INTRODUCTION

Integrating Data in Microsoft Azure

DATA INTEGRATION SERVICES IN AZURE

Marcelo Pastorino
SOFTWARE DEVELOPER / SOLUTIONS ARCHITECT

@evangeloper   softwaredeveloper.io/marcelo

Data Integration

Process of combining data from different data sources into meaningful and valuable information.

Integrate Data in Azure

ETL
Extract, Transform and Load

ELT
Extract, Load and Transform

3 "simple" steps
1. Migrate data
2. Transform it
3. Store it!

Azure Data Integration Services

Azure Data Factory

Azure Databricks

Azure EventHubs

Azure Stream Analytics

Power BI

CLIP 1

Azure Data Factory

Hybrid data integration service

Simplifies ETL at scale

Enables modern data integration

Drag and drop interface

Over 80 connectors available

Move, transform and save data

On-premise data stores

Different cloud providers supported

Data Migration Assistant

Good for on-premise SQL Server migrations

Helps migrating schema and data

And to detect compatibility issues between data sources

Azure Database Migration Service

Recommended for large migrations

Fully managed

Helps migrate databases at scale

Minimal downtime

Azure Data Factory

Integrate data from disparate data sources

Multi-cloud ready
- AWS, GCP, Azure

Copy Data Tool

Part of Azure Data Factory

Eases and optimizes the process of ingesting data

Azure EventHubs

Real-time data ingestion

Fully managed

Azure Stream Analytics

Serverless real-time analytics platform

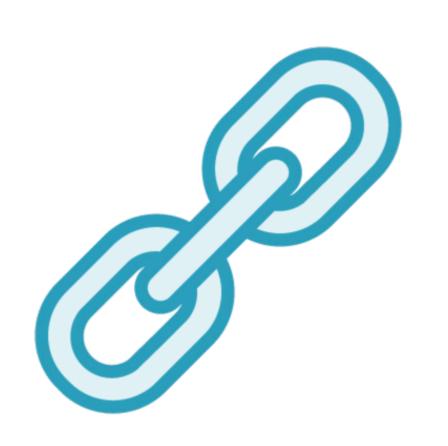**Helps to build an end-to-end streaming pipeline**

Familiar SQL syntax

Azure Databricks

Collaborative Apache Spark-based analytics service

Azure Cloud Services

https://azure.microsoft.com/en-in/services

CLIP 2

A Startup Story

- Collects environmental data

- IoT devices deployed around the world

- Devices with sensors to measure air pollution data

- The goal is to alert individuals and city officials

Proof of Concept Phase

10 selected cities
Devices installed in just 10 selected cities

On-Premise
Leveraged on-premise infrastructure only

Amazon S3
Migrated partial pollution data to Amazon S3

The Cloud Phase

Expand to several countries around the world

Extending on-premise infrastructure unviable

Decided to move the infrastructure to Azure

Want to use Azure Data Integration services

Requirements to Fulfill

Migrate data from on-premise SQL Server

Migrate data from Amazon S3

Transform and enrich data

Send email notifications

Create a real-time data pipeline

Migrate Data on-premise Data

From on-premise SQL server to Azure SQL Database

Ensure data integrity

Ensure server compatibility

Use Data Migration Assistant

Some sensors still pushing data to on-premise infrastructure

Need to migrate data incrementally

Need to migrate existing unprocessed pollution data from on-premise SQL Server

Migrate Data from Amazon S3

Migrate unprocessed pollution data stored in Amazon S3

Migrate it to Azure Blob Storage

Transform and Enrich Data

We need to transform and enrich data

We also need to store this data in Azure SQL Database

Use Azure Data Factory and Azure Databricks to fulfill this requirement We need to transform and enrich data

Send Email Notifications

We need to send pipeline execution status notifications

Integrate Azure LogicApps in Data Factory pipelines to fulfill this requirement

Create a Real-time Data Pipeline

Create a service to ingest a large number of concurrent events

Use Azure EventHubs to fulfill this request

Use Azure Stream Analytics to process and analyze events in real-time

Stream real-time data to Power BI

Persist event in Azure Blob Storage

Requirements to Fulfill

Migrate data from on-premise SQL Server

Migrate data from Amazon S3

Transform and enrich data

Send email notifications

Create a real-time data pipeline

SUMMARY

Azure Data Factory

Data integration service that simplifies ETL at scale.