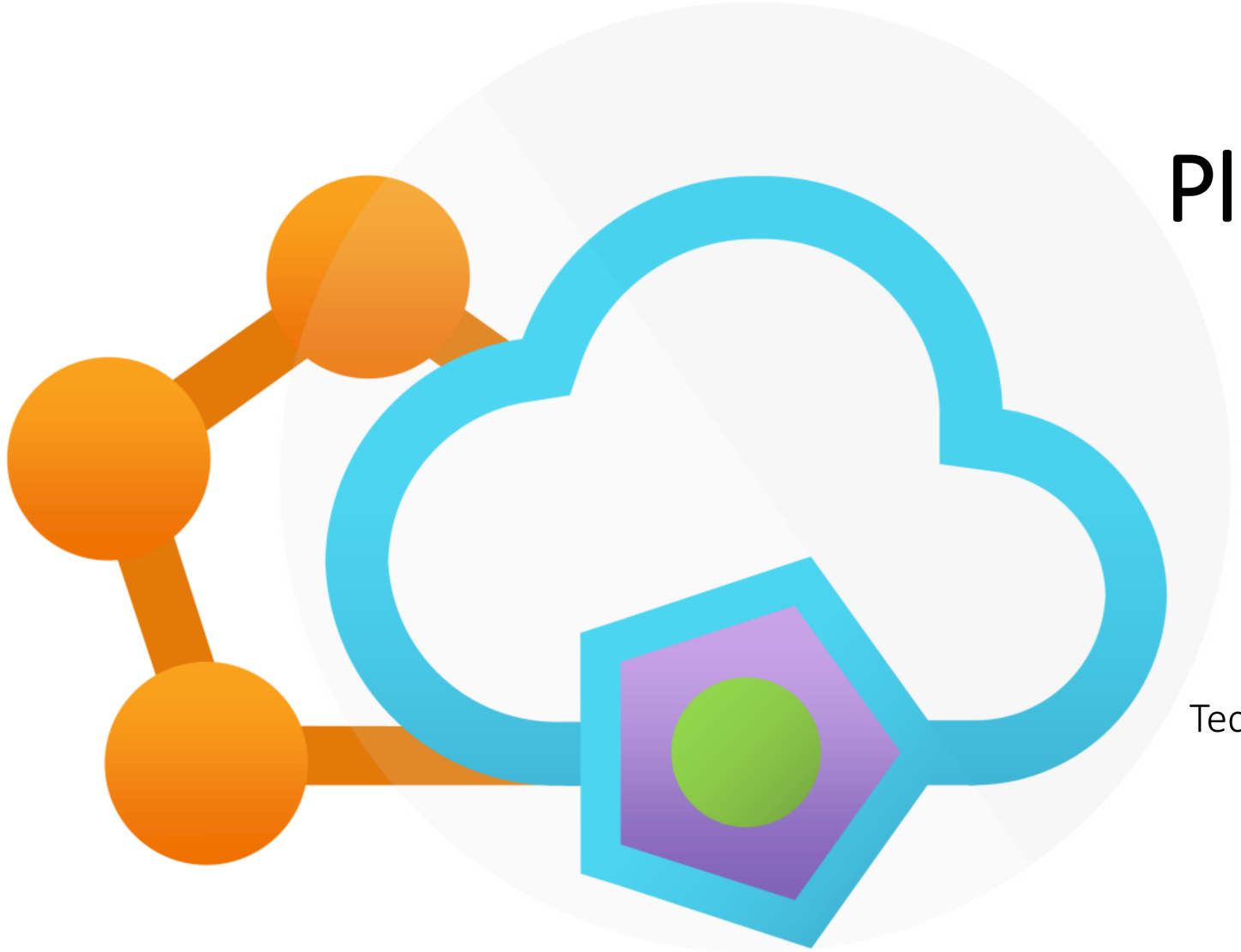


An Evolution of Data Platform Architectures

Lambda, Kappa, Delta, Mesh & Fabric

λ κ δ

Paul Andrew

Technical Architect | Director

MVP Microsoft Most Valuable Professional

Cloud Formations

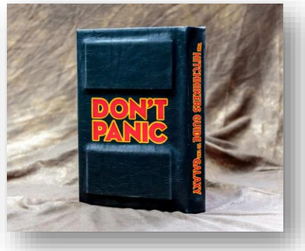
Architecture Agenda:

Lambda, Kappa, Delta, Mesh & Fabric

But first, a couple of questions…

What is the answer to life, the universe and everything?

Answer:
42 ✗



Answer:
It depends! ✓

What is big data?

Answer:
It depends! ✓

Answer:
"Any data that you cannot process
in the time that you have/want
using the technology you have." ✓

- Buck Woody

@BuckWoodyMSFT

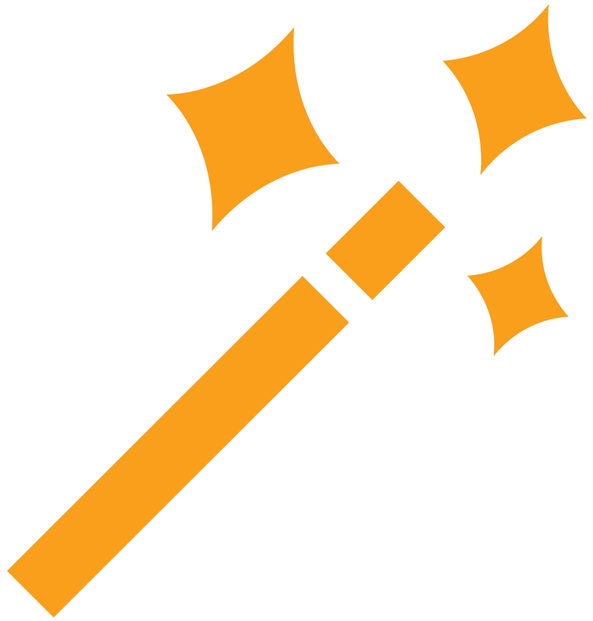
Volume
Velocity
Variety
Veracity
Value

What is the goal of our data solutions?

Data = Information = Knowledge = Power

How do we deliver our data insights?

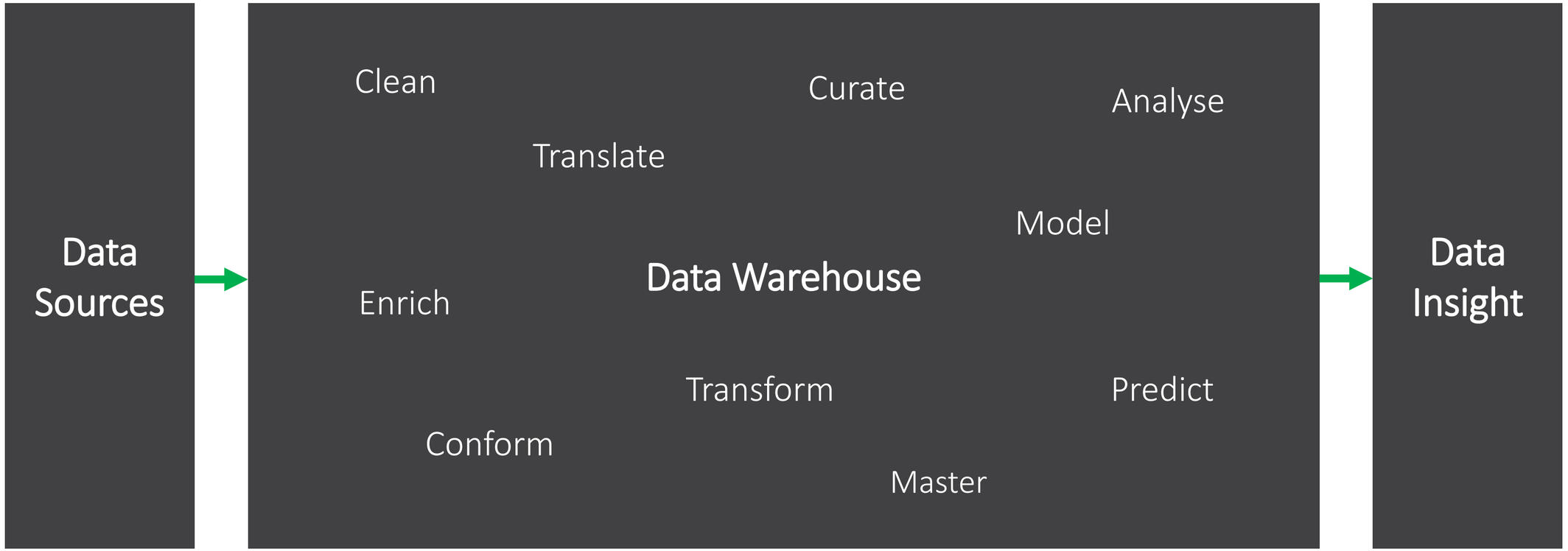
Data Sources



Data Insight

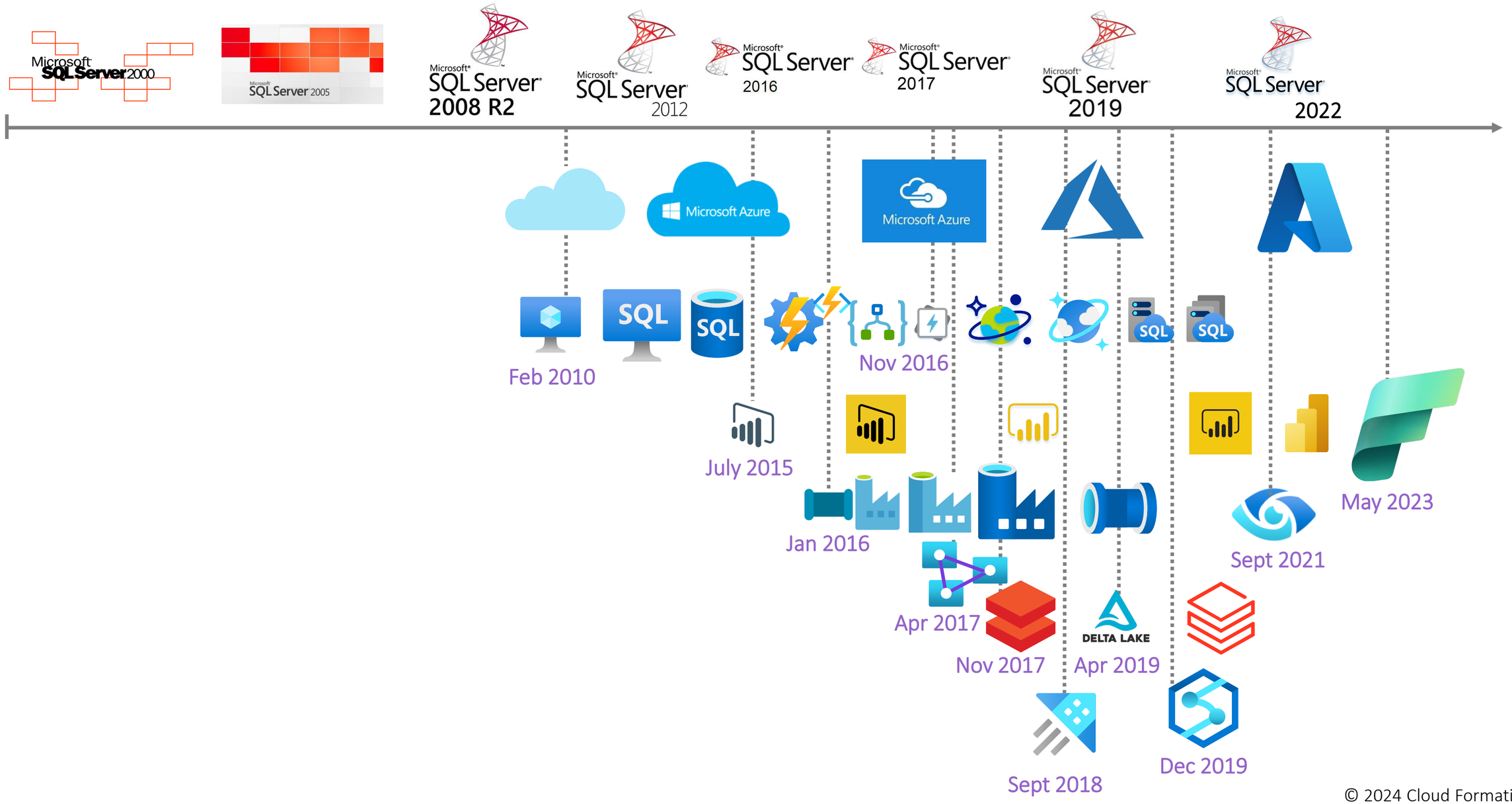
Data = Information = Knowledge = Power

How do we deliver our data insights?

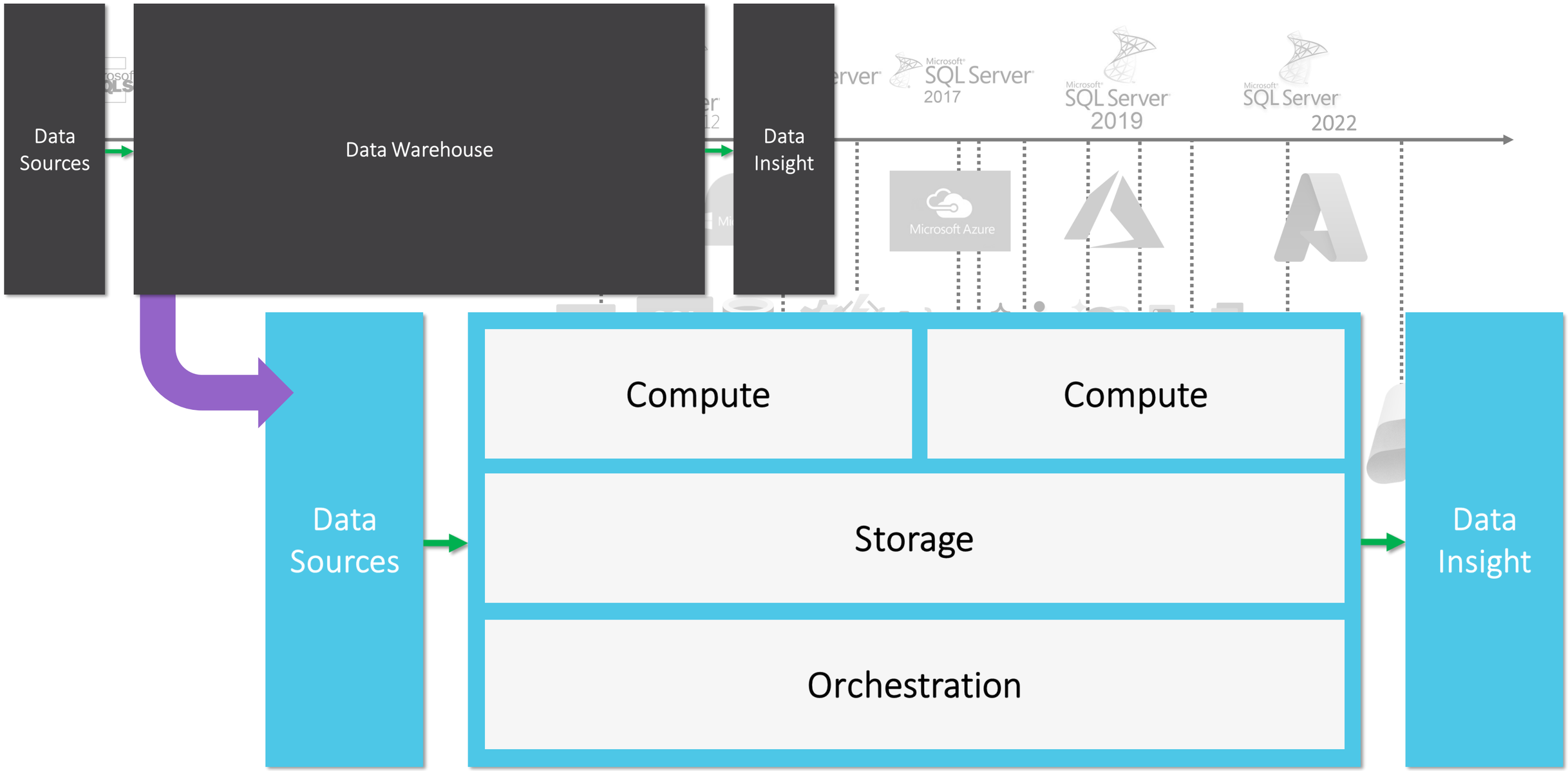


Data = Information = Knowledge = Power

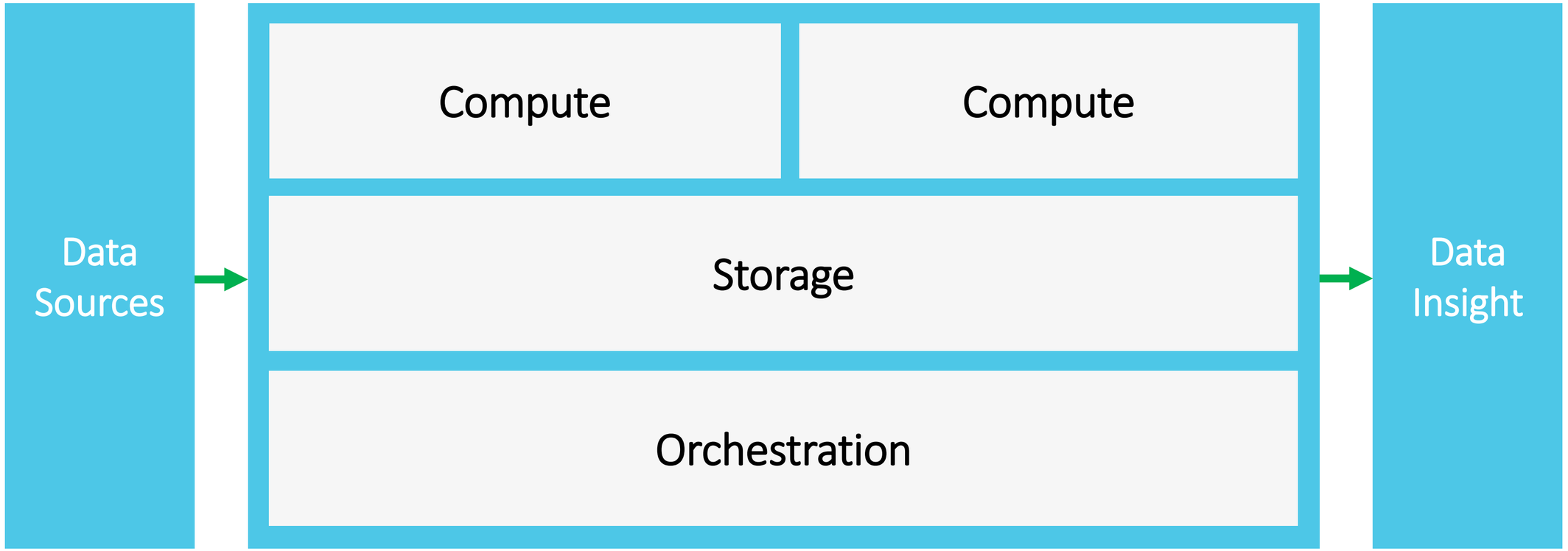
An Evolution of Microsoft Data Technology



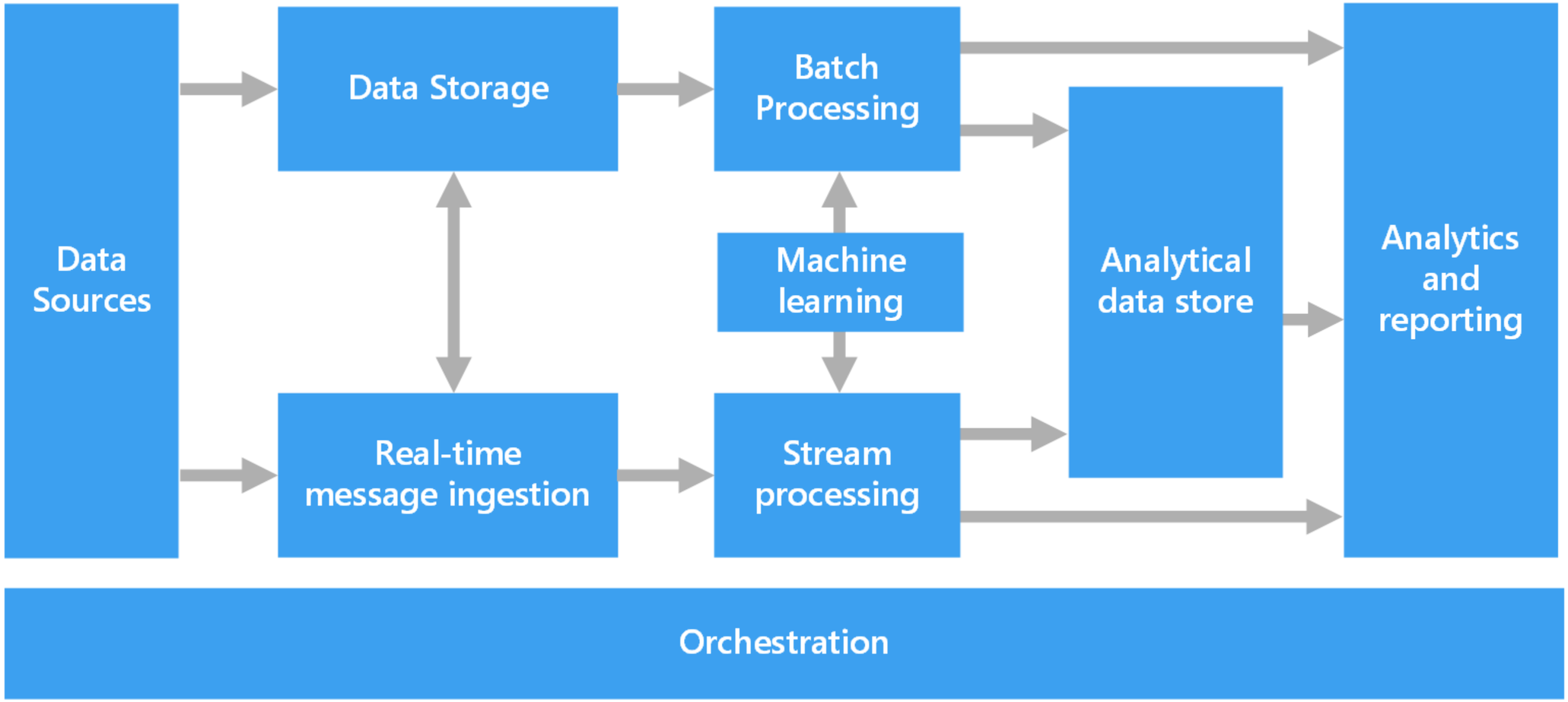
An Evolution of Microsoft Data Technology



My First Reference Architecture

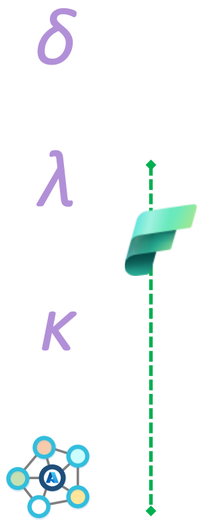


Microsoft's Components of a ~~Big~~ Data Architecture

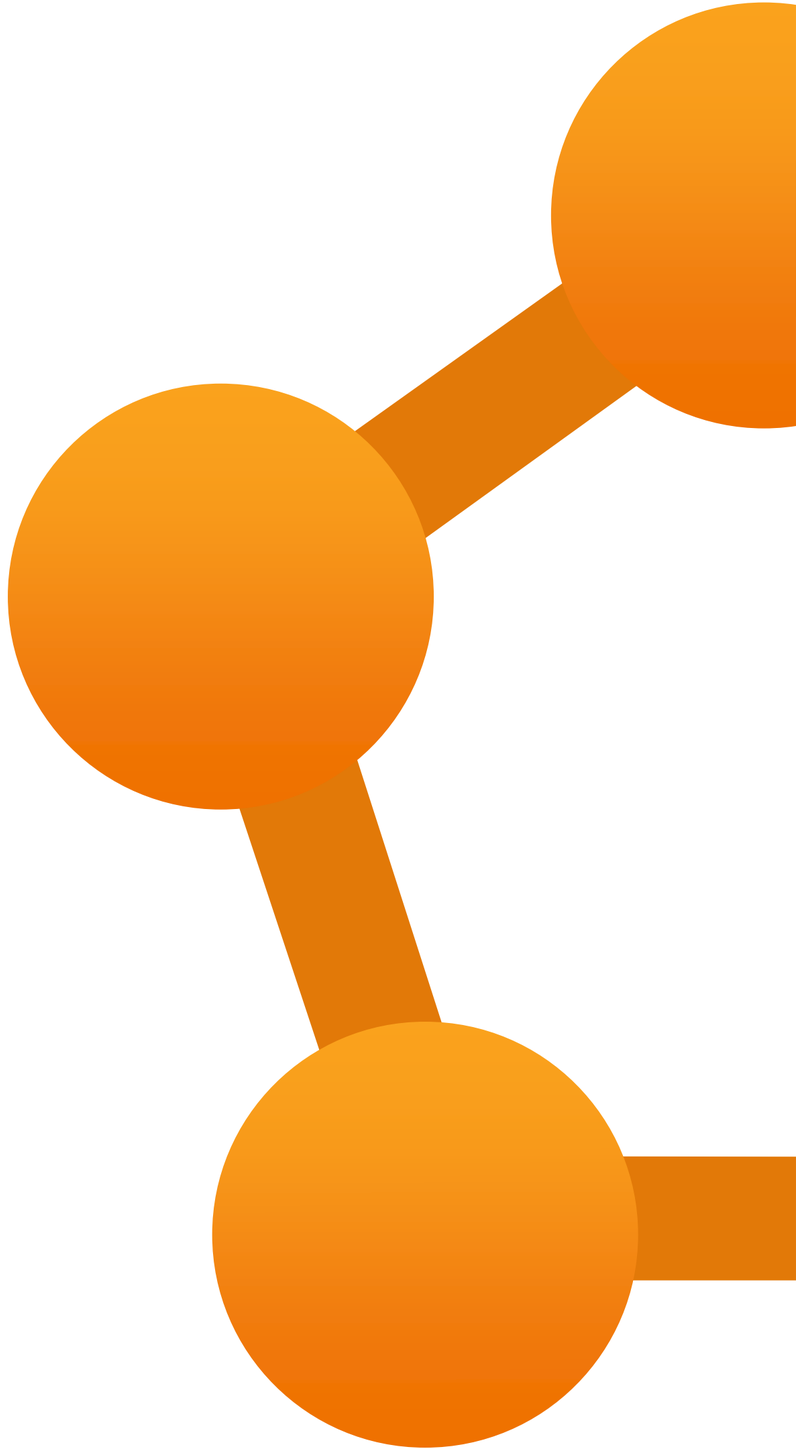


https://docs.microsoft.com/en-us/azure/architecture/data-guide/big-data/

Architecture Agenda:



Delta δ



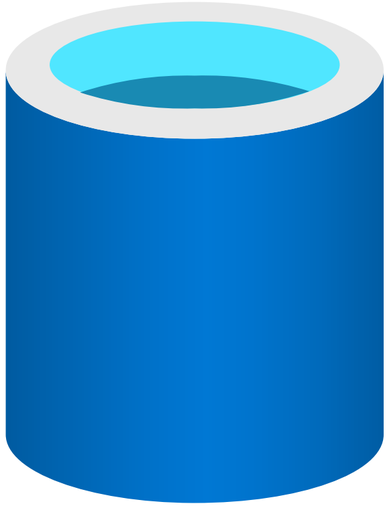
Cloud Formations

Databases

Data**B**ase

Management

System



Atomicity

Consistency

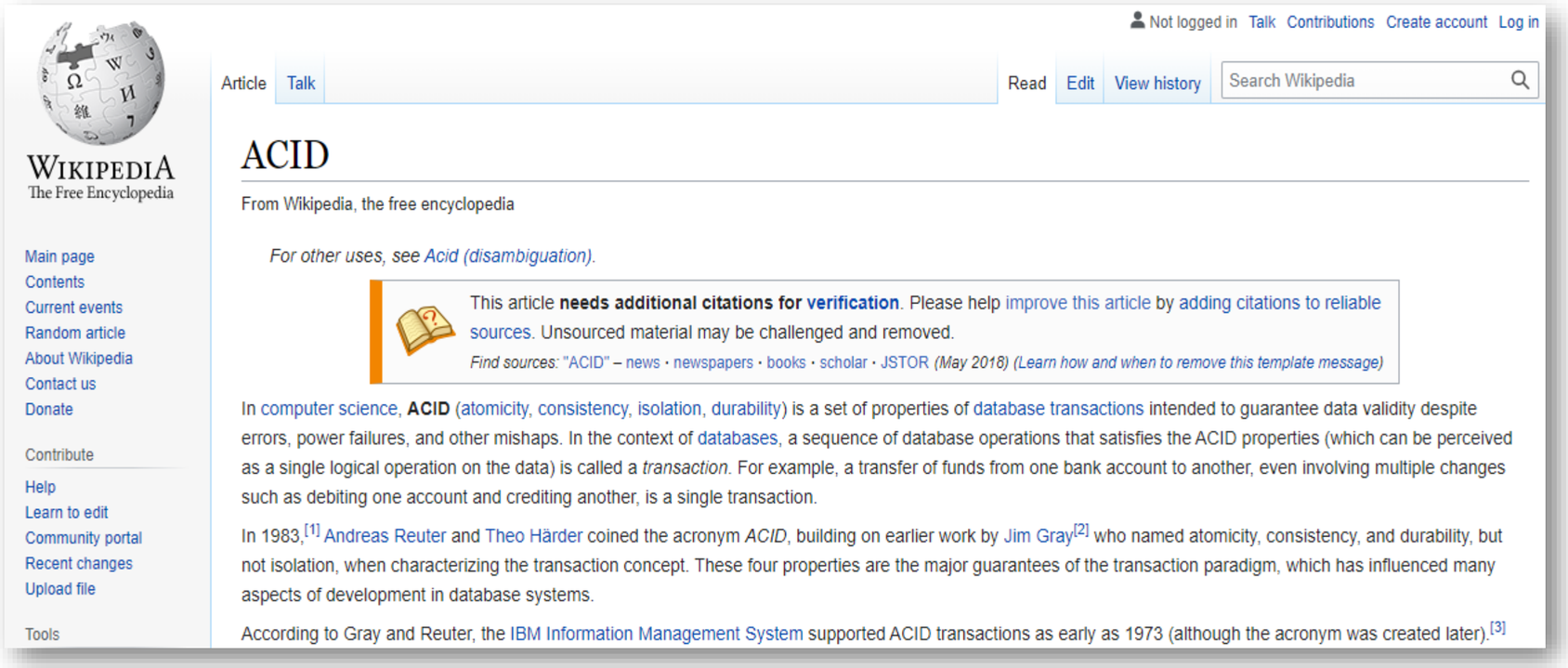
Isolation

Durability

Databases

Data**B**ase
Management
System

"is a set of properties of database transactions intended to guarantee data validity"



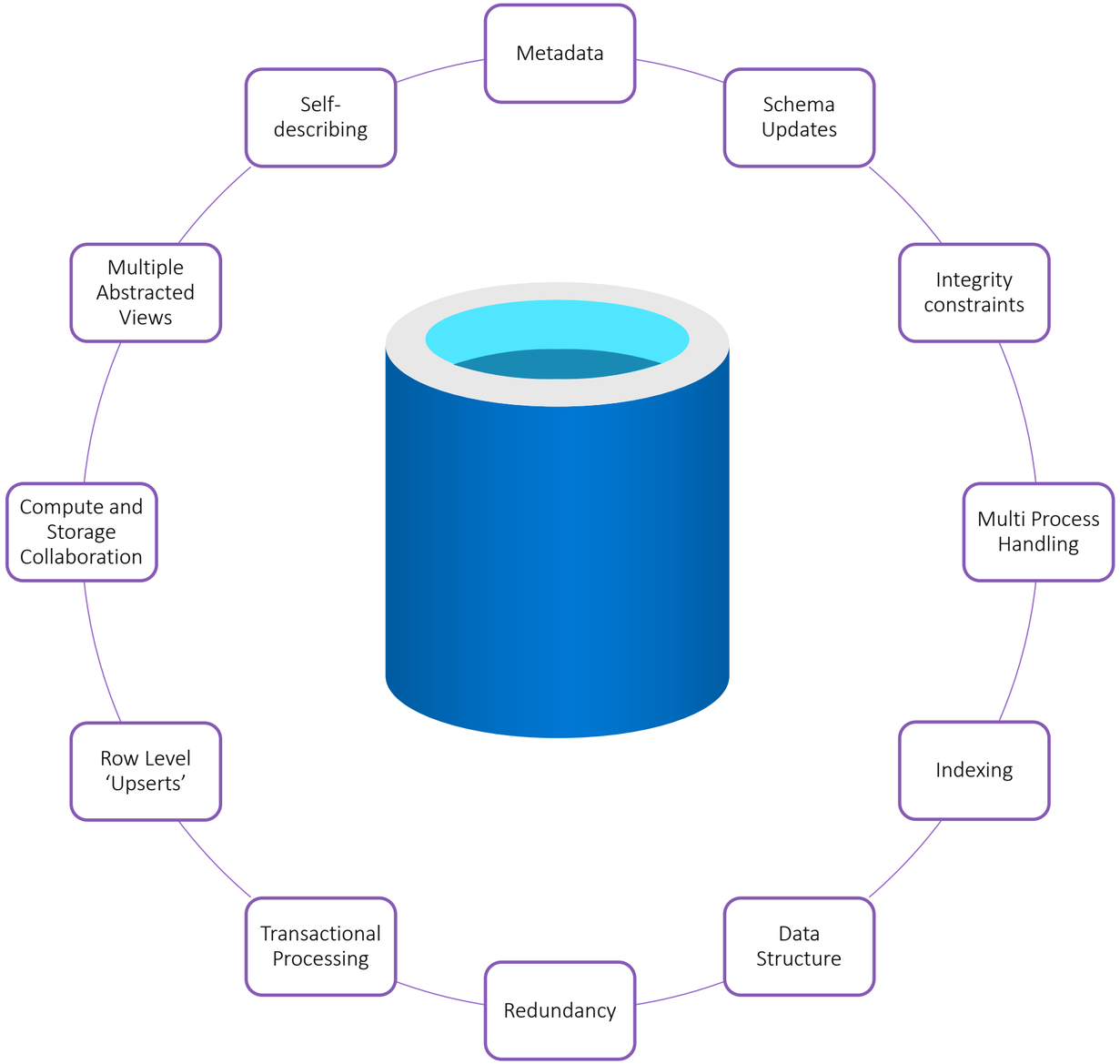
https://en.wikipedia.org/wiki/ACID

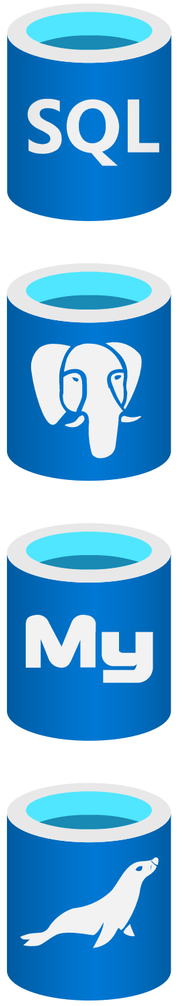
Atomicity
Consistency
Isolation
Durability

Databases

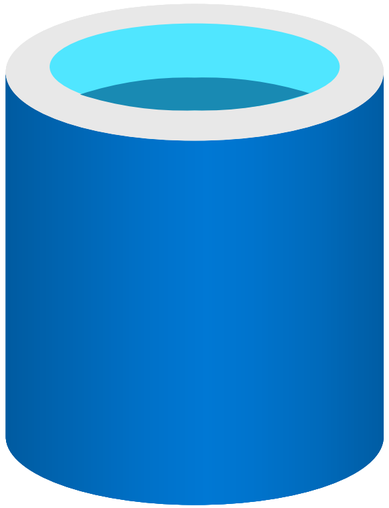
Data**B**ase
Management
System

Atomicity
Consistency
Isolation
Durability



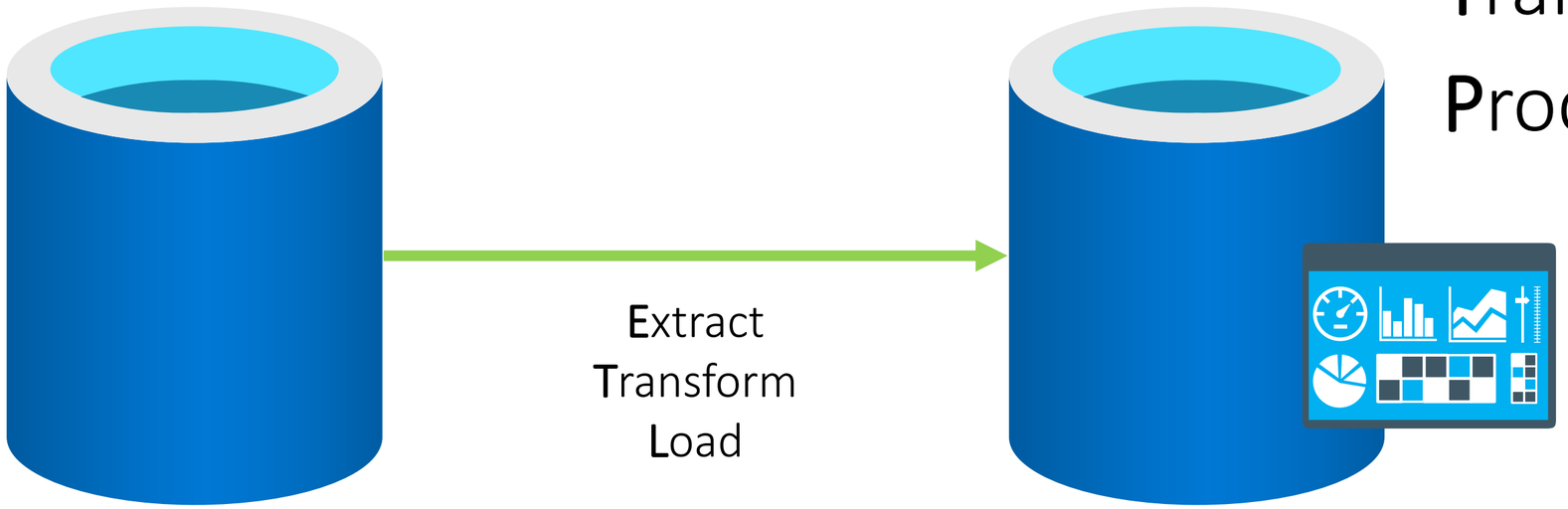


Databases



Creating a Data Warehouse

Online
Line
Transactional
Processing



Extract
Transform
Load

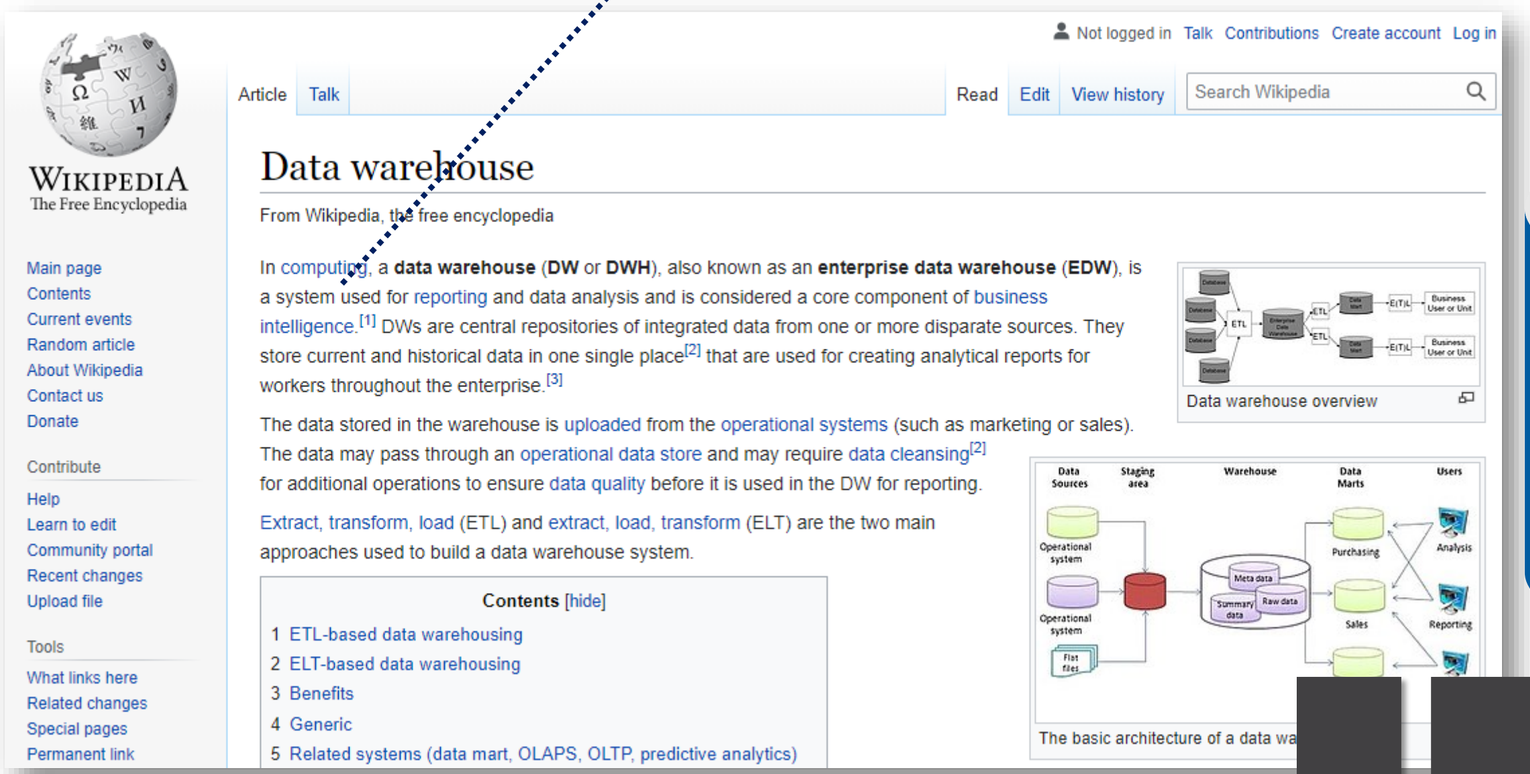
Application
Data

Data
Warehouse

Offline
Analytical
Transactional
Processing

Creating a Data Warehouse

"a system for reporting and data analysis"



Not logged in Talk Contributions Create account Log in

Article Talk | Read Edit View history | Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Learn to edit
Community portal
Recent changes
Upload file

Tools
What links here
Related changes
Special pages
Permanent link

Data warehouse

From Wikipedia, the free encyclopedia

In computing, a **data warehouse** (**DW** or **DWH**), also known as an **enterprise data warehouse** (**EDW**), is a system used for reporting and data analysis and is considered a core component of business intelligence.[1] DWs are central repositories of integrated data from one or more disparate sources. They store current and historical data in one single place[2] that are used for creating analytical reports for workers throughout the enterprise.[3]

Data warehouse overview

The data stored in the warehouse is uploaded from the operational systems (such as marketing or sales). The data may pass through an operational data store and may require data cleansing[2] for additional operations to ensure data quality before it is used in the DW for reporting.

Extract, transform, load (ETL) and extract, load, transform (ELT) are the two main approaches used to build a data warehouse system.

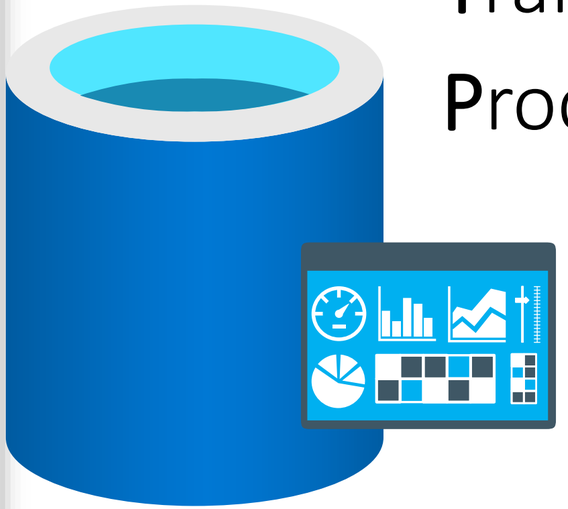
Contents [hide]
1 ETL-based data warehousing
2 ELT-based data warehousing
3 Benefits
4 Generic
5 Related systems (data mart, OLAPS, OLTP, predictive analytics)

Data Sources | Staging area | Warehouse | Data Marts | Users
Operational system
Operational system
Flat files
Meta data
Summary data
Raw data
Purchasing
Sales
Analysis
Reporting

The basic architecture of a data wa

https://en.wikipedia.org/wiki/Data_warehouse

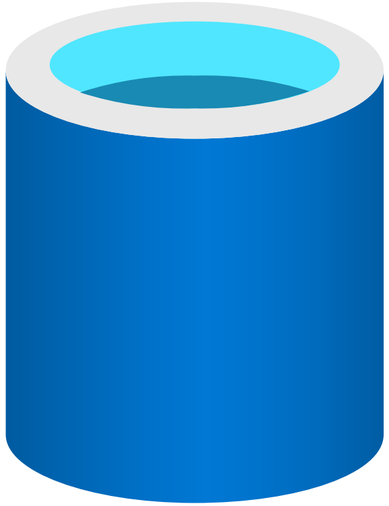
Offline
Analytical
Transactional
Processing



Data Warehouse

Big Data:

- Volume
- Velocity
- Variety
- Veracity
- Value

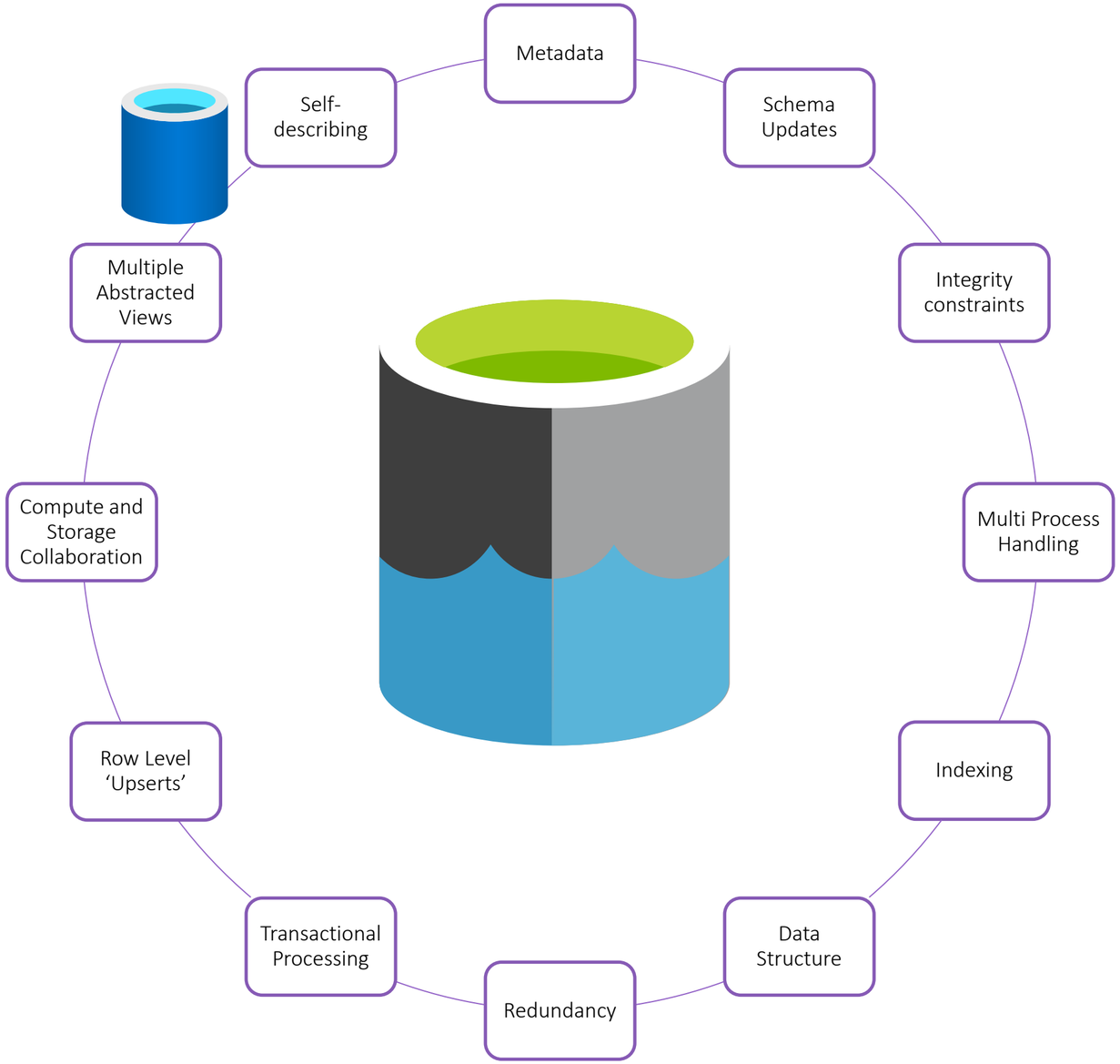


Data Warehouse

Data Lakes

Big Data:

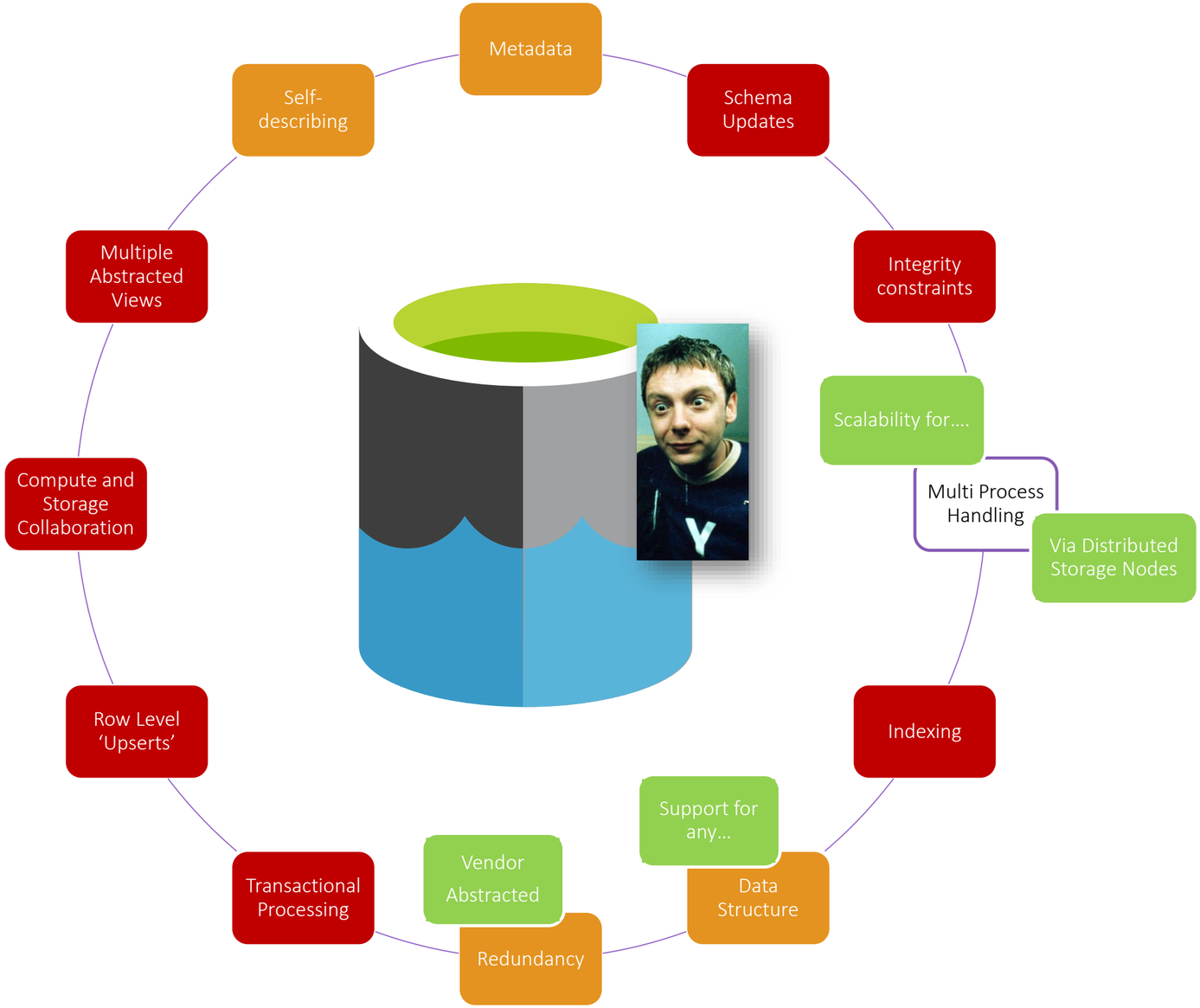
- **V**olume
- **V**elocity
- **V**ariety
- **V**eracity
- **V**alue



Data Lakes

Big Data:

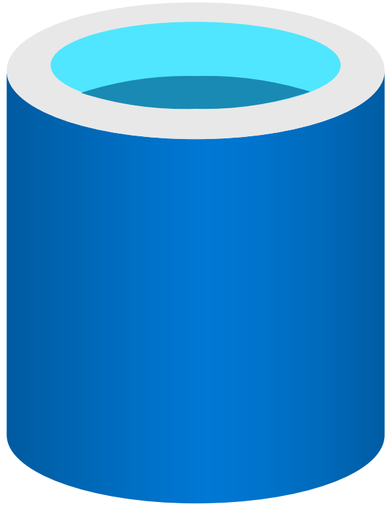
- **V**olume
- **V**elocity
- **V**ariety
- **V**eracity
- **V**alue



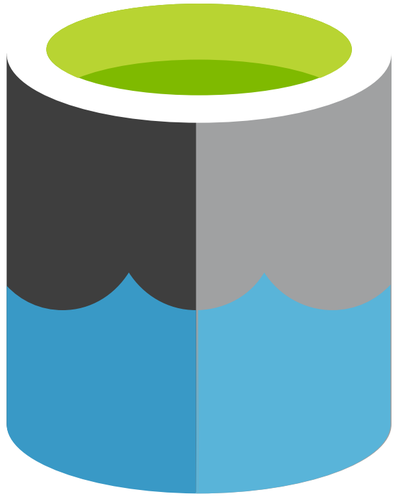
Problem Summary

Data Lakes are good, but they still lack some of the basic ACID functionality needed for data processing.

We are/were trying to use Data Lakes for everything (to replace Databases).



VS

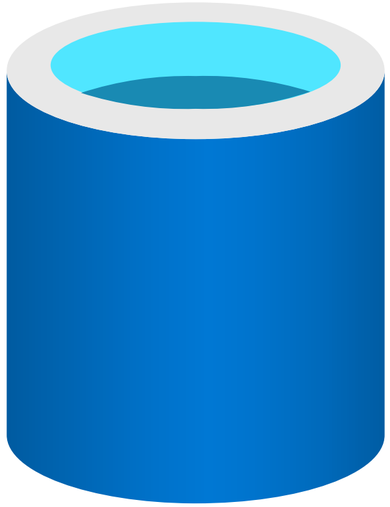


Scales Up	Scales Out
Natural Home for Structured Data	Any Data Structure
Storage Limits	No Storage Limits
Transactional Resilience	No Transactional Handling
Storage & Compute Coupled	Storage & Compute Decoupled

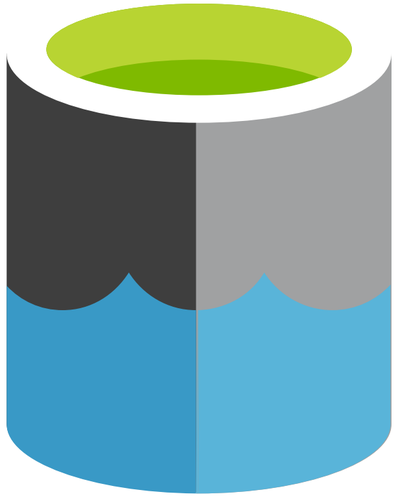
Problem Summary

Data Lakes are good, but they still lack some of the basic ACID functionality needed for data processing.

We are/were trying to use Data Lakes for everything (to replace Databases).



VS



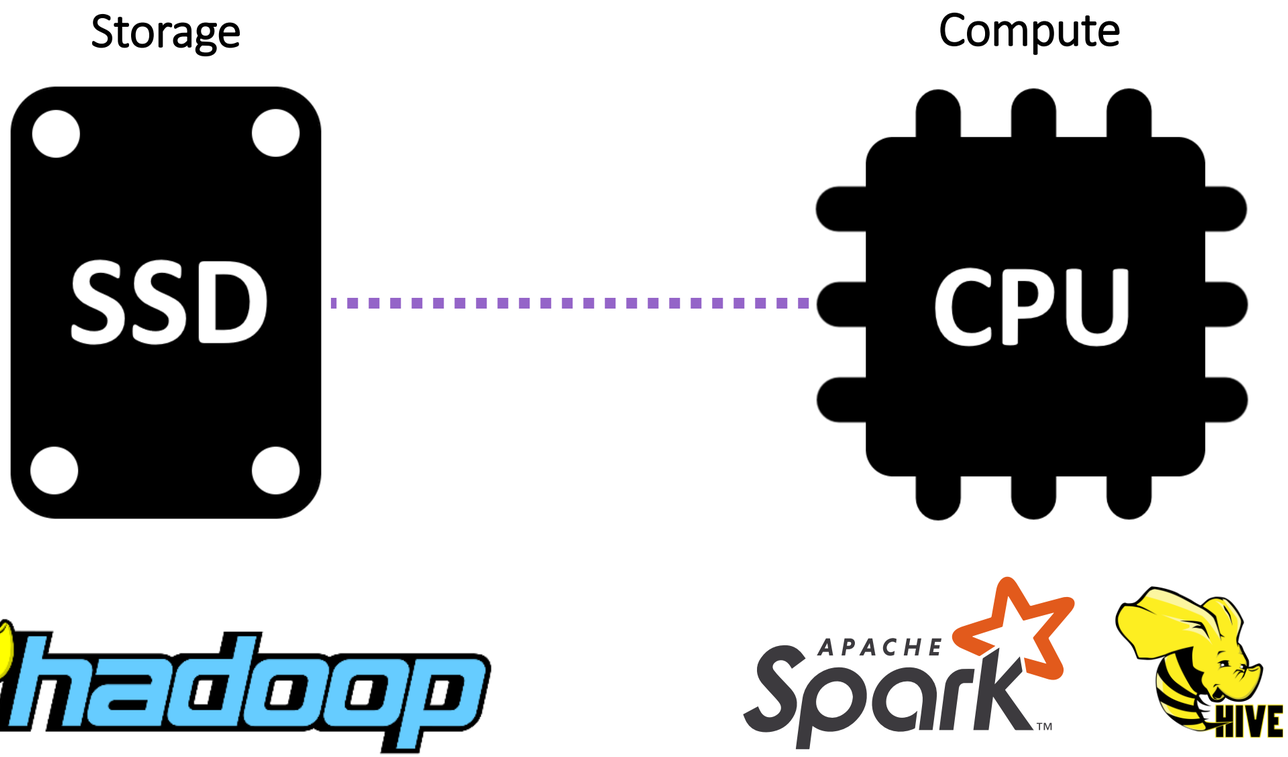
Scales Up	Scales Out
Natural Home for Structured Data	Any Data Structure
Storage Limits	No Storage Limits
Transactional Resilience	No Transactional Handling
Storage & Compute Coupled	Storage & Compute Decoupled

Solution

‘Just’ enable ACID transactional support for Data Lakes...

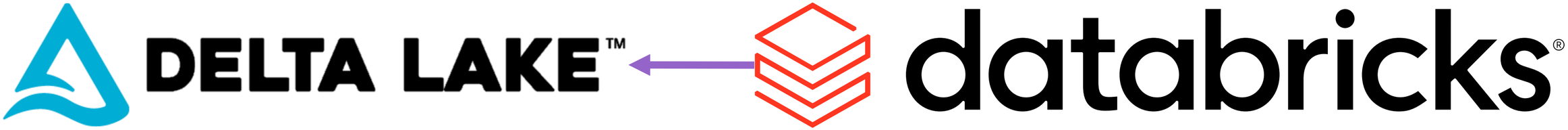
Big Data:

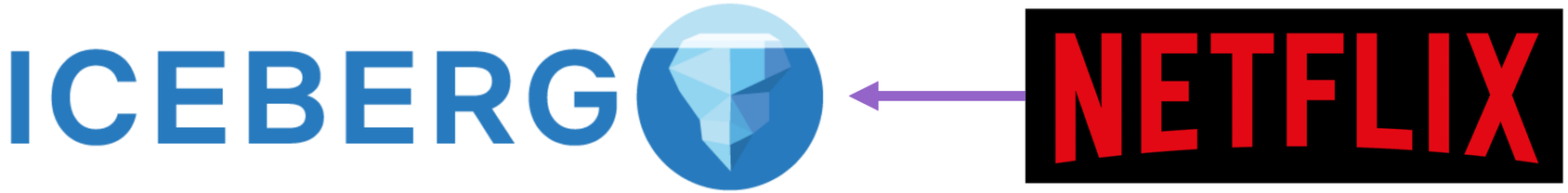
- **V**olume
- **V**elocity
- **V**ariety
- **V**eracity
- **V**alue



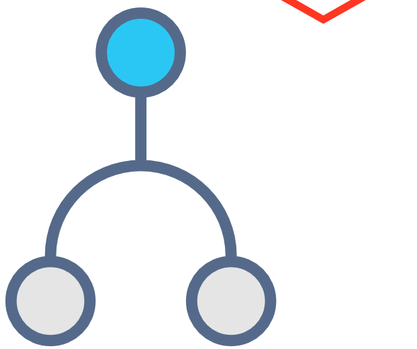
Storage & Compute ~~Decoupled~~ Working Together Again As Friends!

ACID Data Frameworks for Data Lakes





What is Delta Lake?

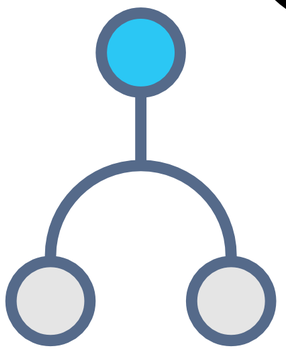


databricks®

February 2019

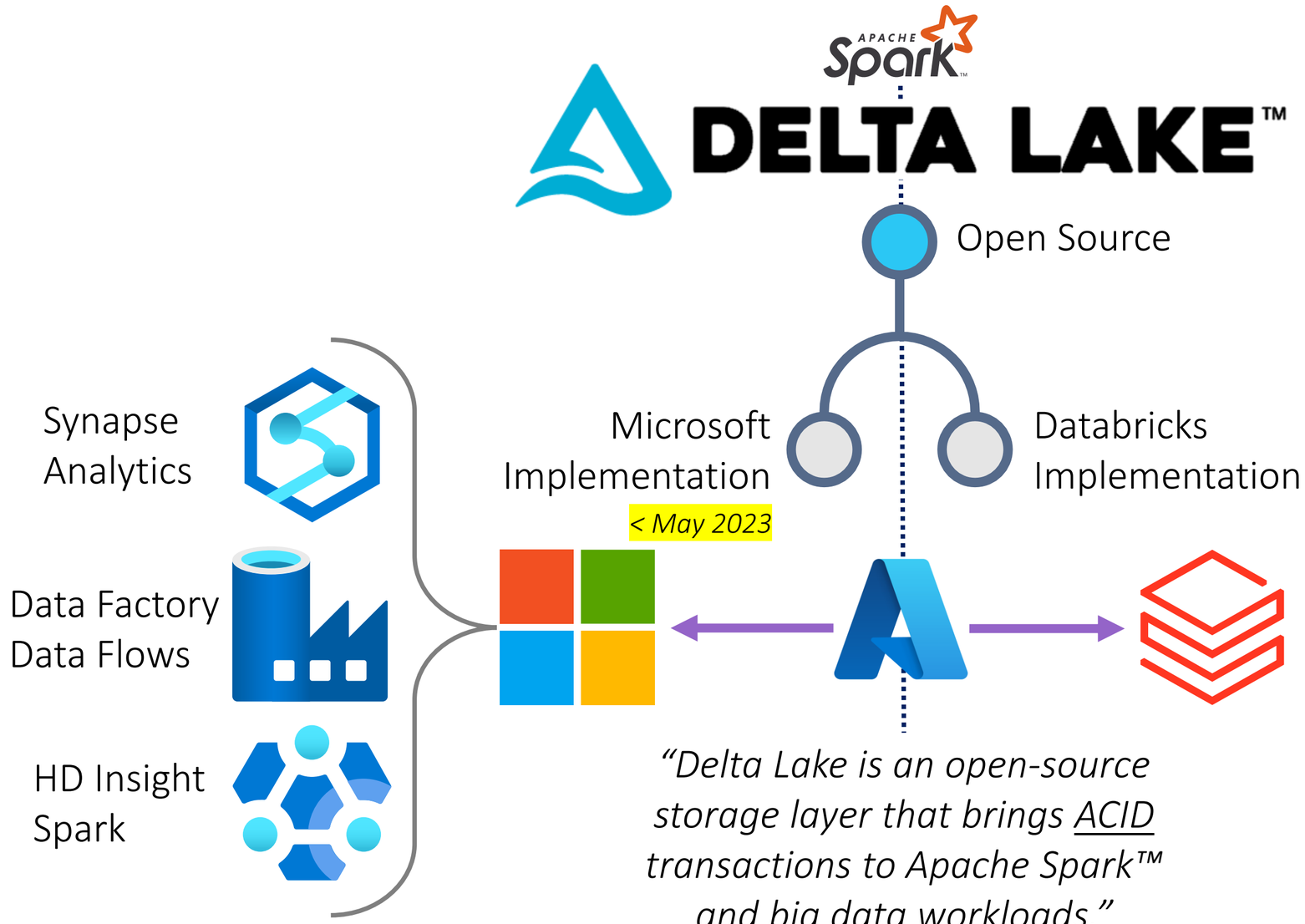
What is Delta Lake?

https://delta.io
https://github.com/delta-io/delta
April 2019



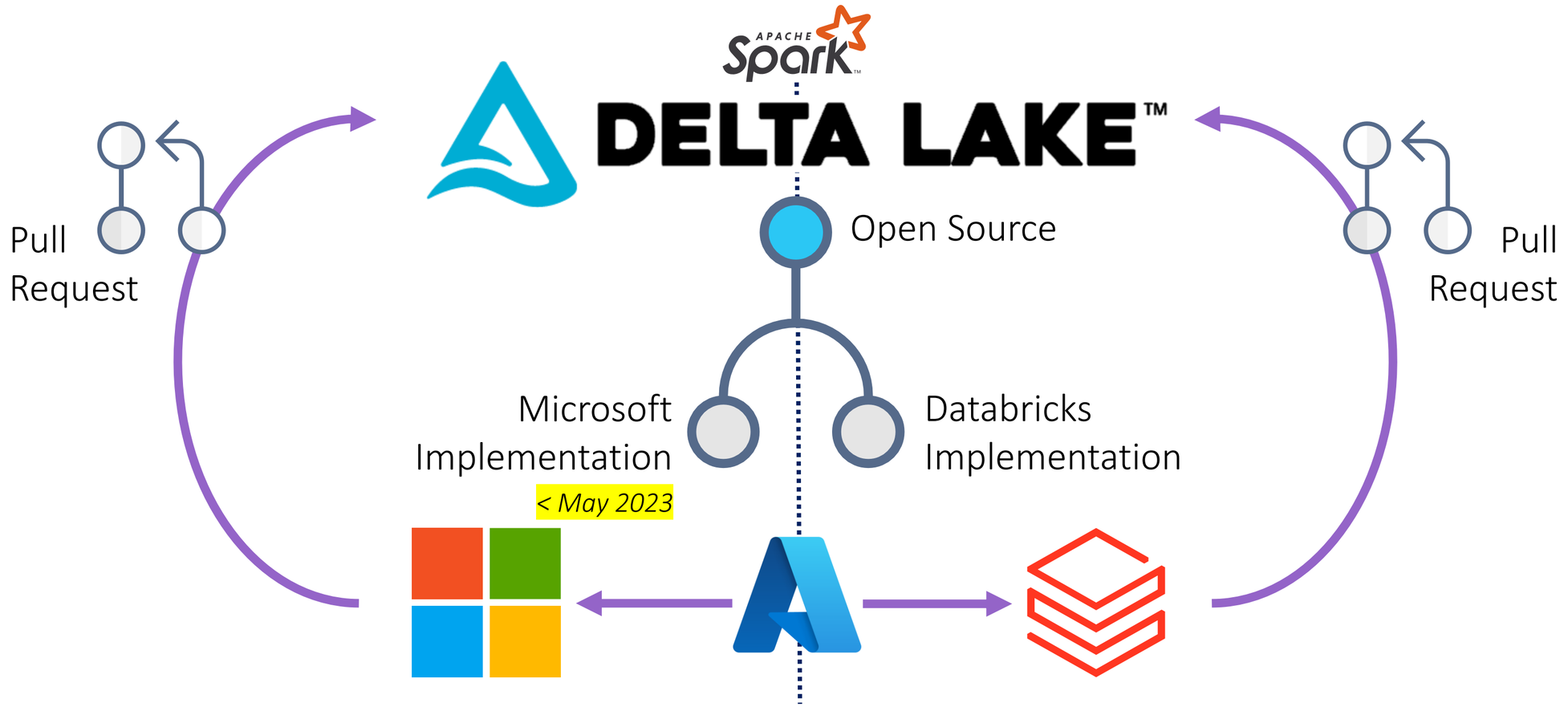
databricks®

What is Delta Lake?



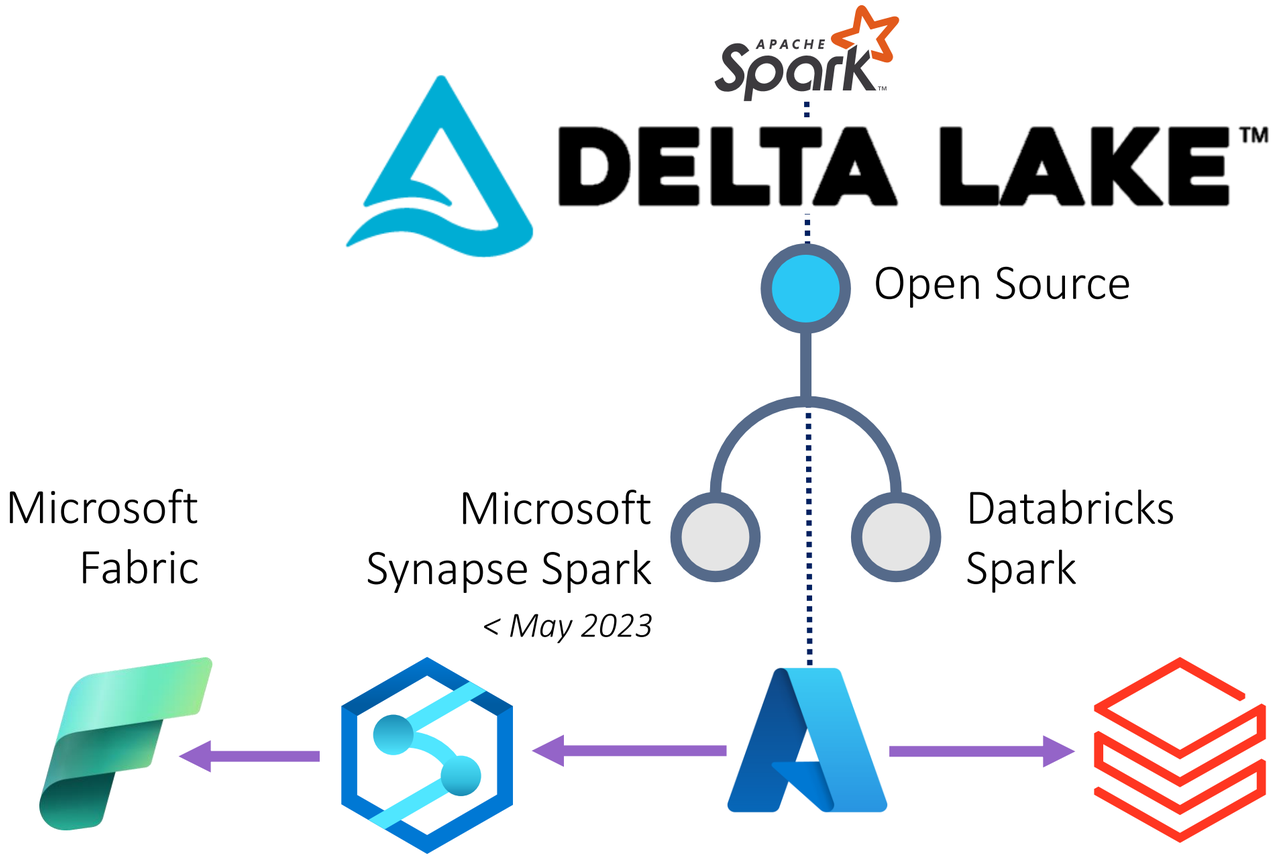
"Delta Lake is an open-source storage layer that brings ACID transactions to Apache Spark™ and big data workloads."

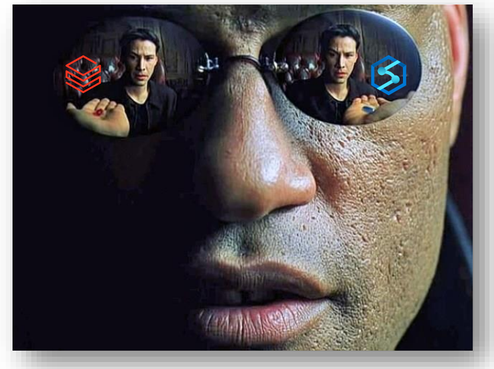
What is Delta Lake?



"Delta Lake is an open-source storage layer that brings ACID transactions to Apache Spark™ and big data workloads."

Which Spark Implementation is Better?





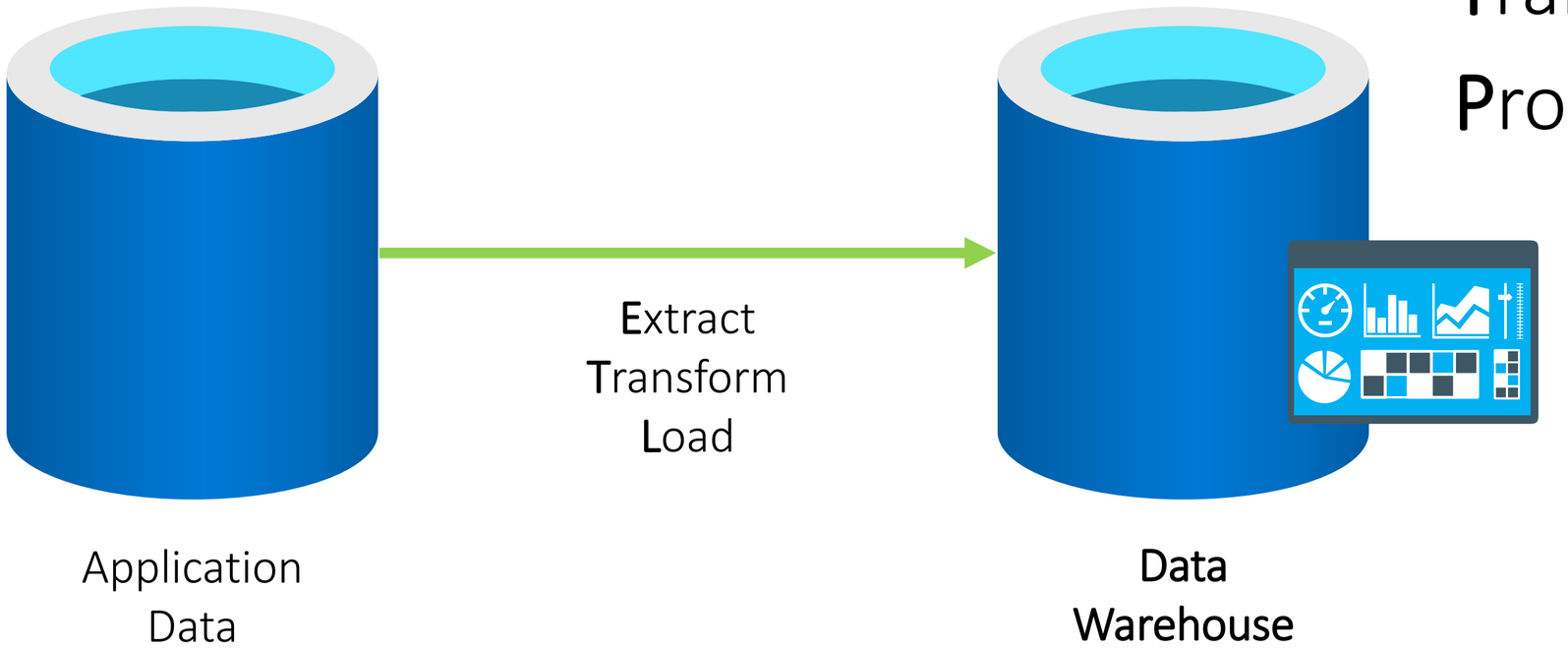
Data Warehouse

Online

Line

Transactional

Processing



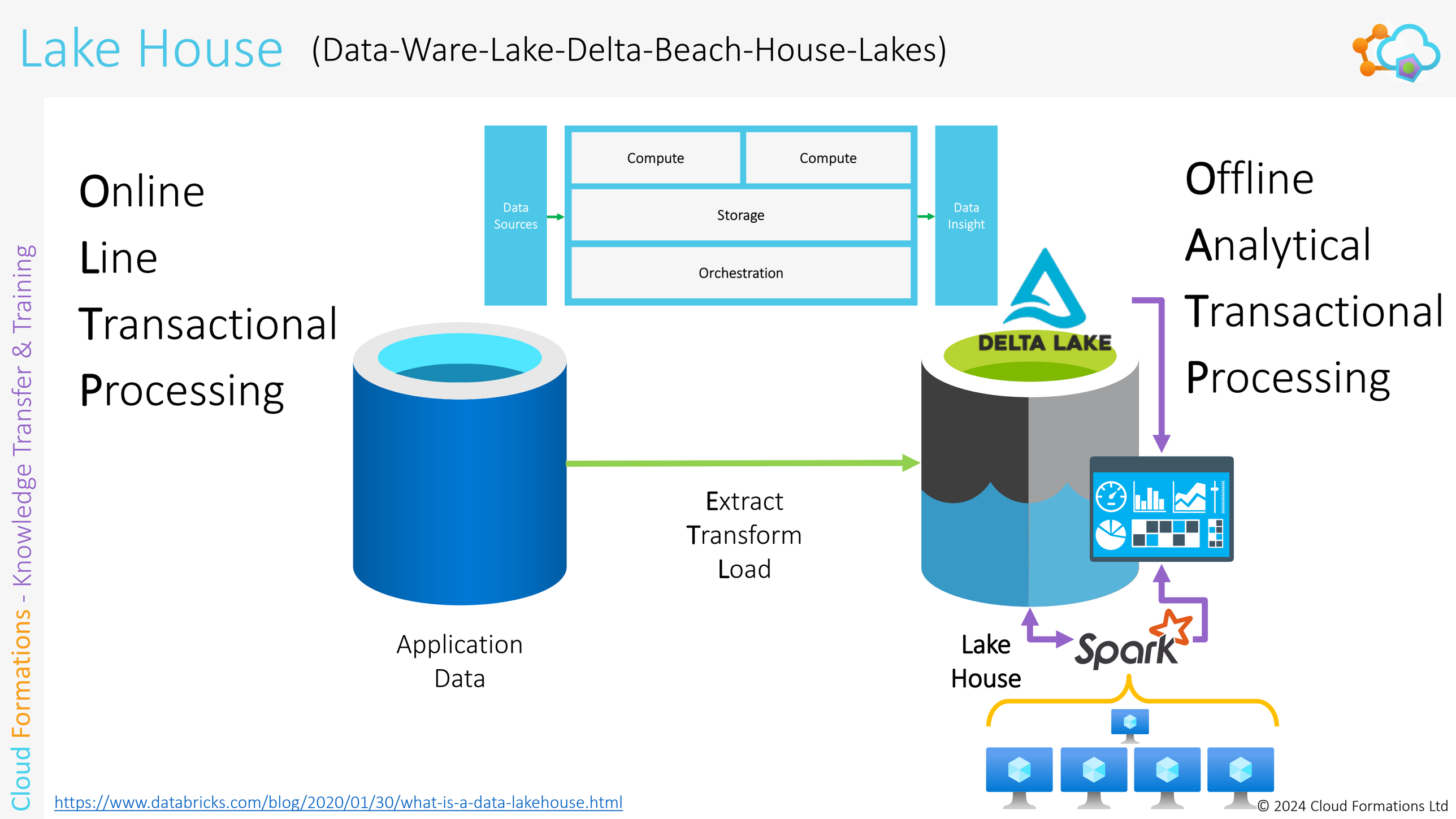
Offline

Analytical

Transactional

Processing

Lake House (Data-Ware-Lake-Delta-Beach-House-Lakes)



https://www.databricks.com/blog/2020/01/30/what-is-a-data-lakehouse.html

Lake House

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Learn to edit
Community portal
Recent changes
Upload file

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Cite this page
Wikidata item

Print/export
Download as PDF
Printable version

Languages
العربية
Deutsch
Español
Français

The Lake House (film)

From Wikipedia, the free encyclopedia

This article includes a list of general references, but it remains largely unverified because **it lacks sufficient corresponding inline citations**. Please help to improve this article by introducing more precise citations. *(October 2017) (Learn how and when to remove this template message)*

The Lake House is a 2006 American fantasy romantic drama film directed by Alejandro Agresti, starring Keanu Reeves and Sandra Bullock (who had previously appeared together in the box office hit *Speed*). It was written by David Auburn.[2] A remake of the South Korean motion picture *Il Mare* (2000), it centers on an architect living in 2004 and a doctor living in 2006 who meet via letters left in a mailbox at the lake house where they have lived at separate points in time. They carry on correspondence over two years, remaining separated by their original difference of two years.[3]

Contents [hide]
1 Plot
2 Cast
3 Production
4 Music
5 Reception
5.1 Box office
5.2 Critical response
5.3 Home media
5.4 Awards
6 References
7 External links

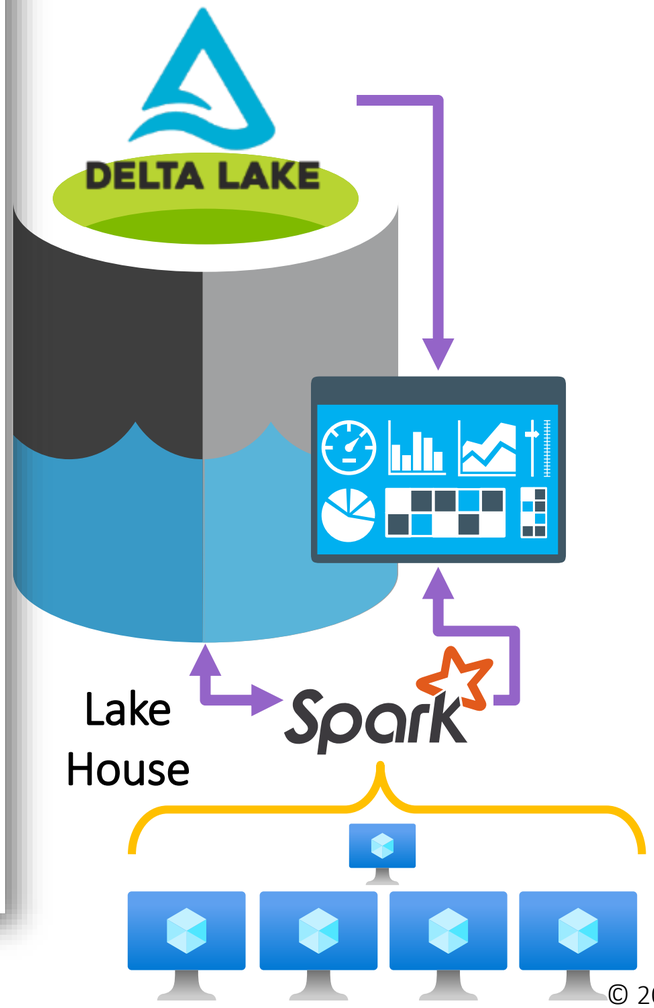
The Lake House

Theatrical release poster

Directed by	Alejandro Agresti
Written by	David Auburn
Based on	*Il Mare* by Kim Eun-jeong Kim Mi-yeong
Produced by	Doug Davison Roy Lee
Starring	Keanu Reeves

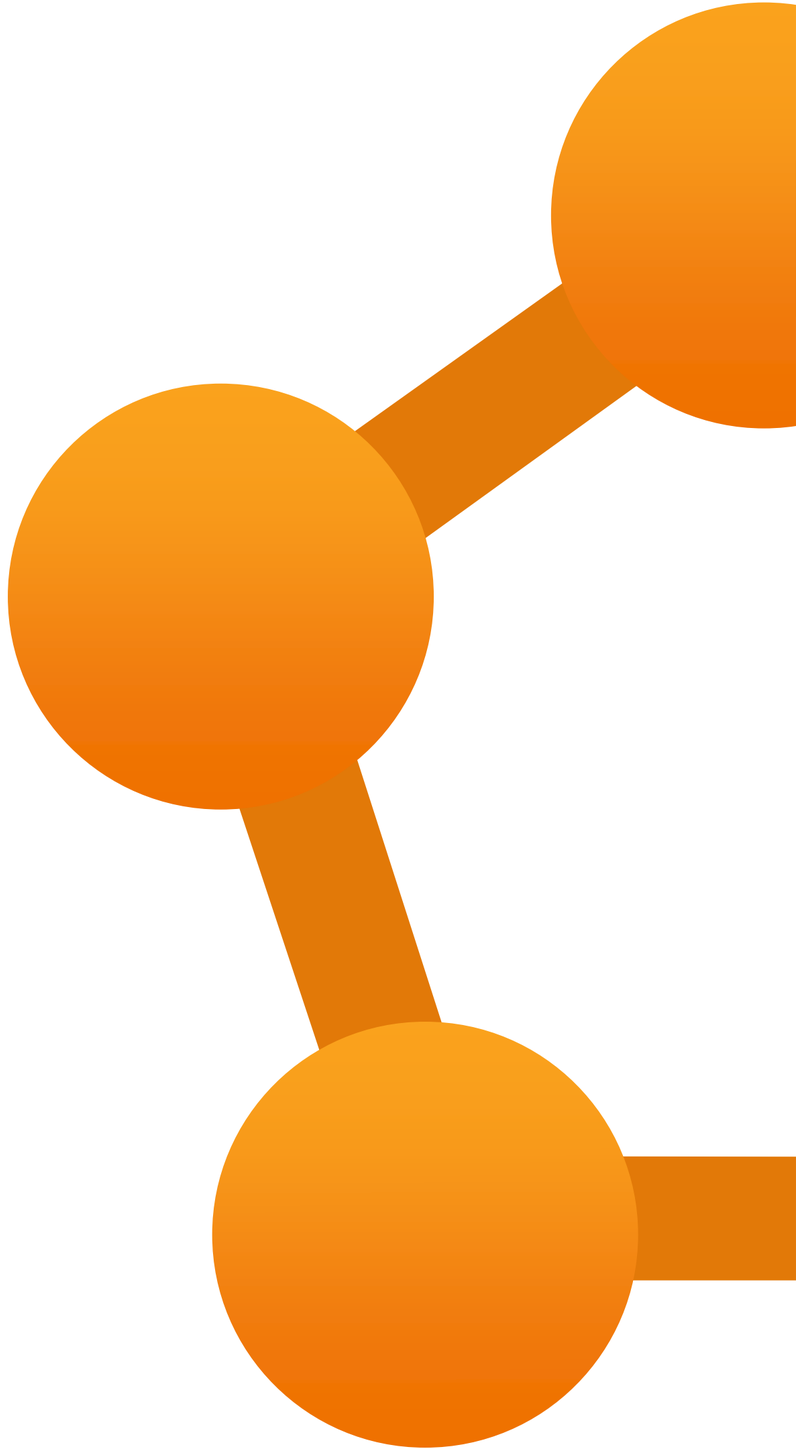
Plot [edit]

In 2006, Dr. Kate Forster (Sandra Bullock) is leaving a lake house that she has been renting in Chicago. Kate leaves a note in the mailbox for the next tenant to forward her mail, adding that the paint-embedded pawprints on the path leading to the house were already there when she arrived.



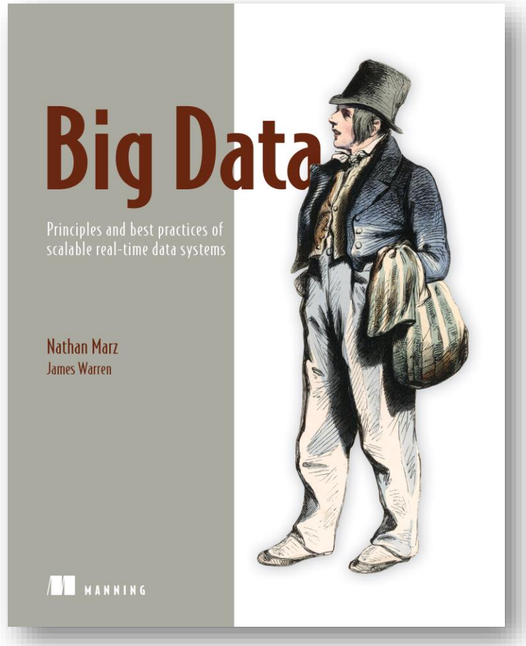
Lambda & Kappa

λ κ



Cloud Formations

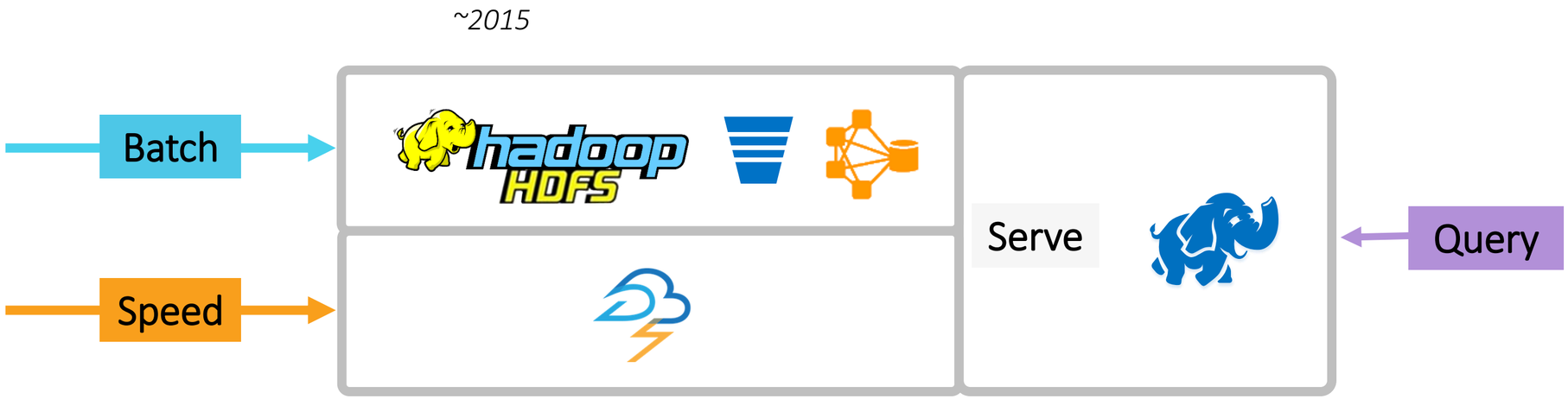
Lambda & Kappa Architectures



Big Data:

- Volume
- Velocity
- Variety
- Veracity
- Value

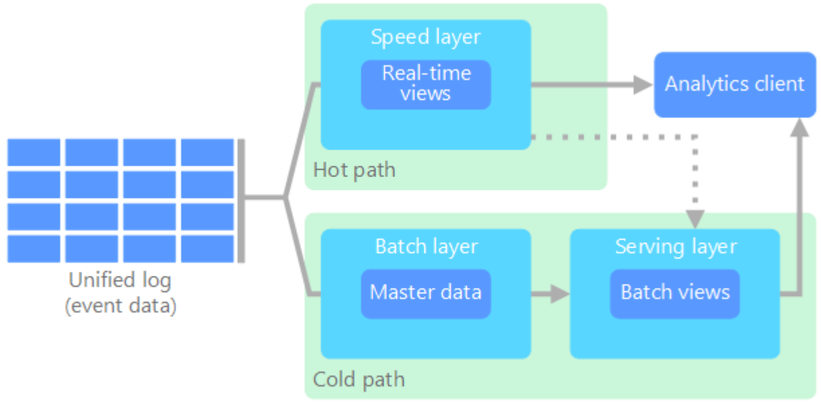
Lambda & Kappa Architectures



Lambda & Kappa Architectures

Lambda & Kappa Architectures

Lambda

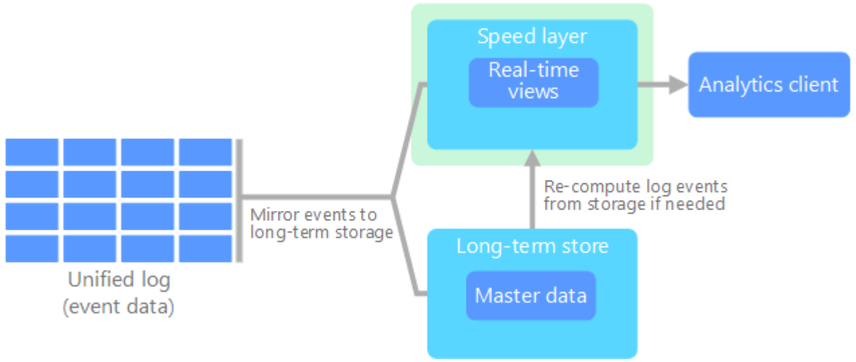


"The **lambda architecture**, first proposed by Nathan Marz, addresses this problem by creating two paths for data flow. All data coming into the system goes through these two paths:

A **batch layer** (cold path) stores all of the incoming data in its raw form and performs batch processing on the data. The result of this processing is stored as a **batch view**.

A **speed layer** (hot path) analyzes data in real time. This layer is designed for low latency, at the expense of accuracy."

Kappa

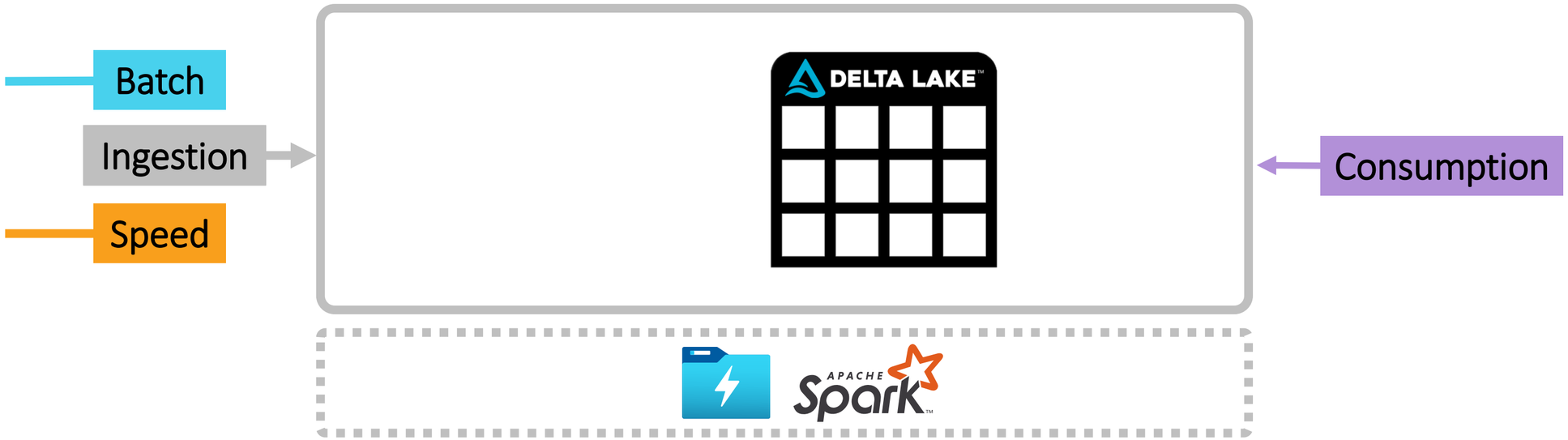


"A drawback to the lambda architecture is its complexity. Processing logic appears in two different places — the cold and hot paths — using different frameworks. This leads to duplicate computation logic and the complexity of managing the architecture for both paths.

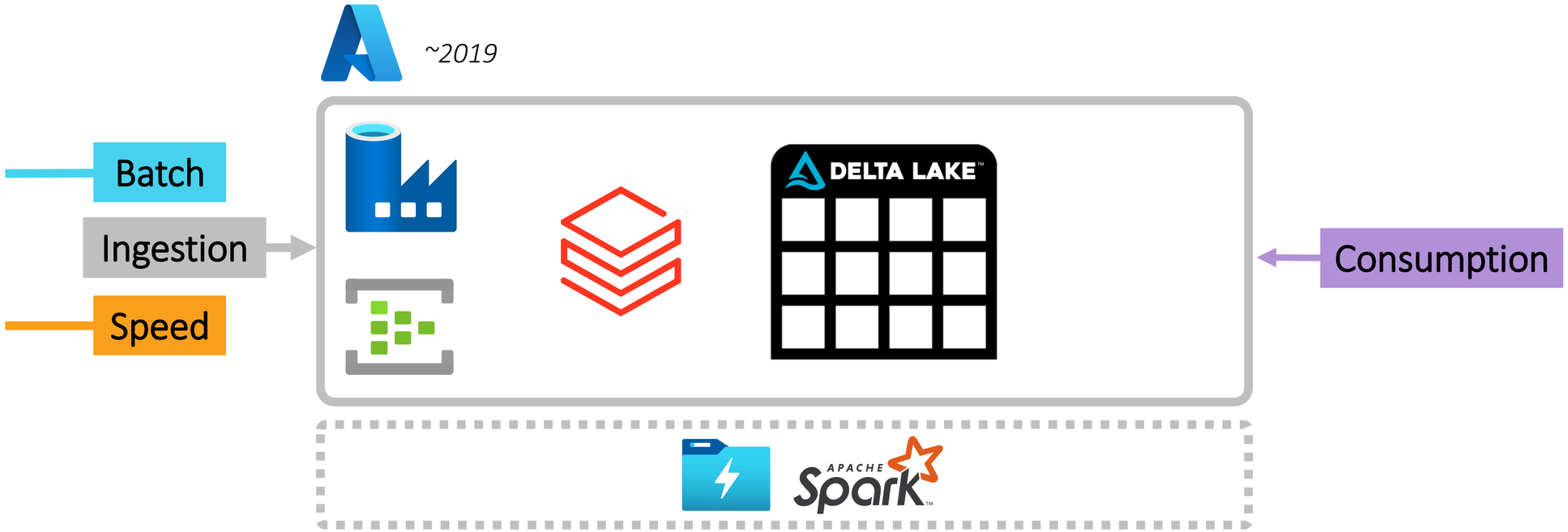
The **kappa architecture** was proposed by Jay Kreps as an alternative to the lambda architecture. It has the same basic goals as the lambda architecture, but with an important distinction: All data flows through a single path, using a stream processing system."

https://docs.microsoft.com/en-us/azure/architecture/data-guide/big-data/

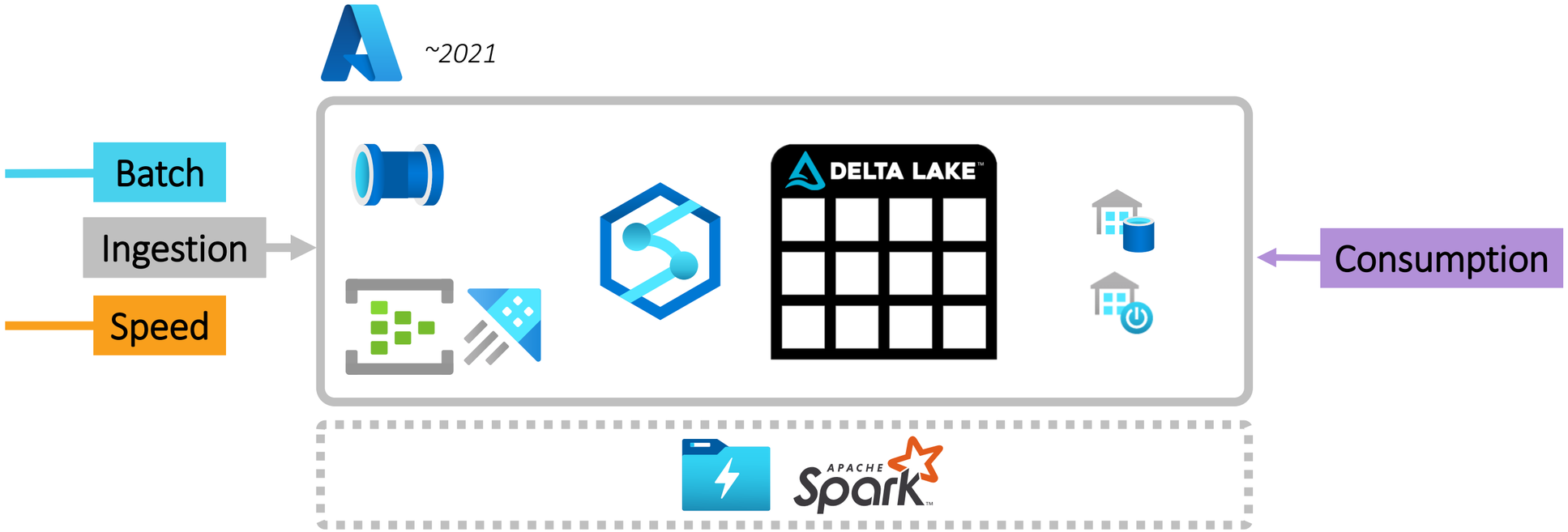
Lambda & Kappa Architectures



Lambda & Kappa Architectures



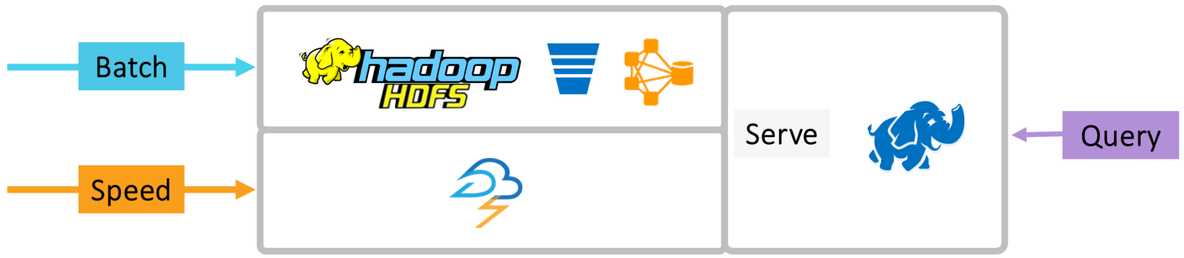
Lambda & Kappa Architectures



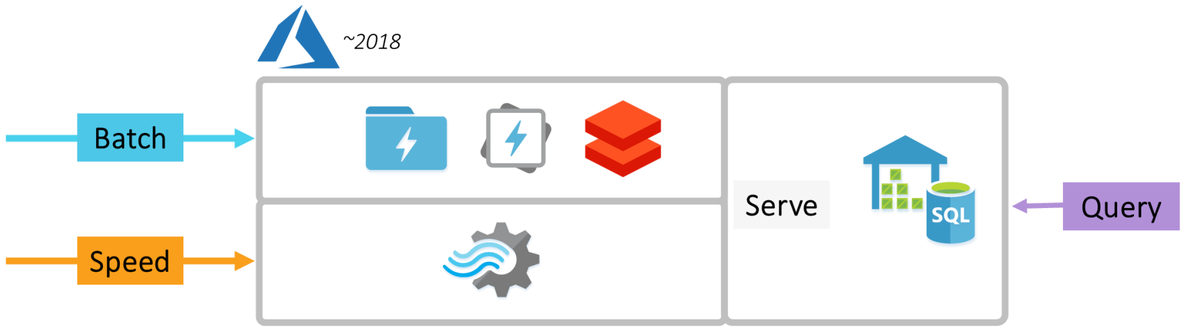
Lambda & Kappa Architectures vs Technology

Lambda

~2015

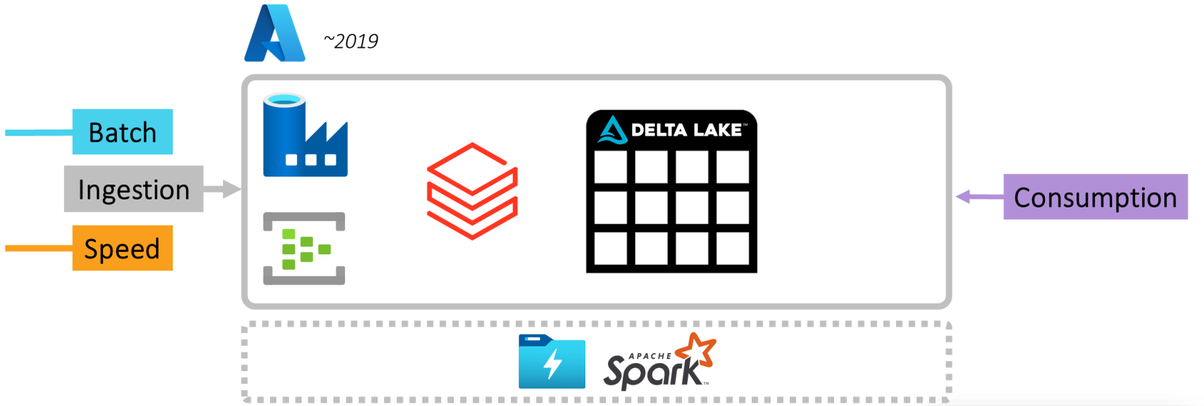


~2018

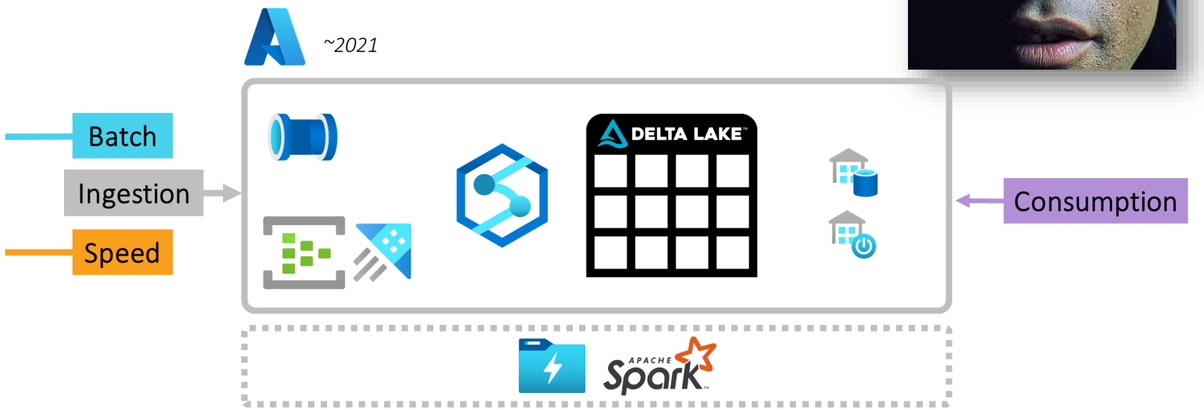


Kappa

~2019

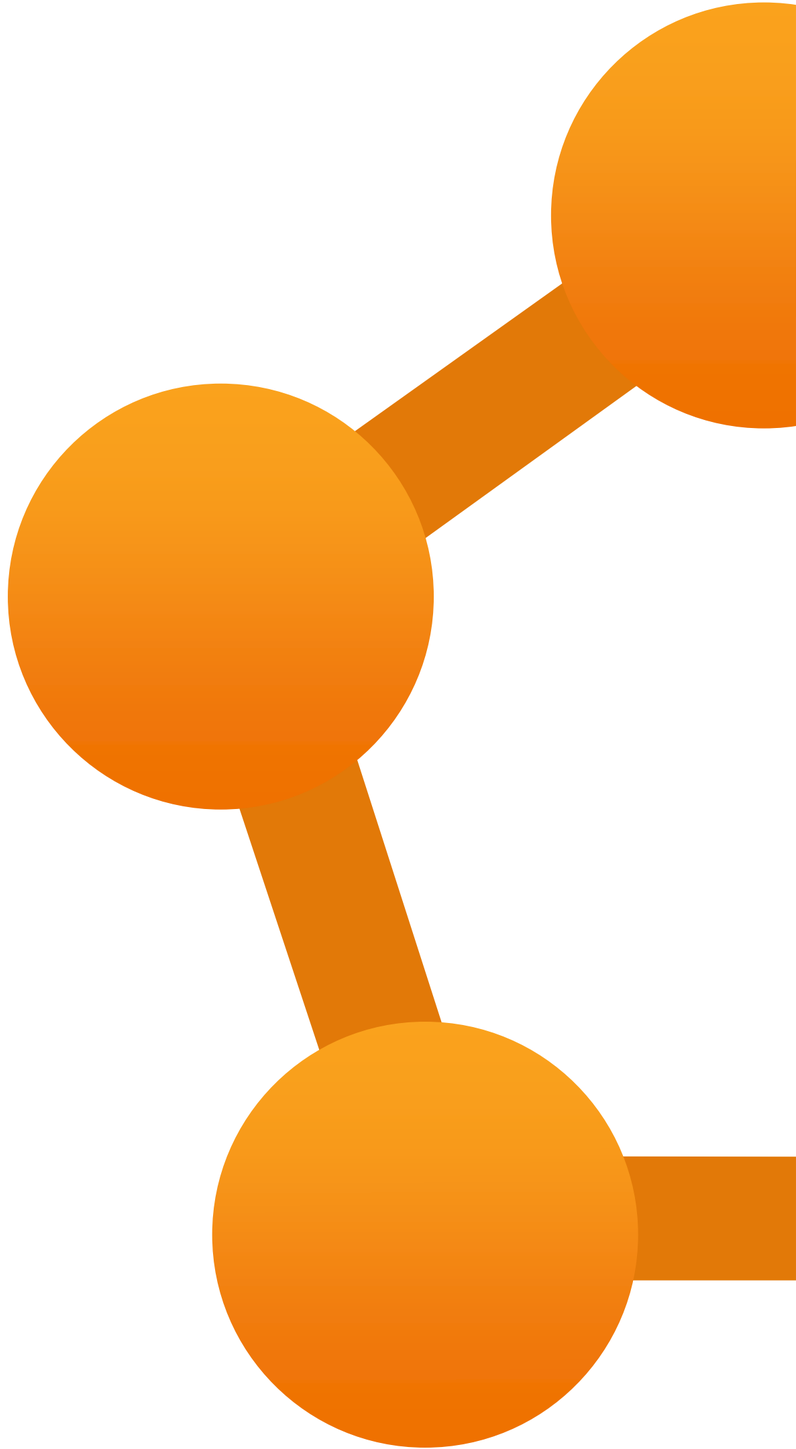


~2021



Cloud Formations - Knowledge Transfer & Training

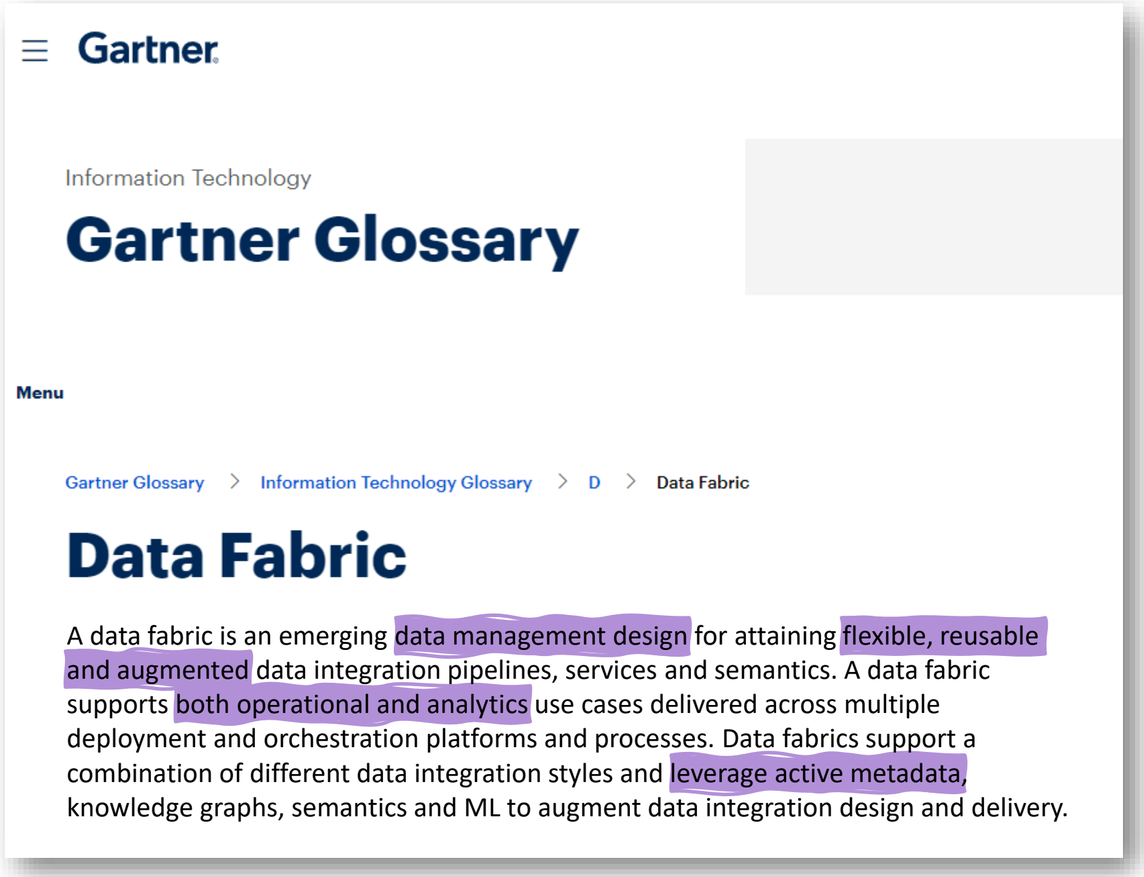
Fabric



Cloud Formations

Different Types of Fabric

Different Types of Fabric



Gartner

Information Technology

Gartner Glossary

Menu

Gartner Glossary > Information Technology Glossary > D > Data Fabric

Data Fabric

A data fabric is an emerging data management design for attaining flexible, reusable and augmented data integration pipelines, services and semantics. A data fabric supports both operational and analytics use cases delivered across multiple deployment and orchestration platforms and processes. Data fabrics support a combination of different data integration styles and leverage active metadata, knowledge graphs, semantics and ML to augment data integration design and delivery.

Ref: https://www.gartner.com/en/information-technology/glossary/data-fabric

Microsoft Documentation

Learn / Microsoft Fabric / Get started /

What is Microsoft Fabric?

Article • 11/15/2023 • 6 contributors

Feedback

In this article

- SaaS foundation
- Components of Microsoft Fabric
- OneLake and lakehouse - the unification of lakehouses
- Fabric solutions for ISVs
- Next steps

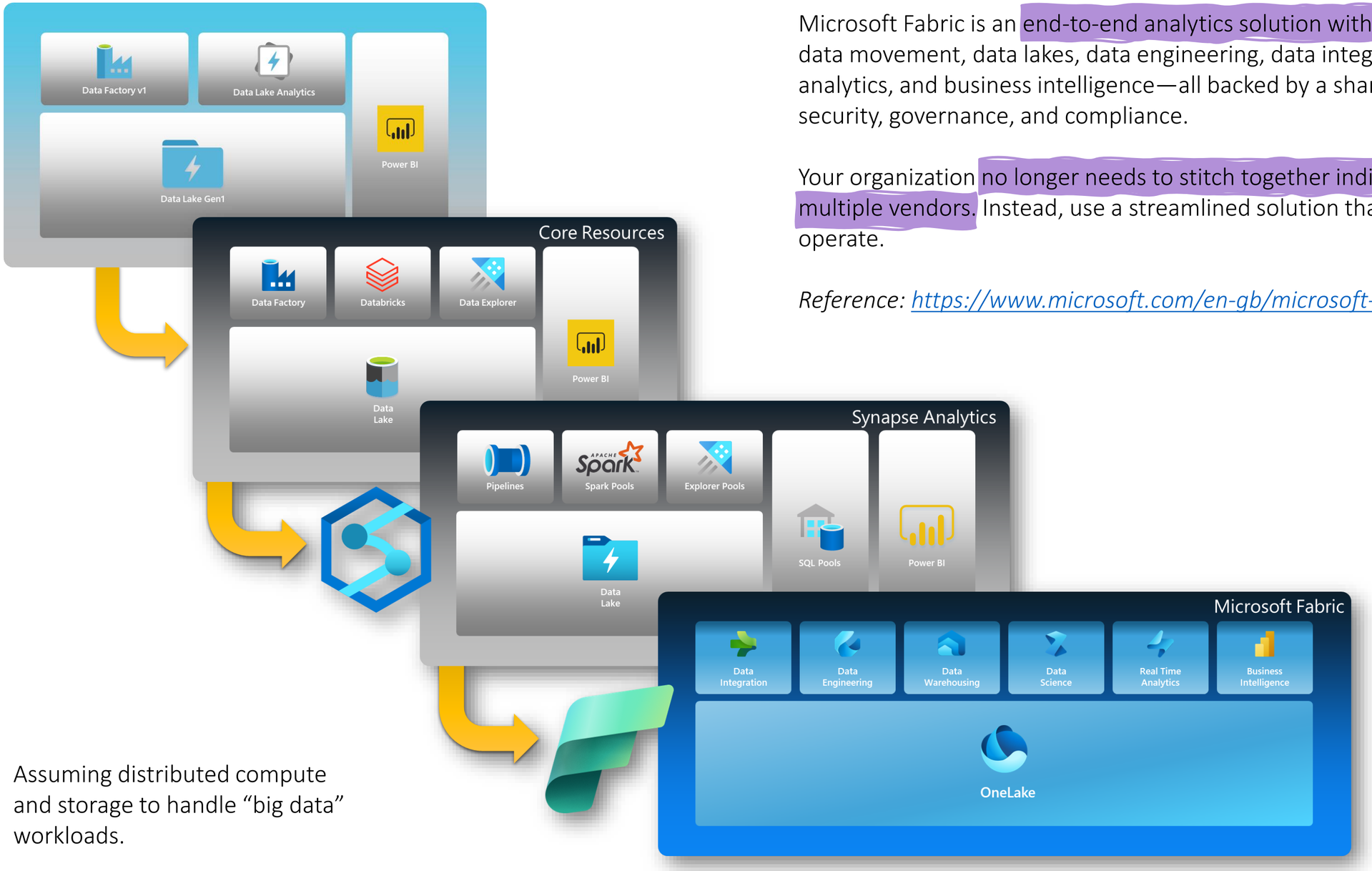
Microsoft Fabric is an all-in-one analytics solution for enterprises that covers everything from data movement to data science, Real-Time Analytics, and business intelligence. It offers a comprehensive suite of services, including data lake, data engineering, and data integration, all in one place.

With Fabric, you don't need to piece together different services from multiple vendors. Instead, you can enjoy a highly integrated, end-to-end, and easy-to-use product that is designed to simplify your analytics needs.

The platform is built on a foundation of Software as a Service (SaaS), which takes simplicity and integration to a whole new level.

Ref: https://www.gartner.com/en/information-technology/glossary/data-fabric

What is Microsoft Fabric? – Vision and Stack Evolution



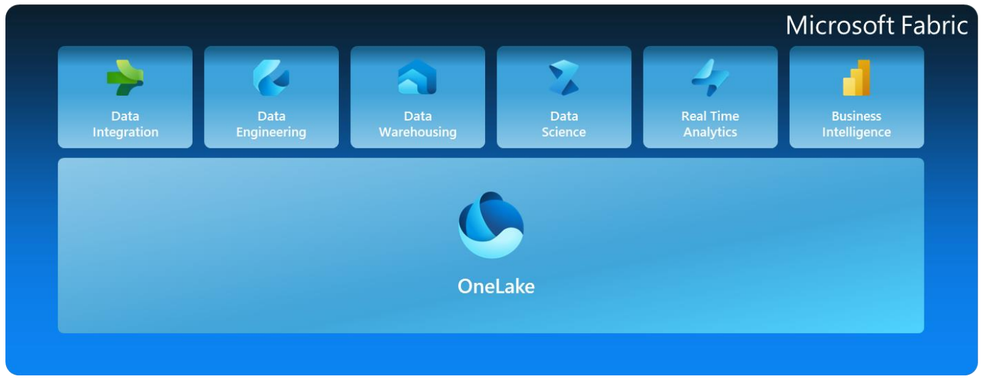
Microsoft Fabric is an end-to-end analytics solution with full-service capabilities including data movement, data lakes, data engineering, data integration, data science, real-time analytics, and business intelligence—all backed by a shared platform providing robust data security, governance, and compliance.

Your organization no longer needs to stitch together individual analytics services from multiple vendors. Instead, use a streamlined solution that's easy to connect, onboard, and operate.

Reference: https://www.microsoft.com/en-gb/microsoft-fabric

Assuming distributed compute and storage to handle "big data" workloads.

What is Microsoft Fabric?



What is Microsoft Fabric? - Experiences vs Technical Capabilities

Data Integration

- Workload management and orchestration with dependency chain handling and scheduling.

Think… Azure Data Factory

Data Engineering

- Low code and full code development in **Python**, **Scala**, **R**, **SQL** executed using Spark clusters.

Think… Azure Databricks

Data Warehouse

- Schema driven relational entities coded using **T-SQL** covering DML and DDL functionality.

Think… Azure Synapse Analytics – SQL Pools.

Data Science

- Predictive analytics and experimentation on data to expose insights and drive outcomes.

Think… Azure Machine Learning.

Real-time Analytics

- Time series data exploration over inbound telemetry and messages coded using **KQL**.

Think… Azure Data Explorer.

Business Intelligence

- Dashboards and metrics created to for the business to consume data, coded using **DAX** and **M**.

Think… Power BI.

Data Activator

- Actions triggered by defined changes in modelled datasets.

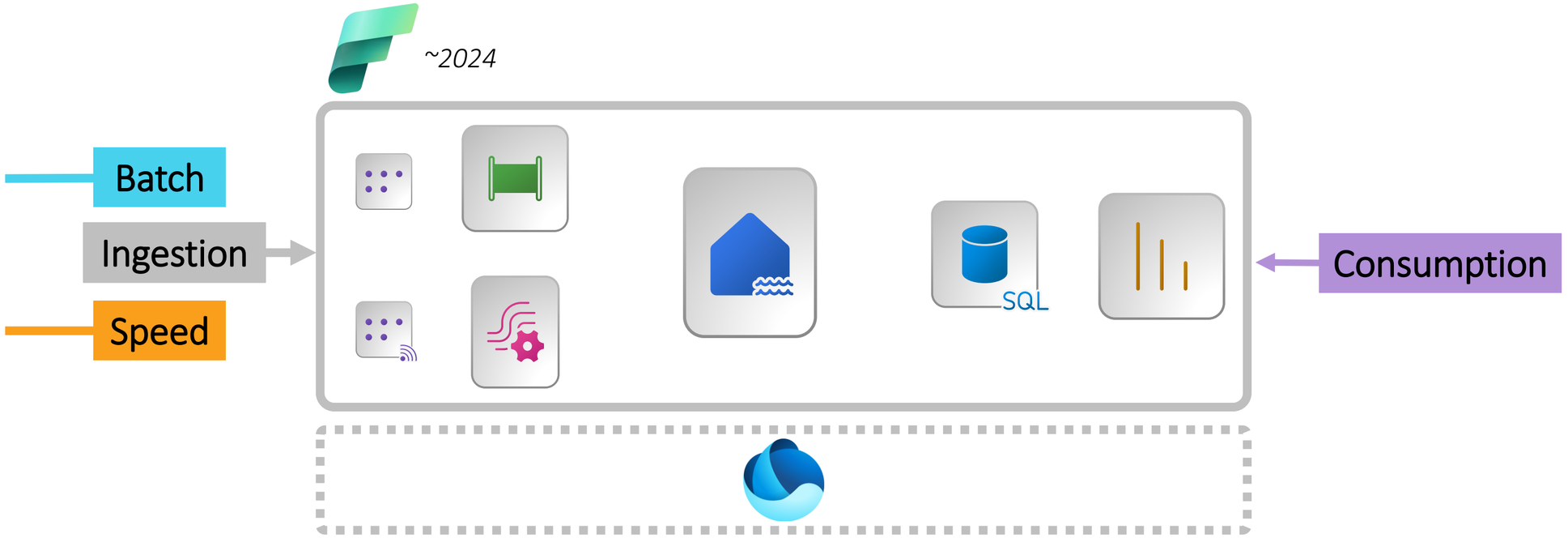
Think… Azure Alerts * *Closest equivalent capability.*

One Lake

- Distributed data storage optimised for analytics and structured as Delta Lake tables.

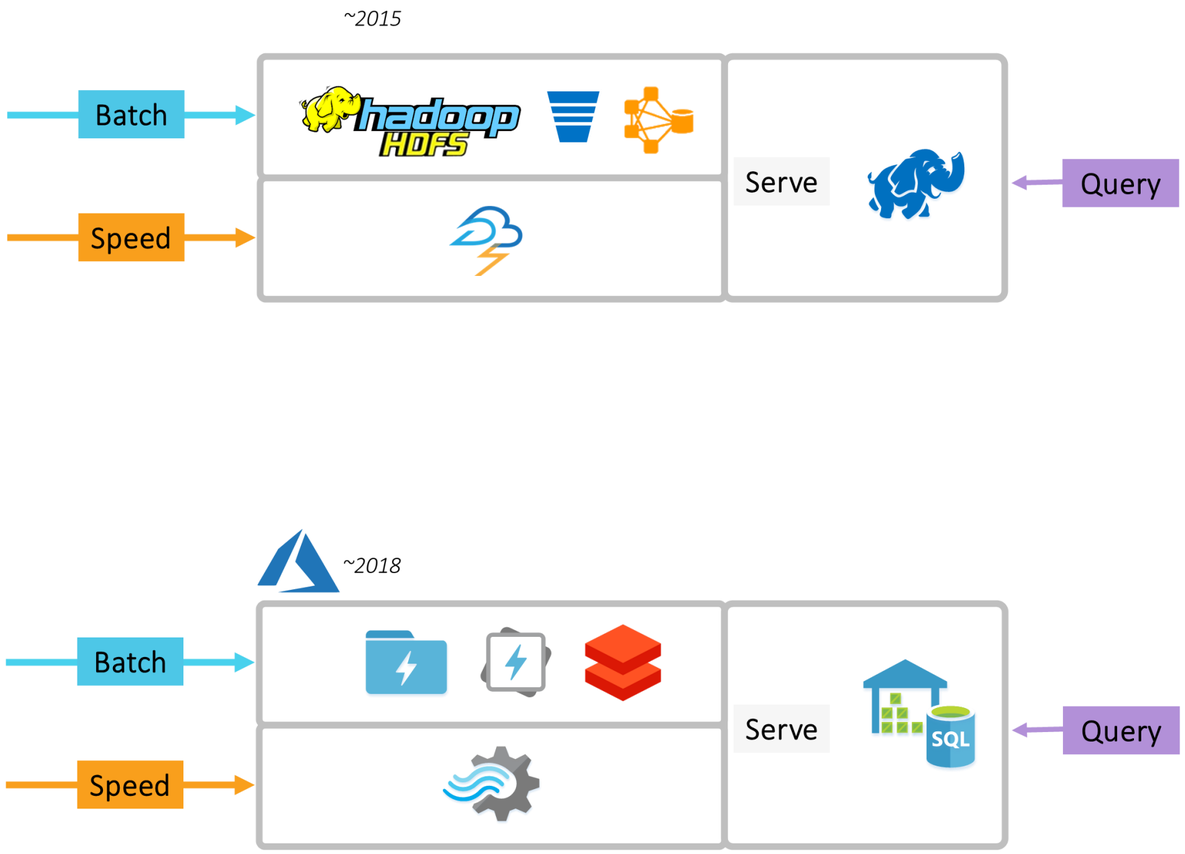
Think… Azure Data Lake Gen2

Microsoft Fabric vs a Kappa Architecture

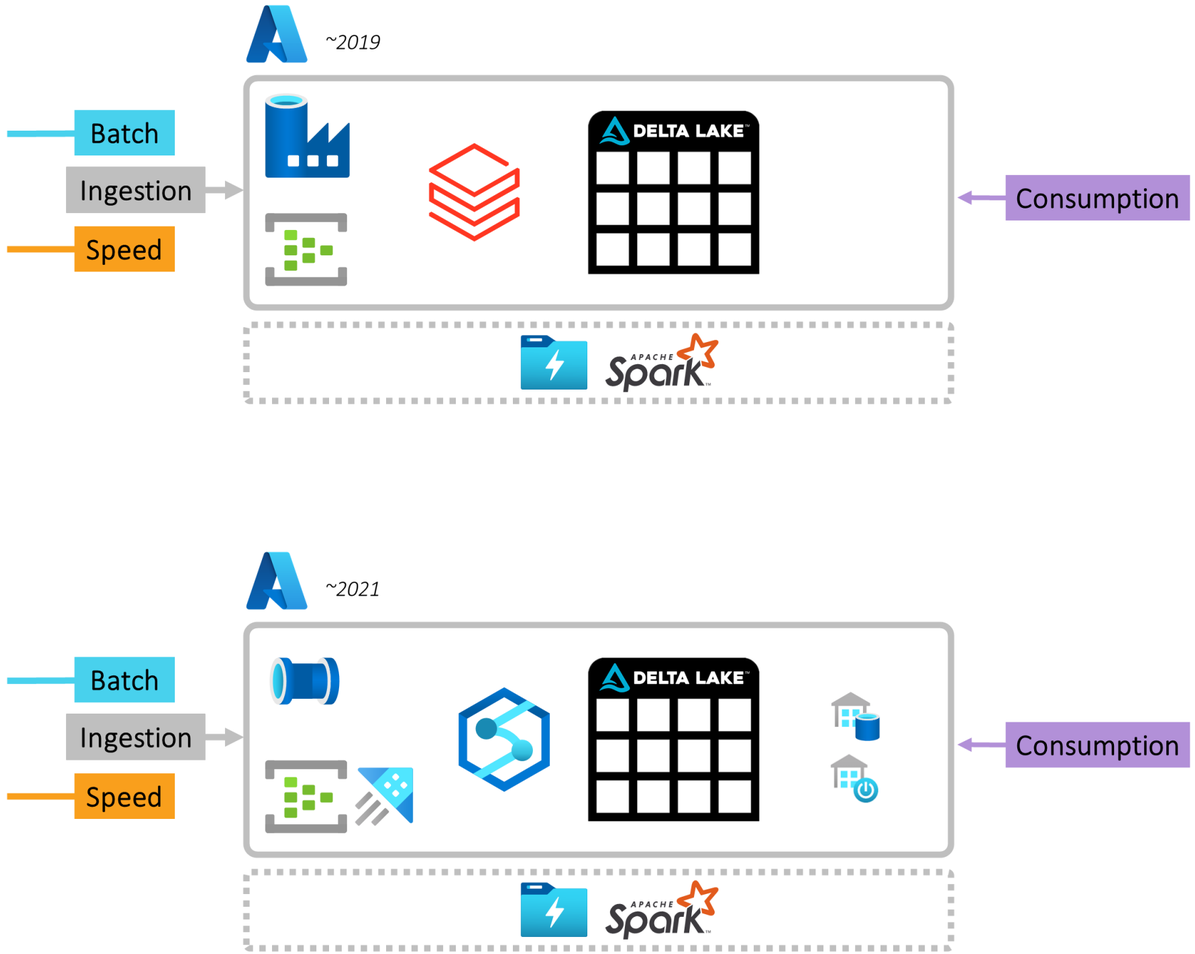


Lambda & Kappa Architectures vs Technology

Lambda



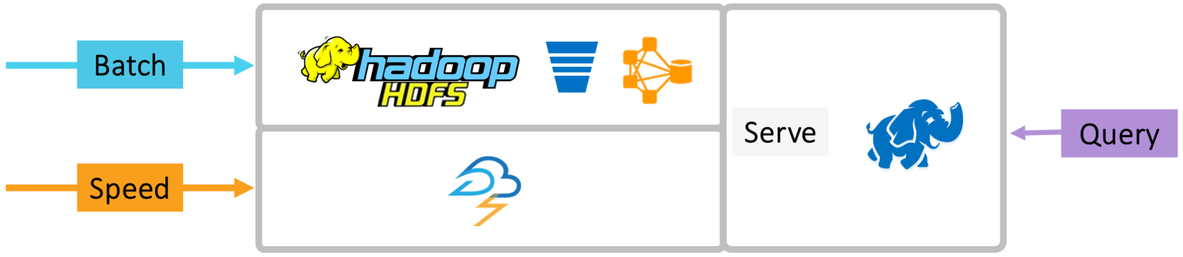
Kappa



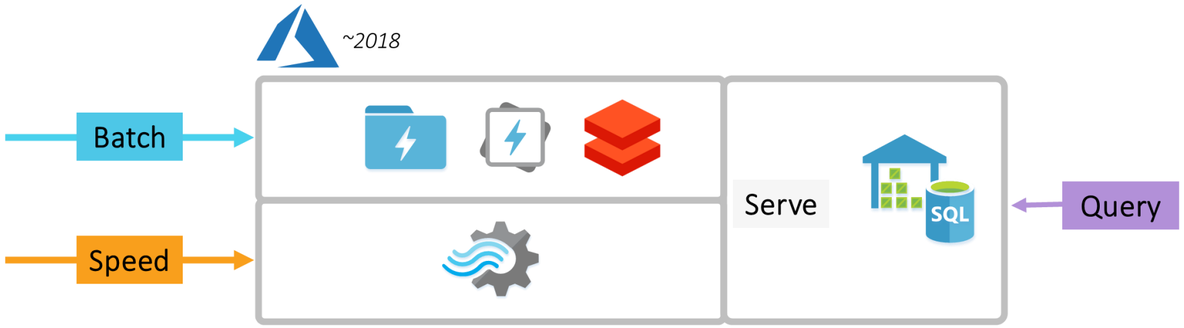
Lambda & Kappa Architectures vs Technology

Lambda

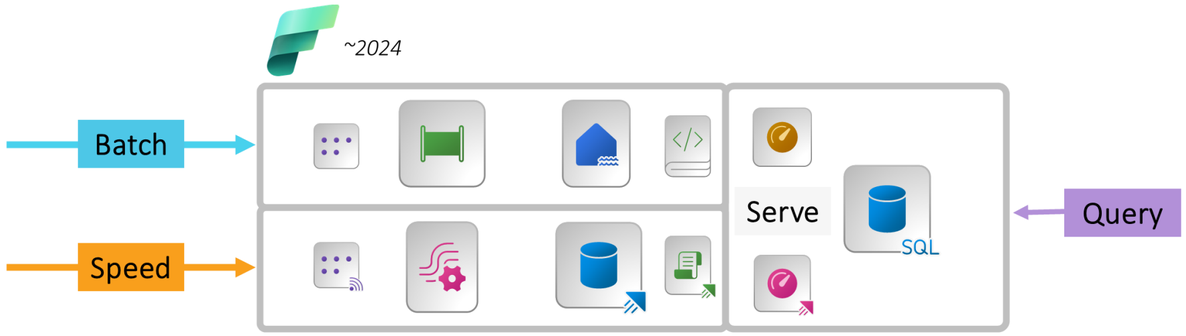
~2015



~2018

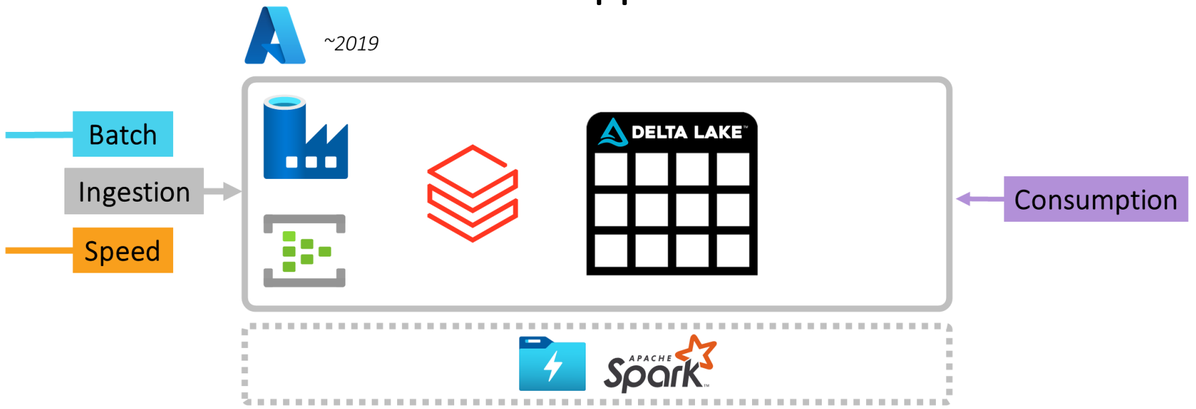


~2024

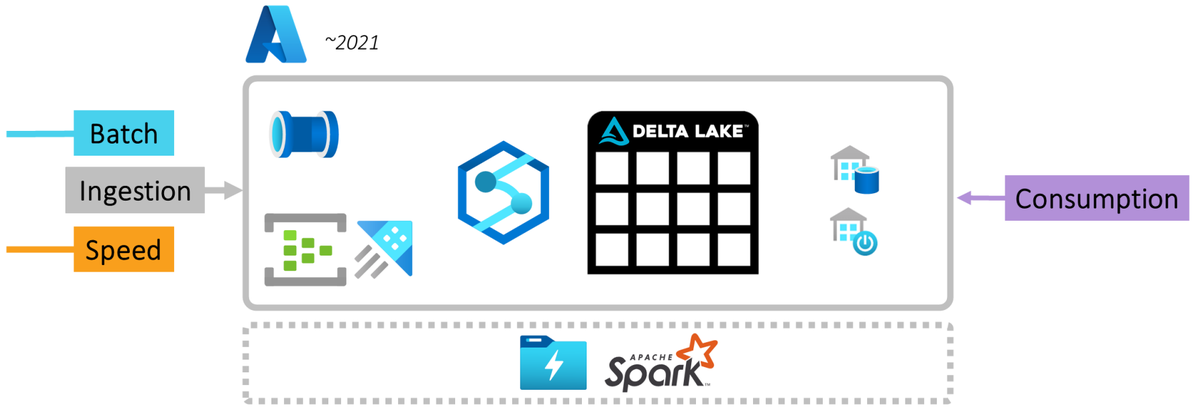


Kappa

~2019



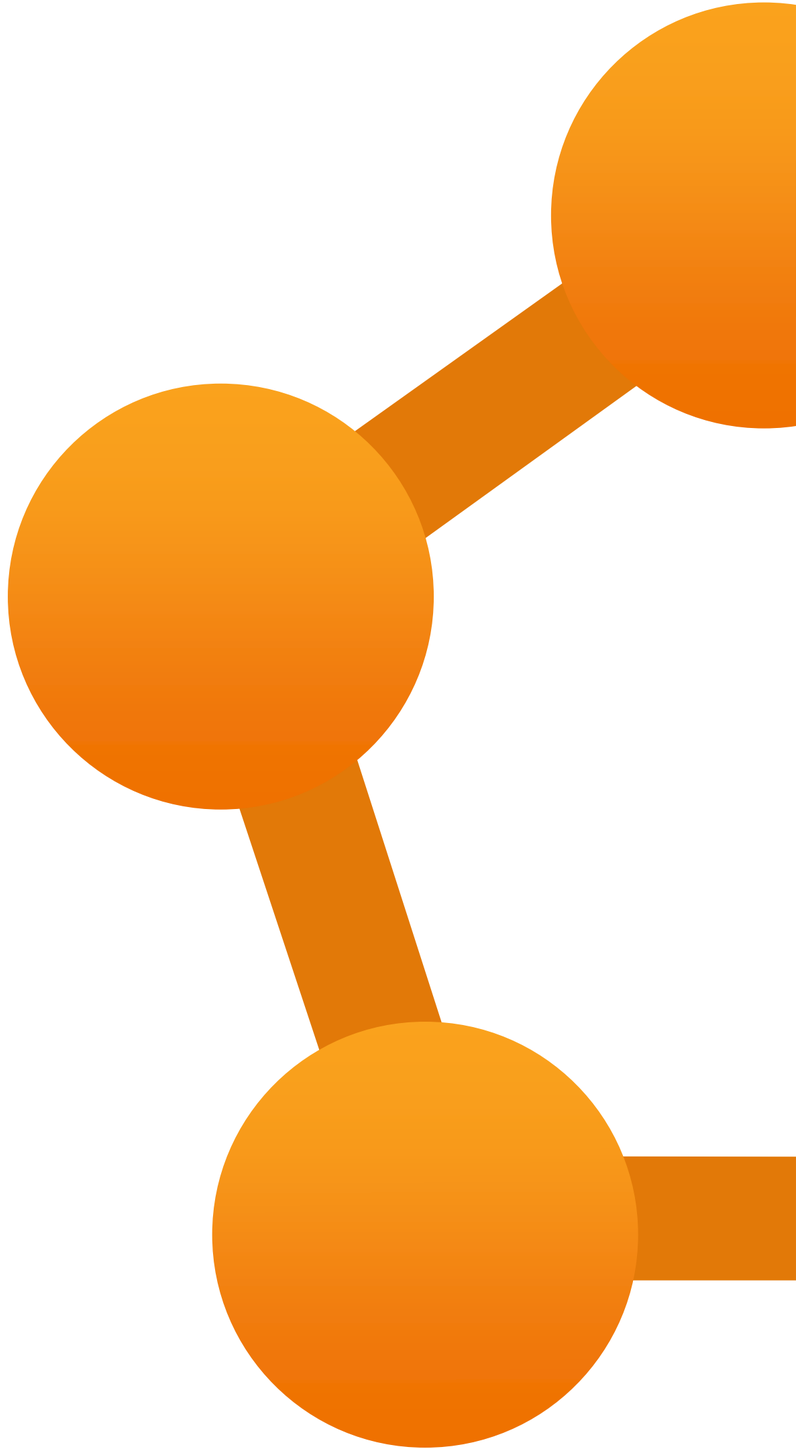
~2021



~2024

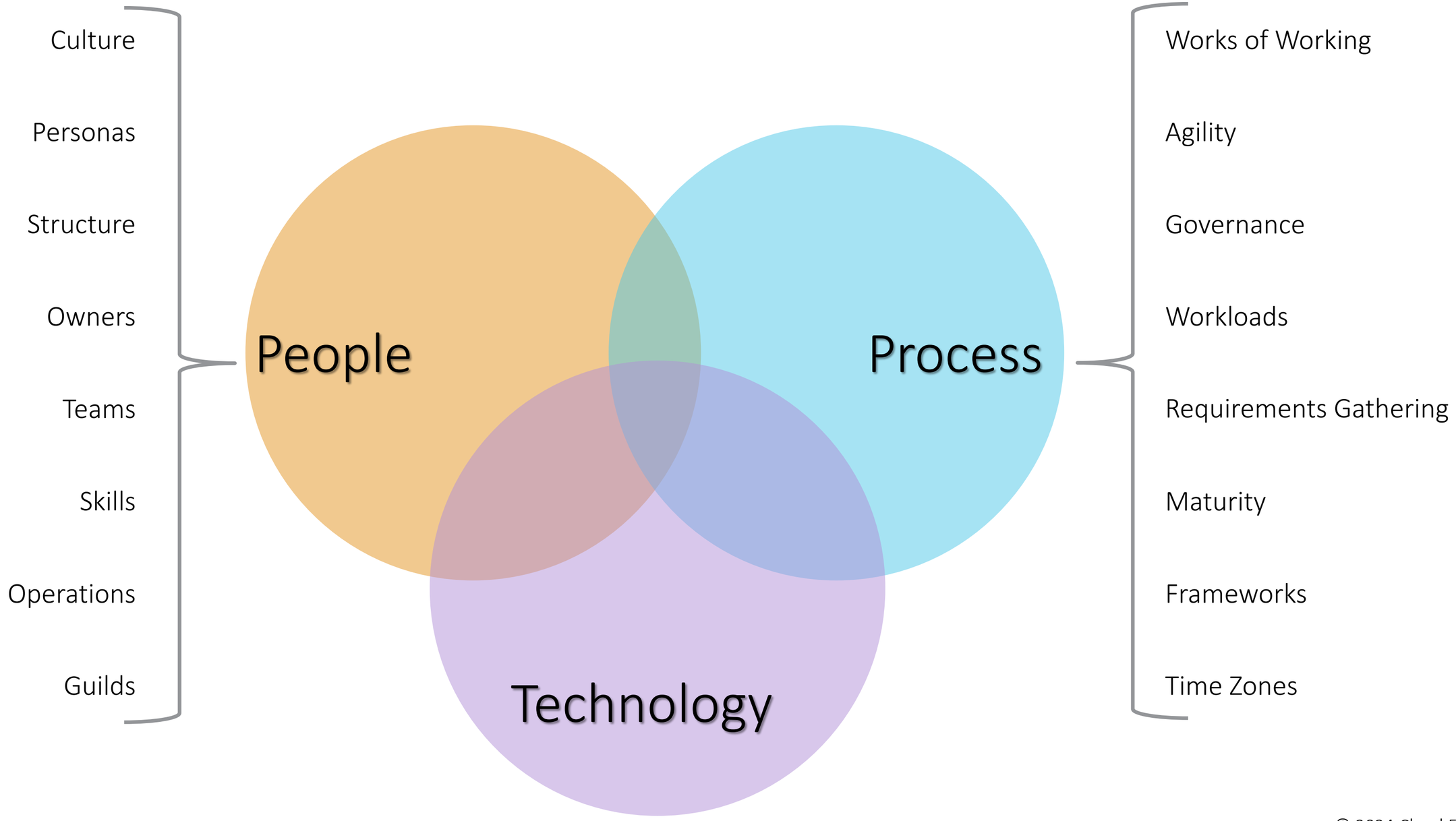
Cloud Formations - Knowledge Transfer & Training

Mesh

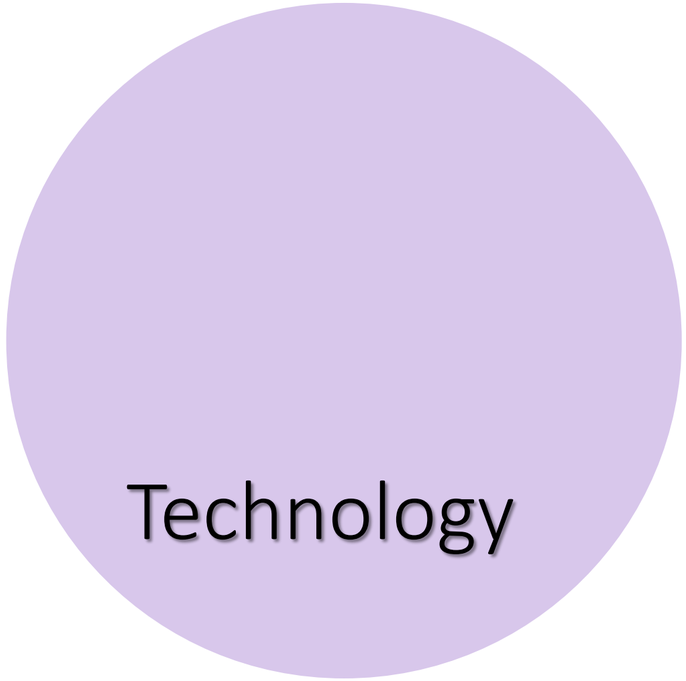


Cloud Formations

Data Mesh – What is it about?



Data Mesh – What is it about?



Data Mesh – What is it about?

Zhamak Dehghani

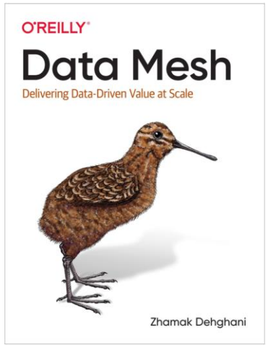
@zhamakd

https://martinfowler.com/articles/data-mesh-principles.html

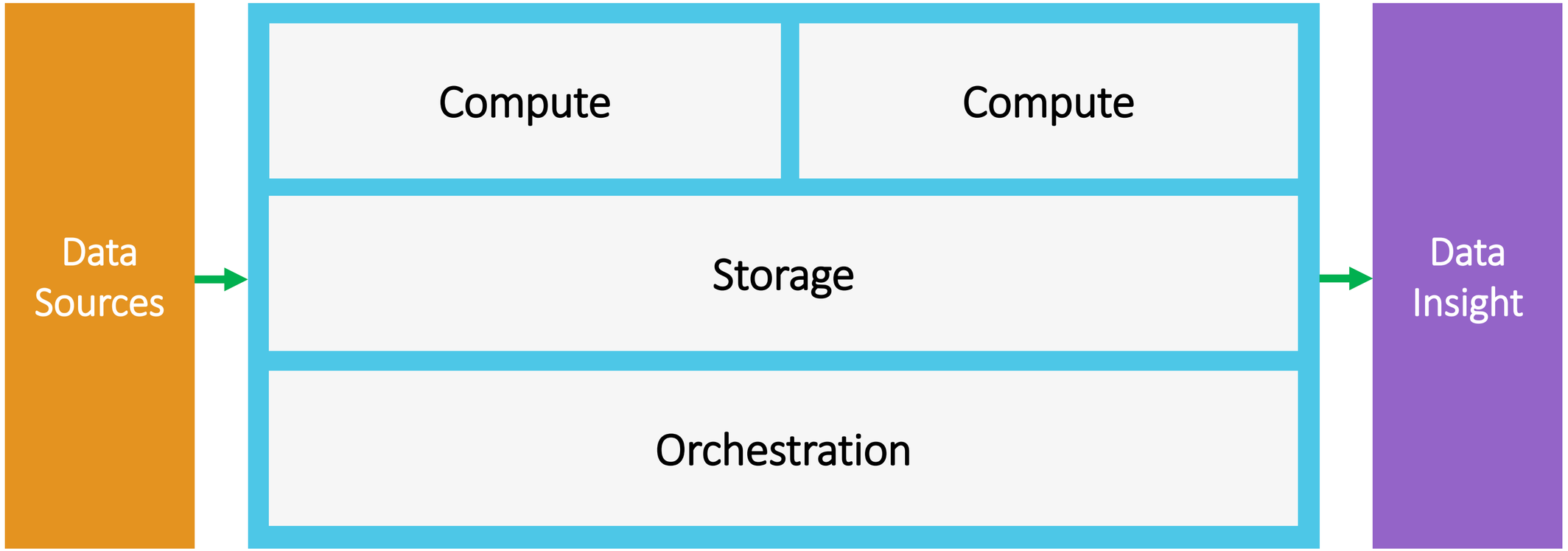
1. Domain-oriented decentralised data ownership and architecture.
2. Data as a product.
3. Self-serve data infrastructure as a platform.
4. Federated computational governance.

ISBN-10
1492092398

ISBN-13
978-1492092391

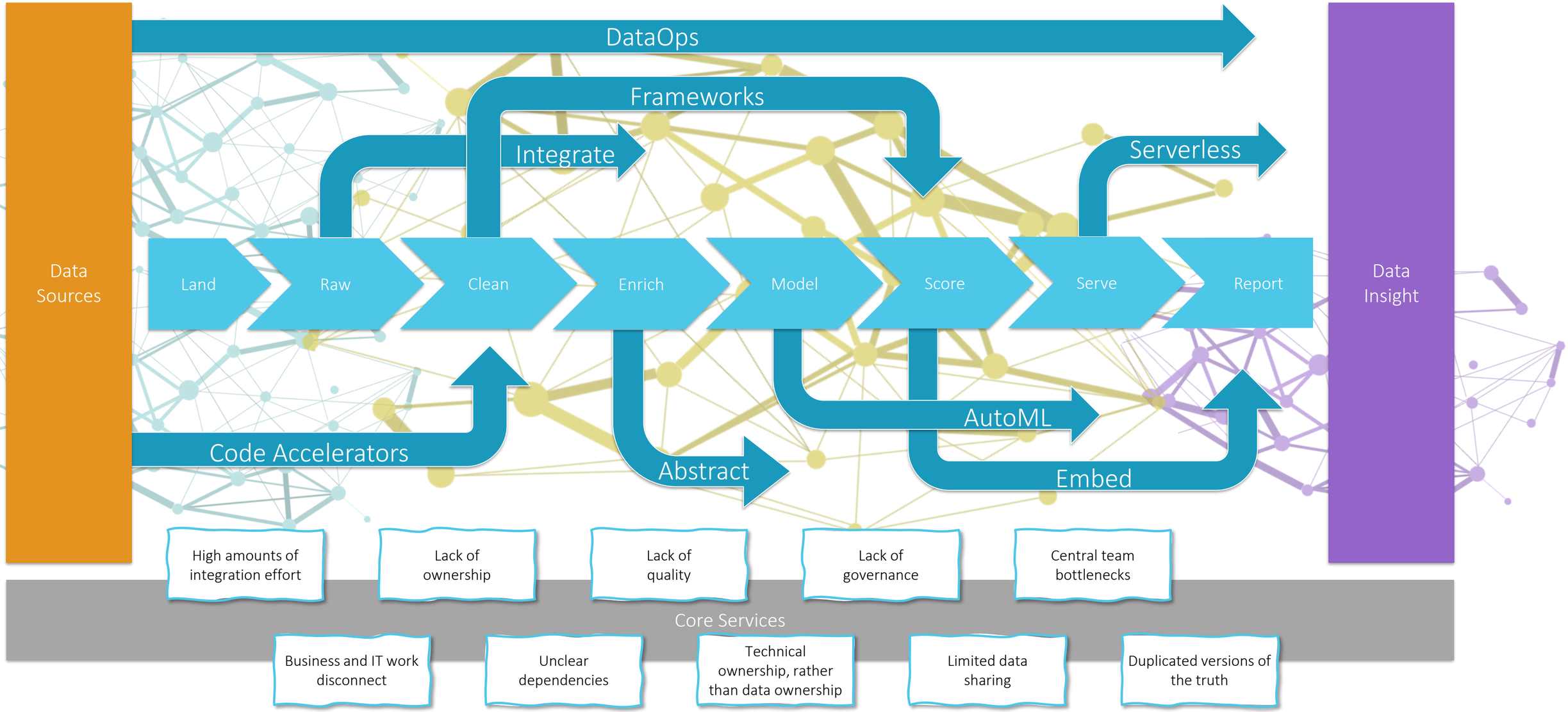


My First Reference Architecture

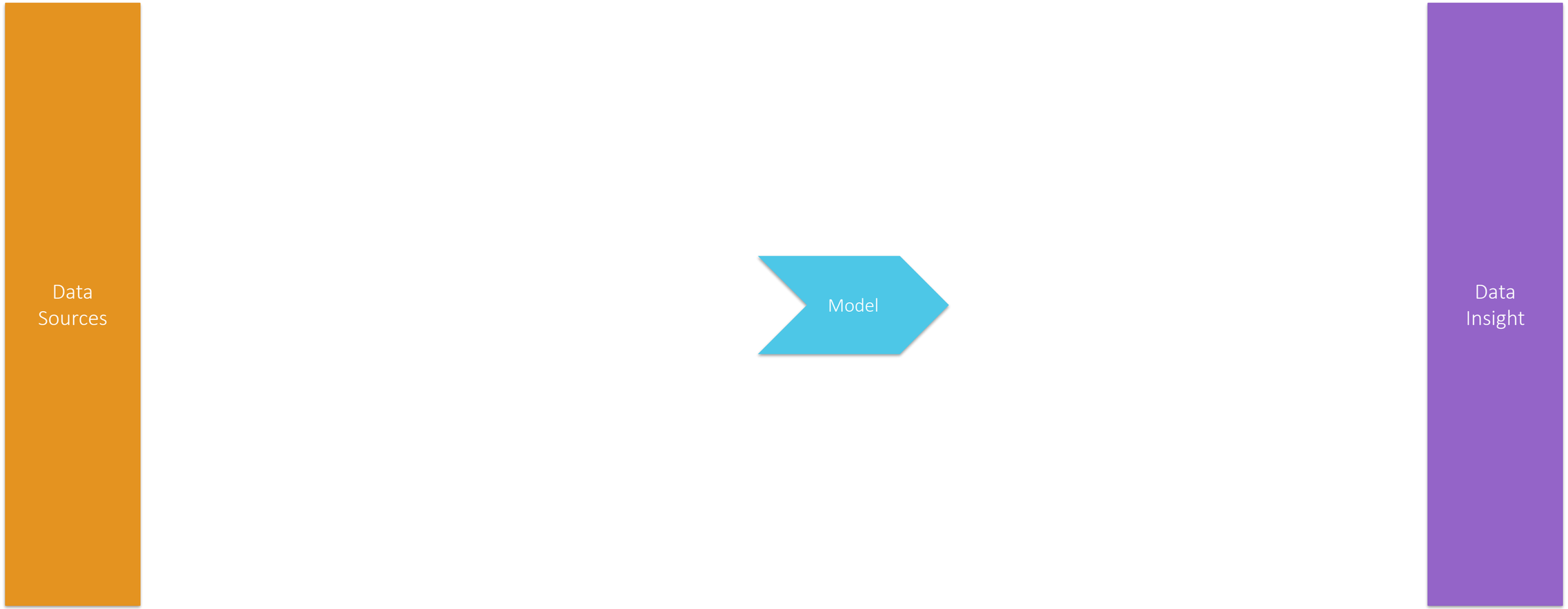


Data Mesh – Why should we build it?

Using a traditional centralised approach, enhanced with cloud scale technologies to create a modern data analytics platform.

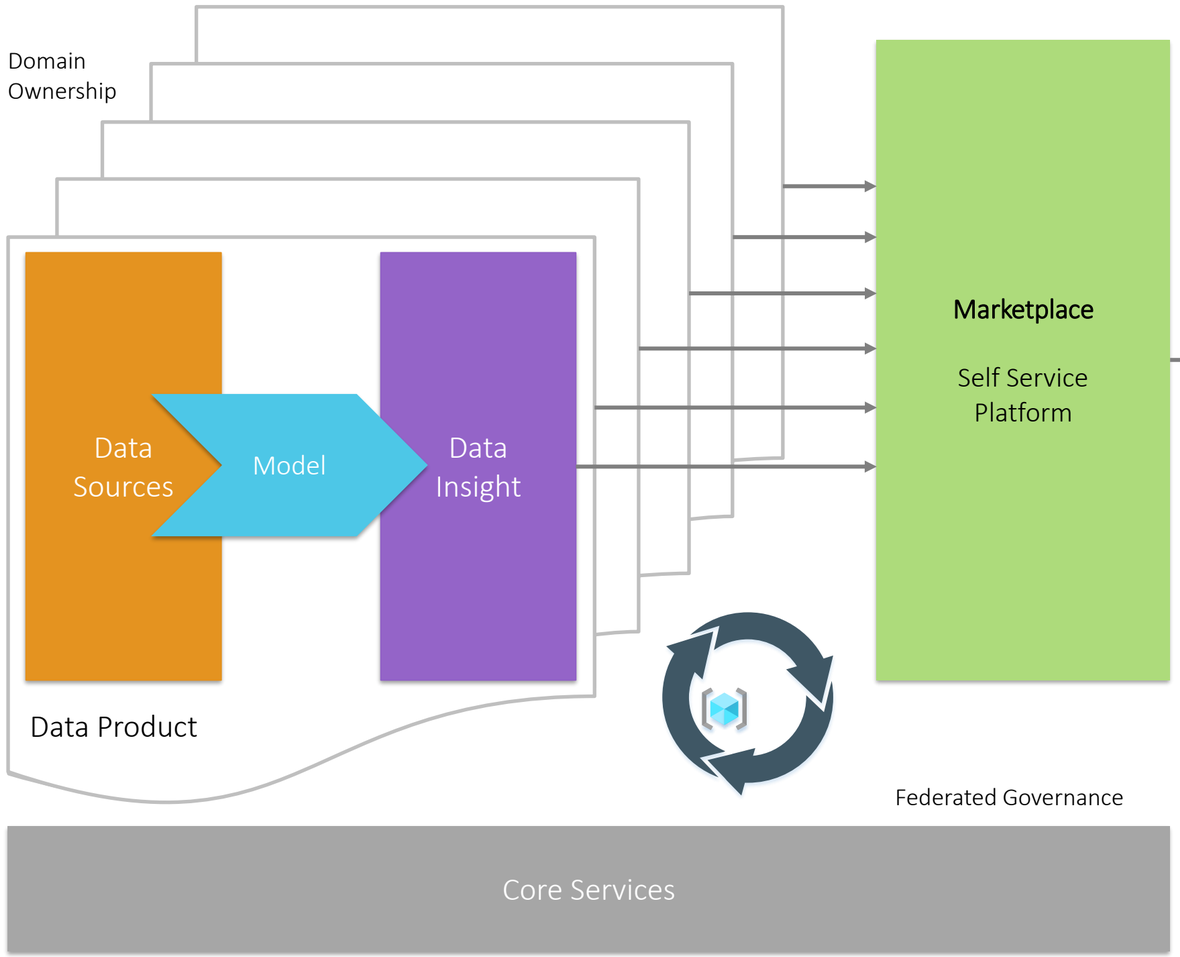


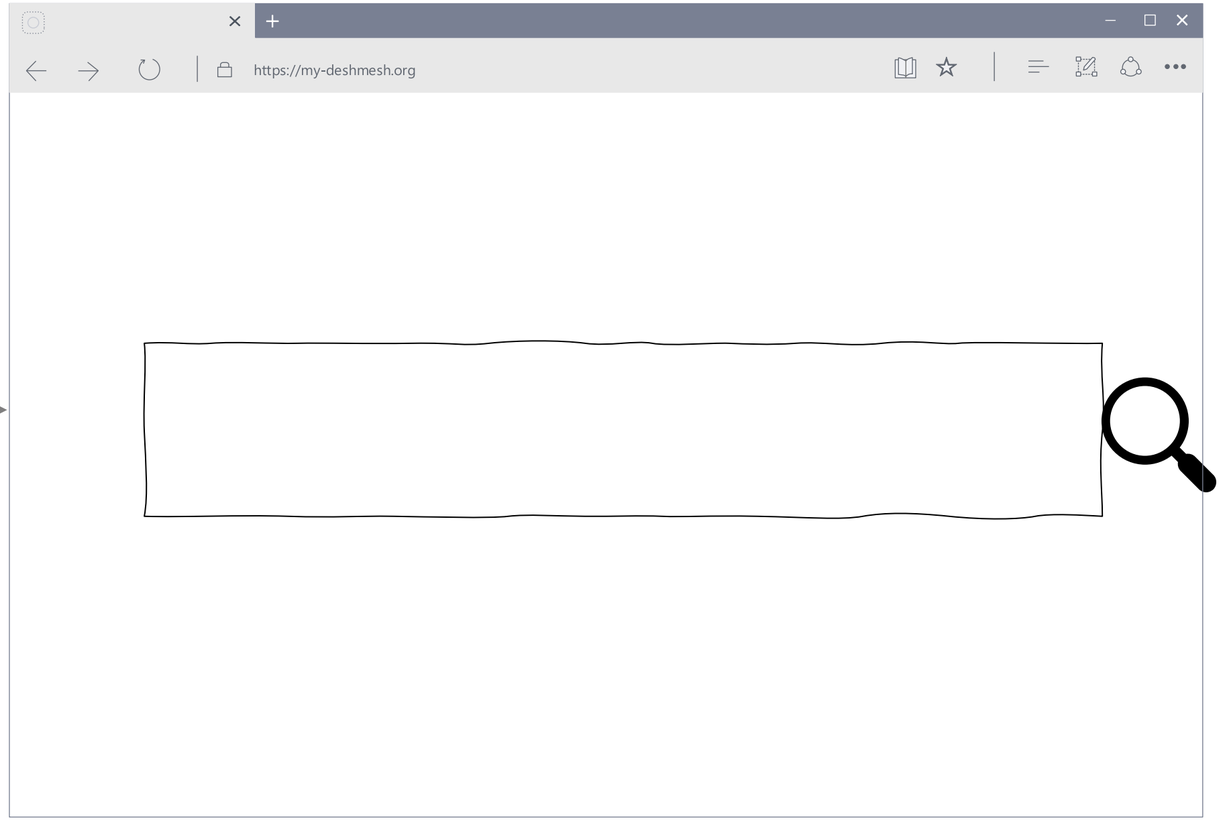
Data Mesh – Why should we build it?



Data Mesh – Why should we build it?

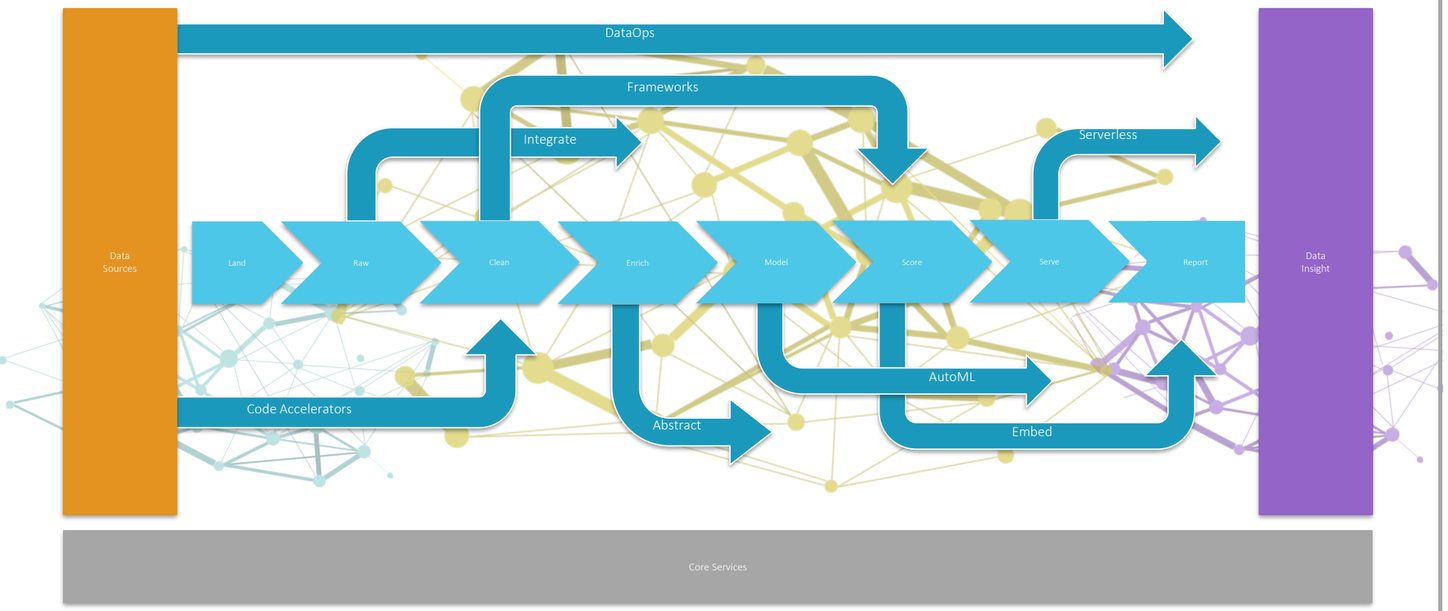
Using a de-centralised approach to cloud scale analytics, empowering users to rapidly gain insights to make strategic business decisions.





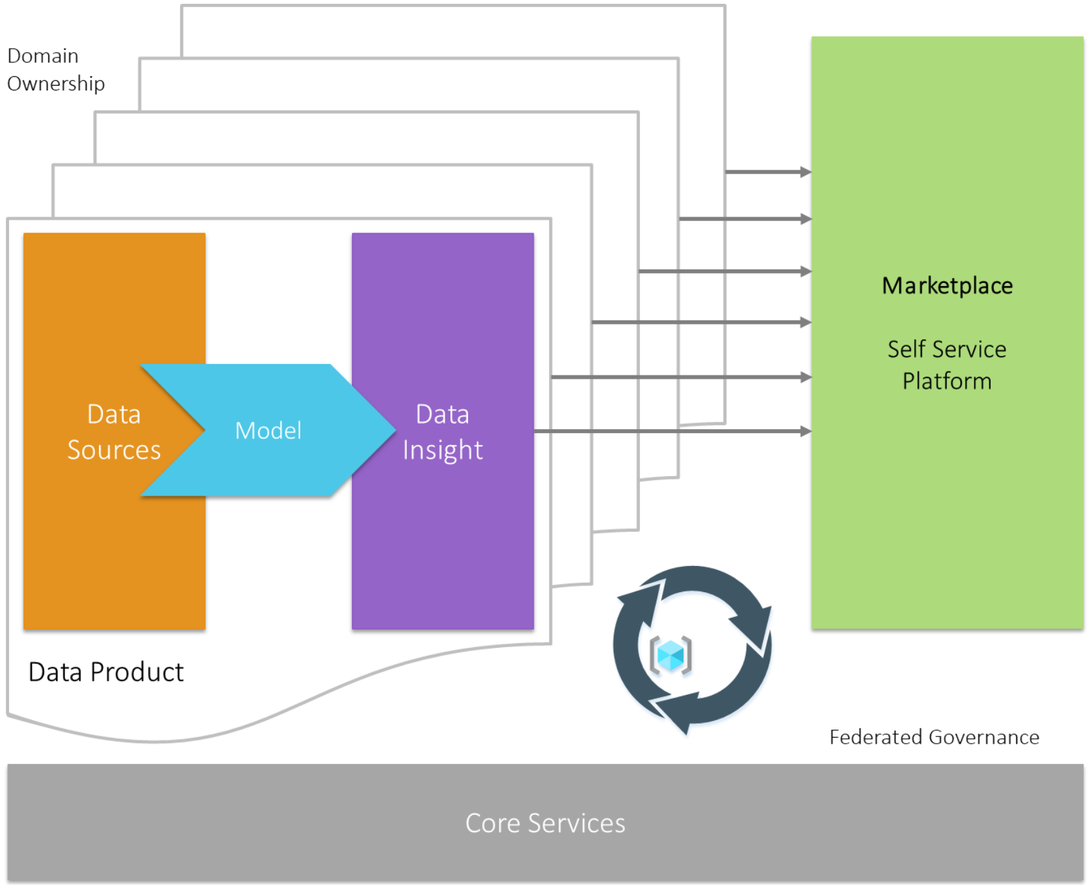
Data Mesh – Why should we build it? A: Time to Insight

Using a traditional centralised approach, enhanced with cloud scale technologies to create a modern data analytics platform.



… Weeks/Months

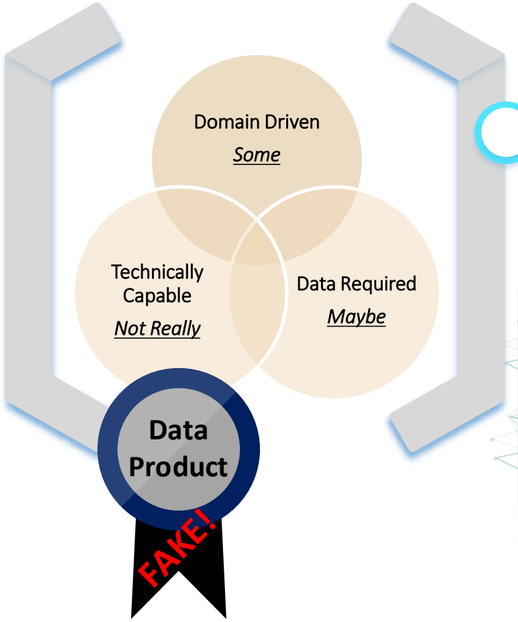
Using a de-centralised approach to cloud scale analytics, empowering users to rapidly gain insights to make strategic business decisions.

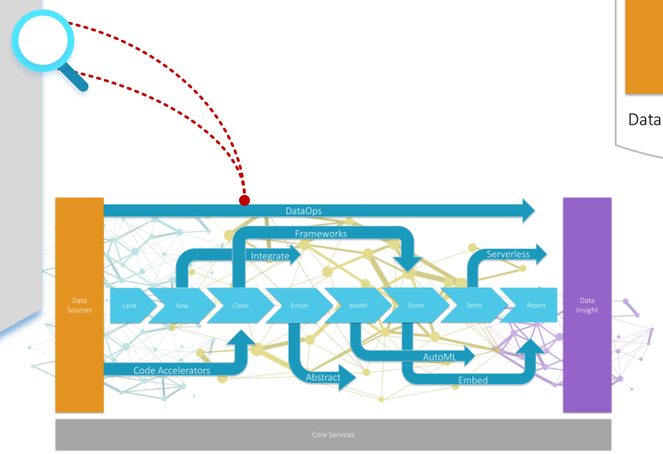


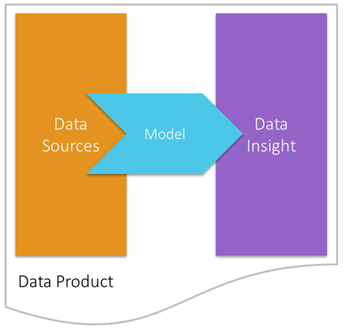
… Hours/Days

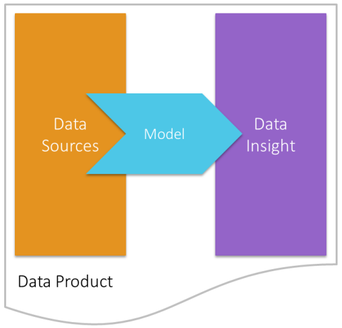
Cloud Formations

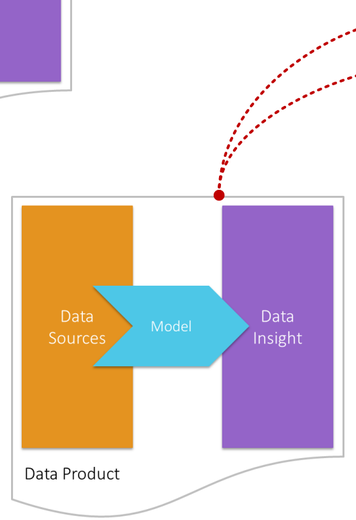
Data Mesh – Data Products

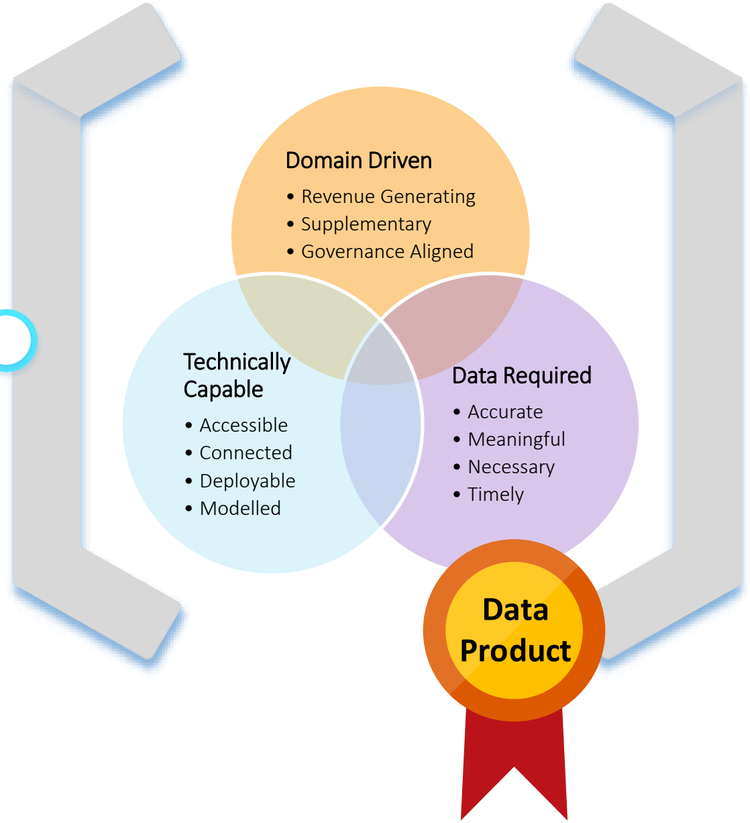




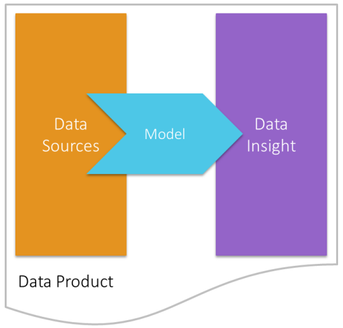


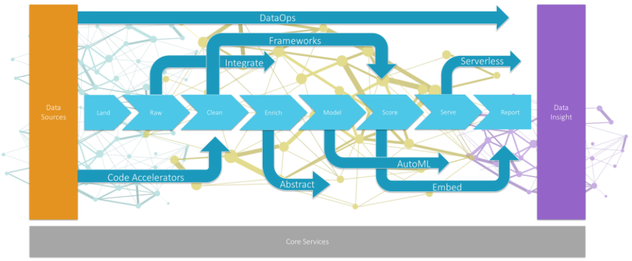


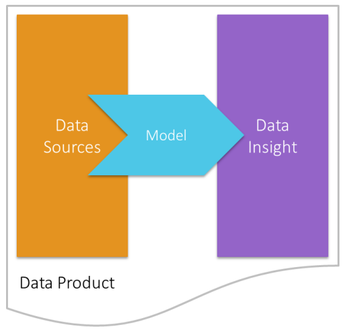


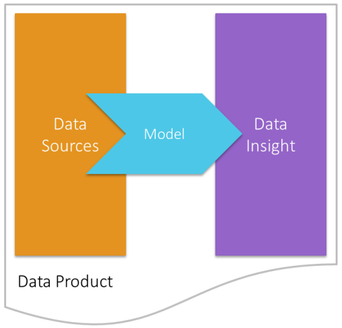


Data Mesh – Data Products in Azure

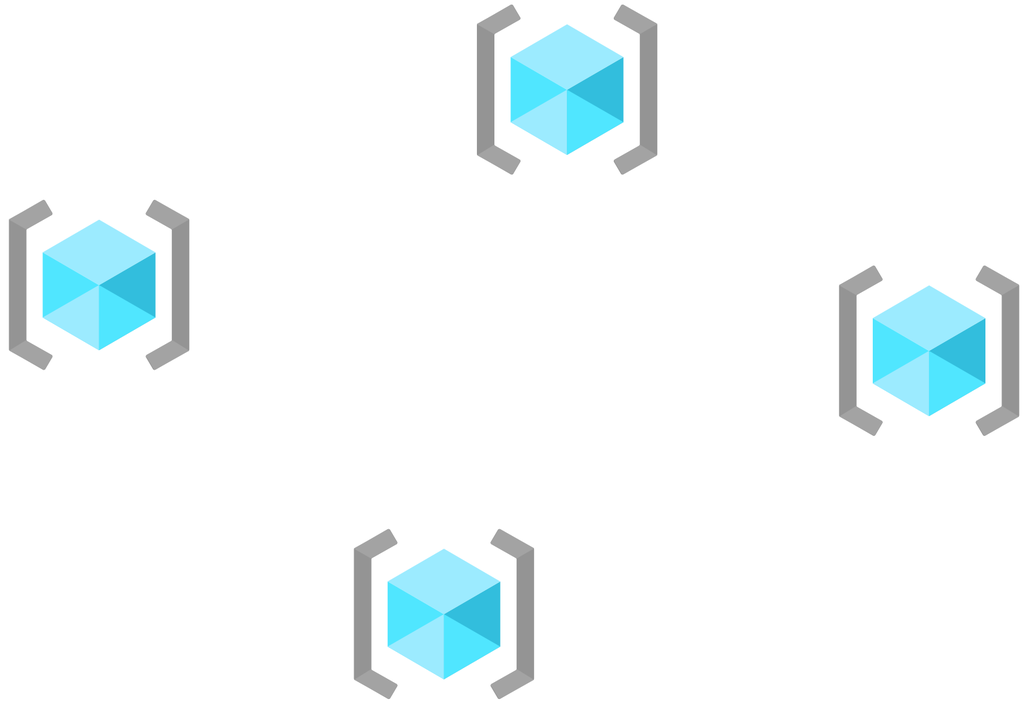




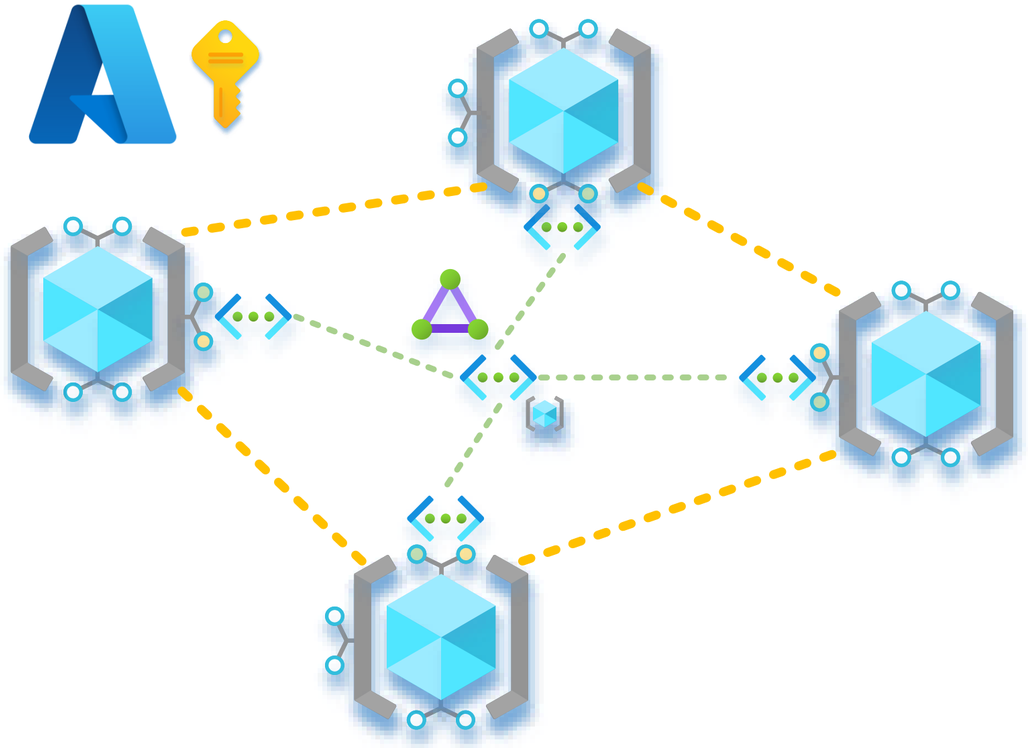




Data Mesh – Data Products in Azure



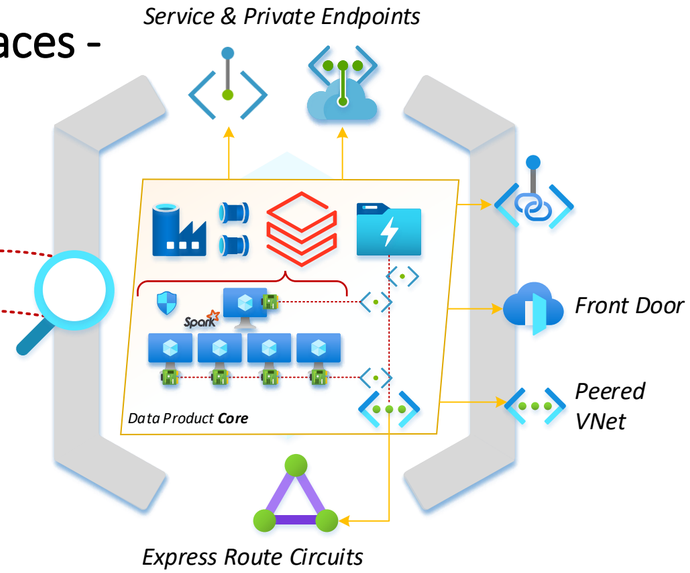
Data Mesh – Data Products in Azure with Interfaces

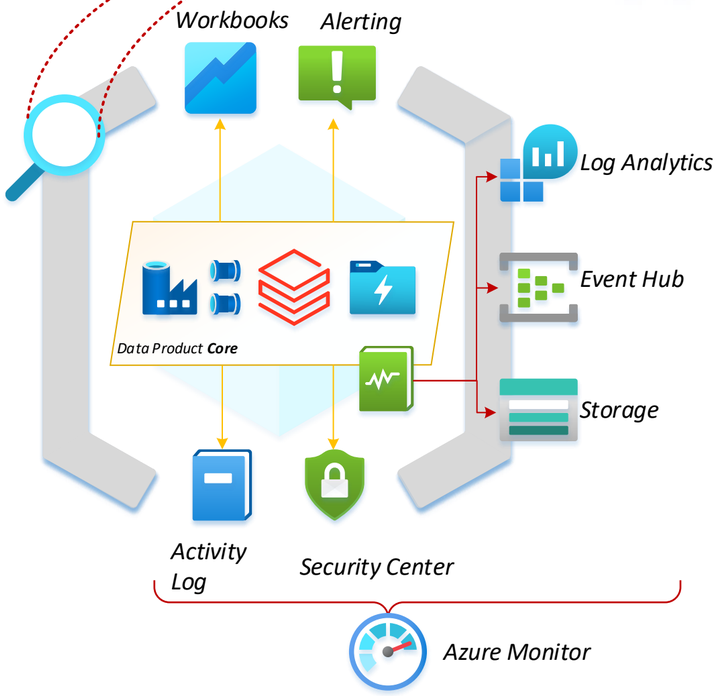


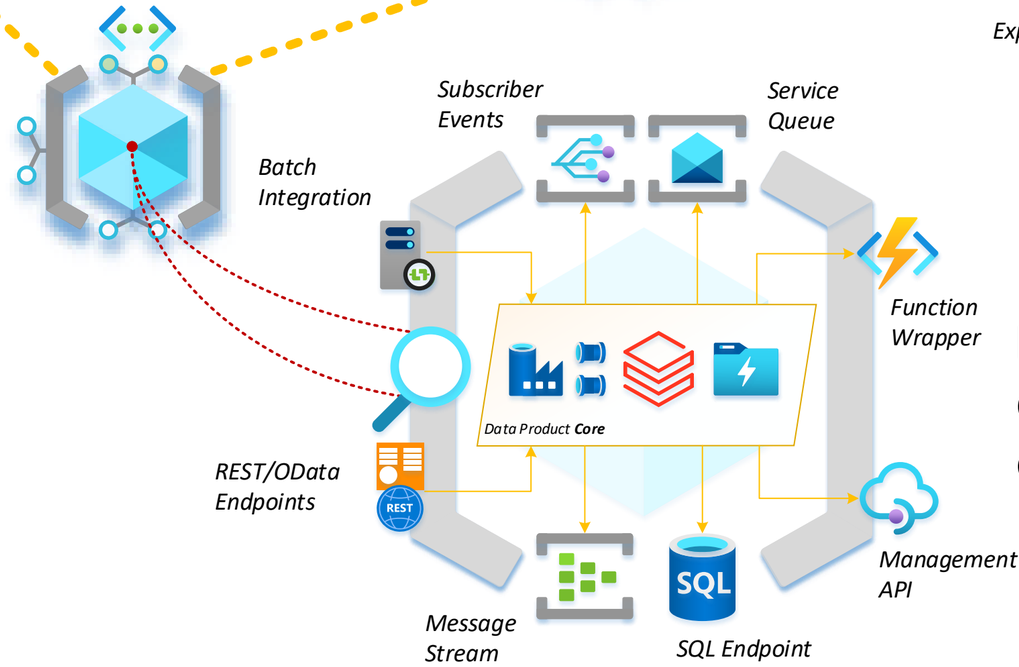
Data Mesh – Data Products in Azure with Interfaces

Secondary Interfaces - operational reporting and logging.

Tertiary Interfaces - resource connectivity.

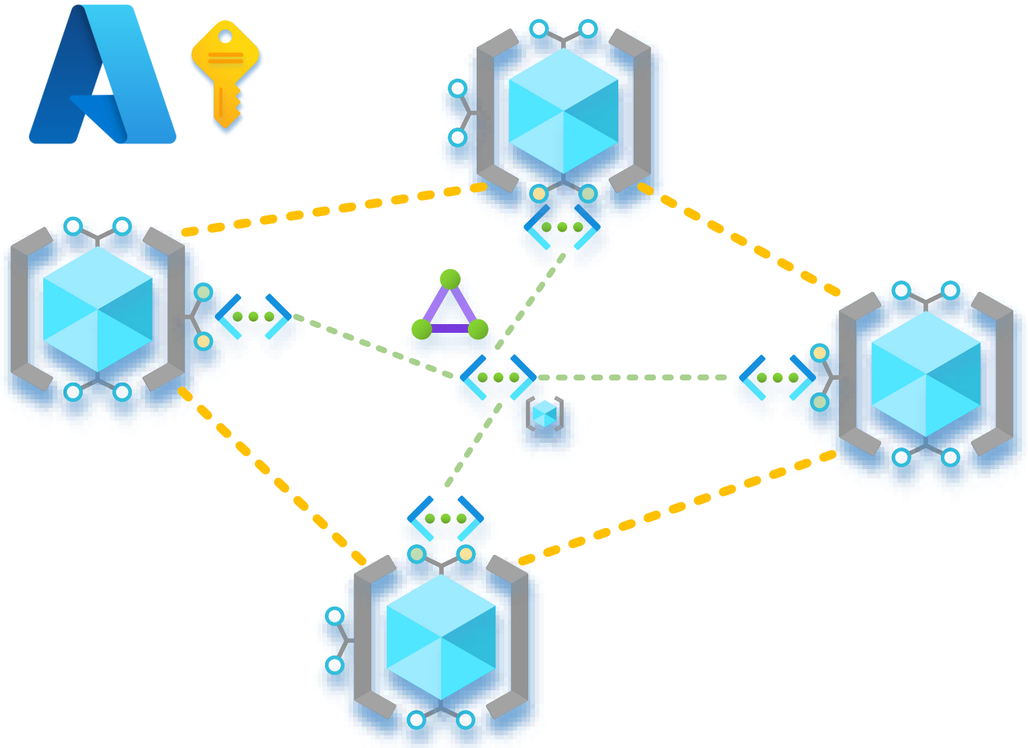




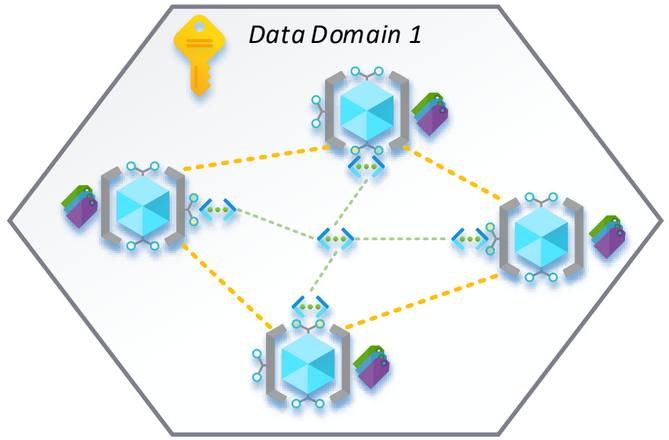


Primary Interfaces – data integration and exchange.

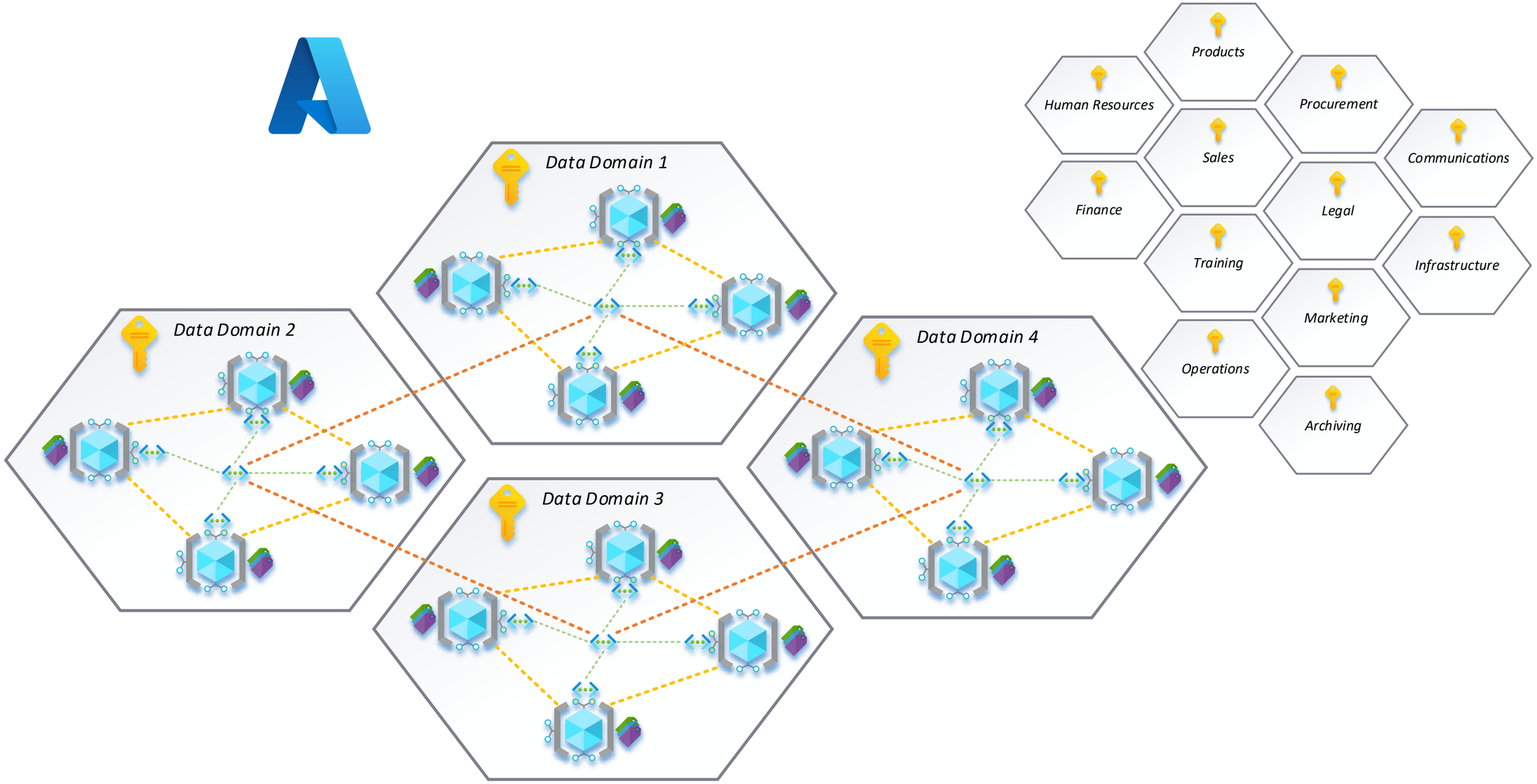
Data Mesh – Data Domains in Azure



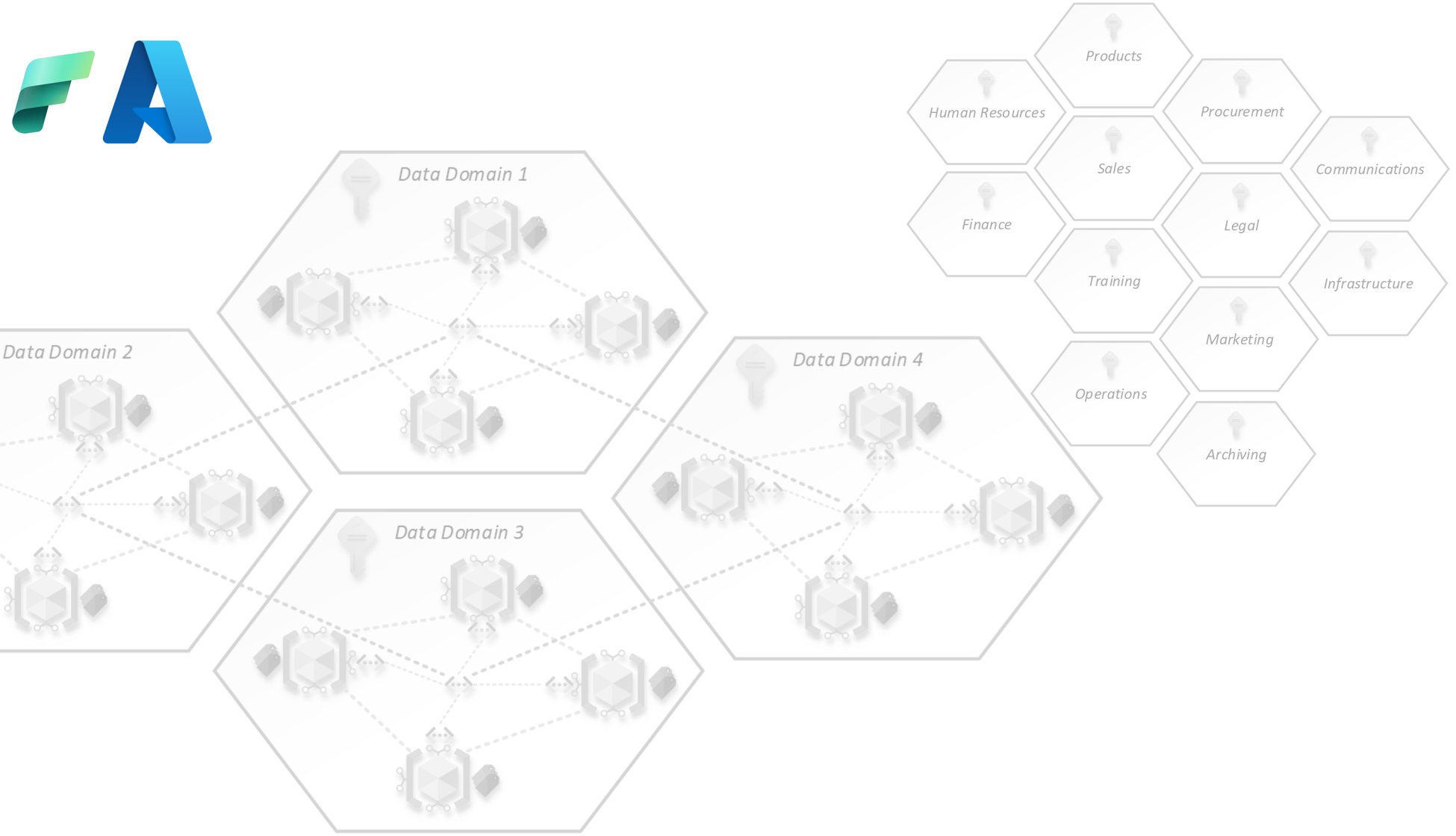
Data Mesh – Data Domains in Azure



Data Mesh – Data Domains in Azure



Data Mesh – Data Domains in Fabric

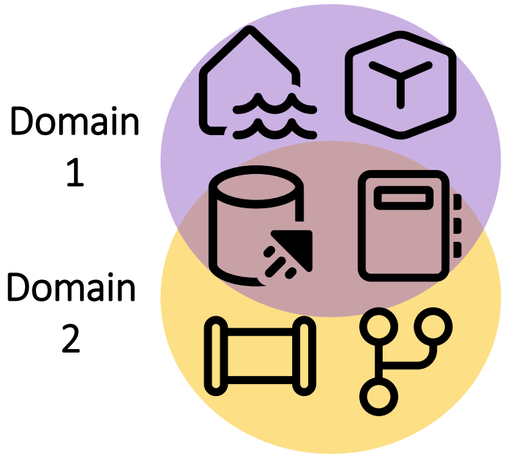


Data Mesh – Data Domains in Fabric

- Data Integration
- Data Engineering
- Data Warehouse
- Data Science
- Real-time Analytics
- Business Intelligence
- Data Activator
- One Lake

Workspaces

Items



…. To meet this challenge, organizations are shifting from traditional IT centric data architectures, where the data is governed and managed centrally, to more federated models organized according to business needs. This federated data architecture is called data mesh. A data mesh is a decentralized data architecture that organizes data by specific business domains, such as marketing, sales, human resources, etc.

What are Fabric domains?

In Fabric, a domain is a way of logically grouping together all the data in an organization that is relevant to a particular area or field. ….

Reference: https://learn.microsoft.com/en-us/fabric/governance/domains

OneLake

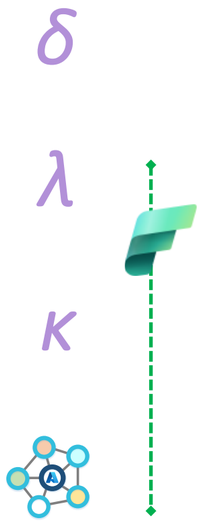
{Domain}

Workspaces

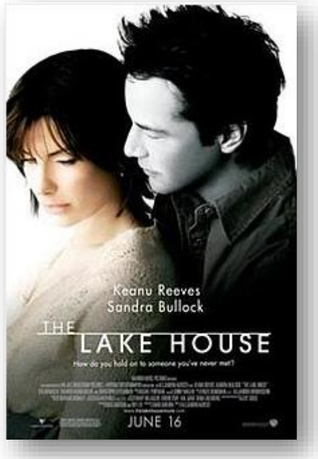
Items

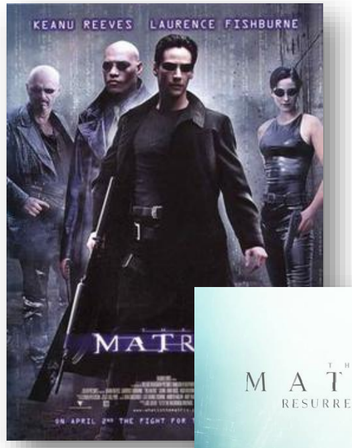
Cloud Formations - Knowledge Transfer & Training

Architecture Agenda:



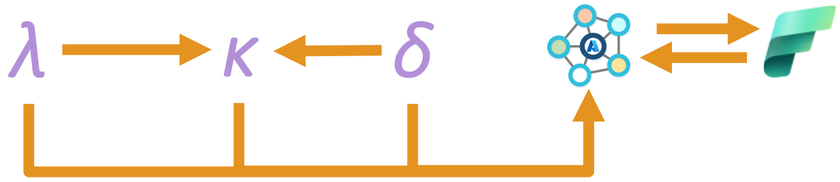
Final thoughts from me...



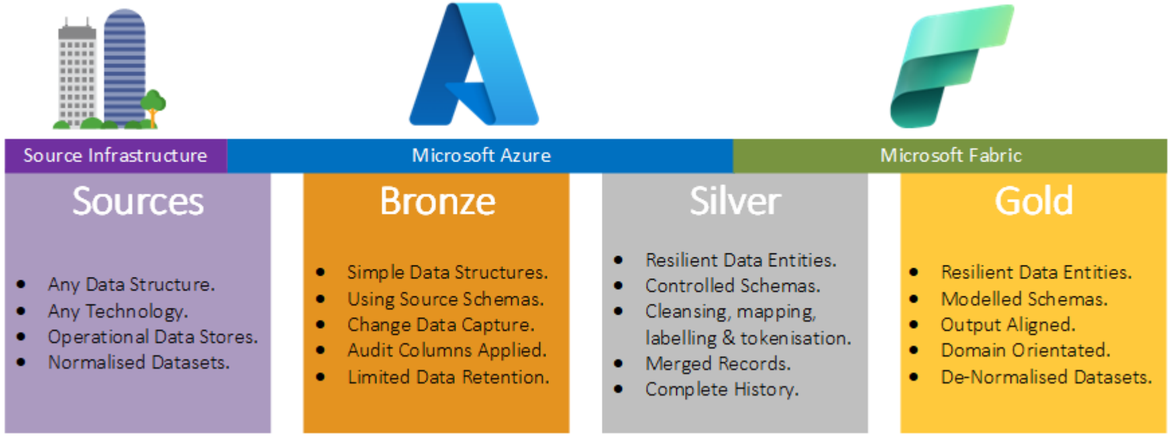


An Evolution of Data Platform Architectures

Lambda, Kappa, Delta, Mesh & Fabric



Q: What about a medallion architecture?

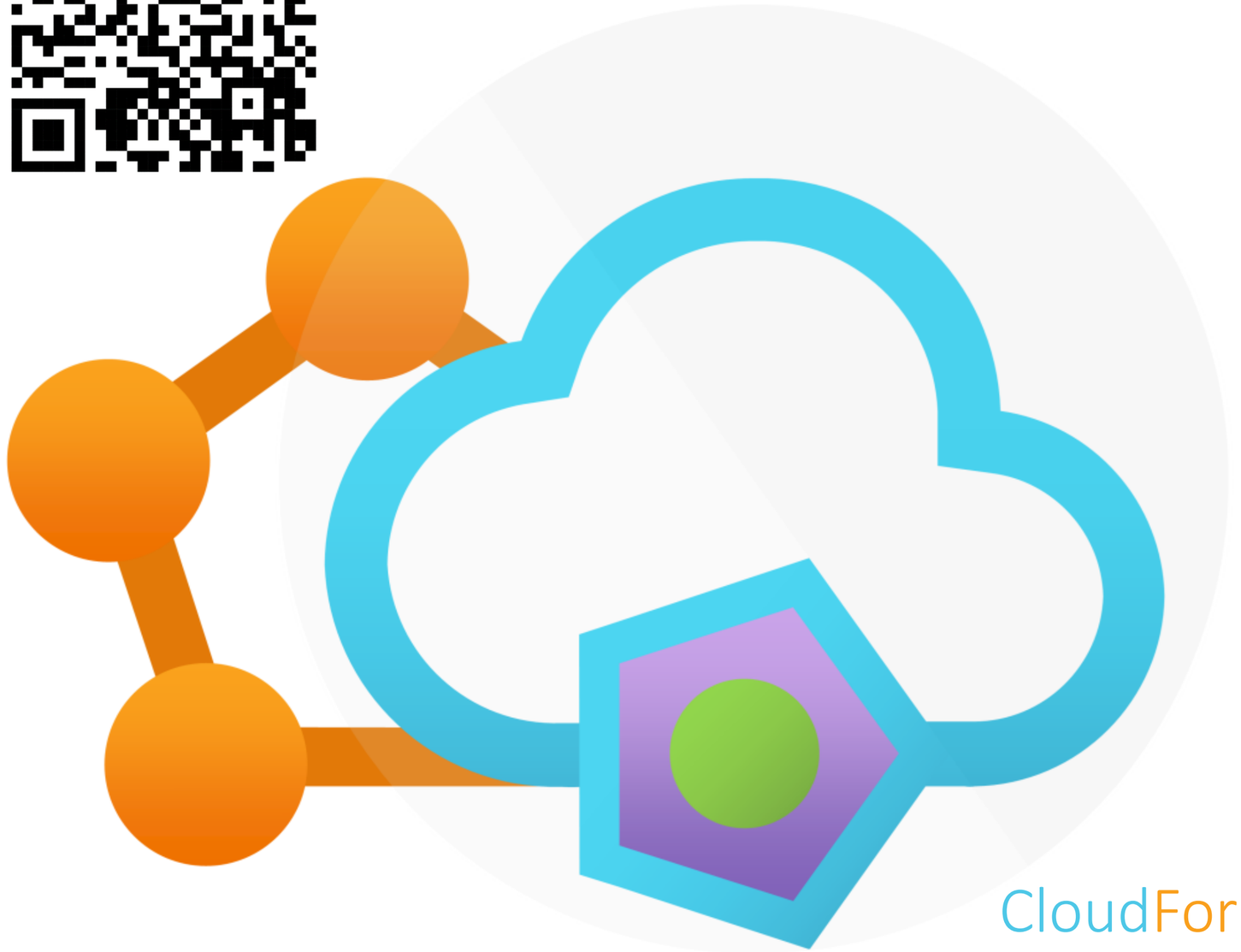


https://mrpaulandrew.com/2023/11/14/considering-a-medallion-architecture-vs-microsoft-fabric/

Q: Should we be considering a solution/technology stack that offers all these capabilities?

A: Yes! ✓

Thank You



mrpaulandrew.com
paul@mrpaulandrew.com
In/mrpaulandrew
@mrpaulandrew

https://cloudformations.org
contactus@cloudformations.org
In/CloudFormations
@CloudFormsLtd
CloudFormationsLtd

CloudFormations.org/Community-Content

PDF