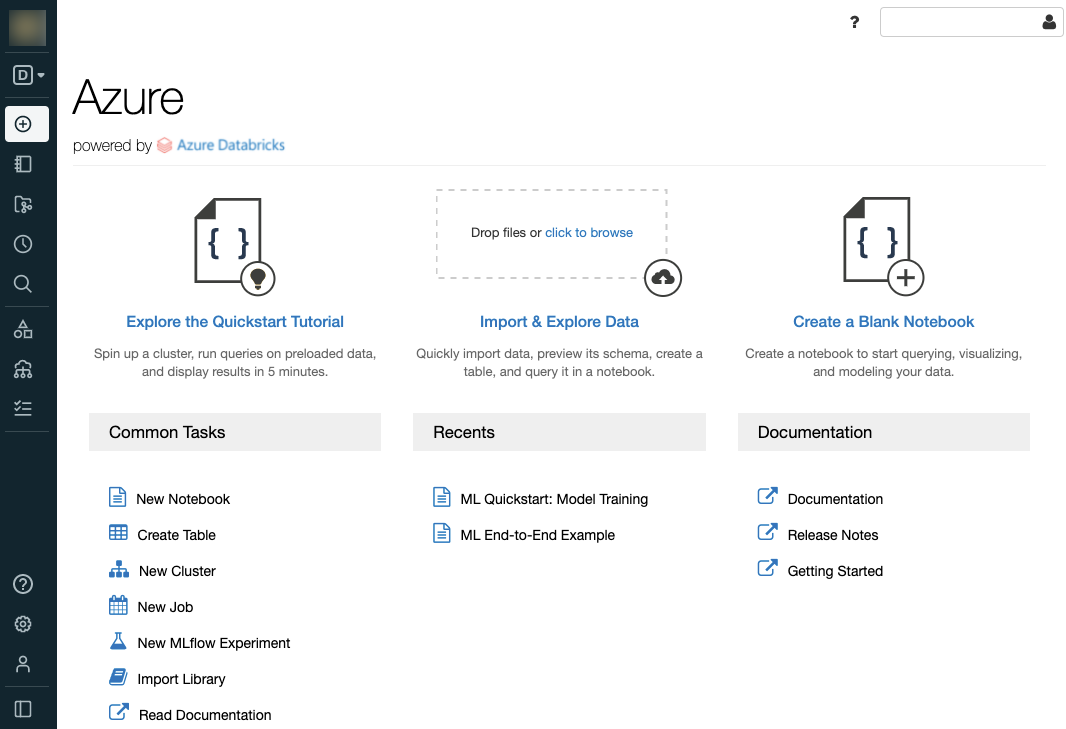
Azure Databricks

Azure Databricks provides the latest versions of Apache Spark and allows you to seamlessly integrate with open-source libraries. Spin up clusters and build quickly in a fully managed Apache Spark environment with the global scale and availability of Azure.

**Databricks SQL** provides an easy-to-use platform for analysts who want to run SQL queries on their data lake, create multiple visualization types to explore query results from different perspectives, and build and share dashboards.

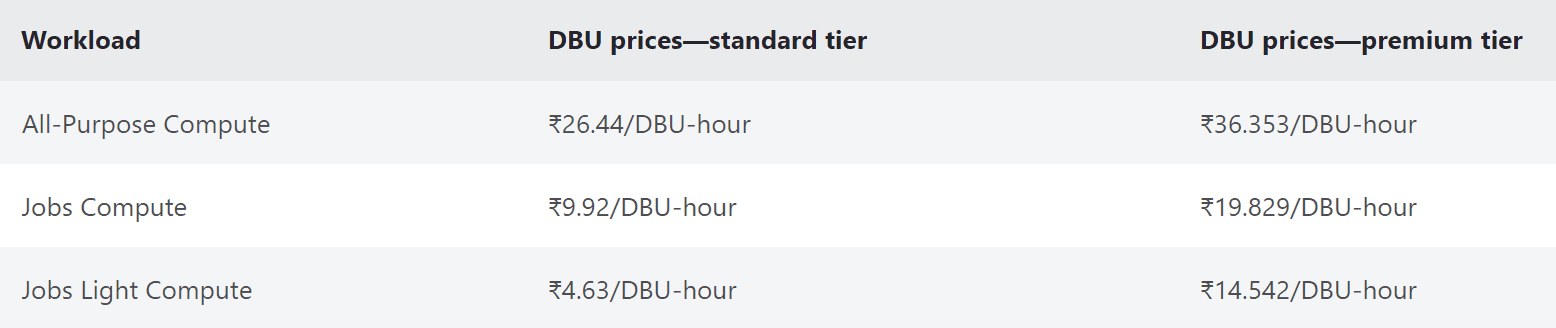
**Databricks Data Science & Engineering** provides an interactive workspace that enables collaboration between data engineers, data scientists, and machine learning engineers. For a big data pipeline, the data (raw or structured) is ingested into Azure through Azure Data Factory in batches, or streamed near real-time using Apache Kafka, Event Hub, or IoT Hub. This data lands in a data lake for long term persisted storage, in Azure Blob Storage or Azure Data Lake Storage. As part of your analytics workflow, use Azure Databricks to read data from multiple data sources and turn it into breakthrough insights using Spark.

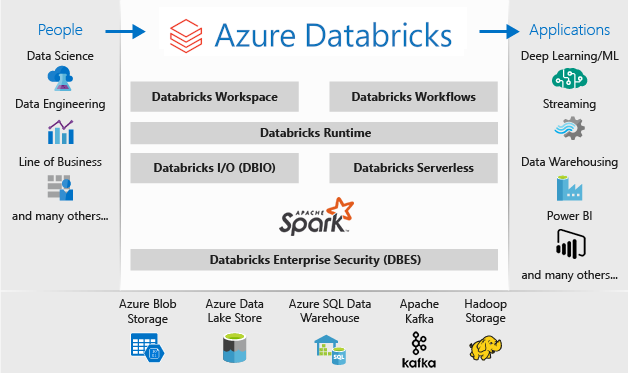


**Databricks Machine Learning** is an integrated end-to-end machine learning environment incorporating managed services for experiment tracking, model training, feature development and management, and feature and model serving.

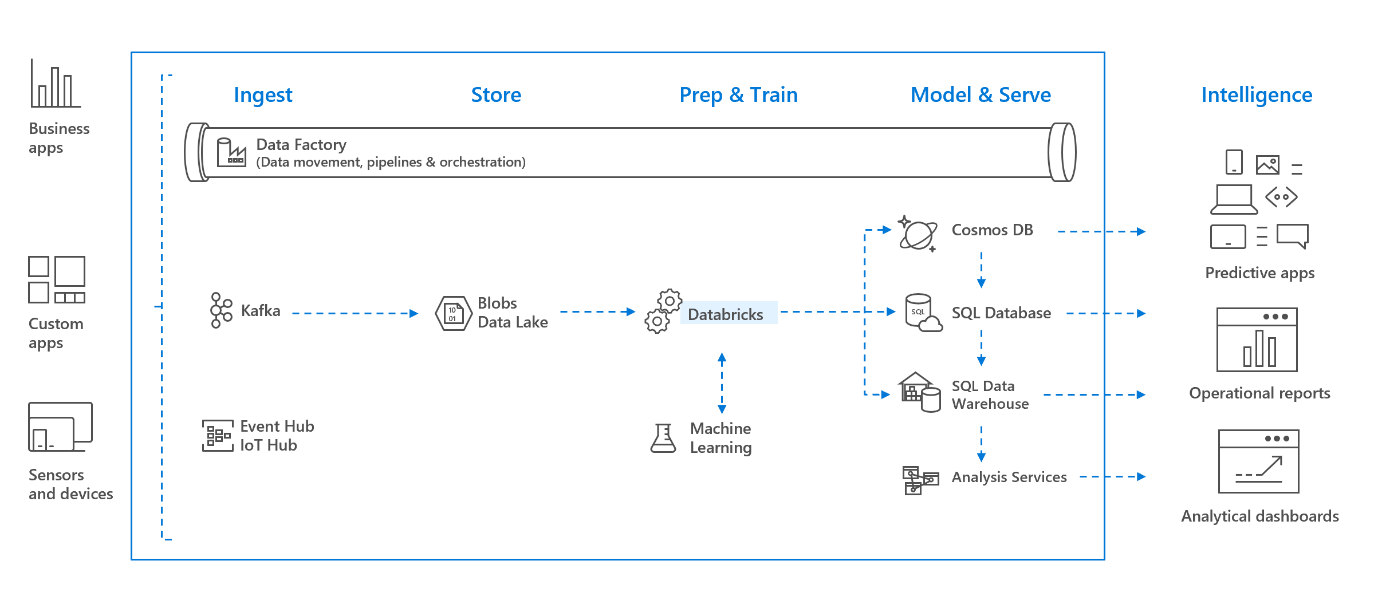
## **Azure Databricks Pricing**

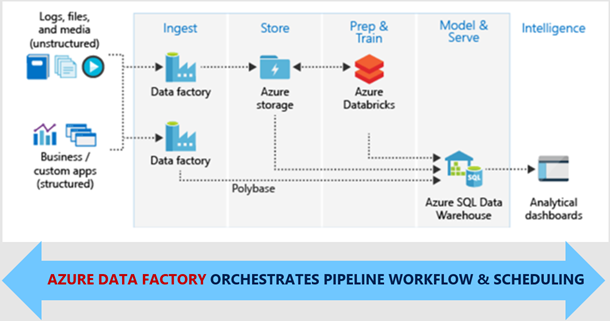
* **Pay as you go:**Azure Databricks cost you for virtual machines (VMs) manage in clusters and Databricks Units (DBUs) depend on the VM instance selected.
* A DBU is a unit of the processing facility, billed on per-second usage, and DBU consumption depends on the type and size of the instance running Databricks.





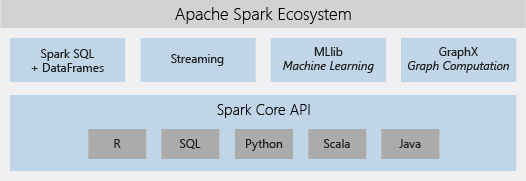
For a big data pipeline, the data (raw or structured) is ingested into Azure through Azure Data Factory in batches, or streamed near real-time using Apache Kafka, Event Hub, or IoT Hub. This data lands in a data lake for long term persisted storage, in Azure Blob Storage or Azure Data Lake Storage. As part of your analytics workflow, use Azure Databricks to read data from multiple data sources such as Azure Blob Storage, Azure Data Lake Storage, Azure Cosmos DB, or Azure SQL Data Warehouse and turn it into breakthrough insights using Spark.





Apache Spark analytics platform

**Databricks Data Science & Engineering comprises** the complete open-source Apache Spark cluster technologies and capabilities. Spark in Databricks Data Science & Engineering includes the following components:



## Apache Spark in Azure Databricks

Azure Databricks builds on the capabilities of Spark by providing a zero-management cloud platform that includes:

* Fully managed Spark clusters
* An interactive workspace for exploration and visualization
* A platform for powering your favourite Spark applications

## Fully managed Apache Spark clusters in the cloud

Azure Databricks has a secure and reliable production environment in the cloud, managed and supported by Spark experts. You can:

* Create clusters in seconds.
* Dynamically auto scale clusters up and down and share them across teams.
* Use clusters programmatically by invoking REST APIs.
* Use secure data integration capabilities built on top of Spark that enable you to unify your data without centralization.
* Get instant access to the latest Apache Spark features with each release.

## Enterprise security

Azure Databricks provides enterprise-grade Azure security, including Azure Active Directory integration, role-based controls, and SLAs that protect your data and your business.

* Integration with Azure Active Directory enables you to run complete Azure-based solutions using Azure Databricks.
* Azure Databricks roles-based access enables fine-grained user permissions for notebooks, clusters, jobs, and data.
* Enterprise-grade SLAs.

## Integration with Azure services

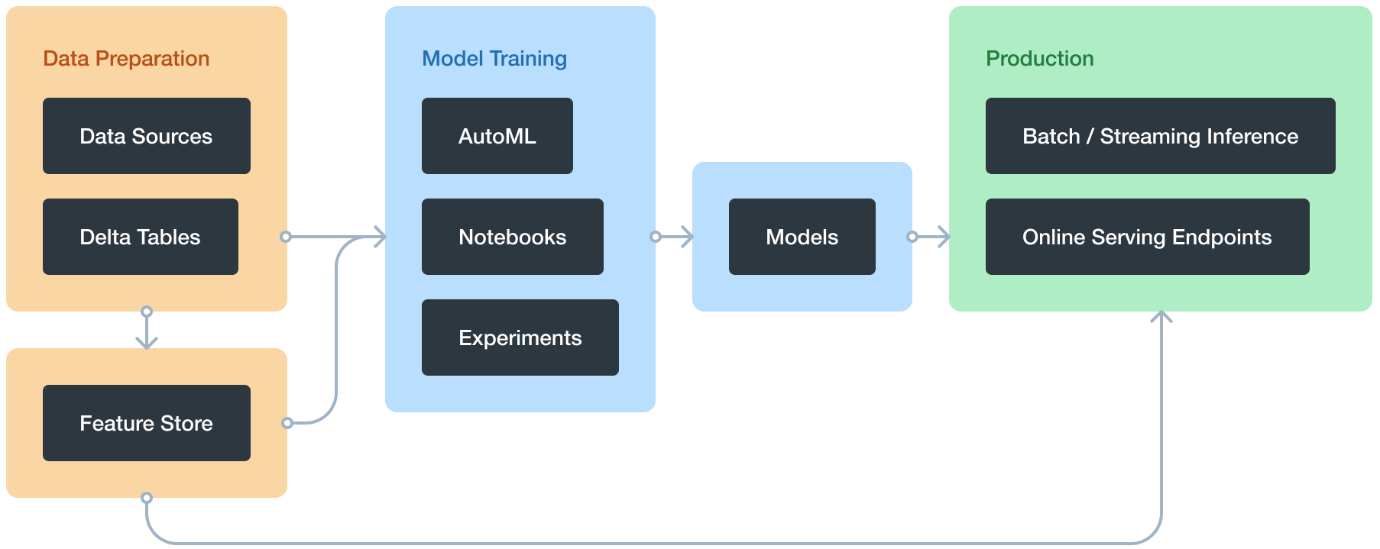
Databricks Data Science & Engineering integrates deeply with Azure databases and stores: Synapse Analytics, Cosmos DB, Data Lake Store, and Blob storage.

## Integration with Power BI

Through rich integration with Power BI, Databricks Data Science & Engineering allows you to discover and share your impactful insights quickly and easily. You can use other BI tools as well, such as Tableau Software.

# What is Databricks Machine Learning?

Databricks Machine Learning (Preview) is an integrated end-to-end machine learning platform incorporating managed services for experiment tracking, model training, feature development and management, and feature and model serving. The diagram shows how the capabilities of Databricks map to the steps of the model development and deployment process.

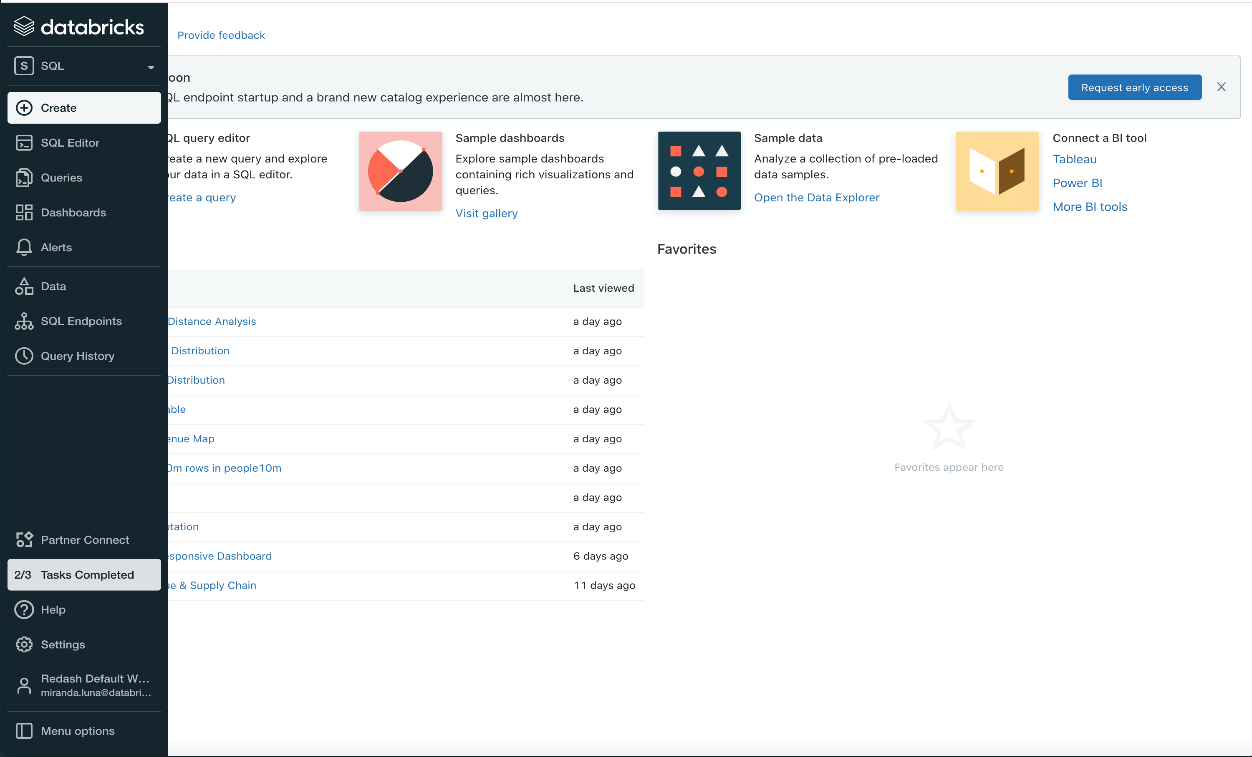


## With Databricks Machine Learning, you can:

1. Train models either manually or with AutoML.
2. Track training parameters and models using experiments with MLflow tracking.
3. Create feature tables and access them for model training and inference.
4. Share, manage, and serve models using Model Registry.
5. For machine learning applications, Databricks provides Databricks Runtime for Machine Learning, a variation of Databricks Runtime that includes many popular machine learning libraries.

# What is Databricks SQL?

Databricks SQL allows you to run quick ad-hoc SQL queries on your data lake. Queries support multiple visualization types to help you explore your query results from different perspectives.



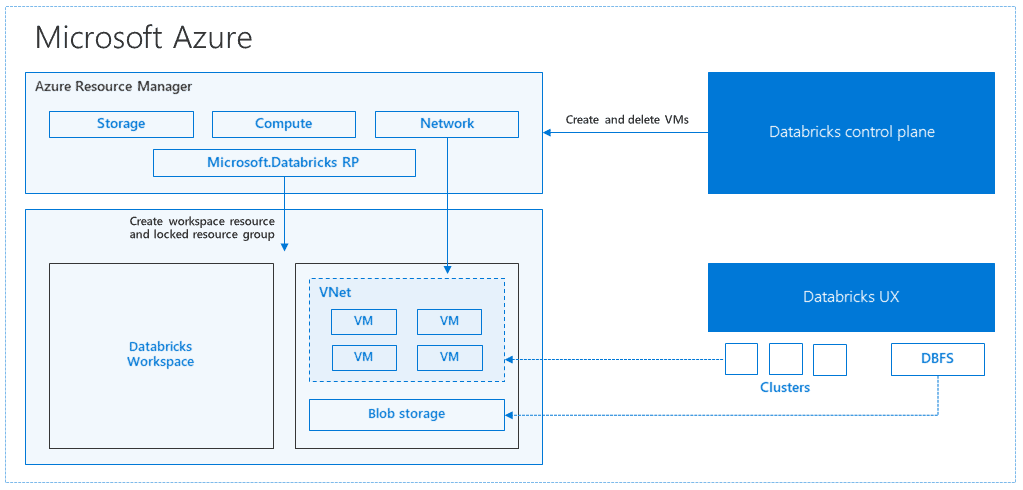
## Azure Databricks architecture

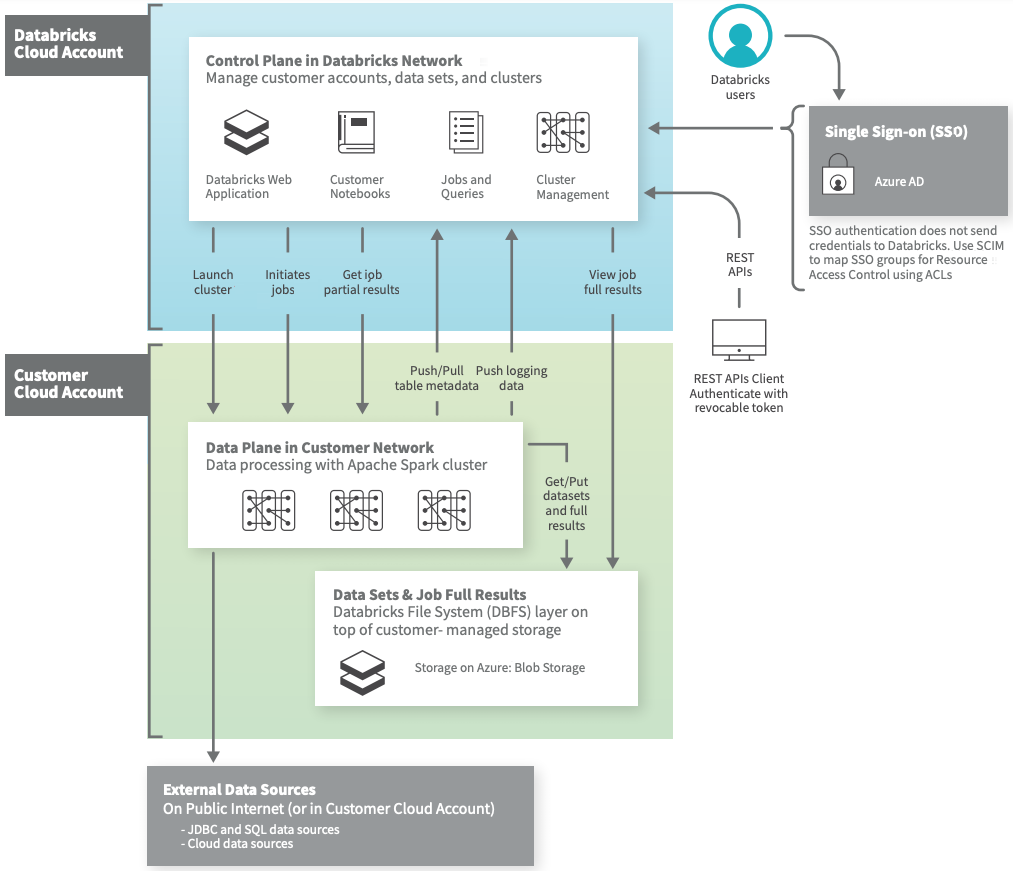
## High-level architecture

Azure Databricks is structured to enable secure cross-functional team collaboration while keeping a significant amount of backend services managed by Azure Databricks so you can stay focused on your data science, data analytics, and data engineering tasks.

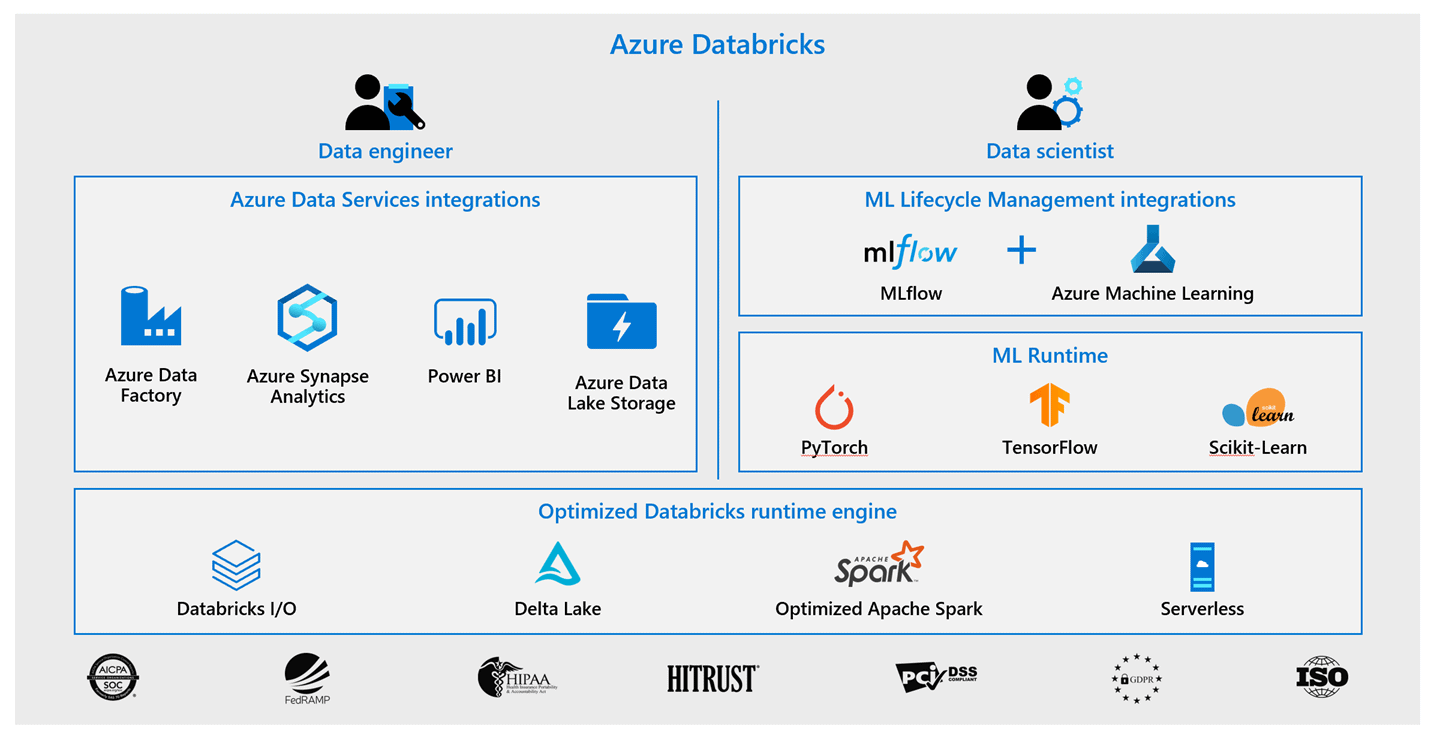
Azure Databricks operates out of a control plane and a data plane.

* The control plane includes the backend services that Azure Databricks manages in its own Azure account. Notebook commands and many other workspace configurations are stored in the control plane and encrypted at rest.
* The data plane is managed by your Azure account and is where your data resides. This is also where data is processed. You can use Azure Databricks connectors so that your clusters can connect to external data sources outside of your Azure account to ingest data or for storage. You can also ingest data from external streaming data sources, such as events data, streaming data, IoT data, and more.





The most common structure and flow of data for Azure Databricks.



## Real-Time Use Cases of Azure Databricks

* As mobile apps and other advances in technology continue to upgrade the way users choose and utilize information, recommendation engines are becoming an essential part of applications and software products.
* Churn analysis also known as customer defection, customer attrition, or customer turnover, is the loss of clients or customers. Forecasting and restricting customer churn are vital to a range of businesses.
* Intrusion detection is required to track network or system activities for malicious activities or policy violations and generate electronic reports to a management station.