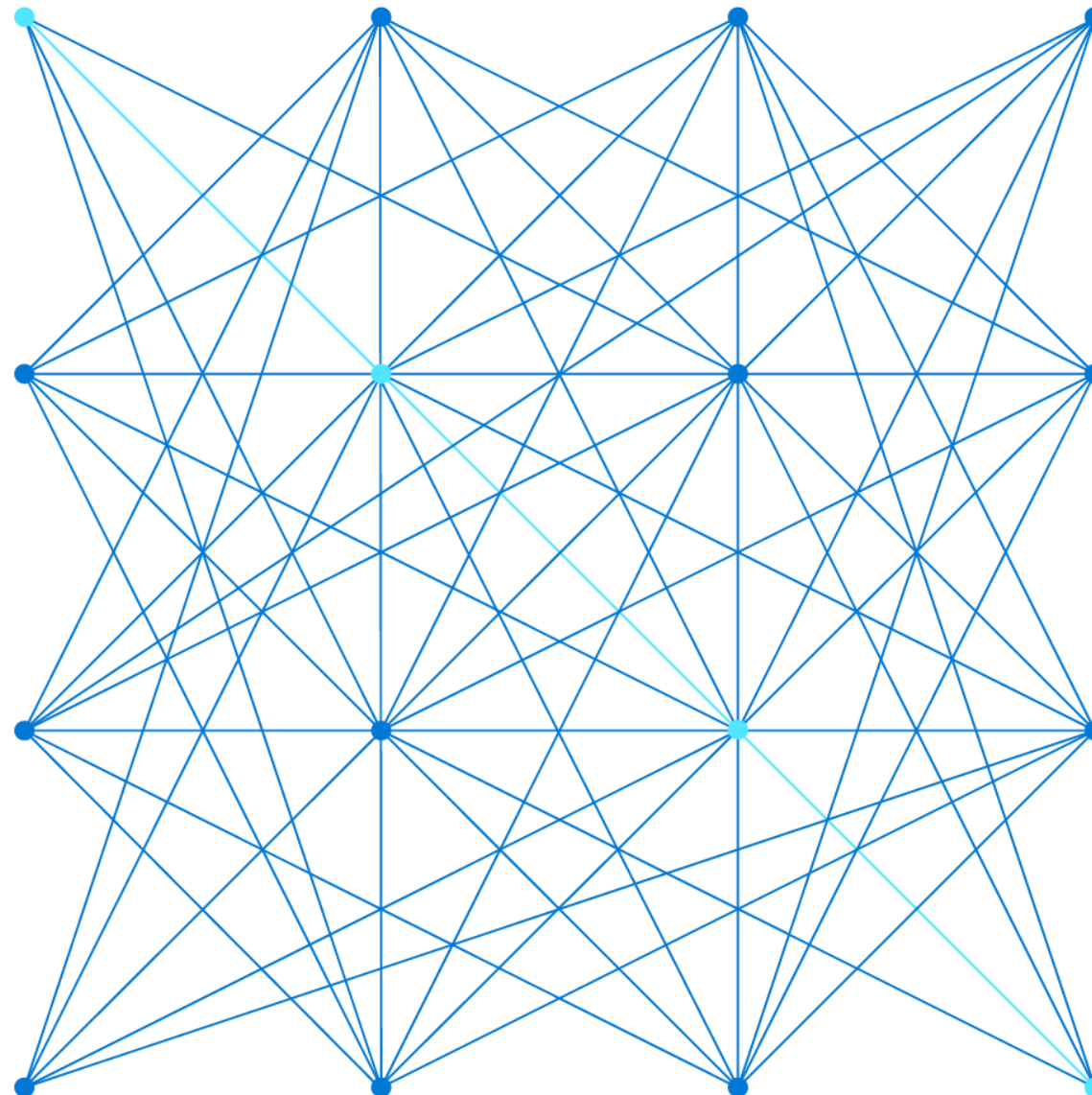


Module 4: Explore modern data warehouse analytics

Mohammed Arif
10/03/2022



Agenda



Examine components of a modern data warehouse



Explore large-scale data analytics



Get started building with Power BI

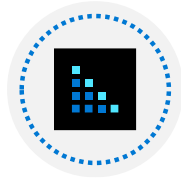
Lesson 1: Examine components of a modern data warehouse



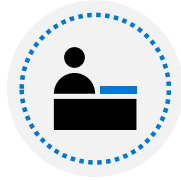
Lesson 1 objectives



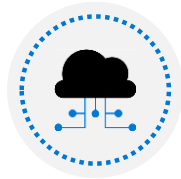
Explore data warehousing concepts



Explore Azure data services for modern data warehousing



Explore modern data warehousing architecture and workload



Explore Azure data services in the Azure portal

Demo: Modern data warehouse components

This video describes the components commonly used to create a data warehouse, and how data might flow through them. This video shows one particular approach



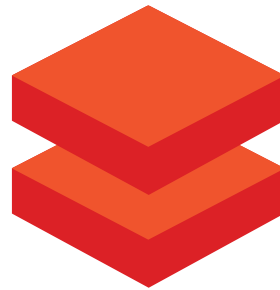
Azure services for the modern data warehouse



**Azure Data
Factory**



**Azure Data
Lake**



**Azure
Databricks**



**Azure
HDInsight**



**Azure Synapse
Analytics**

What is Azure Data Factory?

Azure Data Factory is described as a data integration service.

- Retrieves data from more than one data source and converts it.
- Filters out noise to keep interesting data
- Work is defined as a pipeline operation – runs continuously as data is received



What is Azure Data Lake Storage?

Azure Data Lake Storage is a repository of data for your modern data warehouse.

- Organizes data into directories for improved file access.
- Supports POSIX and RBAC permissions.
- Compatible with the Hadoop Distributed File System



What is Azure Databricks?

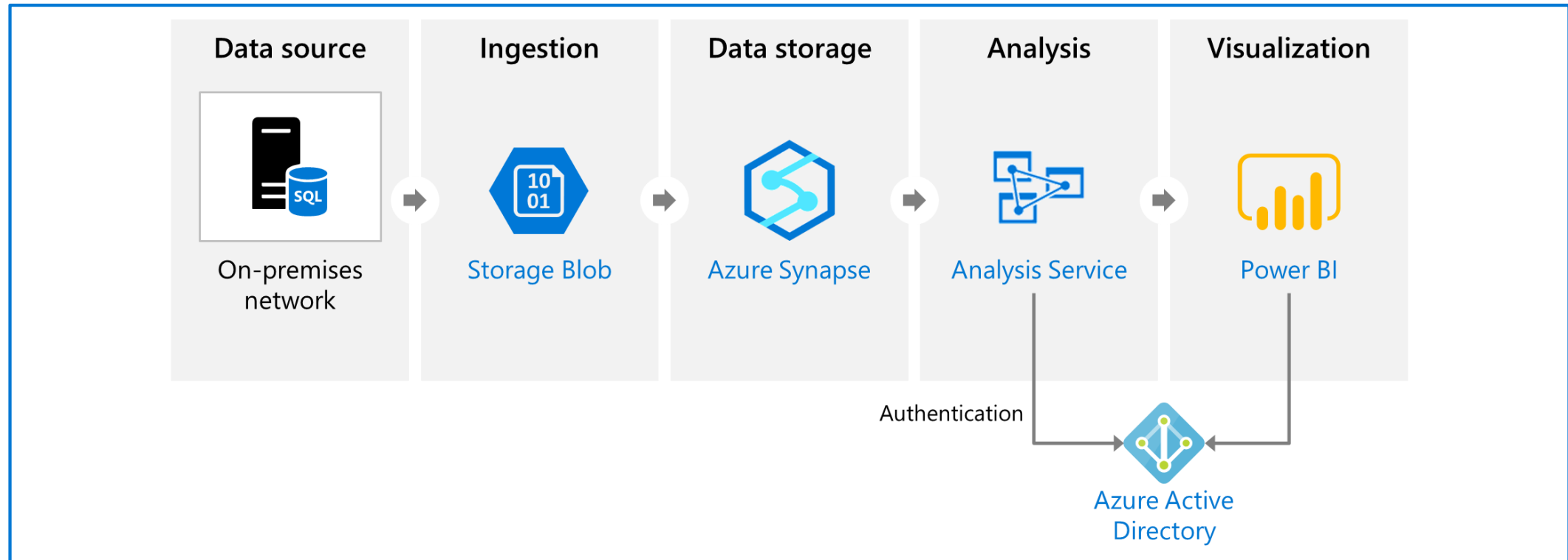
Azure Databricks is an Apache Spark-based platform that provides big data processing and streaming.

- Simplifies the provisioning and collaboration of Apache spark-based analytical solutions.
- Utilizes the security capabilities of Azure.
- Integrates with a variety of Azure data platform services and Power BI.



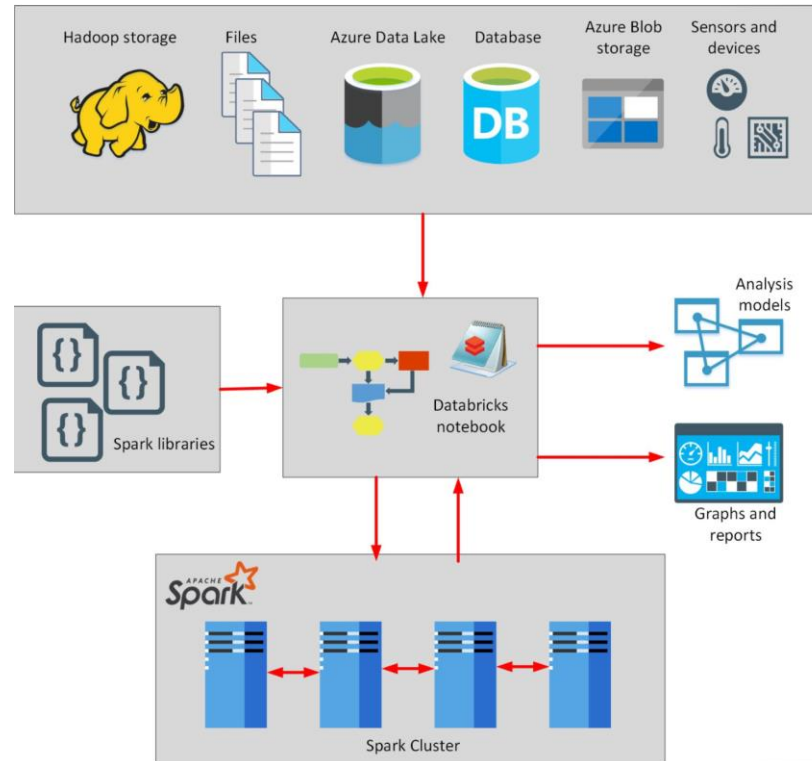
What is Azure Analysis Services?

Azure Analysis Services builds tabular models to support online analytical processing (OLAP) queries. It can combine data from multiple sources, including Azure SQL Database, Azure Synapse Analytics, Azure Data Lake Store, Azure Cosmos DB and others.



What is Azure HDInsight?

Azure HDInsight is a big data processing services which allows you to use open-source libraries on the one platform, in an Azure environment.



Lesson 1: Knowledge check



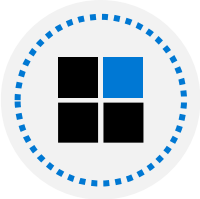
When should you use Azure Synapse Analytics?

- ☒ To perform very complex queries and aggregations
 - ☐ To create dashboards from tabular data
 - ☐ To enable large number of users to query analytics data
-



What is the purpose of data ingestion?

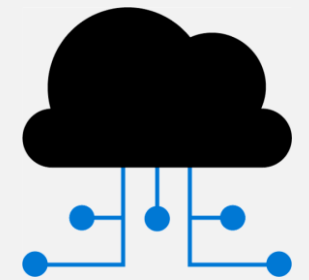
- ☐ To perform complex data transformations over data received from external sources
 - ☒ To capture data flowing into a data warehouse system as quickly as possible
 - ☐ To visualize the results of data analysis
-



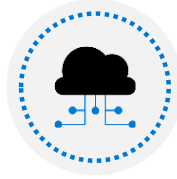
What is the primary difference between a data lake and a data warehouse?

- ☐ A data lake contains **structured information**, but a data warehouse holds **raw business data**
- ☒ A data lake holds **raw data**, but a data warehouse holds **structured information**
- ☐ Data stored in a data lake is dynamic, but information stored in a data warehouse is static

Lesson 2: Explore large-scale data analytics



Lesson 2 objectives



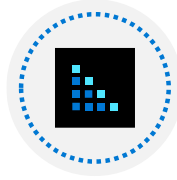
Describe data ingestion in Azure



Describe components of Azure Data Factory



Describe data processing options for performing analytics in Azure



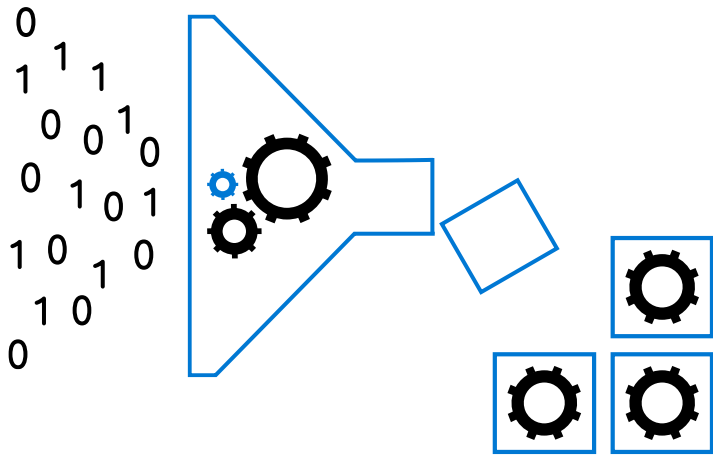
See how to use Azure Data Factory to load data into a data warehouse



Explore Azure Synapse Analytics

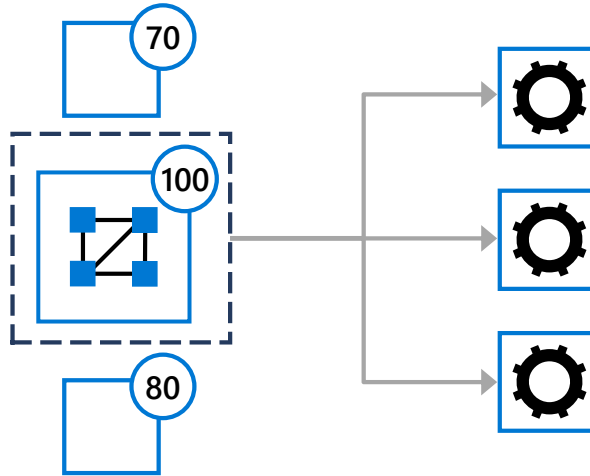
Describe data ingestion in Azure

ADF



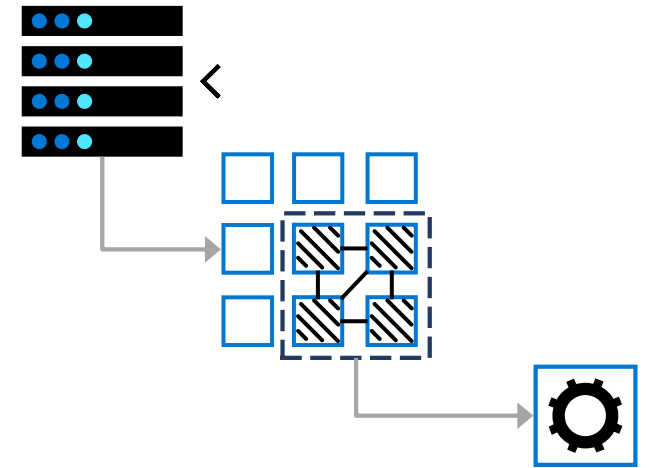
Heterogenous

PolyBase



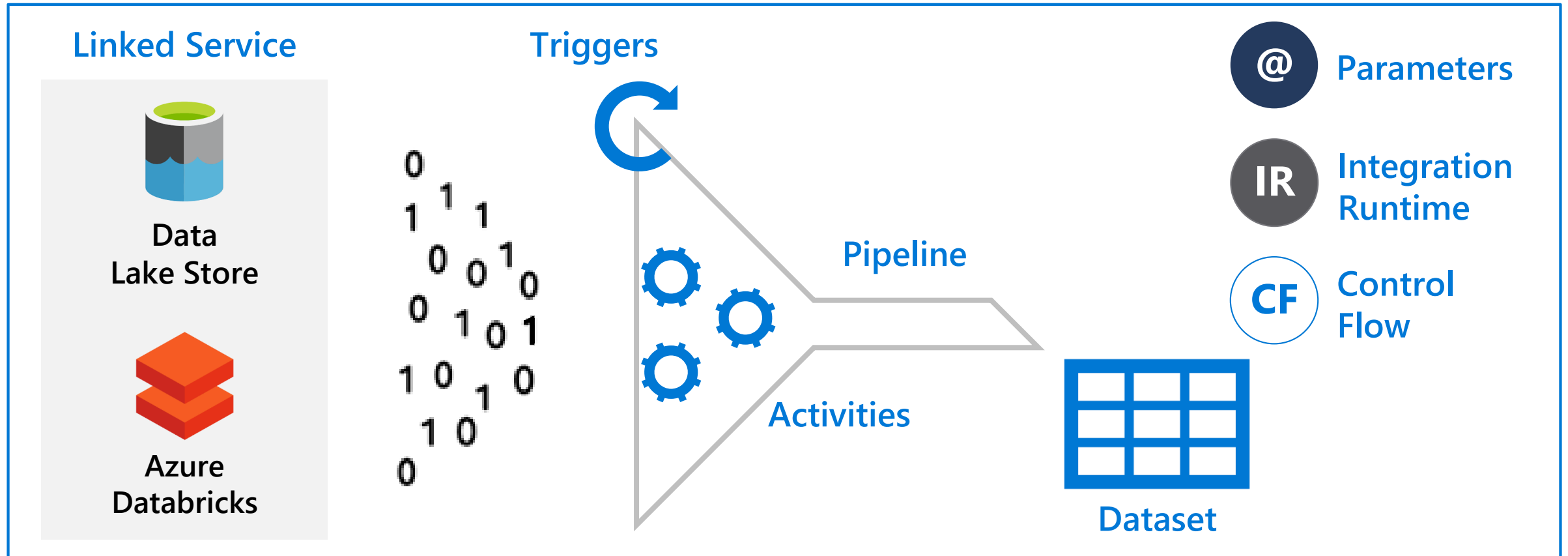
File based

SSIS



Heterogenous

Describe components of Azure Data Factory



Explore Azure Synapse Analytics



Synapse Pipelines



Synapse Link



Synapse Studio



Synapse SQL



Azure Synapse
Analytics



Synapse Spark

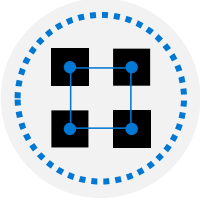
Demo: Create an Azure Synapse Analytics workspace and analyze data



In this demo, we'll create a Synapse Analytics workspace and use it to ingest and analyze some data.

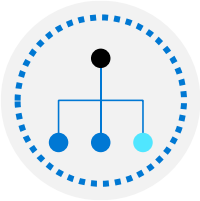
The demo is designed to familiarize you with some key capabilities of Synapse Analytics, not as a comprehensive guide to performing advanced data analysis. The exercise should take around 30 minutes to complete.

Lesson 2: Knowledge check



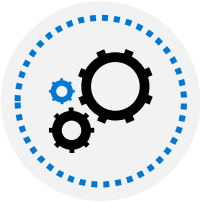
Which component of an Azure Data Factory can be triggered to run data ingestion tasks?

- ☐ CSV File
 - ☒ Pipeline
 - ☐ Linked service
-



When might you use PolyBase?

- ☒ To query data from external data sources from Azure SQL Database
 - ☐ To ingest streaming data using Azure Databricks
 - ☐ To orchestrate activities in Azure Data Factory
-



Which of these services can be used to ingest data into Azure Synapse Analytics?

- ☒ Azure Data Factory
- ☐ Power BI
- ☐ Azure Active Directory

Lesson 2: Knowledge check (continued)



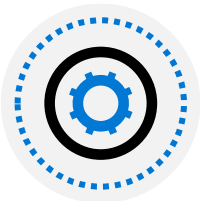
You have a large amount of data held in files in Azure Data Lake storage. You want to retrieve the data in these files and use it to populate tables held in Azure Synapse Analytics. Which processing option is most appropriate?

- ☐ Use Azure Synapse Link to connect to Azure Data Lake storage and download the data
 - ☒ Synapse SQL pool
 - ☐ Synapse Spark pool
-



Which of the components of Azure Synapse Analytics allows you to train AI models using AzureML?

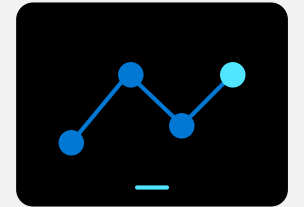
- ☐ Synapse Studio
 - ☐ Synapse Pipelines
 - ☒ Synapse Spark
-



In Azure Databricks how do you change the language a cell uses?

- ☒ The first line in the cell is %language. For example, %scala
- ☐ Change the notebook language before writing the commands
- ☐ Wrap the command in the cell with ##language##

Lesson 3: Get started building with Power BI



Lesson 4 objectives



Learn how Power BI services and applications work together

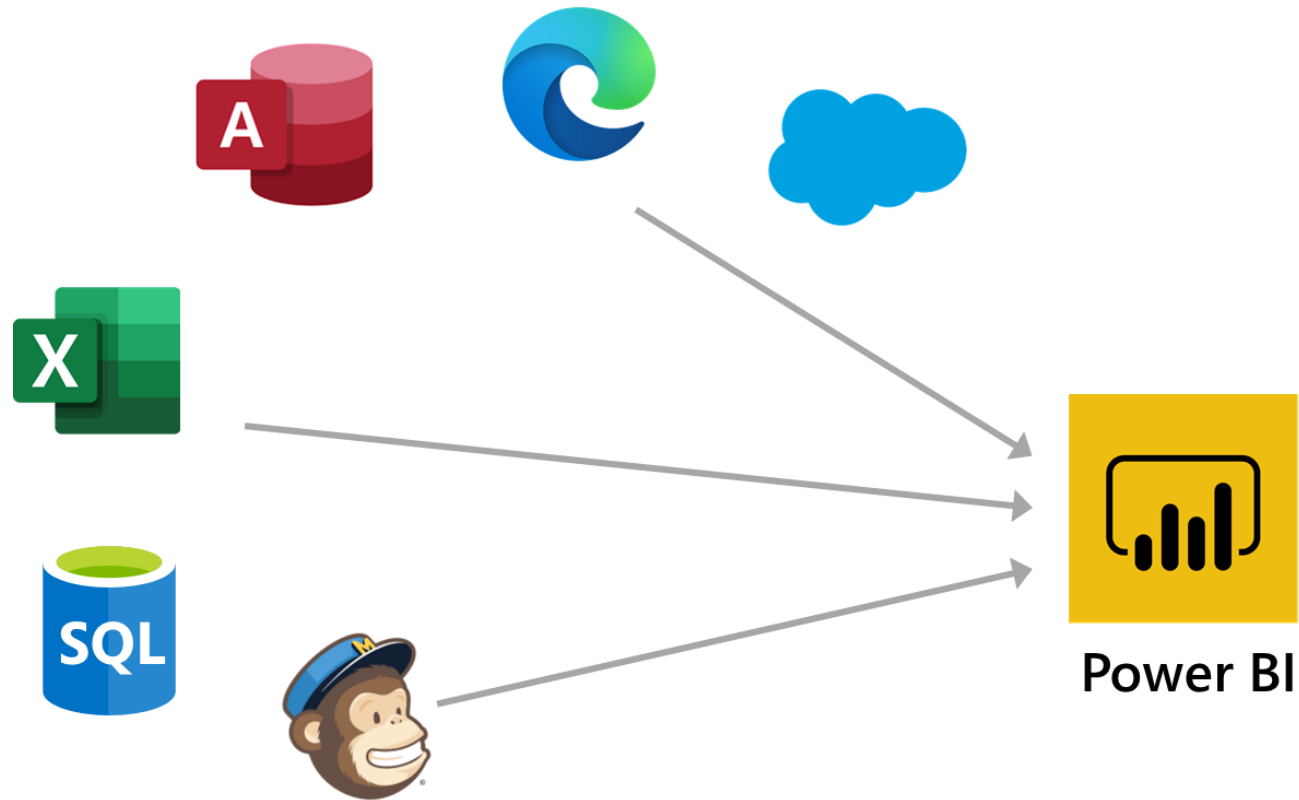


Explore how Power BI can make your business more efficient



Learn how to create compelling visuals and reports

What is Power BI?

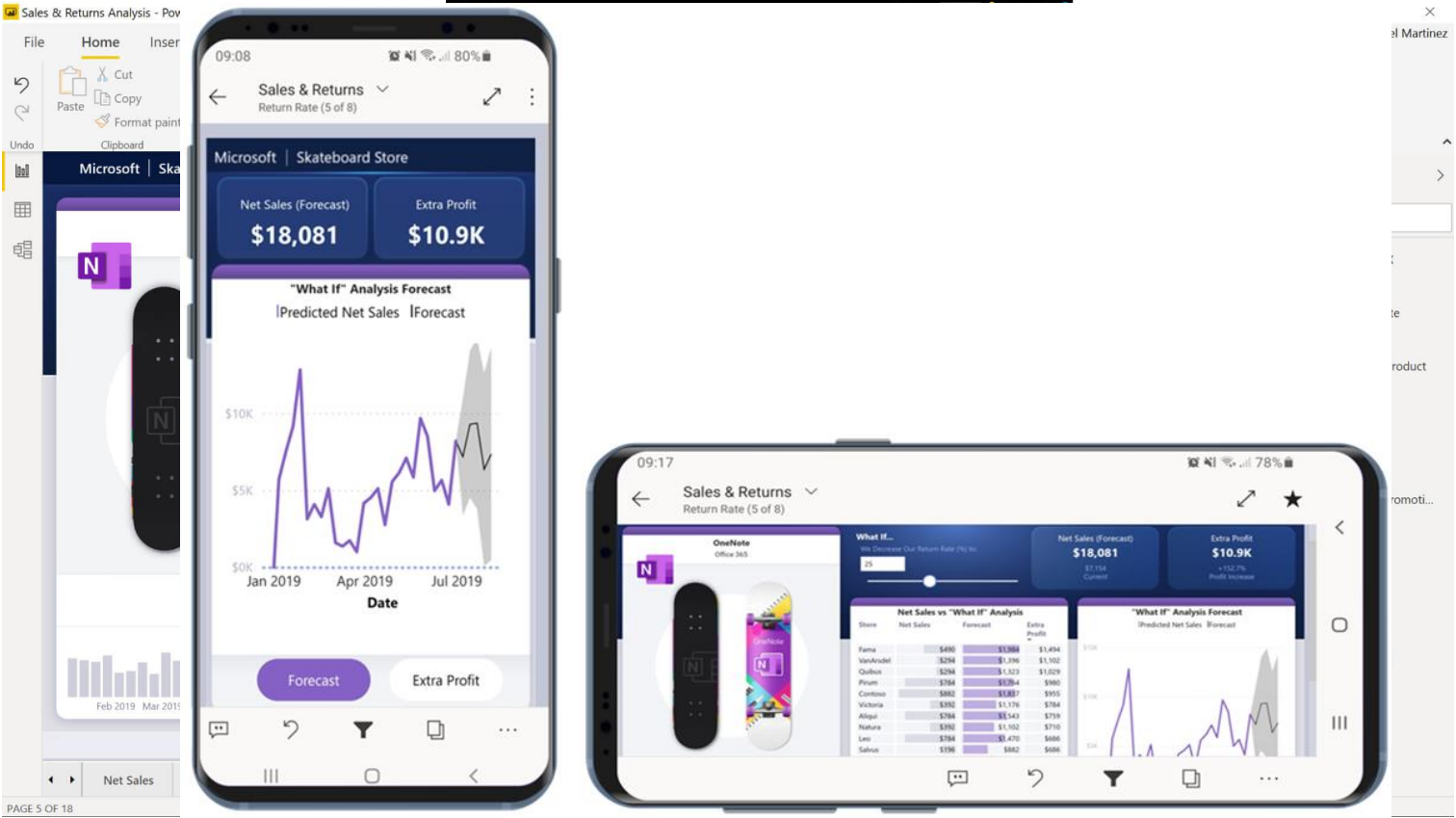


How can you use Power BI?

Power BI Desktop

Power BI Service

Power BI Mobile



The anatomy of a Power BI app

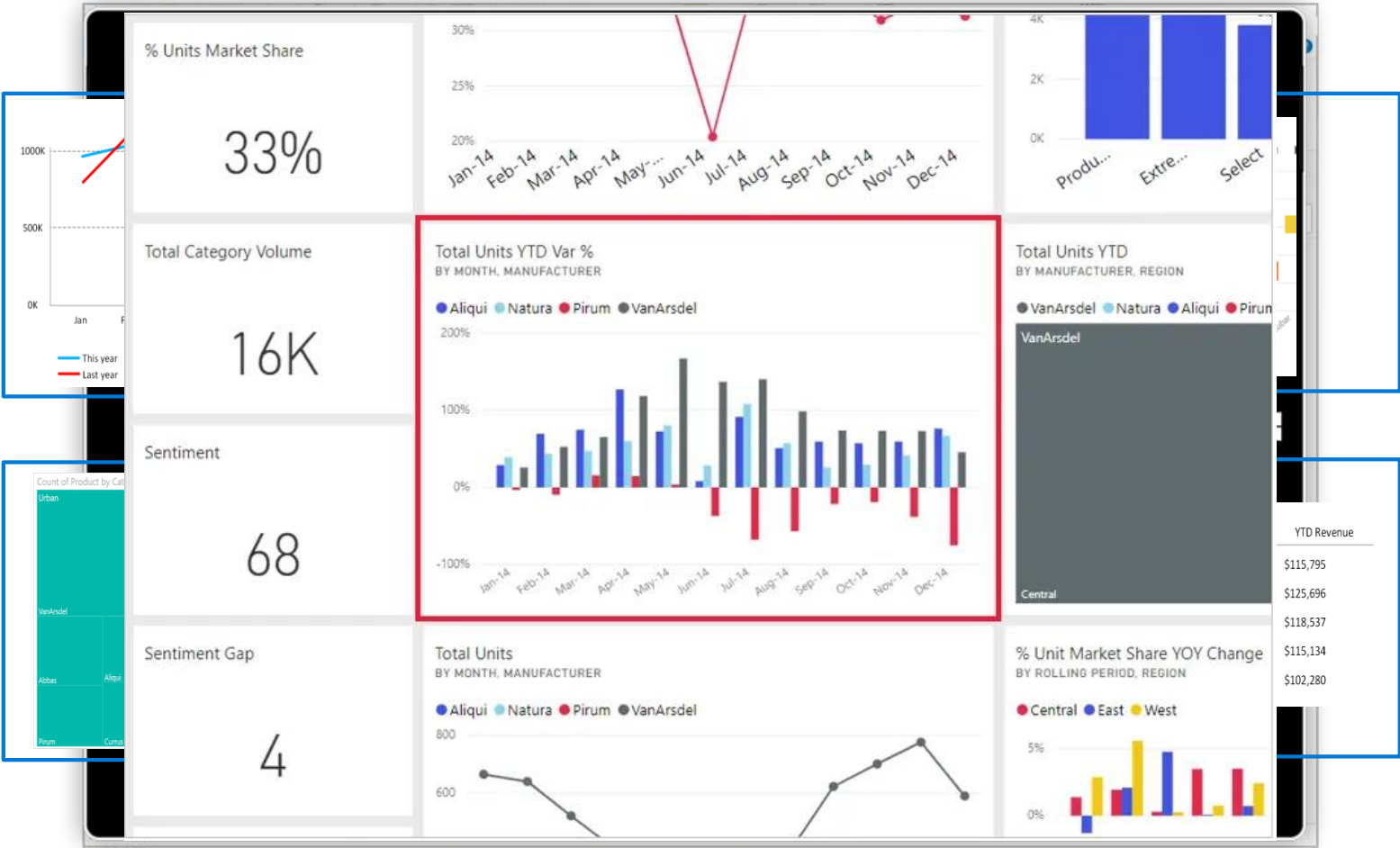
Visualizations

Datasets

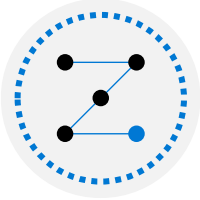
Reports

Dashboards

Tiles

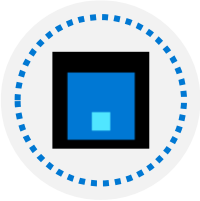


Lesson 4: Knowledge check



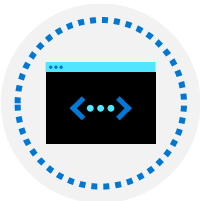
What is the common flow of activity in Power BI?

- ☐ Create a report in Power BI mobile, share it to the Power BI Desktop, view and interact in the Power BI service
 - ☐ Create a report in the Power BI service, share it to Power BI mobile, interact with it in Power BI Desktop
 - ☒ Bring data into Power BI Desktop and create a report, share it to the Power BI service, view and interact with reports and dashboards in the service and Power BI mobile
 - ☐ Bring data into Power BI mobile, create a report, then share it to Power BI Desktop
-



Which of the following are building blocks of Power BI?

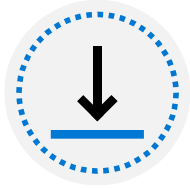
- ☐ Tiles, dashboards, databases, mobile devices
 - ☒ Visualizations, datasets, reports, dashboards, tiles
 - ☐ Visual Studio, C#, and JSON files
-



A collection of ready-made visuals, pre-arranged in dashboards and reports is called what in Power BI?

- ☐ The canvas
- ☐ Scheduled refresh
- ☒ An app

Lab: Explore Power BI



Go to the [Visualize data with Power BI](#) module on Microsoft Learn and follow the instructions in the exercise.

