment in minutes and autoscale quickly and easily.
Using Azure, you can collaborate with data scientists on shared projects and workspaces in a wide range of languages, including SQL, R, Scala, and Python. Because of native integration with Azure Active Directory and other Azure services, you can build diverse solution types. For example, build a modern data warehouse or machine learning and real-time analytics solutions.

Healthcare

In the healthcare industry, use Azure Databricks to accelerate big-data analytics and AI solutions. Apply these technologies to genome studies or pharmacy sales forecasting at a petabyte scale. Using Databricks features, you can set up your Spark environment in minutes and autoscale quickly and easily.
Using Azure, you can collaborate with data scientists on shared projects and workspaces in a wide range of languages, including SQL, R, Scala, and Python. Because of native integration with Azure Active Directory and other Azure services, you can build diverse solution types. For example, build a modern data warehouse or machine learning and real-time analytics solutions.

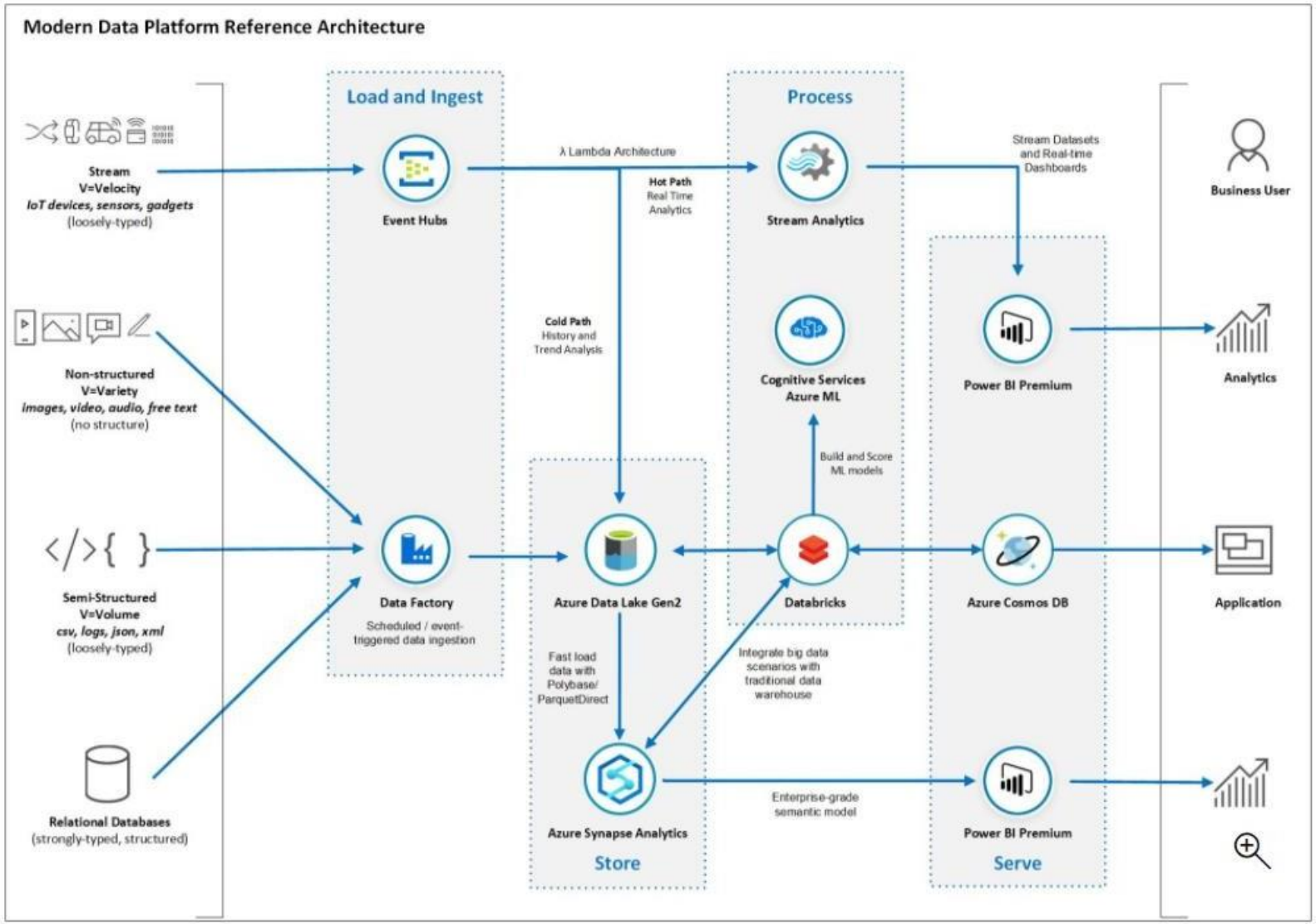
Case Study

IOT Solution

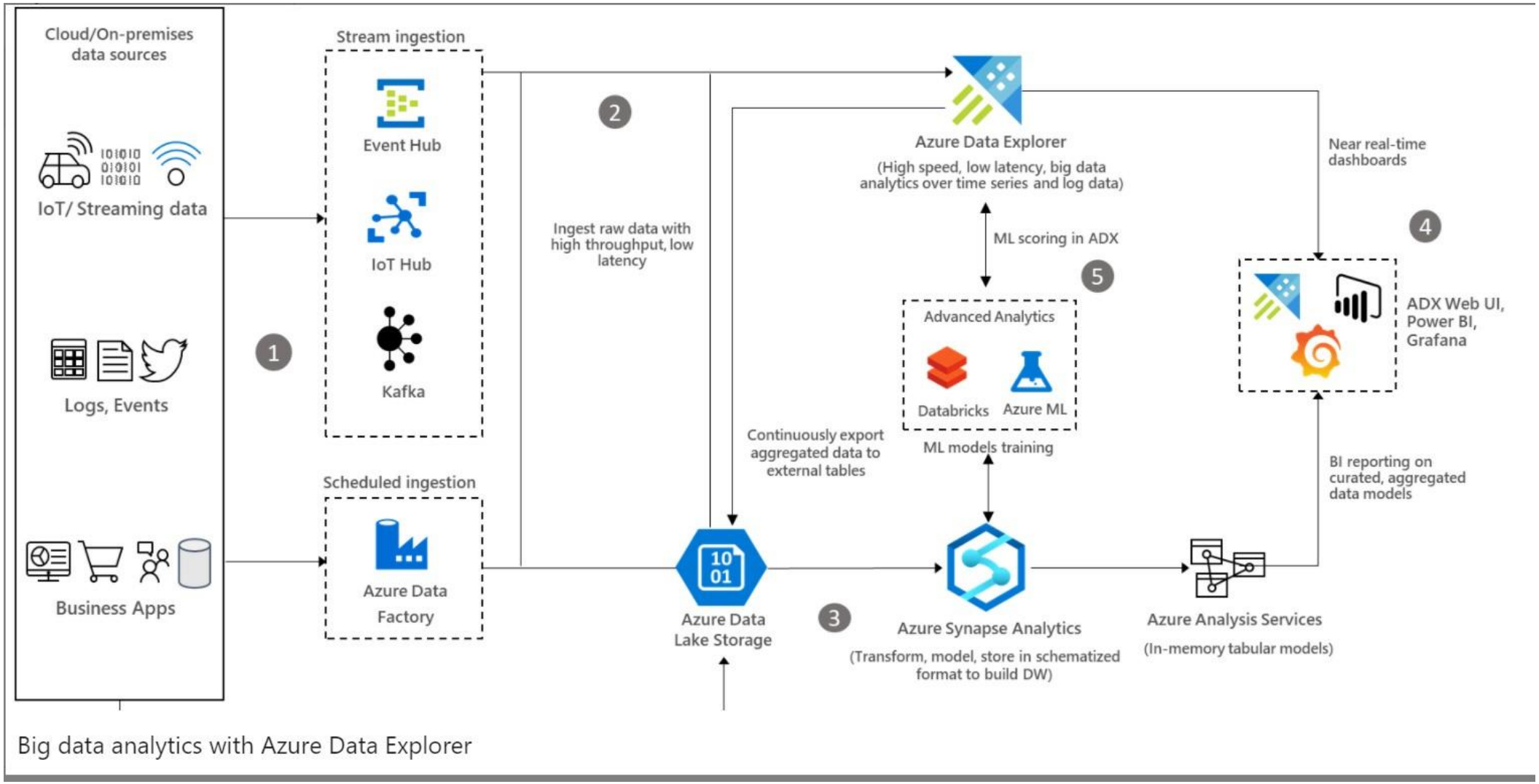
In the healthcare industry, use Azure Databricks to accelerate big-data analytics and AI solutions. Apply these technologies to genome studies or pharmacy sales forecasting at a petabyte scale. Using Databricks features, you can set up your Spark environment in minutes and autoscale quickly and easily.

Using Azure, you can collaborate with data scientists on shared projects and workspaces in a wide range of languages, including SQL, R, Scala, and Python. Because of native integration with Azure Active Directory and other Azure services, you can build diverse solution types. For example, build a modern data warehouse or machine learning and real-time analytics solutions.

Azure Kuberneets Service



Microservices with AKS



Choose a data storage approach in Azure

Great Learning
POWER AHEAD

[Data store decision tree - Azure Application Architecture Guide | Microsoft Docs]

Azure Storage Account

Storage account overview - Azure Storage | Microsoft Docs

Feature	Description	When to use
Azure Files	Offers fully managed cloud file shares that you can access from anywhere via the industry standard Server Message Block (SMB) protocol. You can mount Azure file shares from cloud or on-premises deployments of Windows, Linux, and macOS.	You want to "lift and shift" an application to the cloud that already uses the native file system APIs to share data between it and other applications running in Azure. You want to replace or supplement on-premises file servers or NAS devices. You want to store development and debugging tools that need to be accessed from many virtual machines.
Azure Blobs	Allows unstructured data to be stored and accessed at a massive scale in block blobs. Also supports Azure Data Lake Storage Gen2 for enterprise big data analytics solutions.	You want your application to support streaming and random access scenarios. You want to be able to access application data from anywhere. You want to build an enterprise data lake on Azure and perform big data analytics.

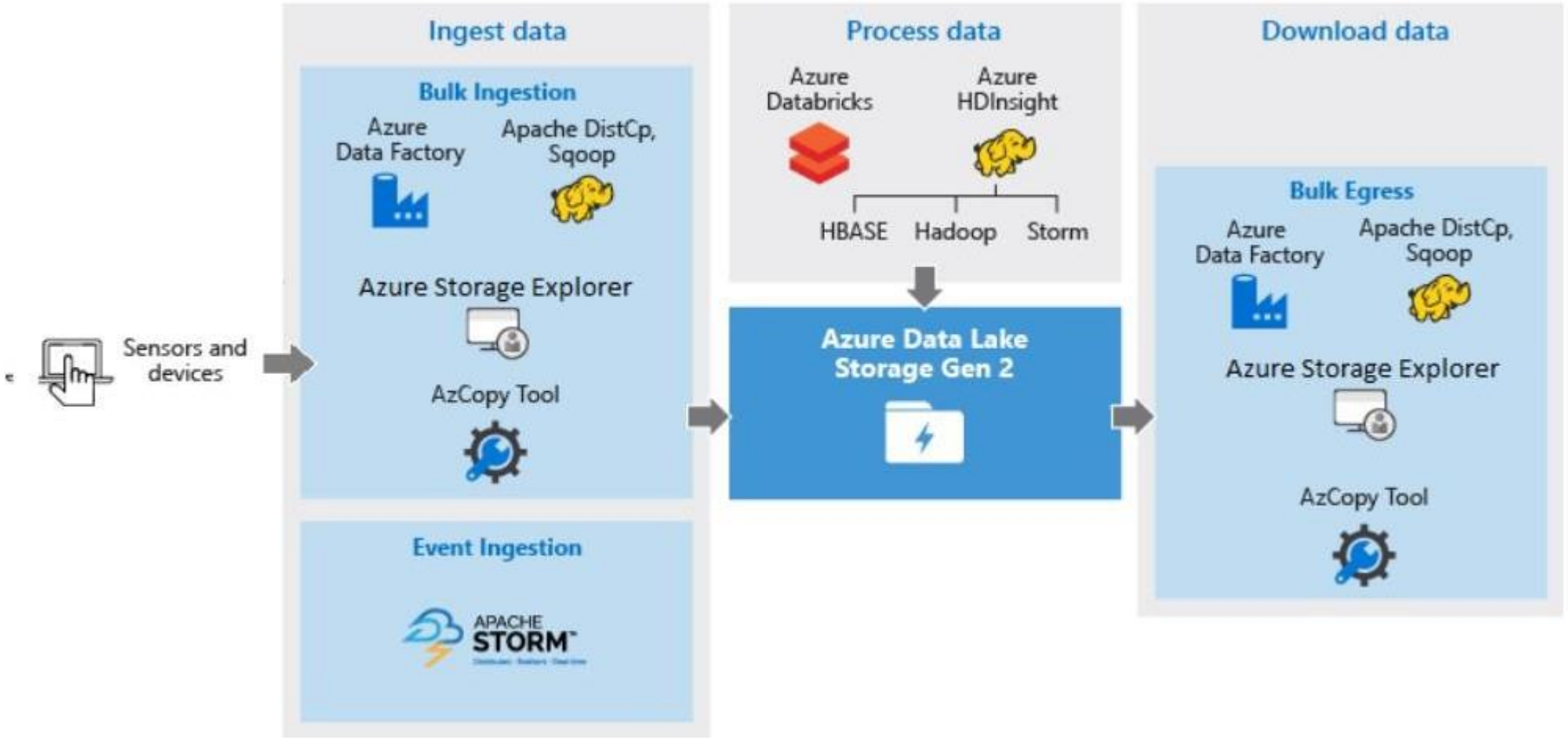
Azure Storage Account

Storage account overview - Azure Storage | Microsoft Docs

Feature	Description	When to use
Azure Disks	Allows data to be persistently stored and accessed from an attached virtual hard disk.	You want to "lift and shift" applications that use native file system APIs to read and write data to persistent disks. You want to store data that is not required to be accessed from outside the virtual machine to which the disk is attached.
Azure Queues	Allows for asynchronous message queueing between application components.	You want to decouple application components and use asynchronous messaging to communicate between them. For guidance around when to use Queue storage versus Service Bus queues, see Storage queues and Service Bus queues - compared and contrasted.
Azure Tables	Allow you to store structured NoSQL data in the cloud, providing a key/attribute store with a schemaless design.	You want to store flexible datasets like user data for web applications, address books, device information, or other types of metadata your service requires. For guidance around when to use Table storage versus the Azure Cosmos DB Table API, see Developing with Azure Cosmos DB Table API and Azure Table storage.

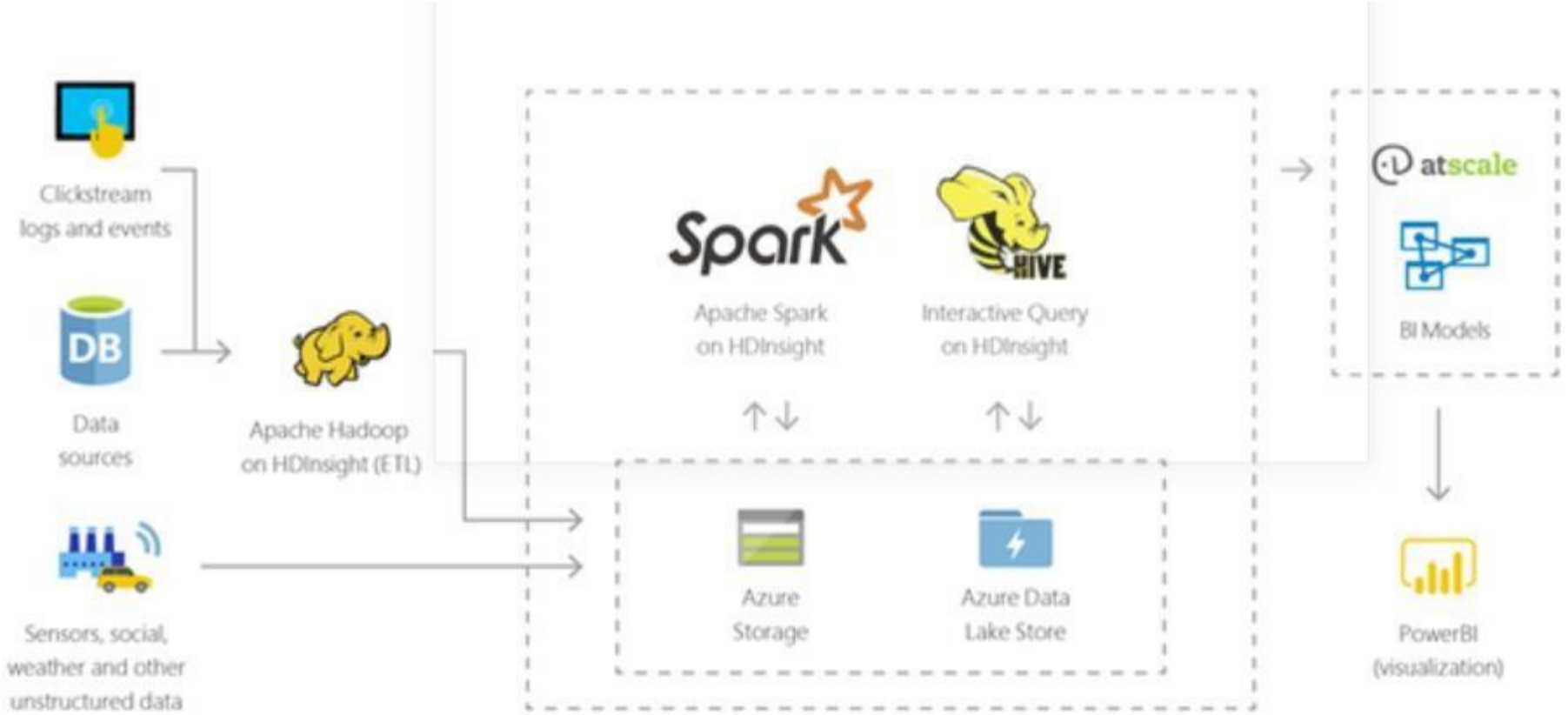
Azure Data Lake Store

https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-introduction



HDI Cluster

https://docs.microsoft.com/en-us/azure/hdinsight/



Great Learning
POWER AHEAD

THANK YOU