

# **AZ-305T00A**

## **Designing Microsoft Azure Infrastructure Solutions**

# Module 05: Design a data integration solution



# Introduction

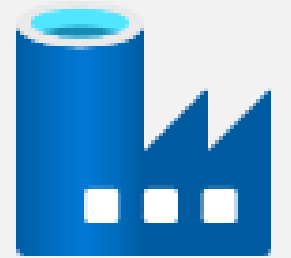
- Design a data integration solution with Azure Data Factory
- Design a data integration solution with Azure Data Lake
- Design a data integration and analytics solution with Azure Databricks
- Design a data integration and analytics solution with Azure Synapse Analytics
- Design Azure Stream Analytics solution for Data Analysis

## AZ-305: Design Data Storage Solutions (25-30%)

### Design Data Integration

- Recommend a solution for data integration
- Recommend a solution for data analysis

# Design a data integration solution with Azure Data Factory

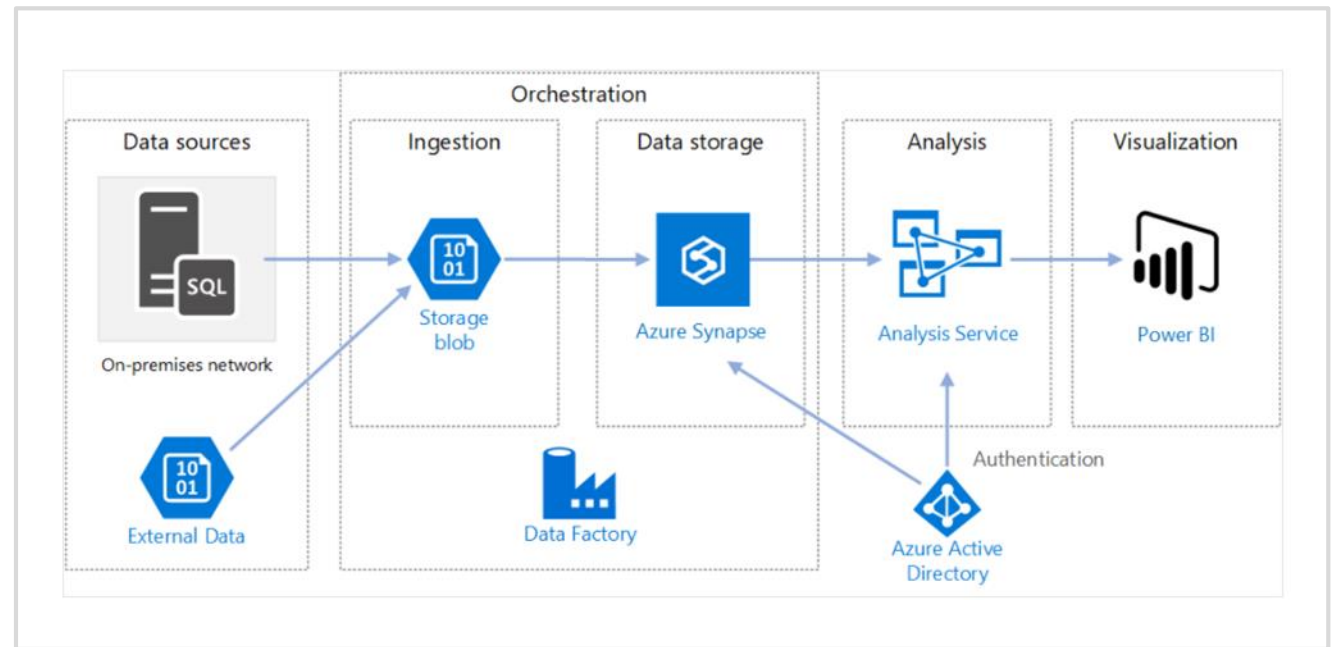


# Data-driven workflows

Azure Data Factory is a cloud-based ETL and data integration service that can help you create and schedule data-driven workflows (called pipelines) that can ingest data from disparate data stores.

You can use Azure Data Factory to:

1. Orchestrate data movement.
2. Transform data at scale.



# Design a data integration solution with Azure Data Lake

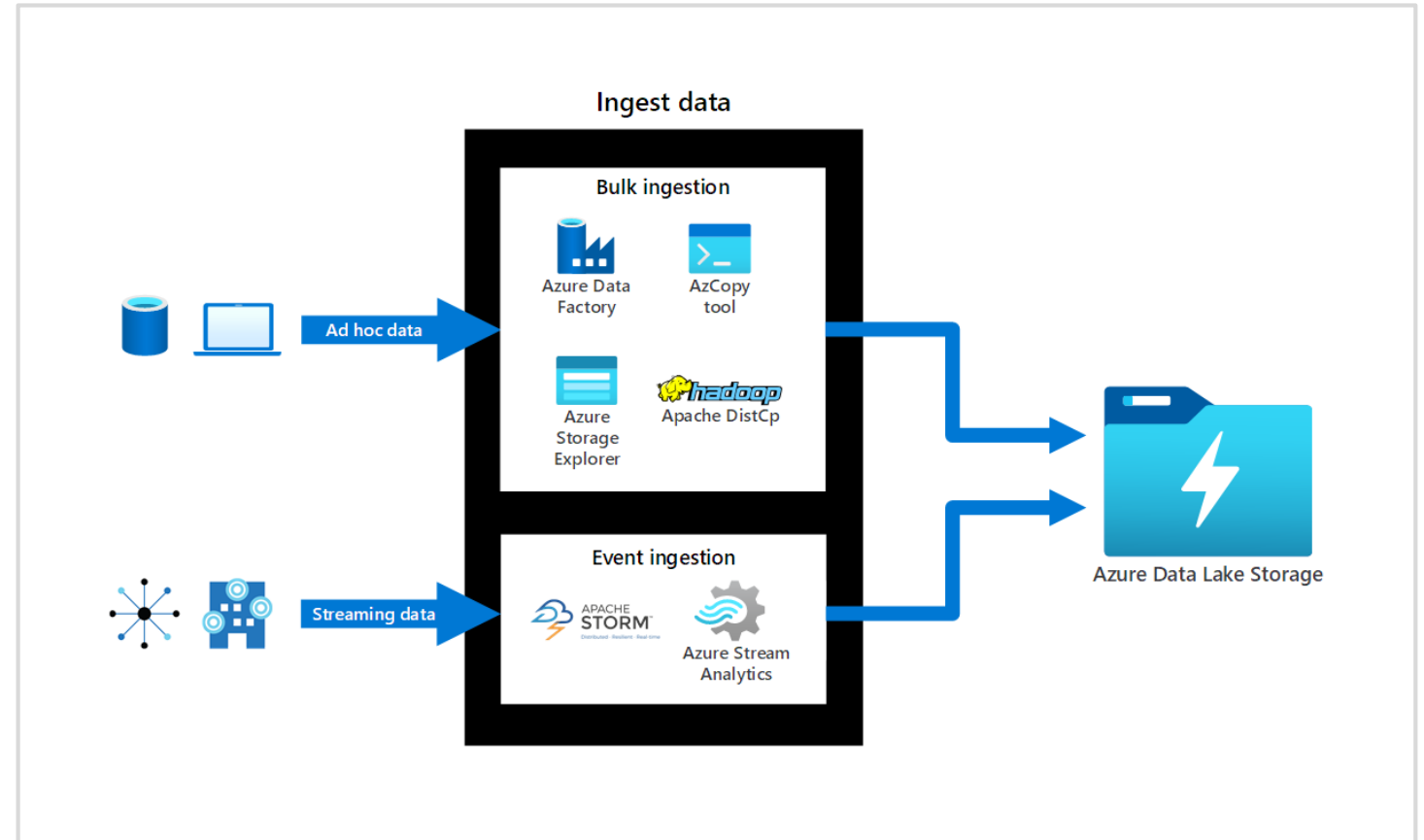


# Azure Data Lake

Azure Data Lake Storage is a comprehensive, scalable, and cost-effective data lake solution for big data analytics built into Azure.

Use Azure Data Lake when you need:

- a data repository on the cloud for managing large volumes of data
- To manage a diverse collection of data types such as JSON files, CSV, log files or other diverse formats
- Real-time data ingestion and storage



# Compare Azure Data Lake to Azure Blob storage

Criteria	Azure Data Lake	Azure Blob Storage
Data type	Good for storing large volumes of text data	Good for storing unstructured non-text-based data such as photos, videos, backup etc.
Namespace support	Supports hierarchical namespaces	Supports flat namespaces
Security	Access Control Lists (ACLs), shared keys, SAS and RBAC	Shared keys, SAS, and RBAC

[Blob Storage feature support in Azure storage accounts | Microsoft Docs](#)



# Design a data integration and analytics solution with Azure Databricks



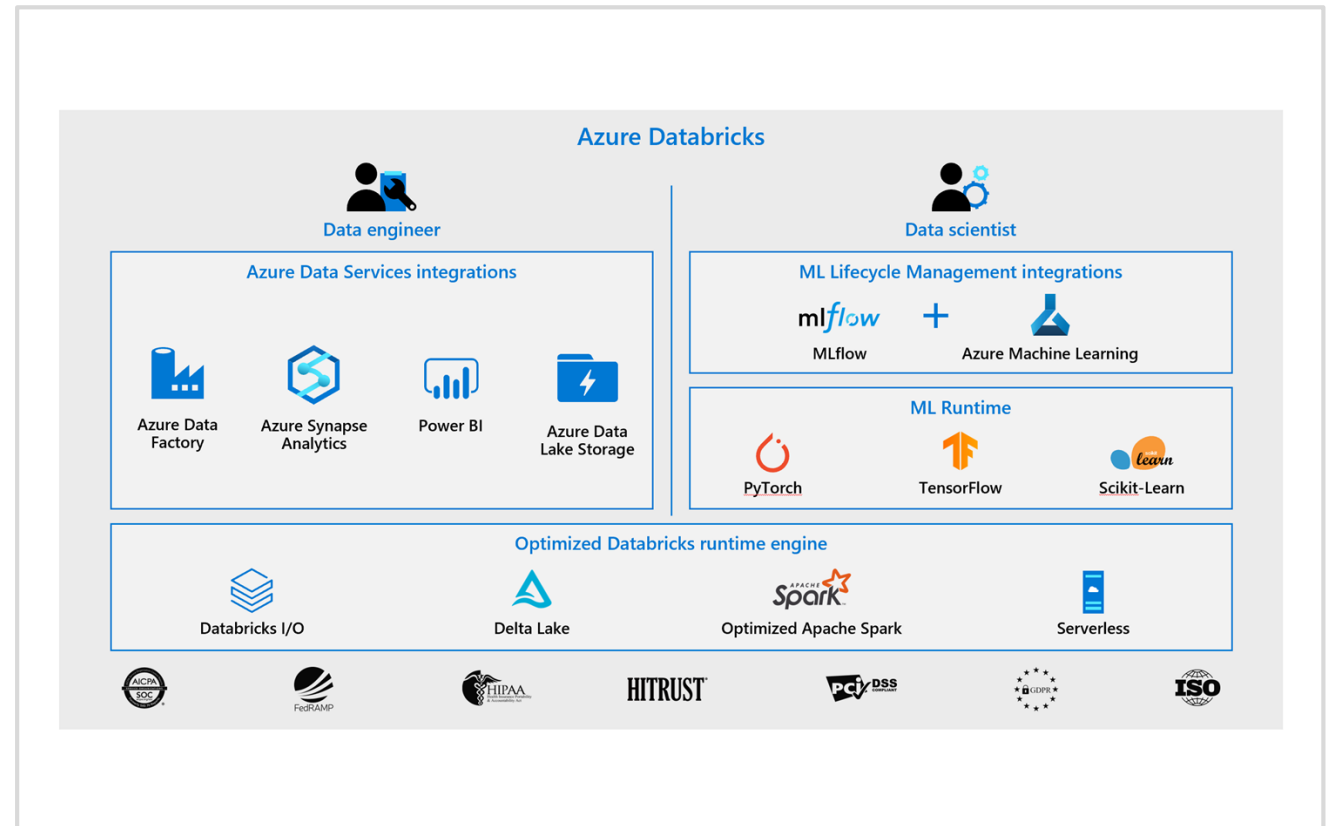
# Azure Databricks

Azure Databricks is a fully managed, cloud-based Big Data and Machine Learning platform, which empowers developers to accelerate AI and innovation.

Provides data science and engineering teams with a single platform for Big Data processing and Machine Learning.

Offers three environments for developing data intensive applications:

- Databricks SQL
- Databricks Data Science & Engineering
- Databricks Machine Learning.



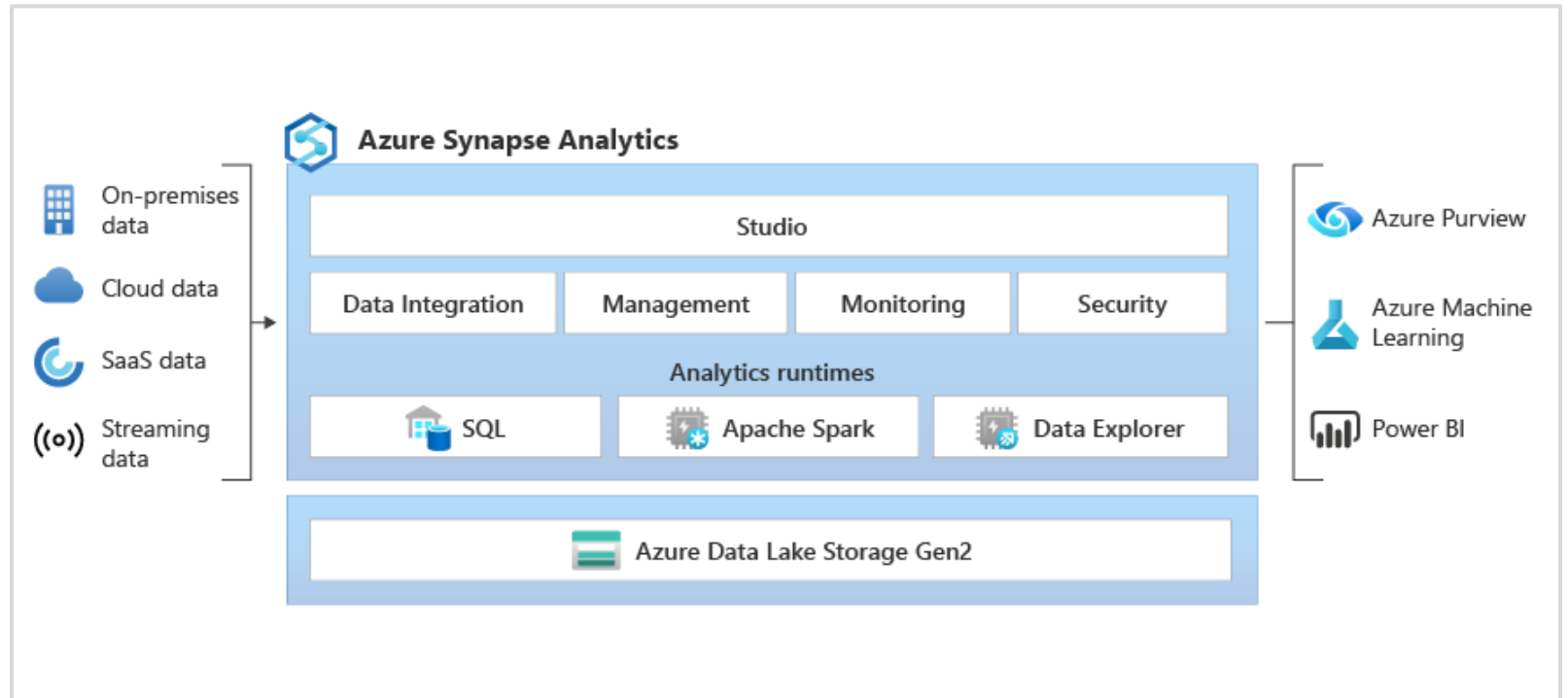
# Design a data integration and analytics solution with Azure Synapse Analytics



# Azure Synapse Analytics

Azure Synapse Analytics is an integrated analytics platform that brings together data integration, enterprise data warehousing, big data analytics and visualization into a single service. Azure Synapse Analytics is an evolution of Azure SQL Data Warehouse.

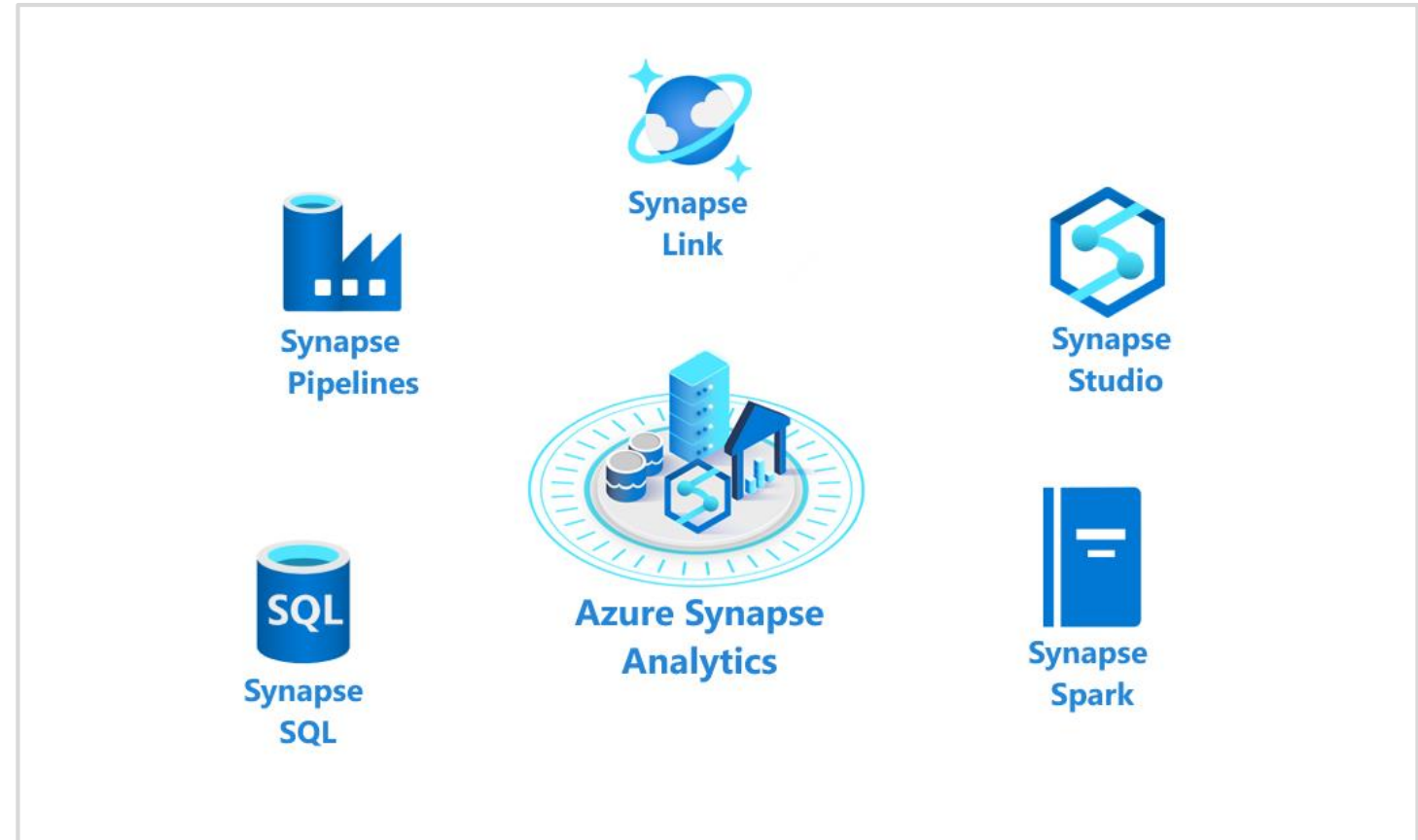
- Modern data warehousing
- Advanced analytics
- Data exploration and discovery
- Real time analytics
- Data integration
- Integrated analytics
- Machine Learning



# Components of Azure Synapse Analytics

## Elements

- Synapse SQL pool
- Synapse Spark pool
- Synapse Pipelines
- Synapse Link
- Synapse Studio

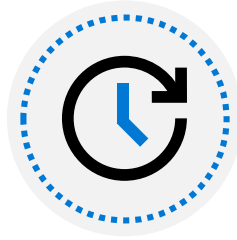


# What kind of analytics can you do with Azure Synapse Analytics?



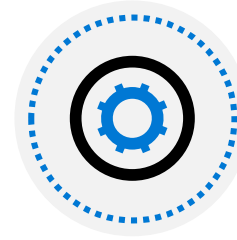
## **Descriptive analytics - "What is happening?"**

Azure Synapse Analytics leverages the dedicated SQL pool capability that enables you to create a persisted data warehouse to perform this type of analysis.



## **Diagnostic analytics - "Why is it happening?"**

You can use the serverless SQL pool capability within Azure Synapse Analytics that enables you to interactively explore data within a data lake.



## **Predictive analytics - "What is likely to happen?"**

Azure Synapse Analytics uses its integrated Apache Spark engine and Azure Synapse Spark pools for predictive analytics with other services such as Azure Machine Learning Services, or Azure Databricks.



## **Prescriptive analytics - "What needs to be done?"**

This type of analytics looks at executing actions based on real-time or near real-time analysis of data, using predictive analytics.

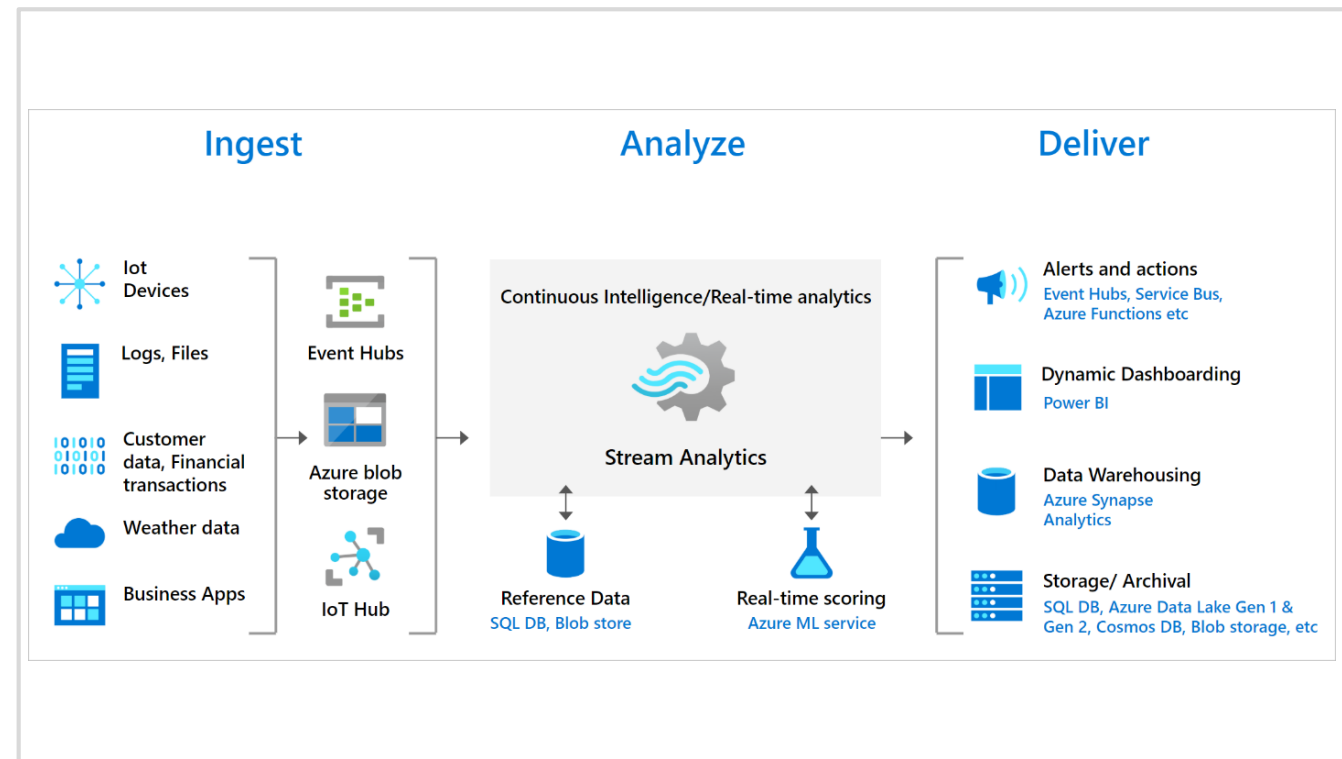
# Design Azure Stream Analytics solution for Data Analysis



# Azure Stream Analytics

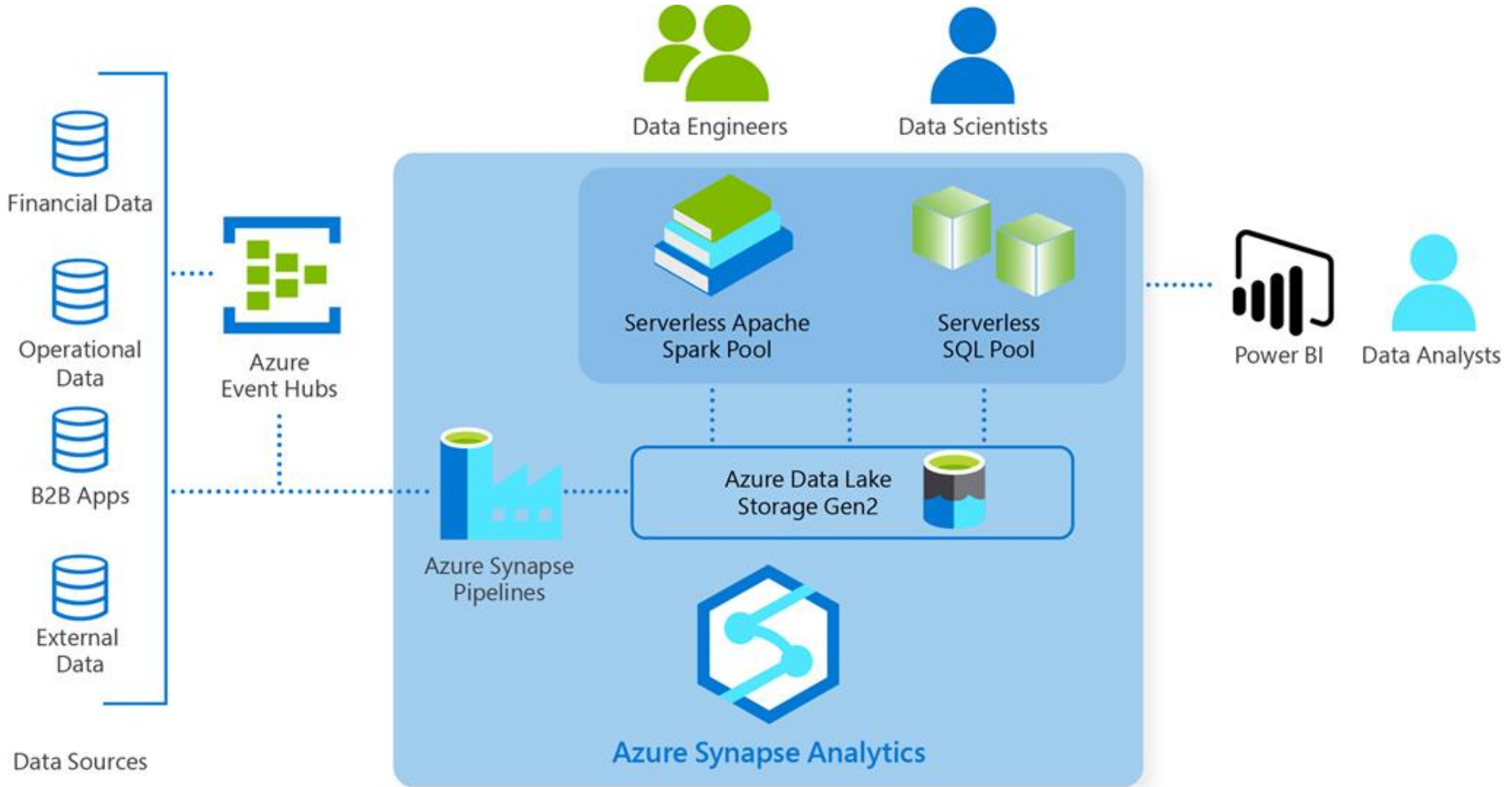
Azure Stream Analytics is a real-time analytics and complex event-processing engine that is designed to analyze and process high volumes of fast streaming data from multiple sources simultaneously.

- Analyze real-time telemetry streams from IoT devices
- Web logs/clickstream analytics
- Geospatial analytics for fleet management and driverless vehicles
- Remote monitoring and predictive maintenance of high value assets
- Real-time analytics on point-of-sale data for inventory control and anomaly detection

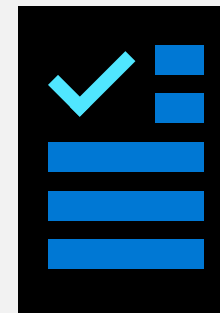




# Use Case: Just-in-time inventory



# Review



# Summary and resources

Check your knowledge



Microsoft Learn Modules ([docs.microsoft.com/Learn](https://docs.microsoft.com/Learn))

[Explore concepts of data analytics - Learn | Microsoft Docs](#)

---

[Explore Azure database and analytics services - Learn | Microsoft Docs](#)

---

[Data integration at scale Azure Data Factory - Learn | Microsoft Docs](#)

---

# End of presentation

