Azure Data s AI Foundations: From ETL to AI
Day-02

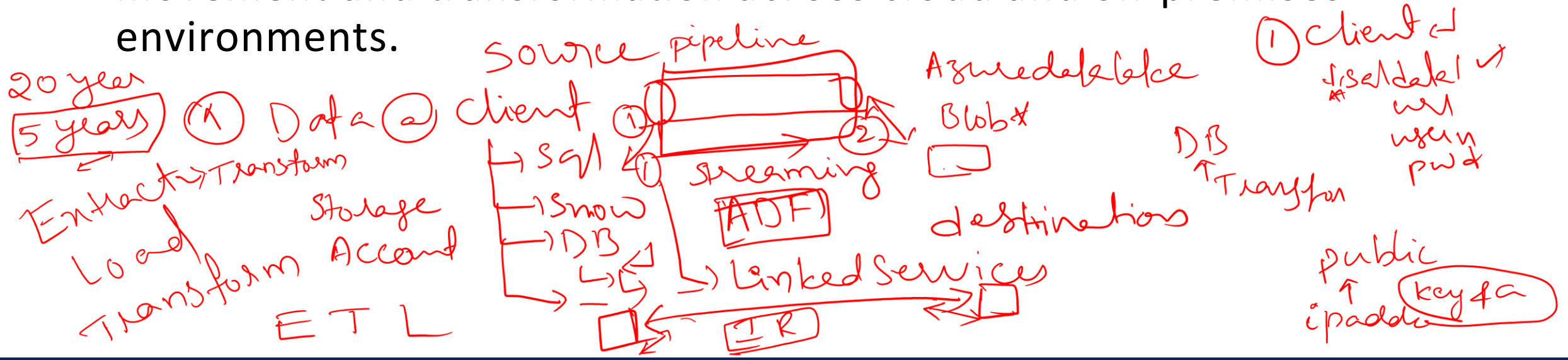
Presenter: Dr. Shiva prasad Koyyada

Slides from:Y. M. Yugandhar Reddy

© NexusIQ Solutions

Introduction to Azure Data Factory

Welcome to the session on **Azure Data Factory (ADF)**. In this session, we will explore ADF concepts, components, and use cases, followed by a hands-on lab to build an ETL pipeline. By the end of this session, you will understand how ADF helps in data movement and transformation across cloud and on-premises environments.



© NexusIQ Solutions

What is Azure Data Factory?

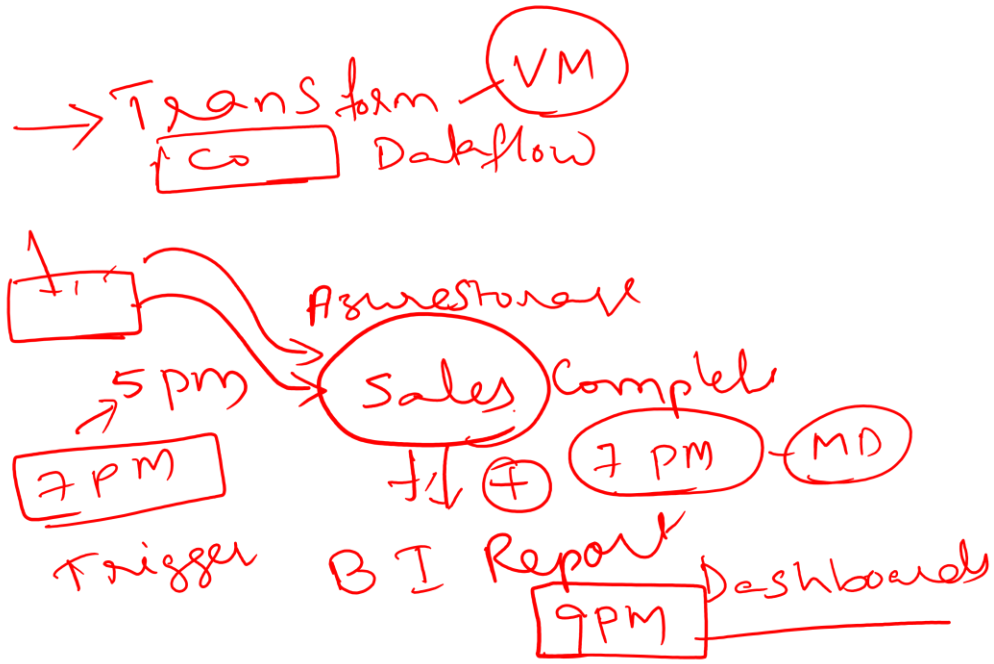
Azure Data Factory (ADF) is a cloud-based data integration service that enables data movement and transformation at scale.

Key Features:

- Fully managed ETL (Extract, Transform, Load) service.
- Supports on-premises and cloud data sources.
- Serverless data transformation using **Mapping Data Flow**.
- Built-in monitoring and logging for pipeline executions.

Common Use Cases:

- Data migration from on-premises to the cloud.
- Data ingestion from multiple sources into a data lake.
- Automated data processing for analytics and reporting.
- ETL and ELT workflows for business intelligence.



© NexusIQ Solutions

Core Components of Azure Data Factory

1. Pipelines:

- A logical grouping of activities that perform a data movement or transformation task.

2. Dataflows:

- Enables data transformation using a visual interface without writing code.

3. Datasets:

- Represents data structures within ADF that define sources and destinations.

© NexusIQ Solutions

Core Components of Azure Data Factory

4. Linked Services:

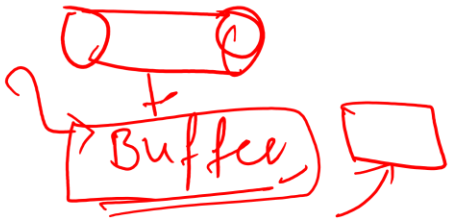
- Connects ADF to data sources like Azure Blob Storage, Azure SQL Database, and on-premises systems.

5. Integration Runtimes:

- **Azure IR:** For cloud-based data movement.
- **Self-Hosted IR:** For on-premises data integration.
- **SSIS IR:** To run SQL Server Integration Services packages in ADF.

6. Triggers:

- Automates pipeline execution based on schedules or events.



© NexusIQ Solutions

Integrating ADF with On-Prem s Cloud Data Sources

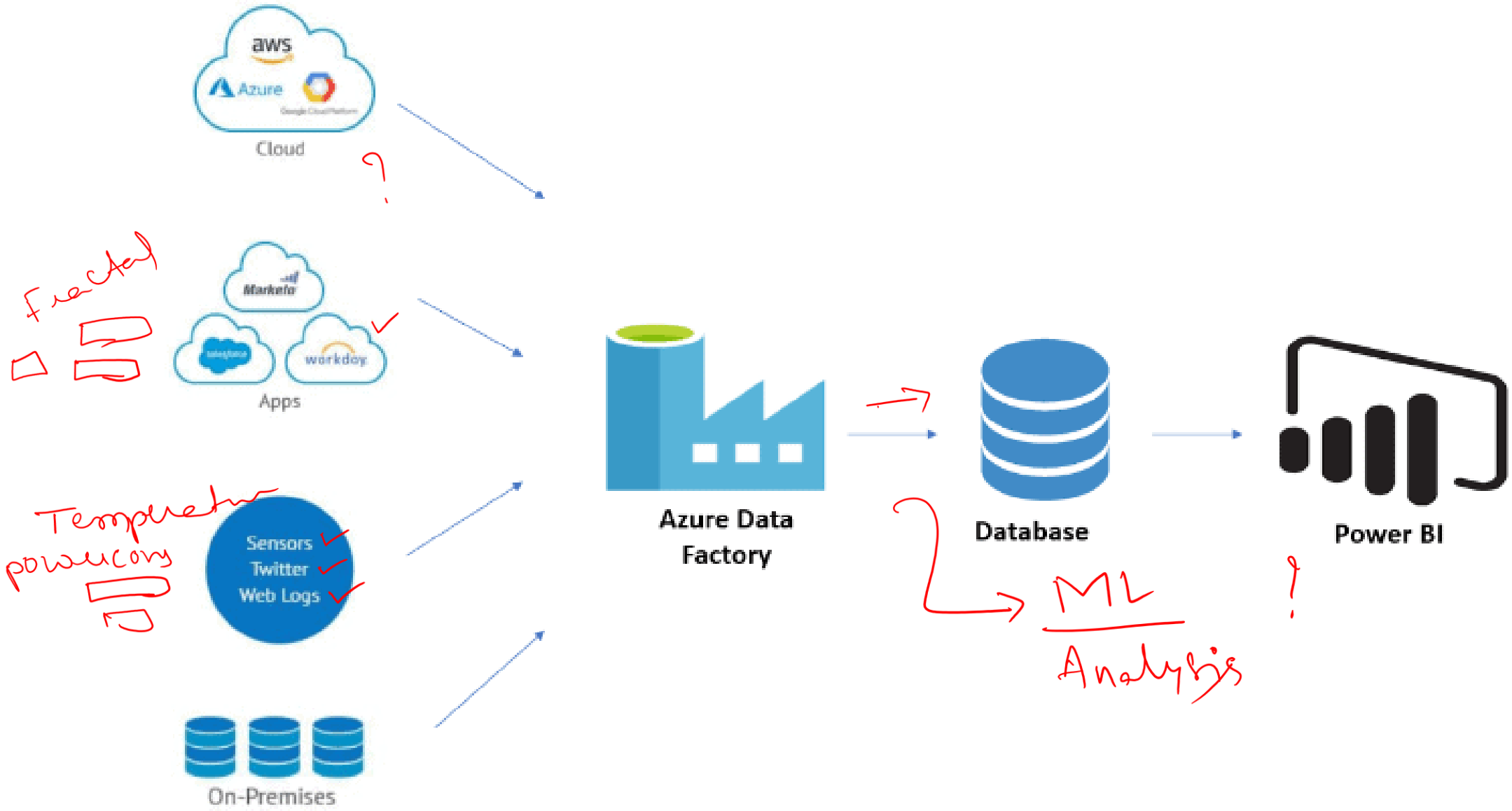
Supported Data Sources:

- **Cloud:** Azure SQL Database, Blob Storage, Data Lake, Cosmos DB, Amazon S3, Google BigQuery, etc.
- **On-Premises:** SQL Server, Oracle, MySQL, SAP, and File Systems.

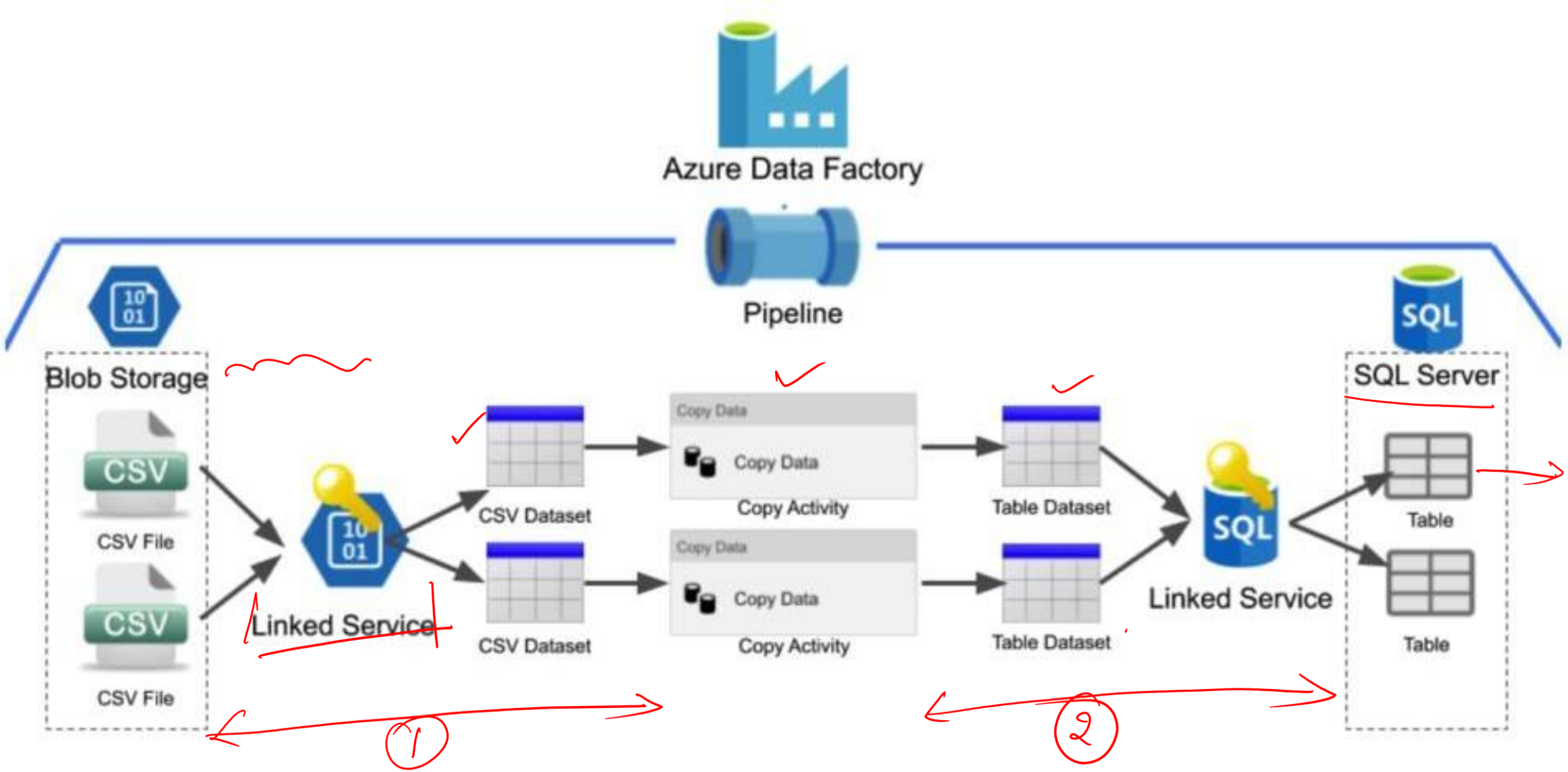
Integration Methods:

- **Self-hosted Integration Runtime (IR)** to access on-premises databases securely.
- **Azure Data Gateway** for secure cloud-to-on-premises connectivity.

© NexusIQ Solutions



© NexusIQ Solutions



© NexusIQ Solutions

Hands-on Lab - Build an ETL Pipeline in ADF

Objective: Build a pipeline to move data from **Azure Blob Storage** to **Azure SQL Database** and apply transformations.

Step 1: Connect ADF to Azure SQL Database and Blob Storage

- Create Linked Services for Blob Storage and Azure SQL Database.
- Define Datasets for source (Blob) and destination (SQL Database).

Step 2: Create a Data Pipeline to Move Data from Blob to SQL

- Add **Copy Data Activity** to transfer data from Blob Storage to Azure SQL Database.
- Configure source and destination datasets.
- Enable logging and monitoring.

© NexusIQ Solutions

Hands-on Lab - Build an ETL Pipeline in ADF

Step 3: Apply Data Transformation using Mapping Data Flow

- Add **Mapping Data Flow** activity.
- Perform data cleansing, filtering, and aggregations.
- Load transformed data into the Azure SQL Database.

Step 4: Use Triggers and Monitor Pipeline Execution

- Configure a **Scheduled Trigger** to automate pipeline execution.
- Use **ADF Monitoring** to track pipeline execution status.
- Debug errors and optimize performance.

© NexusIQ Solutions

© NexusIQ Solutions