Fellowmind

Microsoft Fabric
Guidebooks

Data Factory

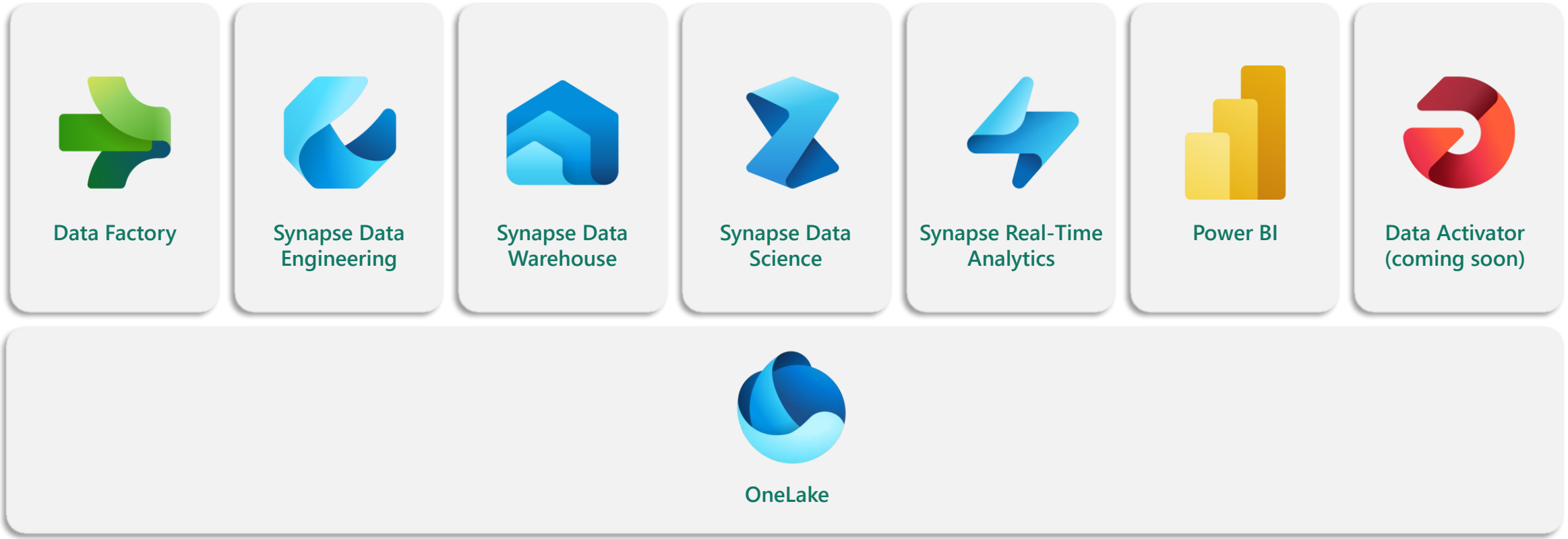
"Microsoft Fabric is a new paradigm in how we work with data – it goes beyond BI as we know it."

"It is probably the biggest innovation in data analytics since Power BI"

Mathias Halkjær
Principal Architect

MVP Microsoft® Most Valuable Professional

Microsoft Fabric



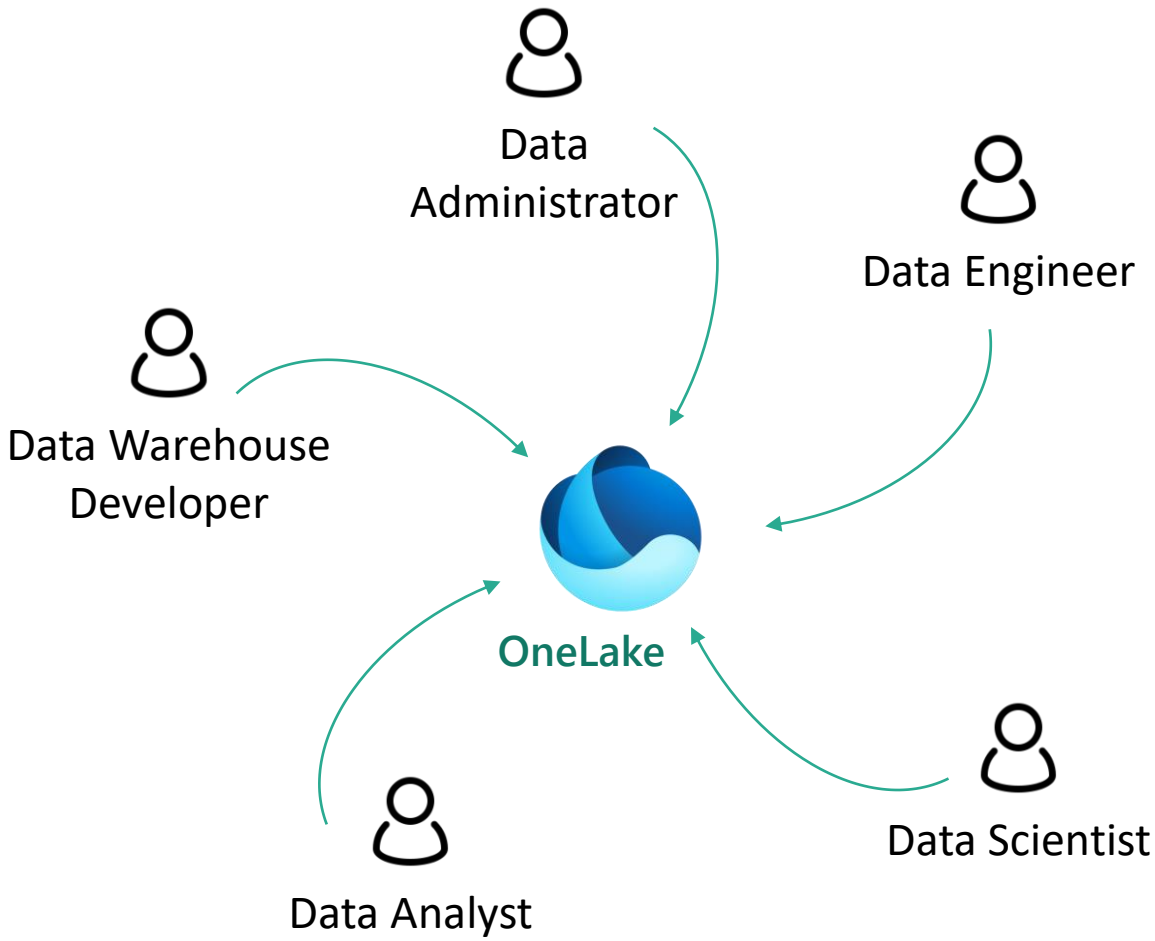
Collaboration

ONELAKE

With OneLake at its core, Microsoft Fabric unifies data disciplines and enhance collaboration across all data professionals.

OneLake both ties together all the tools, experiences and technologies – and by doing so the people working in it.

Never has it been as easy to share ones important and impactful work instantly with the right colleagues.

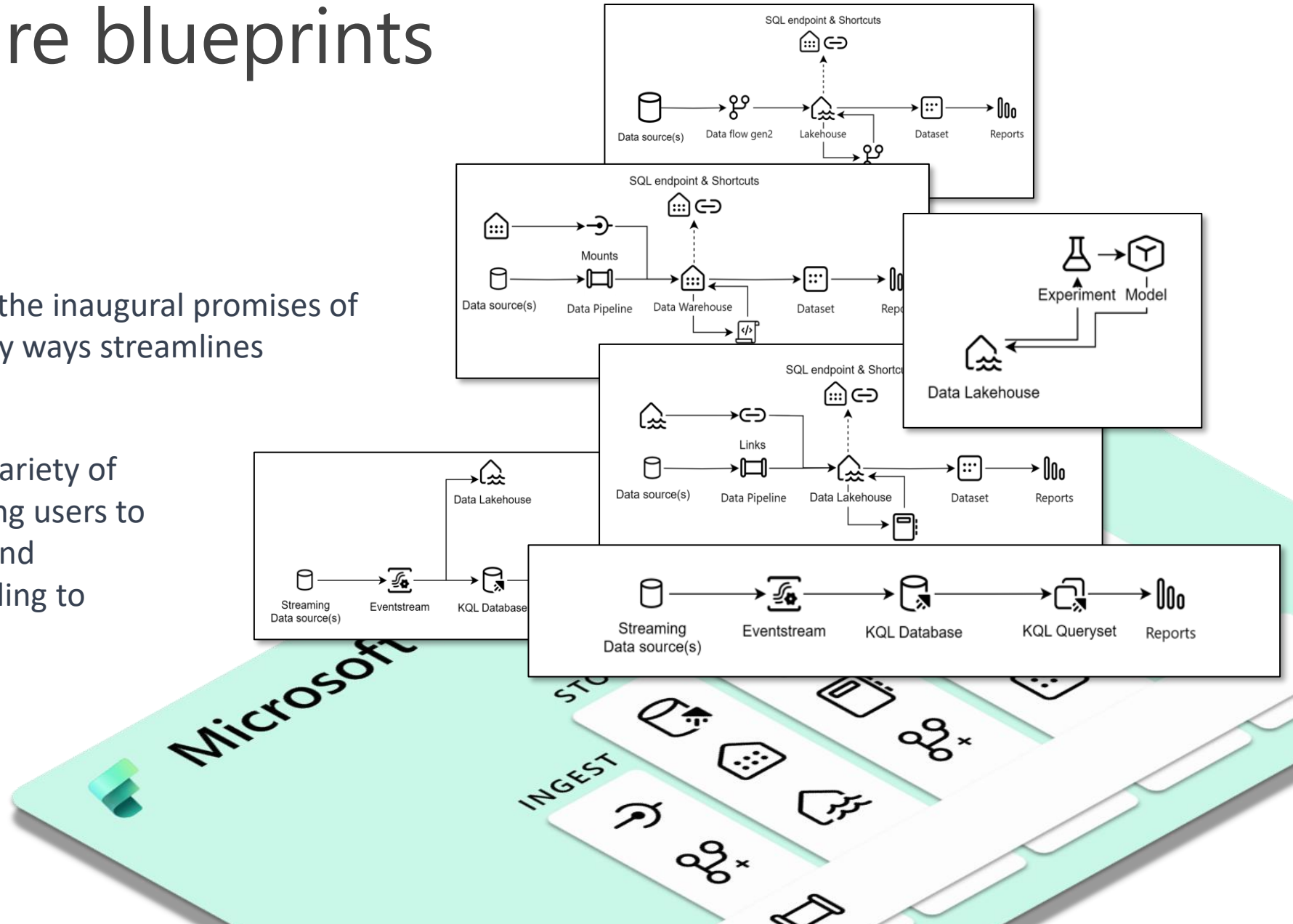


Architecture blueprints

ALL

“One Architecture” is one of the inaugural promises of Microsoft Fabric, that in many ways streamlines architectural complexities.

It does, however, present a variety of options and patterns, enabling users to customize their experience and maximize its potential according to their needs.



Capacity & pricing

CAPACITY | DOMAIN | WORKSPACE

Microsoft Fabric offers a variety of purchasable capabilities divided into SKUs, each providing unique computing power quantified by Capacity Units (CU).

Fabric features two SKU types:

- **Azure** – Billed per second with no commitment.
- **Microsoft 365** – Billed monthly or yearly, with a monthly commitment

SKU*	Capacity Units (CU)	Power BI SKU	Power BI v-cores
F2	2	-	0.25
F4	4	-	0.5
F8	8	EM/A1	1
F16	16	EM2/A2	2
F32	32	EM3/A3	4
F64	64	P1/A4	8
F128	128	P2/A5	16
F256	256	P3/A6	32
F512	512	P4/A7	64
F1024	1024	P5/A8	128
F2048	2048	-	256

Organization

CAPACITY

DOMAIN

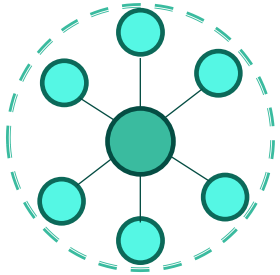
WORKSPACE

Warehouses, Lakehouses, Data Marts, Pipelines and Notebooks. Microsoft Fabrics launched with more new gadgets and technologies than we could have ever dreamed of.

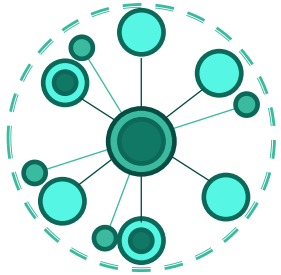
However, it's essential that as organizations, we don't overlook the foundational aspects such as our internal structure, objectives, and strategic planning.

One common organizational decision to consider when deploying a data platform like Fabric, is to choose between a centralized, decentralized, or hybrid implementation approach.

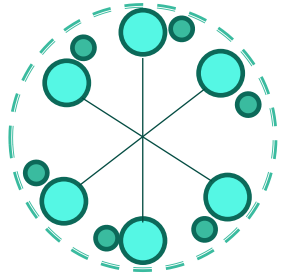
ENTERPRISE



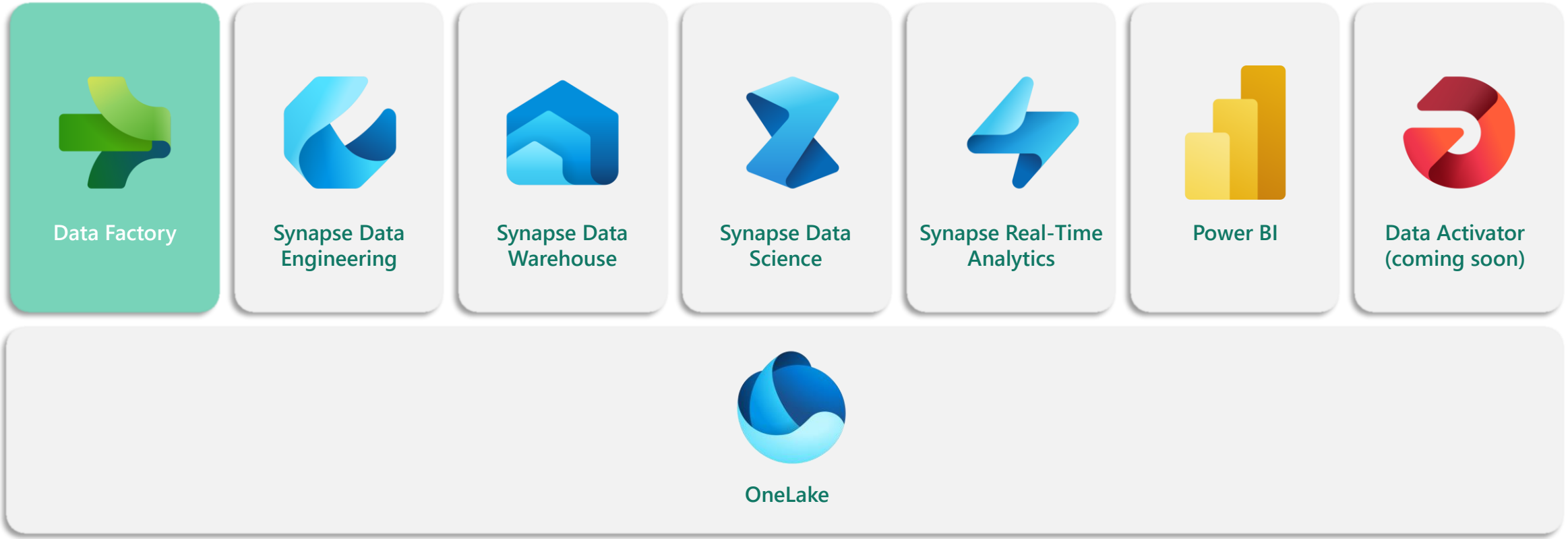
HYBRID



SELF-SERVICE



Microsoft Fabric



Data Factory

Empower data engineers and data warehouse engineers to fetch data from any data source effectively and flexible.

By combining the low-code visual transformation capabilities in the upgraded dataflows with the scalability of data pipelines, Data Factory in Microsoft Fabric offers us the most powerful, wide-ranging and flexible data integration toolbox to date.

Moreover, the powerful orchestration capabilities of pipelines turns Data Factory into the control center of how your data flows.

TOOLS

- DATAFLOW (GEN2)
- DATA PIPELINE
- SPARK JOB DEFINITION
- DATA PIPELINES

Code-free ETL with Dataflows

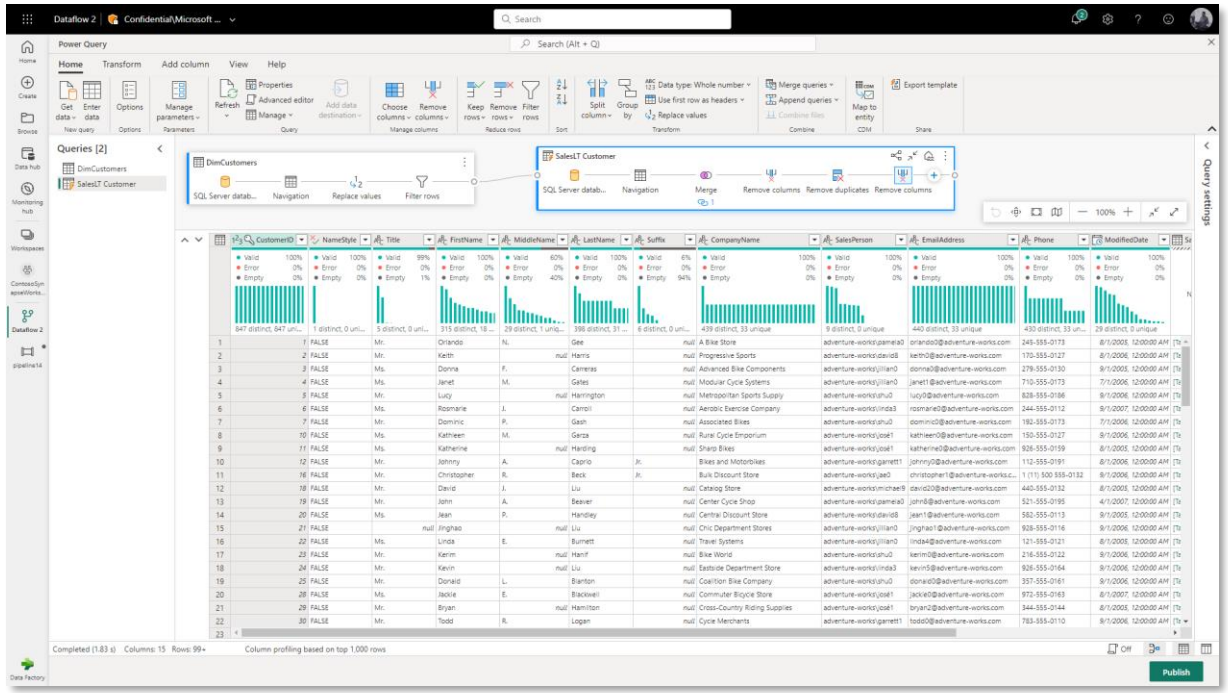
DATAFLOW (GEN2)

Redesigned as to utilize the powerful Spark engine, Dataflows have undergone a huge upgrade in its 2. generation.

Dataflows now offer the following:

- Spark-powered capabilities
- Visual no-code visual transformation
- Flexible data destinations
  (SQL, Lakehouse and Data Explorer)

Use dataflow easily import data from 100s of sources and as easily write to several destinations.



Data Pipelines

DATA PIPELINE

Data Pipelines excel in handling large, complex datasets and support a vast array of data sources. Their superior automation, orchestration, and meta-data driven flexibility allow for dynamic ingestion, enabling large-scale, efficient data integration from a diverse range of data sources.

Picture a scenario where you need to automatically and dynamically extract all tables—potentially in the thousands—from a SQL database. This is where Data Pipelines' capabilities truly excel.

Moreover, Data Pipelines prove exceptionally beneficial in scenarios where you need to retrieve a lot of data from REST APIs by dynamically paging through thousands of API pages.

"Pipelines are particularly effective when dealing with large-scale data processing tasks or complex data sources"

Carsten Koudal
Architect, Fellowmind

Ingest vs. load

Ingestion, as performed in Azure Data Factory, refers to acquiring and importing data in its raw form from diverse sources. It's systematic, scheduled, and supports large volumes.

Loading, as seen in datasets – and sometimes in dataflows, is the simpler process of inputting data into a specific system or database for immediate analysis, often without sophisticated preparation.

While this may be convenient, it can lead to unsystematic ETL (Extract, Transform, Load) processes and might limit data quality and comprehensive insights

Dynamic Pipelines

DATA PIPELINE

In Power Query, we usually work with data that doesn't change much. But the reality is that data is often changing and coming from many places—With dynamic pipelines, we can keep up with these changes.

These pipelines can easily and automatically adjust to new data sources, change their actions based on the type of data, or send data to different places as needed.

This means we can do more with our data, and we can do it more easily and efficiently.

Meta-data framework

DATA PIPELINE

A top-notch metadata framework is the secret sauce behind every smooth-running dynamic data pipeline. It keeps track of data sources, table names, and where and how to store the data. This makes it simple and flexible to tweak or expand your data integration setup, offering an efficient and user-friendly solution.

Suggestions for a meta-data framework:

- Azure SQL
- Data Lake Table Storage
- JSON config files

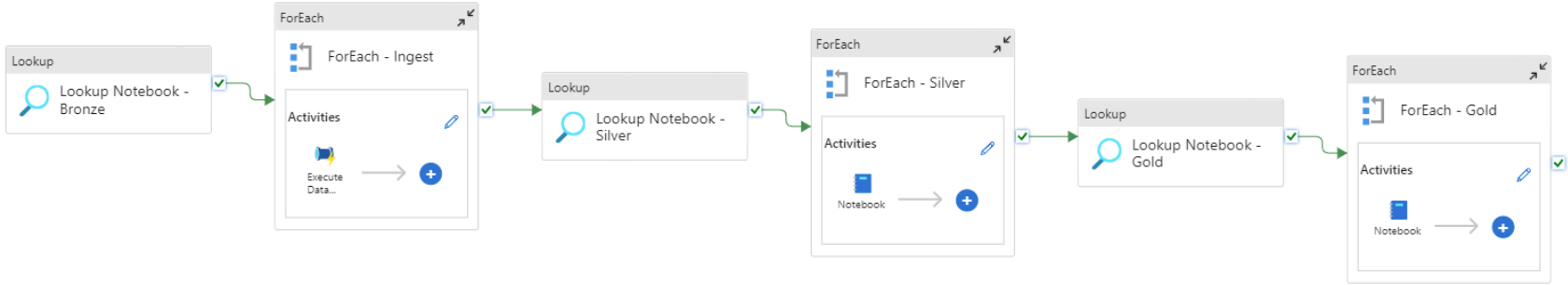
Orchestration

DATA PIPELINE

Orchestration in Microsoft Fabric is essential for managing complex data workflows.

It coordinates tasks, monitors dependencies, and handles failures, ensuring seamless data flow. This process empowers the data platform by minimizing errors, optimizing resource utilization, and enabling scalability.

In any data platform, orchestration ensures smooth data movement and processing, enabling efficient analytics
and business insights.



Example orchestration pipeline

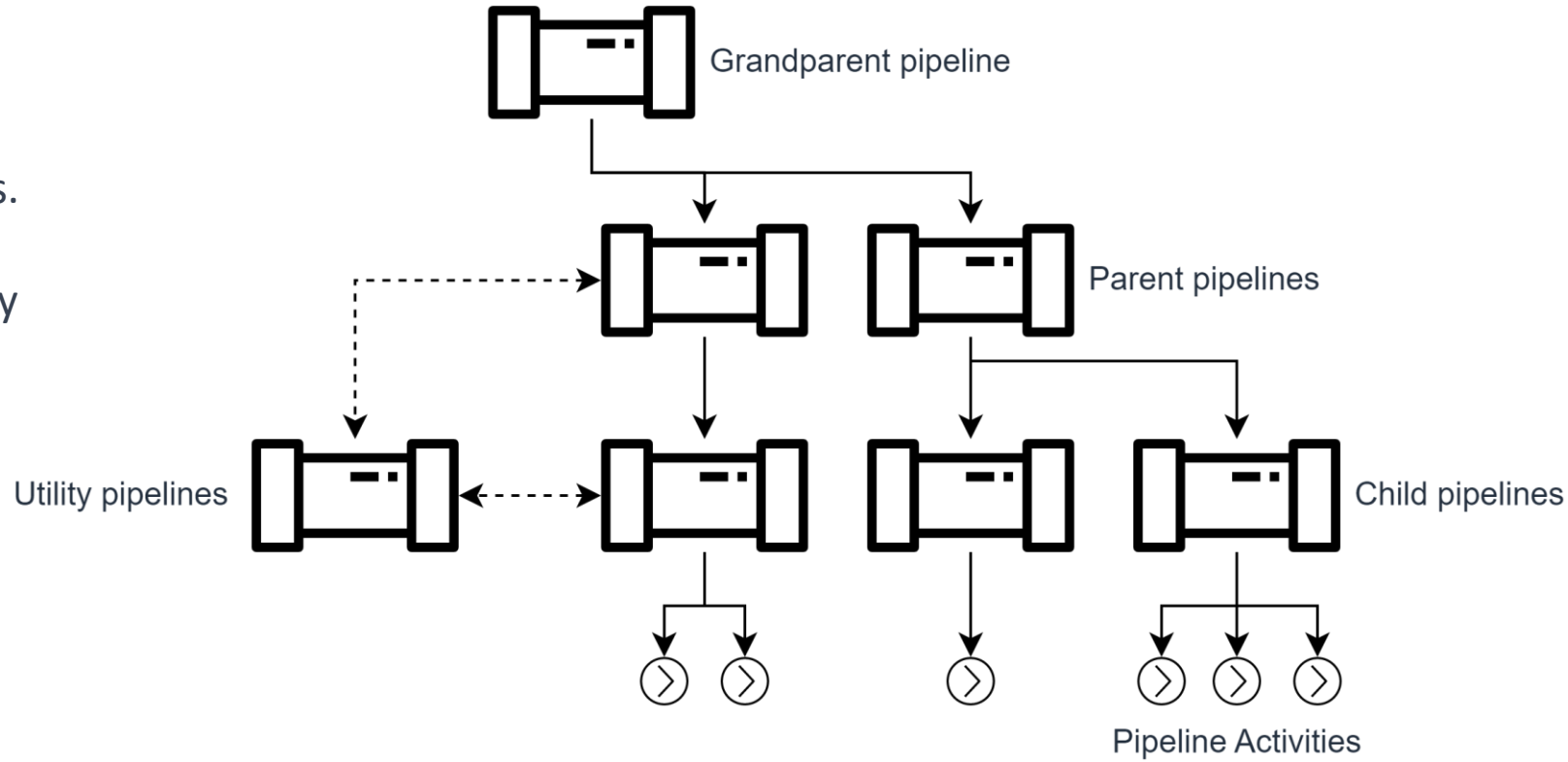
Hierarchy of pipelines

DATA PIPELINE

Pipelines can trigger other pipelines.

A very flexible capability that quickly turn messy if not structured properly.

By structuring pipelines in a managed hierarchy of pipelines it is easier to manage, troubleshoot and govern how data flows through the entire organization.



Get started today

Try Microsoft Fabric

Try Fabric (microsoft.com)

Watch Fellowmind's monthly Power BI Update

Power BI Update (fellowmindcompany.com)

Connect with our Microsoft Data Platform MVPs

https://www.linkedin.com/in/mhalkjaer

https://www.linkedin.com/in/brianbonk