



# AZ-305T00A

## Designing Microsoft Azure Infrastructure Solutions



# Design a data integration solution



# Learning Objectives

- 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
- Case study
- Learning recap

AZ-305: Design Data Storage Solutions  
(20-25%)

## Design Data Integration

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

# Design a data integration solution with Azure Data Lake

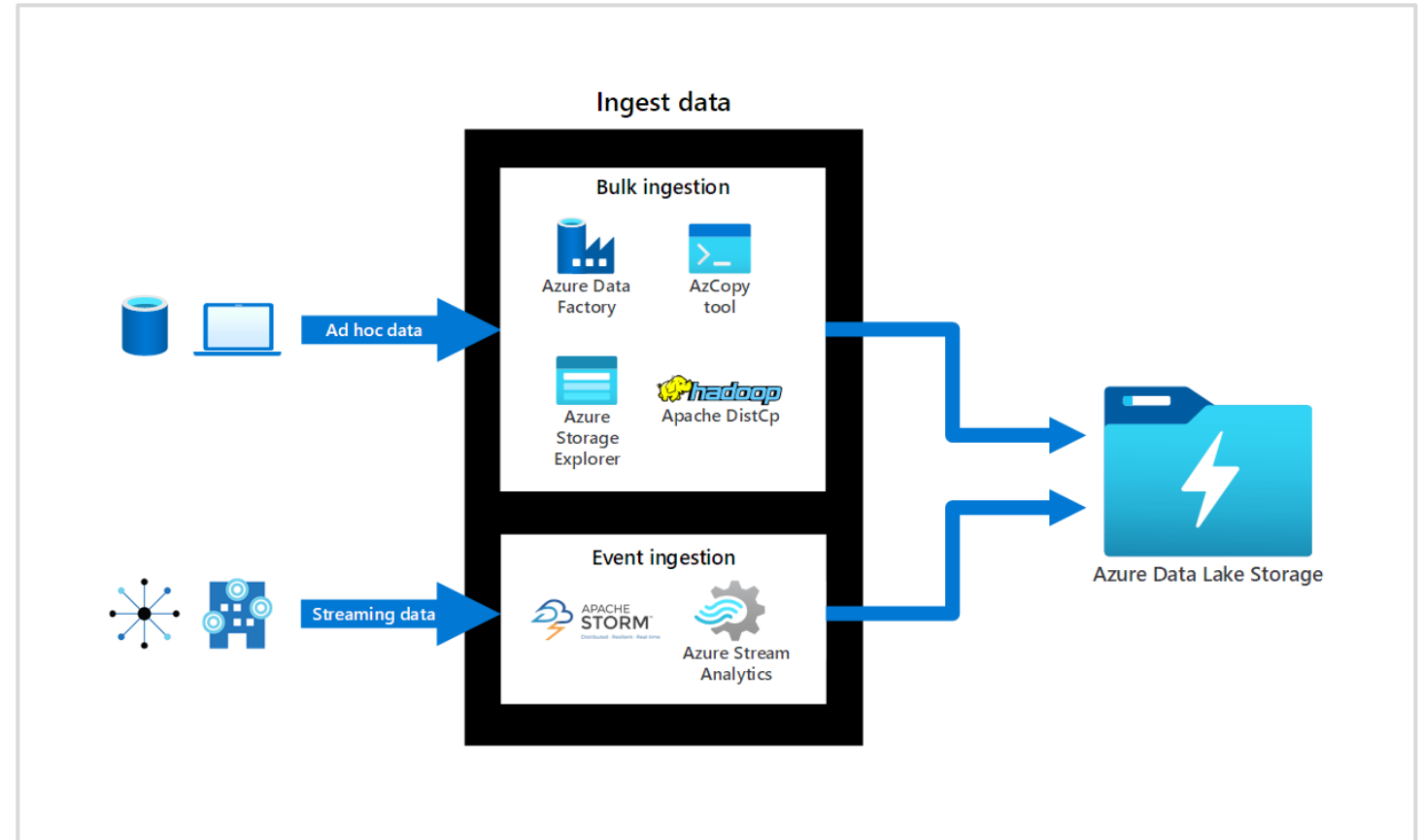


# Azure Data Lake

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
Hadoop compatibility	Optimized for big data, like Hadoop	Is not Hadoop compatible
Security	Access Control Lists (ACLs), shared keys, SAS and RBAC	Shared keys, SAS, and RBAC



# Design a data integration solution with Azure Data Factory

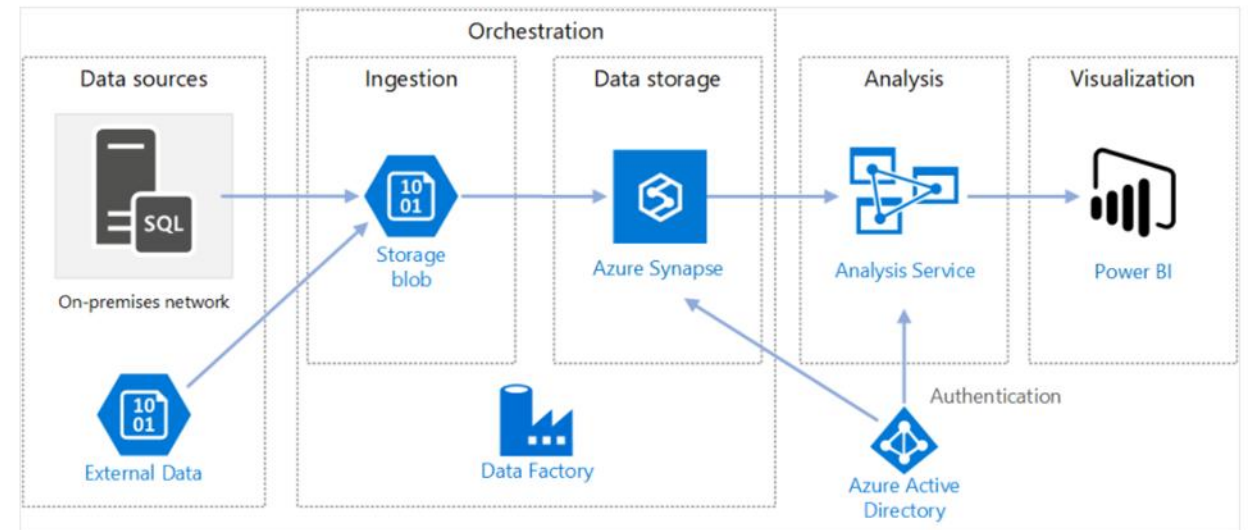


# Data-driven workflows

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 and analytics solution with Azure Databricks



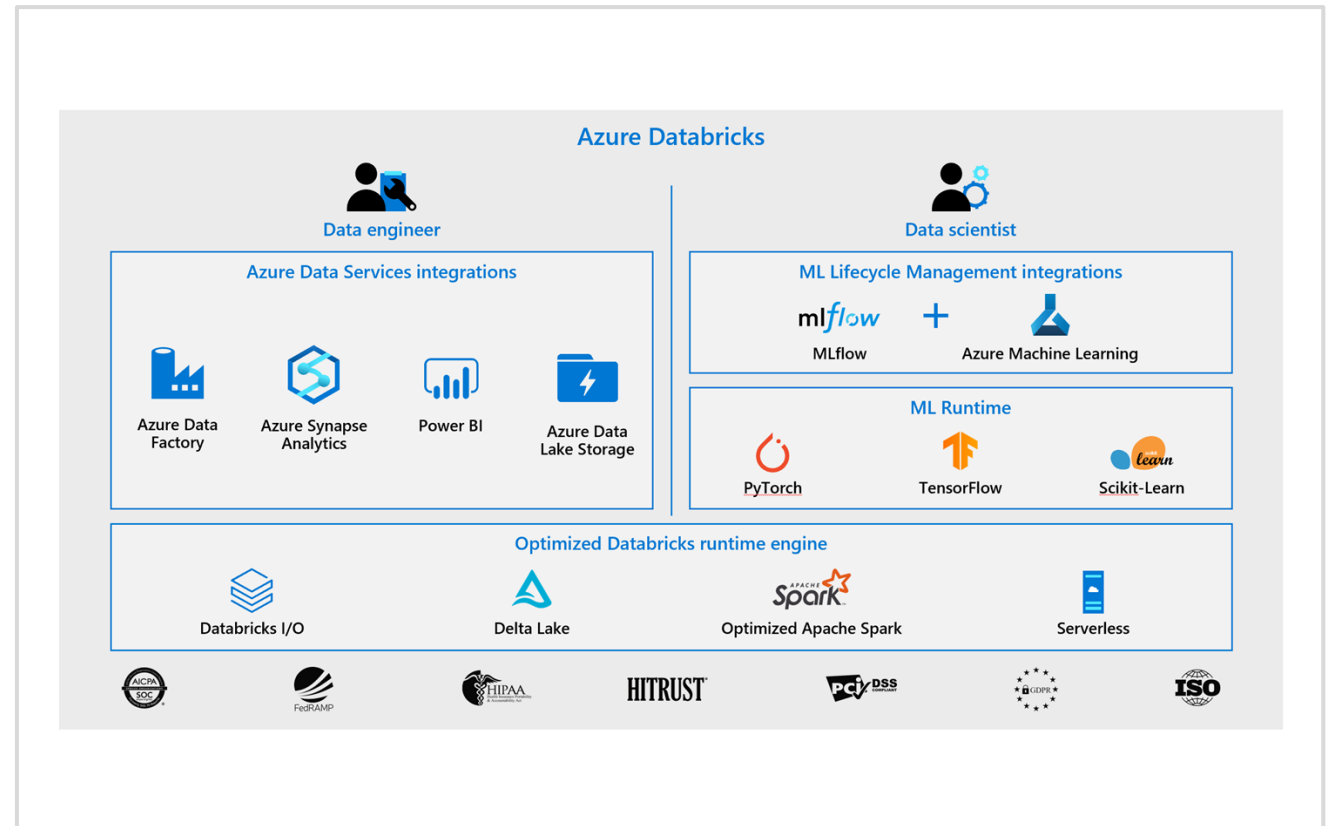
# Azure Databricks

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.



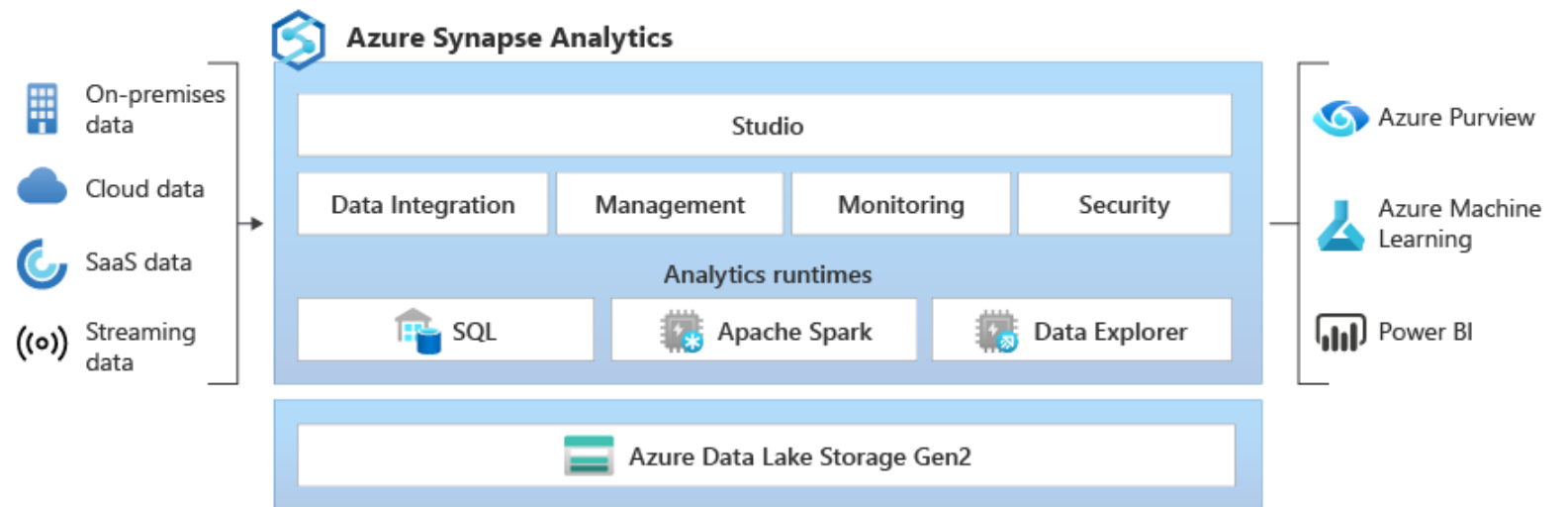
# Design a data integration and analytics solution with Azure Synapse Analytics



# 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.

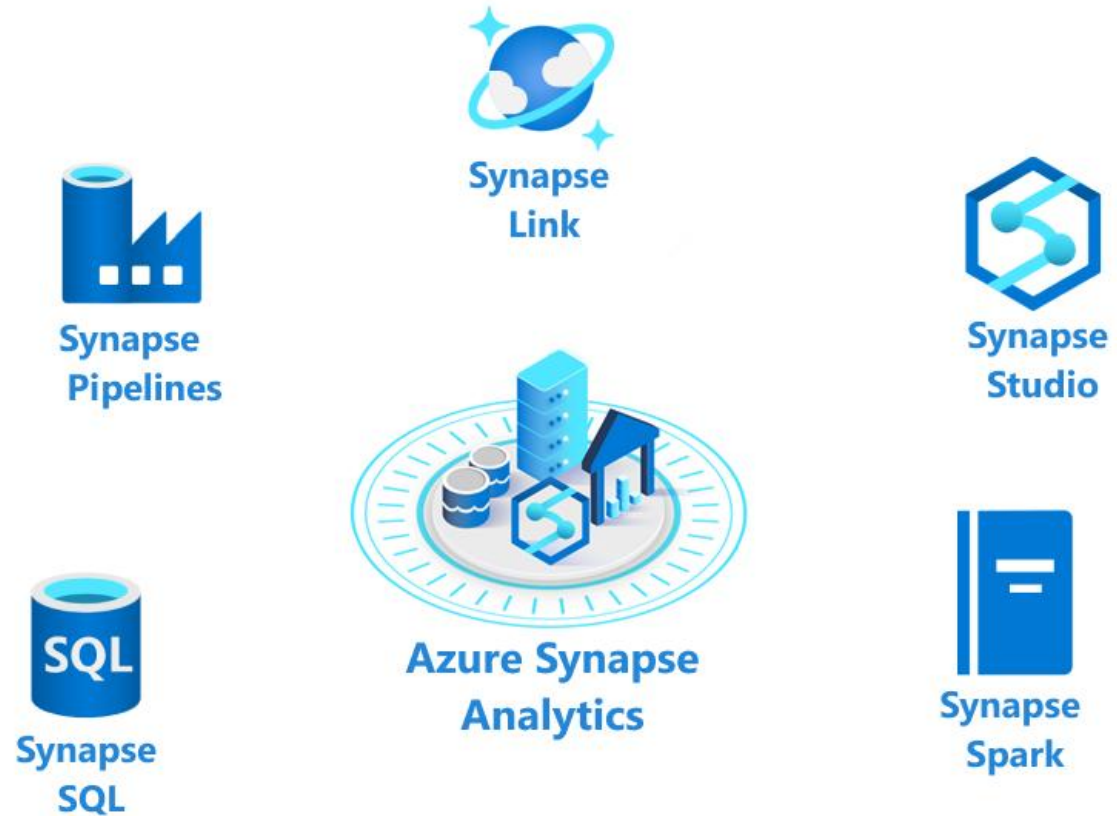
- Modern data warehousing
- Advanced analytics
- Data exploration and discovery
- Real time analytics
- Data integration
- Integrated analytics
- Machine Learning



# 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?



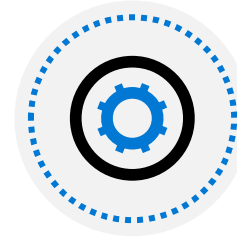
## **Descriptive analytics - "What is happening?"**

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



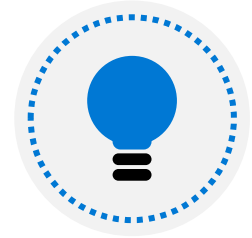
## **Diagnostic analytics - "Why is it happening?"**

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.



# Compare Azure Data Factory to Azure Synapse Analytics

Criteria	Azure Data Factory	Azure Synapse Analytics
Integration runtime sharing	Can be shared across different data factories	No sharing
Solution templates	Provided with Azure Data Factory template gallery	Provided with Synapse Workspace Knowledge center
Integration Runtime cross region support	Support Cross region data flows	Does not support cross region data flows
Monitoring of Spark Jobs for Data Flow	Not supported	Supported by the Synapse Spark pools

# Compare Synapse to Databricks

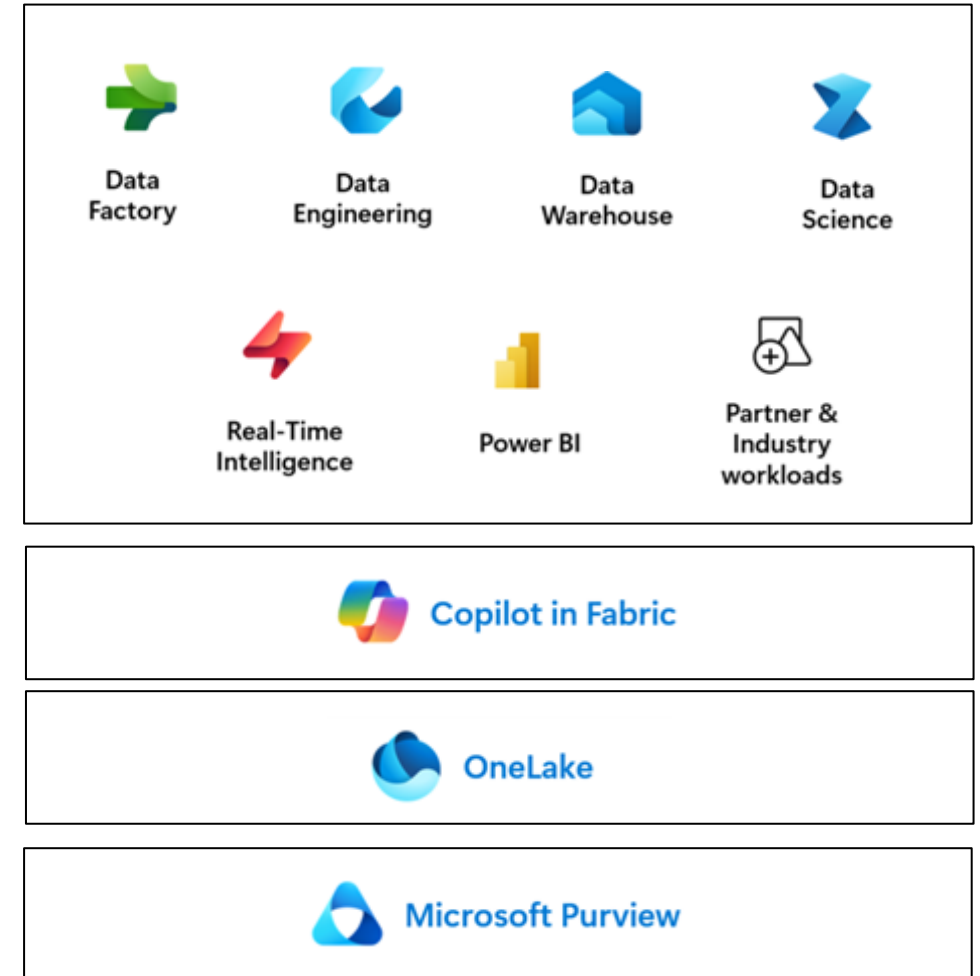
Azure Synapse Analytics and Azure Databricks offer different capabilities which may be combined if required

Capabilities	Databricks	Synapse
Machine Learning	<ul style="list-style-type: none"><li>Optimized runtimes with support for TensorFlow, PyTorch, and Keras.</li><li>GPU support</li></ul>	<ul style="list-style-type: none"><li>Built-in support Azure ML</li><li>Train models using SparkML, MLlib and other open-source libraries</li><li>GPU-accelerated pools</li></ul>
Feature Set	<ul style="list-style-type: none"><li>Optimized Apache Spark environment</li></ul>	<ul style="list-style-type: none"><li>Distributed T-SQL system</li><li>Spark environment</li><li>Data Integration</li><li>Unified experience with Synapse Studio</li></ul>
Reporting	<ul style="list-style-type: none"><li>Azure Databricks connection available in PowerBI</li></ul>	<ul style="list-style-type: none"><li>PowerBI available directly from Synapse Studio</li></ul>

# Microsoft Fabric

## End-to-end analytics and data platform to provide a unified solution

- Combines components from Power BI, Azure Synapse Analytics, Azure Data Factory, and more into a single environment
- Extensive data integration capabilities, including data movement, processing, ingestion, and transformation
- Supports real-time event routing and analytics
- Embeds AI capabilities to assist users with intelligent suggestions and task automation.
- Provides a unified OneLake for centralized data storage and management.



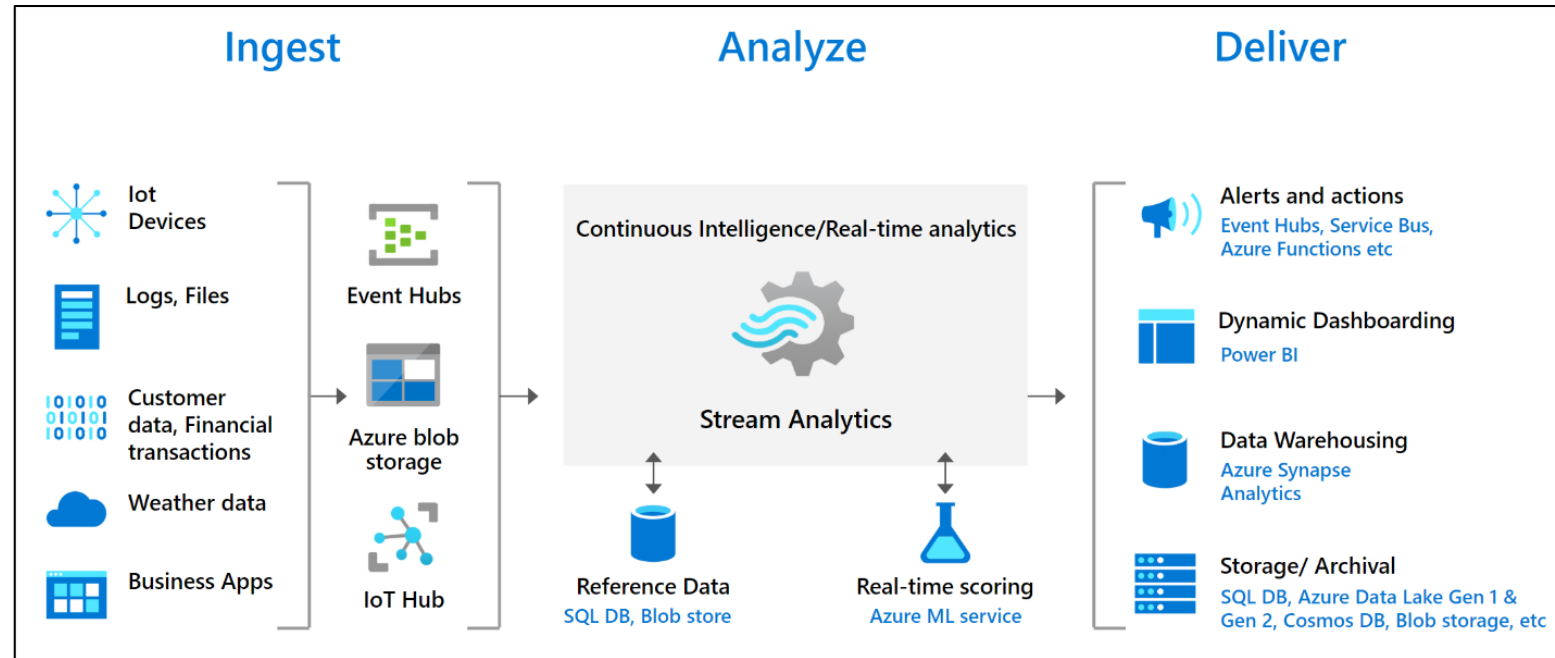
# Design Azure Stream Analytics solution for Data Analysis



# 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

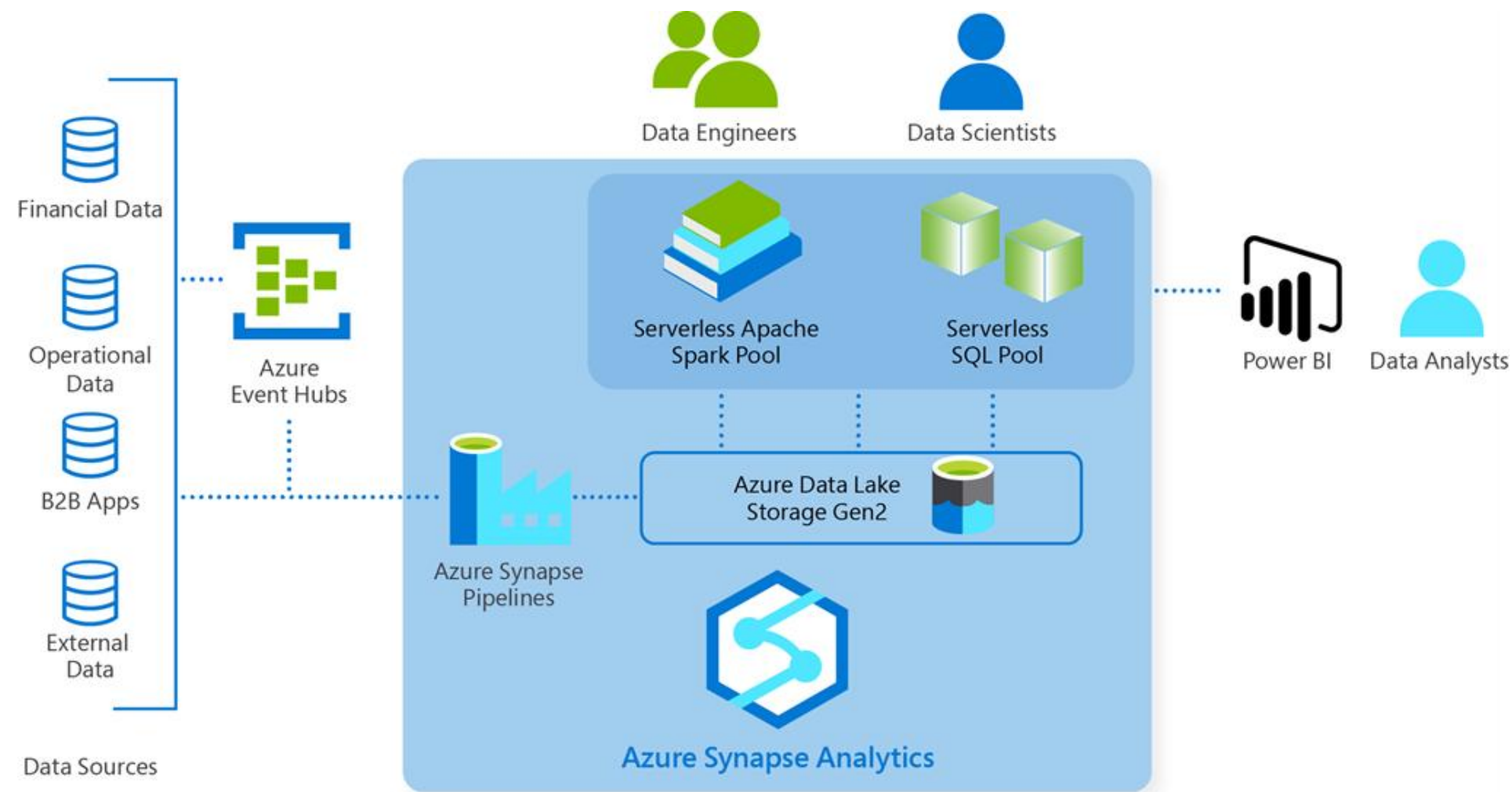


# Case study and review

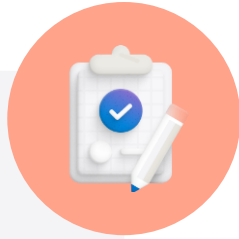




# Use Case: Just-in-time inventory



# Learning Recap – Data integration solutions



Check your  
knowledge  
questions and  
review

## Reference modules

- [Explore concepts of data analytics](#)
- [Data integration at scale Azure Data Factory](#)
- [Explore Azure Databricks](#)
- [Introduction to Azure Data Lake Storage Gen2](#)
- [Introduction to end-to-end analytics using Microsoft Fabric](#)

# End of presentation



Design a strategy for  
hot/warm/cold data path



# When to use Hot/Warm/Cold data path

Path	Requirement
Hot data path	<ul style="list-style-type: none"><li>• When data requirements are known to change frequently</li><li>• When processing or displaying data in real time</li></ul>
Warm data path	<ul style="list-style-type: none"><li>• When you need to store or display a recent subset of data</li><li>• Used for data that is consumed for small analytical and batch processing</li></ul>
Cold data path	<ul style="list-style-type: none"><li>• When data is rarely used. The data might be stored for compliance or legal reasons</li><li>• Used for data that is consumed for long term analytics and batch processing</li></ul>