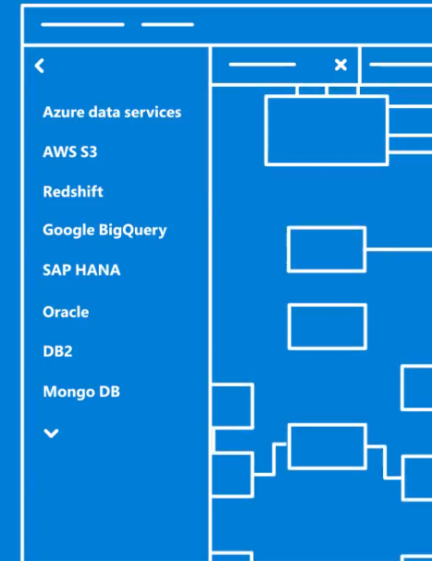
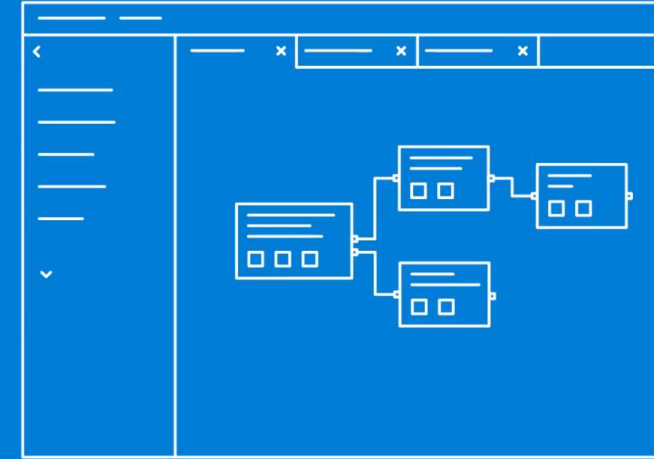


Azure Data Factory



Dishan Rajapaksha | Sameendra Chaturanga
29th May, 2020

Agenda

What is Azure Data Factory?

What are ADF components?

How to build dynamic data pipelines with ADF?

Introduction

Big data

Raw & unorganized data doesn't have context or meaning to provide insights

Extract-Transform-Load

Analyse

Workflow automation

Azure Data Factory

Azure Data Factory

Cloud data integration & transformation service

Complex ETL processes using Databricks, HDInsight or Synapse Analytics

Codeless

Scalable

Automation

Azure Data Factory

Lift and Shift SSIS packages

Cost-efficient and fully managed

Integration with Azure Databricks for transformations

CI/CD integration

How does it work?



Ingest



Transform



CI/CD & Publish



Monitor

Ingest

Ingest data from multiple data sources without custom data movement components

Datasets & linked services

Multi-cloud & on-premise hybrid copy into a centralized data store

Transform

Transform collected data using mapping data flows

Use Hadoop, Spark and other external activities

Code free data pipelines to control the flow

Utilize workflow constructors such as loops & conditions

SSIS integration runtime for SSIS packages

CI/CD & Publish

Integrate into existing DevOps processes

Incremental development

Use Git for version controlling

Monitor

Monitor scheduled activities

Use Azure Monitor & health panels

Proactively monitor using alerts

Use metrics for pipeline analysis

Use diagnostic logs with Azure Log Analytics

Azure Data Factory Components

Activity

Single processing step

Data movement activities

Data transformation activities

Control activities

Pipeline

Logical grouping of activities

Activates can be independent, sequential or parallel

Control Flow

Orchestration of activities & pipelines

Workflow constructors

- Loops

- Branches

- Conditional execution

Parameters & Variables

Parameters are defined in the pipeline

Arguments for parameters can be passed during execution

Activities within the pipeline consume parameters

Variables can store temporary values

Pass values between pipelines, data flows & other activities using variables

Dataset

Represent data structures within the data stores

Reference to data to be used in activities

Point activity outputs

Linked Services

Defines the connection to external data sources

Data store links

- SQL databases

- Blob storage

- File share

Compute resource links

- Azure machine learning activities

- HDInsight Spark activities

- Stored procedures

Data flows

Mapping & Wrangling

Graphical data transformation

- Lookups

- Aggregates

- Filter

- Sort

- Exists

Stored Procedures, Data Bricks, Spark, Hadoop, Functions etc.

Scale-out with Azure Integration Runtime

Integration runtime

Computer environment for ADF

- Data Flow execution

- Activity dispatching

- Data movement

- SSIS package execution

Azure

Self-hosted

Azure-SSIS

Debugging

Debug variables & parameters for dynamic content

Data preview

Pipeline & activity error messages

Diagnostic logs

Execution

On-demand

- API

- PowerShell

- Portal

- SDKs

Triggers

- Wall clock schedule

- Tumbling windows

- Event based

Building Dynamic Data Pipelines with Azure Data Factory

Demo

Questions?

Thank You