

An Evolution of Data Platform Architectures

Lambda, Kappa, Delta, Mesh & Fabric



Paul Andrew

Technical Strategist | Director



Cloud Formations

Paul Andrew



Co-Founder & Director
Chief Technology Officer



/mrpaulandrew



@mrpaulandrew

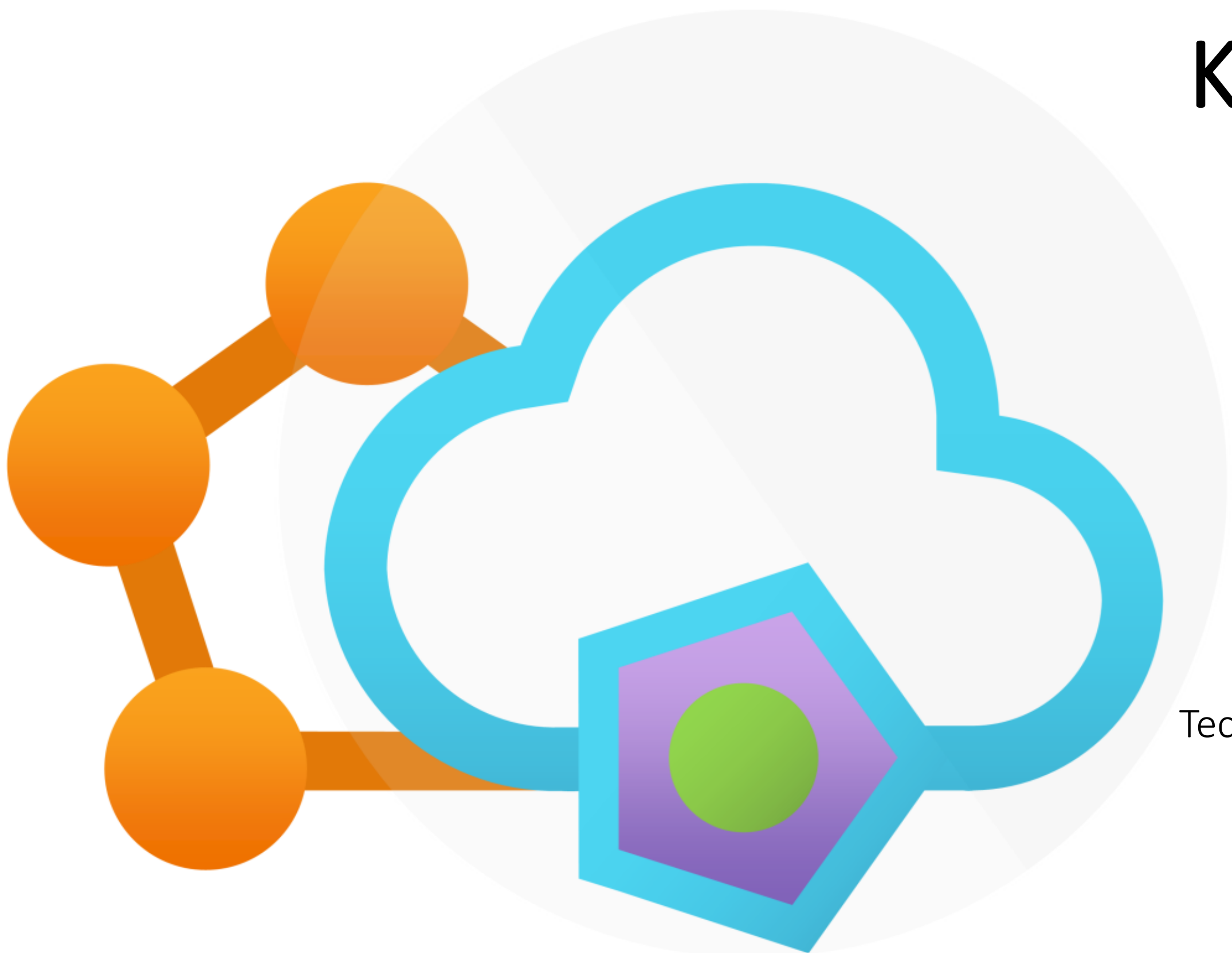


In/mrpaulandrew

- Mentor | Author
- Speaker | Podcast Host
- Event Organiser

SQL Server 2000





Keanu Reeves is really a data engineer!

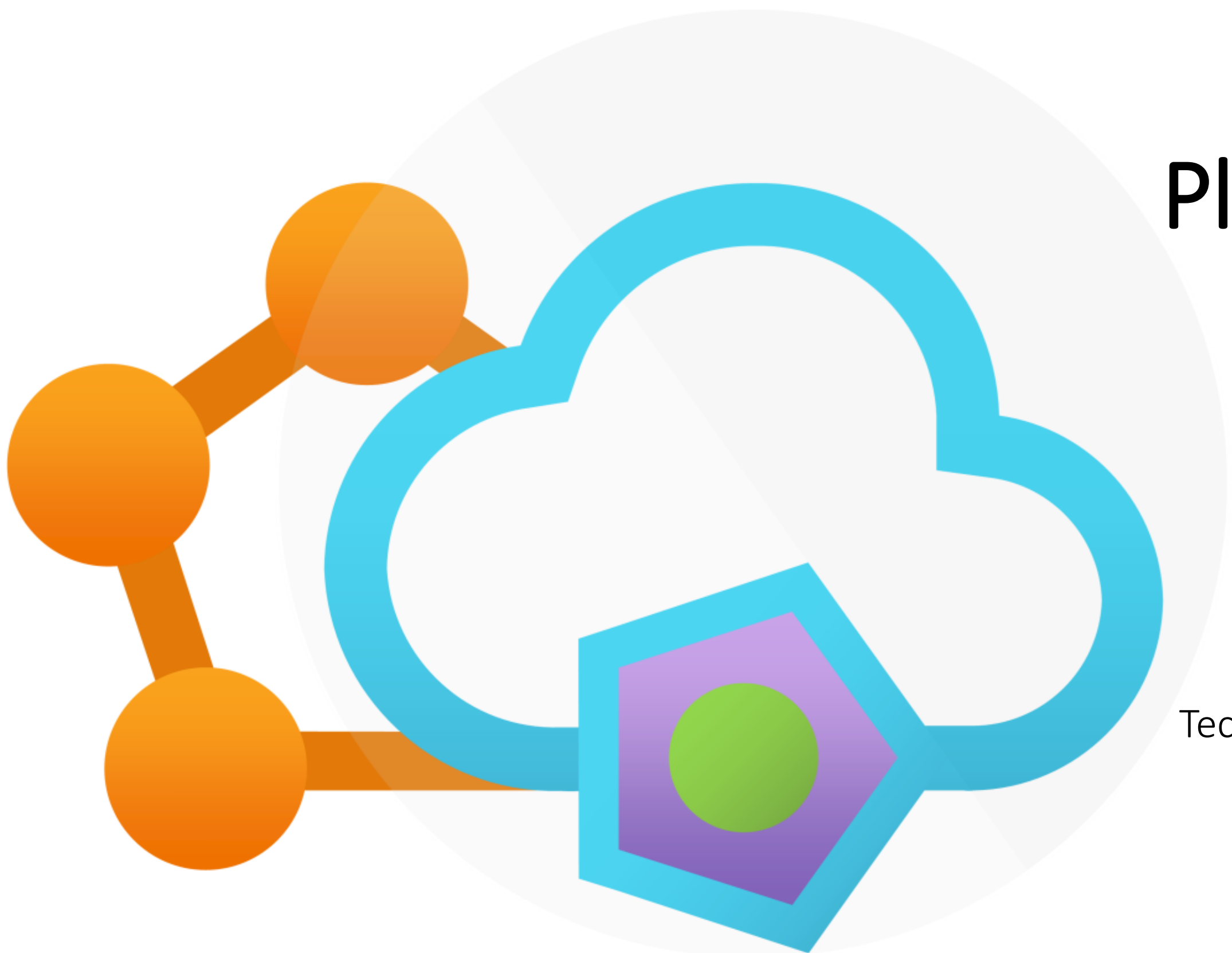


Paul Andrew

Technical Strategist | Director



Cloud Formations



An Evolution of Data Platform Architectures

Lambda, Kappa, Delta, Mesh & Fabric



Paul Andrew

Technical Strategist | Director



Cloud Formations



Architecture Agenda:

Lambda, Kappa, Delta, Mesh & Fabric

Delta Lake

λ

κ

δ



Lambda

Microsoft Fabric

Kappa

Data Mesh



What is the goal of our data solutions?



Data
Sources

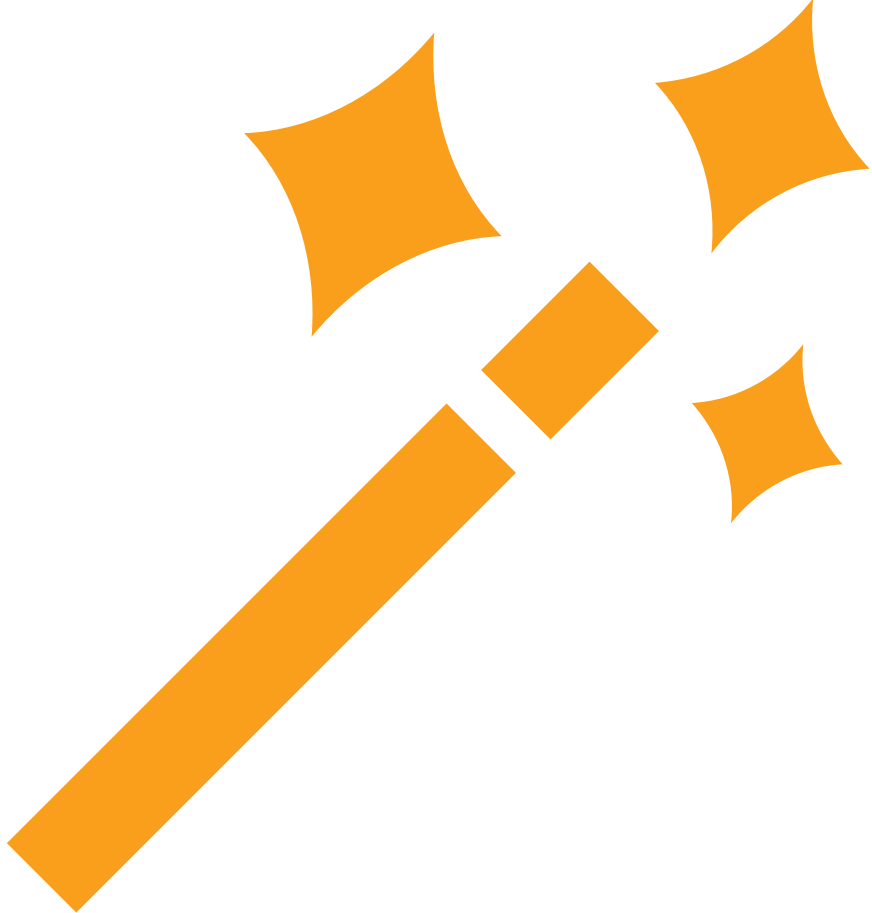
Data
Insight

Data = Information = Knowledge = Power

How do we deliver our data insights?



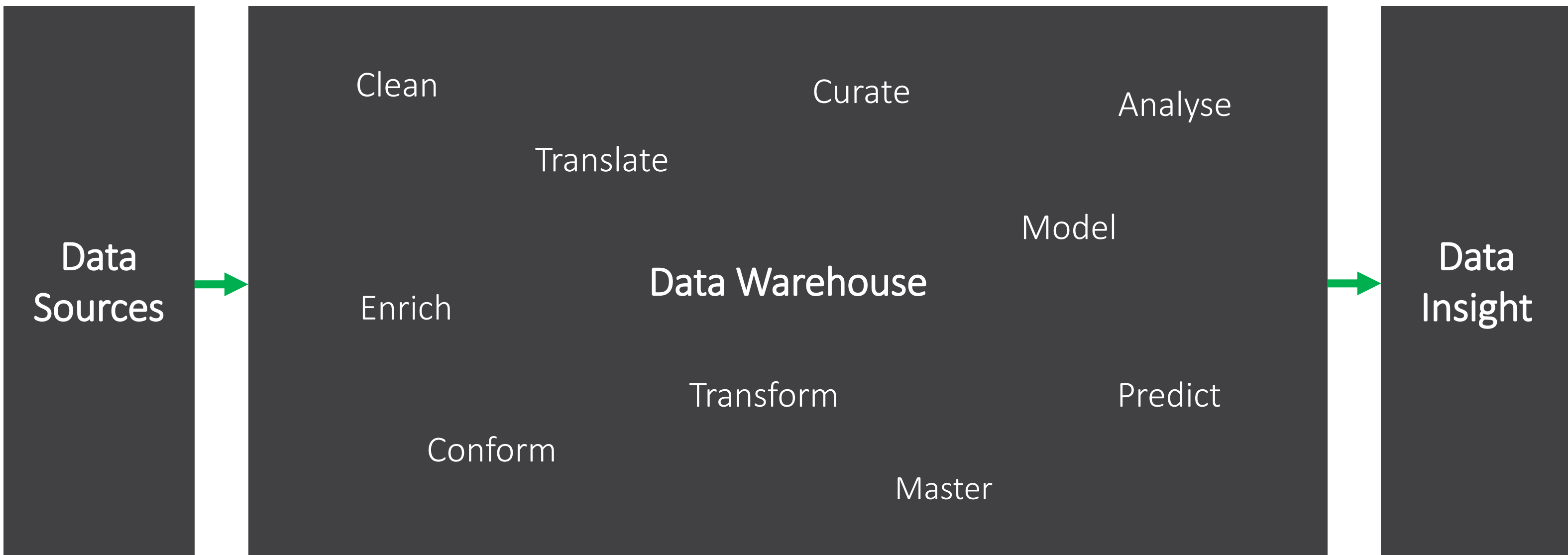
Data Sources



Data Insight

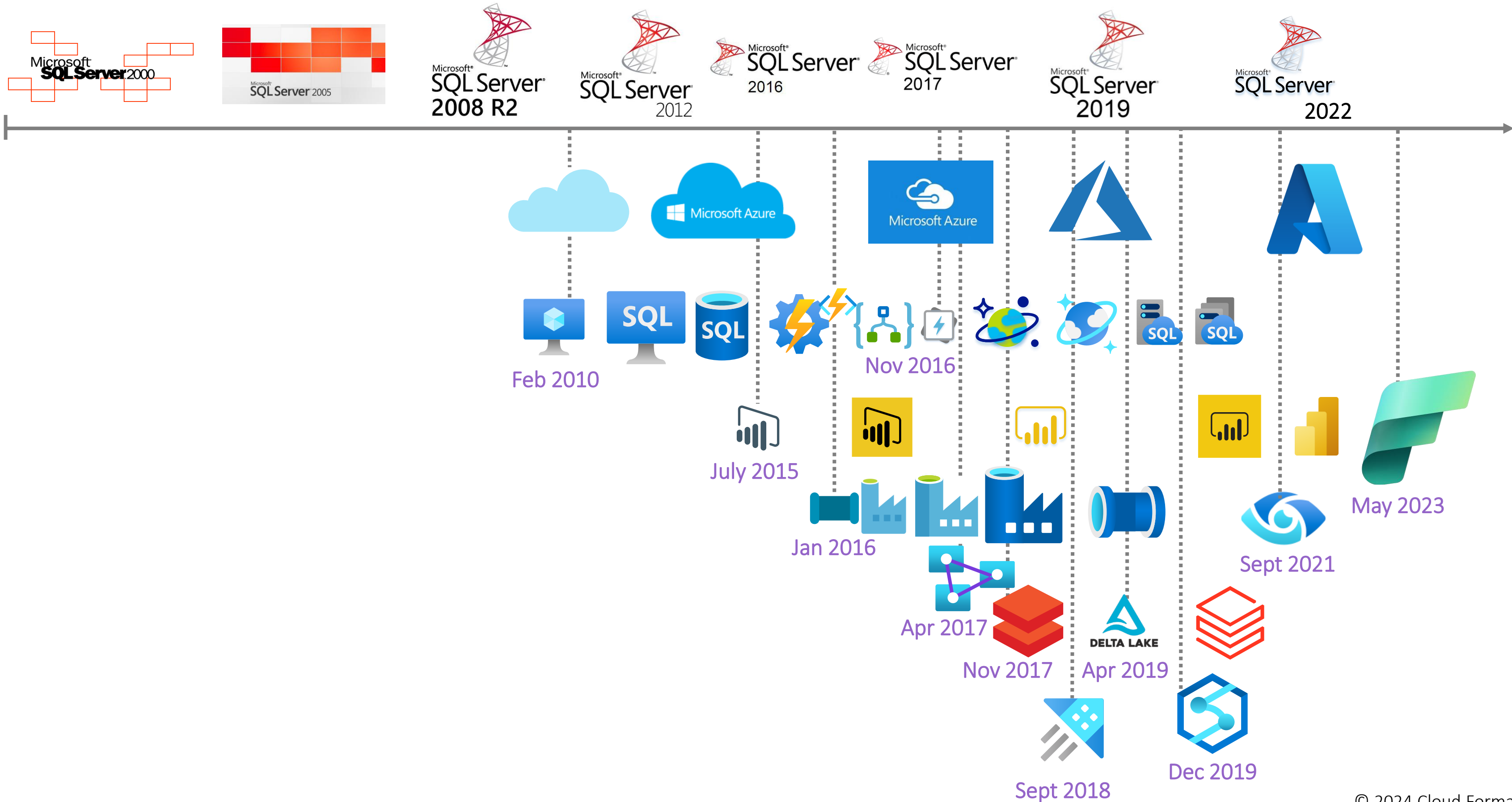
Data = Information = Knowledge = Power

How do we deliver our data insights?



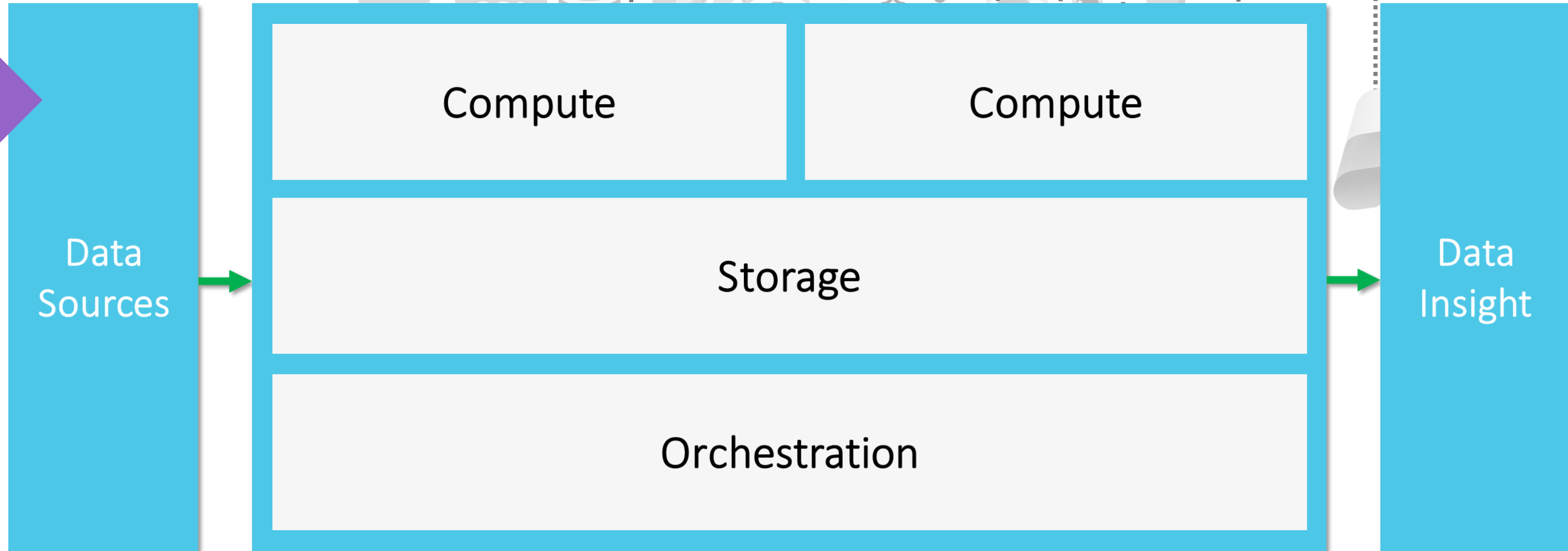
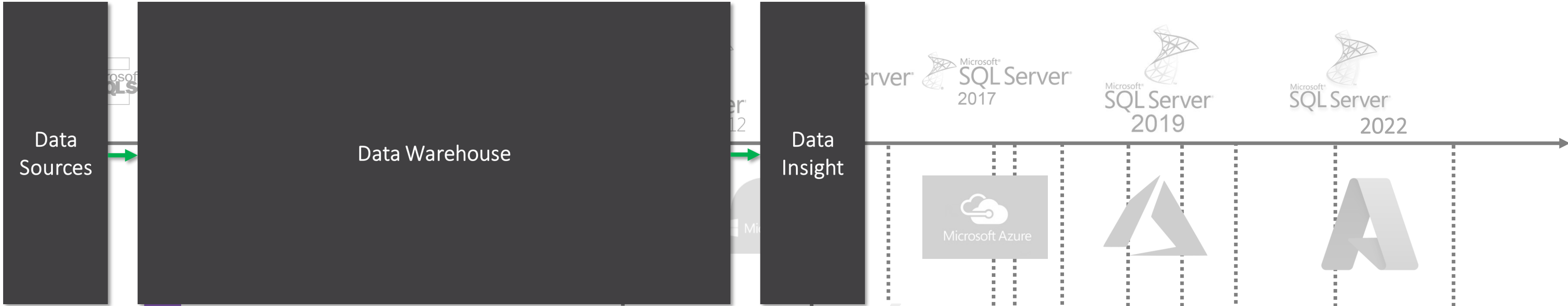
Data = Information = Knowledge = Power

A Timeline of Microsoft Data Technology



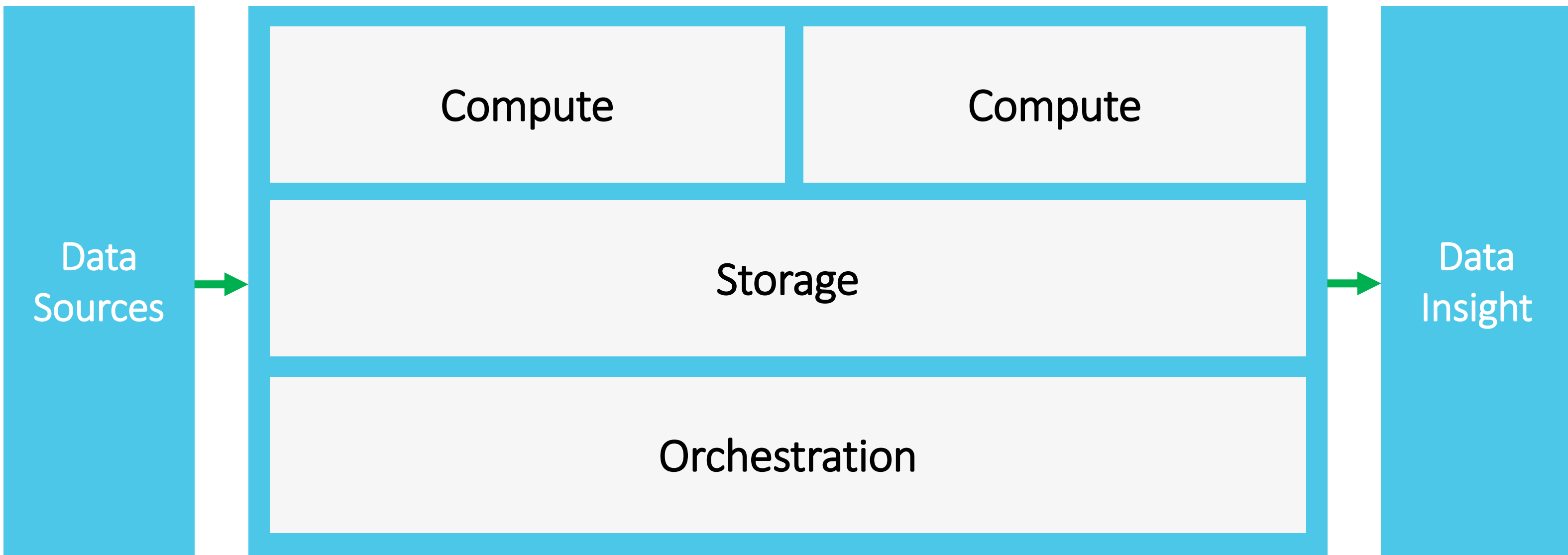
Cloud Formations - Knowledge Transfer & Training

A Timeline of Microsoft Data Technology

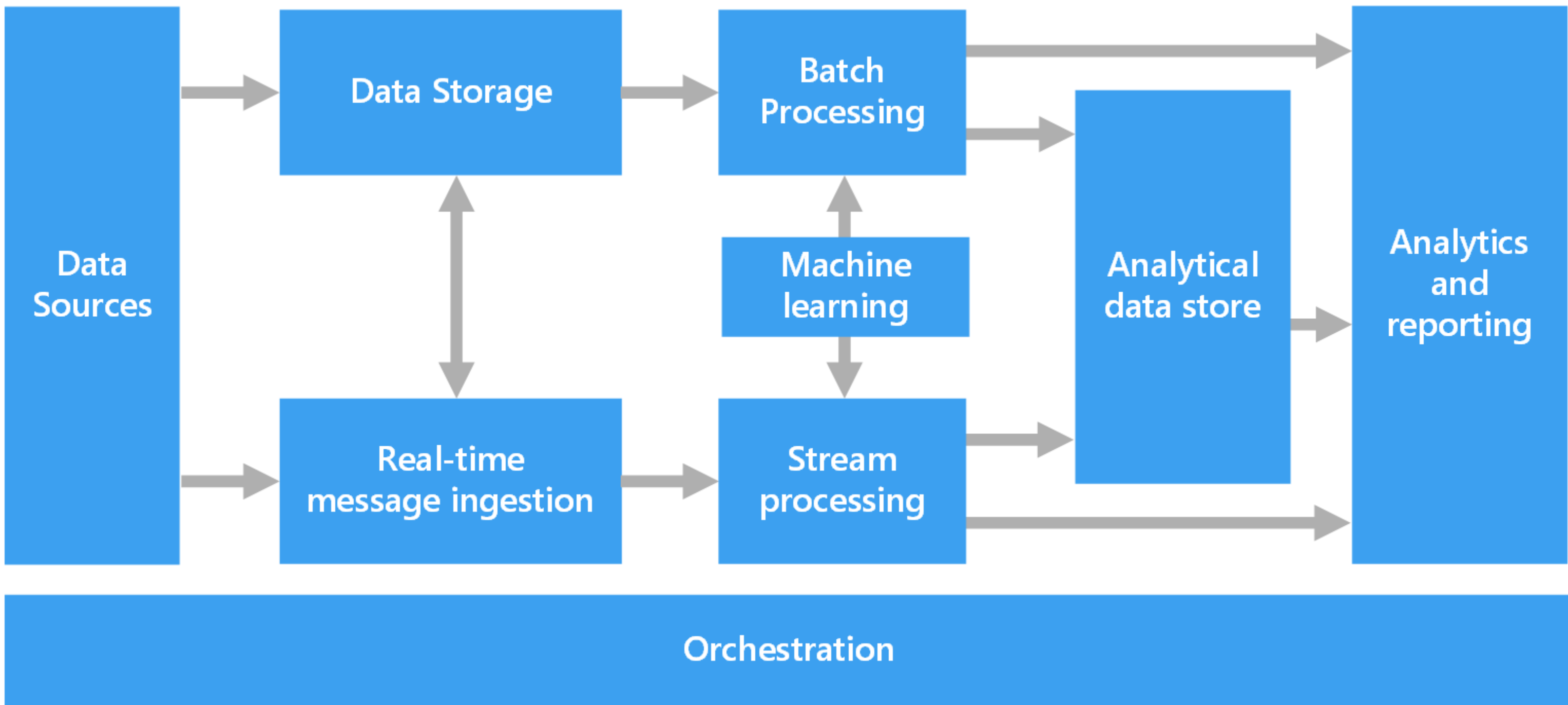


Cloud Formations - Knowledge Transfer & Training

My First Reference Architecture



Microsoft's Components of a Big Data Architecture





Architecture Agenda:

δ

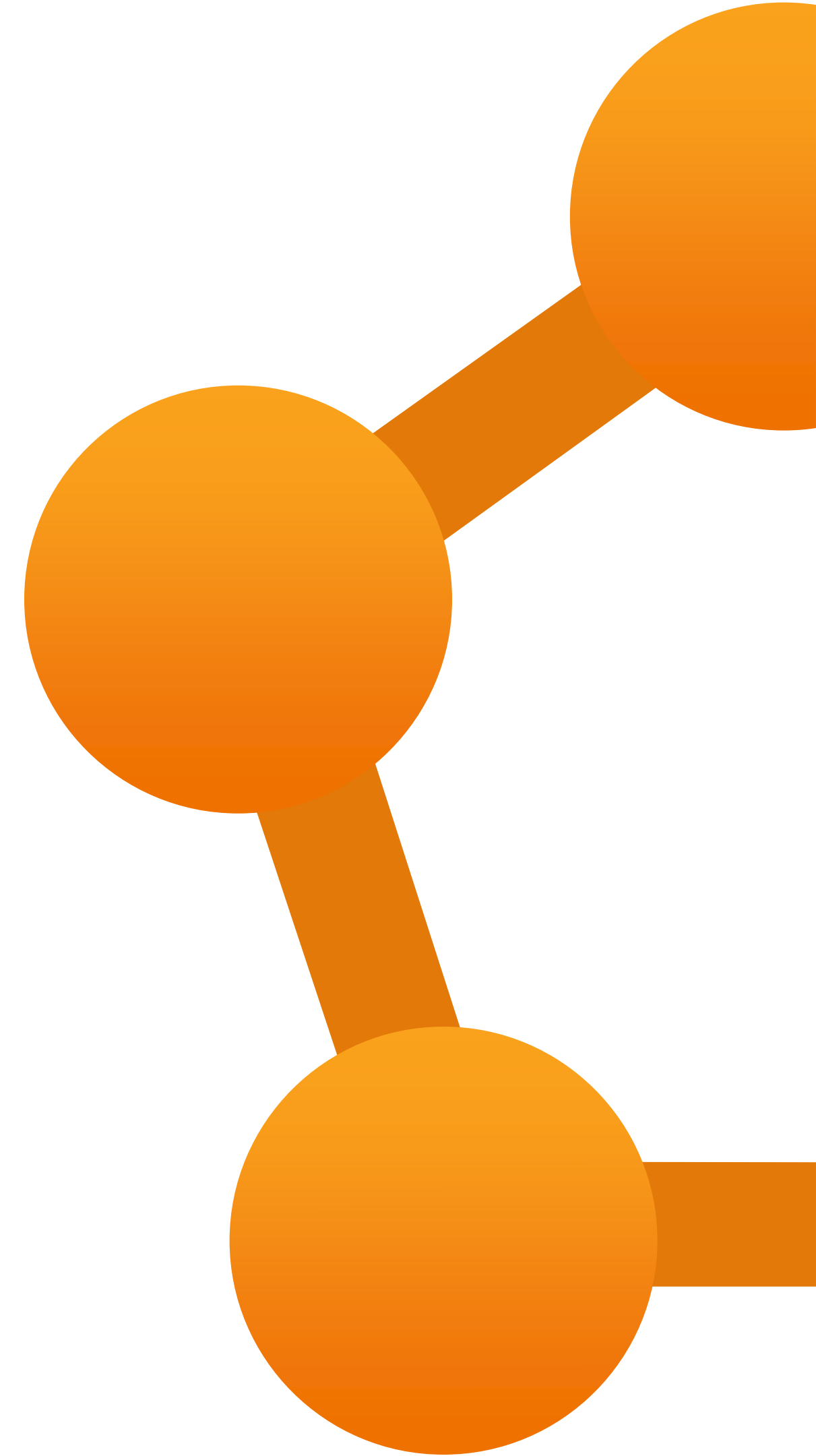
λ

K



Delta δ
Lake

Cloud Formations





DataBase Management System



Atomicity
Consistency
Isolation
Durability



DataBase
Management
System

Atomicity
Consistency
Isolation
Durability

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

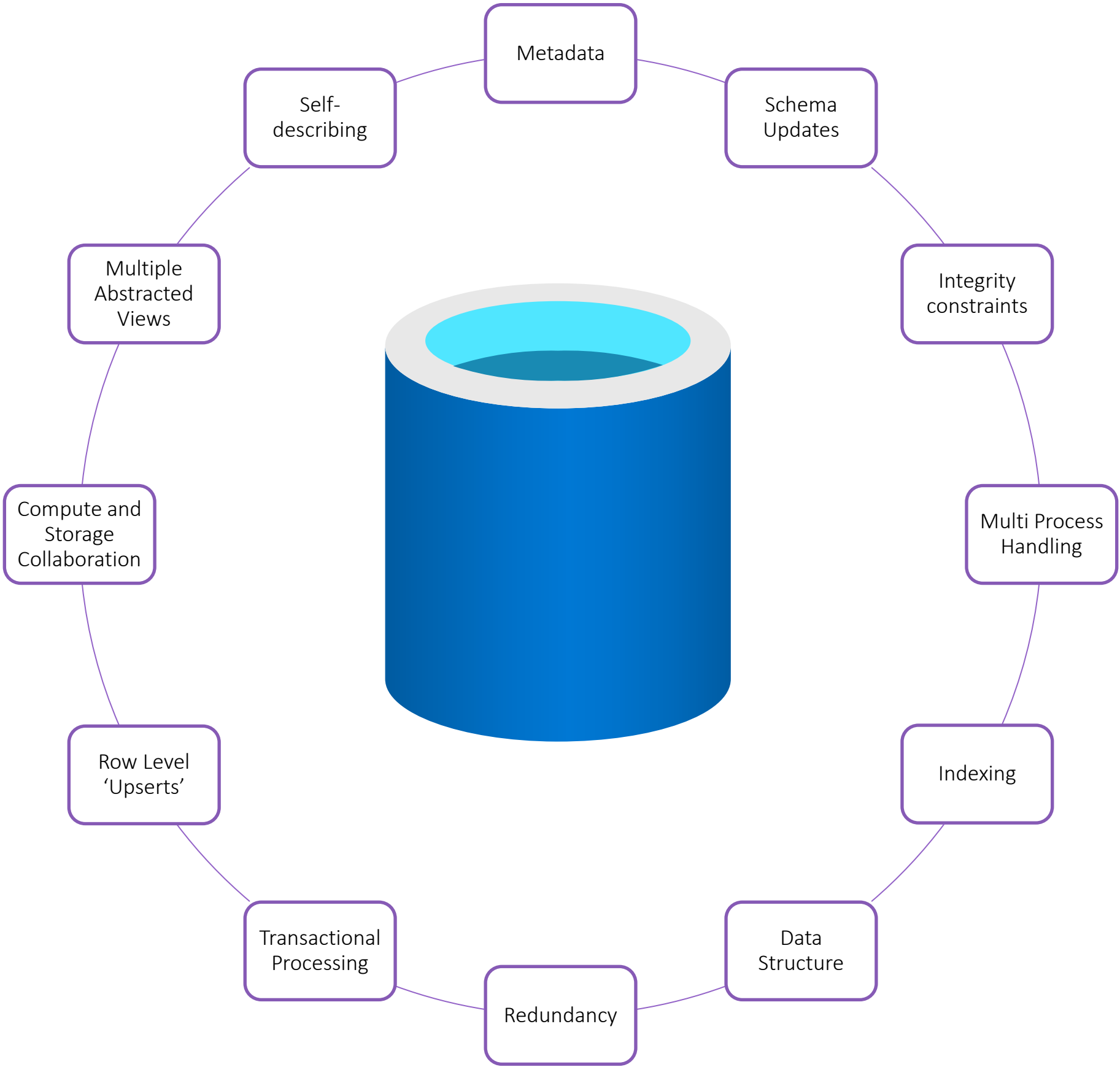
The screenshot shows the Wikipedia article for ACID. The page title is "ACID" and it includes a navigation bar with "Article" and "Talk" tabs. The main content area contains a definition of ACID in computer science, a note about the need for citations, and a historical note about the acronym's origin. The left sidebar contains the Wikipedia logo and various navigation links.

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



DataBase Management System

Atomicity
Consistency
Isolation
Durability



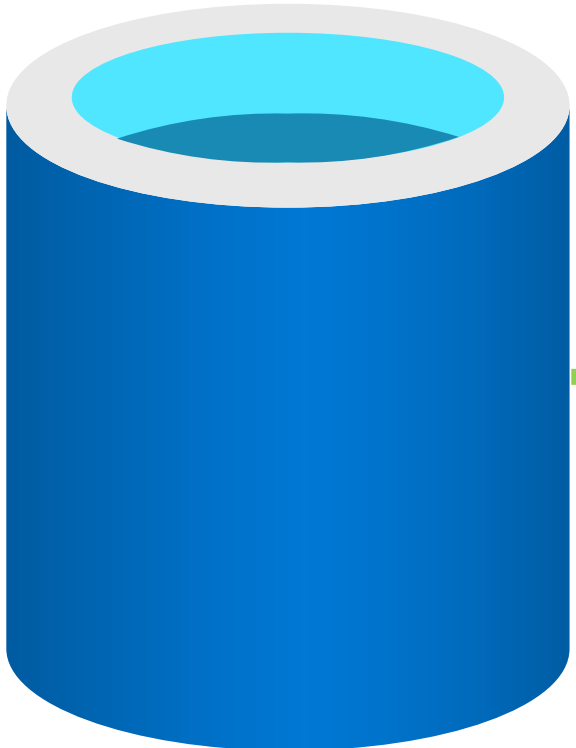
Databases



Creating a Data Warehouse



Online
Line
Transactional
Processing



Application
Data



Extract
Transform
Load



Data
Warehouse



Offline
Analytical
Transactional
Processing

Creating a Data Warehouse



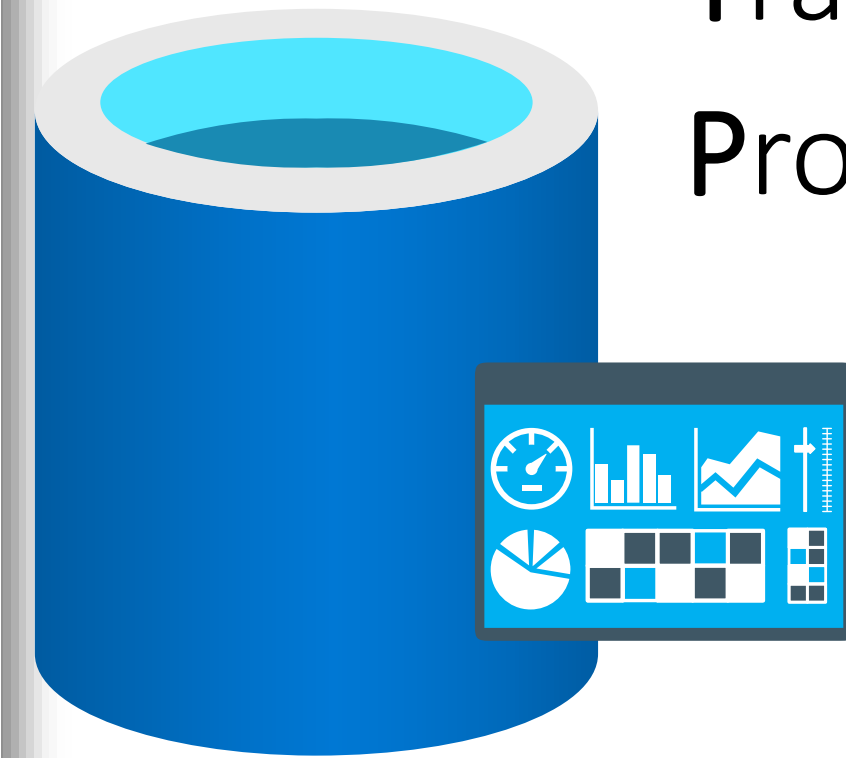
“a system for reporting and data analysis”

Offline
Analytical
Transactional
Processing

The screenshot shows the Wikipedia article for 'Data warehouse'. The article text includes: '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] 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.

The 'Contents' section lists: 1 ETL-based data warehousing, 2 ELT-based data warehousing, 3 Benefits, 4 Generic, 5 Related systems (data mart, OLAPS, OLTP, predictive analytics).

Two diagrams are included: 'Data warehouse overview' showing data sources feeding into a central warehouse, and 'The basic architecture of a data warehouse' showing a flow from Data Sources (Operational systems, Flat files) through a Staging area to a Warehouse (Meta data, Summary data, Raw data), which then feeds into Data Marts (Purchasing, Sales) and finally to Users (Analysis, Reporting).



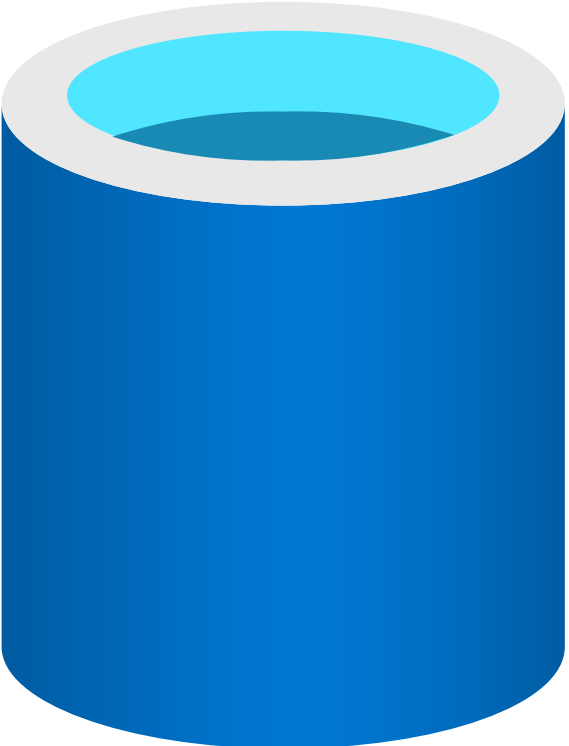
https://en.wikipedia.org/wiki/Data_warehouse





Big Data:

- Volume
- Velocity
- Variety
- Veracity
- Value

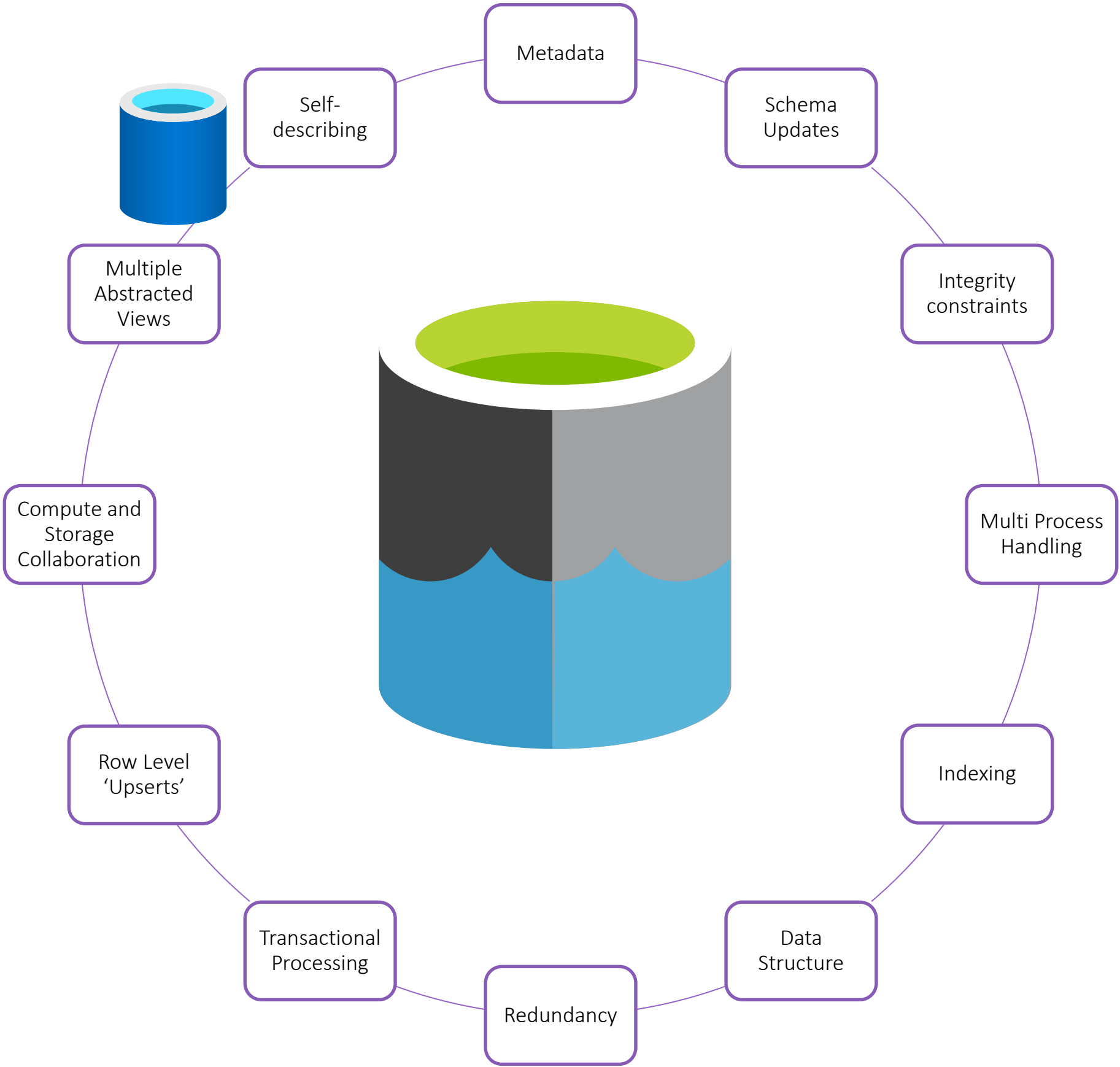


Data Warehouse

Data Lakes

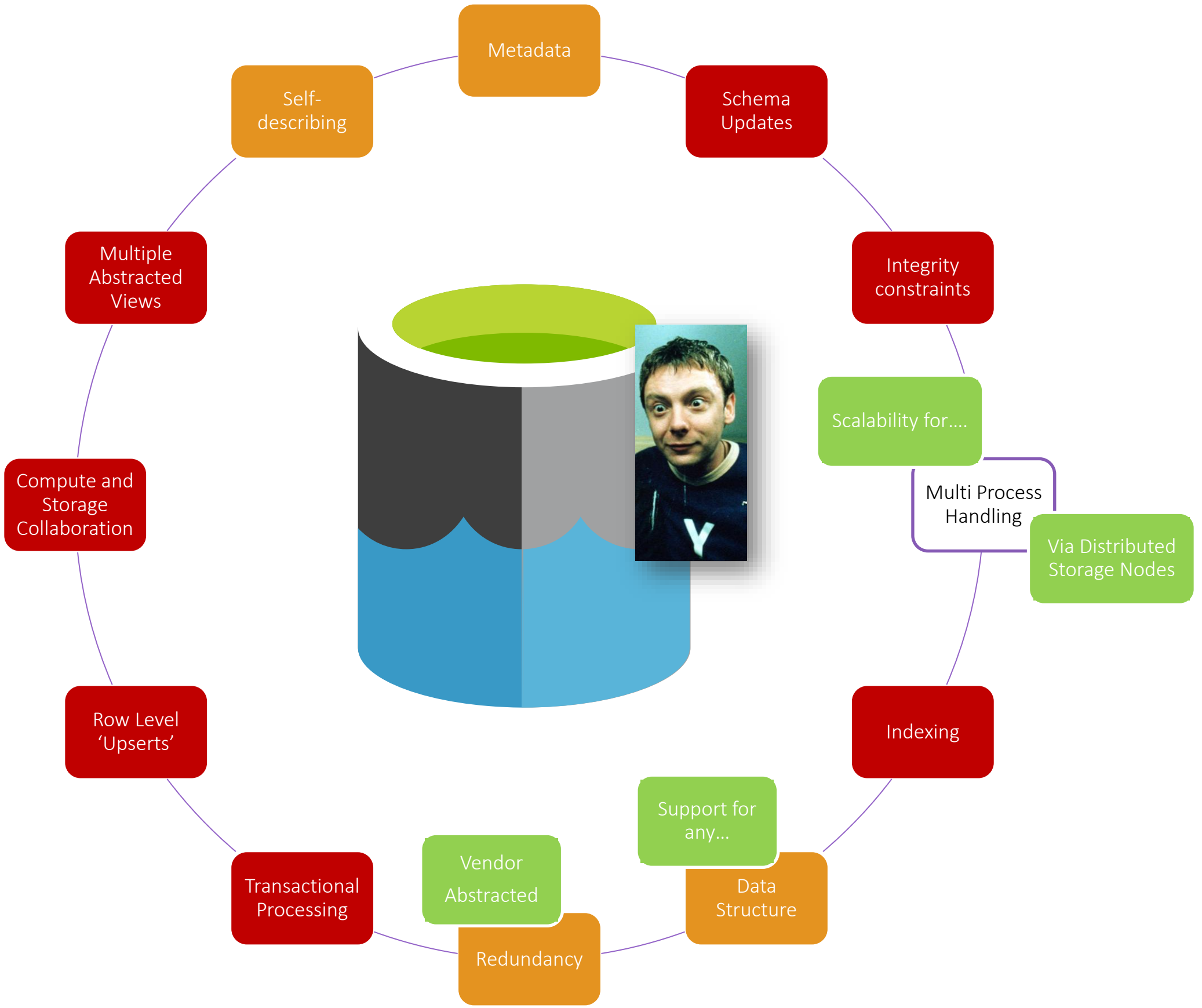


Big Data:
Volume
Velocity
Variety
Veracity
Value





Big Data:
Volume
Velocity
Variety
Veracity
Value



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



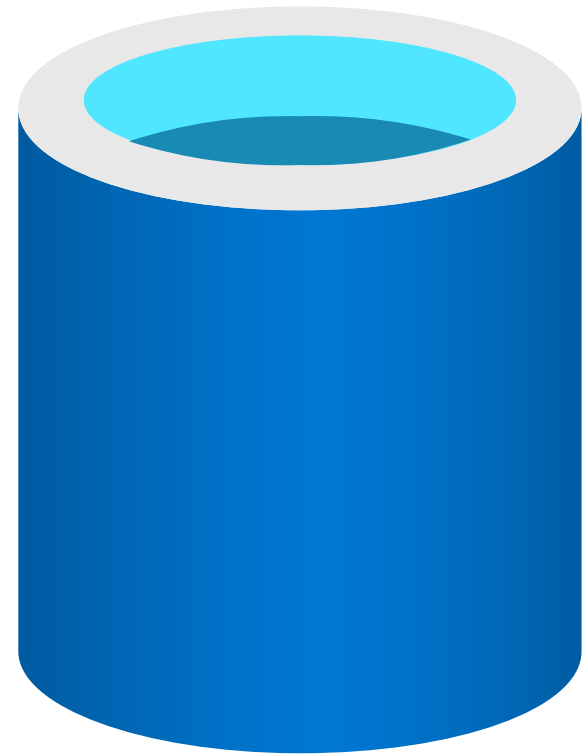
Scales Up	Scales Out
Natural Home for Structured Data	Any Data Structure
Storage Limits	No Storage Limits
Transactional Resilience	No Transactional Handling
Storage & Compute Coupled	Storage & Compute Decoupled

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).



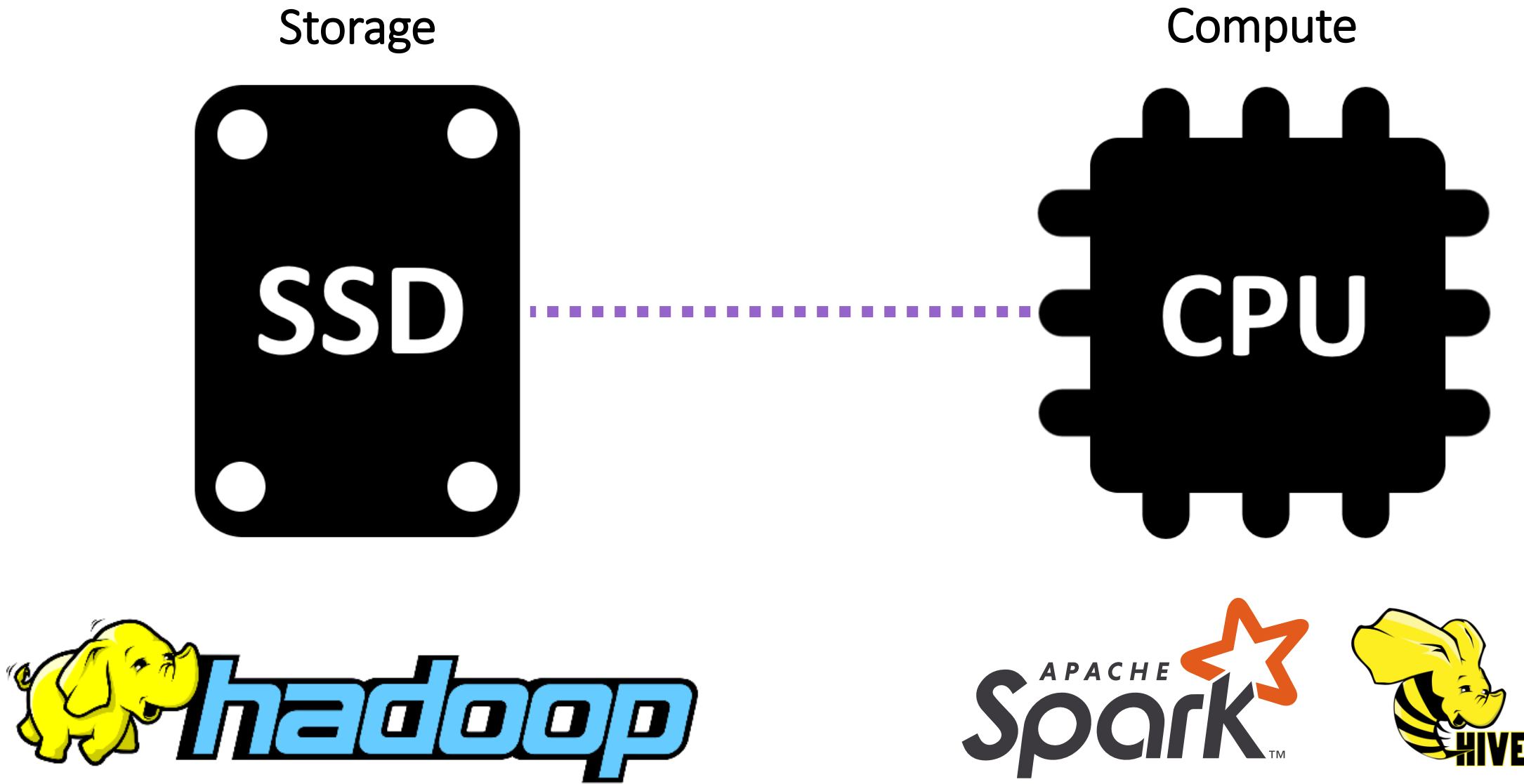
VS



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

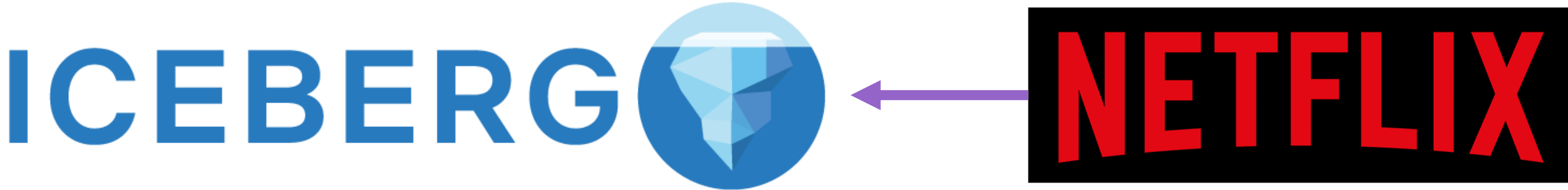
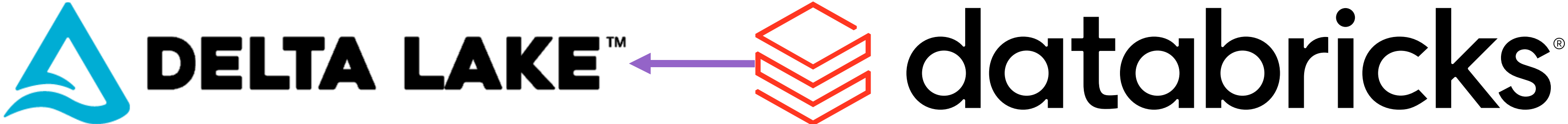


'Just' enable ACID transactional support for Data Lakes...



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

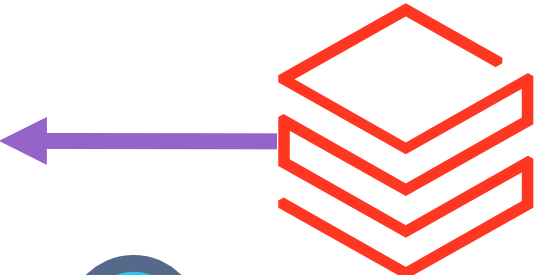
ACID Data Frameworks for Data Lakes



What is Delta Lake?

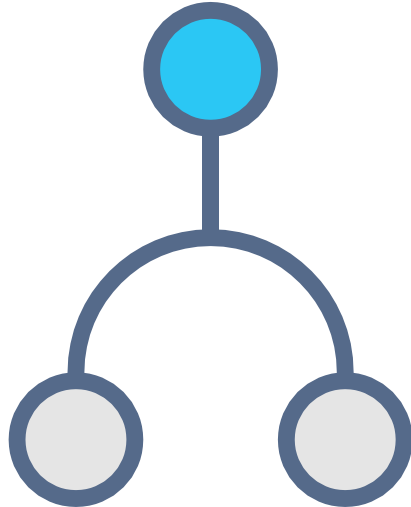


DELTA LAKE™



databricks®

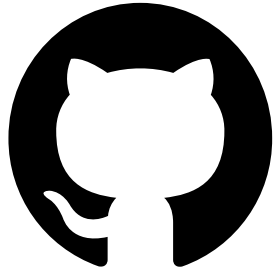
February 2019



What is Delta Lake?



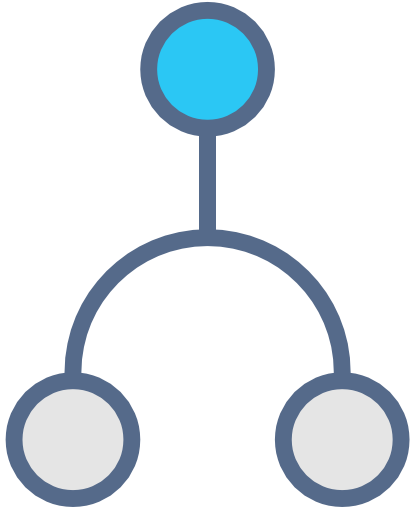
DELTA LAKE™



<https://delta.io>

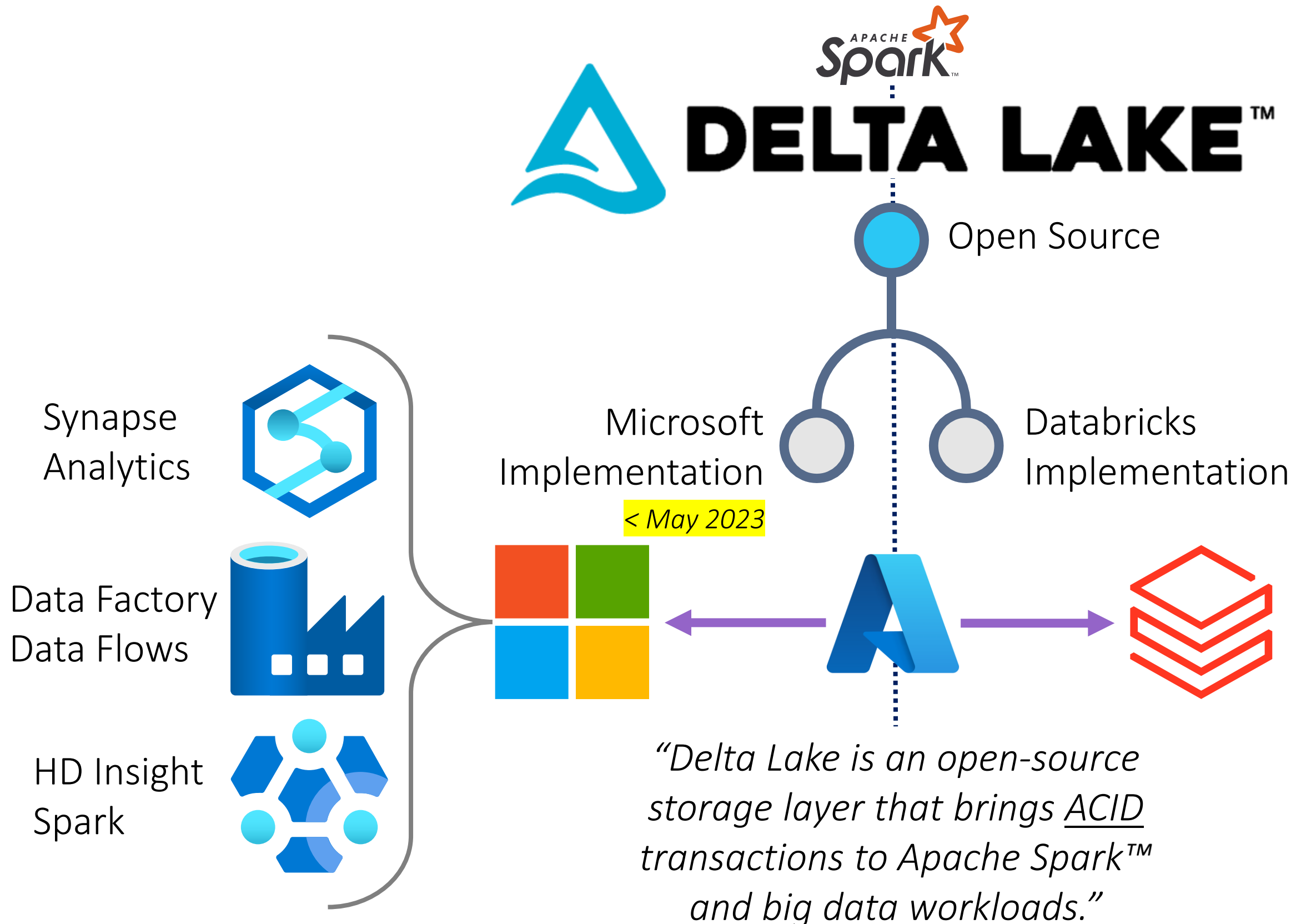
<https://github.com/delta-io/delta>

April 2019

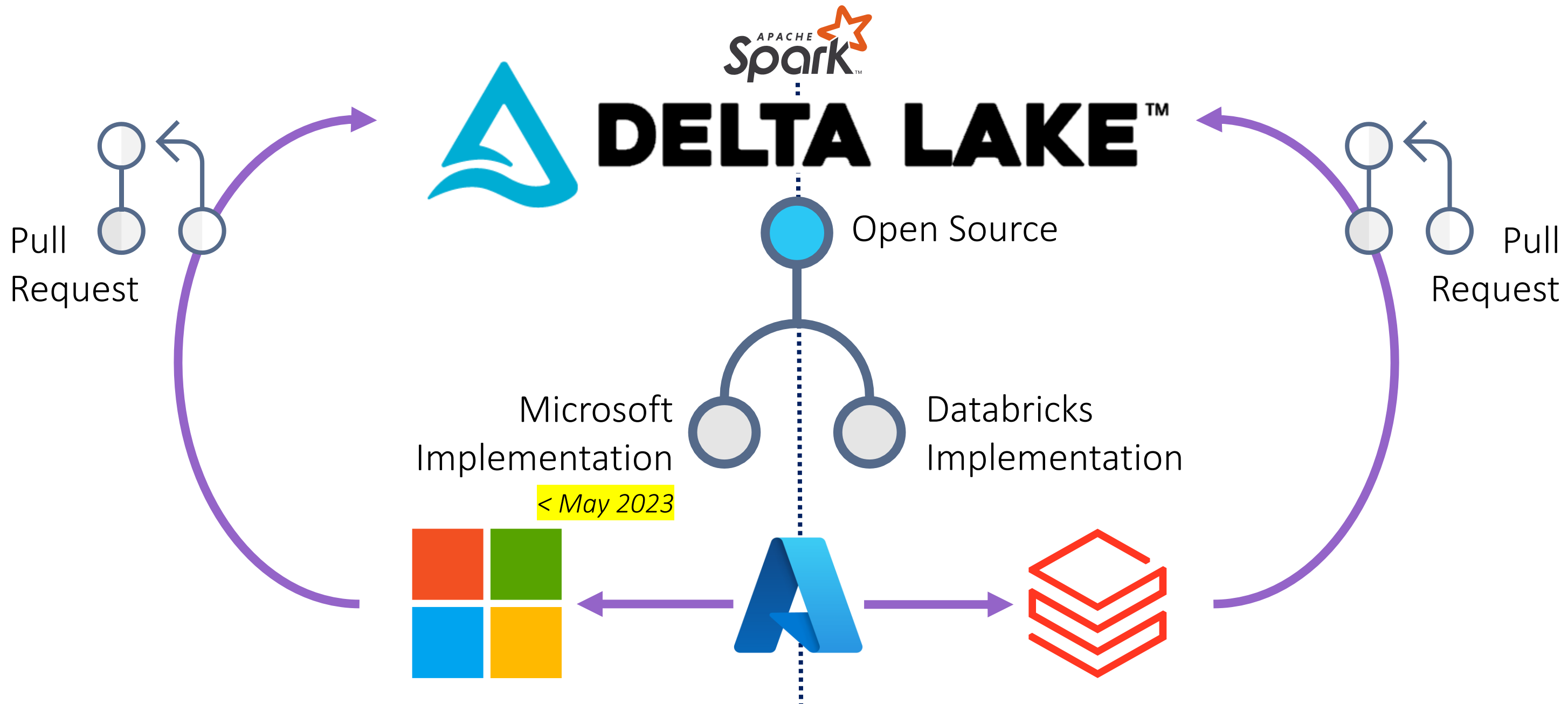


databricks®

What is Delta Lake?

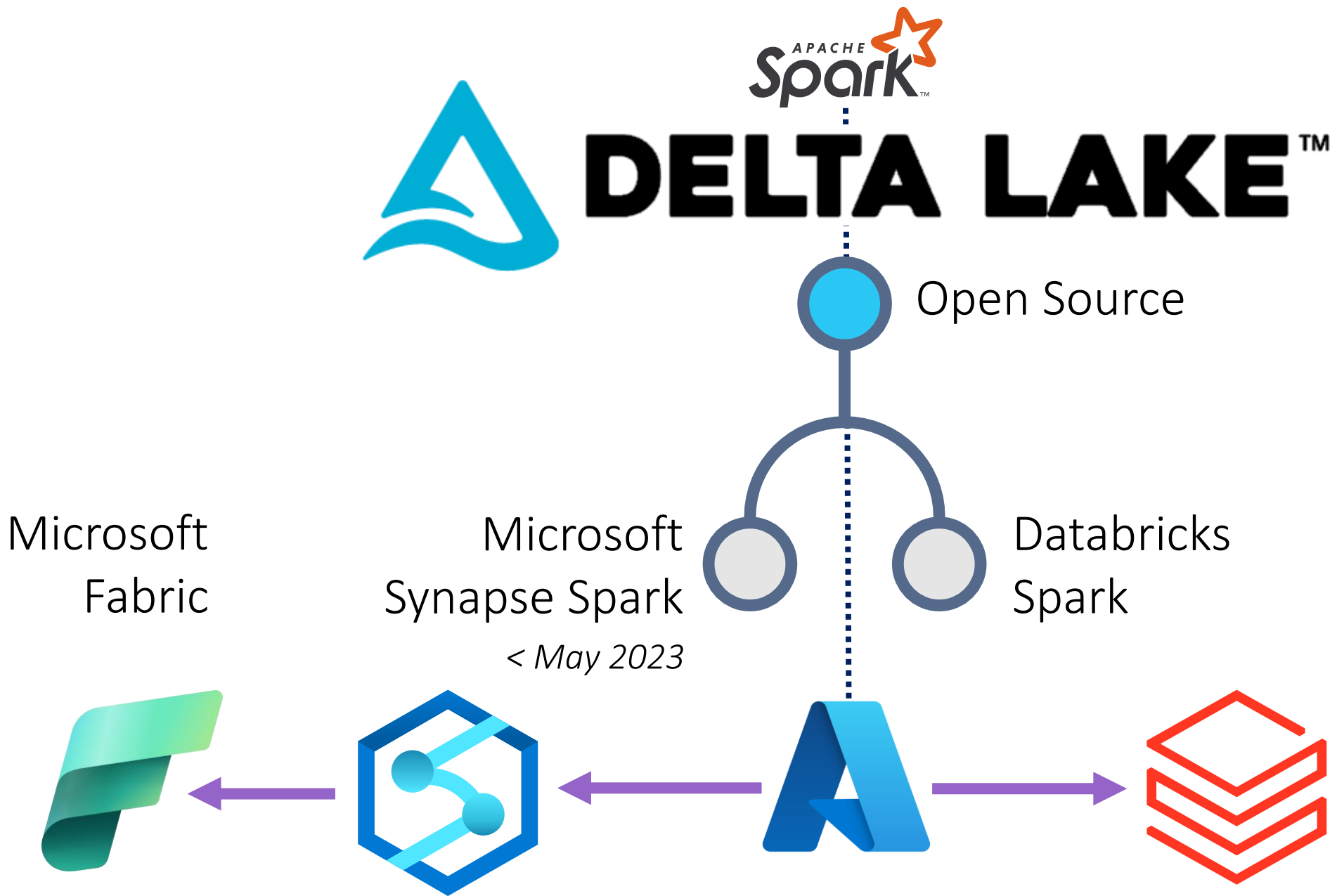


What is Delta Lake?



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

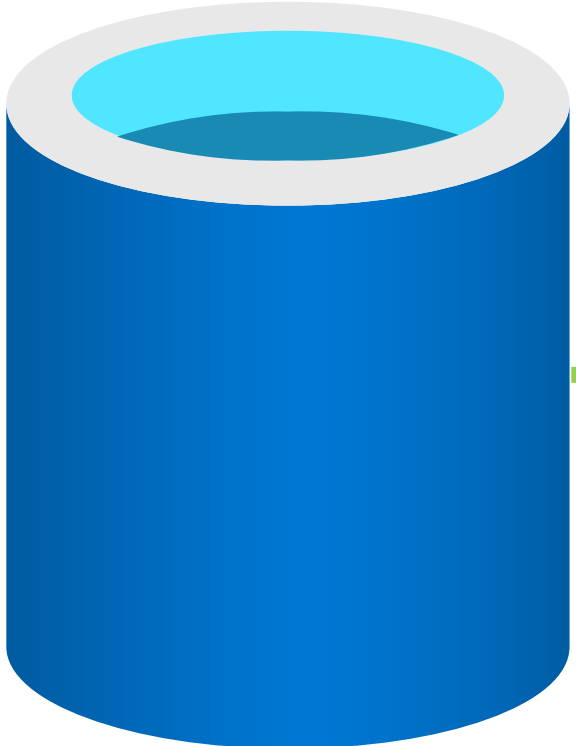
Which Spark Implementation is Better?



Data Warehouse



Online
Line
Transactional
Processing



Application
Data



Extract
Transform
Load



Data
Warehouse



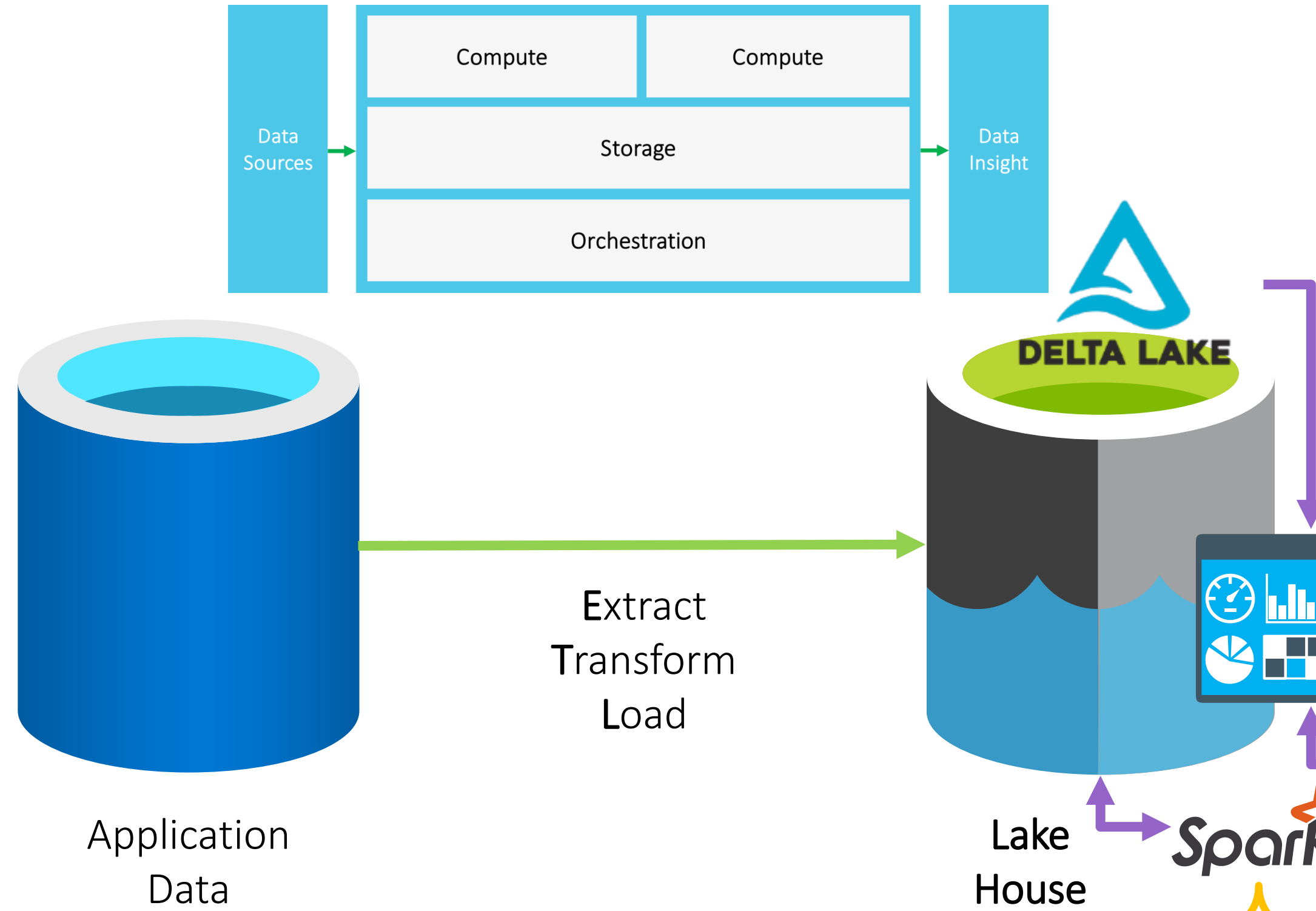
Offline
Analytical
Transactional
Processing

Lake House



Cloud Formations - Knowledge Transfer & Training

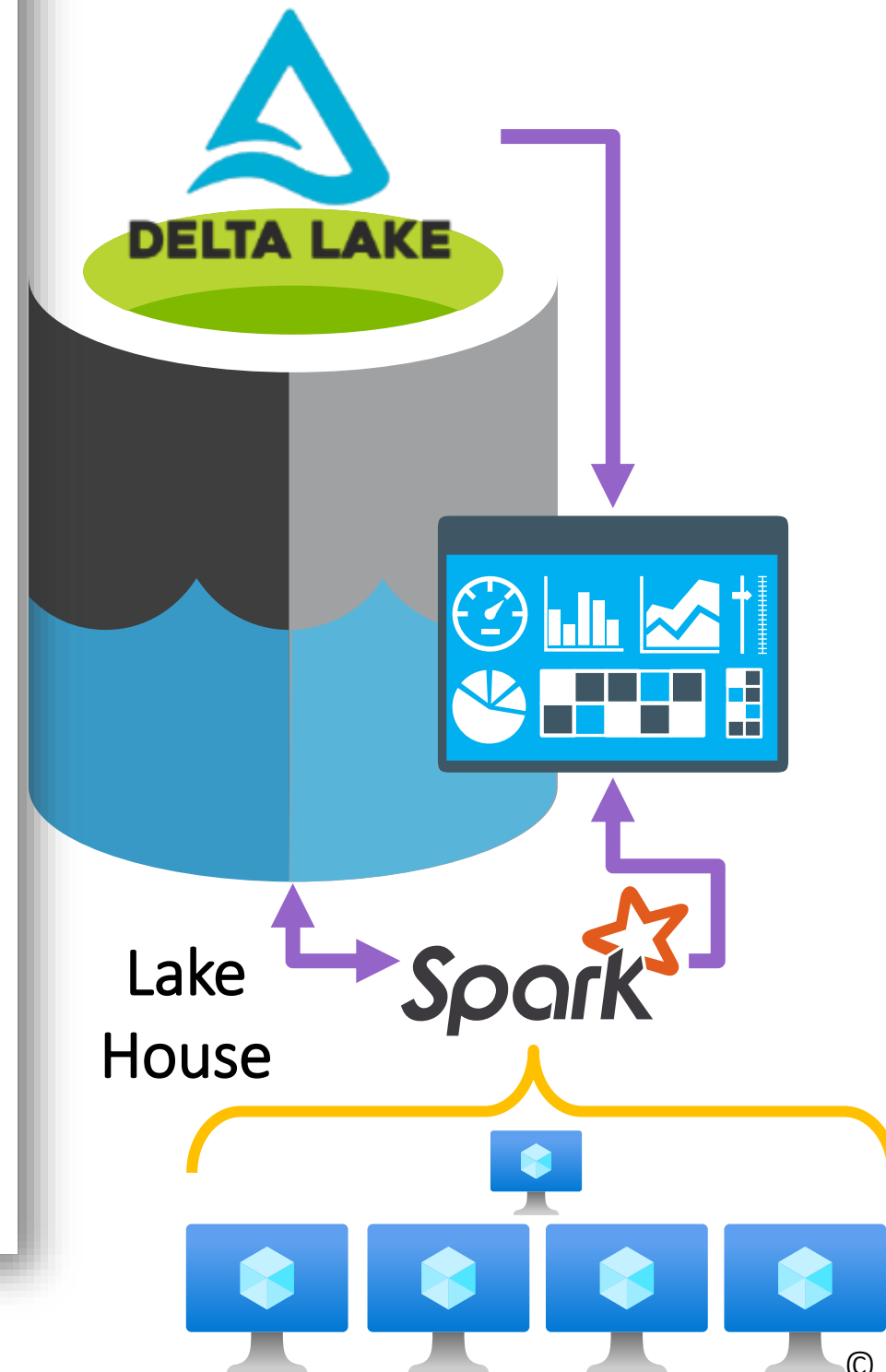
Online
Line
Transactional
Processing



Offline
Analytical
Transactional
Processing



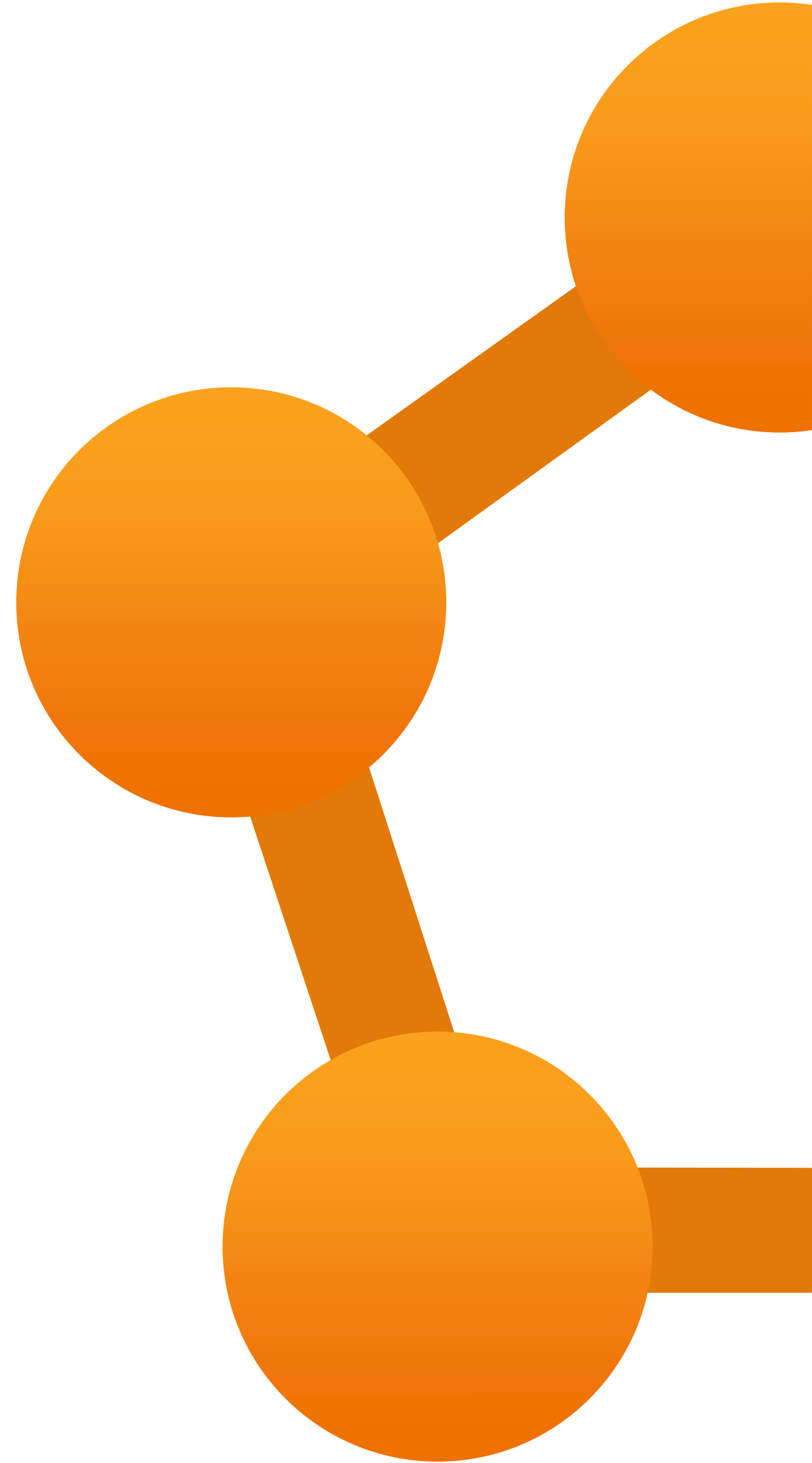
The screenshot shows the Wikipedia article for "The Lake House (film)". At the top, it says "Not logged in" with links for "Talk", "Contributions", "Create account", and "Log in". Below that are tabs for "Article" and "Talk", and buttons for "Read", "Edit", and "View history". A search bar is also present. The article title is "The Lake House (film)" with the subtitle "From Wikipedia, the free encyclopedia". A yellow warning box states: "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 main text describes the film as a 2006 American fantasy romantic drama directed by Alejandro Agresti, starring Keanu Reeves and Sandra Bullock. A "Contents" table of contents is visible on the left, listing sections like Plot, Cast, Production, Music, Reception, References, and External links. The "Plot" section begins with "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." A "Theatrical release poster" is shown on the right, featuring Keanu Reeves and Sandra Bullock. Below the poster, the film's credits are listed: Directed by Alejandro Agresti, Written by David Auburn, Based on Il Mare by Kim Eun-jeong and Kim Mi-yeong, Produced by Doug Davison and Roy Lee, and Starring Keanu Reeves.



Lambda & Kappa

λ K

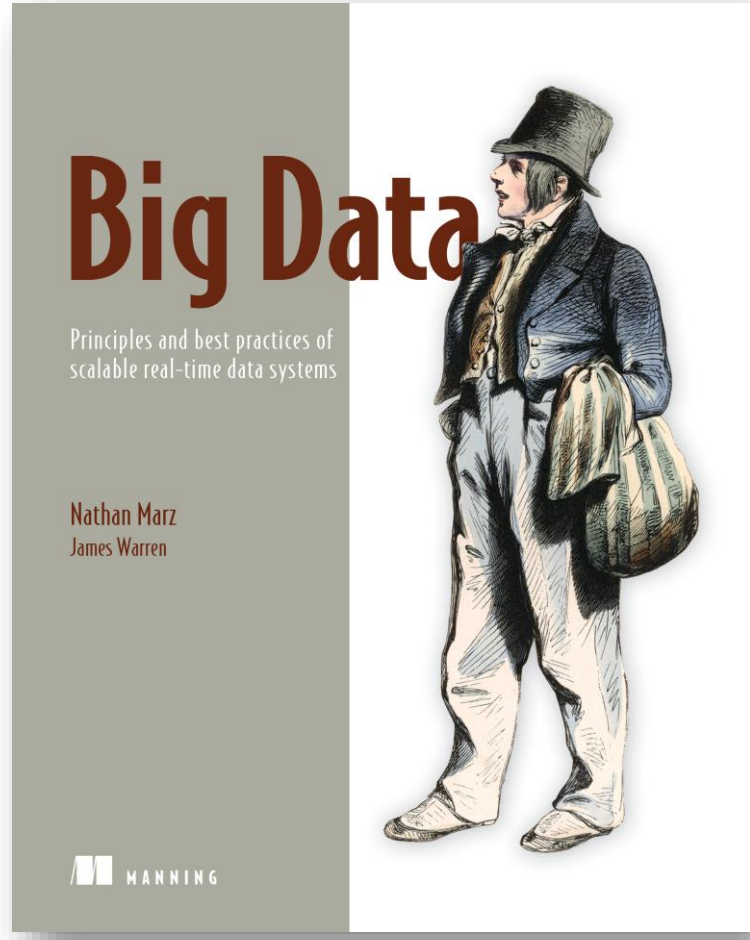
Cloud Formations



Lambda & Kappa Architectures

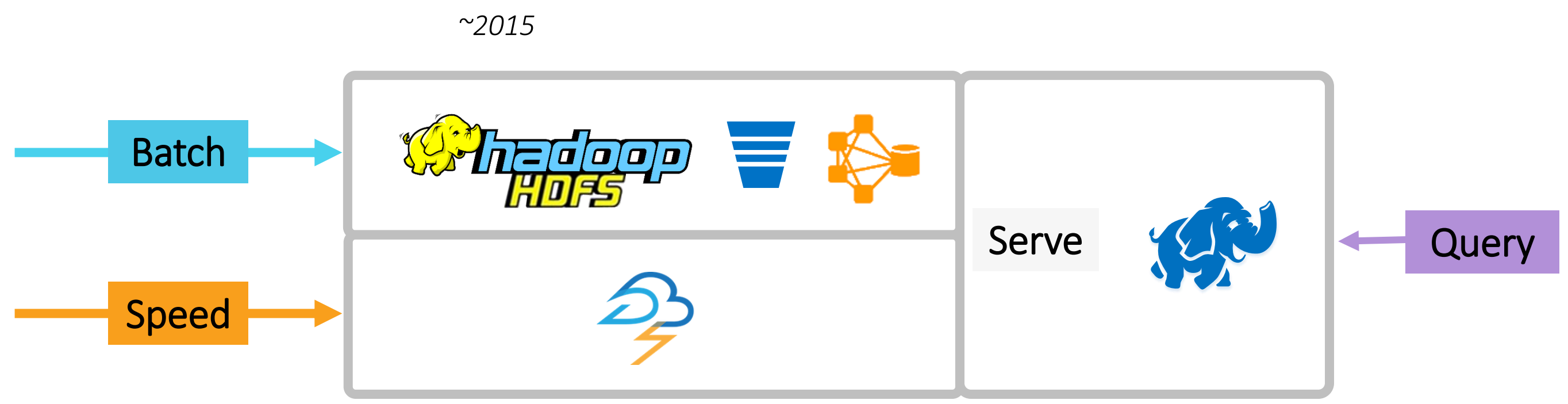


Cloud Formations - Knowledge Transfer & Training

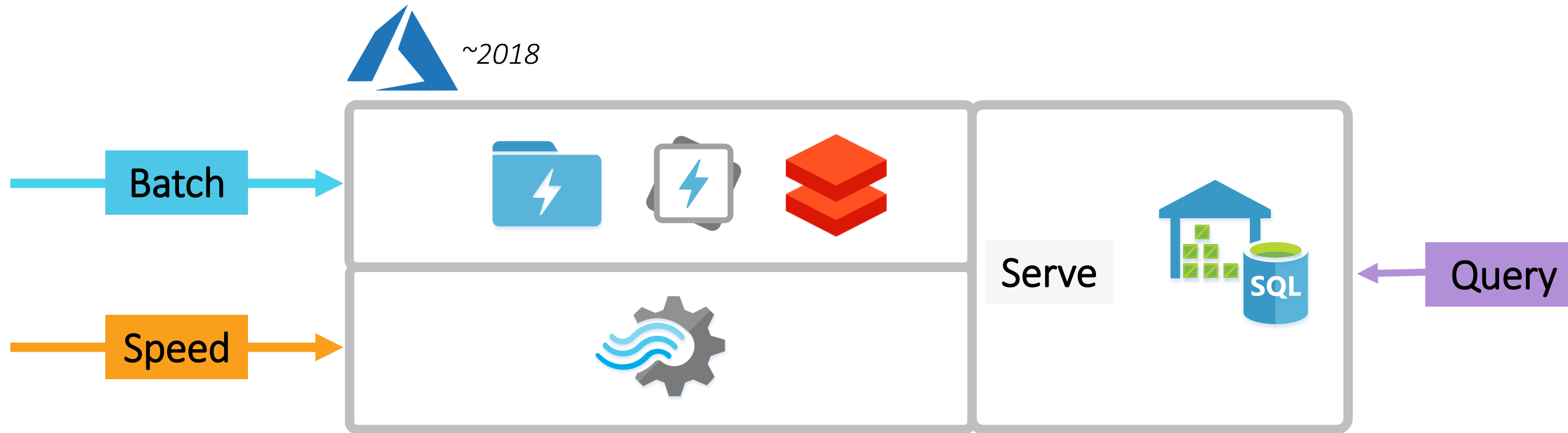


Big Data:
Volume
Velocity
Variety
Veracity
Value

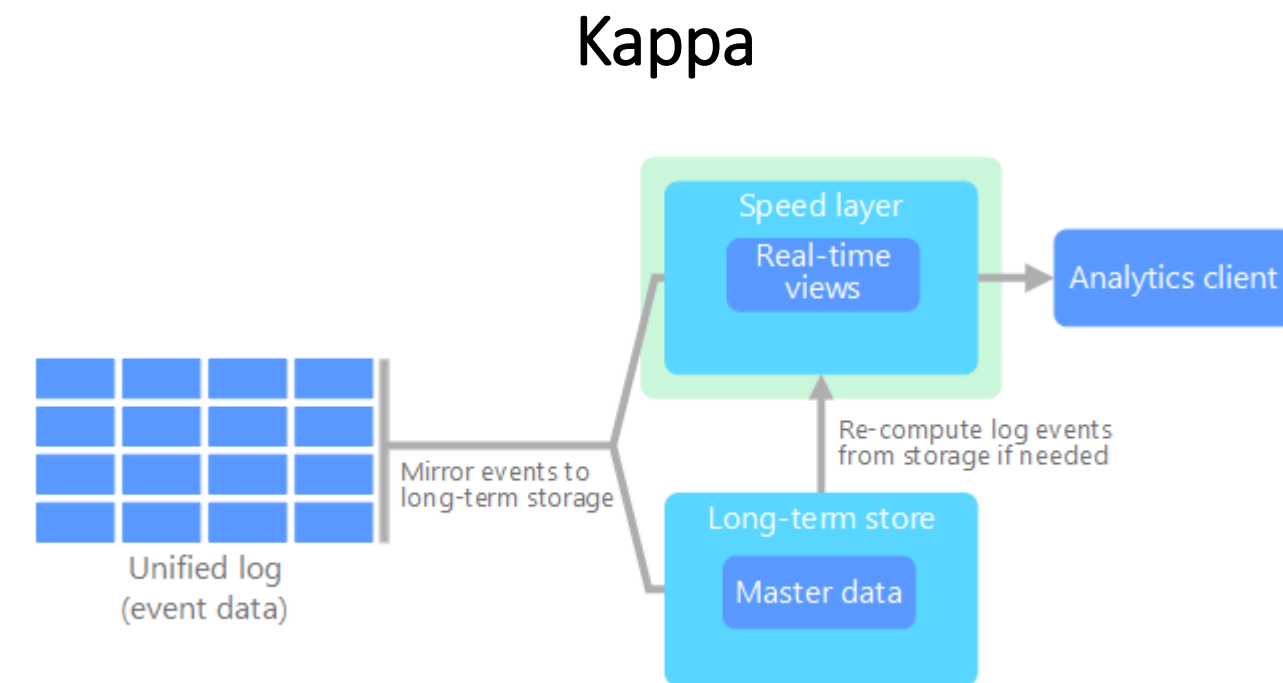
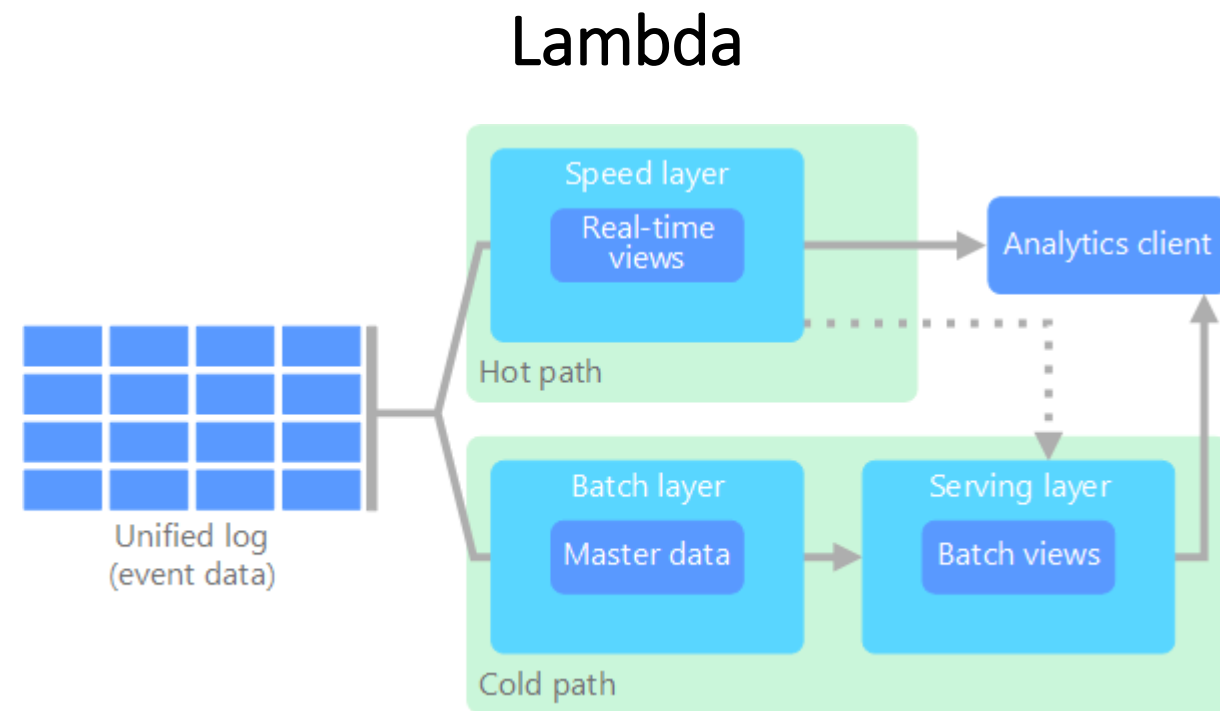
Lambda & Kappa Architectures



Lambda & Kappa Architectures



Lambda & Kappa Architectures



“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.”

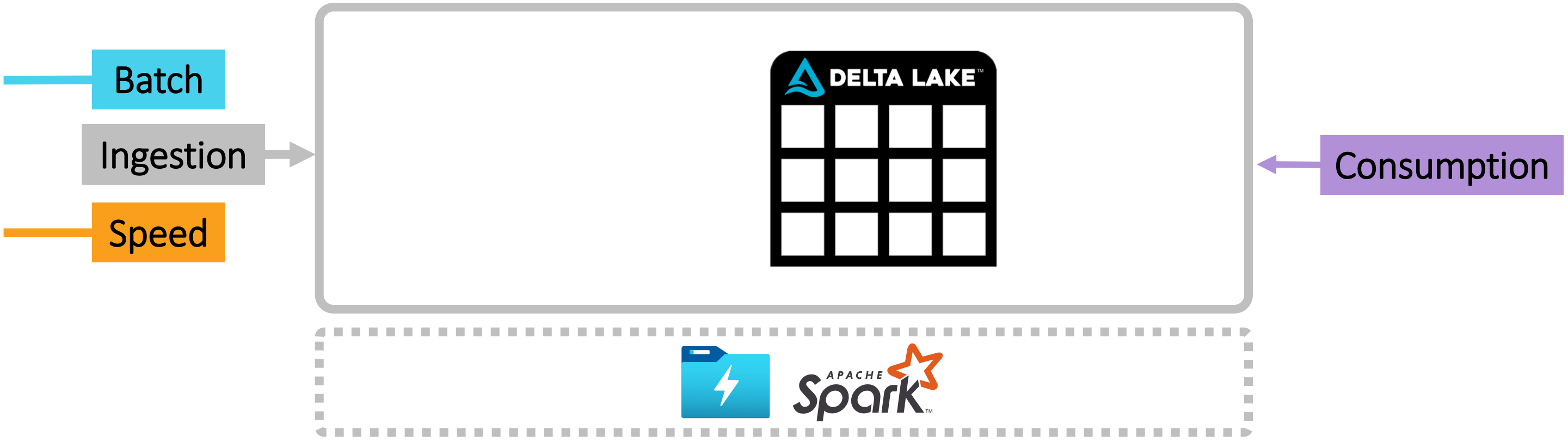
“A drawback to the lambda architecture is its **complexity**. **Processing logic appears in two different places** — the cold and hot paths — using different frameworks. This leads to duplicate computation logic and the complexity of managing the architecture for both paths.

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.”

Lambda & Kappa Architectures



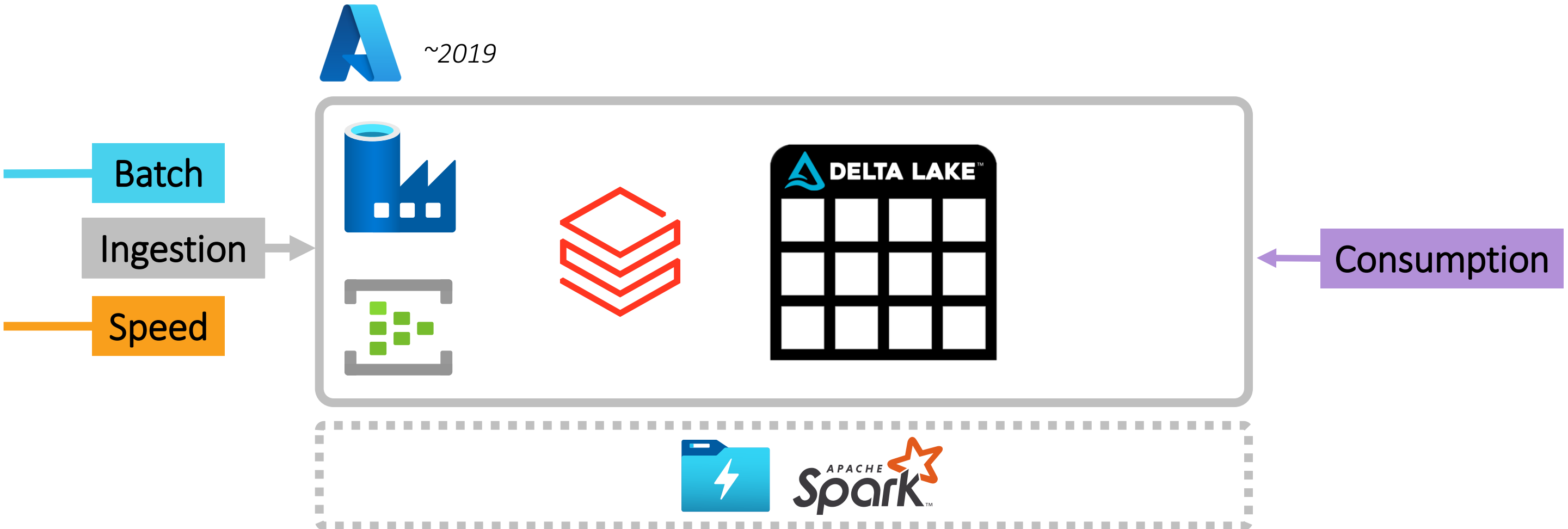
Cloud Formations - Knowledge Transfer & Training



Lambda & Kappa Architectures



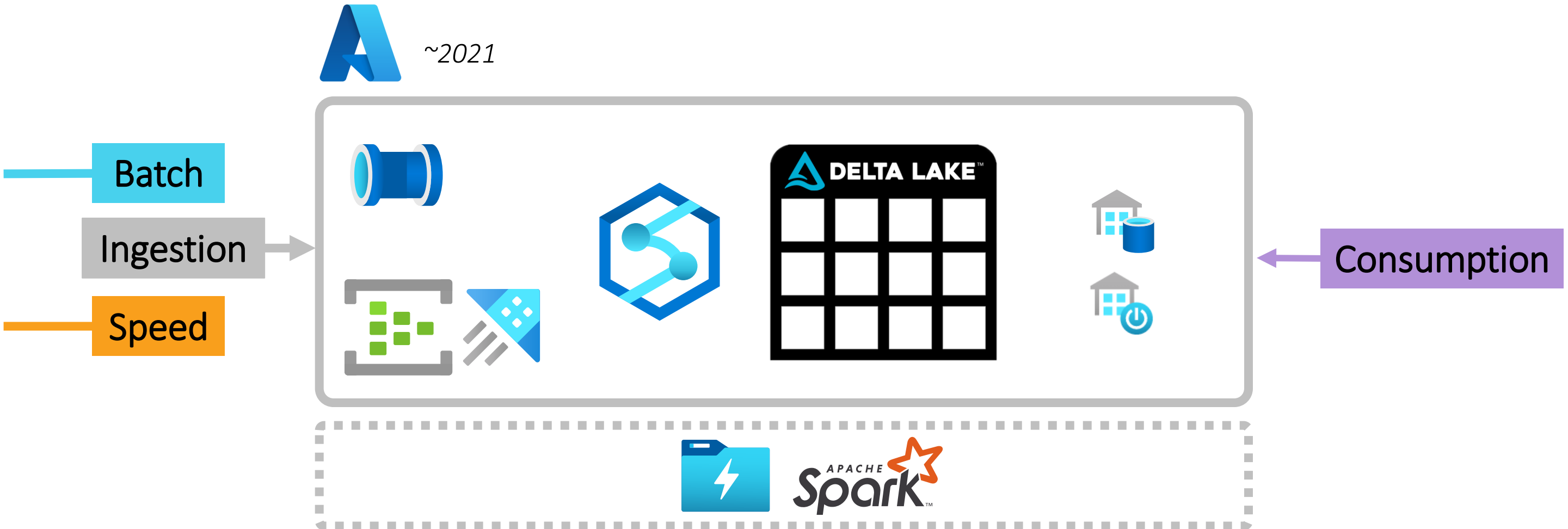
Cloud Formations - Knowledge Transfer & Training



Lambda & Kappa Architectures



Cloud Formations - Knowledge Transfer & Training

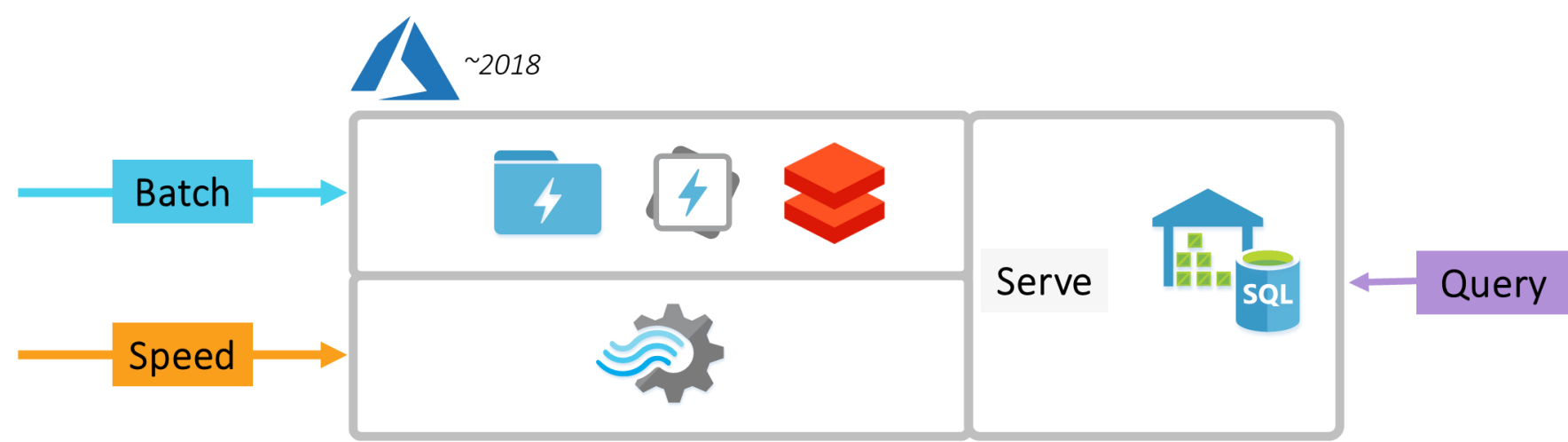
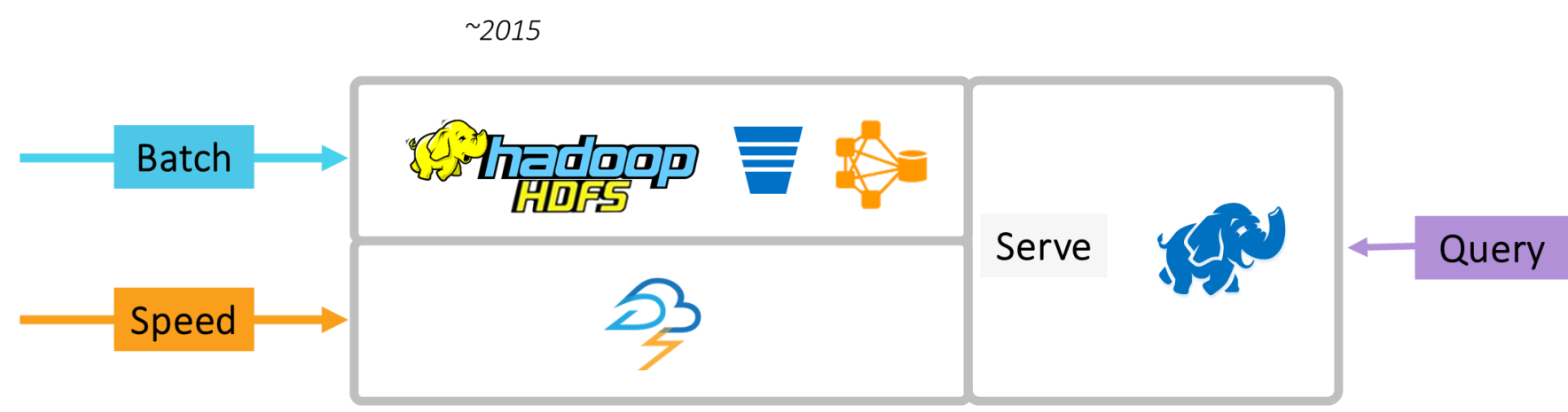


Lambda & Kappa Architectures vs Technology

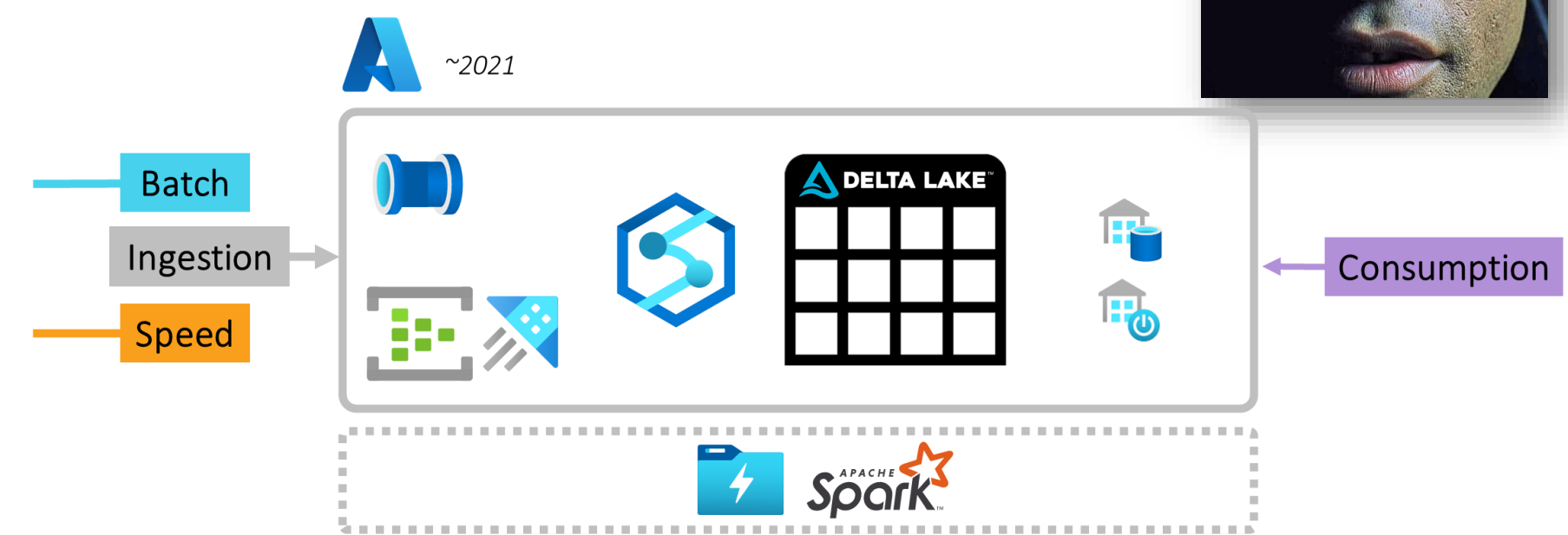
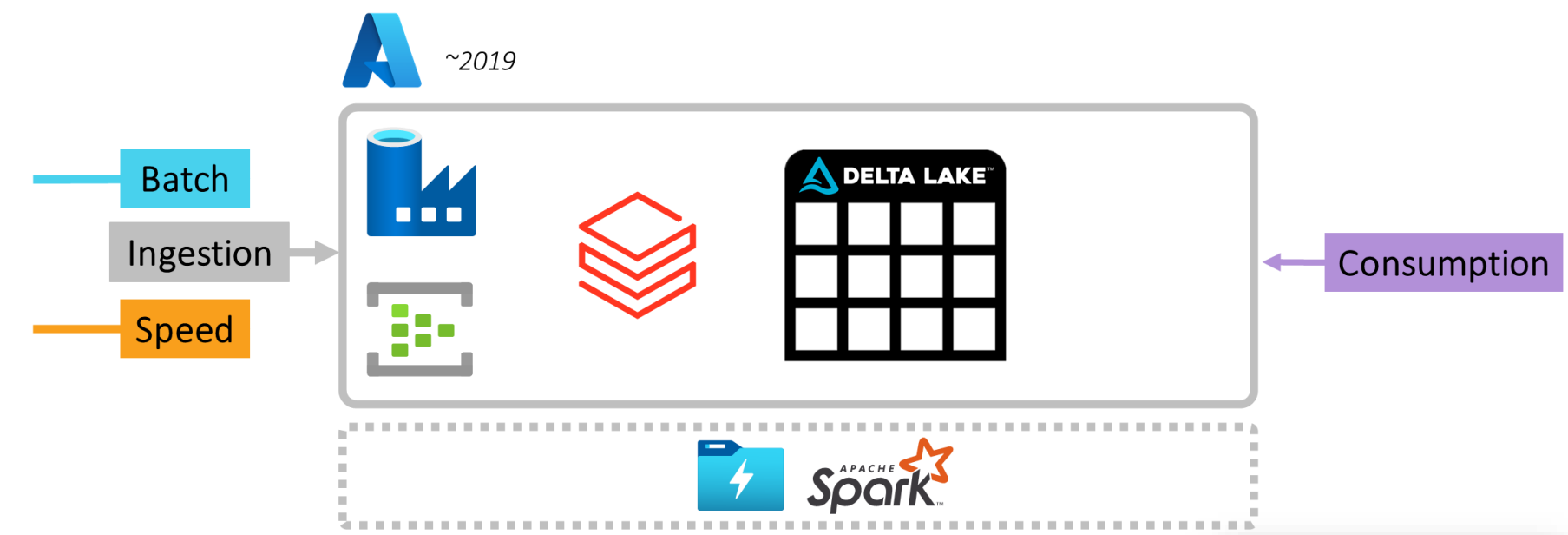


Cloud Formations - Knowledge Transfer & Training

Lambda

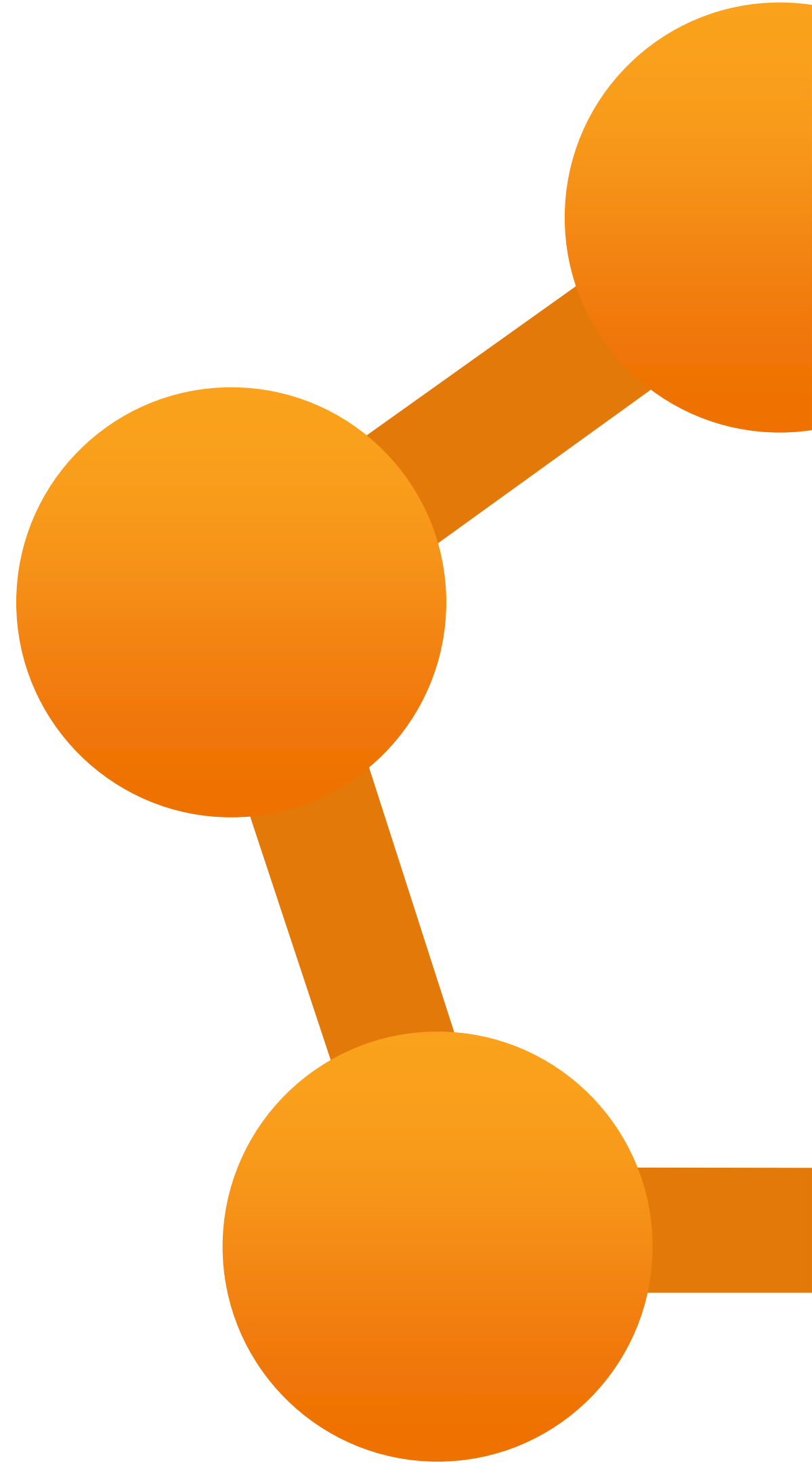


Kappa





Cloud Formations



Different Types of Fabric



Different Types of Fabric



The screenshot shows the Gartner Glossary page for 'Data Fabric'. The page header includes the Gartner logo and 'Information Technology'. The main title is 'Gartner Glossary' with a sub-section for 'Data Fabric'. The breadcrumb trail is 'Gartner Glossary > Information Technology Glossary > D > Data Fabric'. The main heading is 'Data Fabric'. The text describes it as an emerging data management design for attaining flexible, reusable and augmented data integration pipelines, services and semantics. It 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.

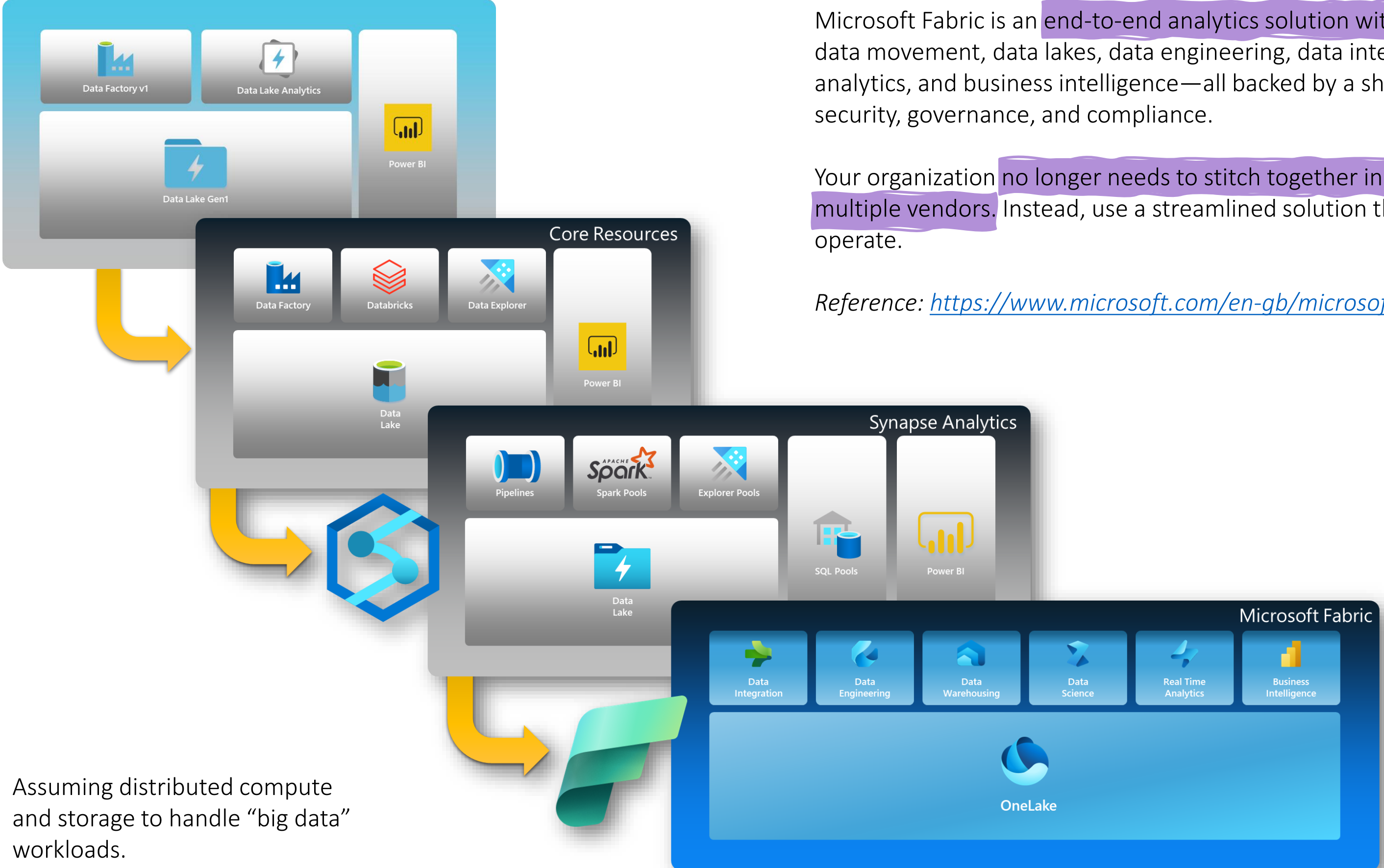
Ref: <https://www.gartner.com/en/information-technology/glossary/data-fabric>

The screenshot shows the Microsoft documentation page for 'What is Microsoft Fabric?'. The page header includes the Microsoft logo and 'Documentation'. The breadcrumb trail is 'Learn / Microsoft Fabric / Get started /'. The main title is 'What is Microsoft Fabric?'. The text describes it as 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.

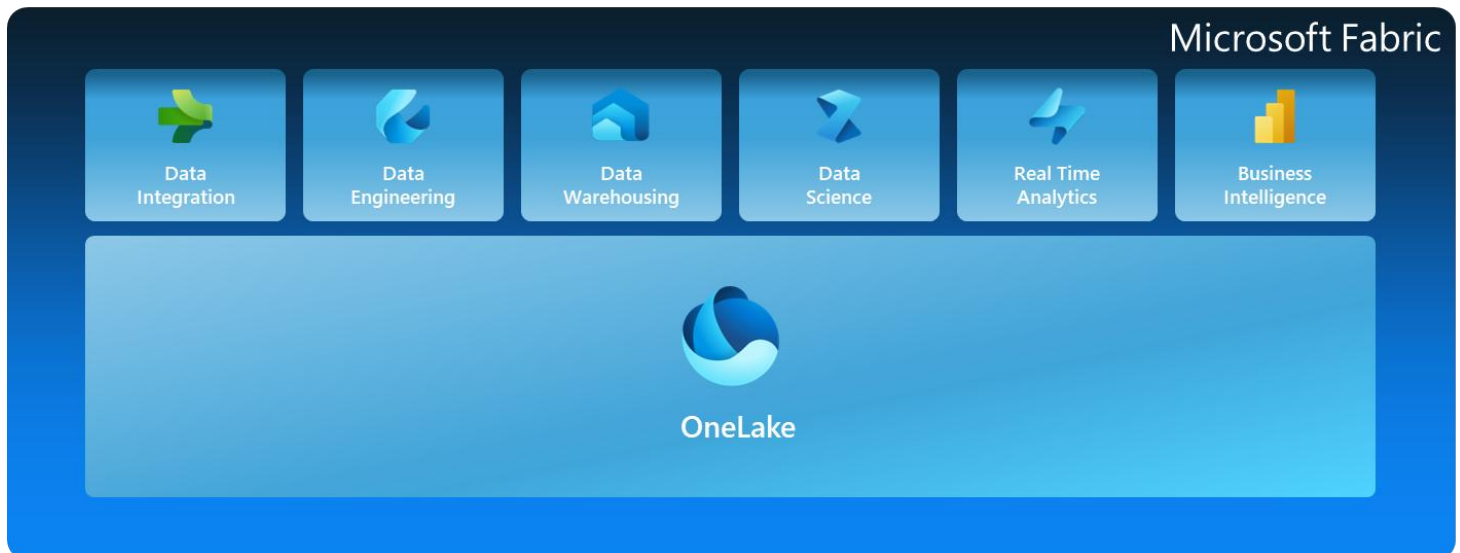
Your organization **no longer needs to stitch together individual analytics services from multiple vendors.** Instead, use a streamlined solution that's easy to connect, onboard, and operate.

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

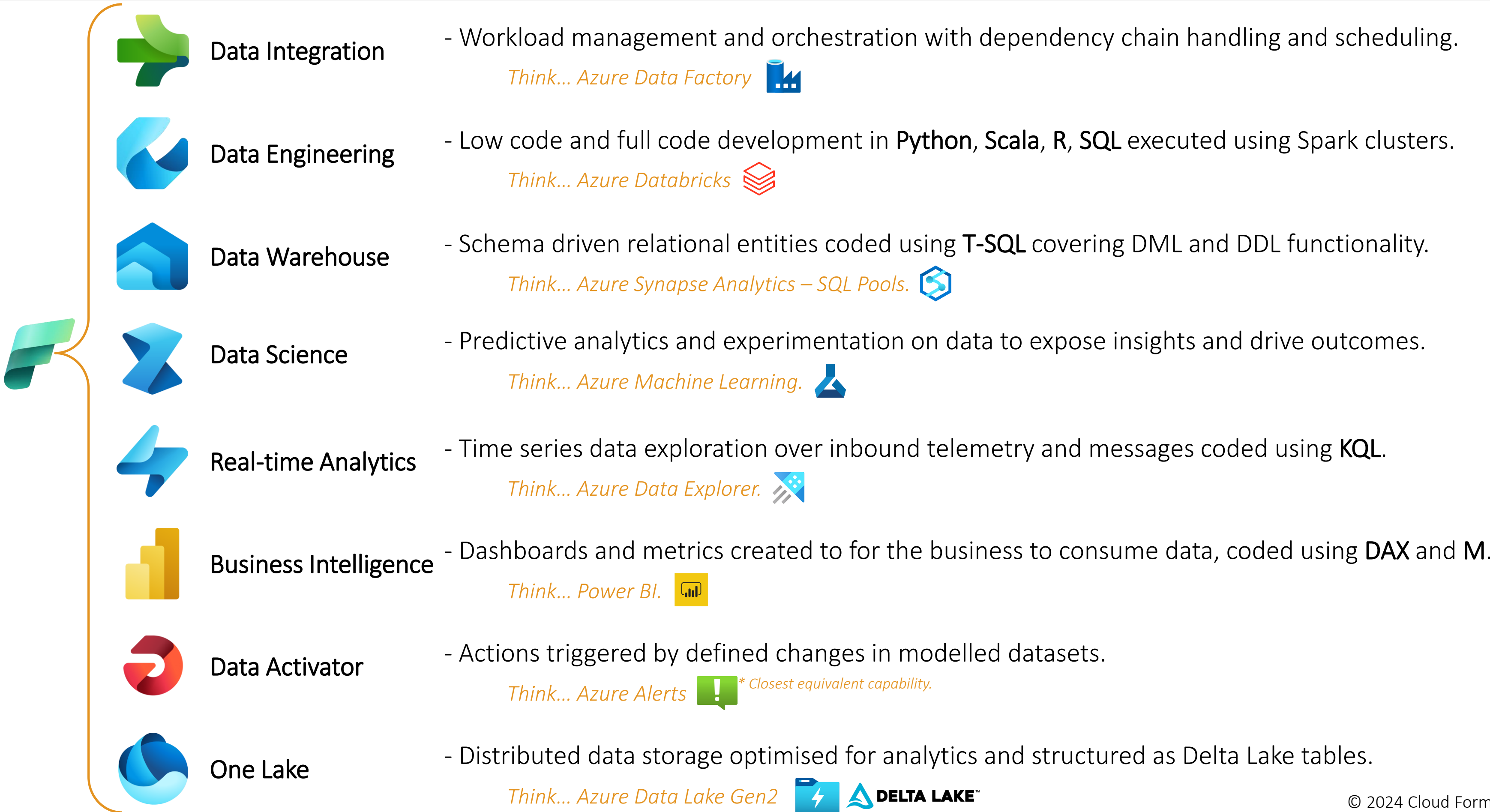
Cloud Formations - Knowledge Transfer & Training

Assuming distributed compute and storage to handle “big data” workloads.

What is Microsoft Fabric?



What is Microsoft Fabric? - Experiences vs Technical Capabilities



Microsoft Fabric vs a Kappa Architecture



Microsoft Fabric vs a Kappa Architecture



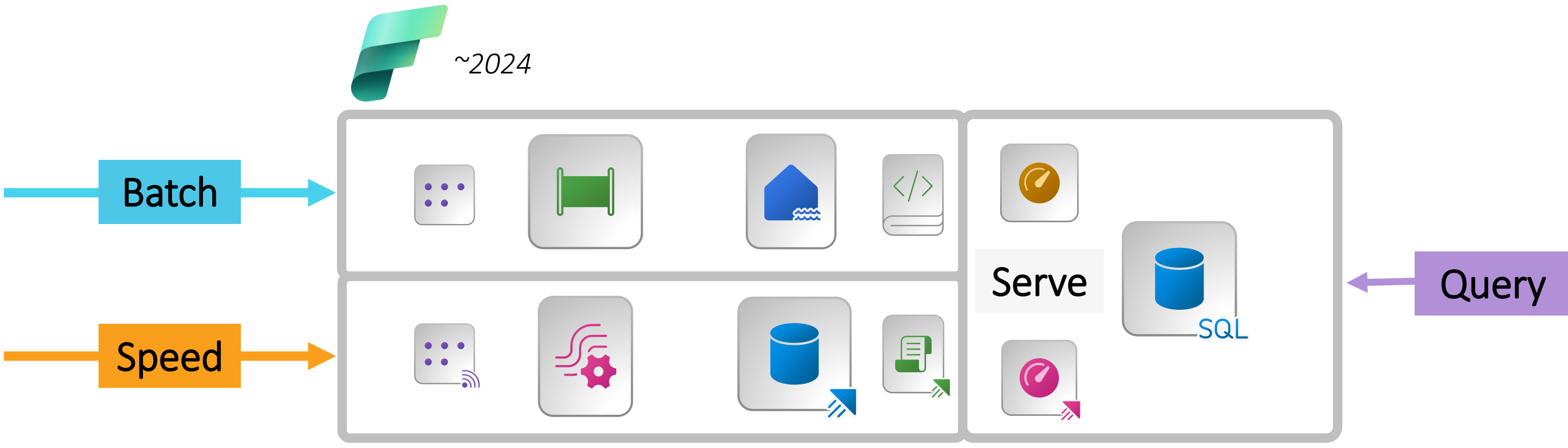
Cloud Formations - Knowledge Transfer & Training



Microsoft Fabric vs a Kappa Architecture



Cloud Formations - Knowledge Transfer & Training

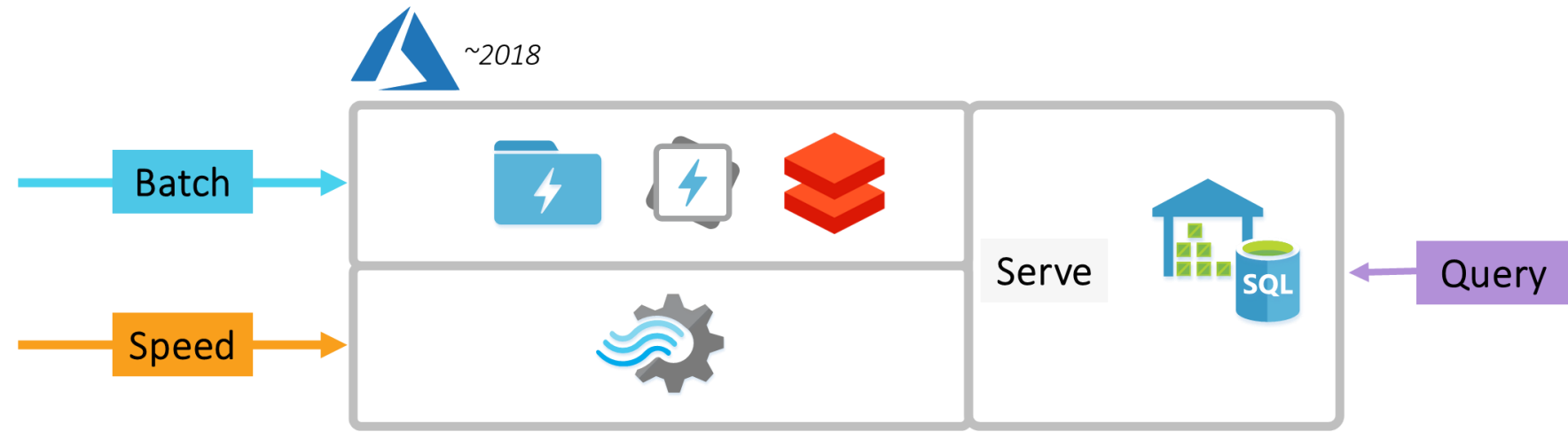


Lambda & Kappa Architectures vs Technology

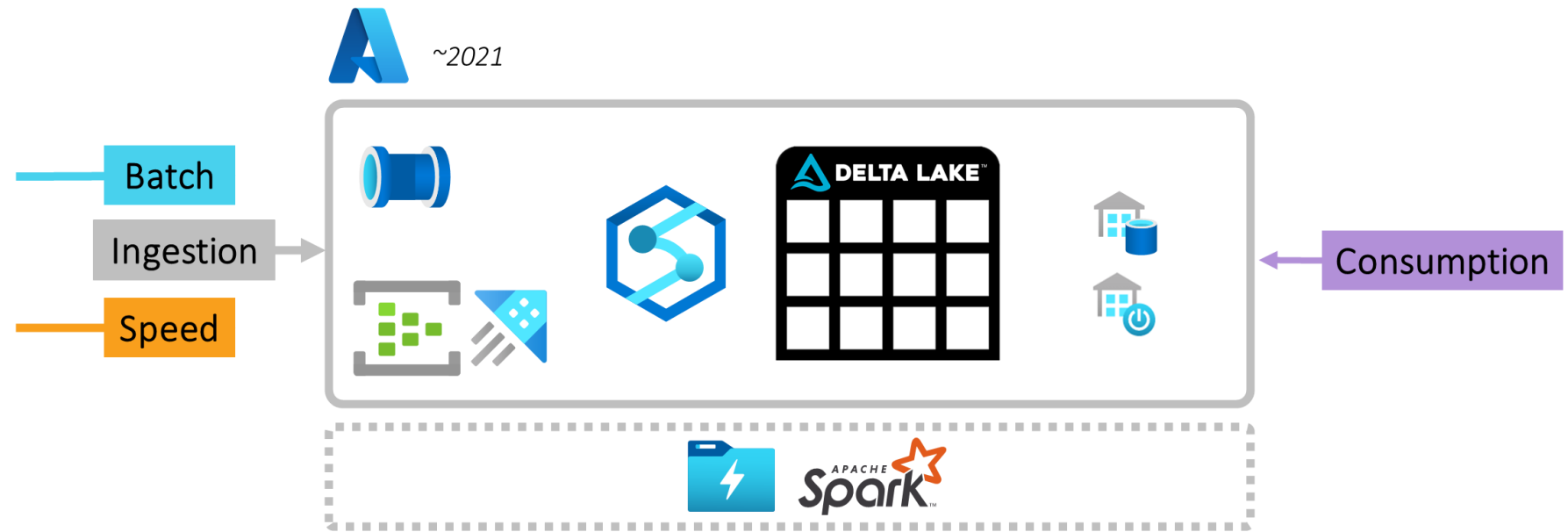
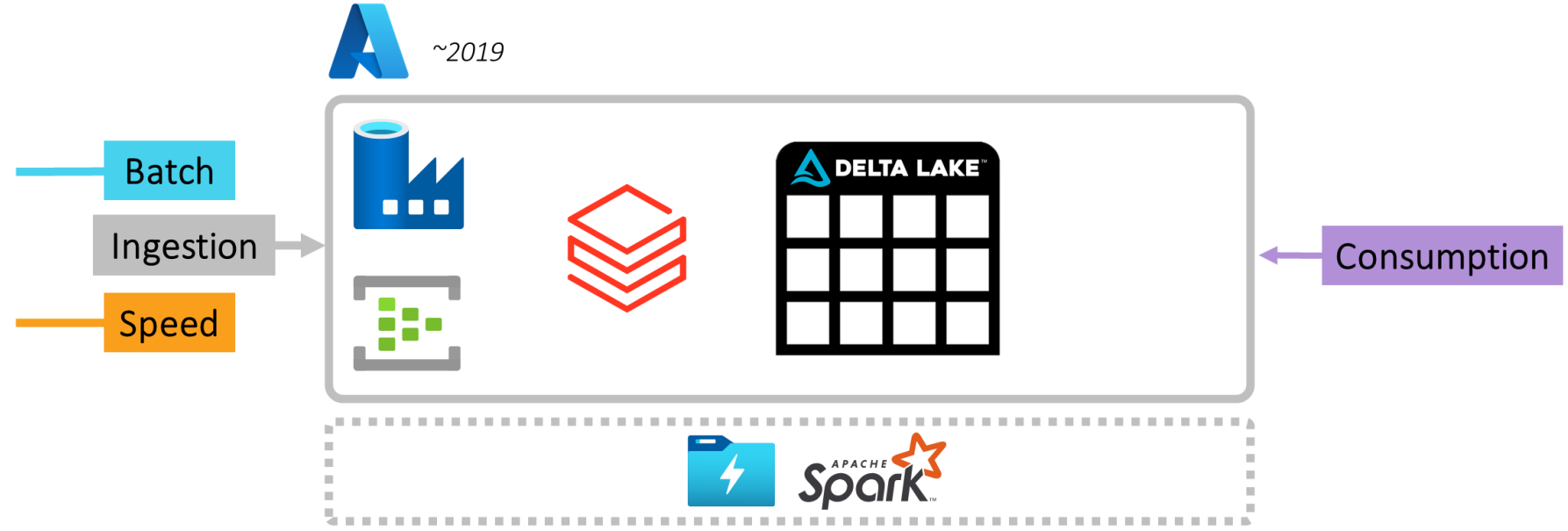


Cloud Formations - Knowledge Transfer & Training

Lambda



Kappa

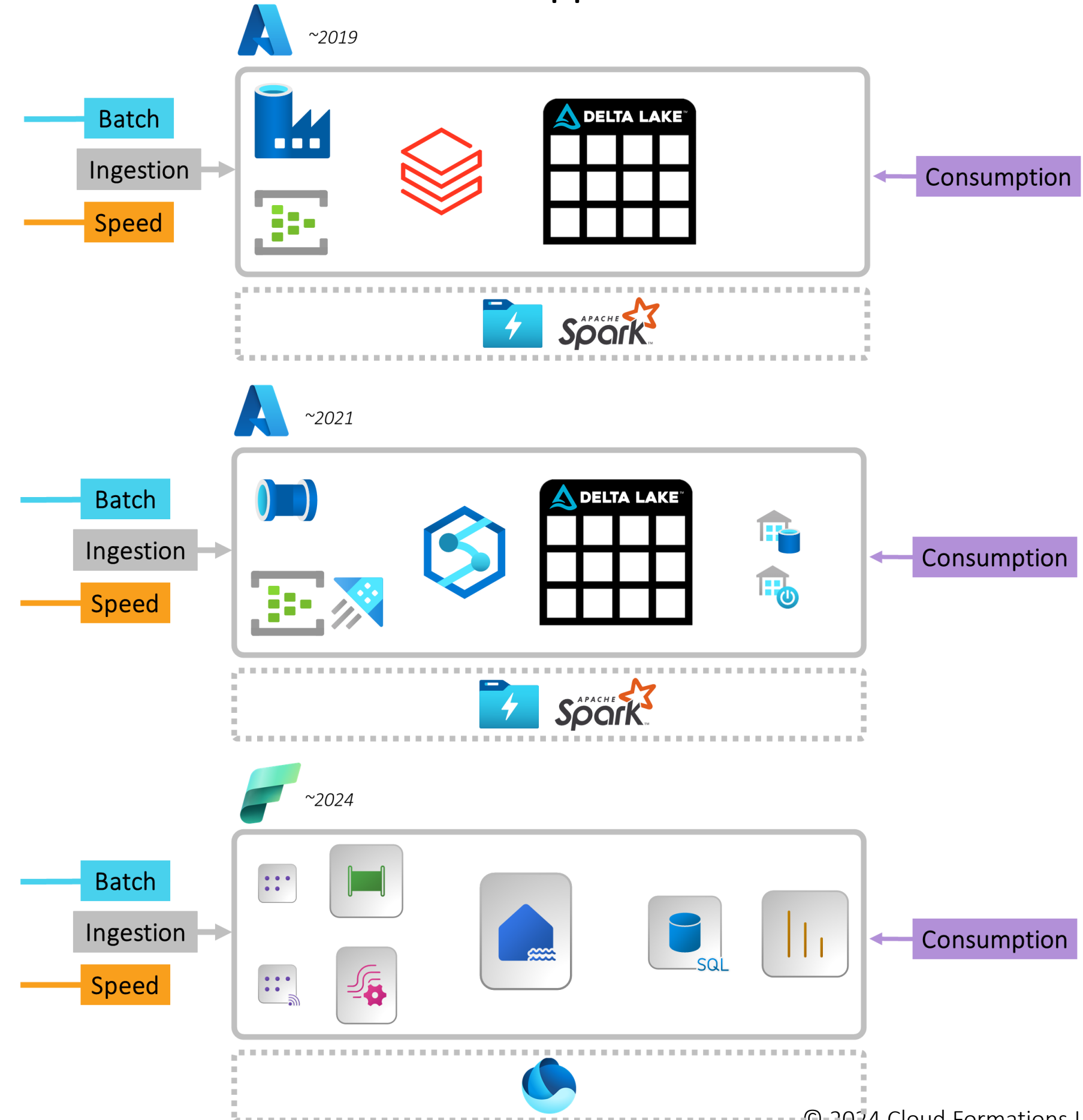
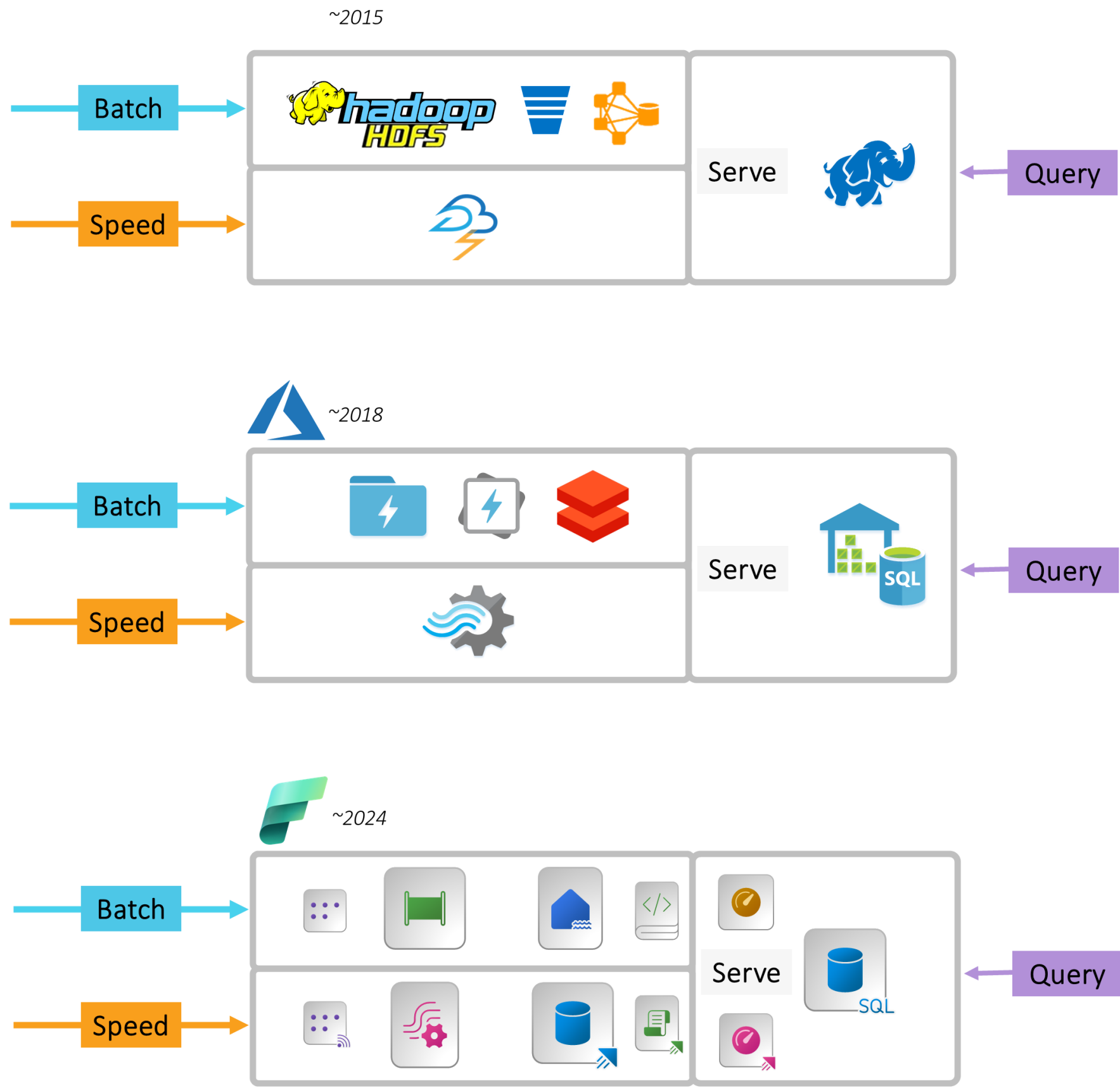


Lambda & Kappa Architectures vs Technology



Lambda

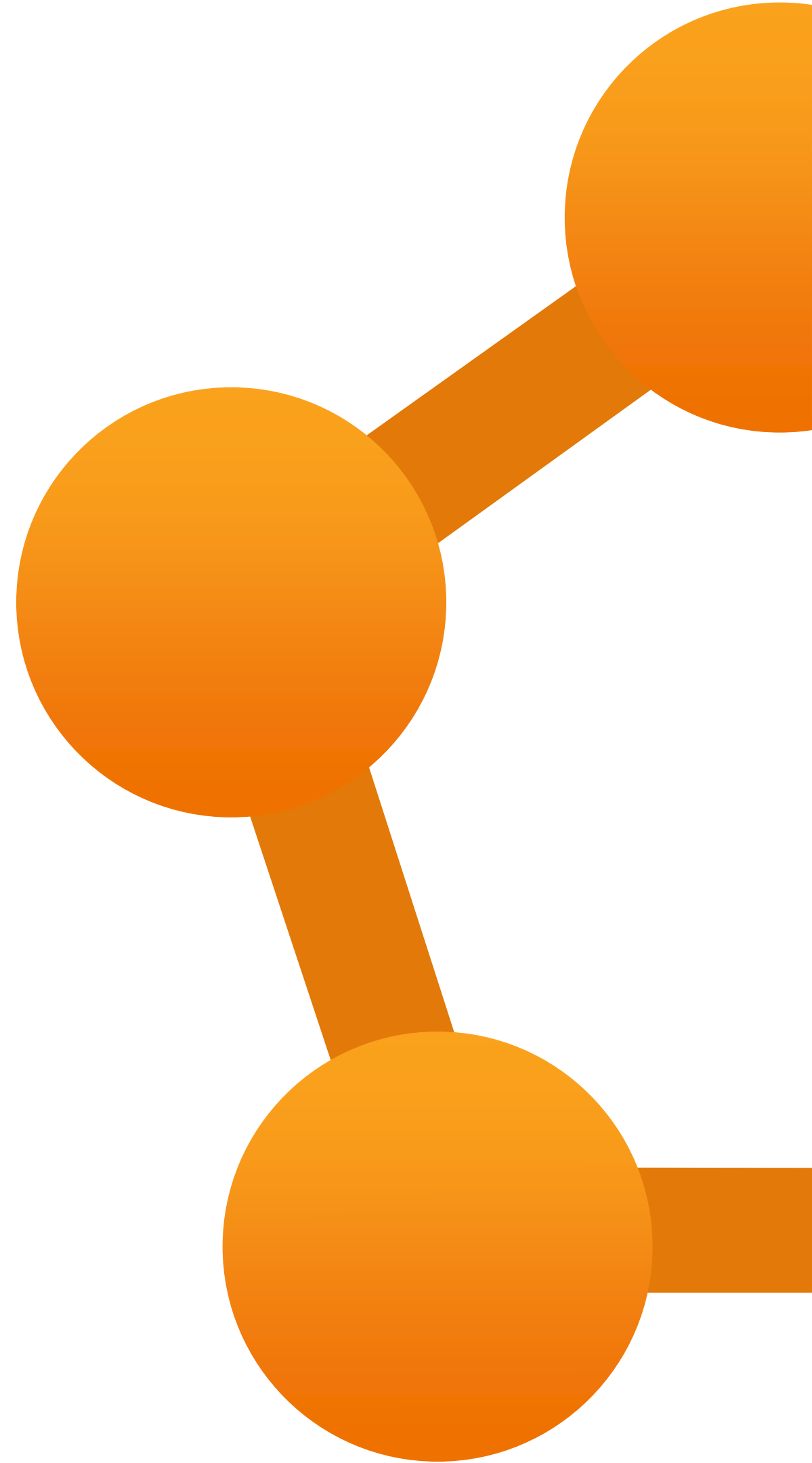
Kappa



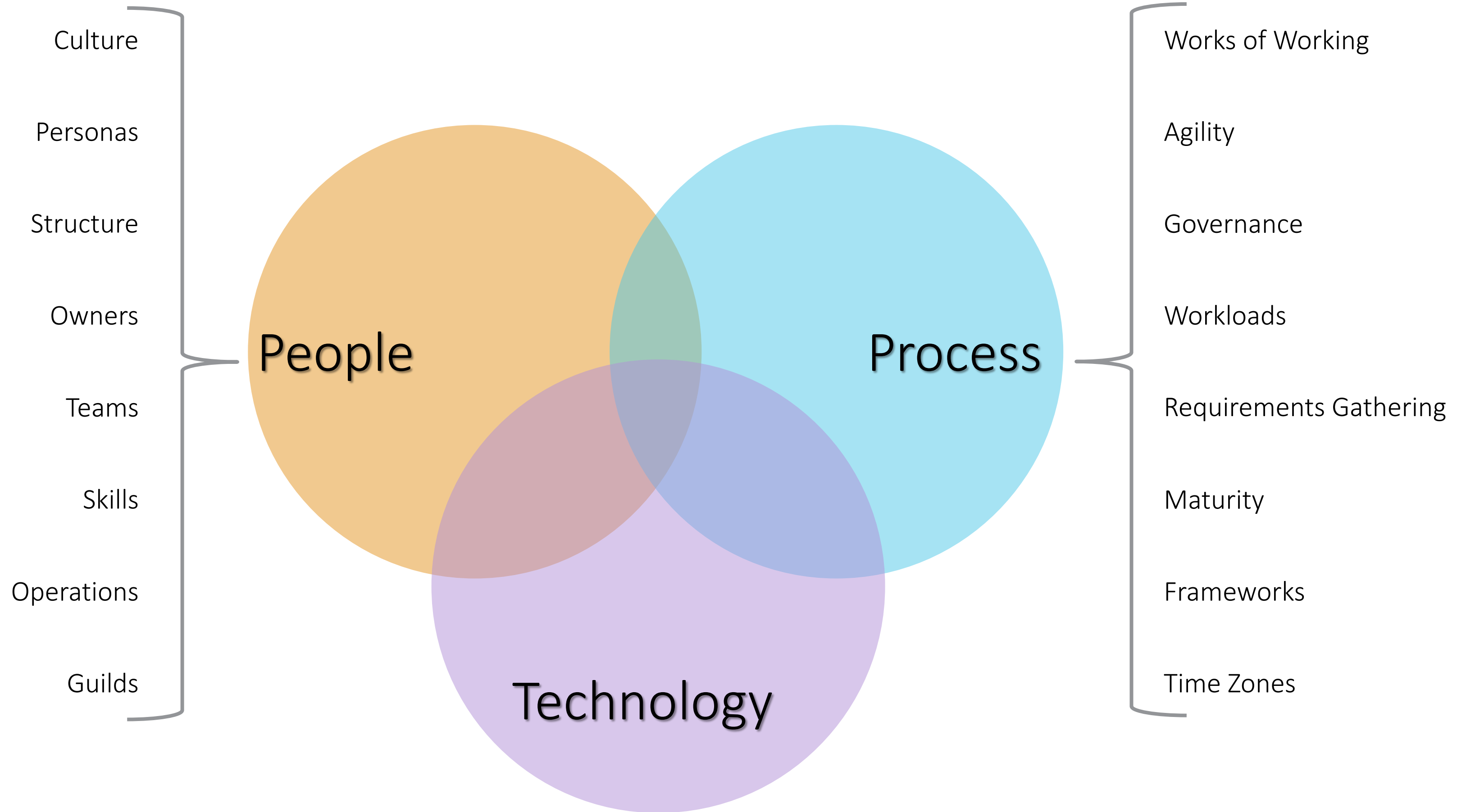
Data Mesh



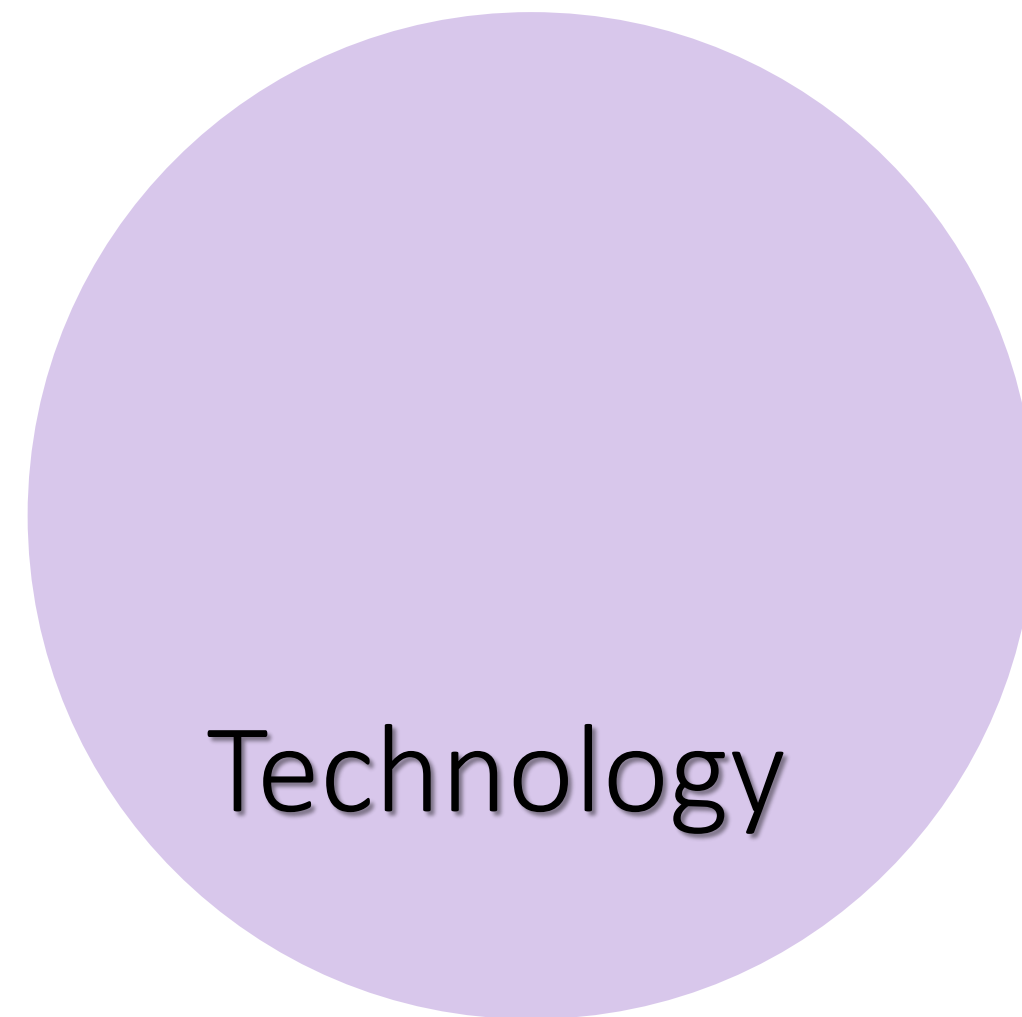
Cloud Formations



Data Mesh – What is it about?



Data Mesh – What is it about?



Data Mesh – What is it about?



Zhamak Dehghani

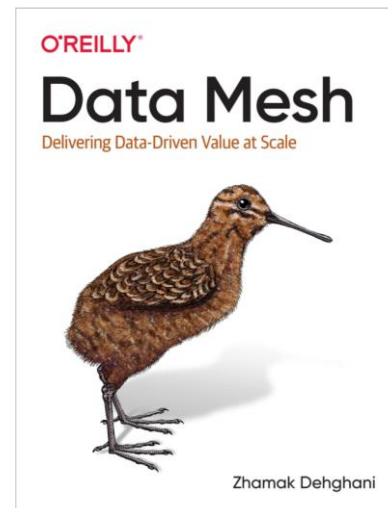
@zhamakd



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

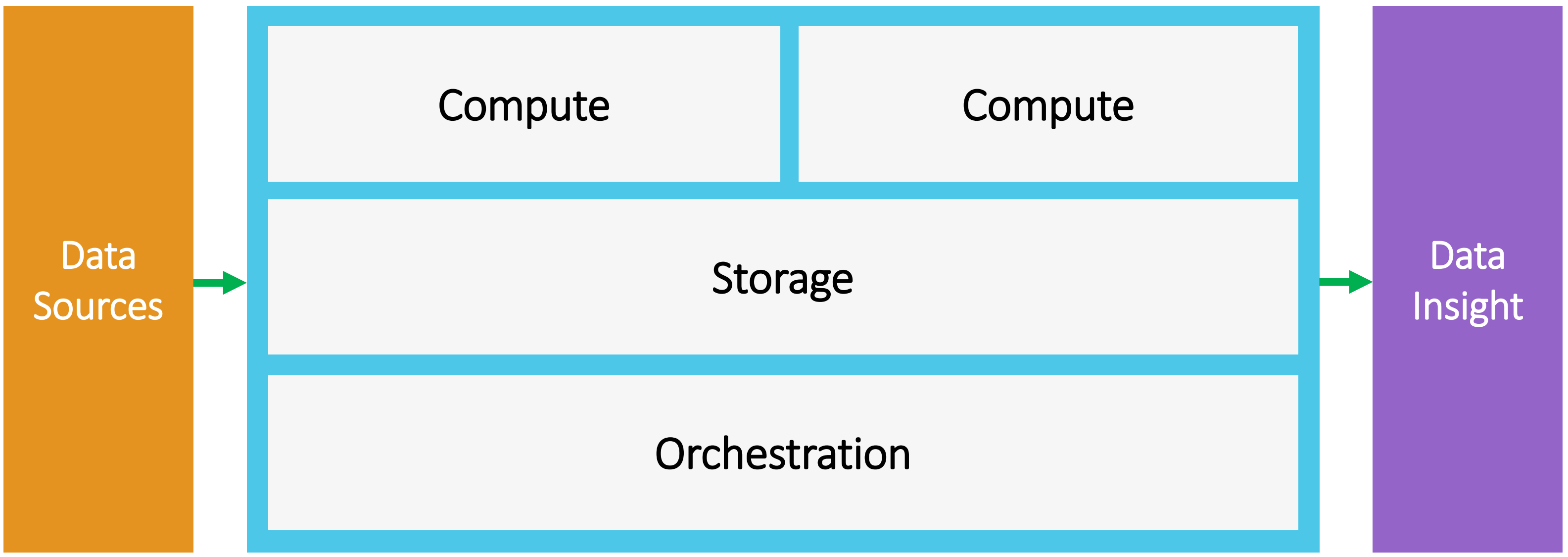
ISBN-10
1492092398

ISBN-13
978-1492092391



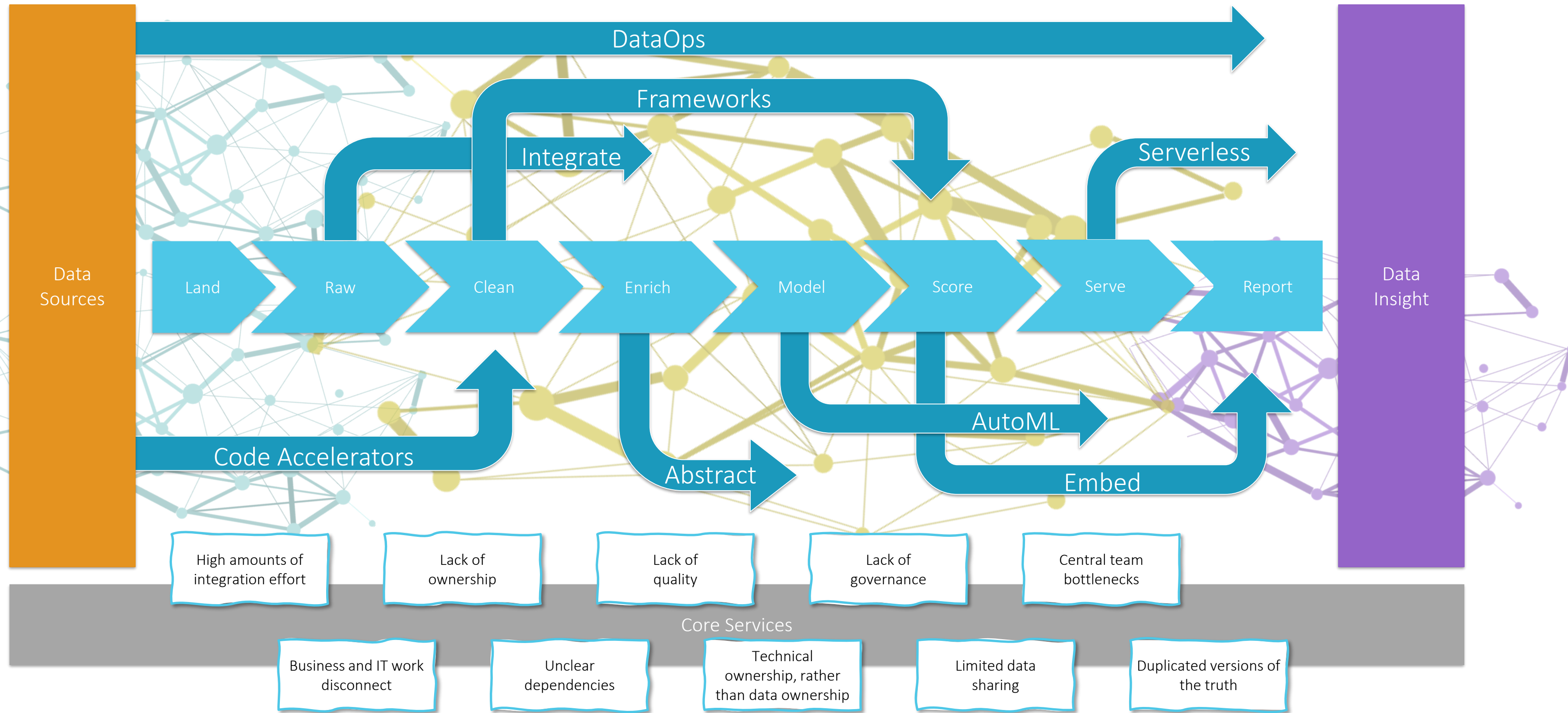
1. Domain-oriented decentralised data ownership and architecture.
2. Data as a product.
3. Self-serve data infrastructure as a platform.
4. Federated computational governance.

My First Reference Architecture

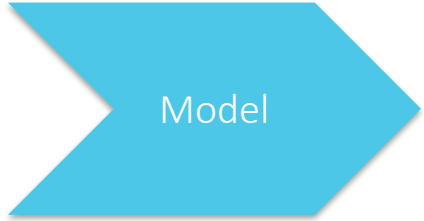


Data Mesh – Why should we build it?

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

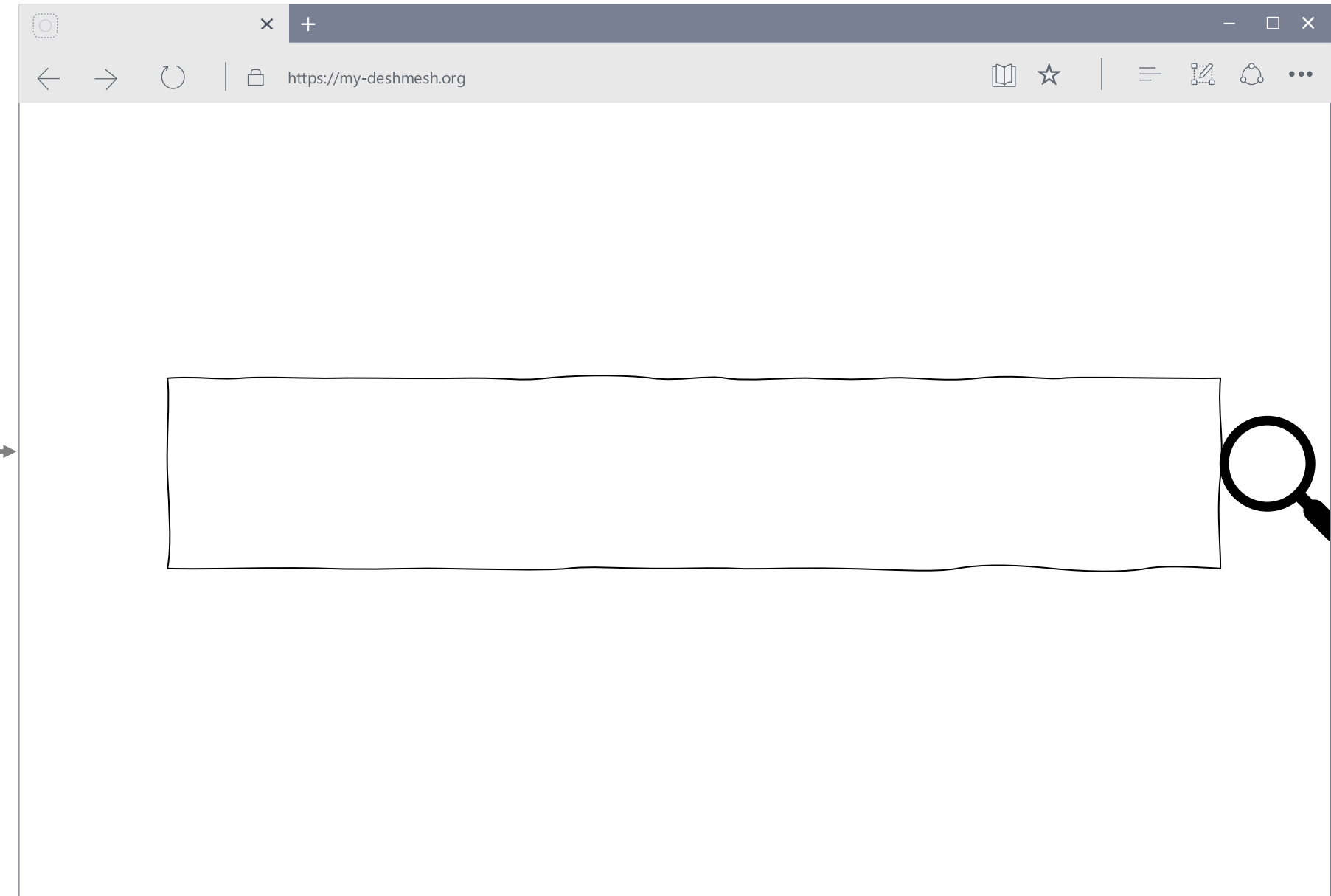
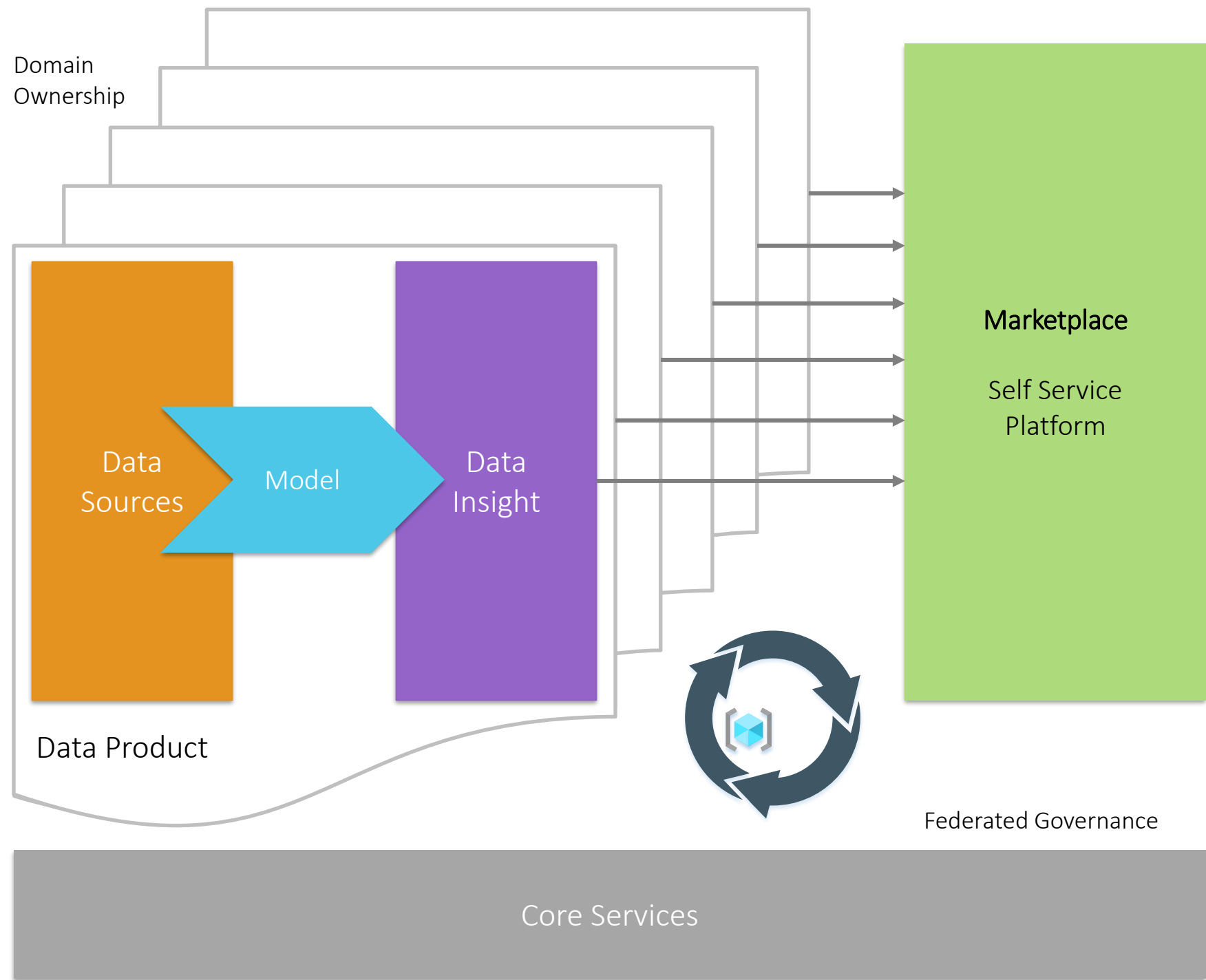


Data Mesh – Why should we build it?



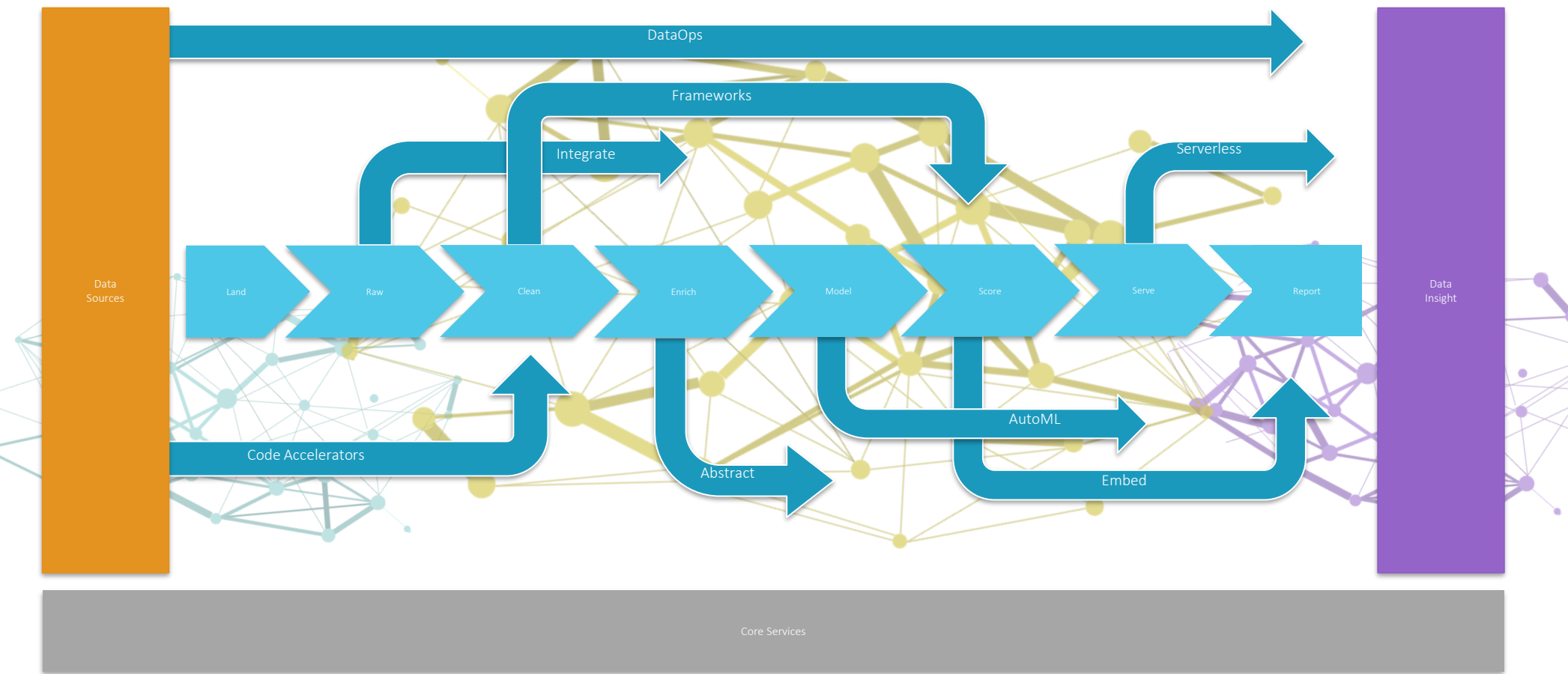
Data Mesh – Why should we build it?

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



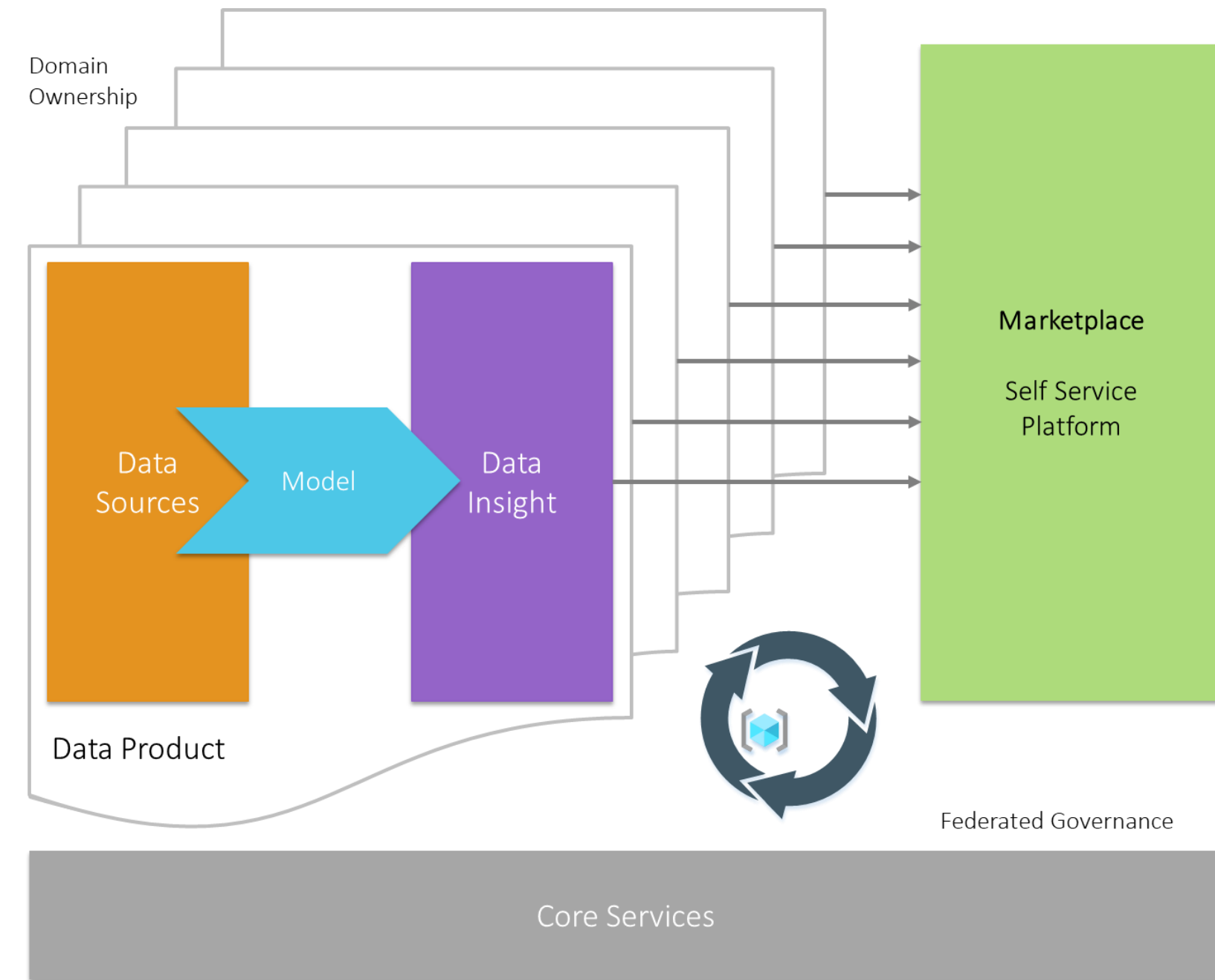
Data Mesh – Why should we build it? A: Time to Insight

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



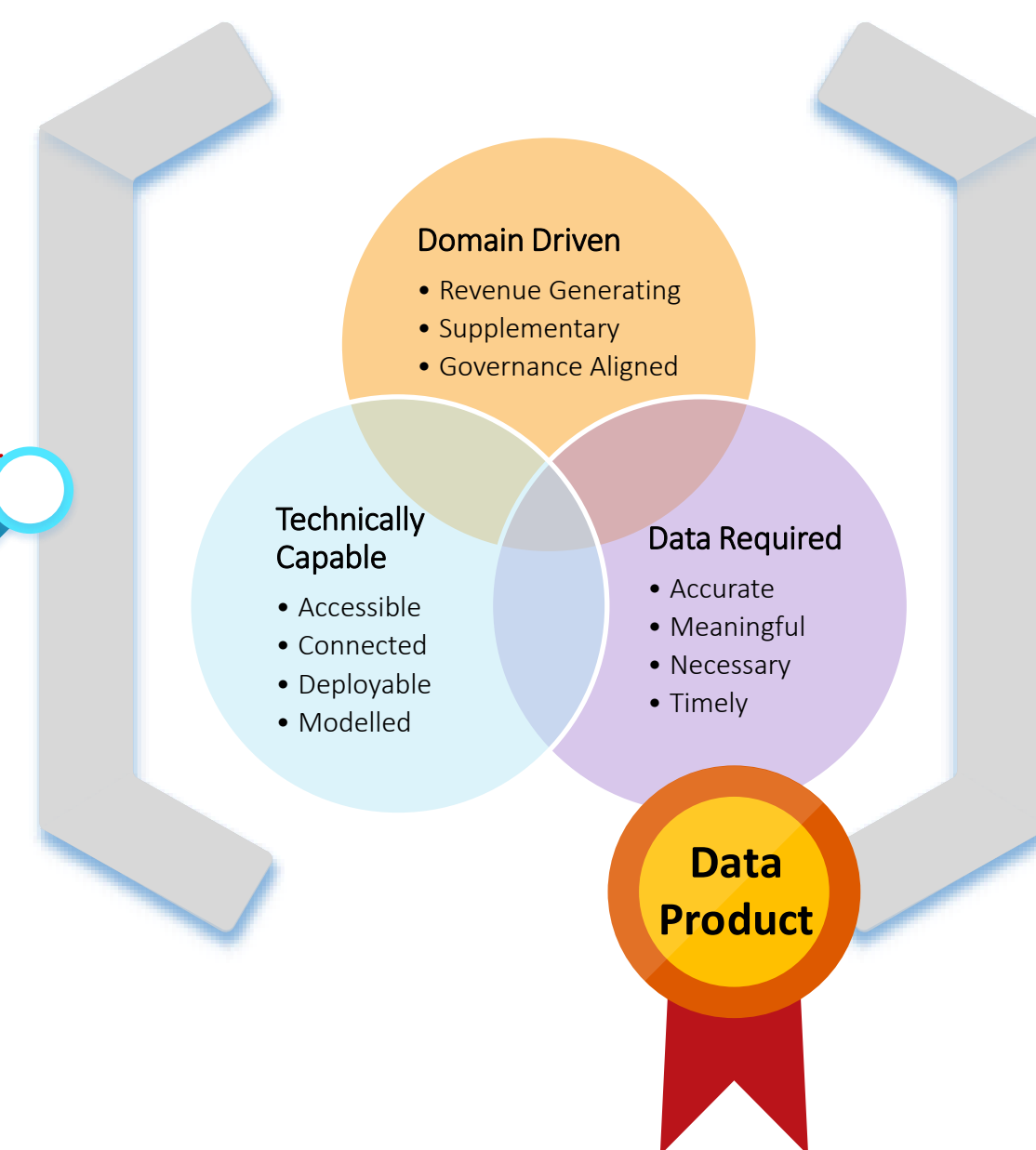
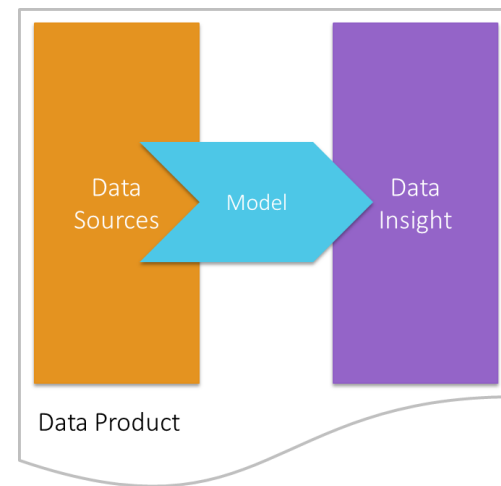
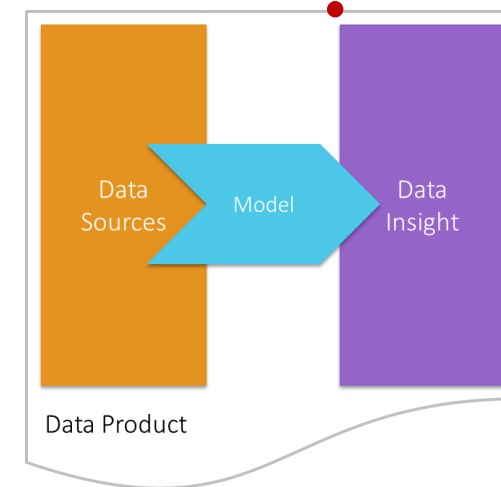
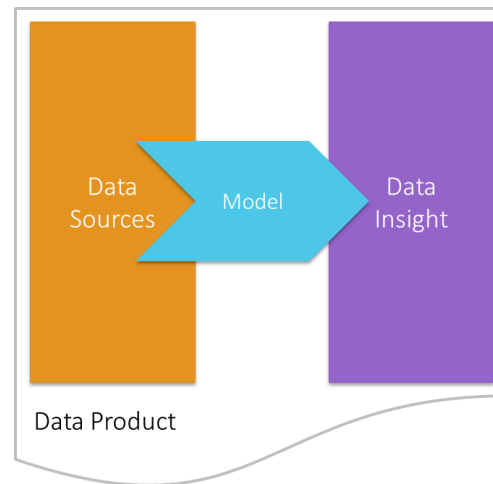
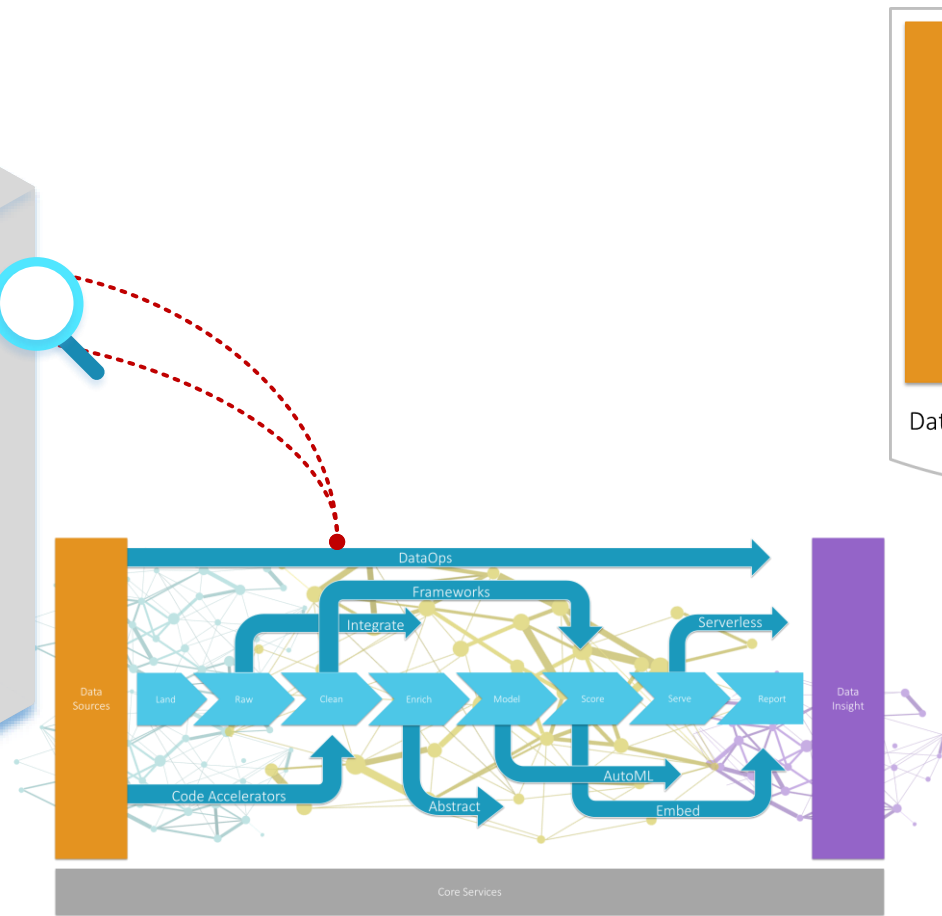
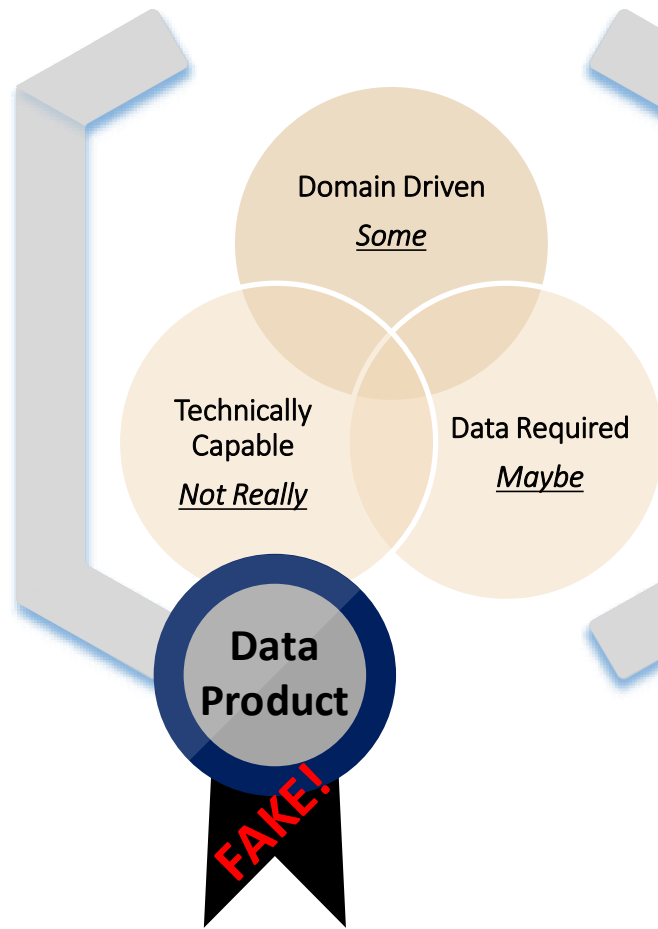
... Weeks/Months

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

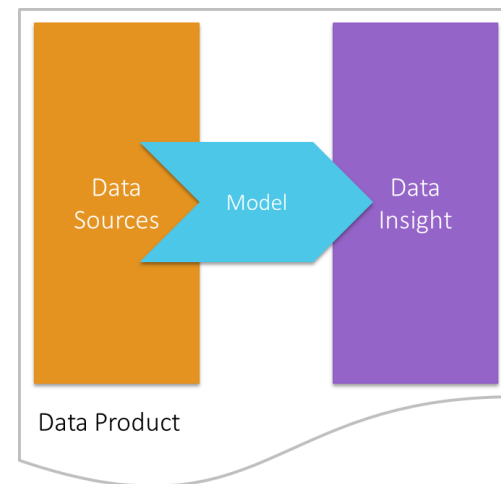
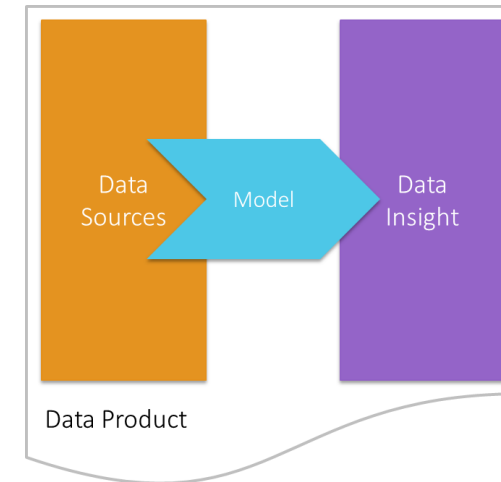
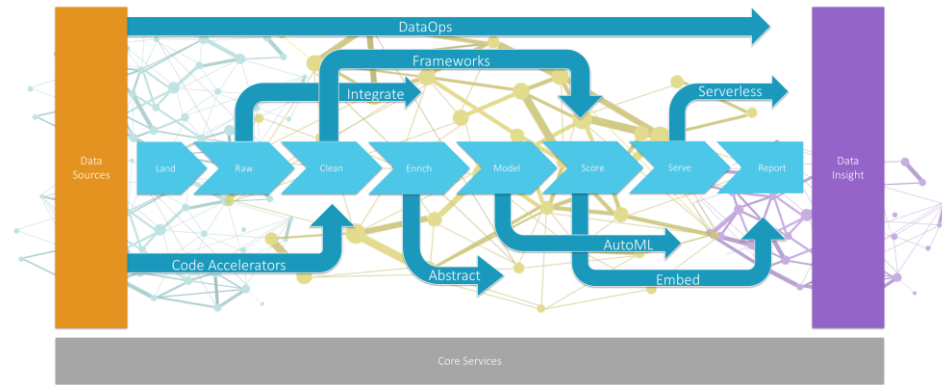
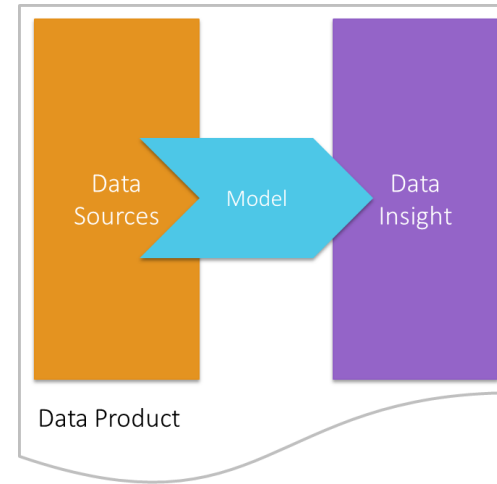


... Hours/Days

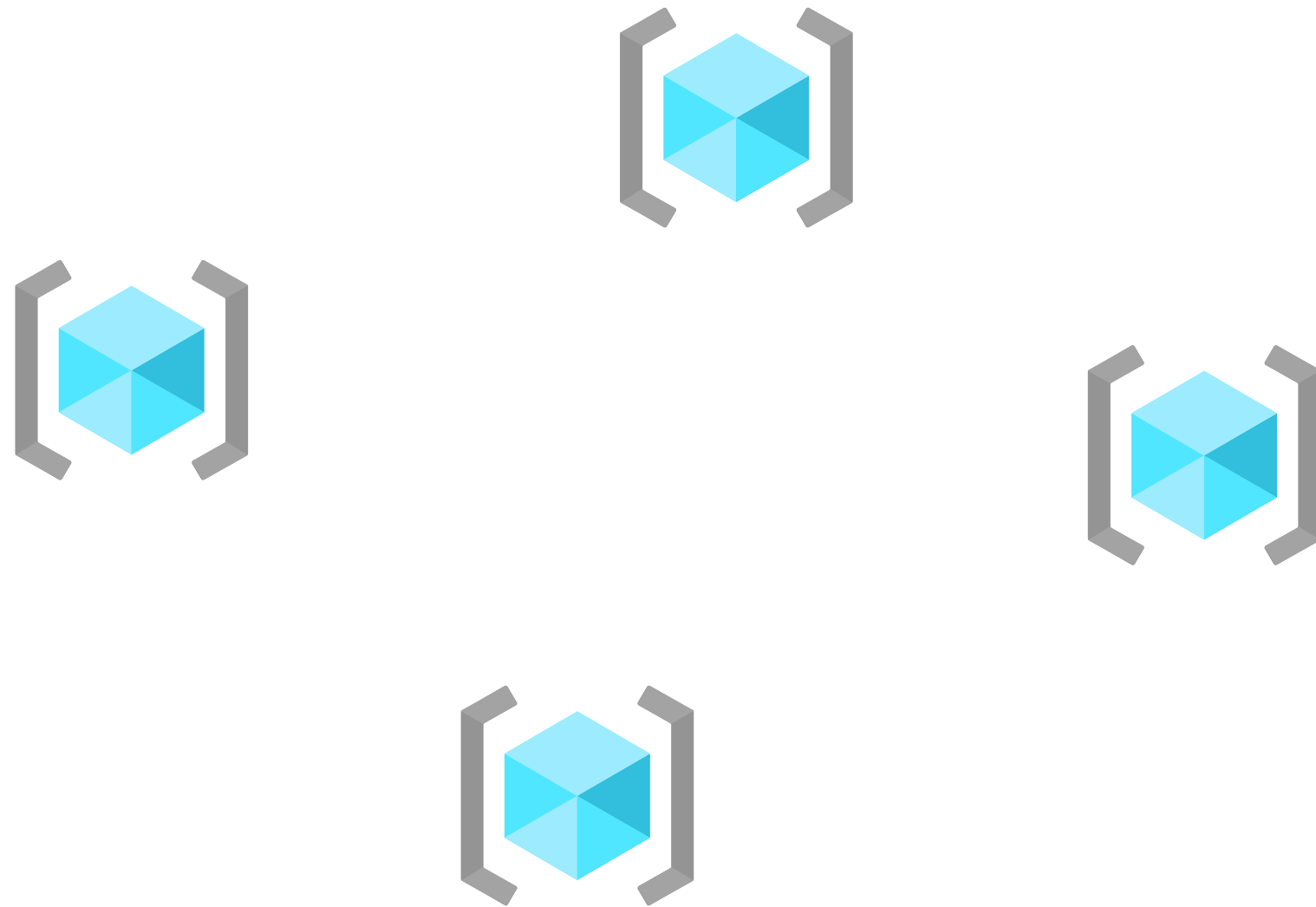
Data Mesh – Data Products



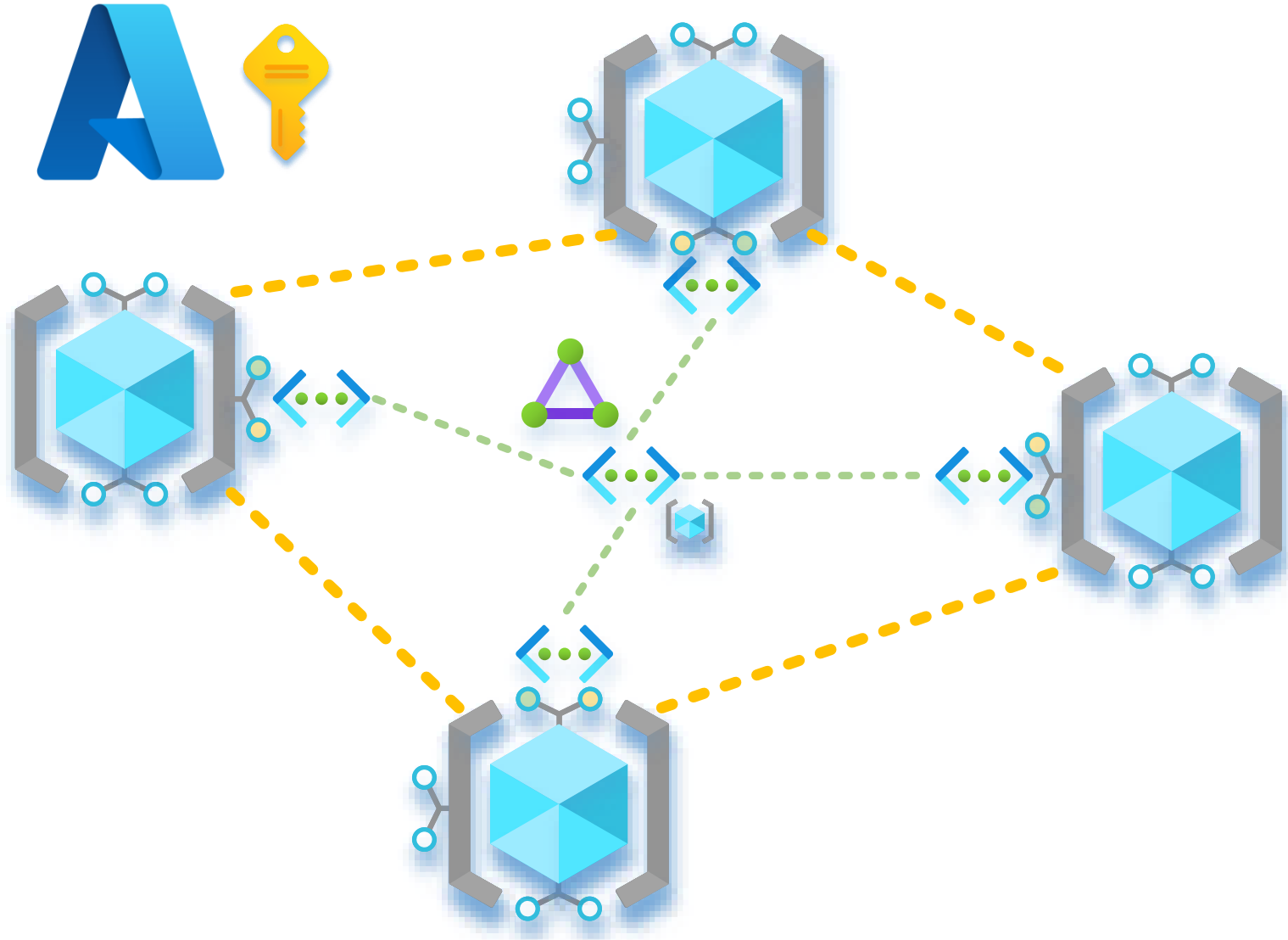
Data Mesh – Data Products in Azure



Data Mesh – Data Products in Azure



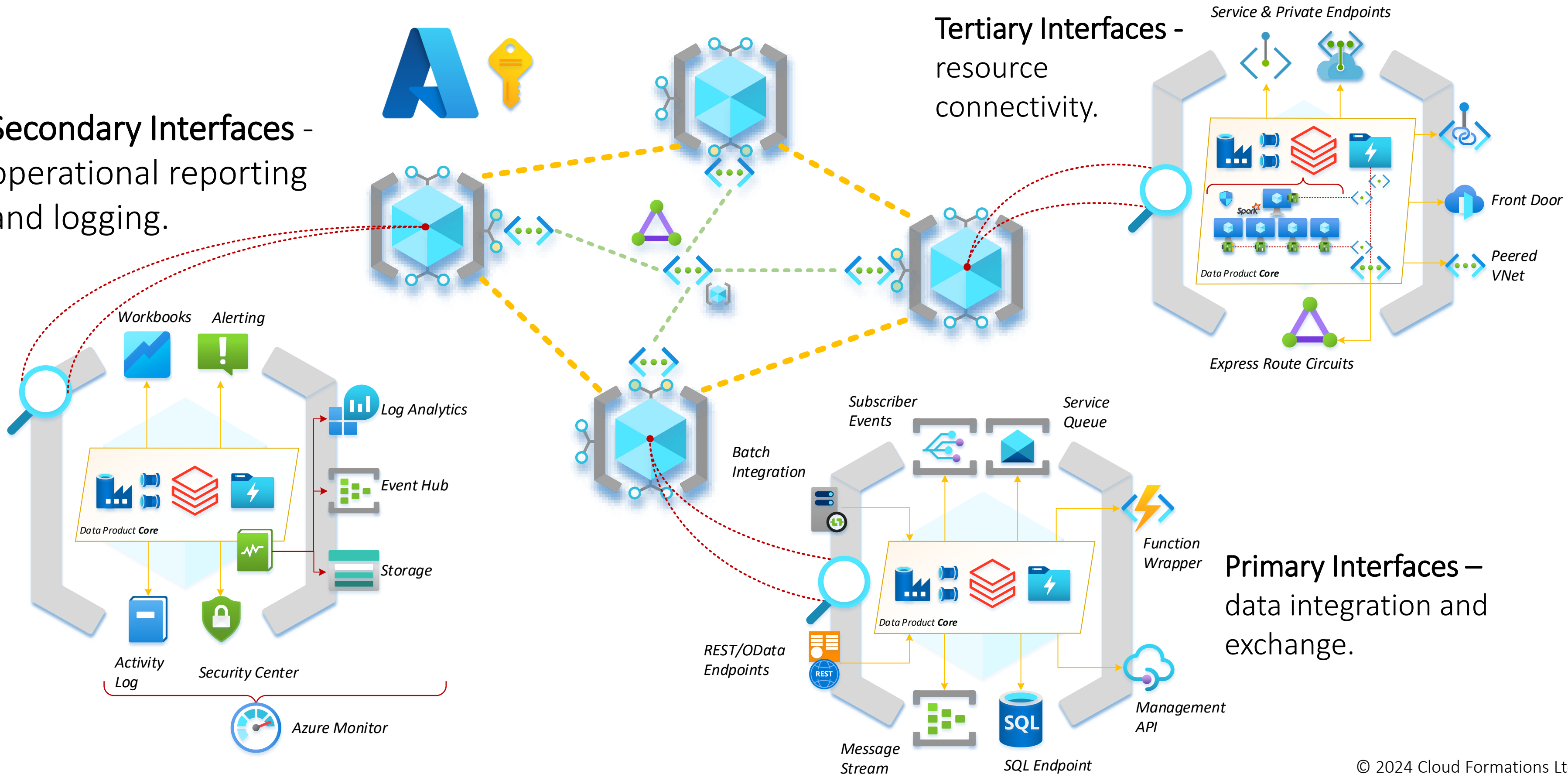
Data Mesh – Data Products in Azure with Interfaces



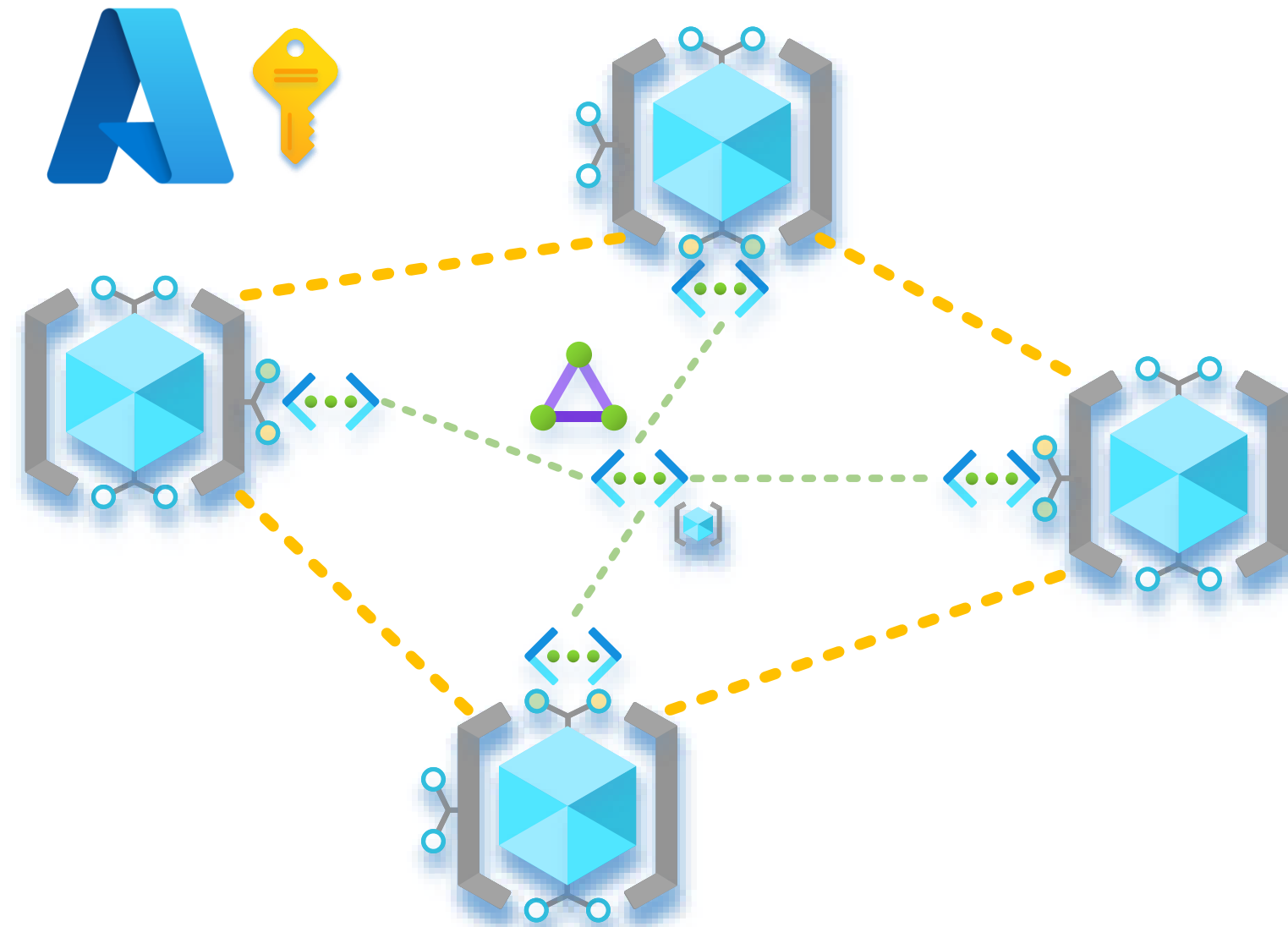
Data Mesh – Data Products in Azure with Interfaces



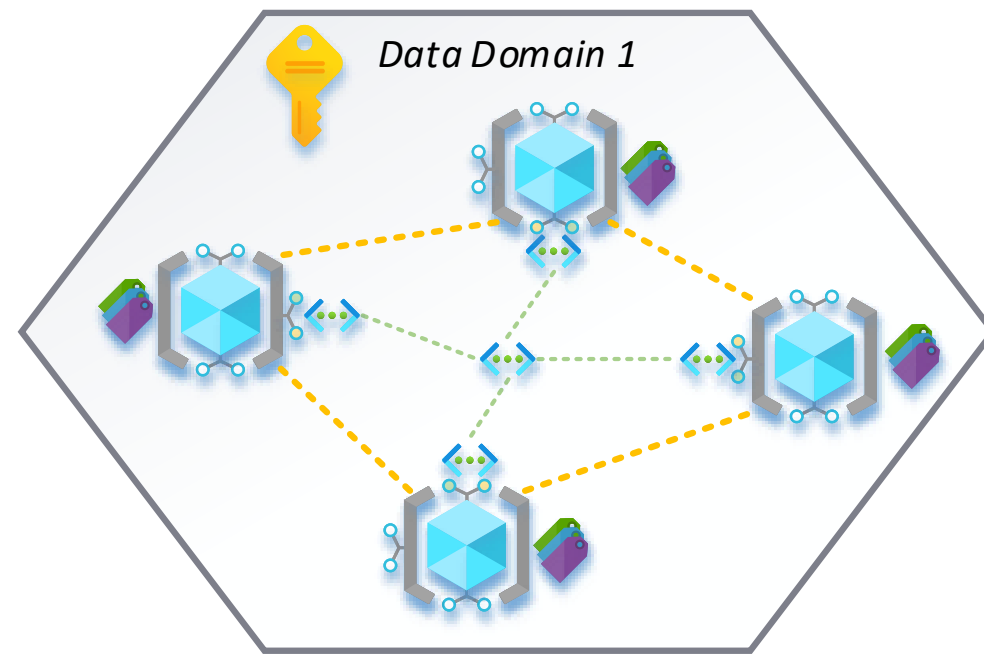
Secondary Interfaces -
operational reporting
and logging.



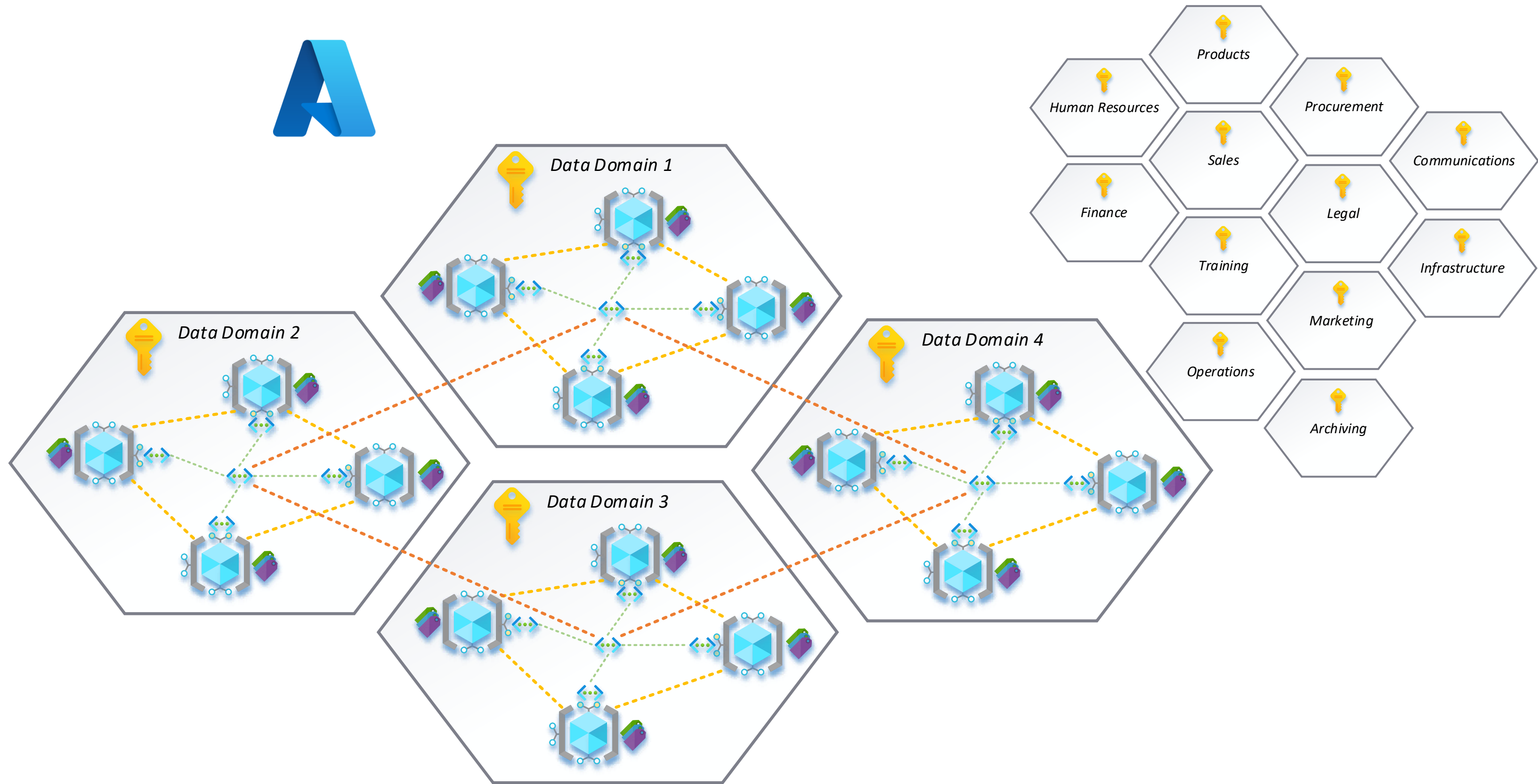
Data Mesh – Data Domains in Azure



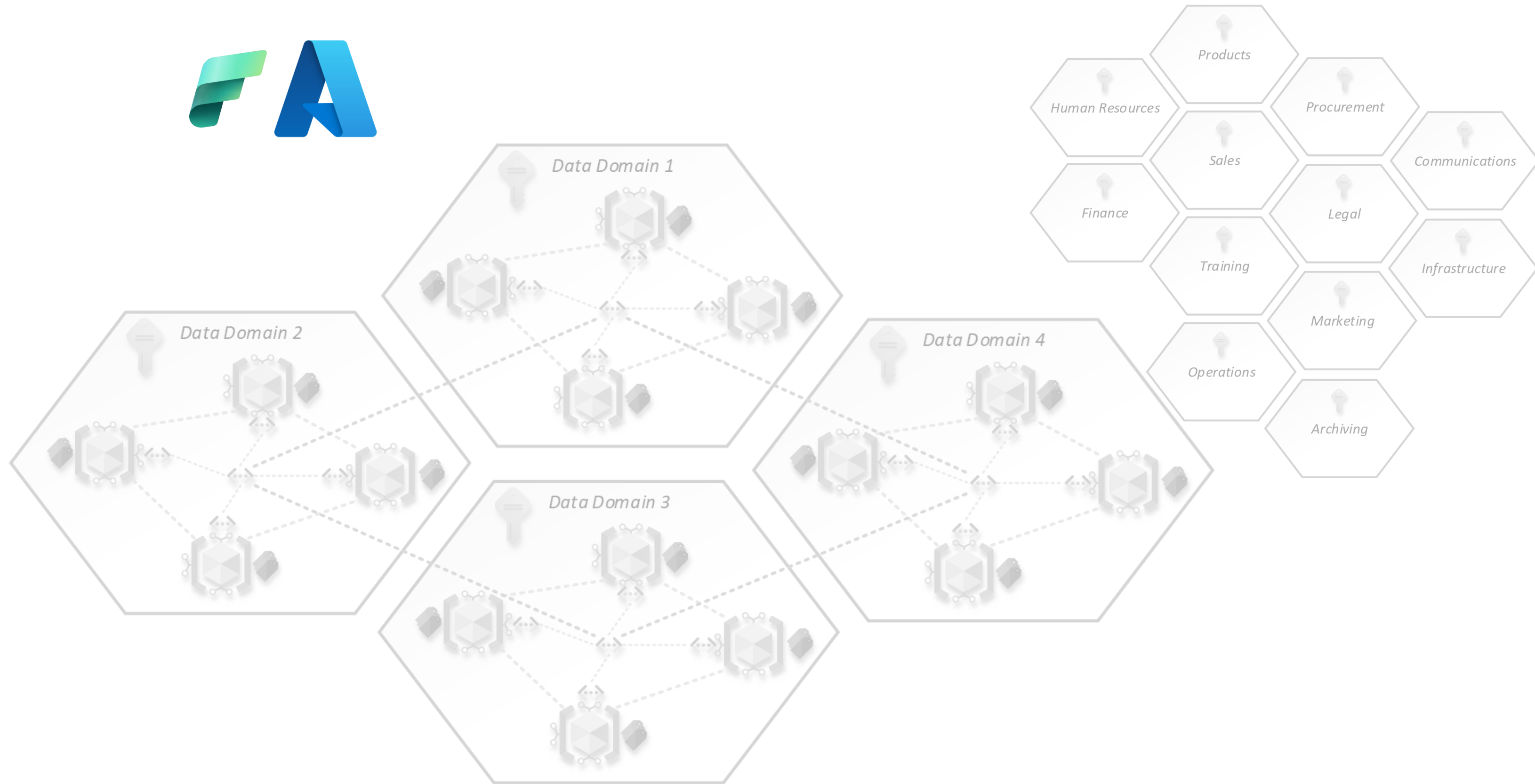
Data Mesh – Data Domains in Azure



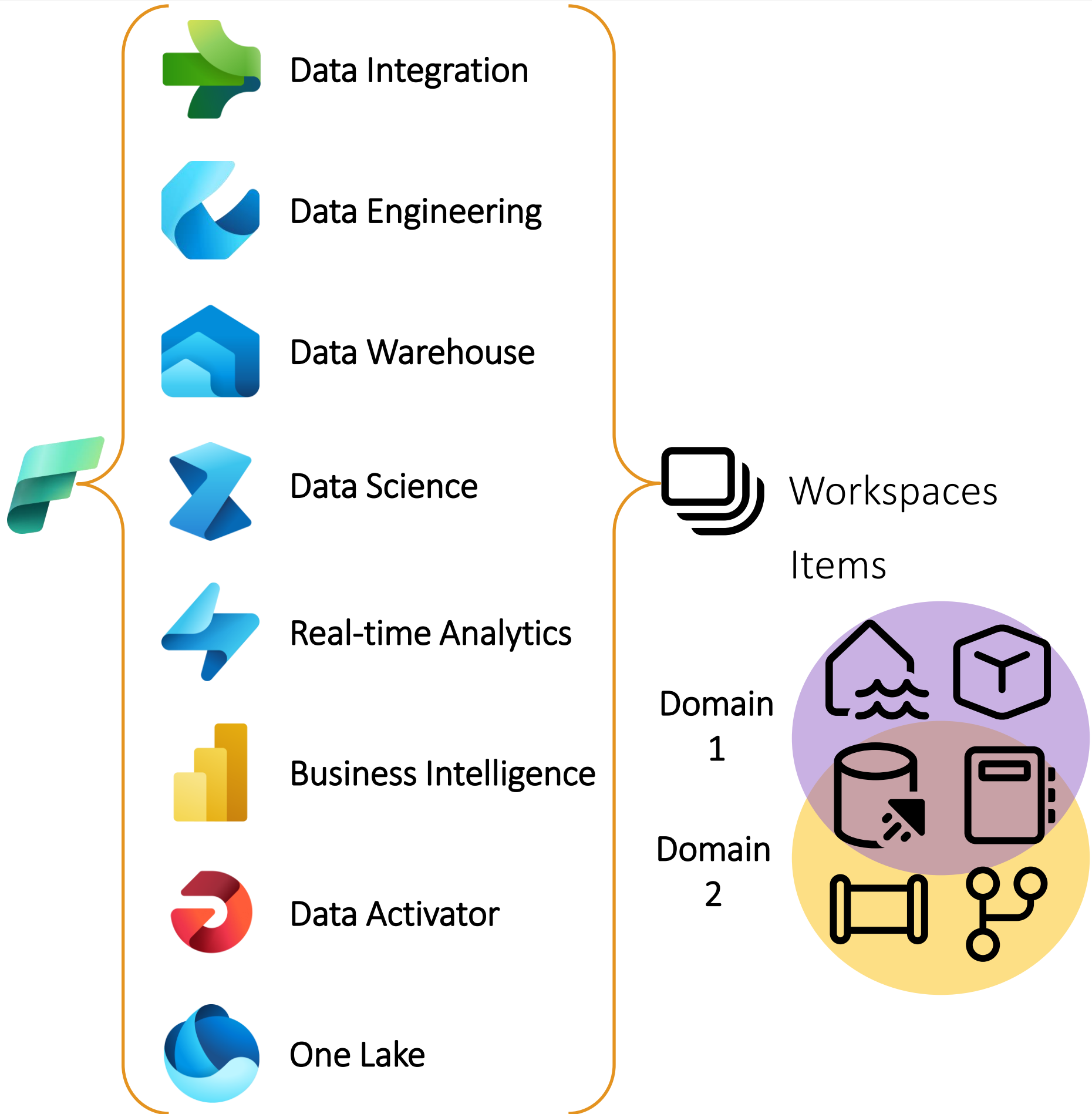
Data Mesh – Data Domains in Azure



Data Mesh – Data Domains in Fabric



Data Mesh – Data Domains in Fabric

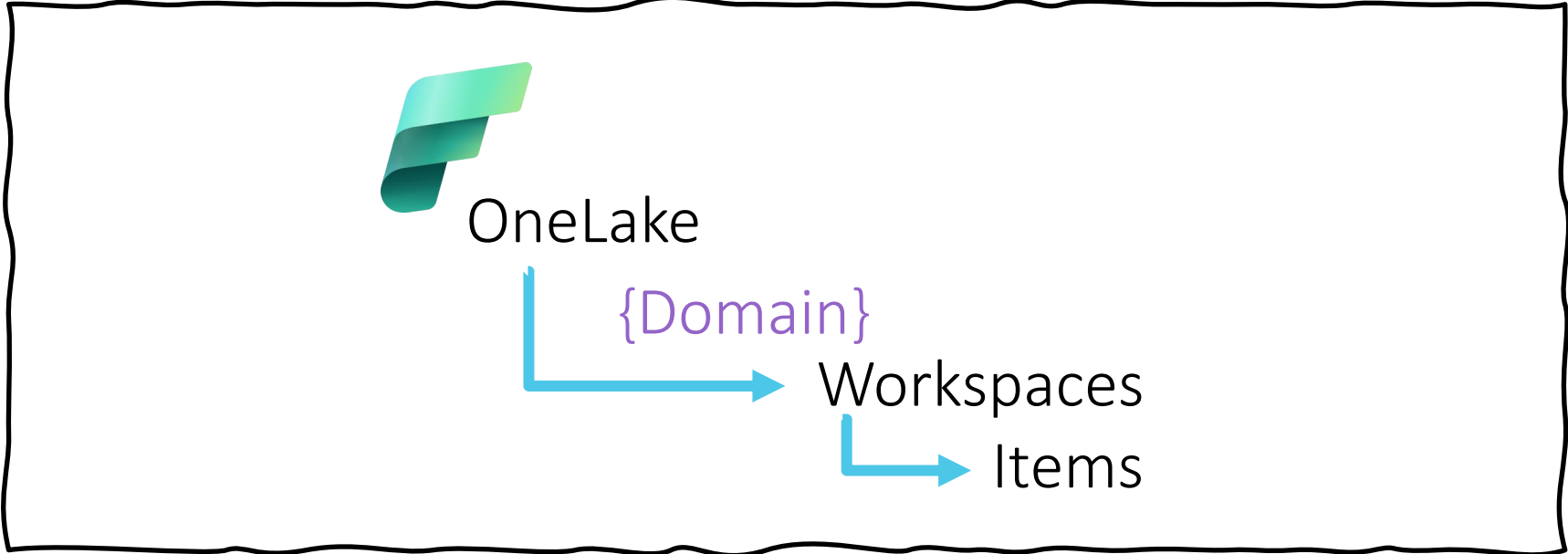


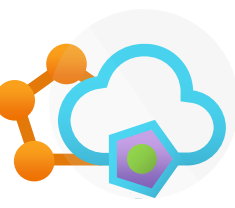
.... To meet this challenge, organizations are shifting from traditional IT centric data architectures, where the data is governed and managed centrally, to more federated models organized according to business needs. This federated data architecture is called data mesh. A data mesh is a decentralized 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>





Architecture Agenda:

δ

λ

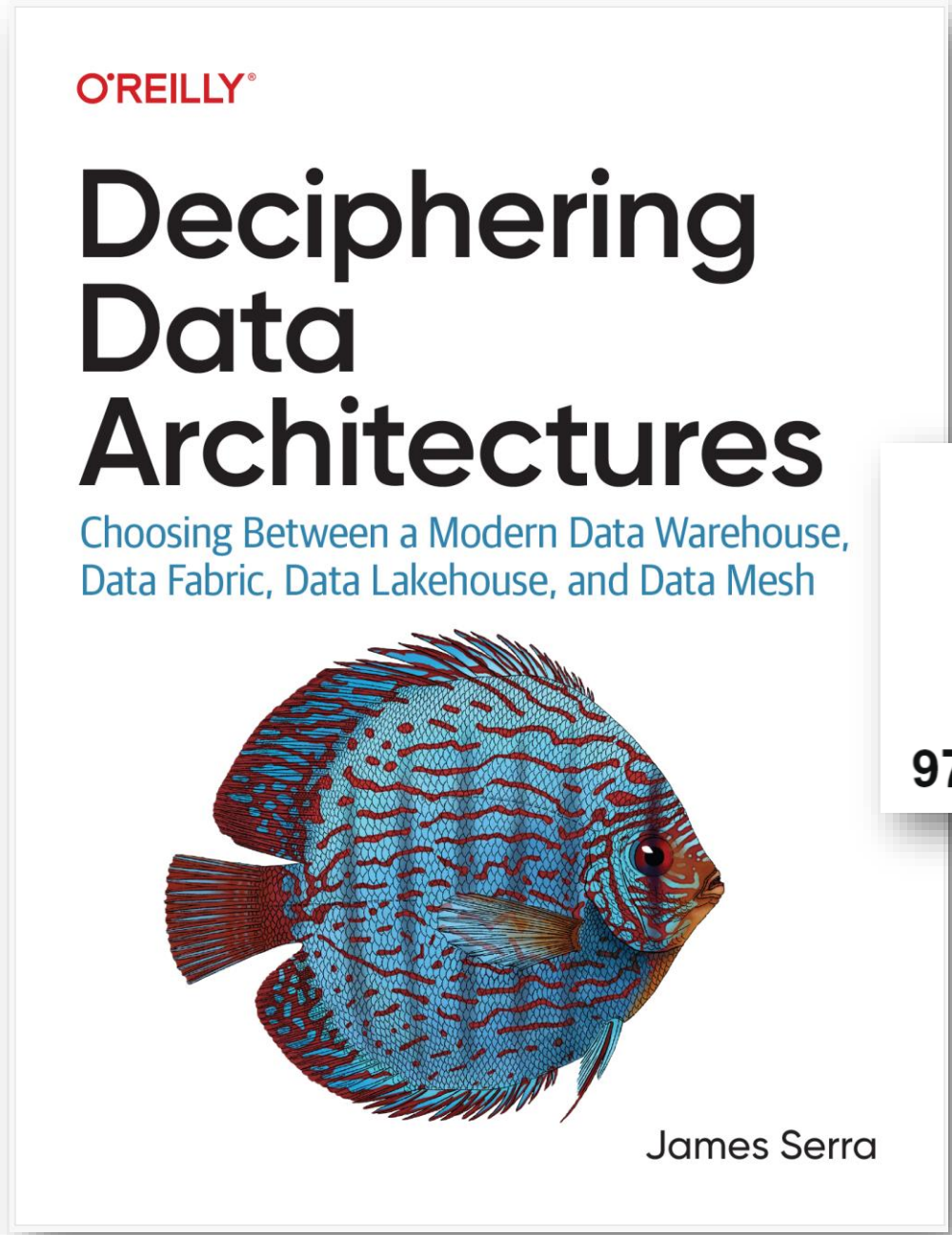
K





Final thoughts from me...

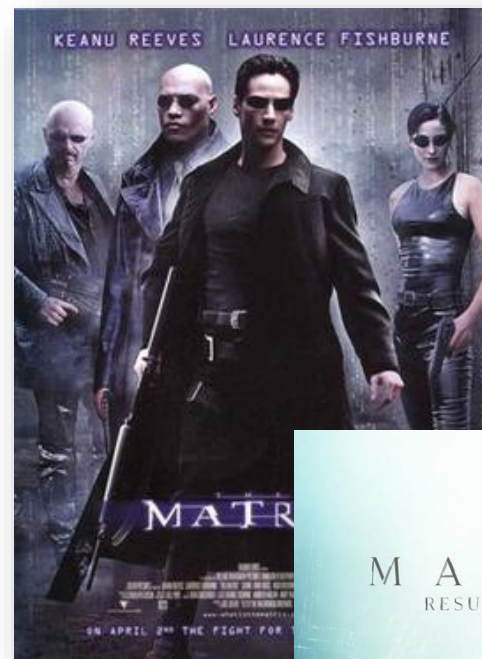
Further Reading



ISBN-13

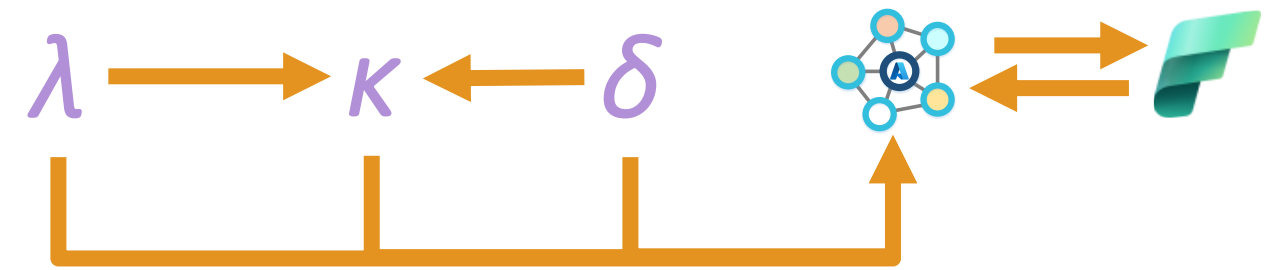
978-1098150761





An Evolution of Data Platform Architectures

Lambda, Kappa, Delta, Mesh & Fabric

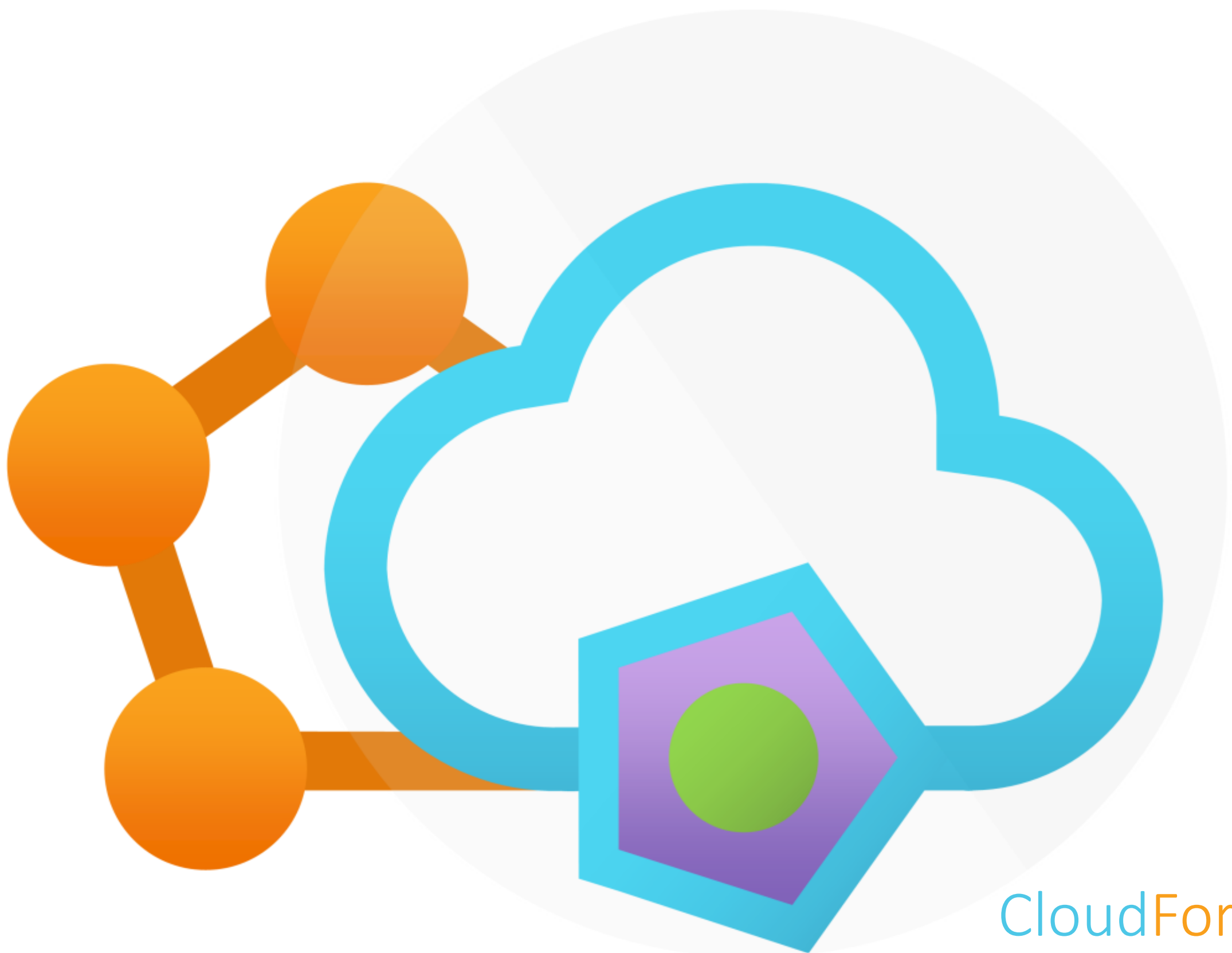


Q: What about a medallion architecture?



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

A: Yes! ✓



Thank You



- mrpaulandrew.com
- paul.andrew@cloudformations.org
- [In/mrpaulandrew](https://www.linkedin.com/in/mrpaulandrew)
- [@mrpaulandrew](https://twitter.com/mrpaulandrew)



- <https://cloudformations.org>
- contactus@cloudformations.org
- [In/CloudFormations](https://www.linkedin.com/company/cloudformations)
- [@CloudFormsLtd](https://twitter.com/CloudFormsLtd)
- [CloudFormationsLtd](https://www.facebook.com/CloudFormationsLtd)

[CloudFormations.org/Community-Content](https://cloudformations.org/Community-Content)

