

DP-900: Microsoft Azure Data Fundamentals

MODULE 4



Module 3: Modern Data Warehouse Analytics

1

**Examine
components of a
Modern Data
Warehouse**

2

**Explore Data
Ingestion in
Azure**

3

**Explore Data
Storage and
Processing in
Azure**

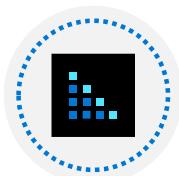
4

**Getting Started
with Power BI**

Examine components of a Modern Data Warehouse



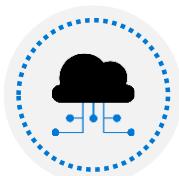
Explore data warehousing concepts



Explore Azure data services for modern data warehousing

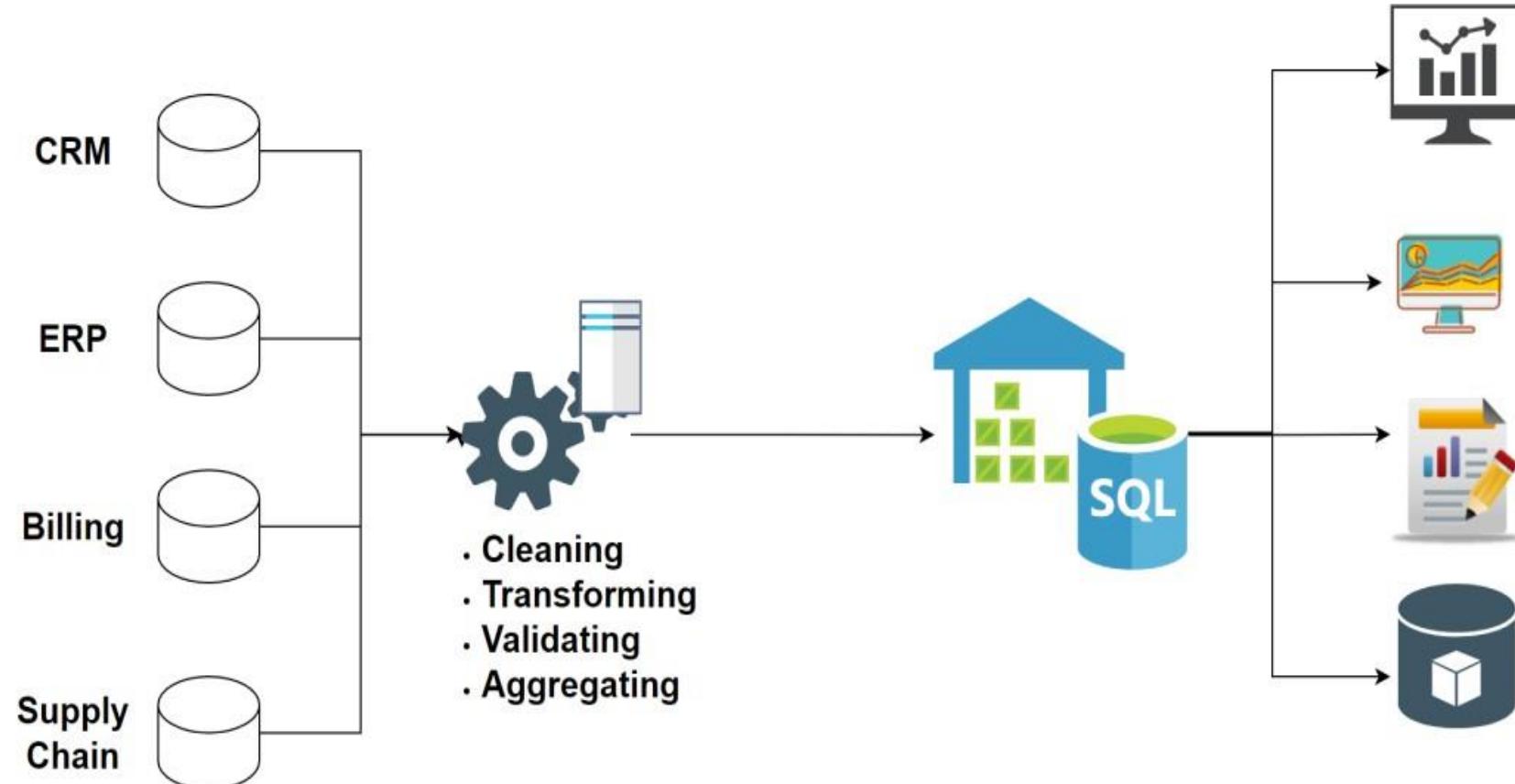


Explore modern data warehousing architecture and workload



Explore Azure data services in the Azure portal

What is a Data Warehouse?



Problems a DWH can solve

We have so much data, but we can't make anything of it

I only want to know what is important

We need to slice and dice the data

Business needs easy access to data

Numbers between departments don't match and we don't know who's right

We want people to make decisions based on facts

OLTP vs OLAP

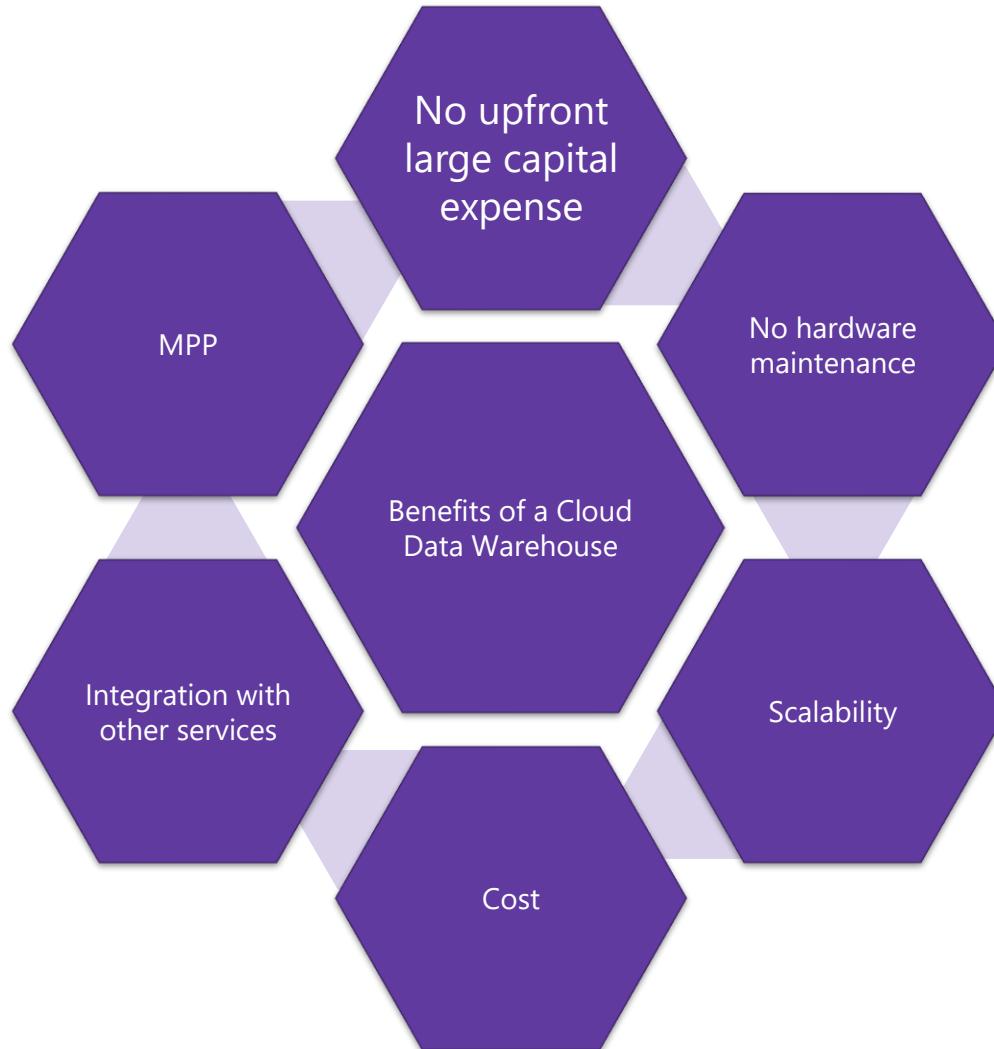
OLTP

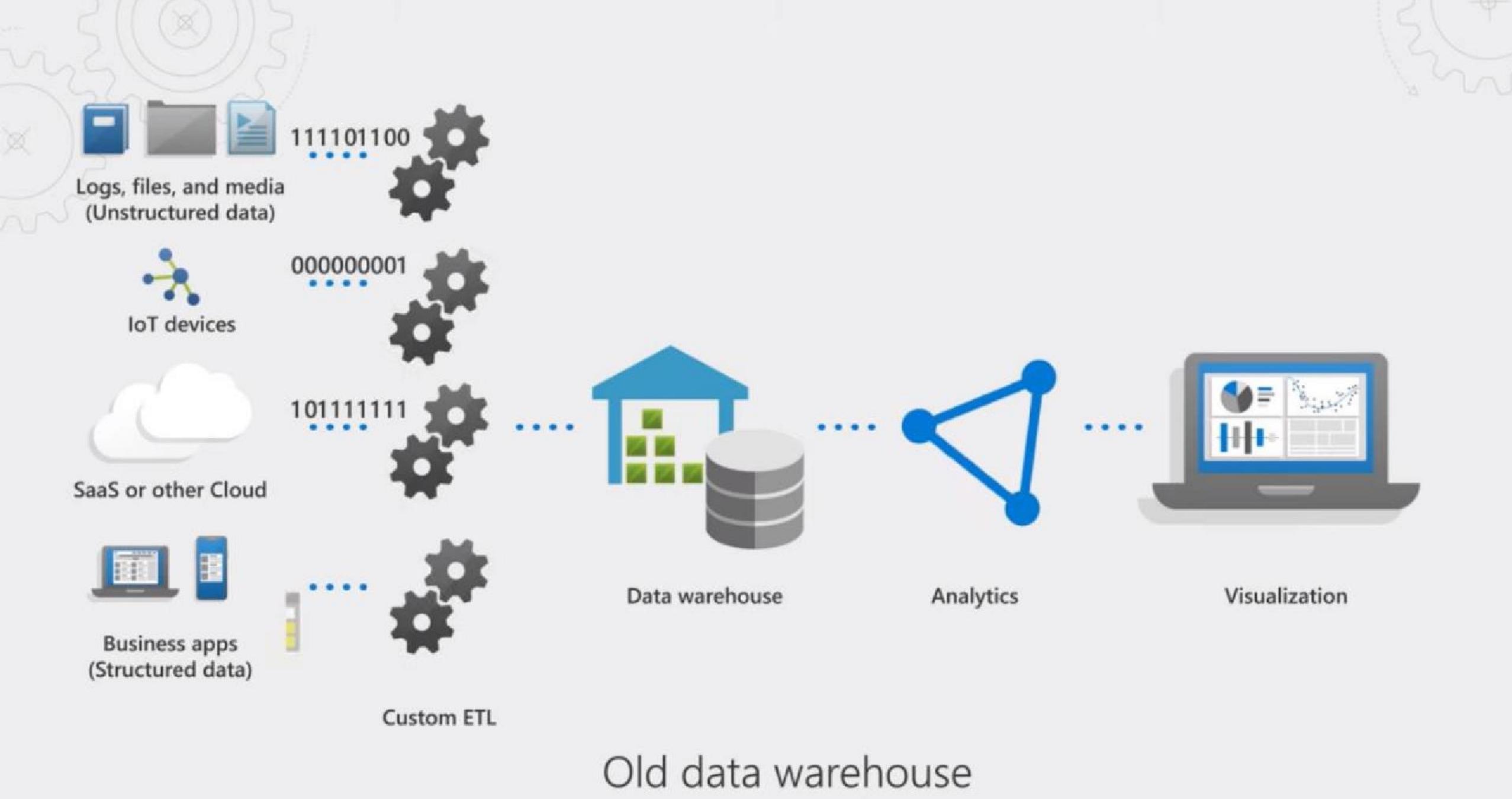
- Many small transaction
- Current data
- Used to run the business
- Highly detailed
- Typically, in the GB scale
- Processing performance limit

OLAP

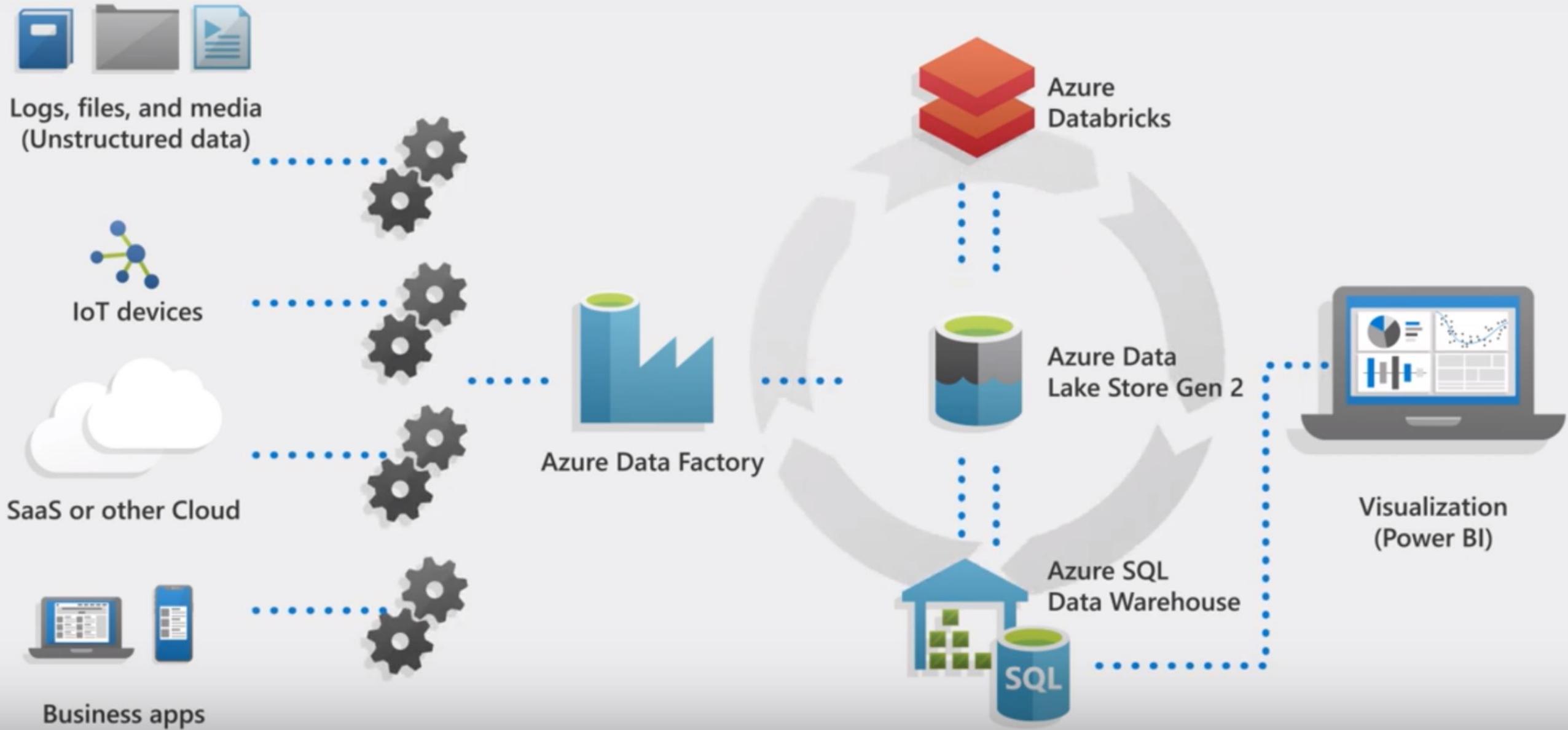
- Low volume but complex queries
- Historic, non-volatile data
- Used to analyze the business
- Consolidated and summarized
- TB and above scale
- No limit, pause/resume compute

Why DWH in Cloud?

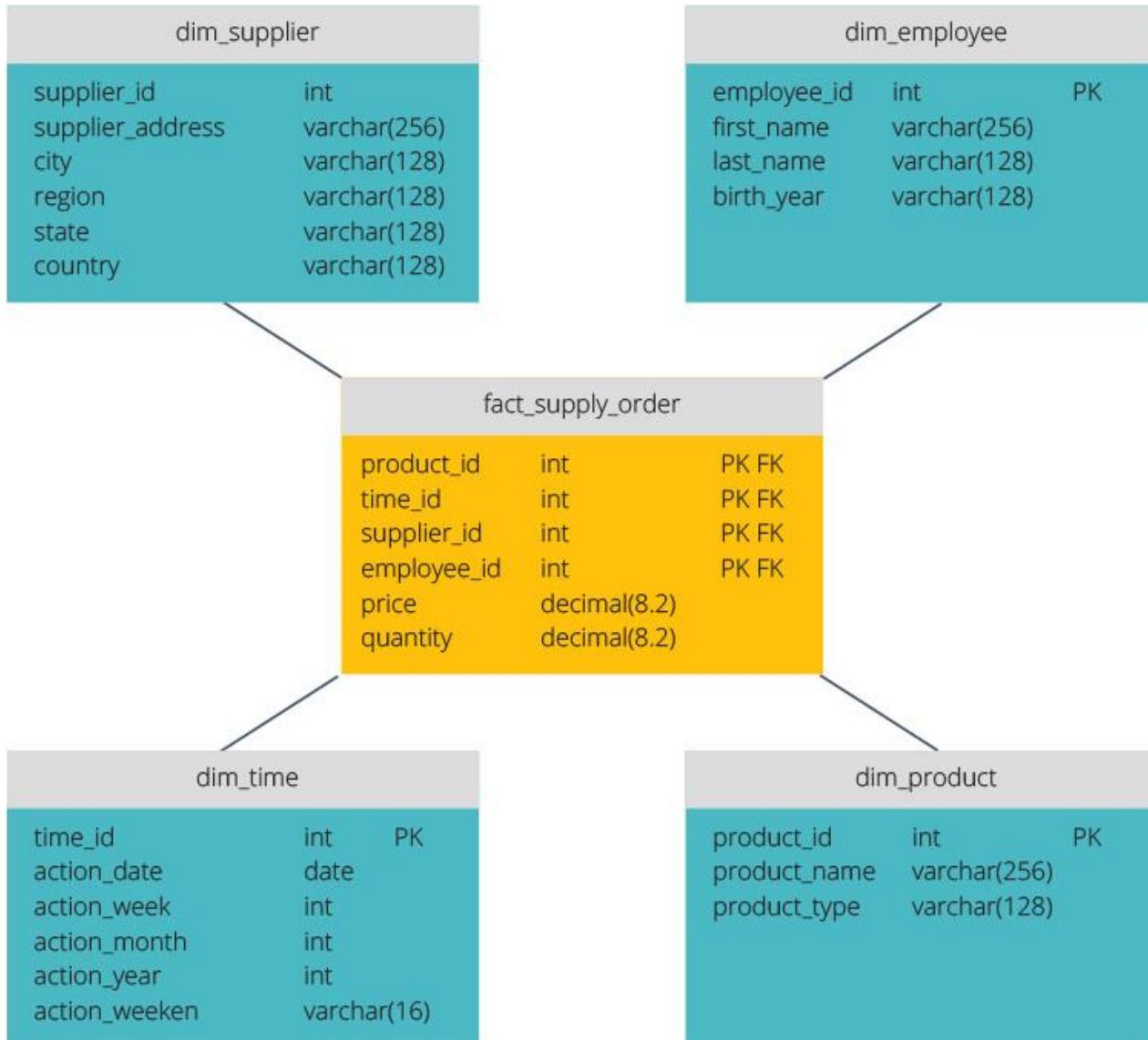




Modern Data Warehouse



Data Warehouse Star Schema

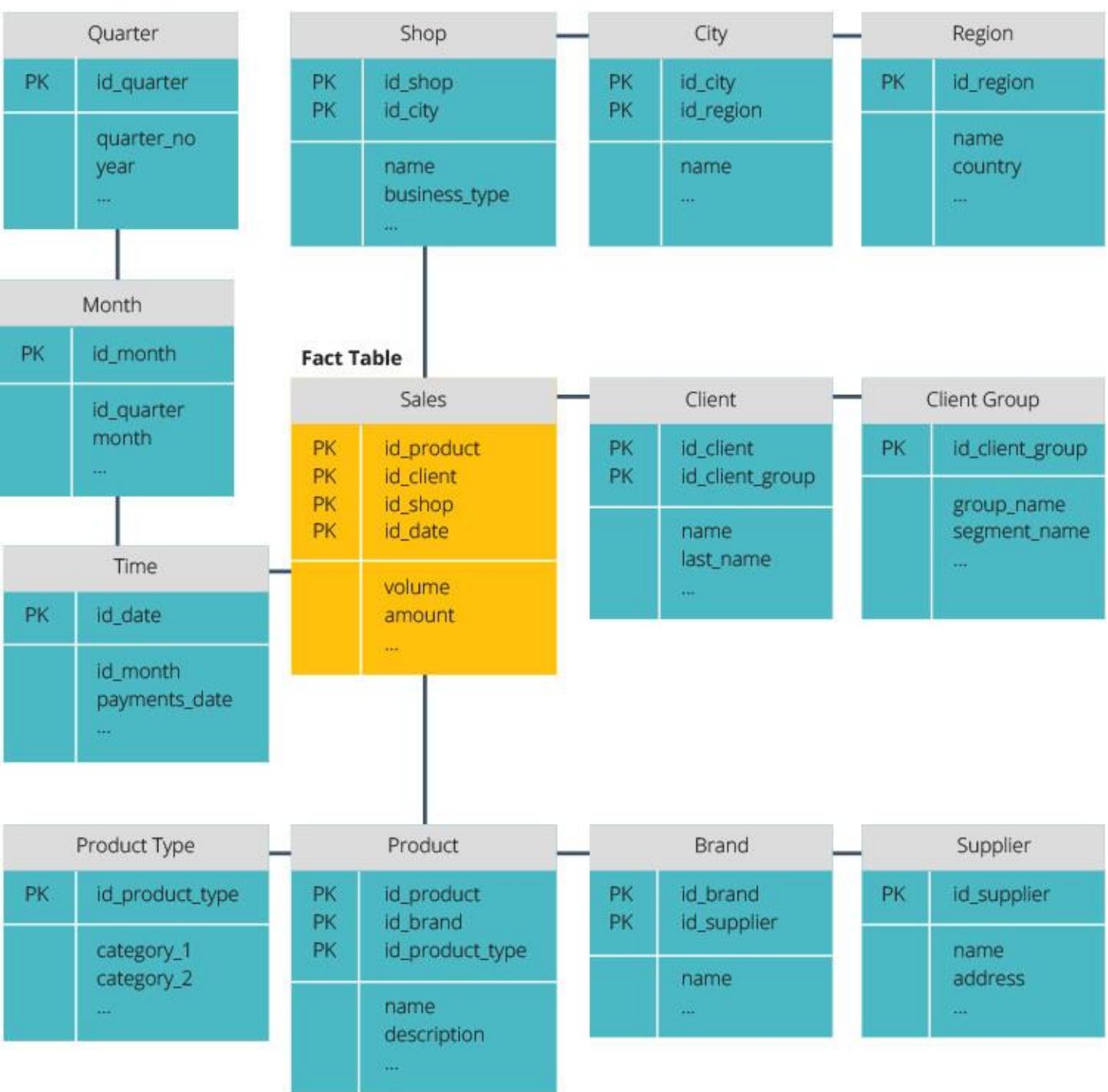


A Data Mart design pattern consisting of fact and dimension tables resembling a star like shape.

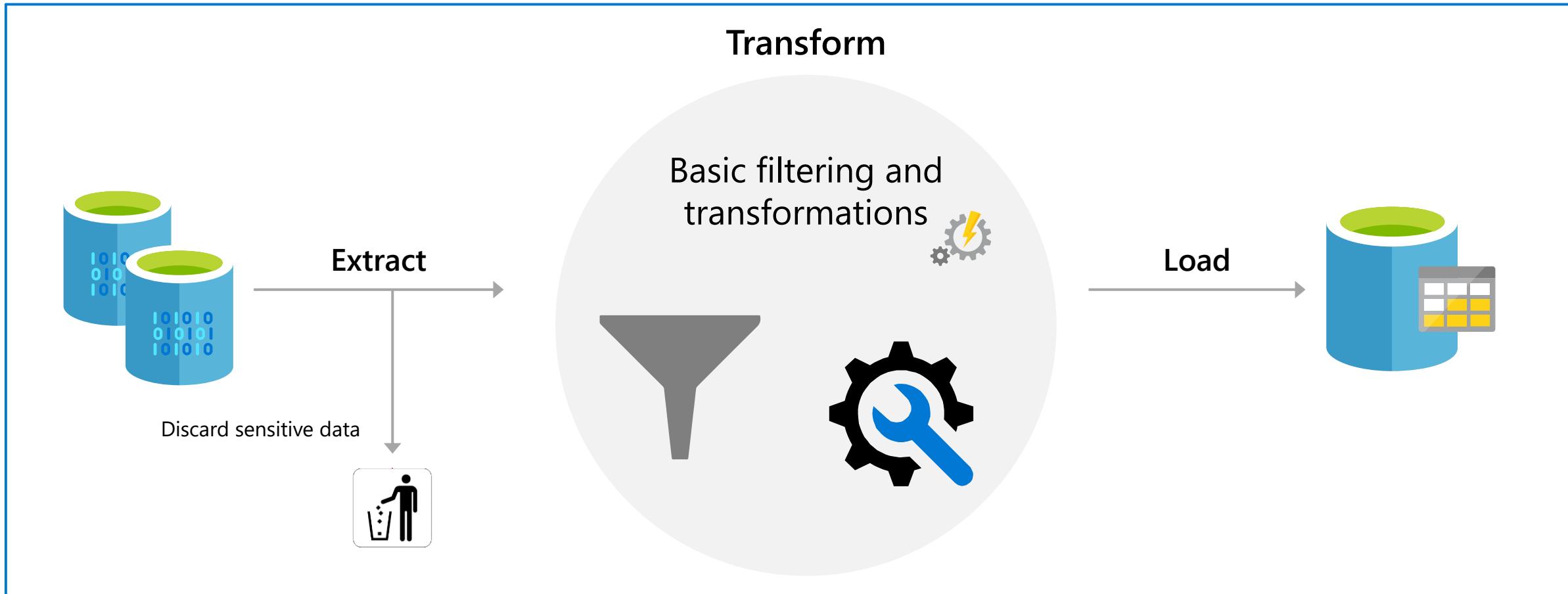
Data Warehouse Star Schema

Facts	Dimensions
Store observations or events	Describe business entities – things you model
Subject of Analysis	Context of Analysis
Numerical	Textual (most often)
Additive and Semi-additive	Contain most attributes
sales orders, stock balances, exchange rates, temperatures, etc.	products, people, places, and concepts including time itself.

Data Warehouse Snowflake Schema



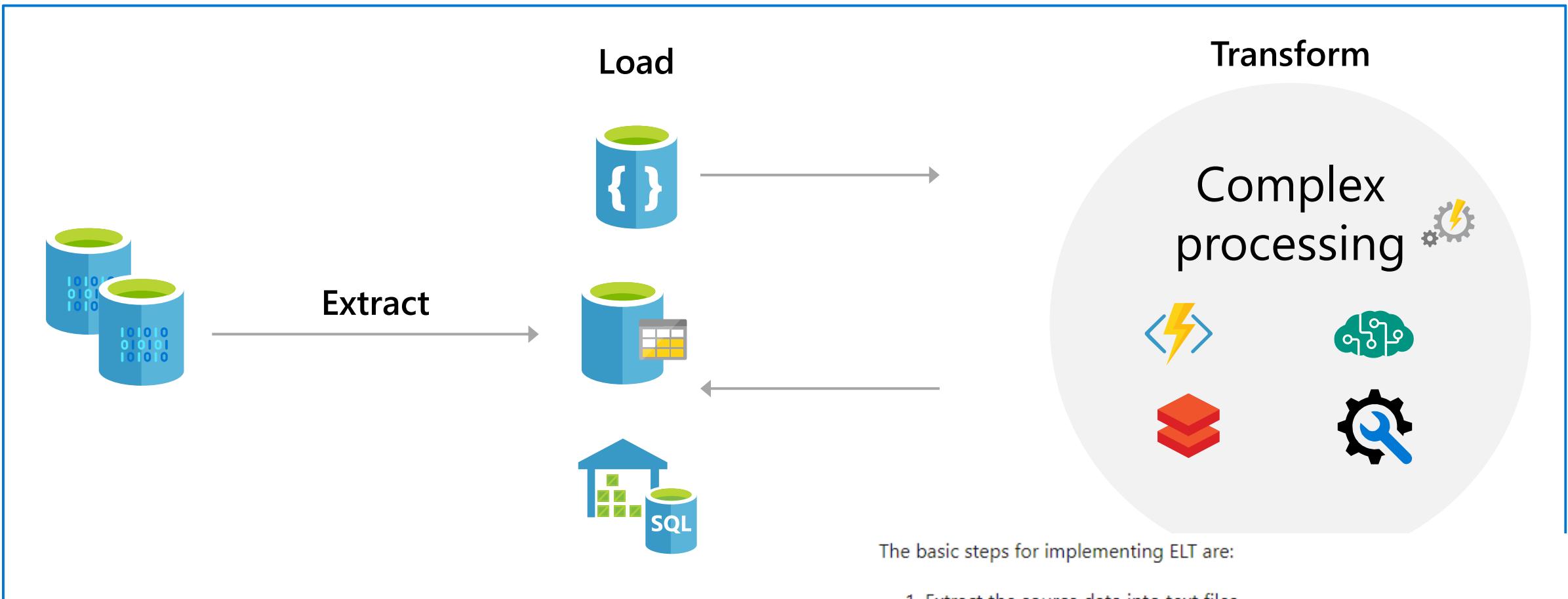
ETL



Azure Data Factory

Azure Synapse Analytics Pipeline

ELT



1. Extract the source data into text files.
2. Land the data into Azure Blob storage or Azure Data Lake Store.
3. Prepare the data for loading.
4. Load the data into staging tables with PolyBase or the COPY command.
5. Transform the data.
6. Insert the data into production tables.

Data Analytics - Types

DESCRIPTIVE ANALYTICS

"What is happening in your business?"

It gives us only insight about whether everything is going well or not in our business without explaining the root cause.

DIAGNOSTIC ANALYTICS

"Why it is happening in your business?"

Diagnostic Analytics explains the root cause behind the outcome of descriptive analytics.

PREDICTIVE ANALYTICS

Explains "what's likely to happen in the future based on previous trends and patterns?"

By utilizing various statistical and machine learning algorithms to provide recommendations and provide answers to questions related to what might happen in the future, that cannot be answered by BI.

PRESCRIPTIVE ANALYTICS

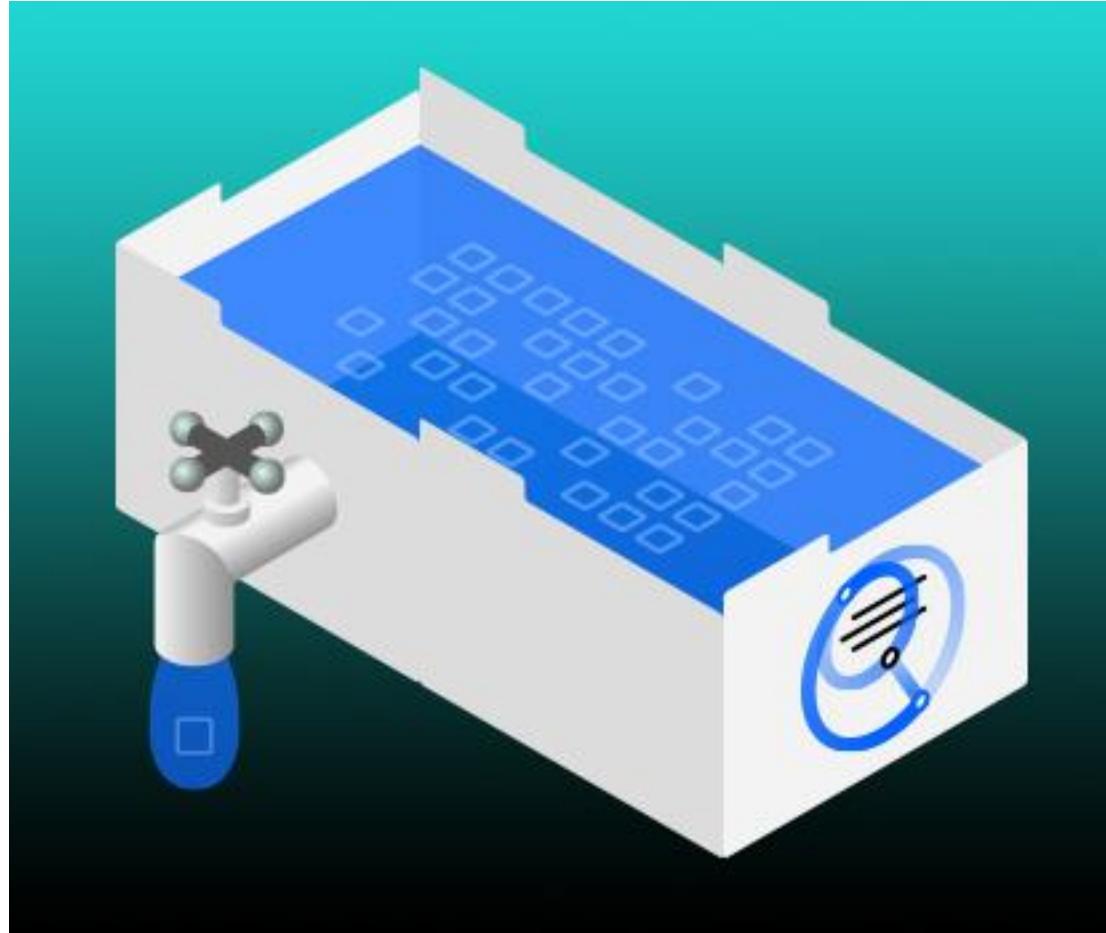
"Helps you to determine the best course of action to choose to bypass or eliminate future issues"

You can use Prescriptive analytics to advise users on possible outcomes and what should they do to maximize their key business metrics.

COGNITIVE ANALYTICS

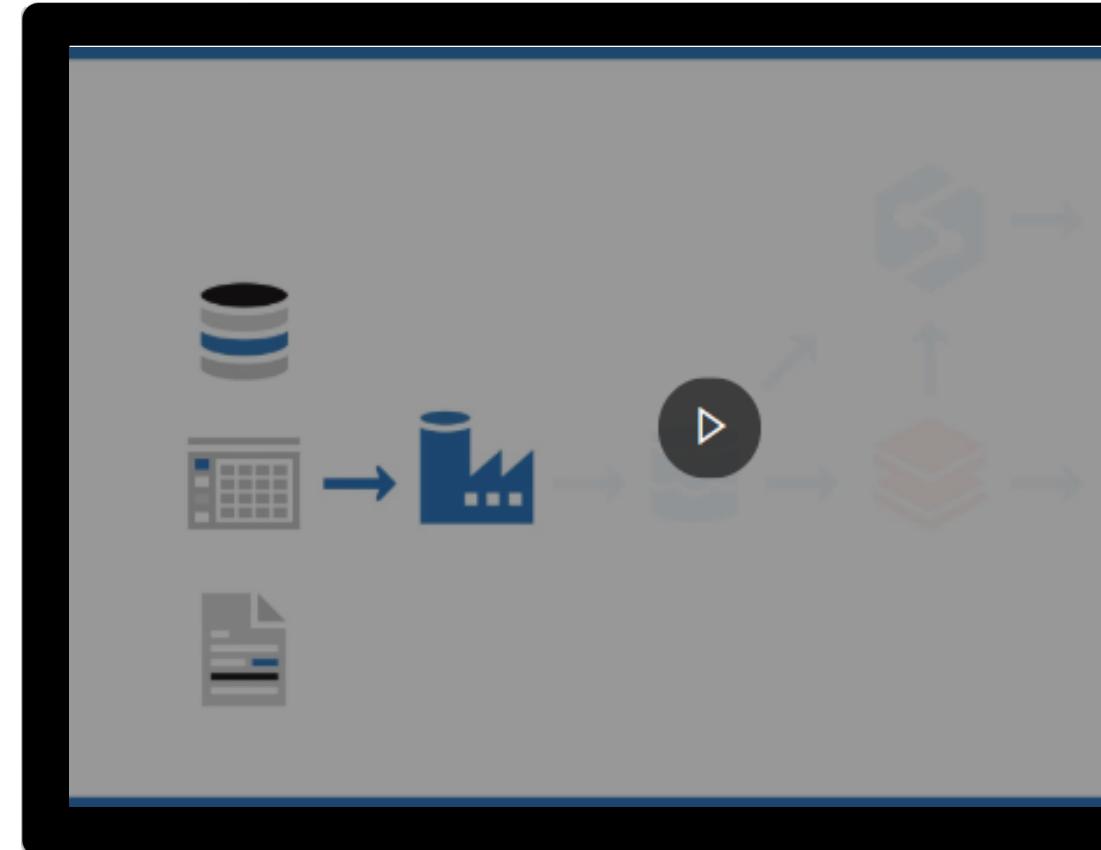
"It combines a number of intelligent technologies like artificial intelligence, machine-learning algorithms, deep learning etc. to apply human brain like intelligence to perform certain tasks."

Data Lake

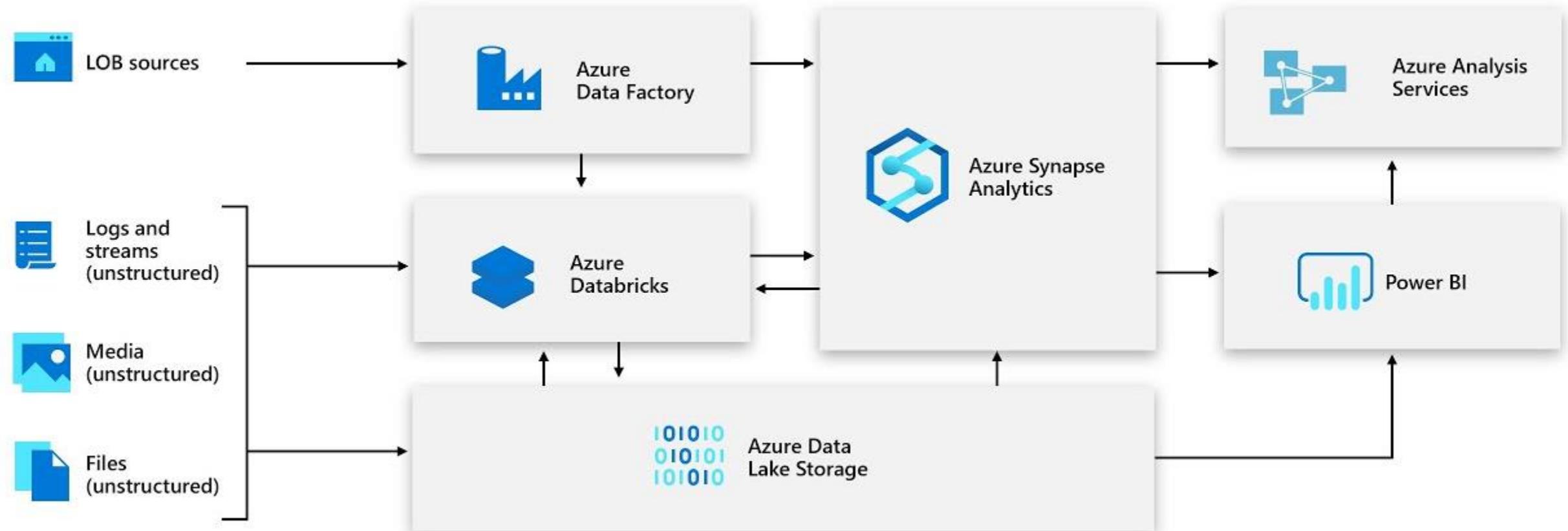


Demo: Modern data warehouse components

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



What is Azure Data Factory?

Azure Data Factory (ADF) is a hybrid data integration service that enables you to quickly and efficiently create automated data pipelines – without having to write any code!

- 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



Azure Data Lake Storage

A repository of data
for your Modern
Data Warehouse

Organises data into
directories for
improved file access

Supports POSIX and
RBAC permissions

It is compatible with
Hadoop Distributed
File System

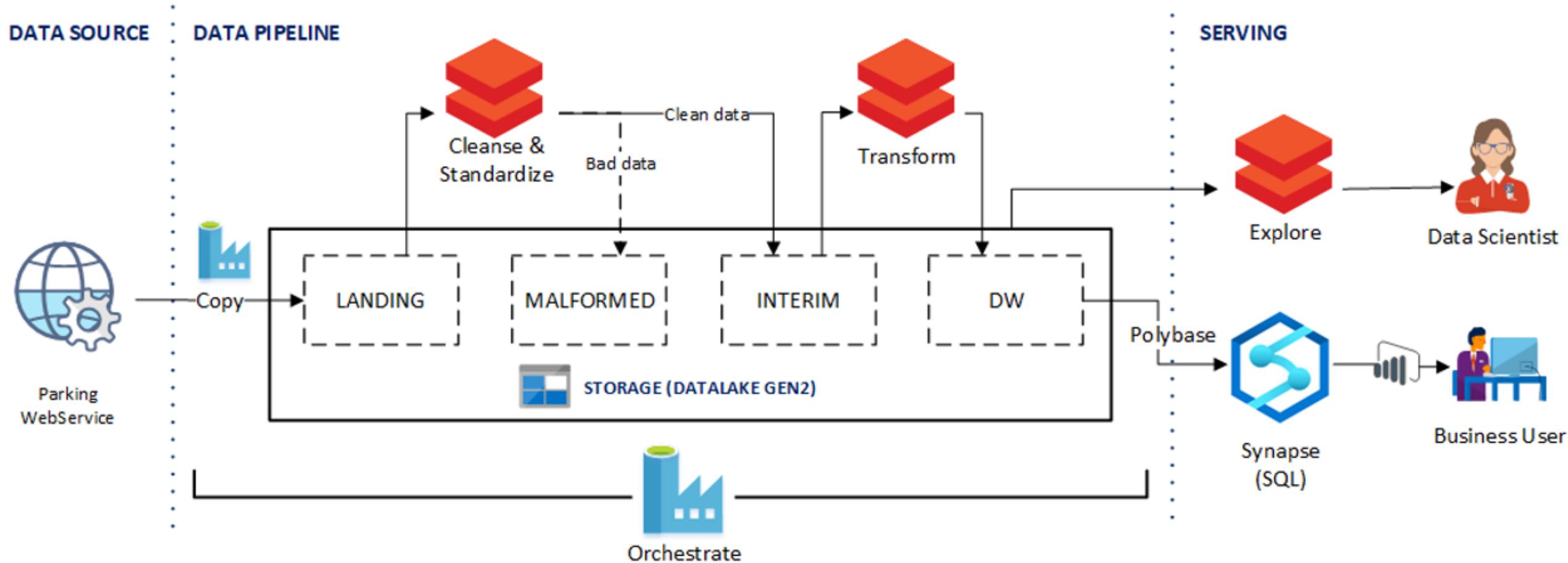
Store



Azure Data Lake Storage

High performance data lake available
in all 54 Azure regions

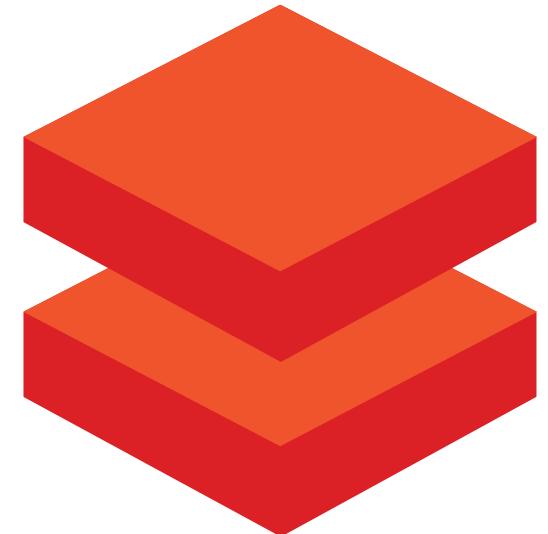
Azure Data Lake Storage



What is Azure Databricks?

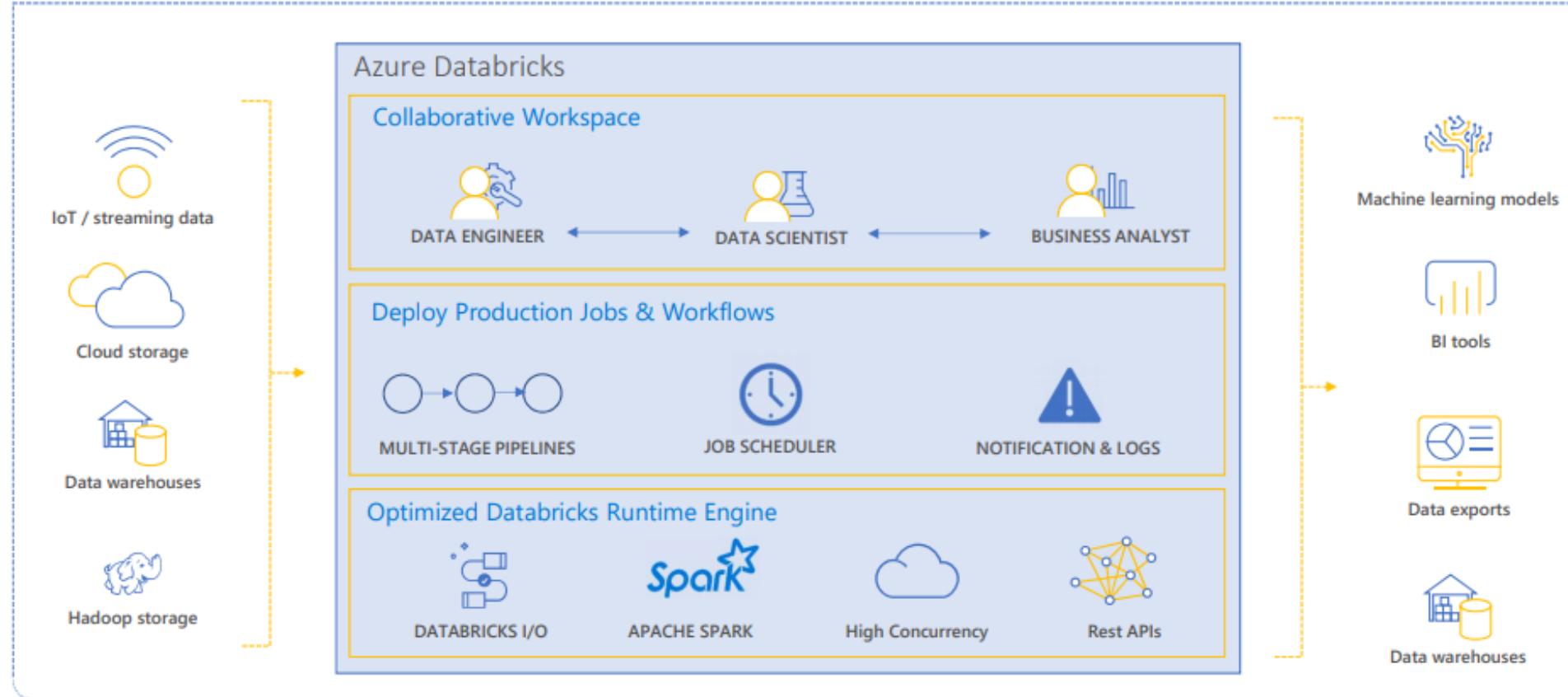
Azure Databricks is an Apache Spark-based platform that provides big data processing and streaming.

- 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.



Azure Databricks

AZURE DATABRICKS



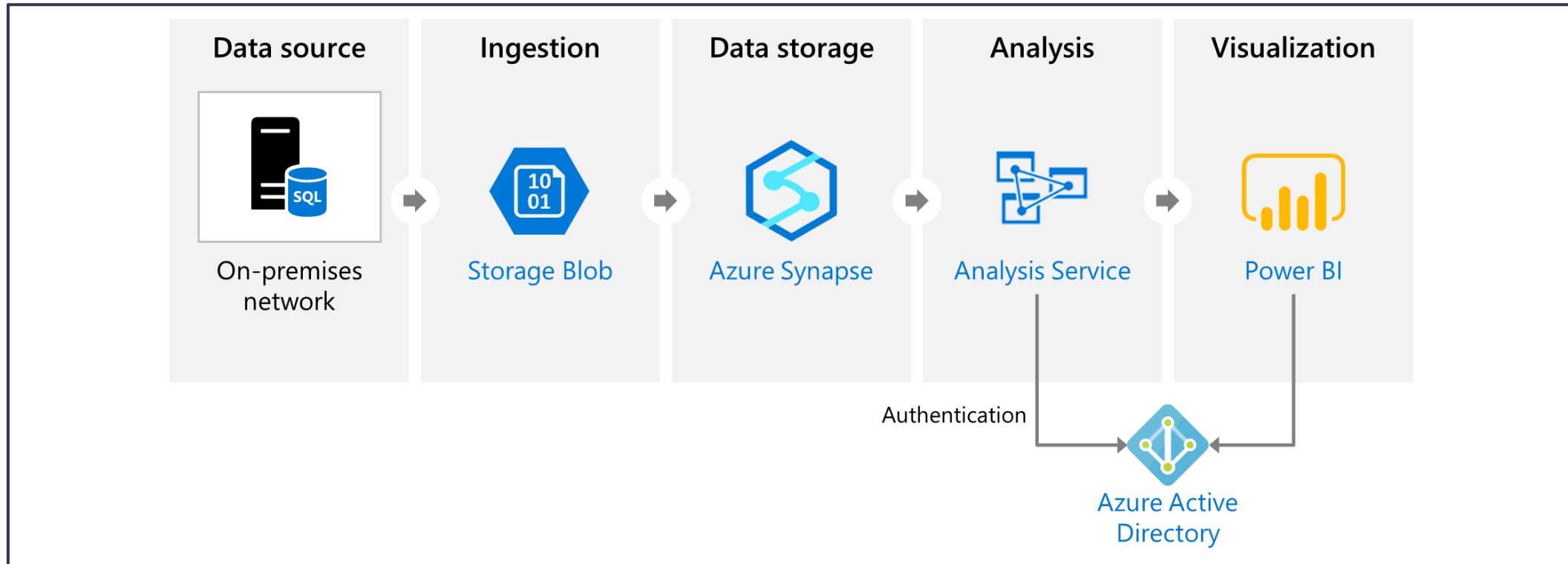
Enhance Productivity

Build on secure & trusted cloud

Scale without limits

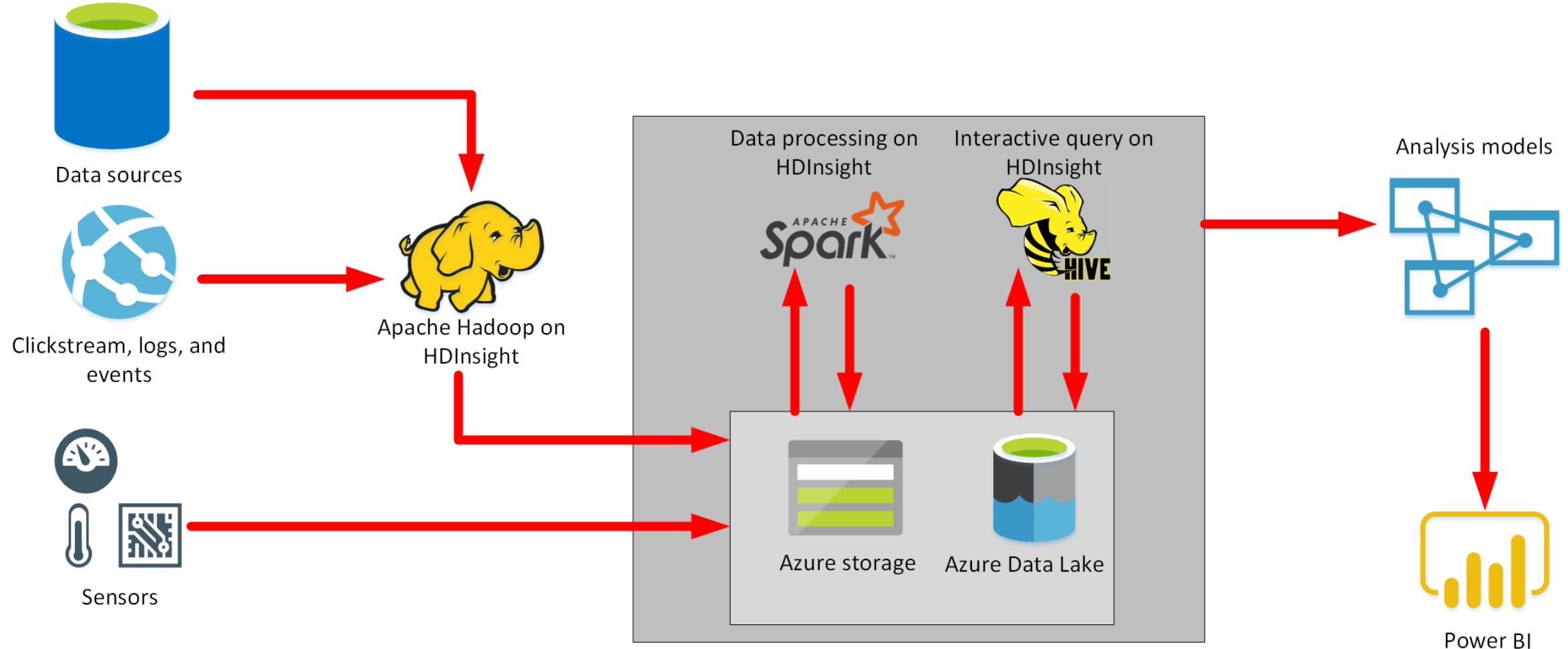
What is Azure Analysis Services?

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.

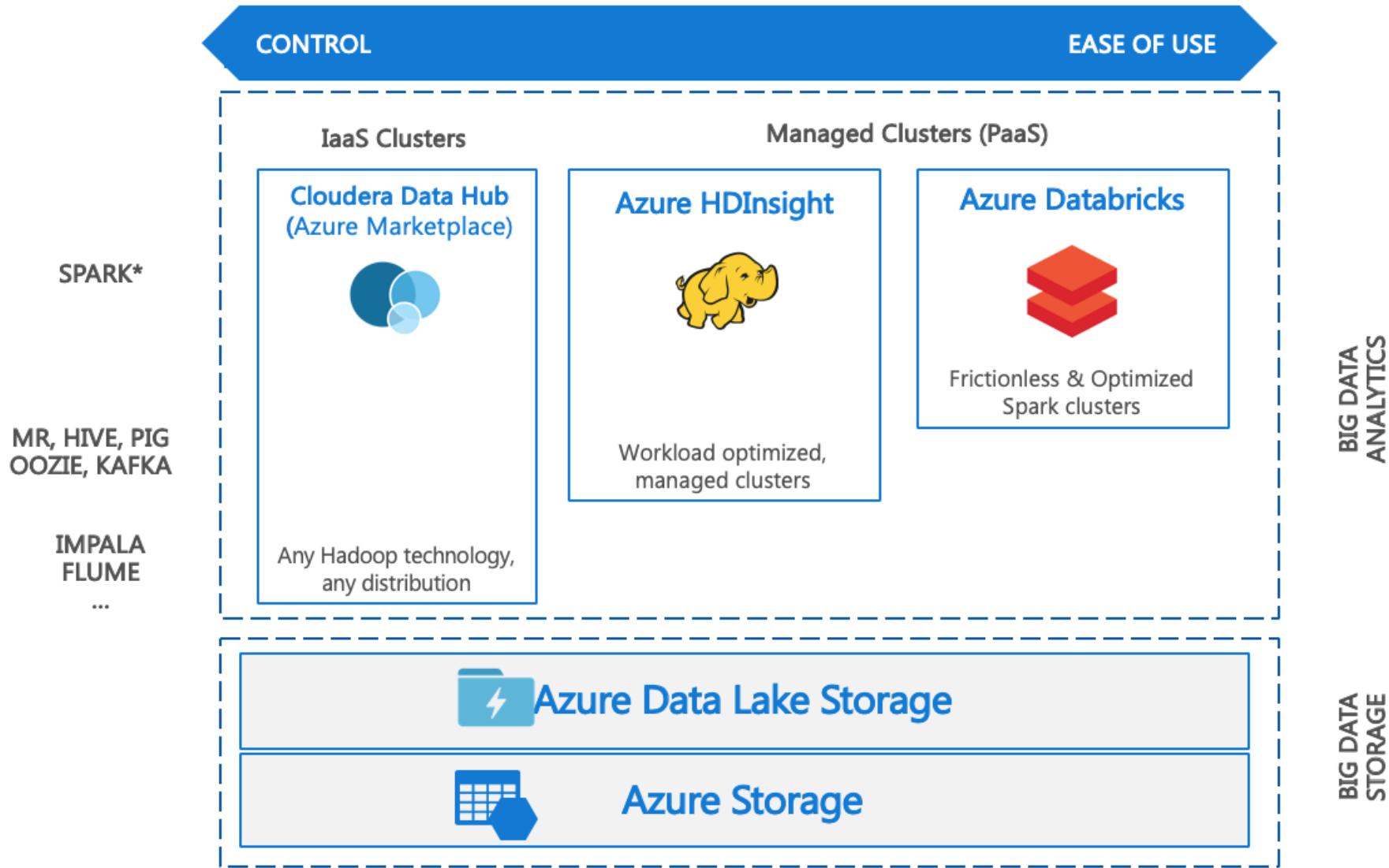


What is Azure HDInsight?

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



Choosing a Big Data Environment



*Spark SQL supports the vast majority of Hive features

Azure Synapse



Synapse Pipelines



Synapse Link



Synapse Studio



Synapse SQL



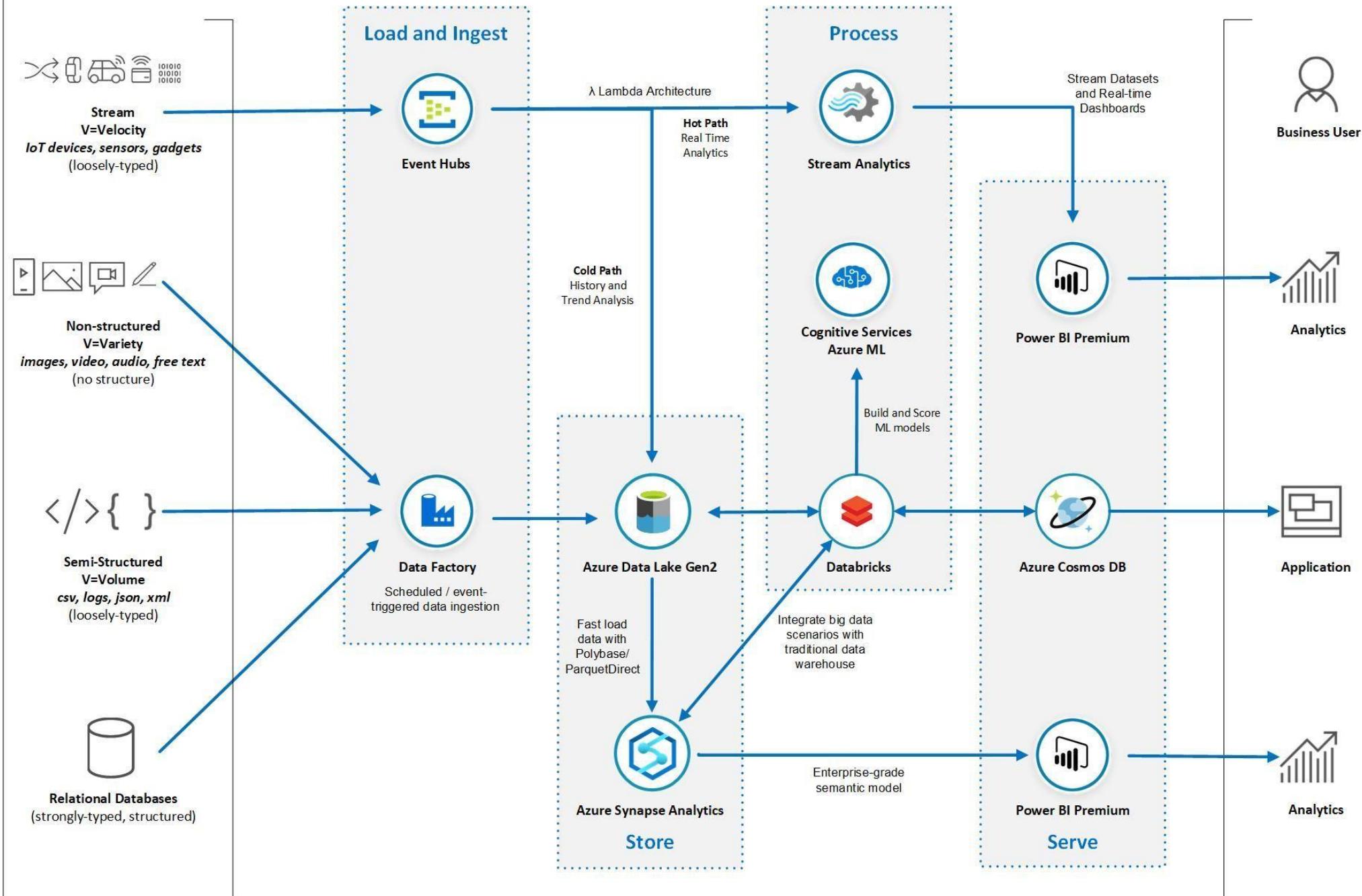
Azure Synapse
Analytics



Synapse Spark

PolyBase

Modern Data Platform Reference Architecture



Lesson 1: Knowledge check



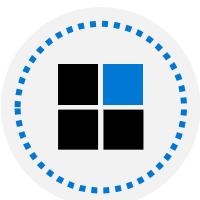
When should you use Azure Synapse Analytics?

- 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

Module 3: Modern Data Warehouse Analytics

1

Examine components of a
Modern Data Warehouse

2

Explore Data Ingestion in
Azure

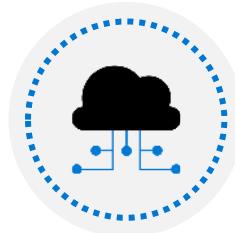
3

Explore Data
Storage and
Processing in
Azure

4

Getting Started
with Power BI

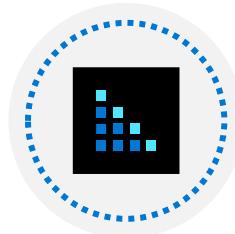
Explore data ingestion in Azure



Describe data ingestion in Azure



Describe components of Azure Data Factory

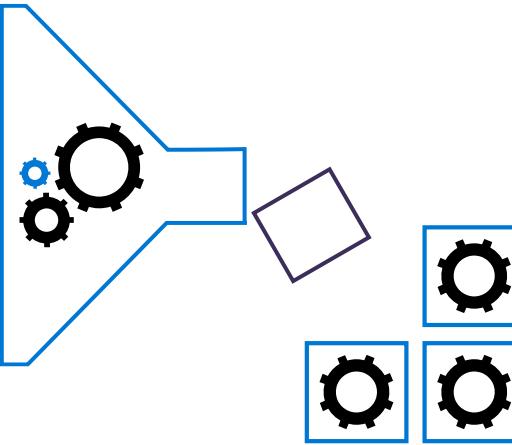


**See how to use Azure Data Factory to load data
into a data warehouse**

Describe data ingestion in Azure

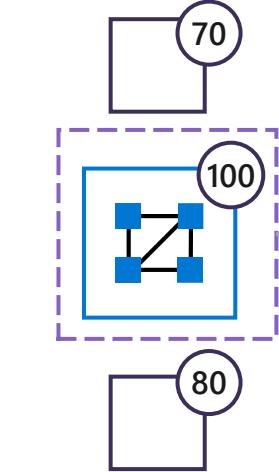
ADF

```
0  
1 1 1  
1 0 0 1 0  
0 1 0 1 0  
1 0 1 0 0  
1 0 0 1 0  
0
```



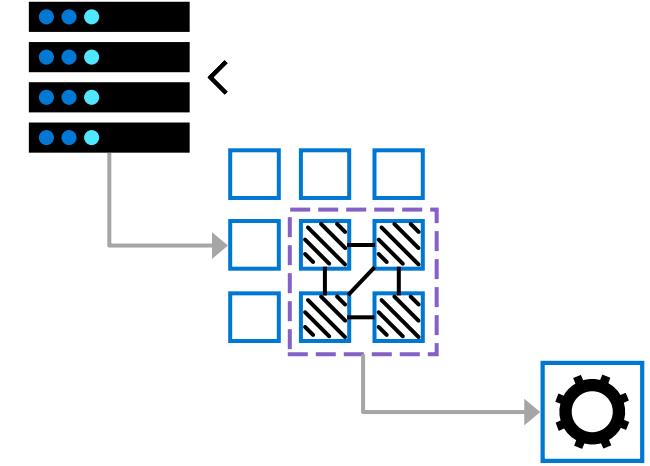
Heterogenous

PolyBase



File based

SSIS



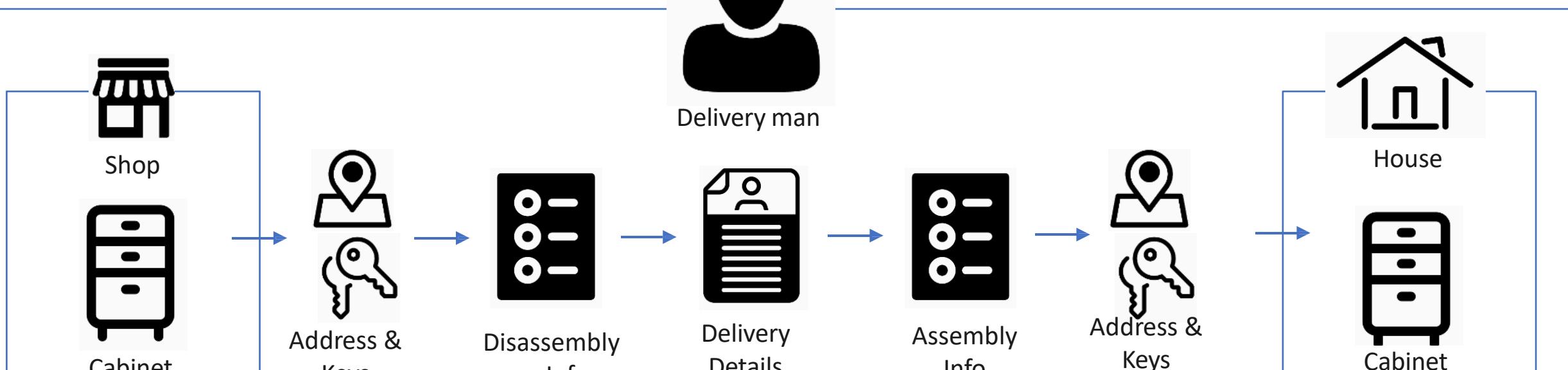
Heterogenous

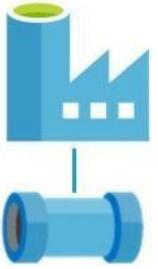


Delivery Manager



Delivery man

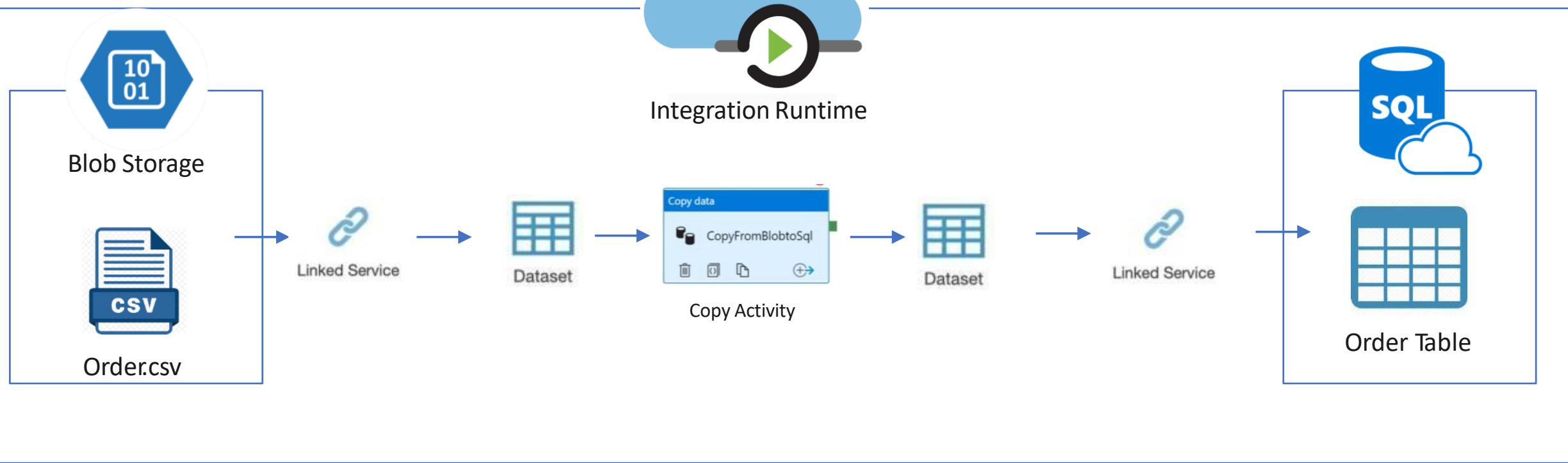




Data Factory Pipeline



Integration Runtime



Data Factory vs SSIS

Azure Data Factory	SSIS
Pipeline	Package
Linked Service	Connection manager
Source	Source
Sink	Destination
Activity	Control flow task
Data Flow	Data flow

ADF Pipelines

Screenshot of the Microsoft Azure Data Factory interface showing the creation and configuration of a pipeline.

The left sidebar shows the navigation menu:

- Home
- Data Factory
- UsmanAhmed-DataFactoryDemo
- Factory Resources
 - Pipelines (1)
 - Datasets (2)
 - AzureSQLTableDest
 - AzureSQLTableSrc
 - Data flows (1)
 - DataFlow
- Connections
- Triggers

The main workspace displays the details for "pipeline1".

Toolbar buttons include:

- Publish all
- Validate all
- Refresh
- Discard all
- Data flow debug (switched on)
- ARM template

Search bar: Search resources

Activity pane:

- Save as template
- Validate
- Debug
- Add trigger

Code button

The pipeline structure is shown as a flow:

```
graph LR; A[Copy data] --> B[Mapping Data Flow]; B --> C[Data Transformation]
```

Activity details:

- General tab:
 - Name: pipeline1
 - Description: (empty)
 - Concurrency: (empty)
 - Annotations: + New
- Parameters tab
- Variables tab
- Output tab

ADF Activities

Microsoft Azure | Data Factory > UsmanAhmed-DataFactoryDemo

Search resources

Data flow debug ARM template

uahmed@alphabold.com ALPHABOLD

Data Factory Publish all Validate all Refresh Discard all

Factory Resources

Pipelines pipeline1

Datasets AzureSQLTableDest AzureSQLTableSrc

Data flows DataFlow

Activities

Move & transform

Azure Data Explorer

Azure Function

Batch Service

Databricks

Data Lake Analytics

General

HDInsight

Iteration & conditionals

Machine Learning

Copy data

Copy Data

Mapping Data Flow

Data Transformation

Save as template Validate Debug Add trigger

Code

Copy data

Copy Data

Mapping Data Flow

Data Transformation

General Parameters Variables Output

Name * pipeline1

Description

Concurrency

Annotations + New

Connections

Triggers

```
graph LR; Start(( )) --> Copy[Copy Data]; Copy --> Mapping[Mapping Data Flow]; Mapping --> Transformation[Data Transformation];
```

ADF Data Flows

Screenshot of the Microsoft Azure Data Factory interface showing the creation of a Data Flow.

The left sidebar shows Factory Resources with:

- Pipelines: 1 (pipeline1)
- Datasets: 2 (AzureSQLTableDest, AzureSQLTableSrc)
- Data flows: 1 (DataFlow, selected and highlighted with a yellow box)

The main workspace displays the DataFlow pipeline1. The pipeline consists of three stages: Source, Transformation, and Sink, connected sequentially. A yellow box highlights the Source stage, which is configured to "Import data from AzureSQLTableSrc".

The DataFlow configuration pane is open, showing the General tab with the following settings:

- Name: DataFlow
- Description: (empty)
- Custom sink ordering:

A red arrow points from the "Name" field back up towards the Source stage of the pipeline diagram.

ADF Datasets

Microsoft Azure | Data Factory ▶ UsmanAhmed-DataFactoryDemo ⌂ ⌃ ⌄ ? uahmed@alphabold.com ALPHABOLD

Data Factory ▾ Publish all ▾ Validate all Refresh Discard all Data flow debug ARM template

Factory Resources ▾ << Filter resources by name +

Pipelines 1 pipeline1

Datasets 2 AzureSQLTableDest AzureSQLTableSrc

Data flows 1 DataFlow

AzureSQLTableSrc X AzureSQLTableDest

Azure SQL Database AzureSQLTableSrc

General Connection Schema Parameters

Name * AzureSQLTableSrc

Description

Annotations + New

Connections

ADF Linked Services

Microsoft Azure | Data Factory ▶ UsmanAhmed-DataFactoryDemo 6 ? uahmed@alphabold.com ALPHABOLD

» Data Factory ▾ Publish all Validate all Refresh Discard all Data flow debug ARM template ...

Factory Resources Filter resources by name + New

- Pipelines** 1
 - pipeline1
- Datasets** ...
 - DestConn
 - SourceConn
- Data flows** 1
 - DataFlow

Connections Search to filter items...

Linked services Integration runtimes

New

Showing 1 - 2 of 2 items

NAME ↑↓	TYPE ↑↓	ANNOTATIONS ↑↓
DestConn	Azure SQL Database	
SourceConn	Azure SQL Database	

ADF Integration Runtimes

Microsoft Azure | Data Factory ▶ UsmanAhmed-DataFactoryDemo ≡ ? uahmed@alphabold.com ALPHABOLD

Data Factory Publish all Validate all Refresh Discard all Data flow debug ARM template

Factory Resources +

Connections

Linked services **Integration runtimes**

+ New Refresh

Showing 1 - 1 of 1 items

NAME ↑↓	TYPE ↑↓	SUB-TYPE ↑↓	STATUS ↑↓	REGION ↑↓
AutoResolveIntegrationRuntime	Azure	Public	Running	Auto Resolve

Pipelines 1
pipeline1

Datasets 2

Data flows 1
DataFlow

Connections

Triggers

ADF Triggers

Microsoft Azure | Data Factory | UsmanAhmed-DataFactoryDemo | Search resources | Publish all 1 | Validate all | Refresh | Discard all | Data flow debug | ARM template | Triggers | New | Filter resources by name | Search to filter items... | + | Pipelines 1 | pipeline1 | Datasets 2 | Data flows 1 | DataFlow | Connections | Triggers

Showing 1 - 1 of 1 items

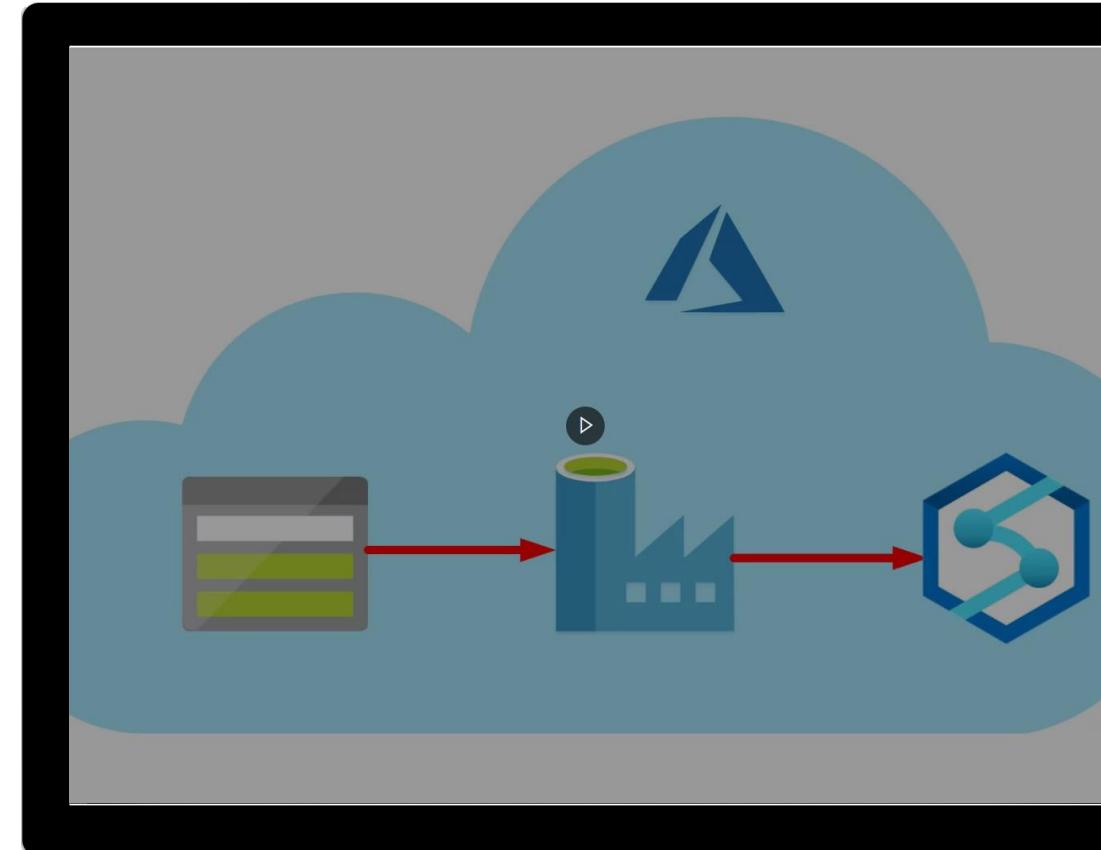
NAME ↑↓	TYPE ↑↓	STATUS ↑↓	NUMBER OF PIPELINES ↑↓	ANNOTATIONS ↑↓
Schedule Run	Schedule	Started	0	

Demo: Load data into Azure Synapse Analytics

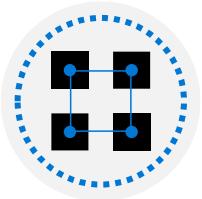
Imagine that you're part of a team that is analyzing house price data. The dataset that you receive contains house price information for several regions. Your team needs to report on how the house prices in each region have varied over the last few months. To achieve this, you need to ingest the data into Azure Synapse Analytics. You've decided to use Azure Data Factory to perform this task

In this video, you'll see how to use Azure Data Factory to ingest and process house price data for analysis

You'll store the data in Azure Synapse Analytics for later analysis

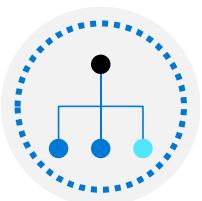


Lesson 2: Knowledge check



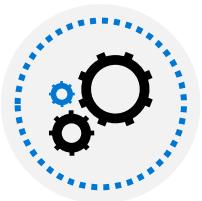
Which component of an Azure Data Factory can be triggered to run data ingestion tasks?

- CSV File
 - Pipeline
 - Linked service
-



When might you use PolyBase?

- To query data from external data sources from Azure Synapse Analytics
 - 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

Module 3: Modern Data Warehouse Analytics

1

Examine components of a
Modern Data Warehouse

2

Explore Data
Ingestion in Azure

3

Explore Data Storage and
Processing in Azure

4

**Getting
Started with
Power BI**

Explore data storage and processing in Azure



Describe data processing options for performing analytics in Azure

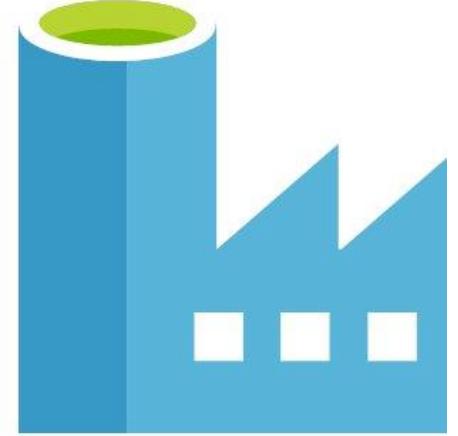
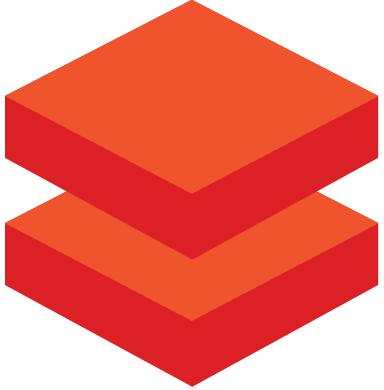


Explore Azure Synapse Analytics

Storage & Processing Services in Azure



Azure Synapse Analytics



Explore Azure Synapse Analytics



Synapse Pipelines



Synapse Link



Synapse Studio



Synapse SQL

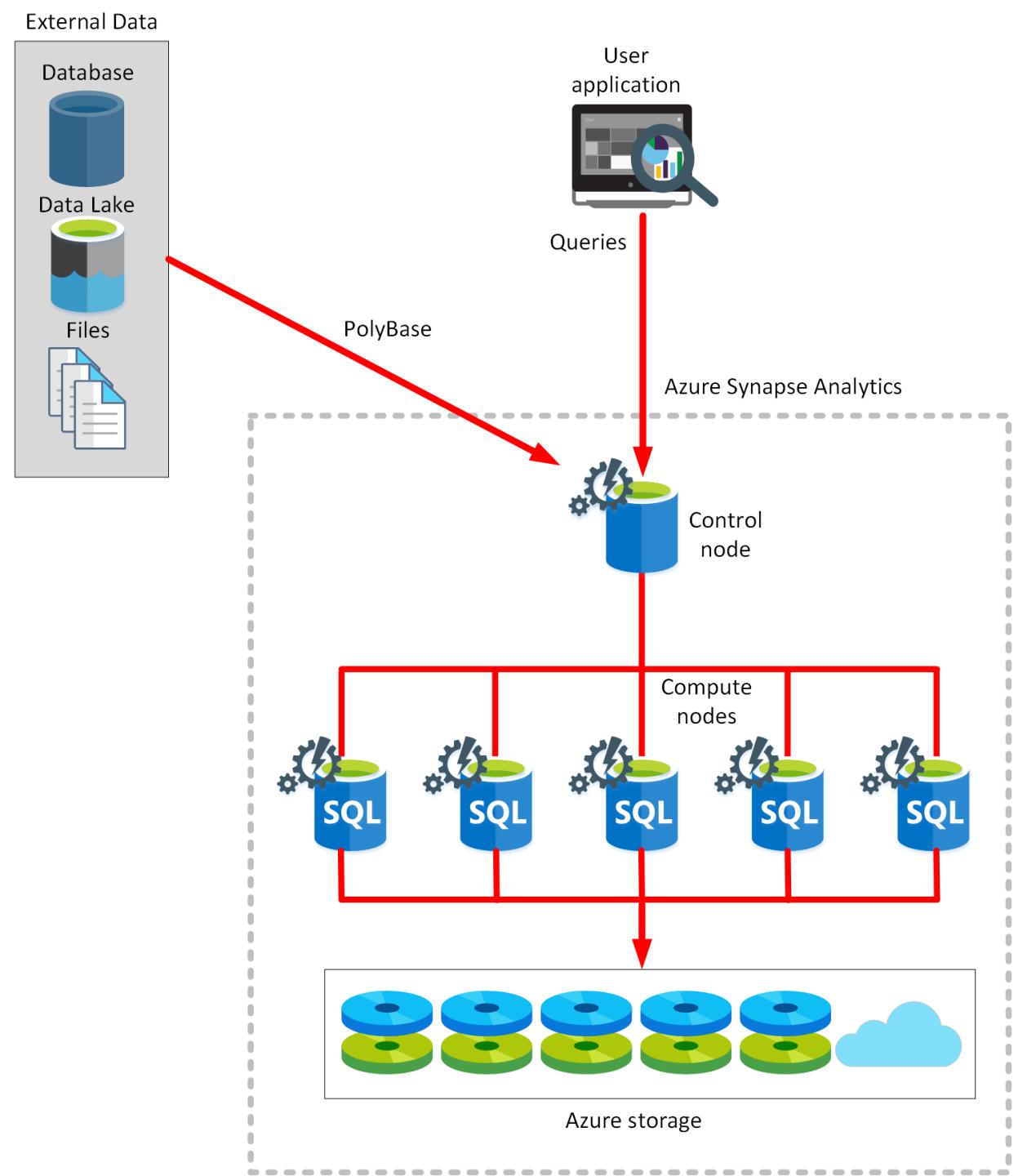


Azure Synapse
Analytics



Synapse Spark

Azure Synapse SQL Pool - MPP

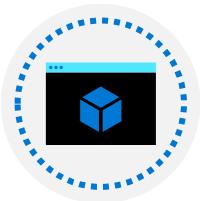


Lesson 3: Knowledge check



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##

Module 3: Modern Data Warehouse Analytics

1

Examine components of a
Modern Data Warehouse

2

Explore Data
Ingestion in Azure

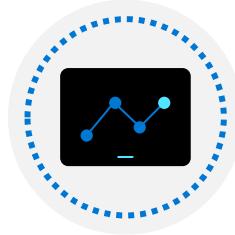
3

Explore Data Storage
and Processing in
Azure

4

Getting Started with
Power BI

Get started building with Power BI



Learn how Power BI services and applications work together



Explore how Power BI can make your business more efficient



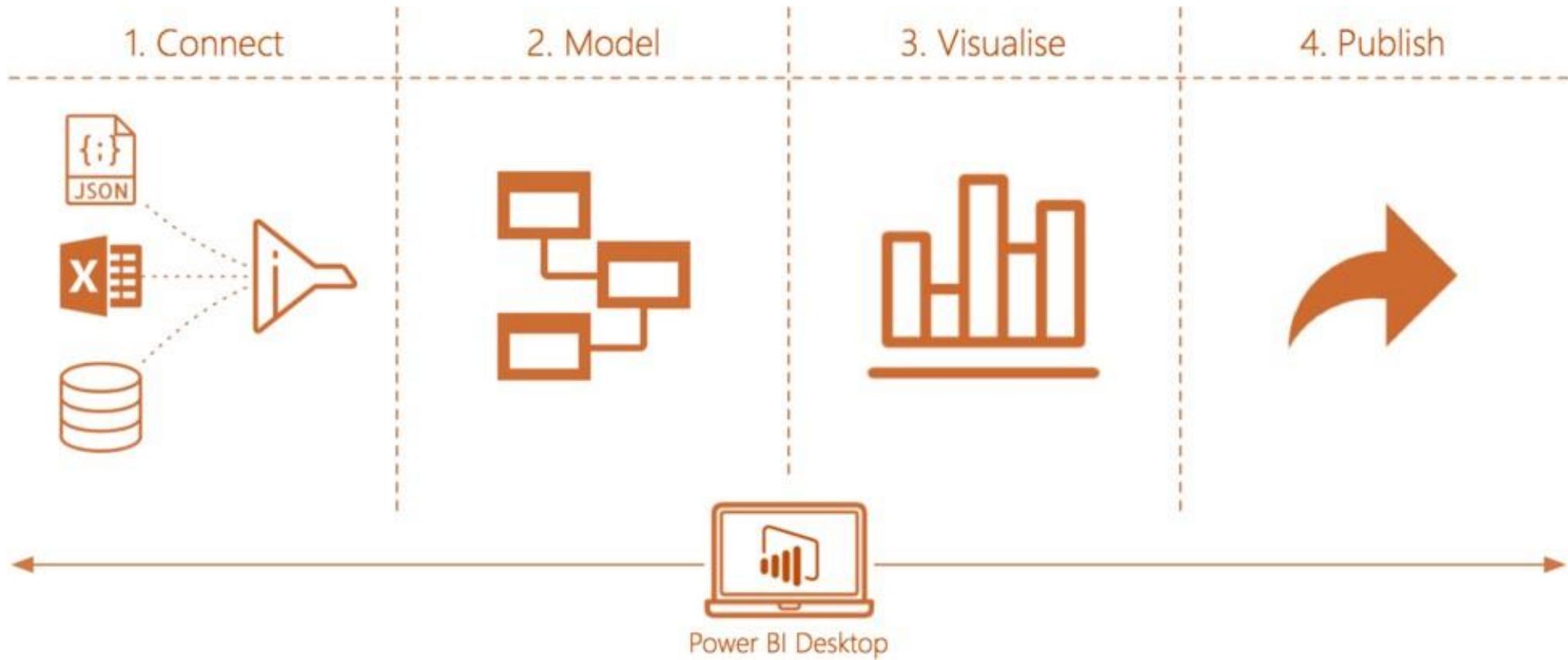
Learn how to create compelling visuals and reports



Power BI



PowerBI Workflow



How can you use Power BI?

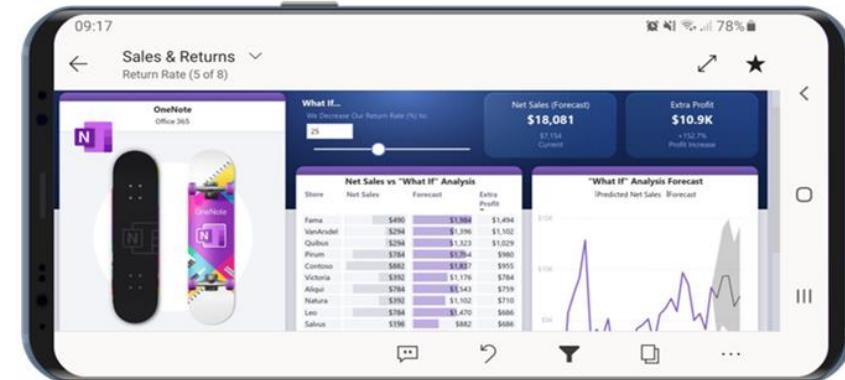
Power BI Desktop



Power BI Service



Power BI Mobile



The anatomy of a Power BI app

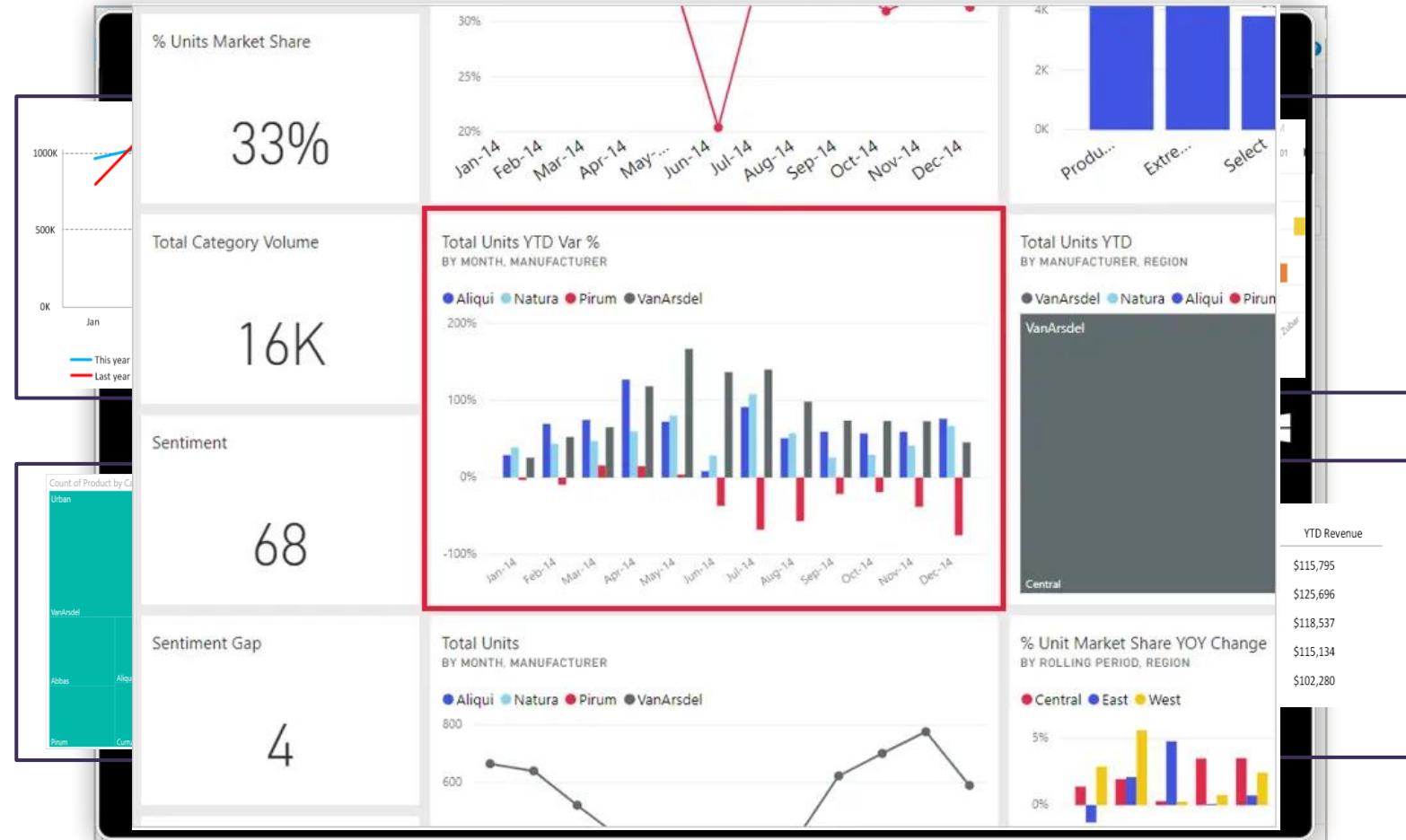
Visualizations

Datasets

Reports

Dashboards

Tiles



Paginated Reports

The screenshot shows a Power BI report titled "Buying Group Account Statement" for the "Tailspin Toys" buying group. The report is displayed in a browser window with a dark theme. At the top, there are navigation buttons for "File", "Export", and "Parameters". The main content area shows the report's title, a date range from May 01, 2016, to May 31, 2016, and a count of 201 customers selected. The report is presented in a tabular format with a header row and several data rows. The total value for the report is \$1,434.05.

Buying Group Account Statement
Invoice Range: May 01, 2016 to May 31, 2016
Buying Group: Tailspin Toys, 201 Customers Selected

Invoice No: 73507-70431

Invoice 2 of 3

Item	Quantity	Unit Price	Tax	LineTotal
DBA joke mug - you might be a DBA if (White)	7	\$13.00	\$13.65	\$104.65
Ogre battery-powered slippers (Green) XL	8	\$32.00	\$38.40	\$294.40
Packing knife with metal insert blade (Yellow) 18mm	15	\$2.40	\$5.40	\$41.40
"The Gu" red shirt XML tag t-shirt (White) 3XS	48	\$18.00	\$129.60	\$993.60
TOTAL			\$187.05	\$1,434.05

Interactive Reports

Sales & Returns Analysis - Power BI Desktop

Miguel Martinez

File Home Insert Modeling View Help

Cut Copy Format painter Paste Get data Excel Power BI datasets Server Enter data Recent sources Transform Refresh data New visual Text box More visuals Insert New measure Quick measure Calculations Publish

Microsoft | Skateboard Store

Last Refresh: Jun 30th, 2019 / Chicago, IL, USA

What If... We Decrease Our Return Rate (%) to: 25

Net Sales (Forecast) \$30,772 Extra Profit \$0 0.0% Profit Increase

Return Rate 24%

Net Sales (Forecast) by Location

Park Ridge Elmhurst Park Cicero Burbsk Oak Park Skokie Chicago Wilmette Skokie Lincoln Park

"What If" Analysis Forecast Predicted Net Sales Forecast

Jan 2015 Mar 2019 May 2019 Jul 2019 Date

Visualizations Fields

Search

Analysis DAX Design DAX % Return Rate Age Associated Product Association Calendar Customer Details Issues and Promotions Product Sales Stable Store Tooltip Info Tooltip Info2

Add data fields here

Values

Product is OneNote

Drillthrough

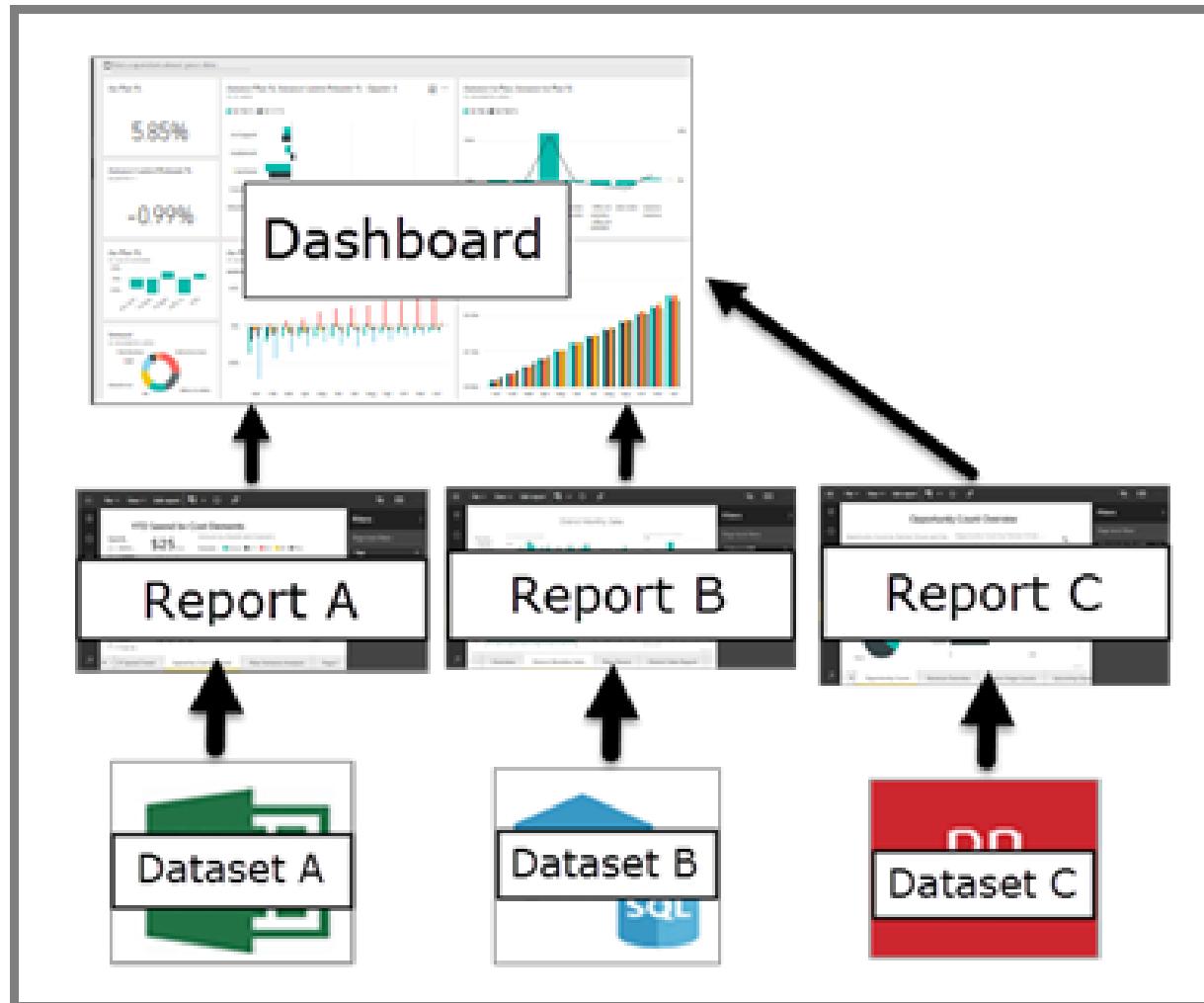
Cross-report Off

Keep all filters Off

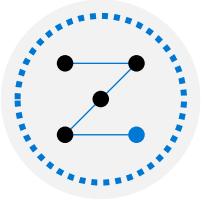
Product is OneNote

PAGE 5 OF 18

Dashboard Reports

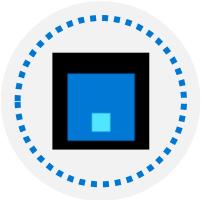


Lesson 4: Knowledge check



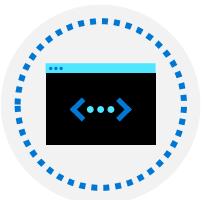
What is the common flow of activity in Power BI?

- 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