

unit it



Mikkel Kjærulff Leth
Dataplatform Specialist

Data factory for begyndere

Teknisk



Agenda

01 Hvad er data factory?

02 Funktioner i data factory

03 Demo

Hvem er jeg?

Mikkel Kjærulff Leth

28 år

Bor i Aarhus

Nørd både professionelt og i fritiden

Dataplatform specialist

6 år i SQL verdenen (på vej mod skyen)





Hvad er data factory?

- Orkestrering af dataflows, azure komponenter, services og meget mere
- ETL uden kode
- 100+ native connectors

Code-Free ETL as a service

Ingest



- Multi-cloud and on-premise hybrid copy data
- 100+ native connectors
- Serverless and auto-scale
- Use wizard for quick copy jobs

Control Flow



- Design code-free data pipelines
- Generate pipelines via SDK
- Utilize workflow constructs: loops, branches, conditional execution, variables, parameters, ...

Data Flow



- Code-free data transformations that execute in Spark
- Scale-out with Azure Integration Runtimes
- Generate data flows via SDK
- Designers for data engineers and data analysts

Schedule



- Build and maintain operational schedules for your data pipelines
- Wall clock, event-based, tumbling windows, chained

Monitor



- View active executions and pipeline history
- Detail activity and data flow executions
- Establish alerts and notifications

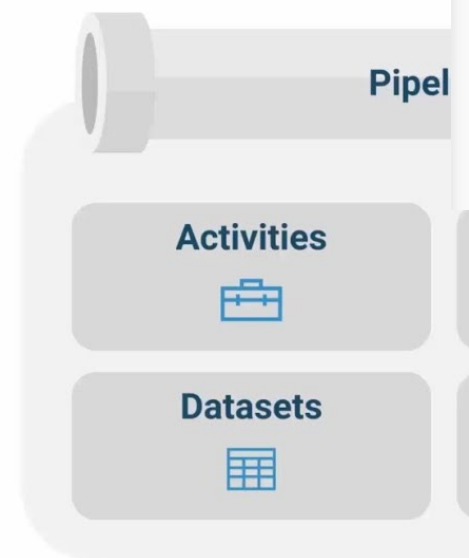
Hvad er data factory?

- Pipelines
- Activities
- Datasets
- Linked services
- Data Flows
- Integration Runtimes

Pipeline og Activities

Funk data

Azure Data Factory



Activities



Data Flows



Triggers



Datasets



Linked Services



Integration Runtimes



Factory Resources

- Pipeline (2)
 - pipeline1
 - SendEmail
- Dataset (0)
- Data flows (0)
- Power Query (0)

Activities

- Move & transform
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks
- Data Lake Analytics
- General
- HDInsight
- Iteration & conditionals
- Machine Learning
- Power Query

pipeline1

Validate Debug Add trigger

Settings

Concurrency ①

Elapsed time metric ① ☐ Emit a metric after specified duration

0.00:10:00

Properties

General Related

Name *

pipeline1

Description

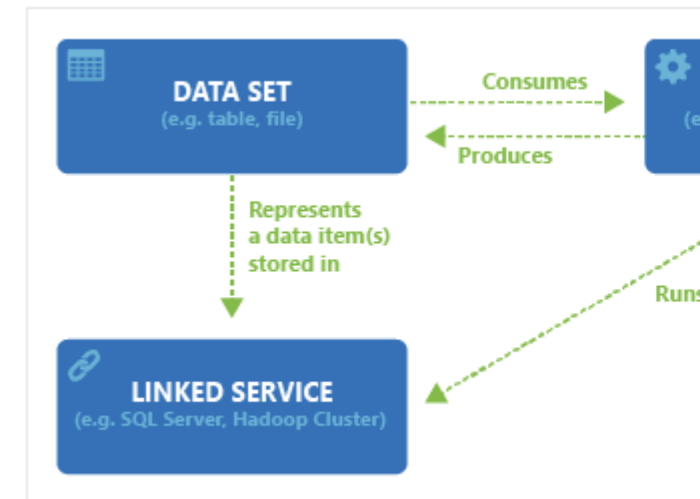
Annotations

+ New

Funktioner i data factory

Linked Services

- Connection til data
- Tænk på det som dine data sources



Edit linked service

Azure SQL Database [Learn more](#)

Name *
AzureSqlDatabase

Description

Connect via integration runtime * ^①
☒ AutoResolveIntegrationRuntime-2

Connection string **Azure Key Vault**

Account selection method ^①
☐ From Azure subscription ☒ Enter manually

Fully qualified domain name *
adf.database.windows.net

Database name *
ADF

Authentication type *
SQL authentication

User name *
mester

Password **Azure Key Vault**

Password *
.....

Always encrypted ^① ☐

Additional connection properties
[+ New](#)

Annotations
[+ New](#)

> Parameters

> Advanced ^①

[Apply](#) [Cancel](#) [Test connection](#)

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Data Factory
Validate all
Publish all

Factory Resources

Filter resources by name

- Pipelines 4
- Change Data Capture (preview) 0
- Datasets 6
- Data flows 1
 - dataflow1
- Power Query 0

Copytestdataentity
Copytestdata
CopyreusePipeline
Dataflow
ADLS_Sinkreusable
ADLS_TSV_Adventur...

Validate
Data flow debug

Product
Update Product

ProductSelect
Renaming Product to ProductSelect with columns 'ProductID, Product, Subcategoryid'

lookupProductSubcata...
Lookup on 'ProductSelect' from 'lookupProductCategory'

Removeduplicaterows
Renaming lookupProductSubcategory to Removeduplicaterows with columns 'ProductID, Product,

sink1
Add sink dataset

lookupProductCatego...

ProductSubcategory
Import data from ADLS_TSV_AdventureWorks

ProductSubcategorySel...
Renaming ProductSubcategory to ProductSubcategorySelect with columns 'Subcategoryid,

lookupProductCategory
Lookup on 'ProductSubcategorySelect' from 'Productcatagoryselect'

ProductCategory
Import data from ADLS_TSV_AdventureWorks

Productcatagoryselect
Renaming ProductCategory to Productcatagoryselect with columns 'Categoryid, Category'

Add Source

Parameters
Settings

+ New

Integration Runtimes

- Computerkraften bag infrastrukturen, der tillader ETL processor.

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

»

Home

Dashboards

Runs

Pipeline runs

Trigger runs

Change Data Capture (previ...

Runtimes & sessions

Integration runtimes

Data flow debug

Notifications

Alerts & metrics

«

Integration runtimes

All

Self-Hosted

Azure

Azure-SSIS

Airflow

Refresh

Edit columns

Time zone : Brussels, Copenhagen, Madrid, P...

Showing 1 - 3 of 3 items

Name ↑↓	Type ↑↓	Sub-type ↑↓	Status ↑↓	Region ↑↓	Created on ↑↓
AutoResolveIntegrationRuntime	Azure	Managed Virtual Network	Running	Auto Resolve	8/21/2023, 3:28:14 PM
AutoResolveIntegrationRuntime-2	Azure	Public	Running	Auto Resolve	---
CKS-IR	Self-Hosted	---	Unavailable More	---	8/22/2023, 8:59:44 AM

Demo

Demo tid

- Copy pipeline
 - Linked service
 - Dataset
 - Pipeline
 - Activities
- Parameter
 - Lave parameter ved copy job
 - Kalde pipeline i pipeline og brug af meta data
- Data flows
 - Opbygning af simpel transformation