



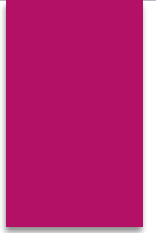
Microsoft Fabric and Azure Data Factory Revealed

DARIUS LIKTORIUS

Please Silence
Your Devices



Introductions



My Background

- ▶ Senior Director, Solutions Architecture @ PwC
- ▶ Providing strategic and tactical leadership to create an AI-enabled next generation global auditing platform
- ▶ Certified on Microsoft Data Products and Azure – Solutions Architect & DevOps Engineer Expert
- ▶ MCSE, MCSD, MCDBA & MCT – since 1996
- ▶ Hobby: Professional Photography

About You

- ▶ DBAs?
- ▶ DB or BI Developers?
- ▶ Who's used Azure Data Factory? MS Fabric?

Our Agenda

- ▶ Data Engineering Primer
- ▶ Intro to Data Factory (Azure & Fabric)
- ▶ Components – Defined & Compared
- ▶ Demo: Azure Data Factory
- ▶ Demo: Data Factory in Fabric
- ▶ Pricing
- ▶ Q & A

Data Engineering Primer

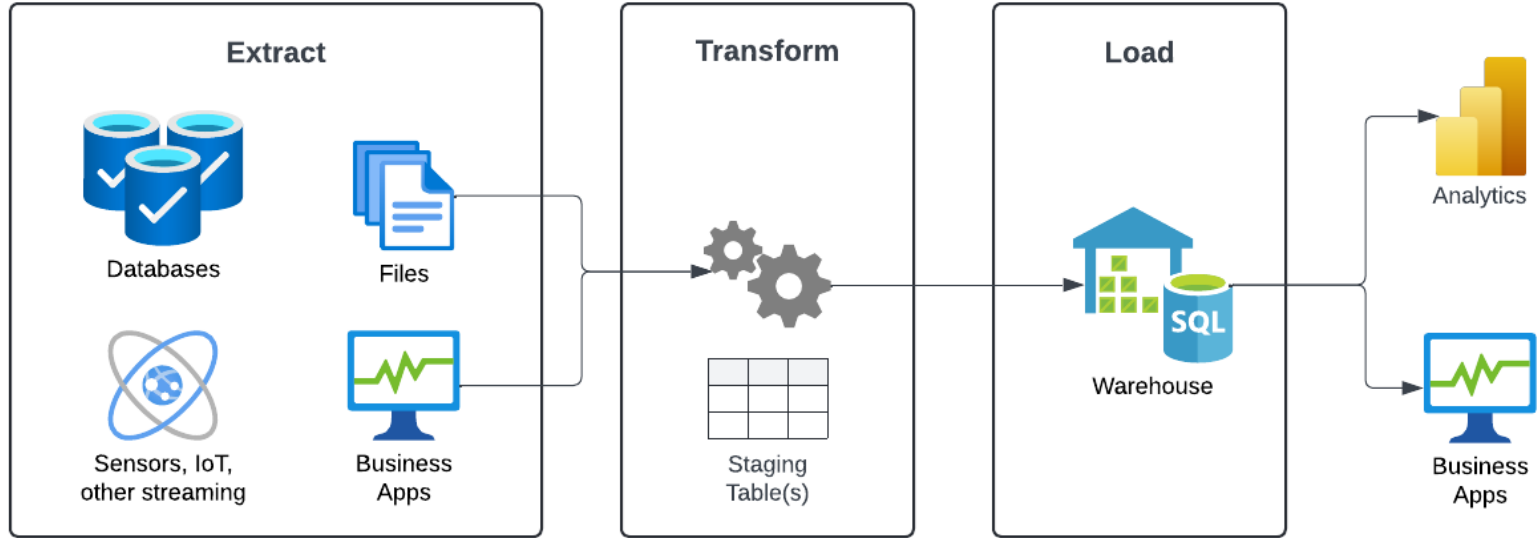


Data Engineering Terminology

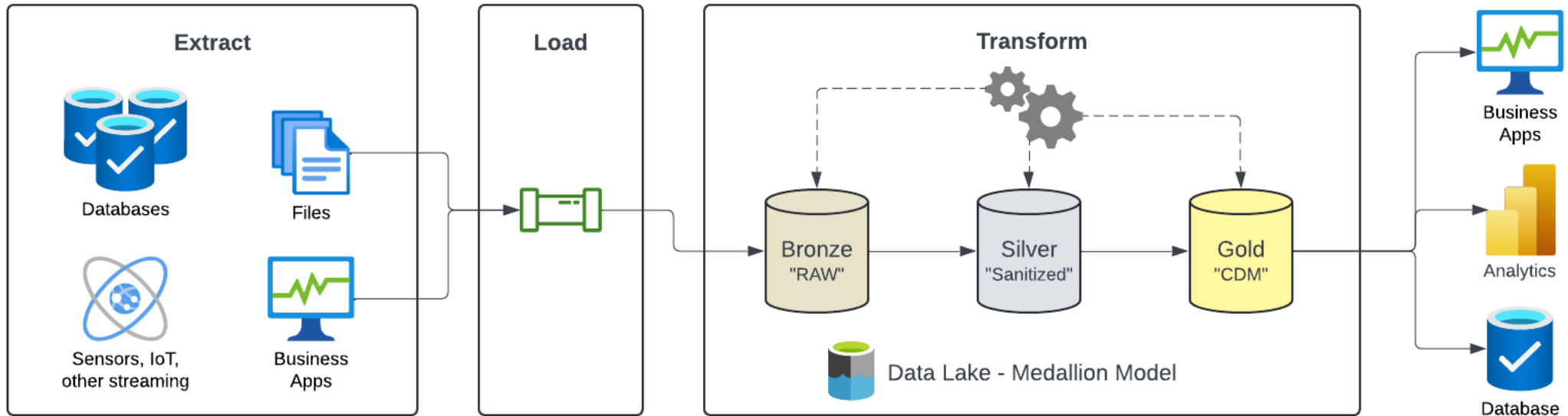
- ▶ Extract, Transform & Load (ETL)
- ▶ Extract, Load & Transform (ELT)
- ▶ Data Mapping & Wrangling
- ▶ Relational Databases
- ▶ Data Lake & Medallion Model

Fabric and Azure Data Factory Revealed

ETL



ELT





Intro to Data Factory (Azure & Fabric)

What is Data Factory?

- ▶ Microsoft Data Platform offering
 - ▶ Runs in the Cloud but Hybrid with “on-prem” features
- ▶ Create pipelines to Copy and Transform your data
 - ▶ ETL (Extract Transform and Load)
 - ▶ ELT (Extract Load and Transform)
- ▶ Numerous integrations & connectivity options

Two Flavors

- ▶ **Azure Data Factory (ADF)** – Established
 - ▶ Platform as a Service (PaaS) offering
 - ▶ Additional integrations and configurability
- ▶ **Data Factory in Fabric (“DFF”)** – Recent
 - ▶ Software as a Service (SaaS) offering
 - ▶ Fabric service & interface with reduced learning curve

Microsoft Fabric - Overview

Intelligent data foundation



Data
Factory



Data
Engineering



Data
Warehouse



Data
Science



Real-Time
Intelligence



Power BI



Data
Activator



Powered by AI with Copilot in Microsoft Fabric



Single data foundation

OneLake

UNIFIED

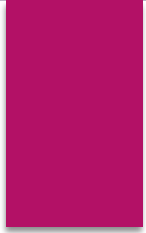
SaaS product experience

Security and governance

Compute and storage

Business model

Data Factory Components



Components of Data Factory

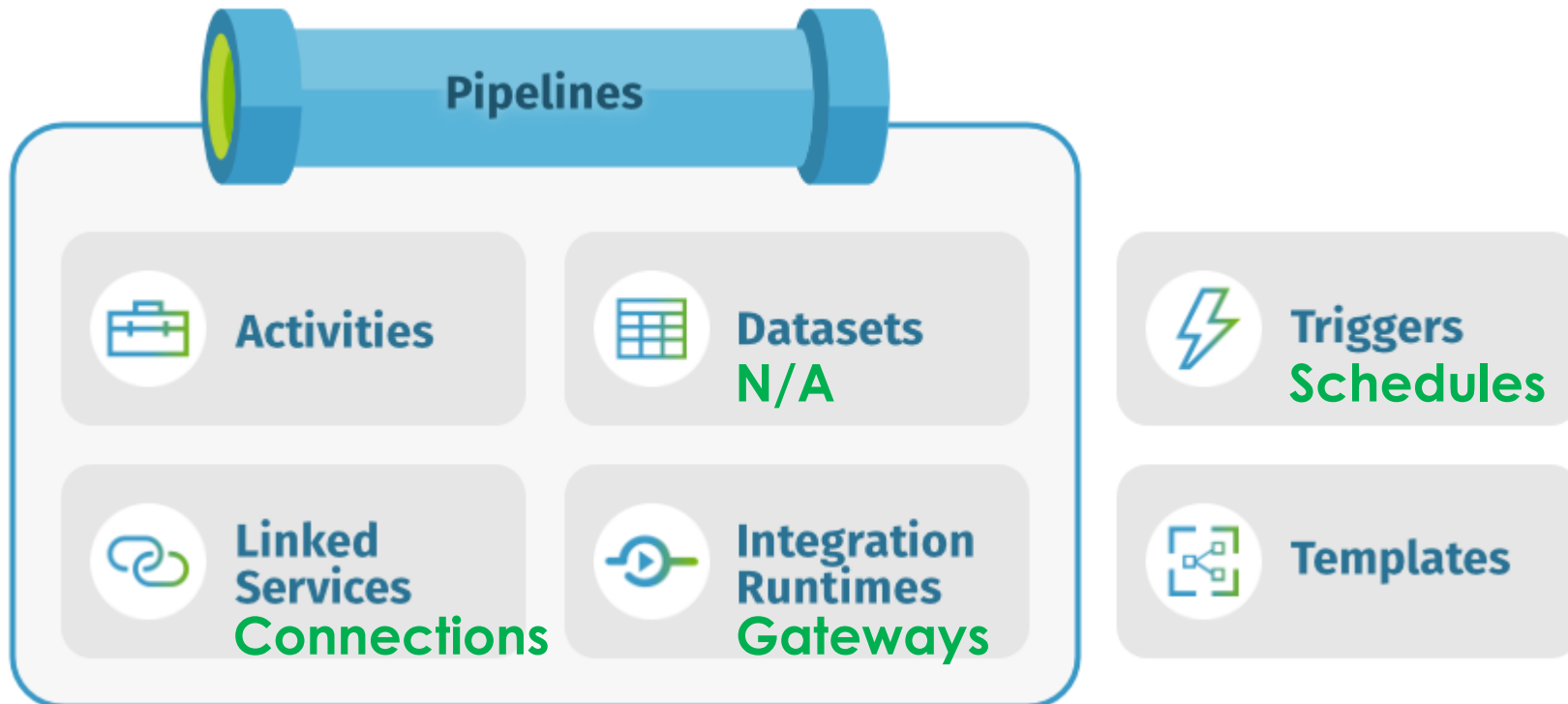
- ▶ Data Factory
- ▶ Pipeline | Data pipeline
- ▶ Activities
- ▶ Mapping Dataflow | Dataflow Gen2
- ▶ Datasets | (N/A)

* Fabric Equivalent

Components of Data Factory

- ▶ Templates * Fabric Equivalent
- ▶ Linked Service | Connections
- ▶ Triggers | Schedules
- ▶ Integration Runtimes | On-premises Data Gateway

Components of Data Factory



Credit: cathrinewilhelmsen.net

Components – Data Factory

- ▶ Environmental Boundary | *Per Workspace*
- ▶ Container to hold all other components
- ▶ Deploy per environment (Dev, QA, Prod) | *Per Workspace*
- ▶ Security (IAM) Boundary
- ▶ Integrates with Git and Azure DevOps
- ▶ Deployable as ARM template | *Deployment Pipelines*

Components – Pipelines

- ▶ Definition of executable ETL/ELT workflow
- ▶ Similar to an SSIS package
- ▶ Canvas designer interface
- ▶ Collection of units of work (Activities)
- ▶ Create new or from a Template

Components – Activities

- ▶ Single unit of work / task
- ▶ Chained in sequence or Parallelized
- ▶ Multiple categories:
 - ▶ Move & transform
 - ▶ Iteration & conditionals
 - ▶ Azure Function, Databricks/Spark Notebook, ML
 - ▶ *more...*

Components – Data Flows

- ▶ Separate type of activities with their own canvas
- ▶ No code / Low code
- ▶ Mapping Data Flows – Transform Data
- ▶ Wrangling Data Flows – Prepare Data | [Notebook](#)
- ▶ Executed from a Pipeline
- ▶ Run on Spark (Synapse) & Databricks environments

Components – Datasets

- ▶ Data reference objects
- ▶ Define format of input/output data
- ▶ Represent single table/view, file or folder
- ▶ ADF only

Components – Linked Services / Connections

- ▶ Data source / destination connection objects
- ▶ ADF: Ninety (114) supported connectors
- ▶ DFF: Forty (40) supported connectors (13 source only)

<https://learn.microsoft.com/en-us/fabric/data-factory/connector-parity>

Components – Integration Runtimes

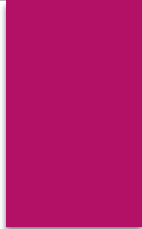
- ▶ Infrastructure to run Activities on
- ▶ Three types:
 - ▶ Azure Hosted | Fabric Capacities
 - ▶ Self-Hosted | On-premises + Virtual network Gateways
 - ▶ Azure-SSIS (ADF only)

Components – Triggers / Schedules

- ▶ Define when to execute your Pipelines
- ▶ Use a recurring schedule, interval or event based
- ▶ Specific Day & Time, or manually (UI or API)
- ▶ Tumbling Windows (ADF only)
- ▶ Storage Event Triggers – ADLS Gen2 & General Purpose v2
- ▶ Event Grid (ADF only)

Components – Templates

- ▶ Created from your own Pipeline
- ▶ Samples available from Microsoft



DEMO

Azure Data Factory

Azure Data Factory - Pricing

- ▶ Prices vary based on Azure vs Self-hosted Runtime
- ▶ Orchestration charges (per 1,000 runs)
- ▶ Execution charges (per hour, by activity type)
- ▶ Data Flows (per vCore-hour)
- ▶ Data Factory Operations
 - ▶ Read/Write: per 50k modified/referenced entities
 - ▶ Monitoring: per 50k run records retrieved

DEMO

Data Factory in Fabric

Data Factory in Fabric - Pricing

- ▶ **Data Pipelines in Fabric:**
 - ▶ Data Movement (1.5 CU's / hour)
 - ▶ Data Orchestration (0.0056 CU's each non-copy activity)
 - ▶ OneLake Storage
- ▶ Pay per Fabric "Capacity Unit" (CU)
- ▶ Billed as "Pay-as-you-go" or "Reserved" Fabric Capacity

Data Factory in Fabric - Pricing

- ▶ **Dataflow Gen2 in Fabric:**
 - ▶ Standard (Mashup) Compute (16 CU's per hour)
 - ▶ High Scale (Lake/Warehouse) Compute (6 CU's per hour)
 - ▶ OneLake Storage
- ▶ Pay per Fabric "Capacity Unit" (CU)
- ▶ Billed as "Pay-as-you-go" or "Reserved" Fabric Capacity



Q&A

Don't be shy!

Contact & Follow Me

Presentation Landing Page & Resources:

[Liktorius.com/go/SQLSAT1080](https://liktorius.com/go/SQLSAT1080)

- ▶ Twitter: **[@DLiktorius](https://twitter.com/DLiktorius)**
- ▶ Blog: **[Liktorius.com](https://liktorius.com)**
- ▶ Follow me on LinkedIn:
linkedin.com/in/DariusLiktorius

