

An Engineers Guide To Real-time Data Analytics

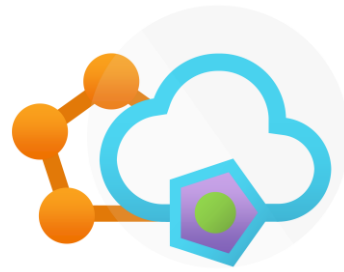


Paul Andrew
Technical Architect



Cloud Formations

Paul Andrew



Co-Founder & Director
Chief Technology Officer



/mrpaulandrew



@mrpaulandrew



In/mrpaulandrew

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

SQL Server 2000





Add a slide about the advantages of combining real-time data with batch data in a lambda architecture.

I completed some of your request, but I'll need more practice before I can do everything in it. What else can I help you with?



By using a lambda architecture, we can take advantage of both real-time and batch data to gain a 360-degree view of our data. Real-time data allows us to monitor system performance and detect issues as they arise, while batch data provides a comprehensive view of our data. By combining these two types of data in a lambda architecture, we can gain a 360-degree view of our data.

f
e
a
a
e
S

An Engineers Guide To Real-time Data Analytics

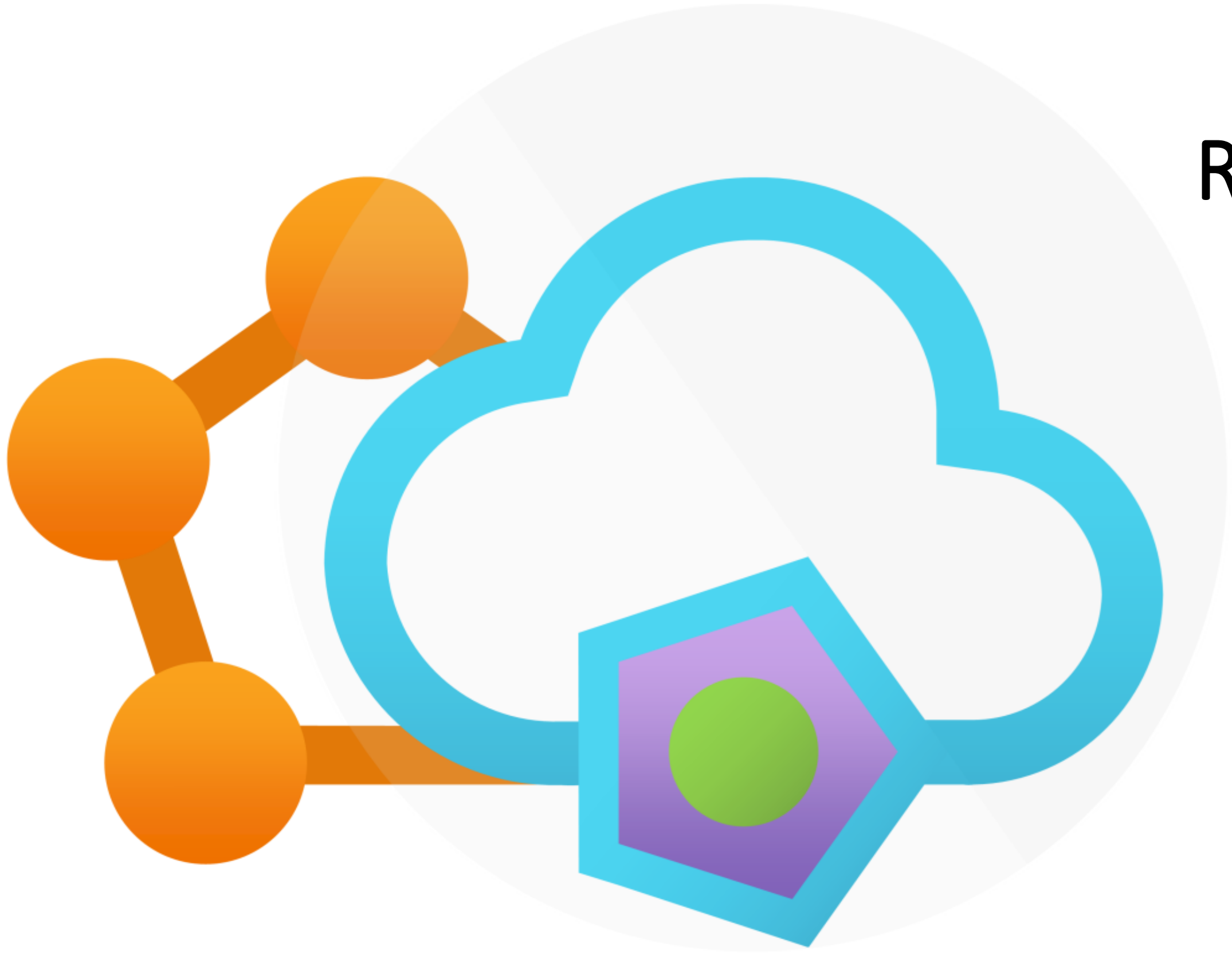


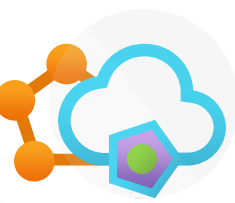
Paul Andrew

Technical Architect



Cloud Formations





Agenda:

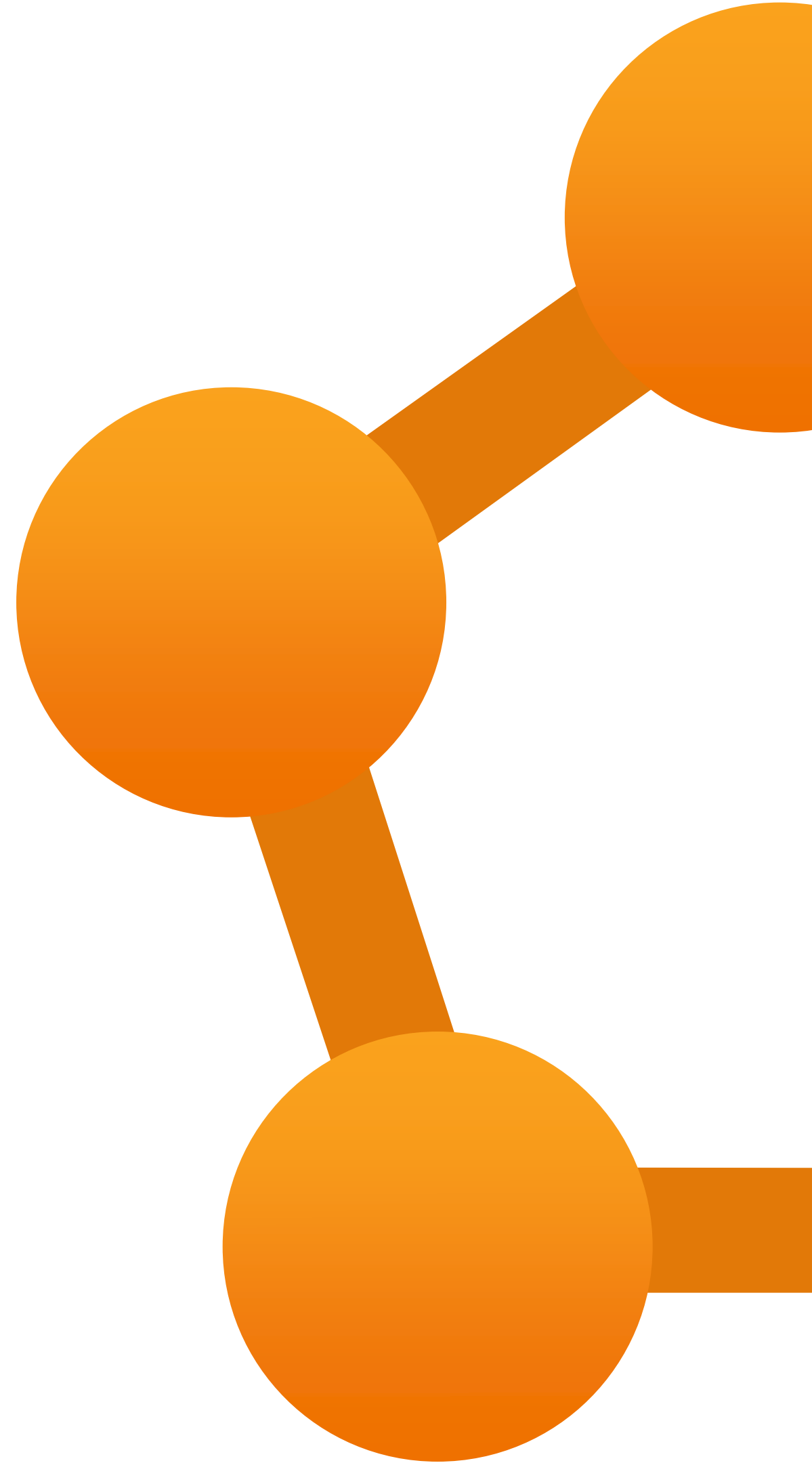
Theory | Tooling | Patterns | Concepts

Practical

Conclusions

An Engineers Guide To ~~Real-time~~ Data Analytics

Cloud Formations



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.”

Volume

Velocity

Variety

Veracity

Value

- Buck Woody

@BuckWoodyMSFT

What is the goal of our data solutions?



Cloud Formations - Knowledge Transfer & Training

Data Sources

Data Insight

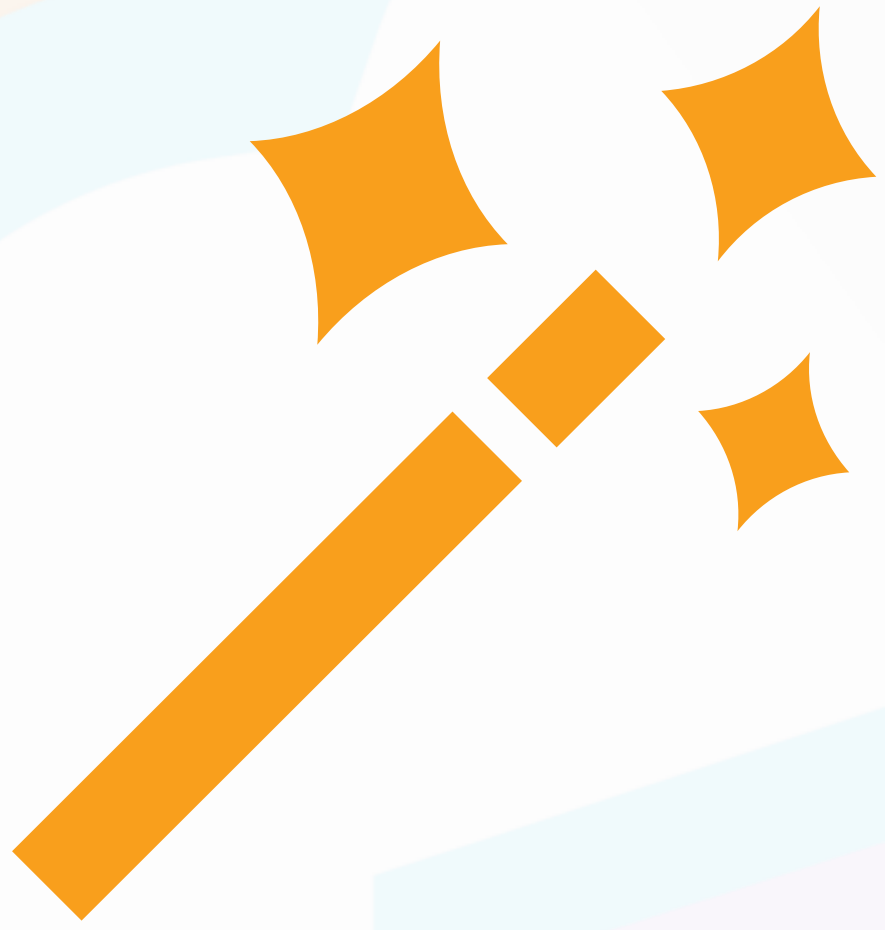
Data = Information = Knowledge = Power

How do we deliver our data insights?



Cloud Formations - Knowledge Transfer & Training

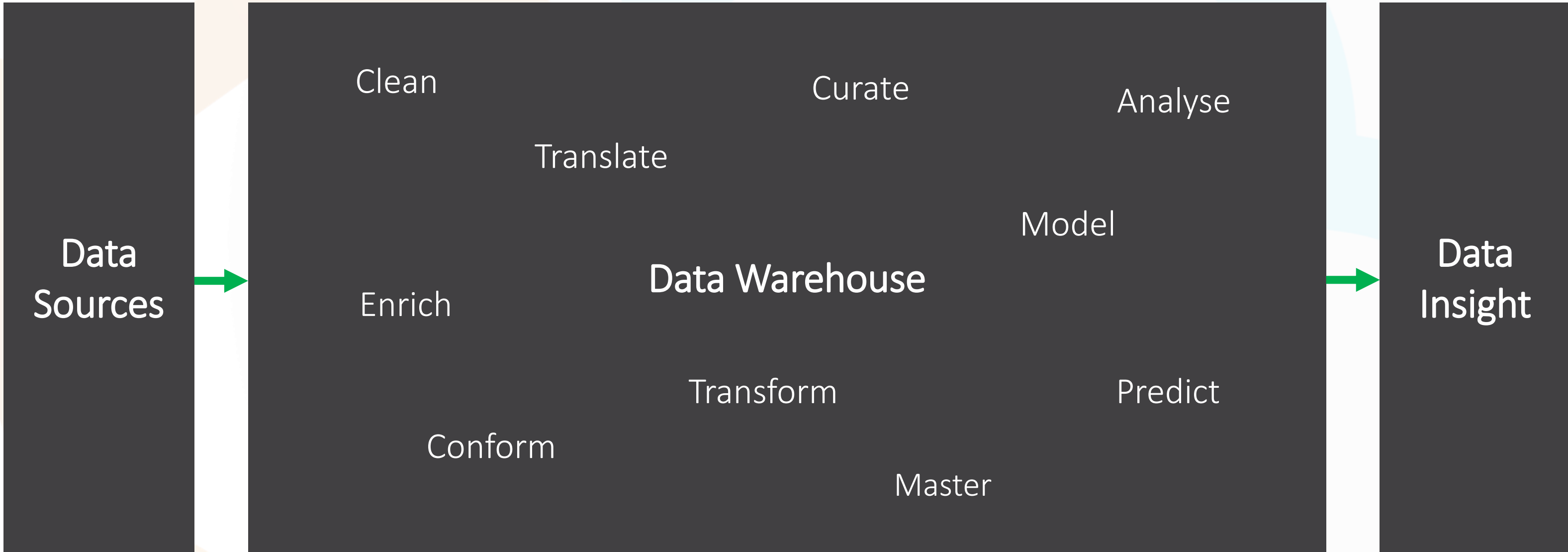
Data Sources



Data Insight

Data = Information = Knowledge = Power

How do we deliver our data insights?



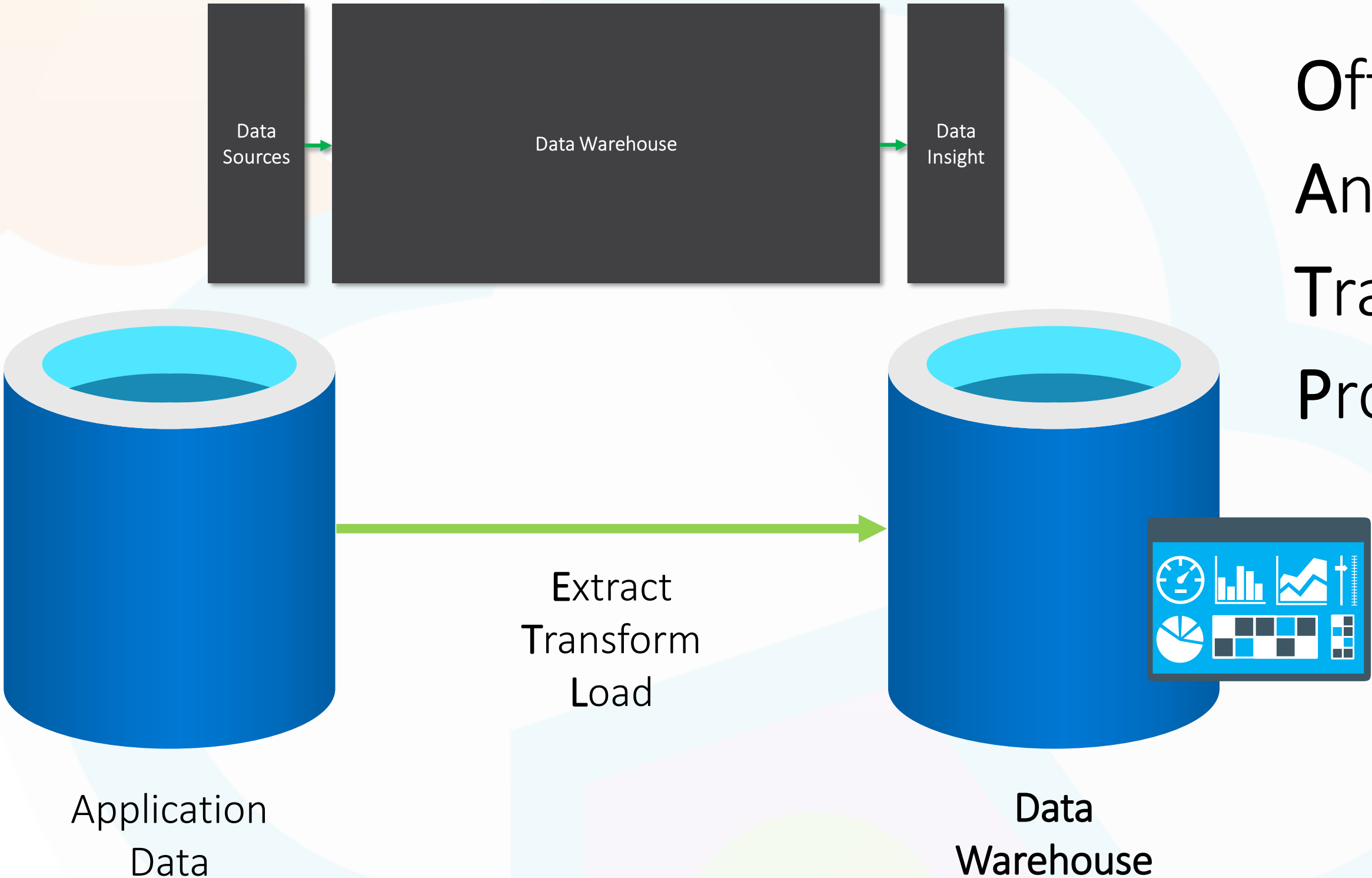
Data = Information = Knowledge = Power

Data Warehouse



Online
Line
Transactional
Processing

Offline
Analytical
Transactional
Processing



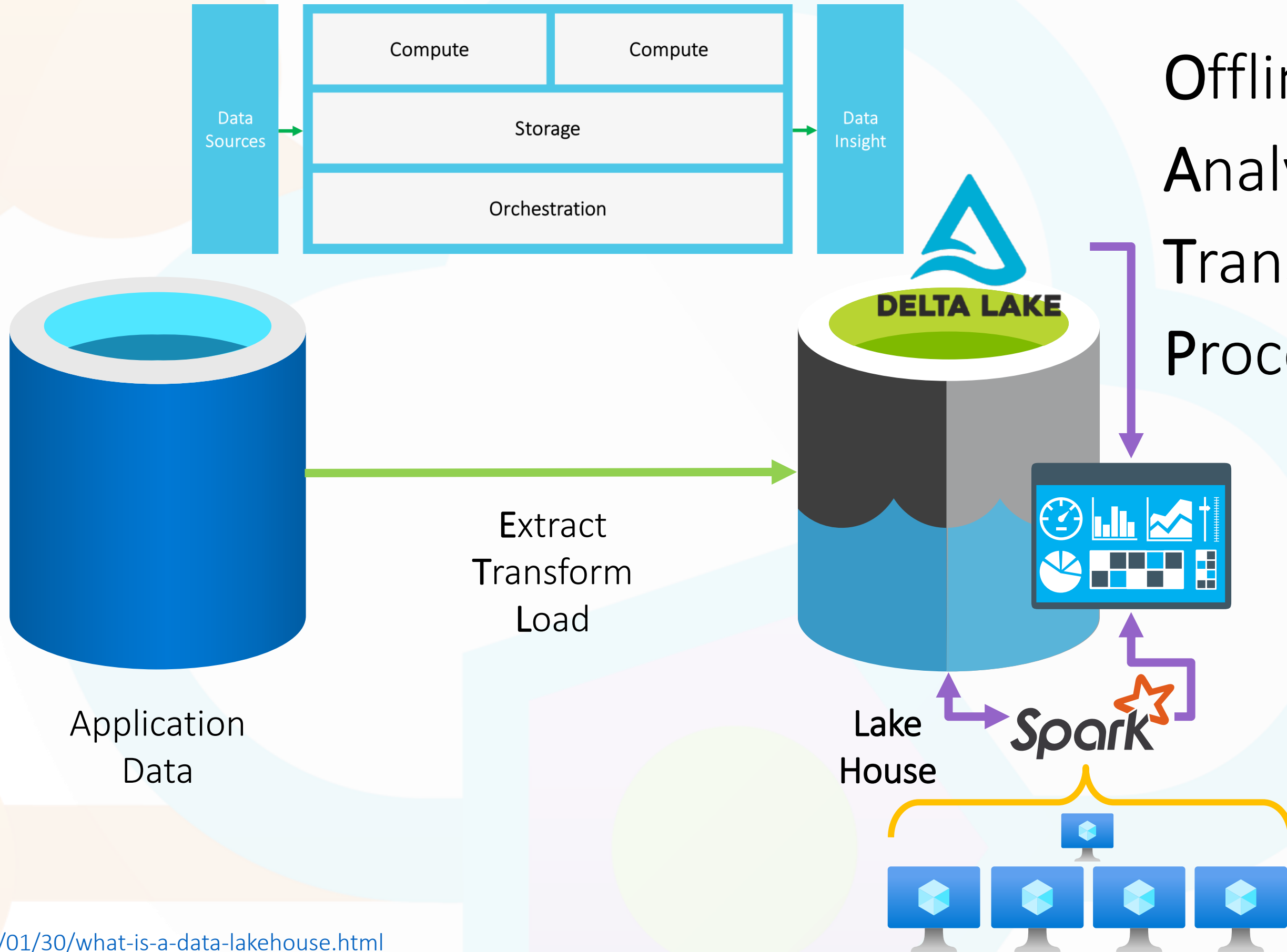
Lake House (Data-Ware-Lake-Delta-Beach-House-Lakes)



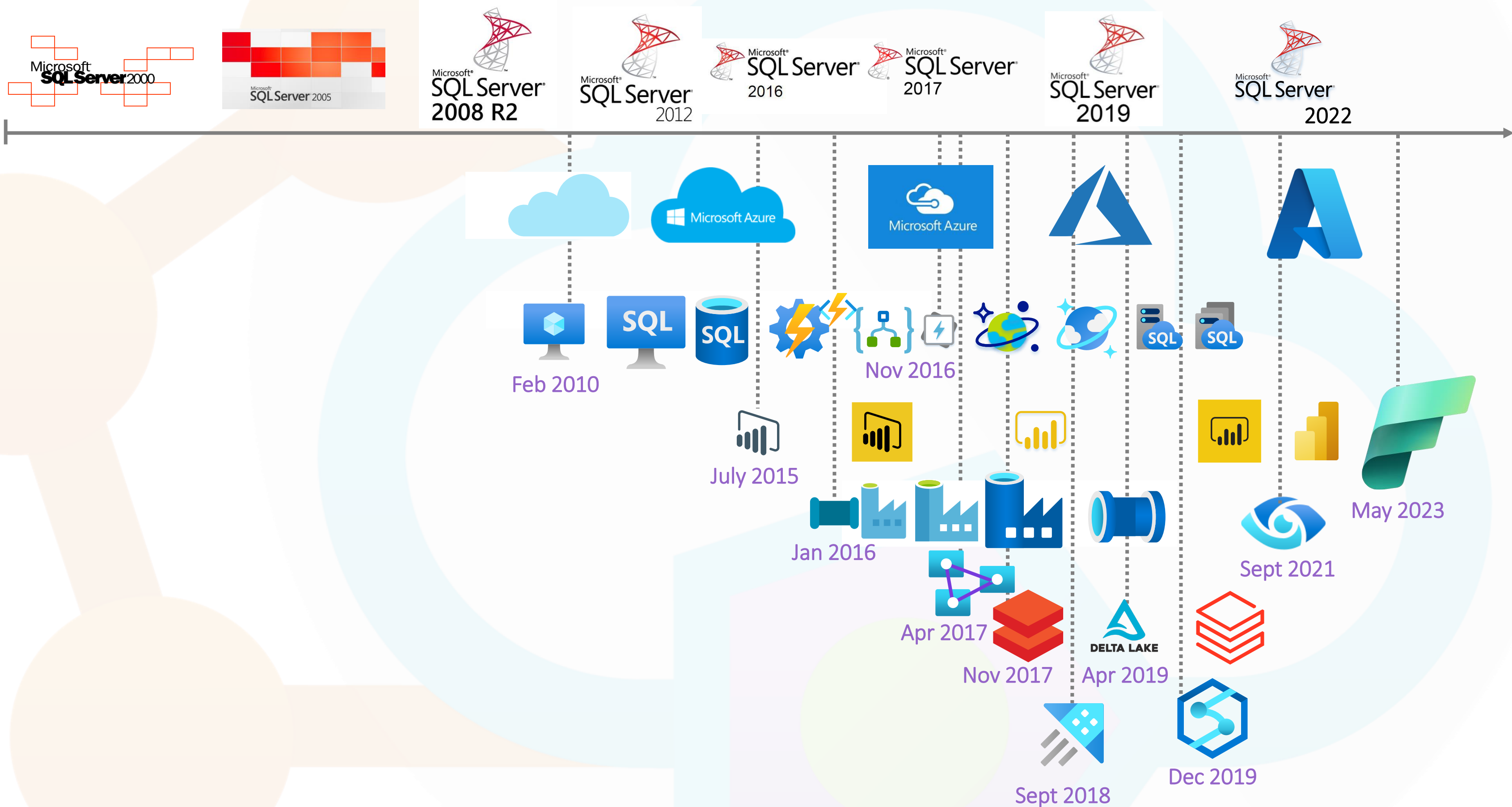
Cloud Formations - Knowledge Transfer & Training

Online
Line
Transactional
Processing

Offline
Analytical
Transactional
Processing

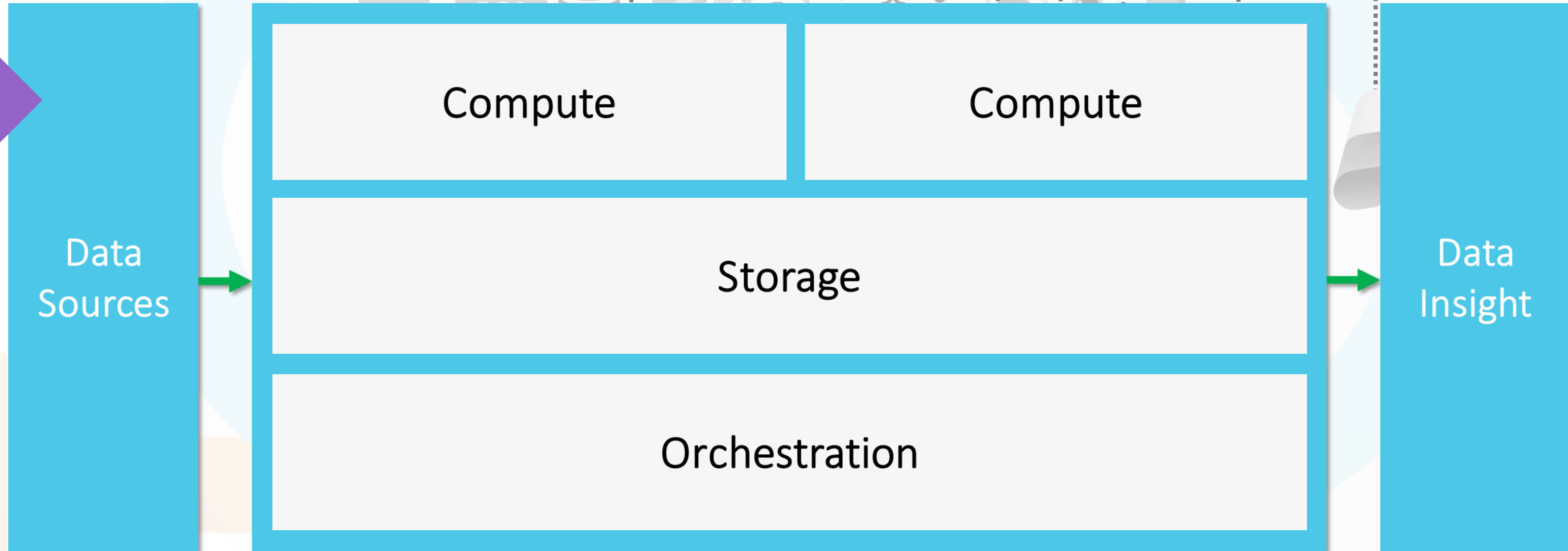
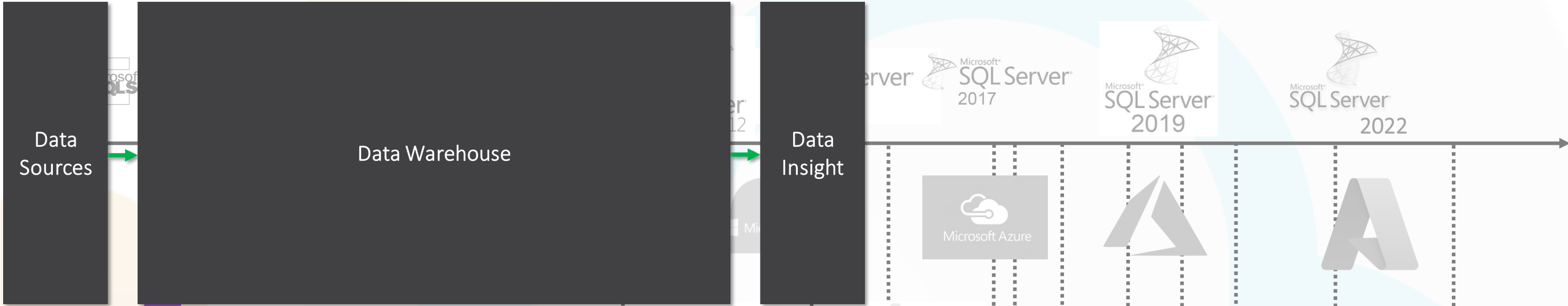


A Timeline of Microsoft Data Technology



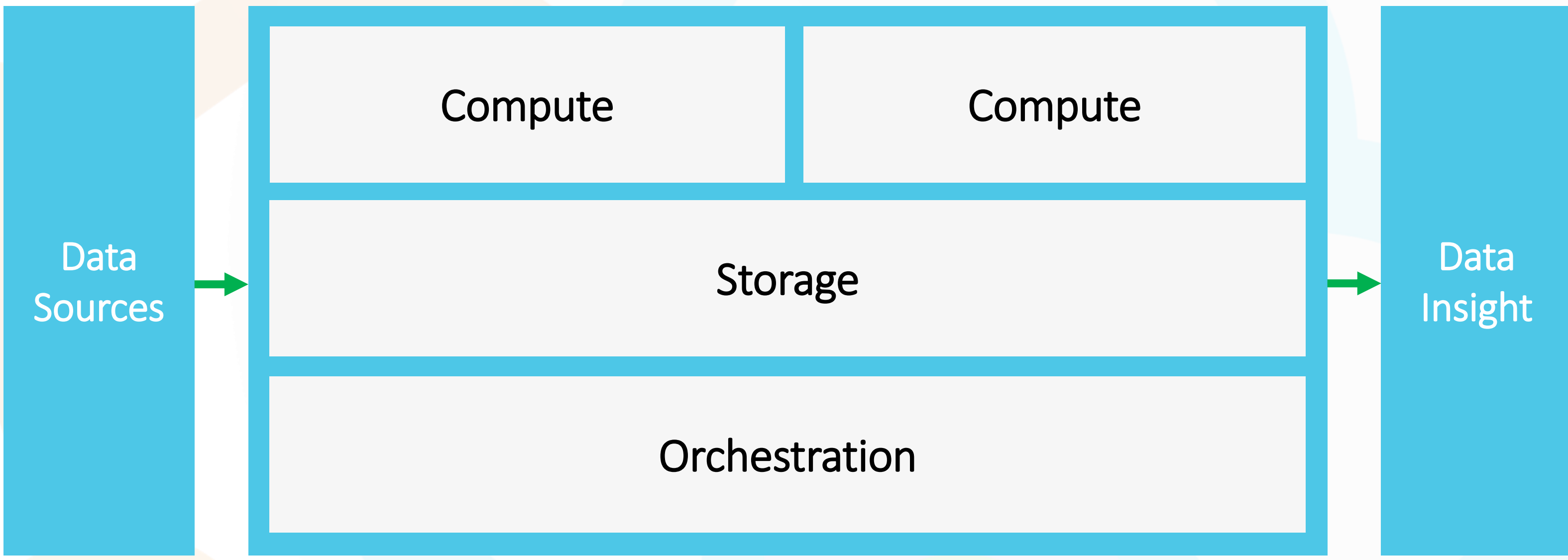
Cloud Formations - Knowledge Transfer & Training

A Timeline of Microsoft Data Technology

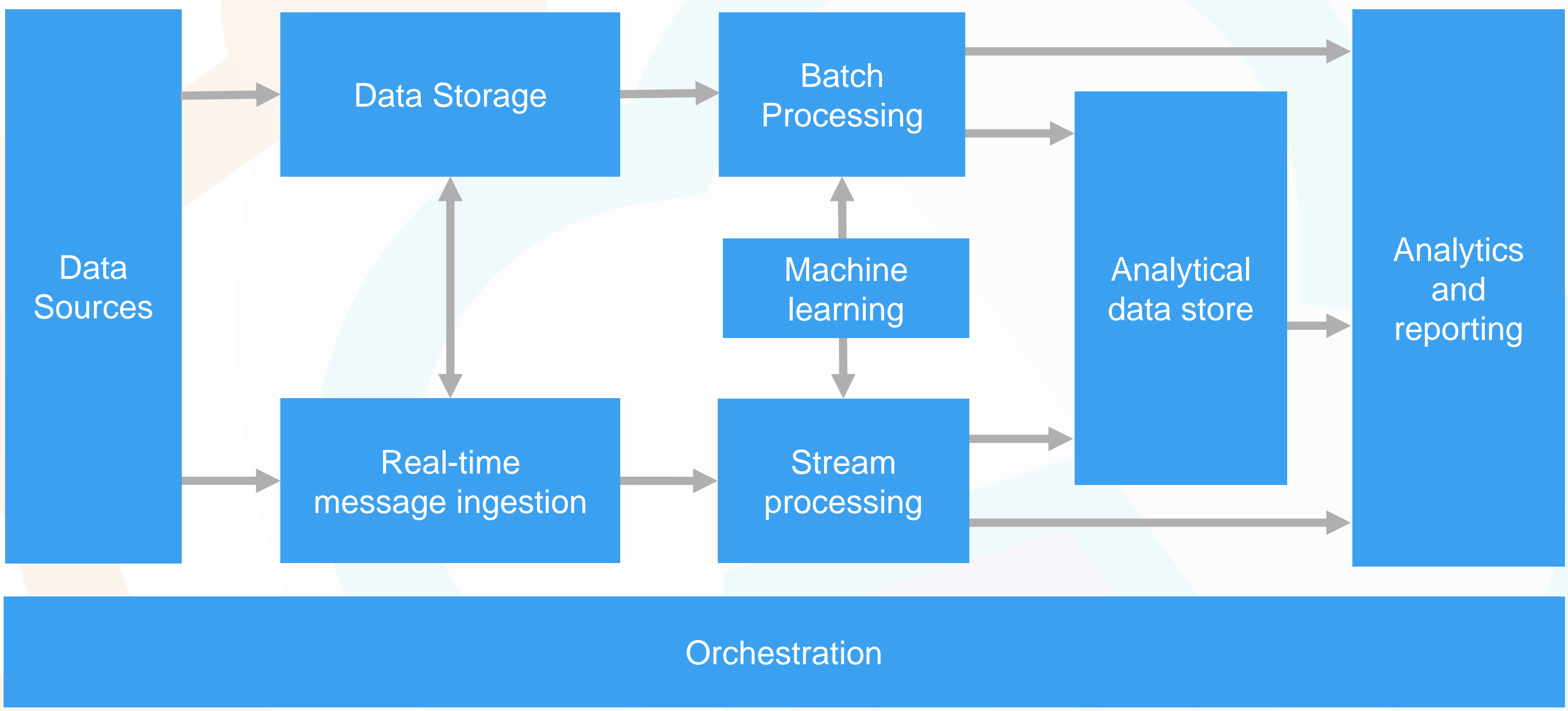


Cloud Formations - Knowledge Transfer & Training

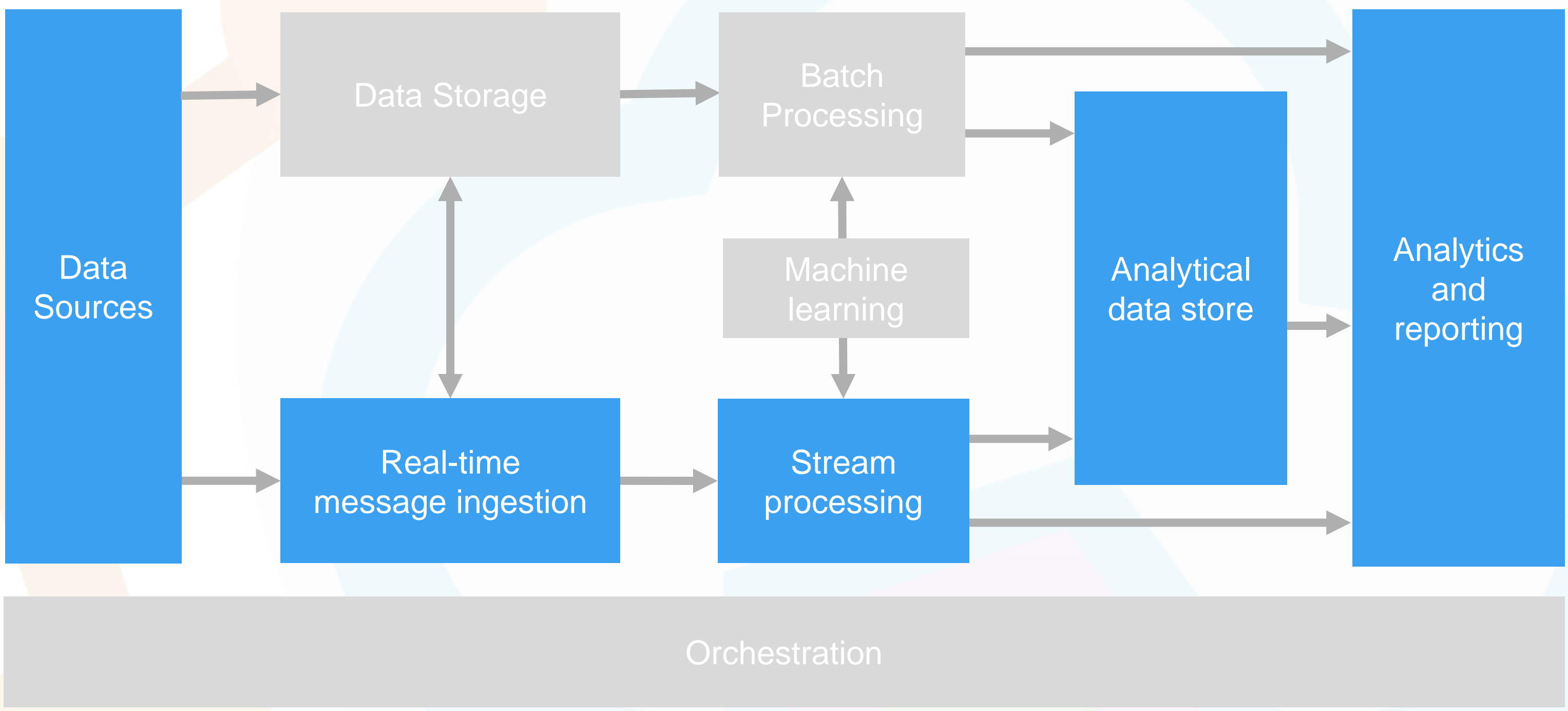
My First Reference Architecture



Components of a Big Data Architecture

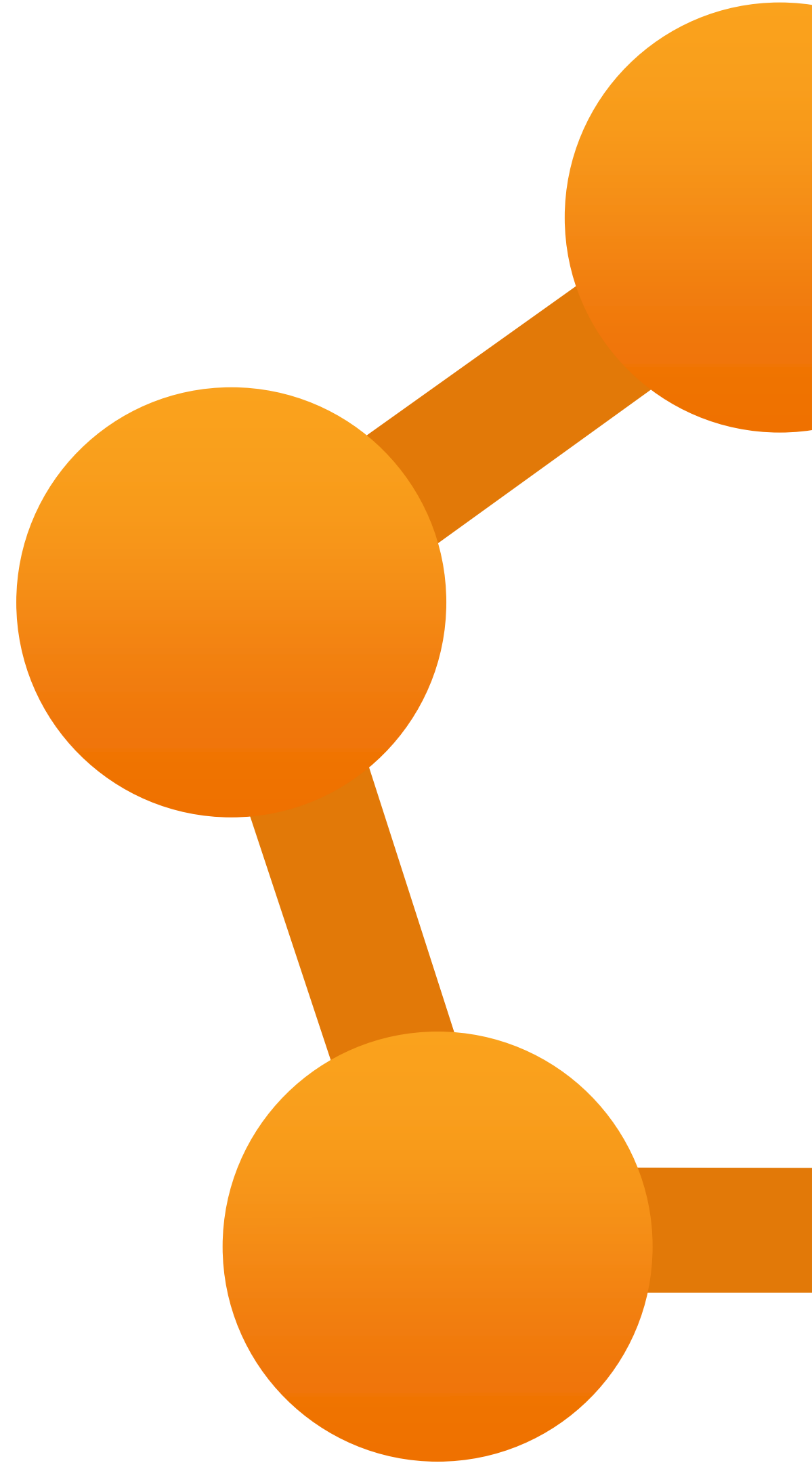


Components of a Big Data Architecture



~~An Engineers Guide To~~ ~~Real-time Data Analytics~~

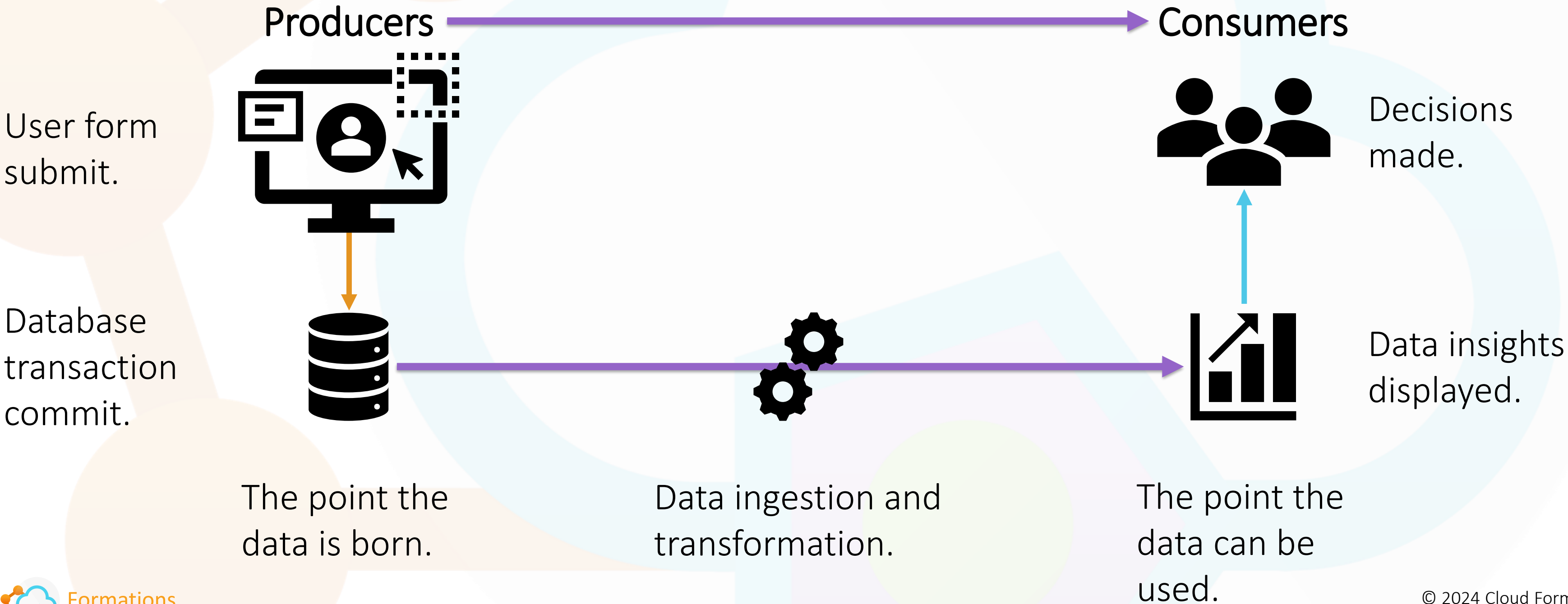
Cloud Formations



What do we mean by real-time data?

Answer (big data):

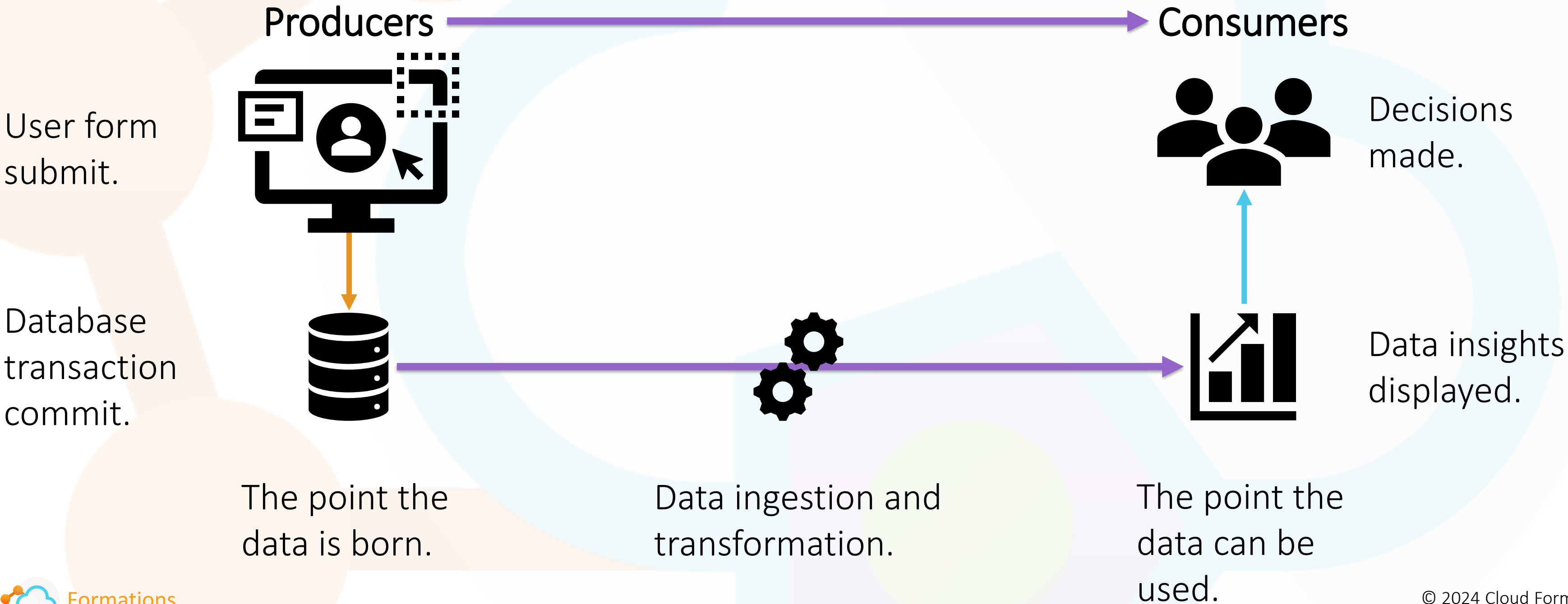
~~“Any data that you cannot process in the time that you have/want using the technology you have.”~~



What do we mean by real-time data?

Answer:

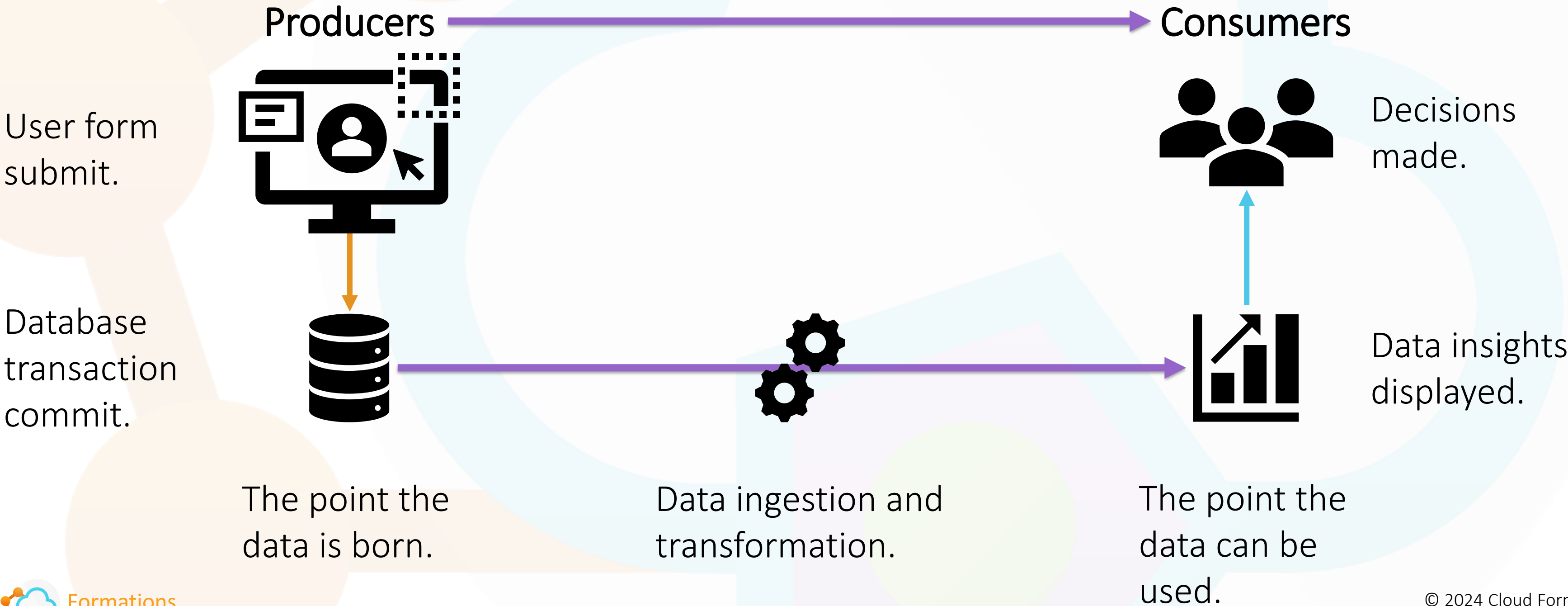
“Delivering data from the producer to consumer as fast as possible using the technology you have.”



What do we mean by near real-time data?

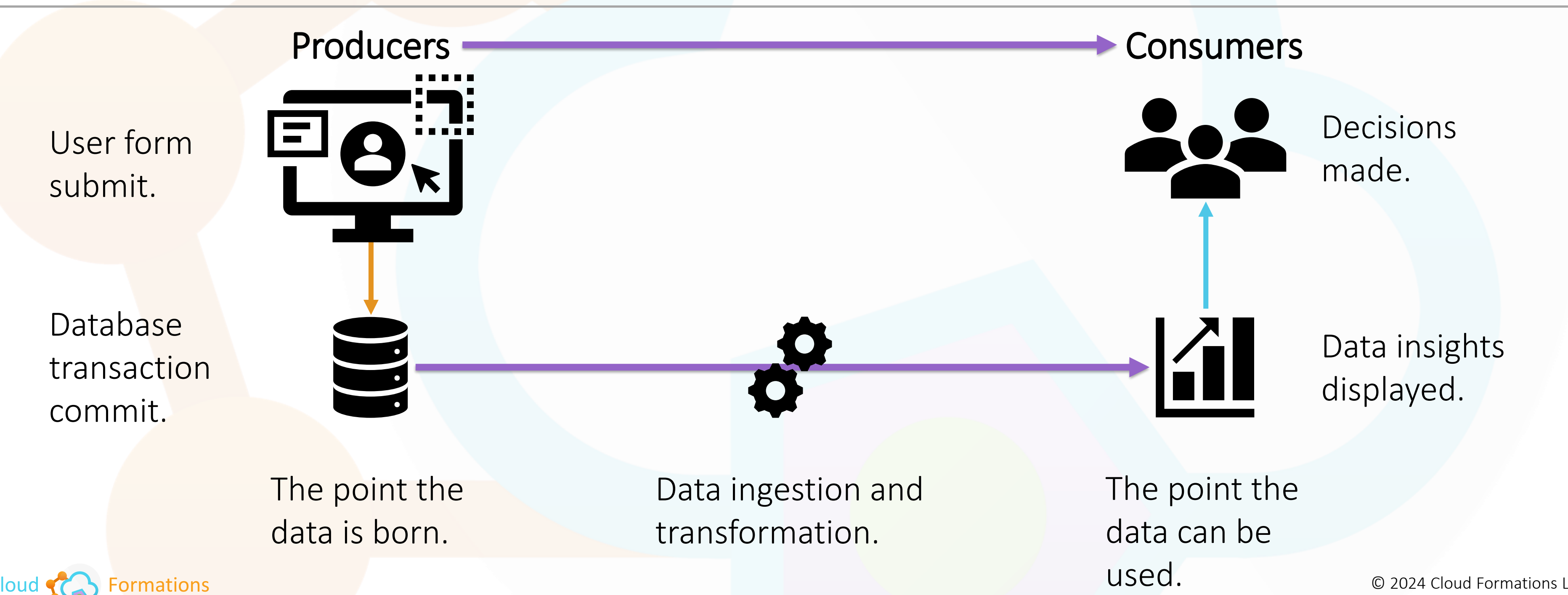
Answer:

“Delivering data from the producer to consumer within 1 minute of it being created (born).”



What do we mean by a data stream?

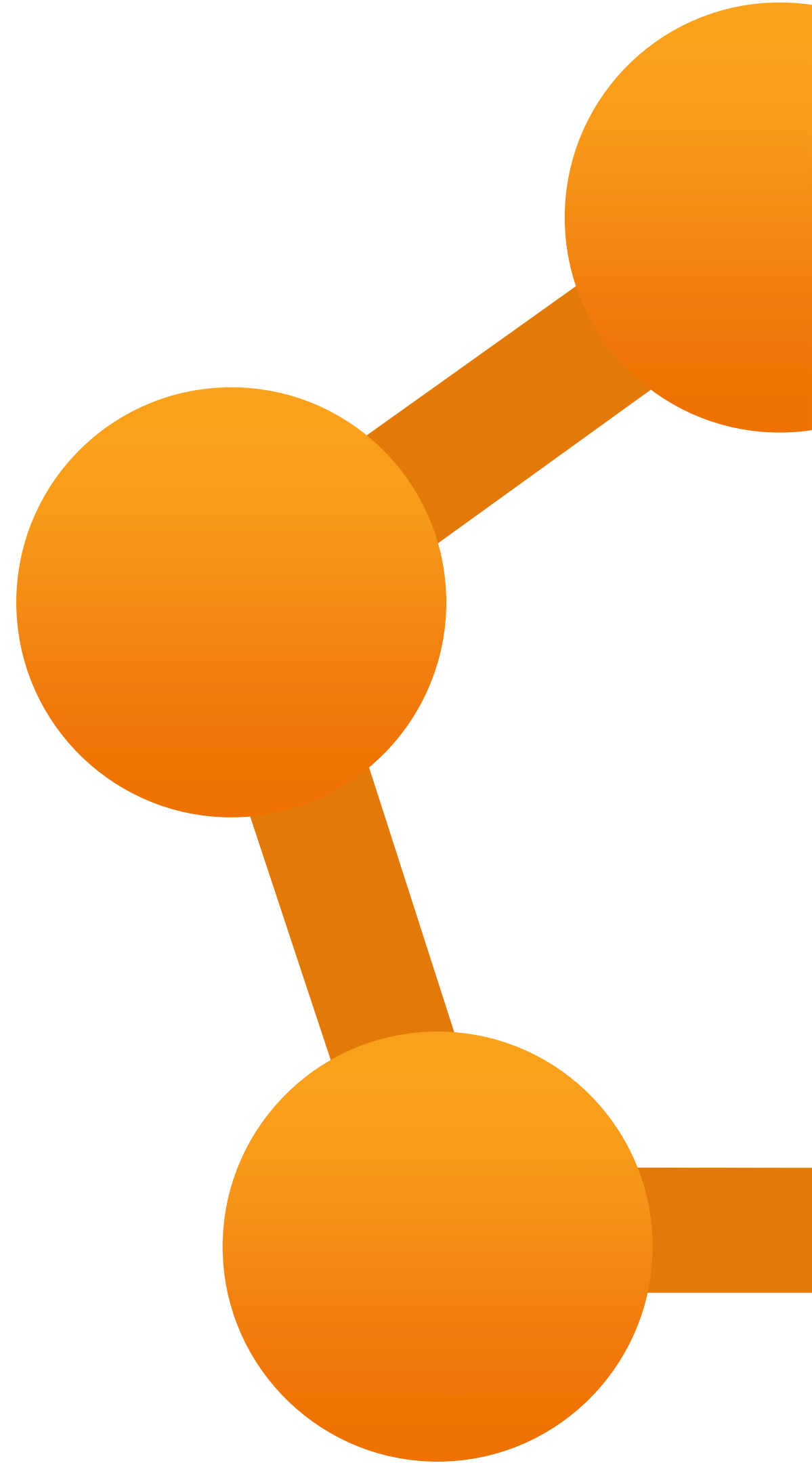
Answer:
“Data that is constantly flowing from producer to consumer in near real-time.”



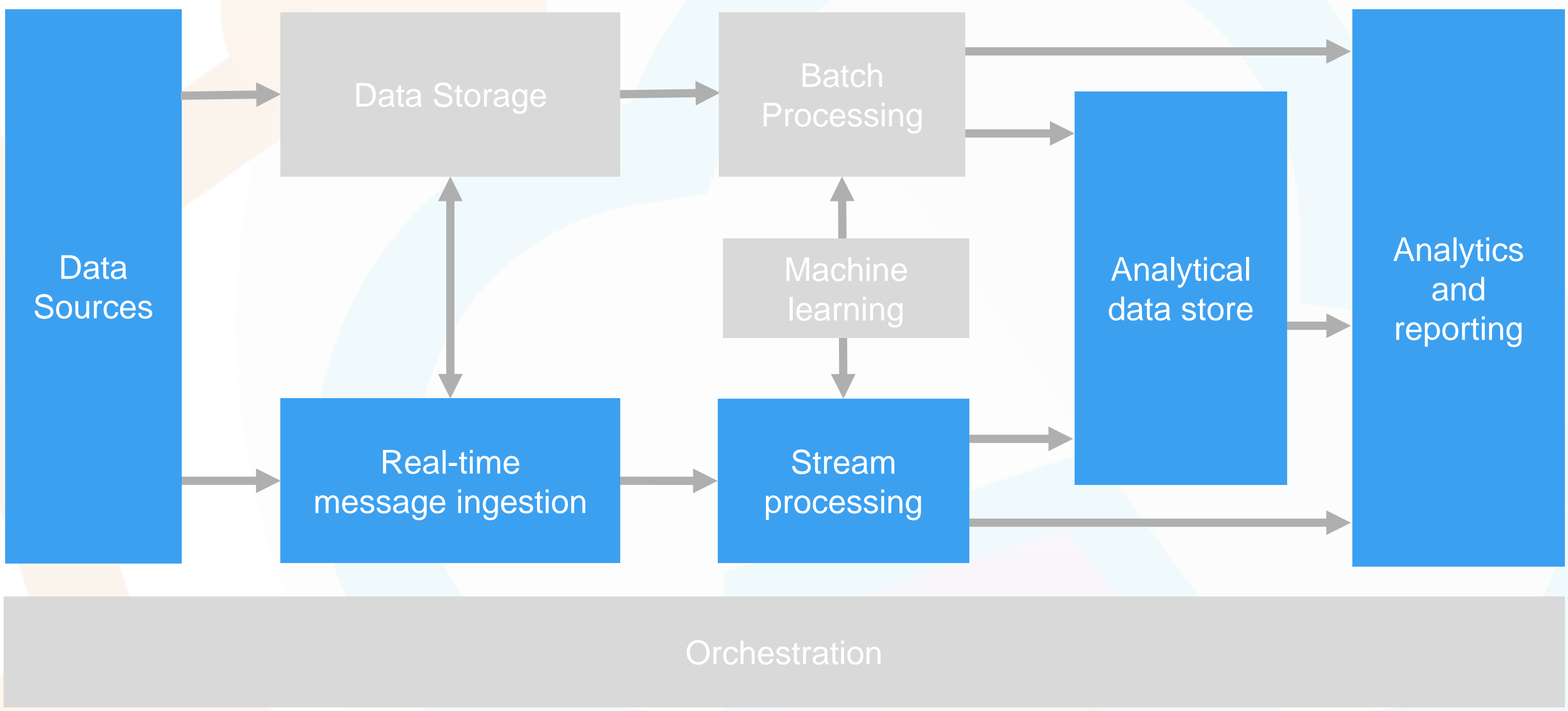


Tooling

Cloud Formations



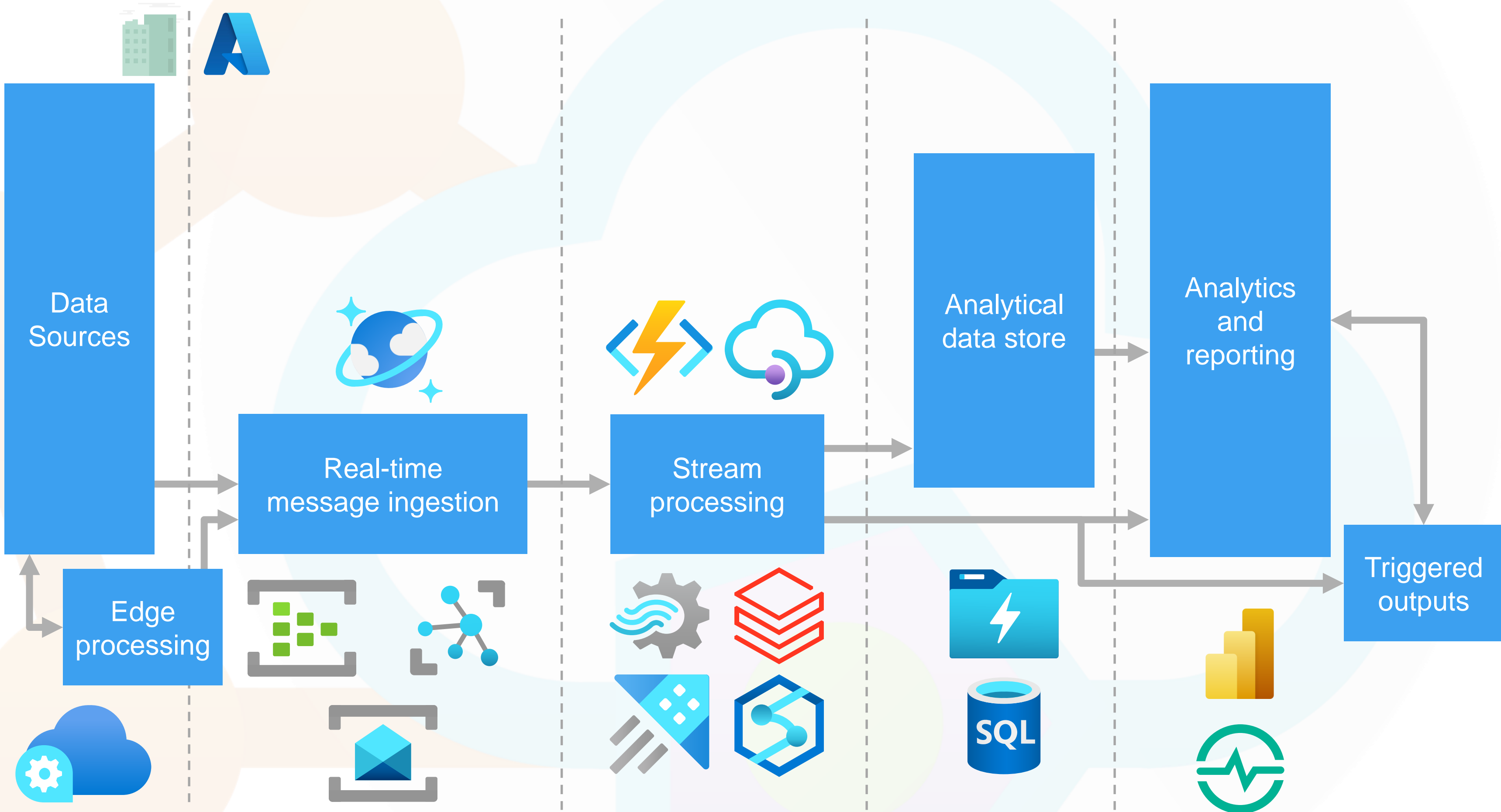
Components of a Big Data Architecture



Azure Tooling



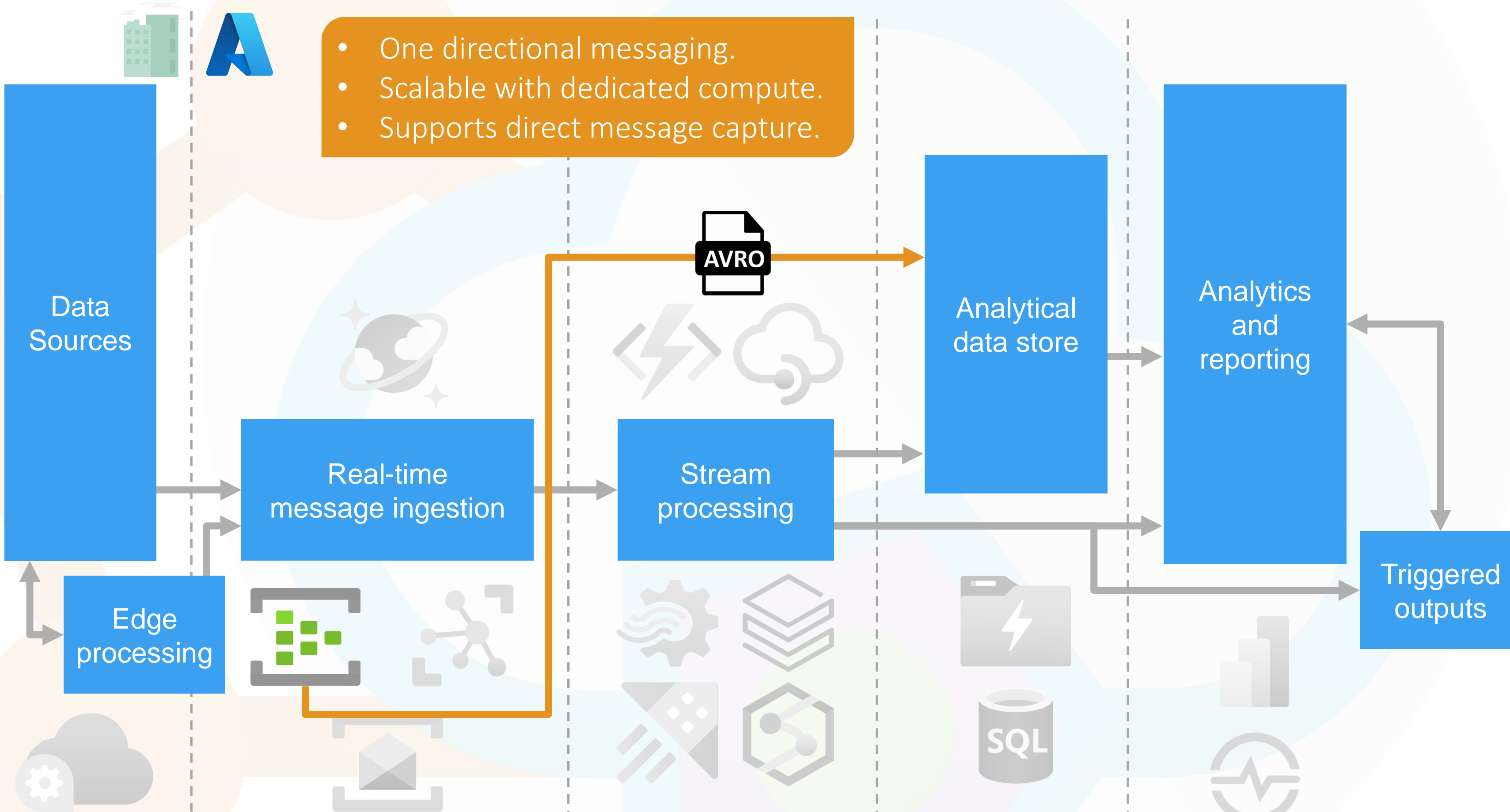
Cloud Formations - Knowledge Transfer & Training



Azure Tooling – Event Hub



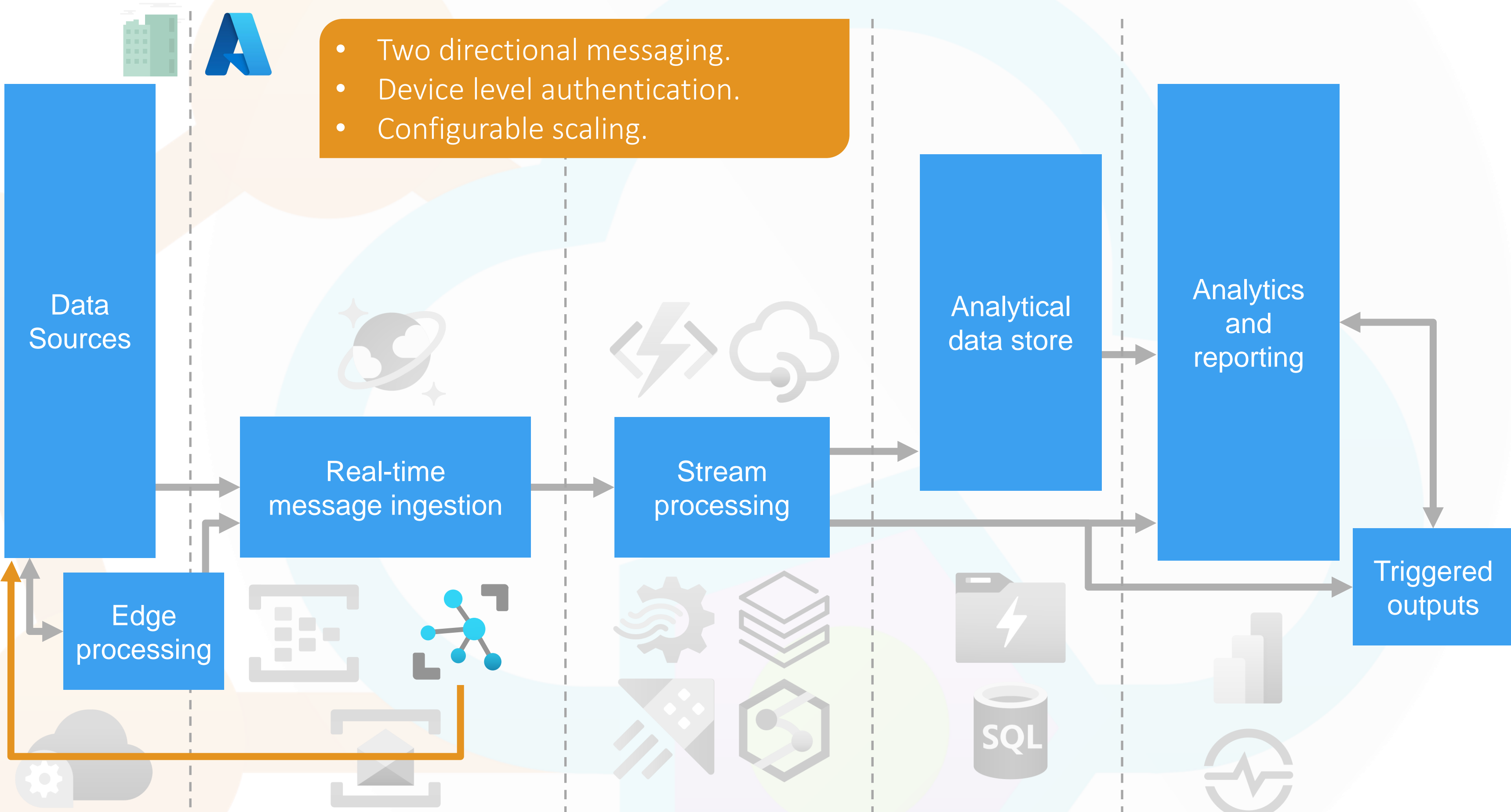
- One directional messaging.
- Scalable with dedicated compute.
- Supports direct message capture.



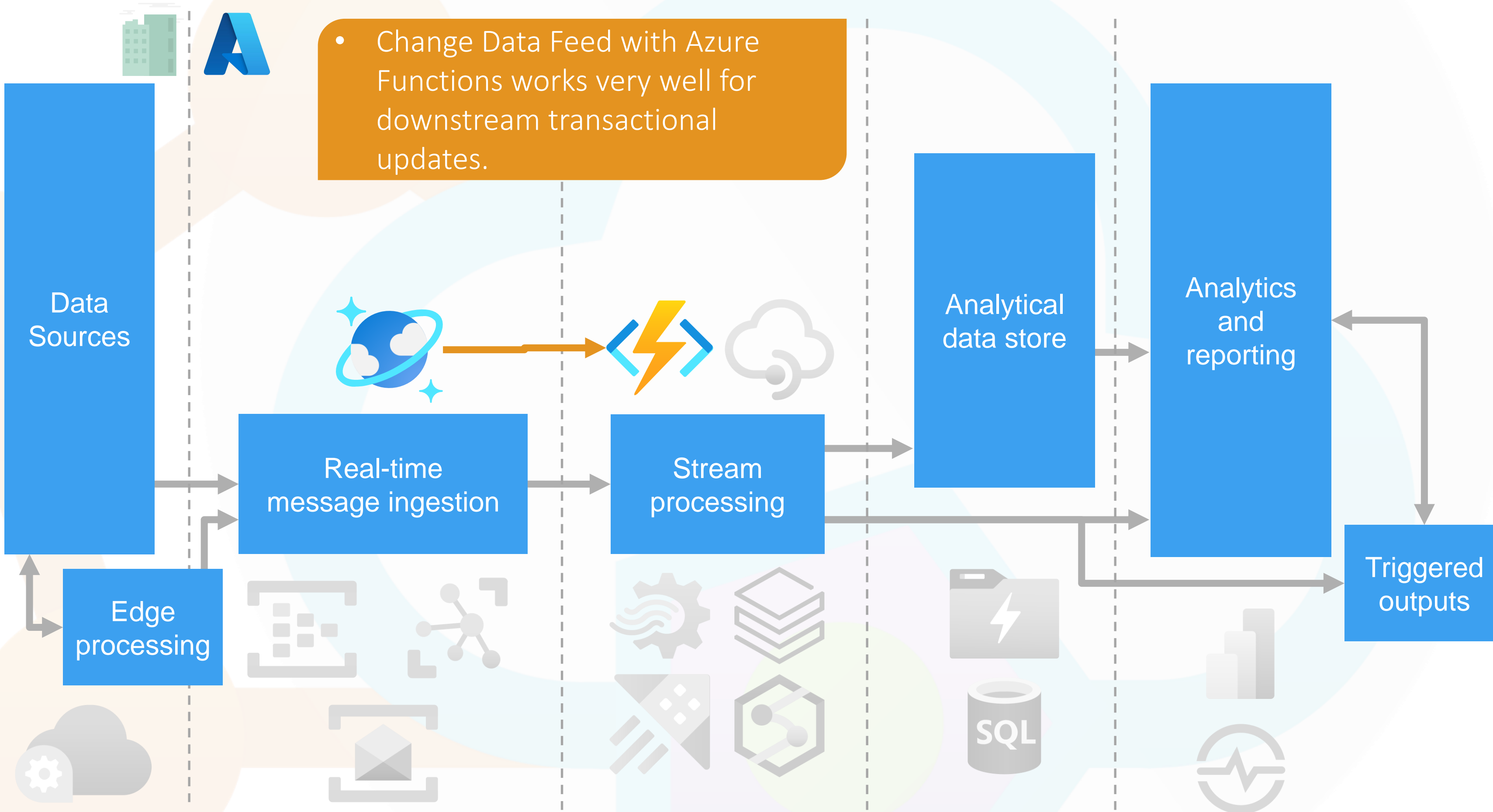
Azure Tooling – IoT Hub



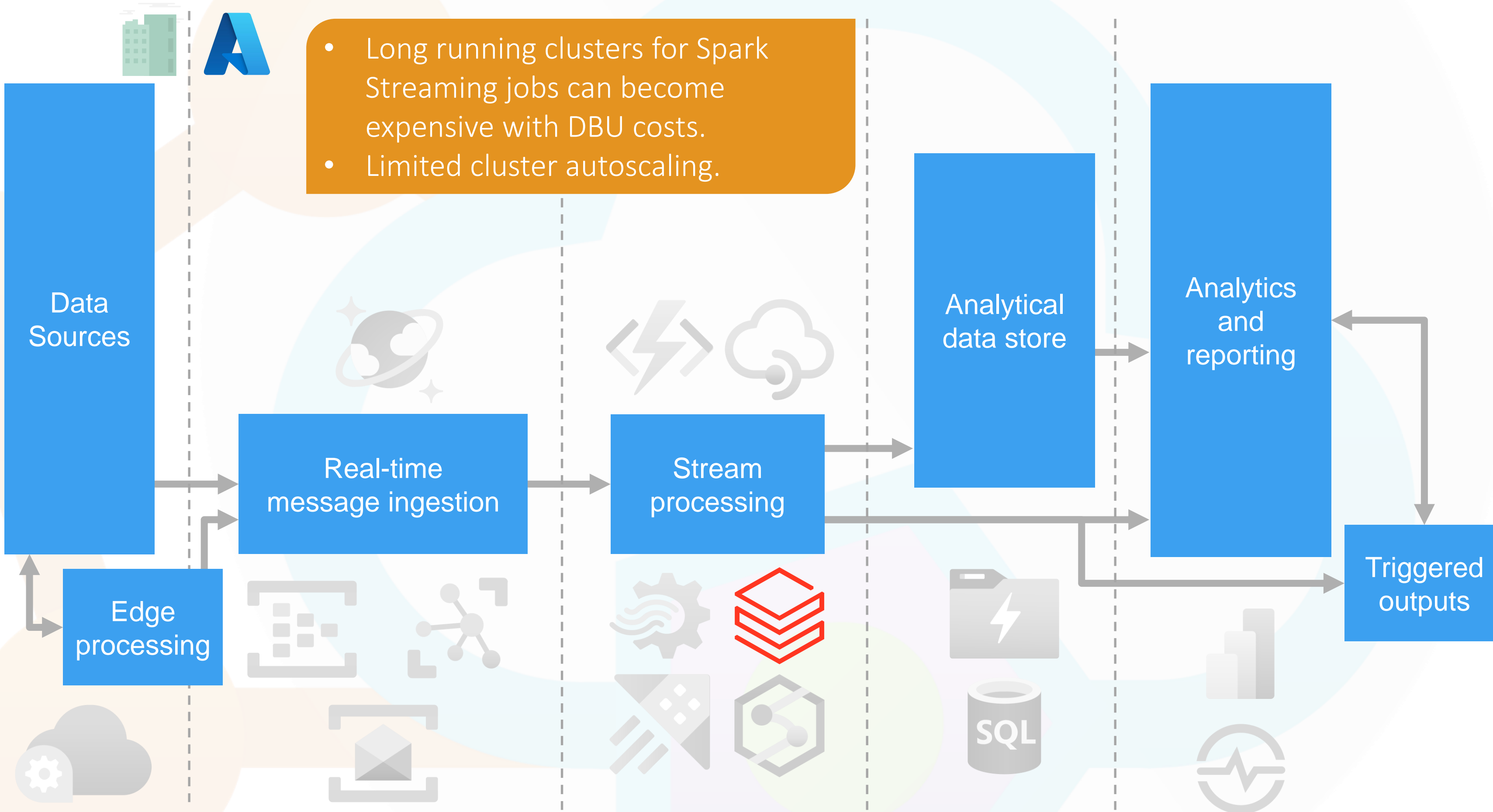
- Two directional messaging.
- Device level authentication.
- Configurable scaling.



Azure Tooling – Cosmos DB



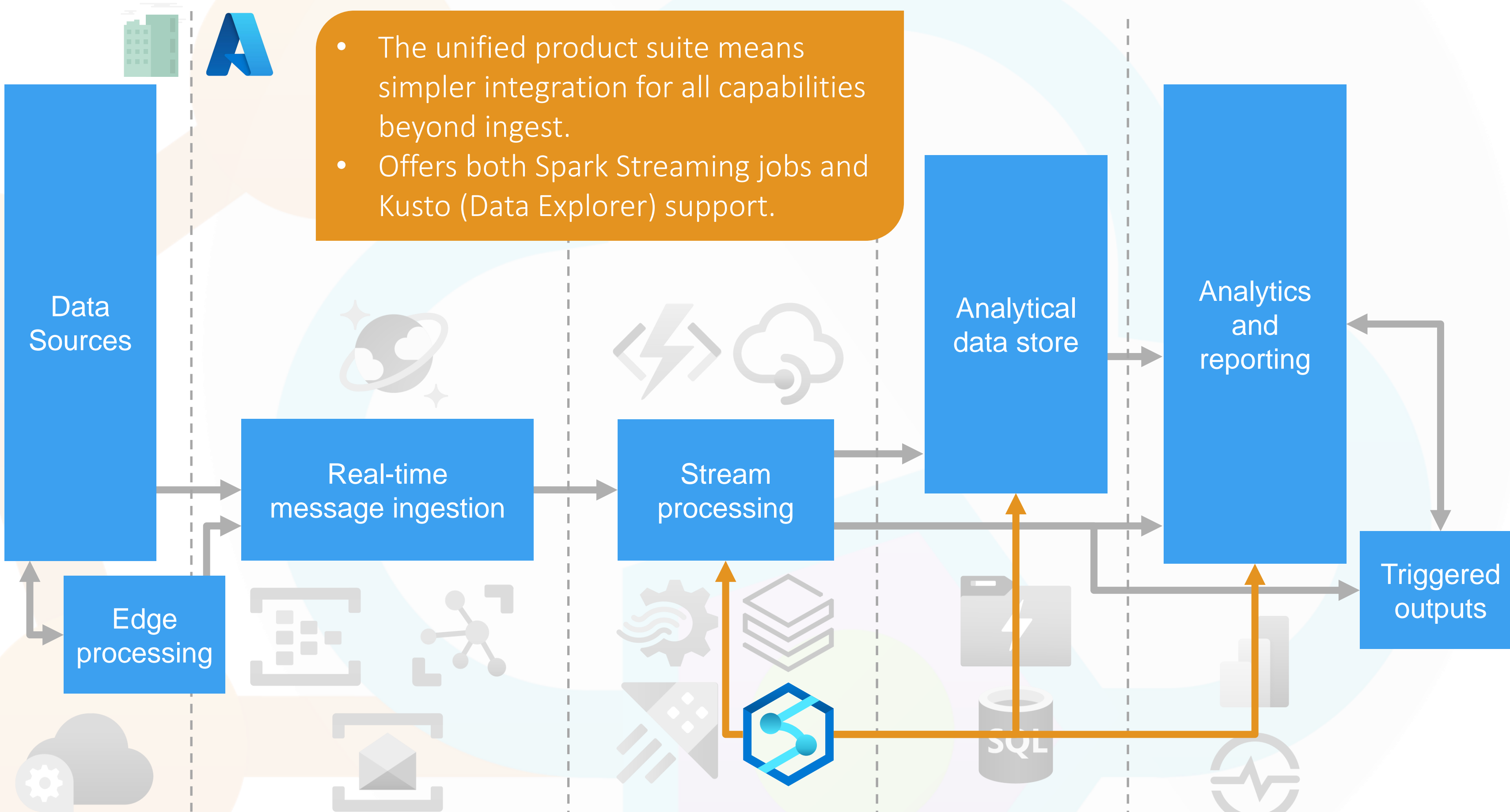
Azure Tooling – Databricks



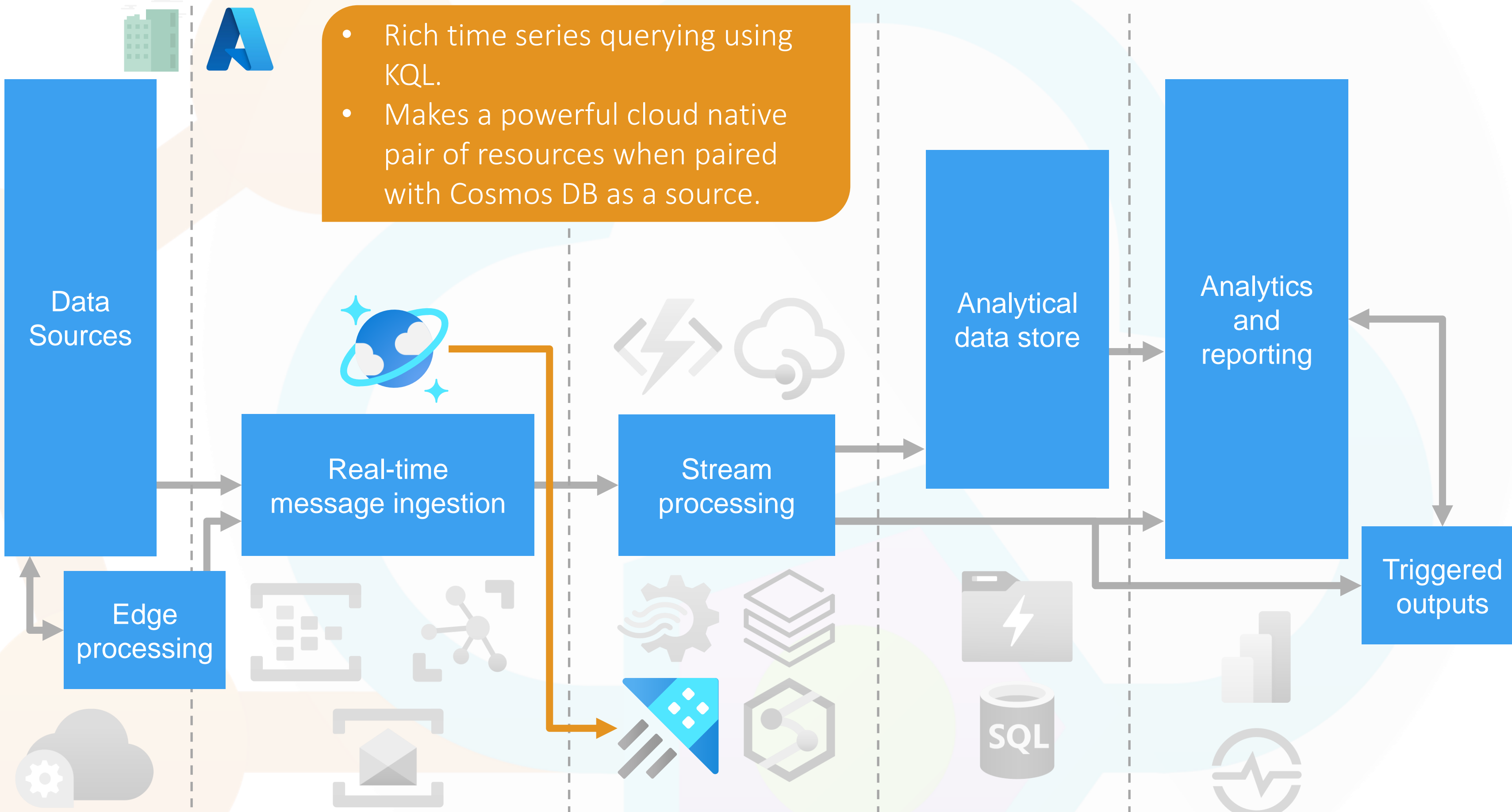
Azure Tooling – Synapse Analytics



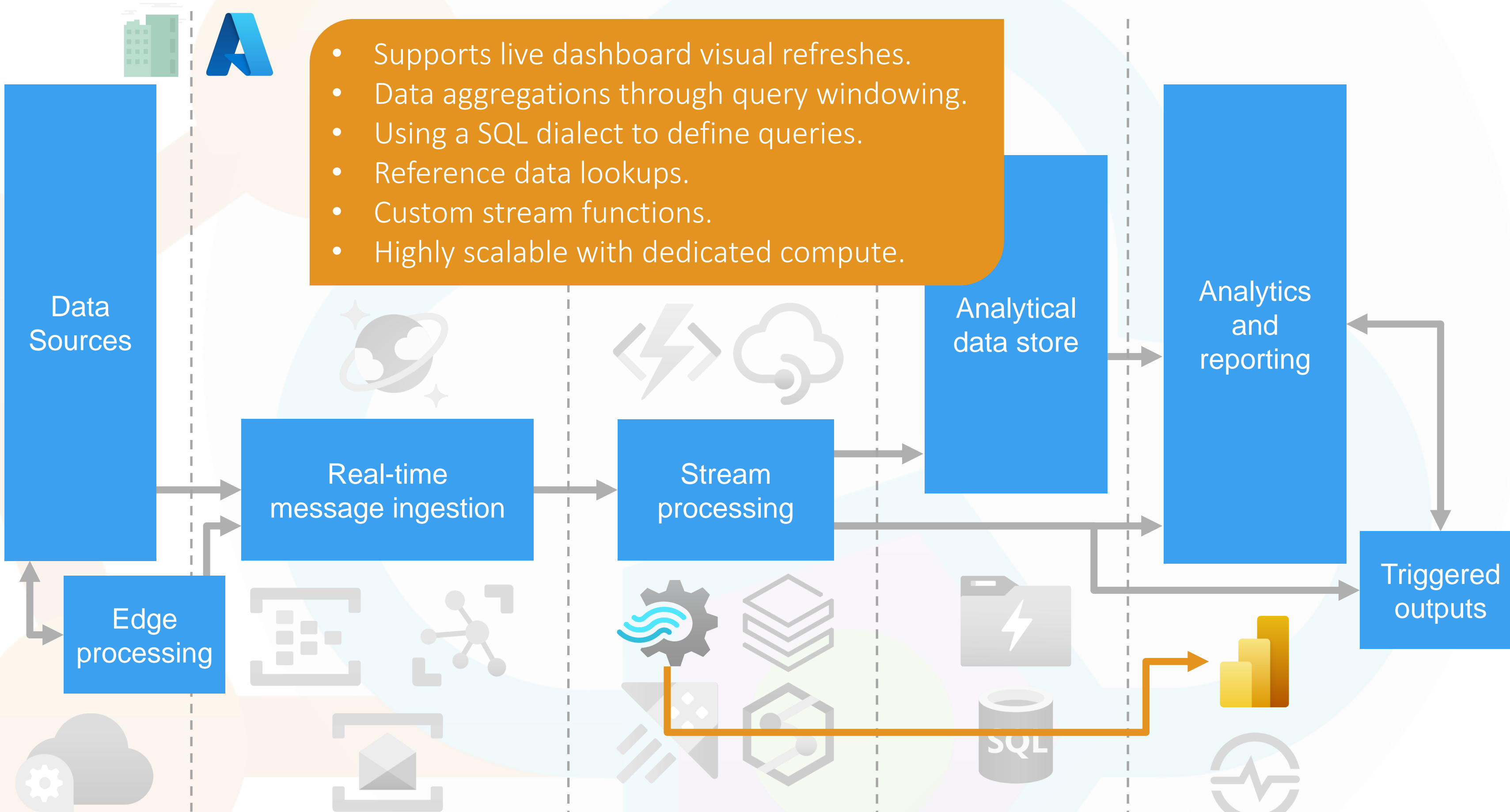
- The unified product suite means simpler integration for all capabilities beyond ingest.
- Offers both Spark Streaming jobs and Kusto (Data Explorer) support.



Azure Tooling – Data Explorer



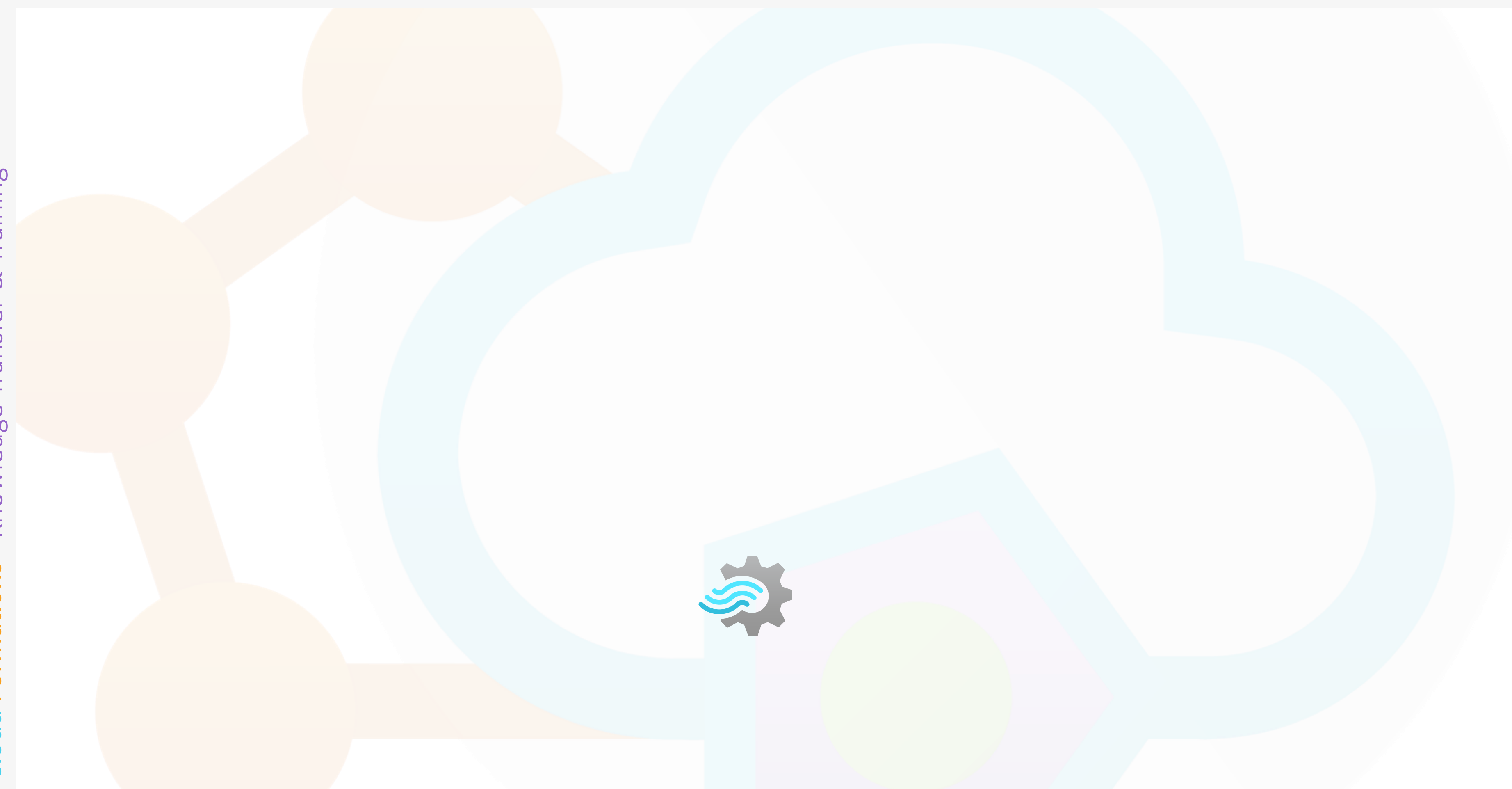
Azure Tooling – Stream Analytics



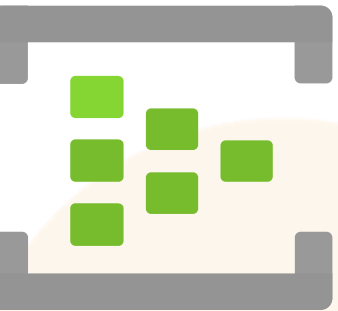
Azure Tooling – Stream Analytics



Cloud Formations - Knowledge Transfer & Training



Azure Stream Analytics



Inputs

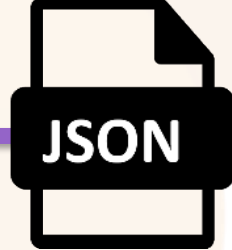
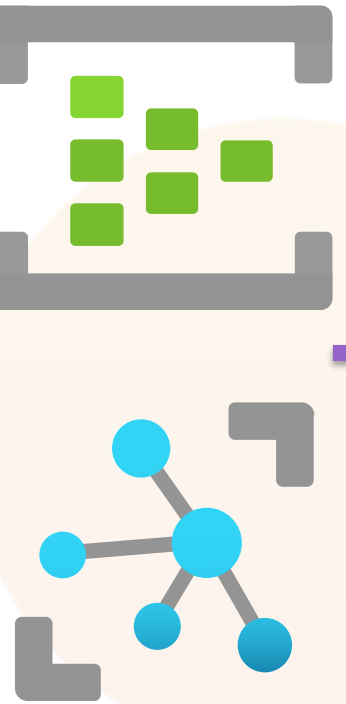


Outputs



Azure Stream Analytics

Query



Inputs

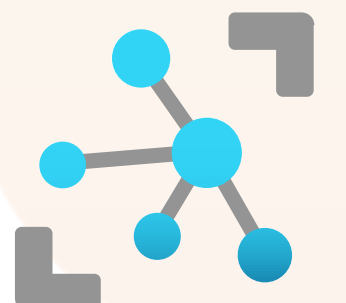
Outputs



Azure Stream Analytics

Query

```
SELECT  
  
    SUM(CAST(eh.UnitPrice AS float)) AS UnitPrice,  
    SUM(CAST(eh.LineTotal AS float)) AS LineTotal,  
    SUM(CAST(eh.OrderQty AS float)) AS OrderQty,  
    COUNT(*) AS RecordCount  
  
INTO  
    [powerbi]  
  
FROM  
    [eventhub] AS eh  
  
GROUP BY  
    eh.EventEnqueuedUtcTime,  
    SlidingWindow(second, 30)
```



Inputs



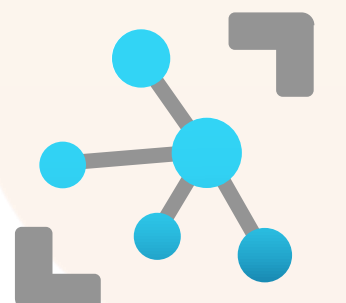
Outputs



Azure Stream Analytics

Query

```
SELECT
  eh.EventEnqueuedUtcTime,
  prd.Name AS ProductName,
  SUM(CAST(eh.UnitPrice AS float)) AS UnitPrice,
  SUM(CAST(eh.LineTotal AS float)) AS LineTotal,
  SUM(CAST(eh.OrderQty AS float)) AS OrderQty,
  COUNT(*) AS RecordCount
INTO
  [powerbi]
FROM
  [eventhub] AS eh
  INNER JOIN [Products] AS prd
    ON eh.[ProductId] = prd.[ProductId]
GROUP BY
  eh.EventEnqueuedUtcTime,
  prd.Name,
  SlidingWindow(second, 30)
```



Inputs



Outputs



Azure Stream Analytics

Query

```
SELECT
  eh.EventEnqueuedUtcTime,
  udf.CleanString(prd.Name) AS ProductName,
  SUM(CAST(eh.UnitPrice AS float)) AS UnitPrice,
  SUM(CAST(eh.LineTotal AS float)) AS LineTotal,
  SUM(CAST(eh.OrderQty AS float)) AS OrderQty,
  COUNT(*) AS RecordCount
INTO
  [powerbi]
FROM
  [eventhub] AS eh
  INNER JOIN [Products] AS prd
    ON eh.[ProductId] = prd.[ProductId]
GROUP BY
  eh.EventEnqueuedUtcTime,
  prd.Name,
  SlidingWindow(second, 30)
```

JS

CSV



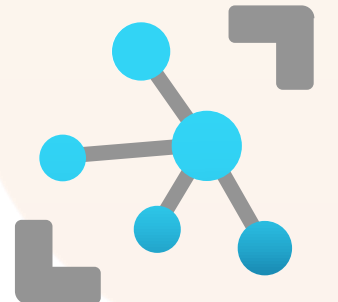
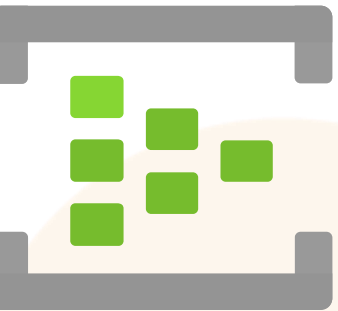
Inputs ... Source Data
Reference Data
Custom Functions

Outputs



Azure Stream Analytics

Query

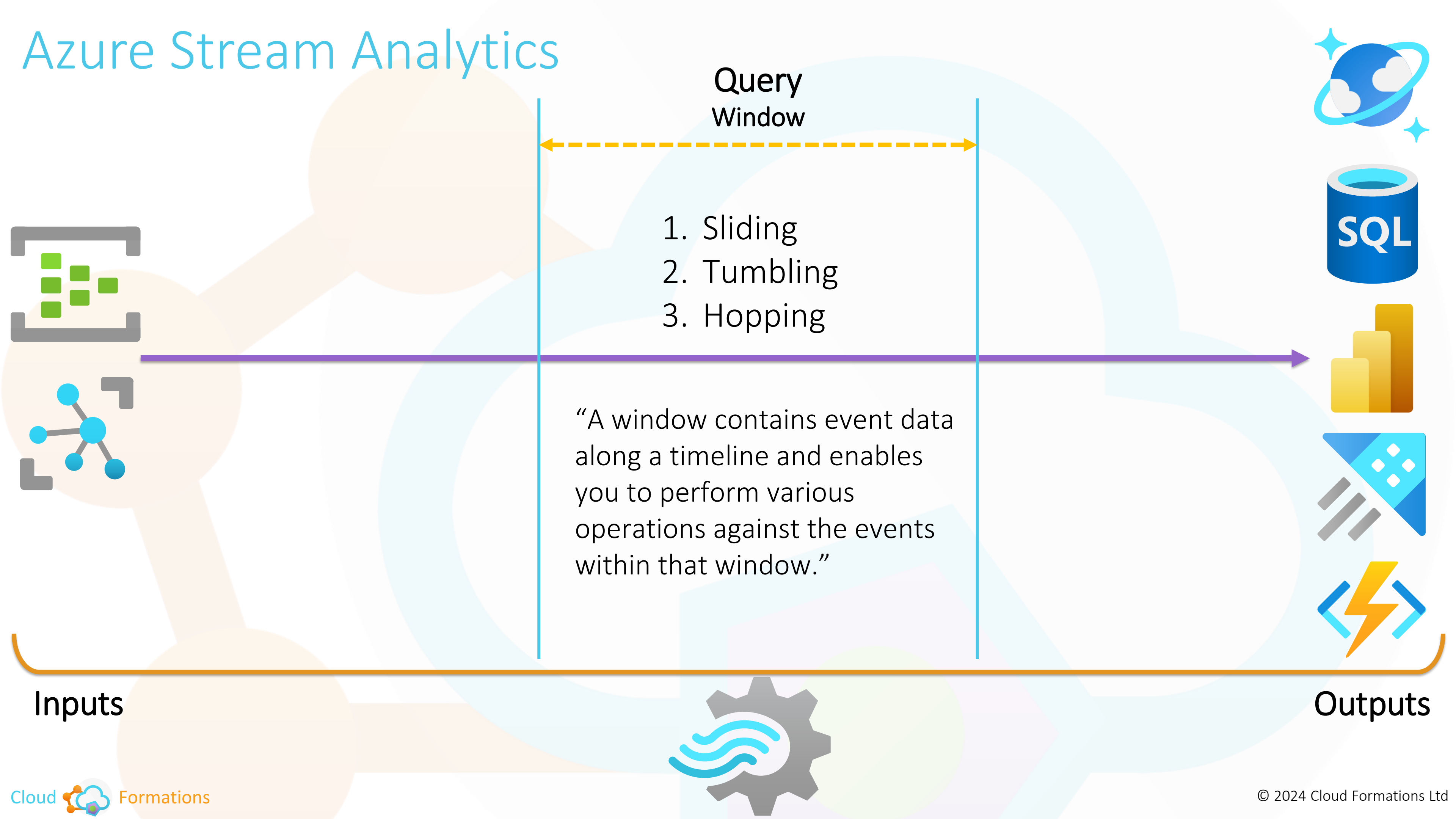


Inputs

Outputs



Azure Stream Analytics



Query Window

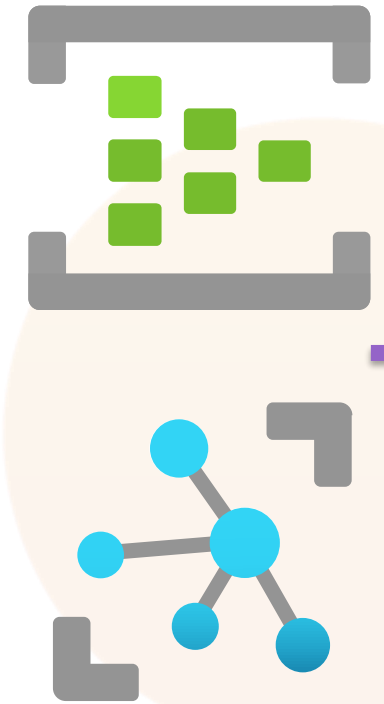
- 1. Sliding
- 2. Tumbling
- 3. Hopping

“A window contains event data along a timeline and enables you to perform various operations against the events within that window.”

Inputs

Outputs

Azure Stream Analytics – Sliding Window



Inputs

Sliding Window

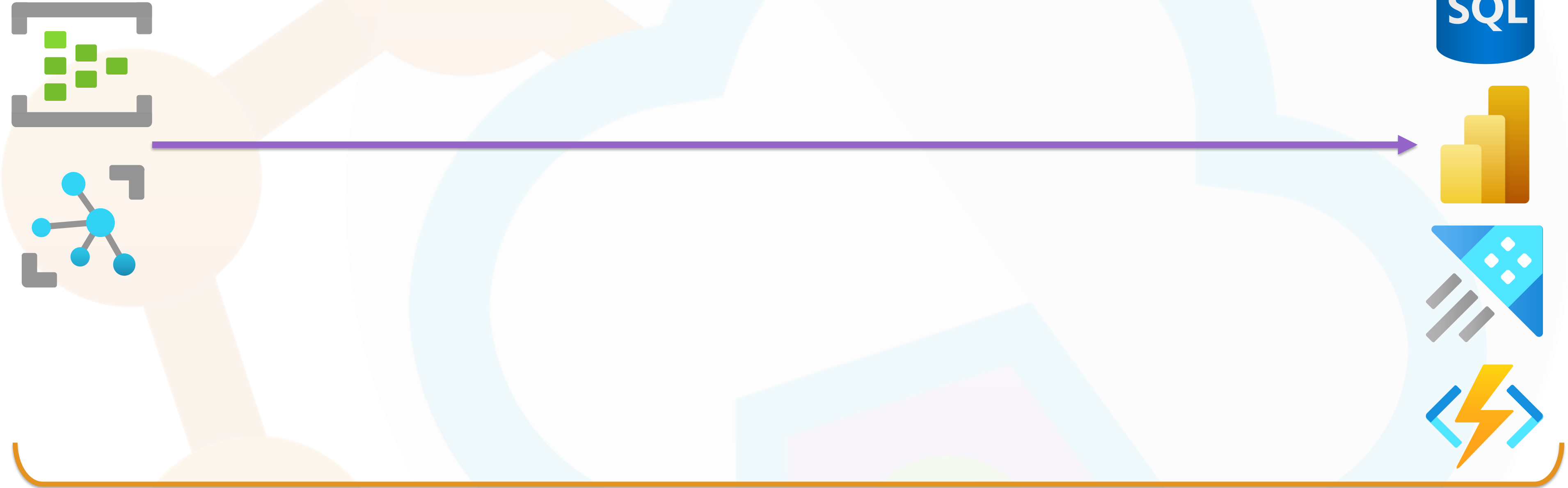
```
SELECT  
    MIN([value]) AS 'Min',  
    MAX([value]) AS 'Max'  
INTO  
    SQLDBAvg  
FROM  
    IoTHub  
TIMESTAMP BY  
    timecreated  
GROUP BY  
    SlidingWindow(second, 5))
```



Outputs



Azure Stream Analytics – Tumbling Window



Inputs

Outputs

Azure Stream Analytics – Tumbling Window



Tumbling Window

Tumbling Window

Tumbling Window

Tumbling Window

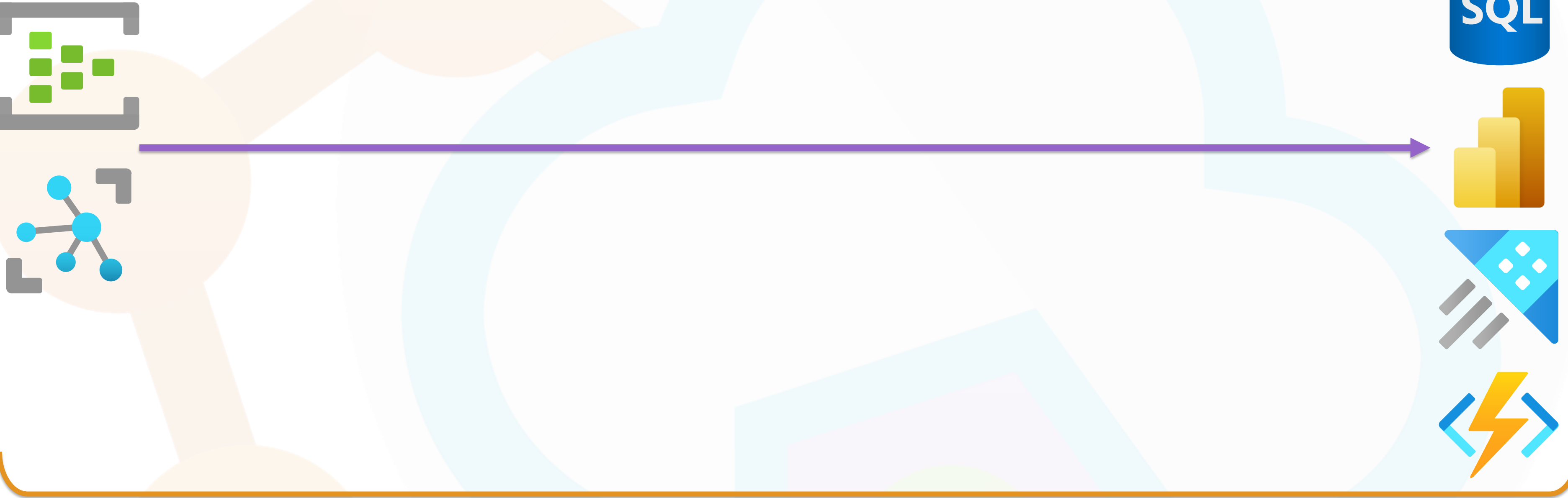
```
SELECT
  MIN([value]) AS 'Min',
  MAX([value]) AS 'Max'
INTO
  SQLDBAvg
FROM
  IoTHub
TIMESTAMP BY
  timecreated
GROUP BY
  TumblingWindow(Duration(MINUTE, 15), Offset(millisecond, -1))
```

Inputs

Outputs



Azure Stream Analytics – Hopping Window



Inputs

Outputs

Azure Stream Analytics – Hopping Window



Hopping Window

Hopping Window

Hopping Window

Hopping Window

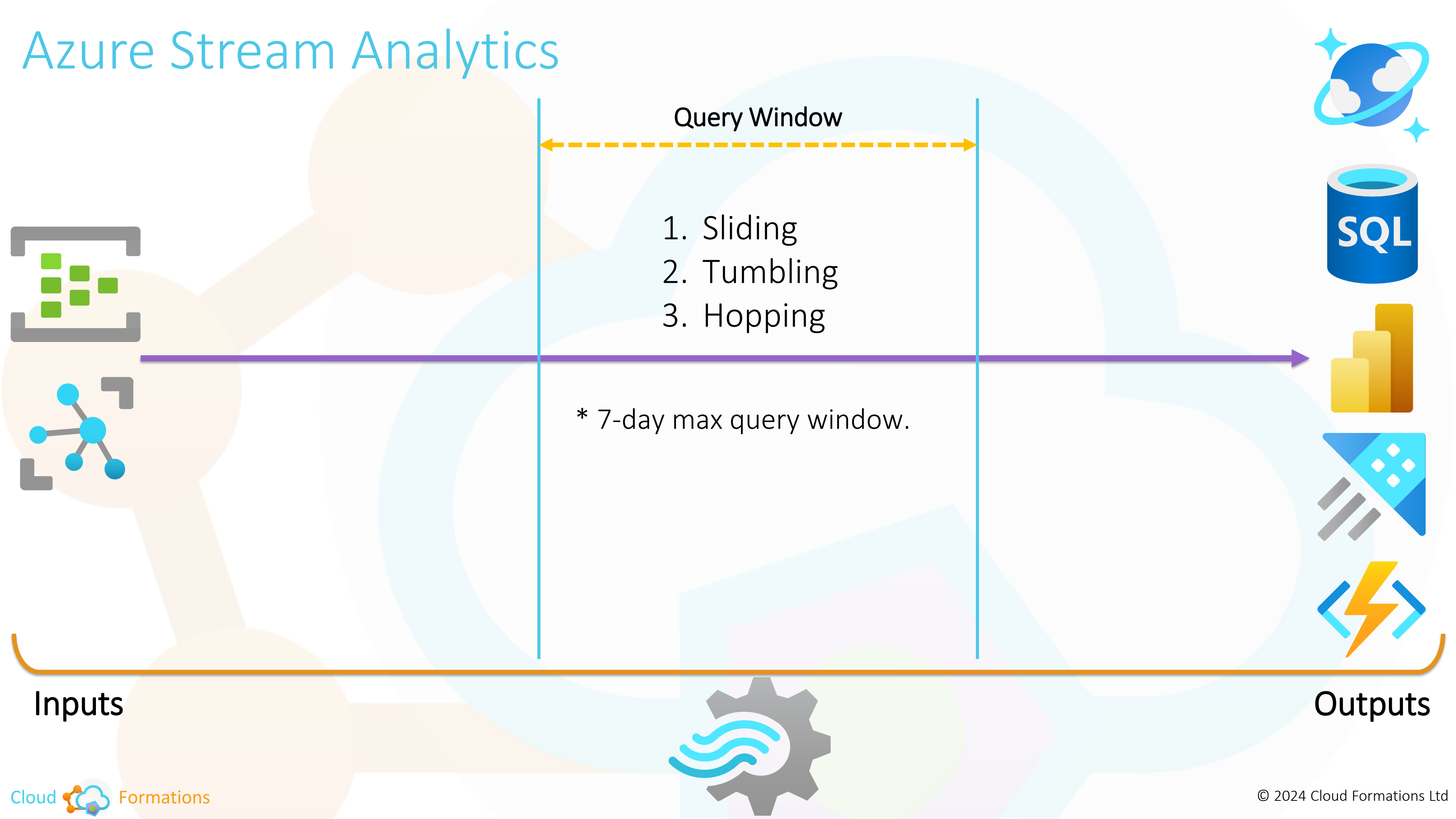
```
SELECT
  MIN([value]) AS 'Min',
  MAX([value]) AS 'Max'
INTO
  SQLDBAvg
FROM
  IoTHub
TIMESTAMP BY
  timecreated
GROUP BY
  HoppingWindow(Duration(MINUTE, 15), Hop(MINUTE, 5), Offset(millisecond, -1))
```

Inputs

Outputs



Azure Stream Analytics



Query Window

- 1. Sliding
- 2. Tumbling
- 3. Hopping

* 7-day max query window.

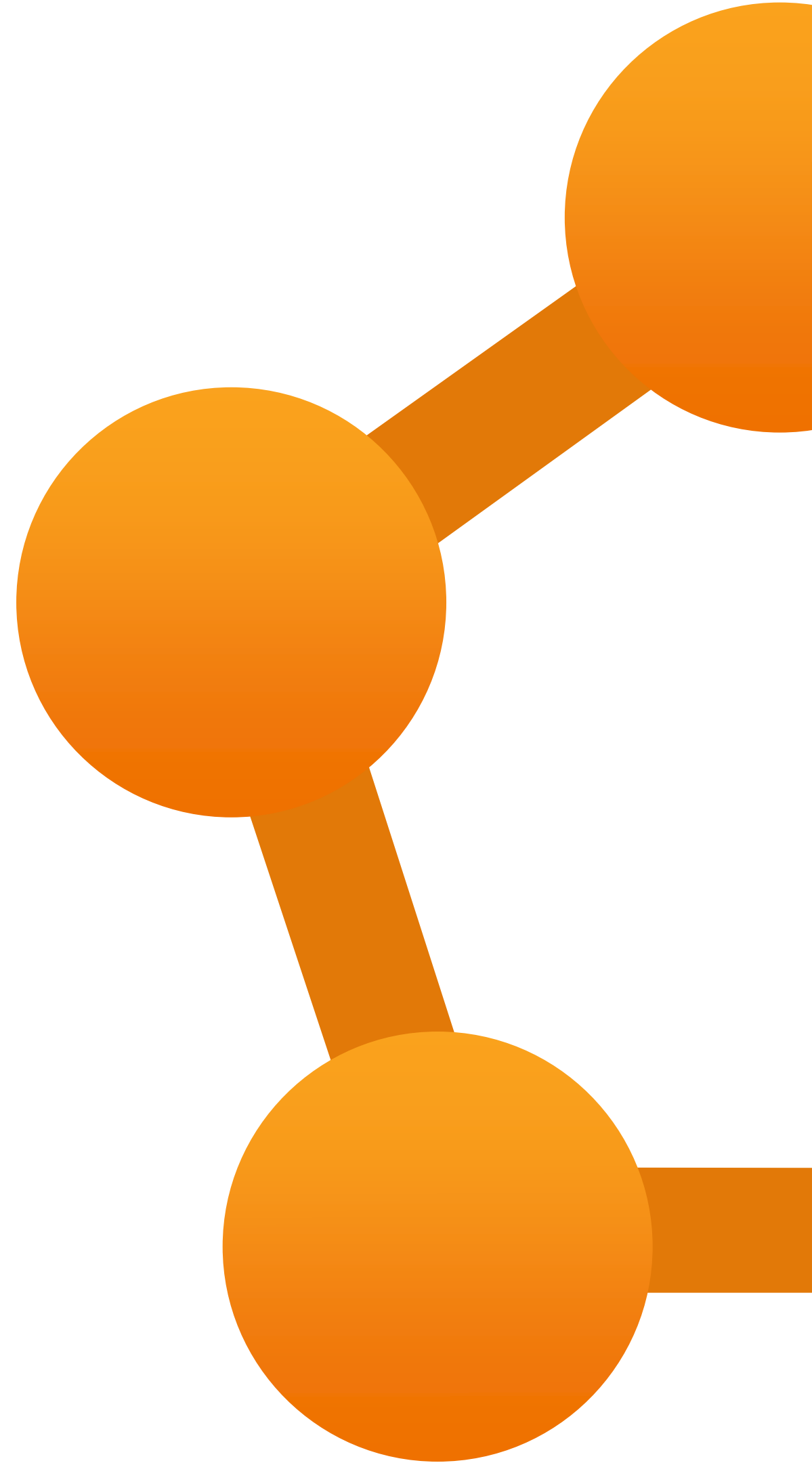
Inputs

Outputs



Tooling

Cloud Formations



Different Types of Fabric



The screenshot shows the Gartner website's Information Technology Glossary page for 'Data Fabric'. The page title is 'Gartner Glossary' under 'Information Technology'. The breadcrumb trail is 'Gartner Glossary > Information Technology Glossary > D > Data Fabric'. The main heading is 'Data Fabric'. The definition states: '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>

The screenshot shows the Microsoft documentation page titled 'What is Microsoft Fabric?'. The breadcrumb trail is 'Learn / Microsoft Fabric / Get started /'. The article is dated 11/15/2023 and has 6 contributors. A 'Feedback' button is visible. The 'In this article' section lists: 'SaaS foundation', 'Components of Microsoft Fabric', 'OneLake and lakehouse - the unification of lakehouses', 'Fabric solutions for ISVs', and 'Next steps'. The main text describes Microsoft Fabric as an 'all-in-one analytics solution' for enterprises covering data movement, data science, Real-Time Analytics, and business intelligence. It offers a 'comprehensive suite of services' including data lake, data engineering, and data integration. The text notes that with Fabric, users don't need to piece together services from multiple vendors, instead enjoying a highly integrated, end-to-end, and easy-to-use product. The platform is built on a SaaS foundation, which takes 'simplicity and integration to a whole new level.'

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



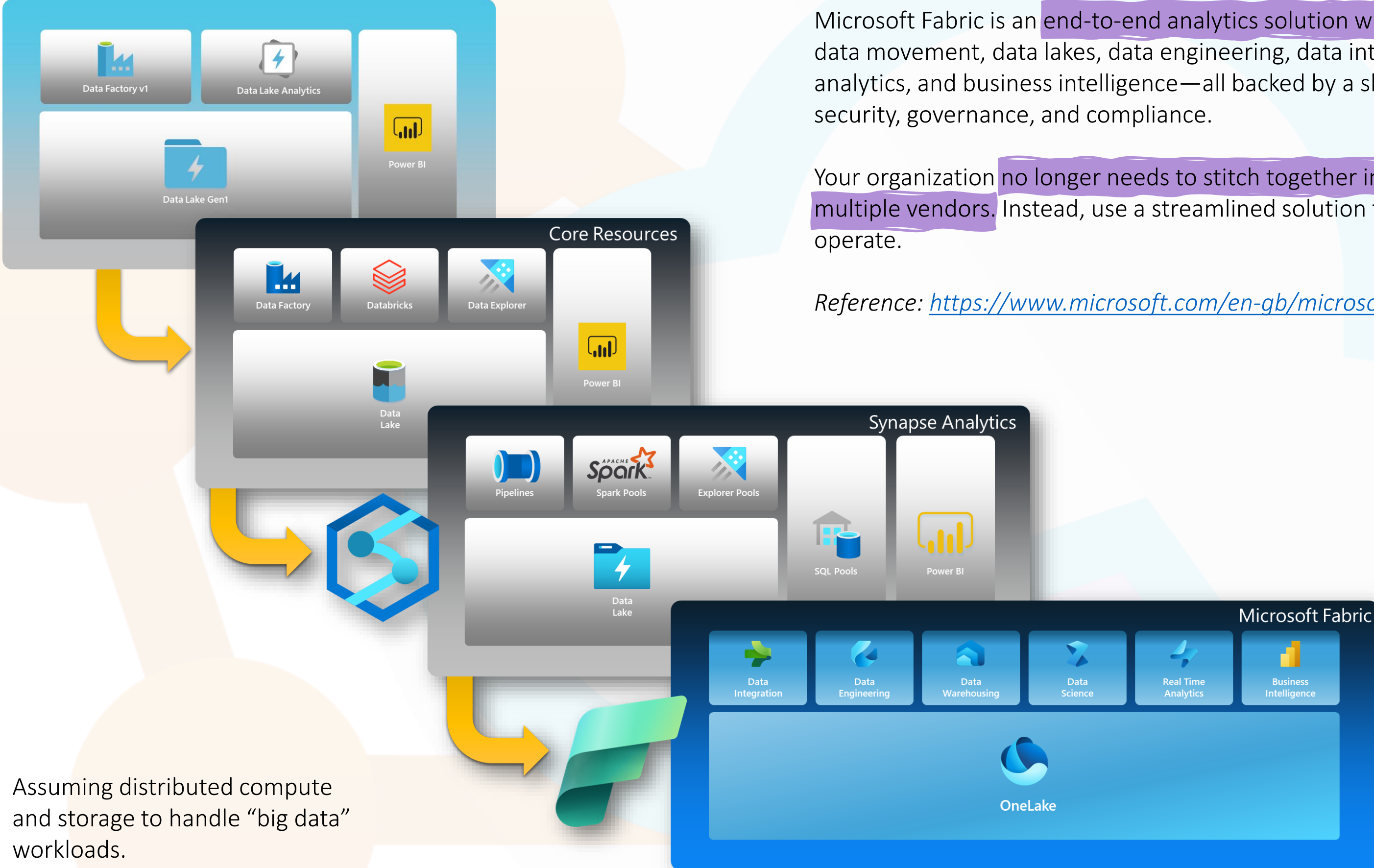
What is Microsoft Fabric? – Vision and Stack Evolution



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?



Cloud Formations - Knowledge Transfer & Training

A screenshot of the Microsoft Fabric dashboard. The dashboard is dark blue and features a grid of six icons representing different data services: Data Integration, Data Engineering, Data Warehousing, Data Science, Real Time Analytics, and Business Intelligence. Below these icons is a large blue box with the OneLake logo and the text 'OneLake'. The text 'Microsoft Fabric' is visible in the top right corner of the dashboard area.

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

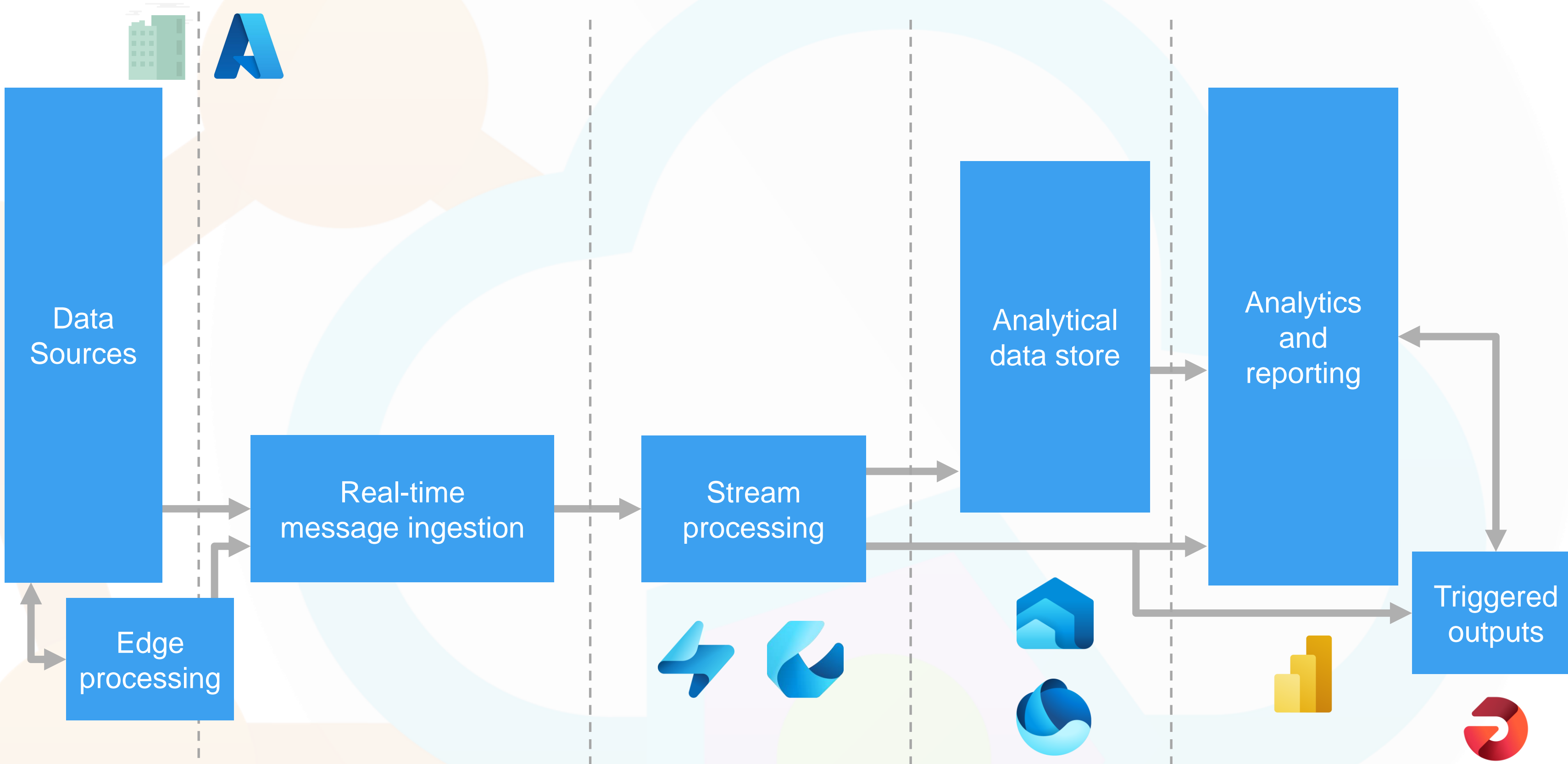
- Distributed data storage optimised for analytics and structured as Delta Lake tables.

Think... Azure Data Lake Gen2   **DELTA LAKE™**

Fabric Tooling



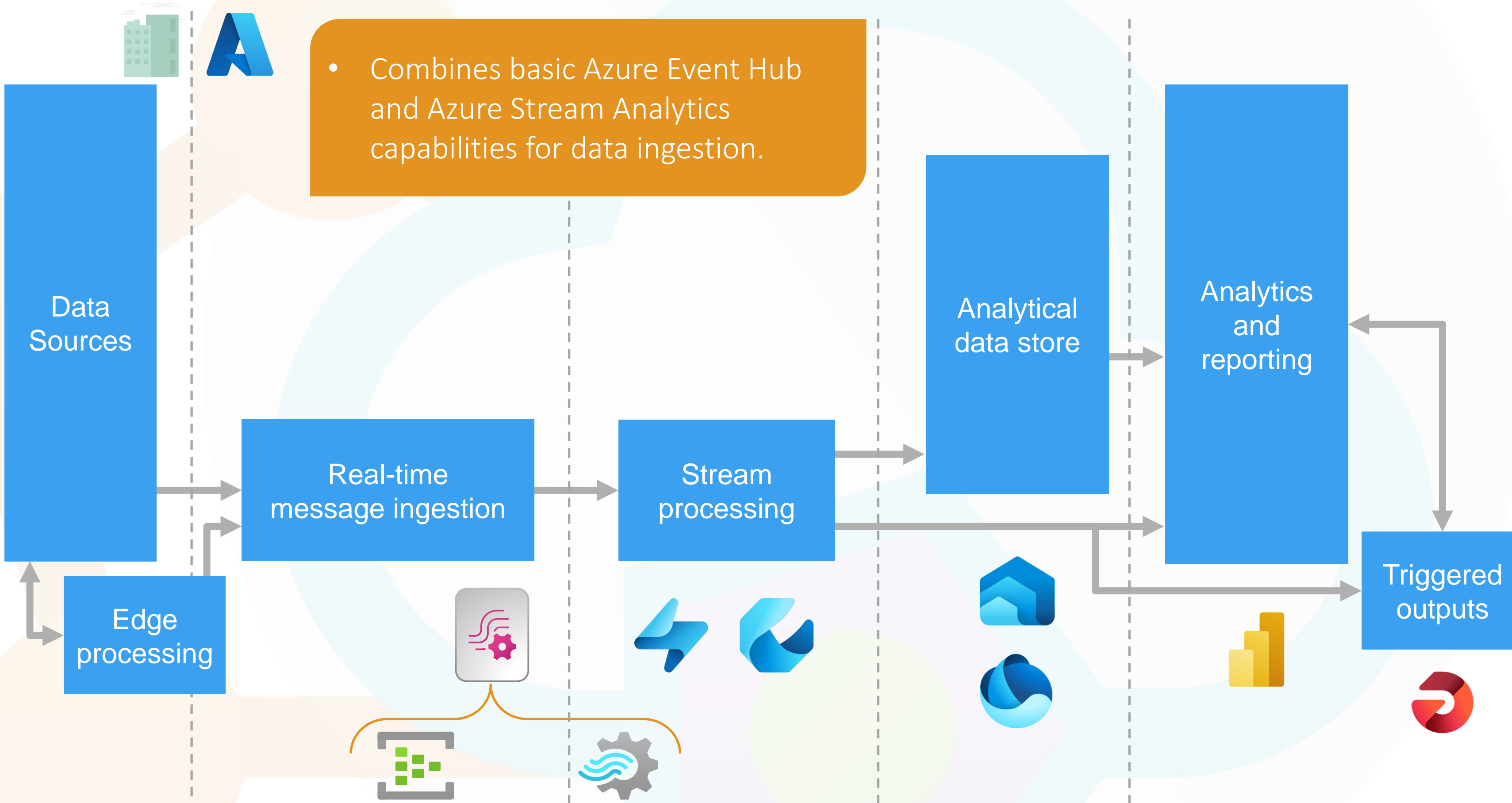
Cloud Formations - Knowledge Transfer & Training



Fabric Tooling – Event Stream



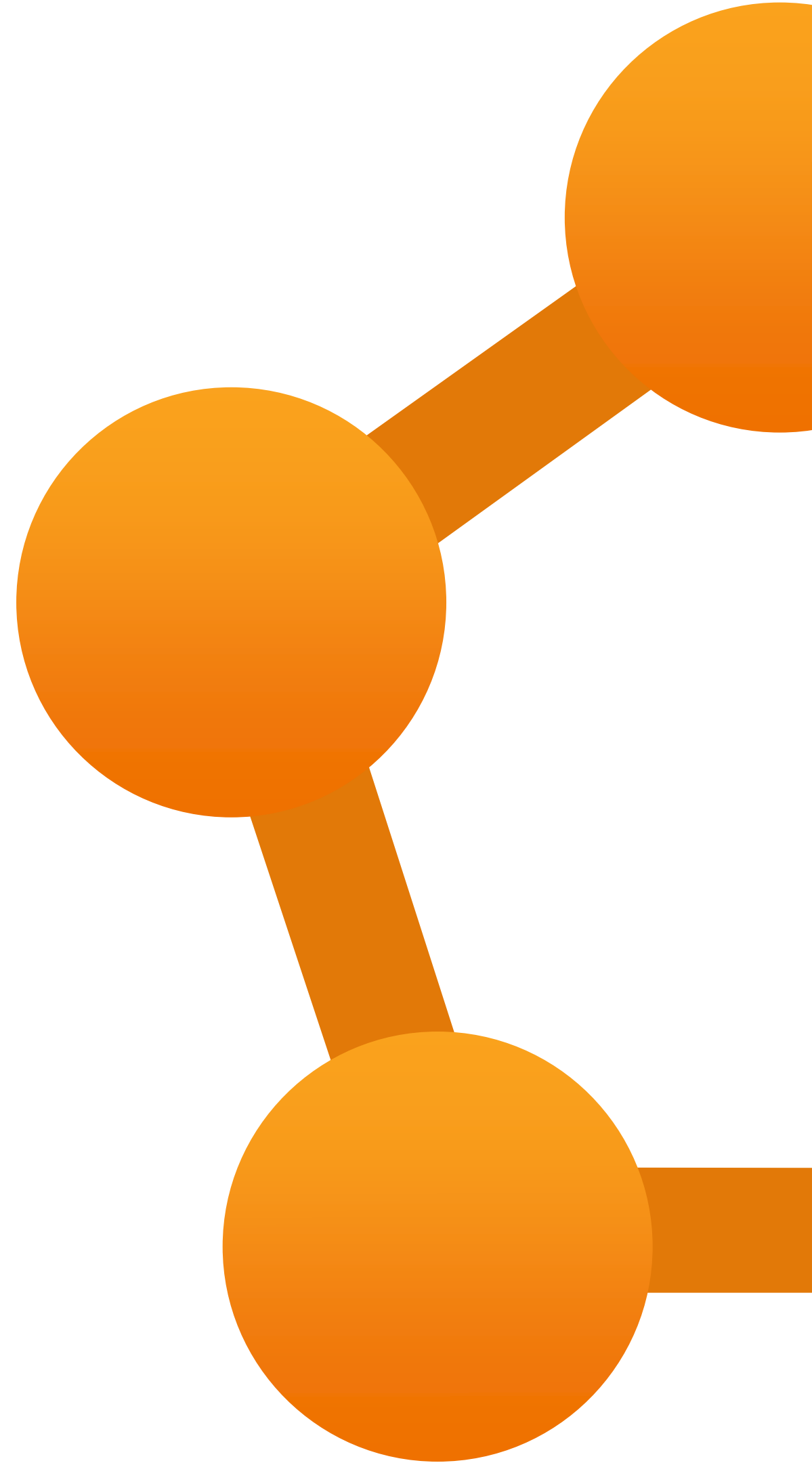
Cloud Formations - Knowledge Transfer & Training



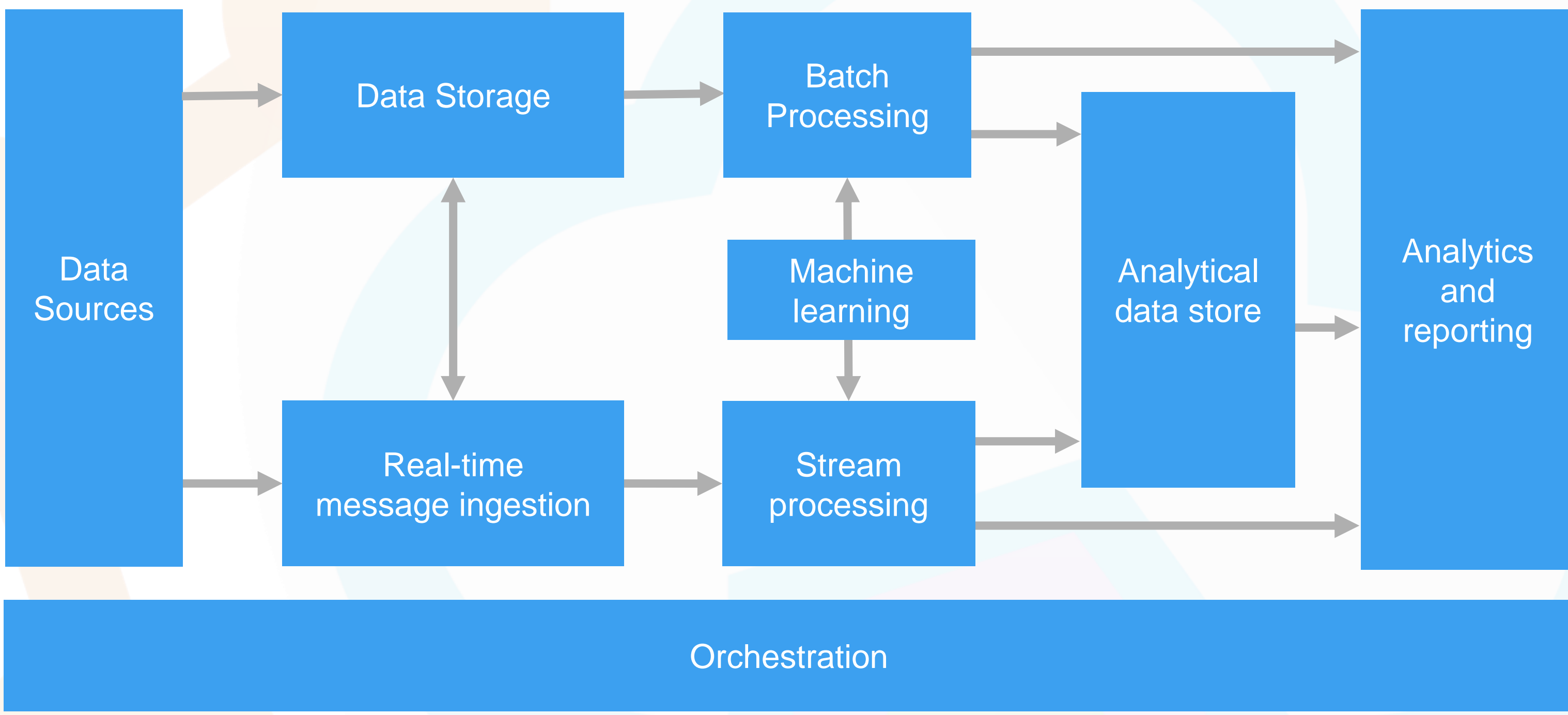
Lambda & Kappa

λ K

Cloud Formations



An Evolution Of Bringing Batch & Stream Data Together

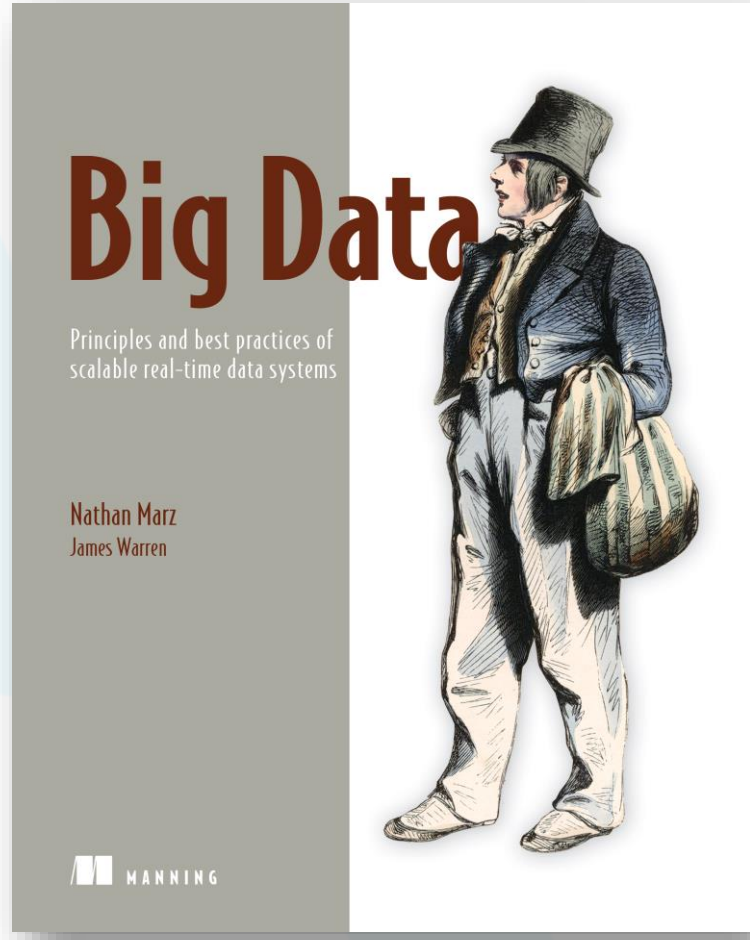
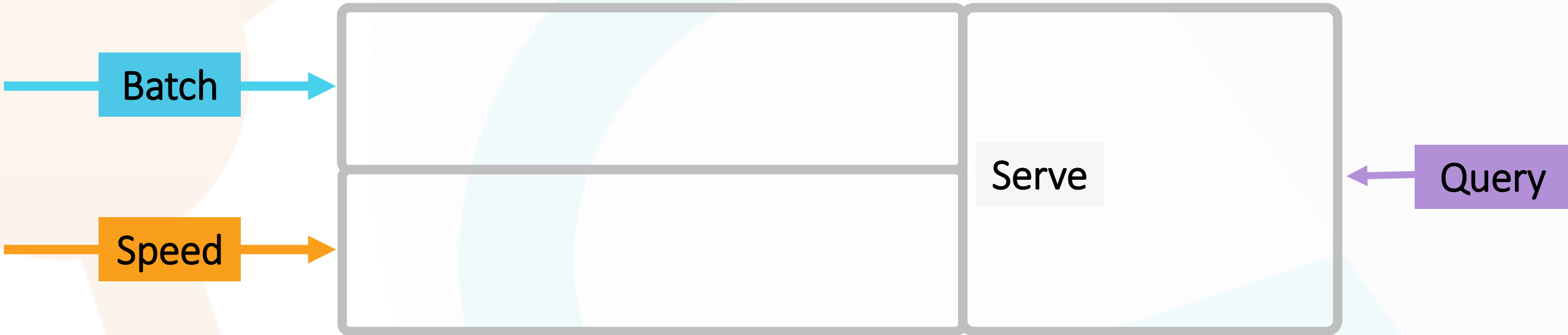


Cloud Formations - Knowledge Transfer & Training

Lambda & Kappa Architectures



Cloud Formations - Knowledge Transfer & Training

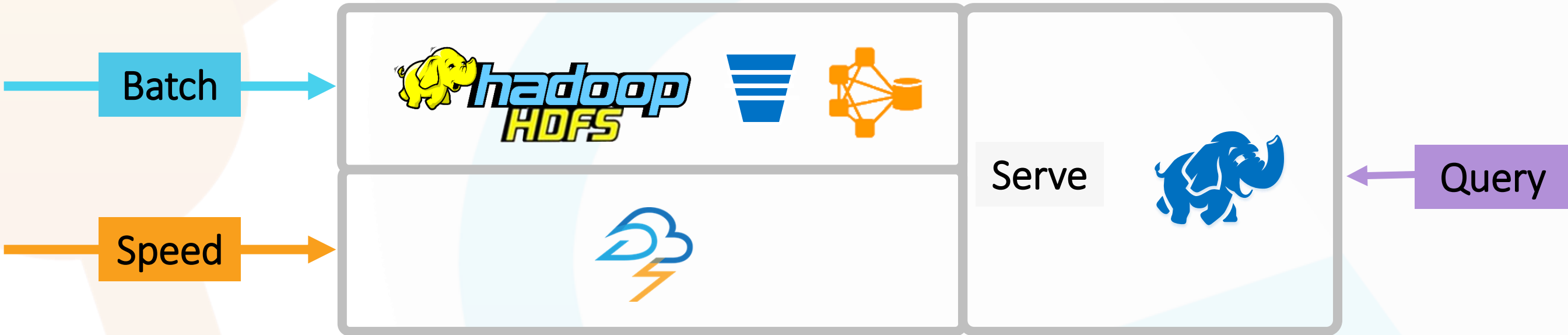


Lambda & Kappa Architectures



Cloud Formations - Knowledge Transfer & Training

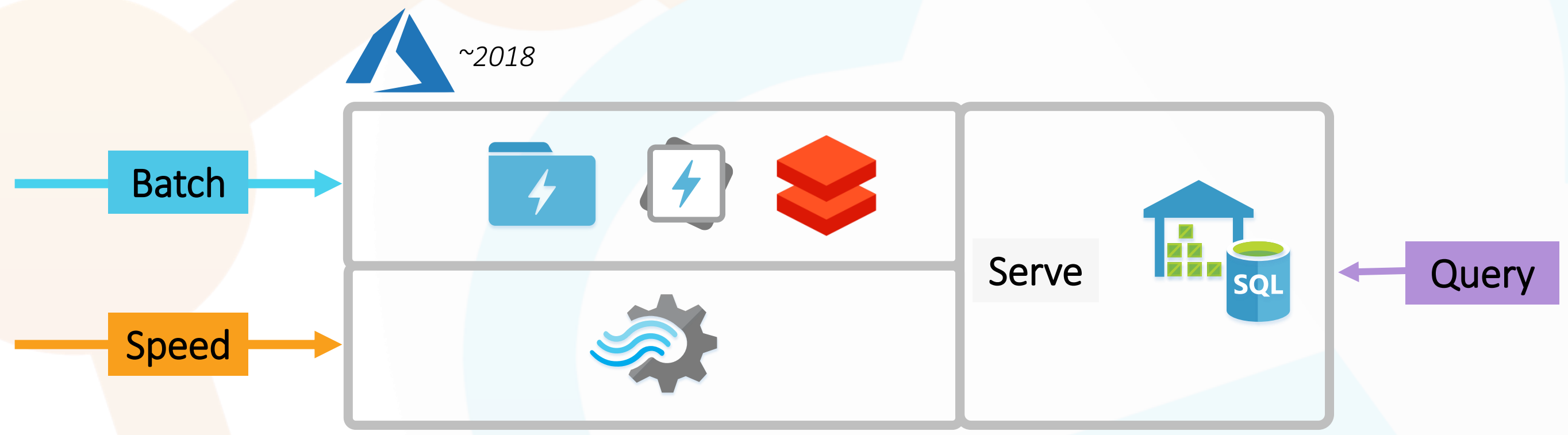
~2015



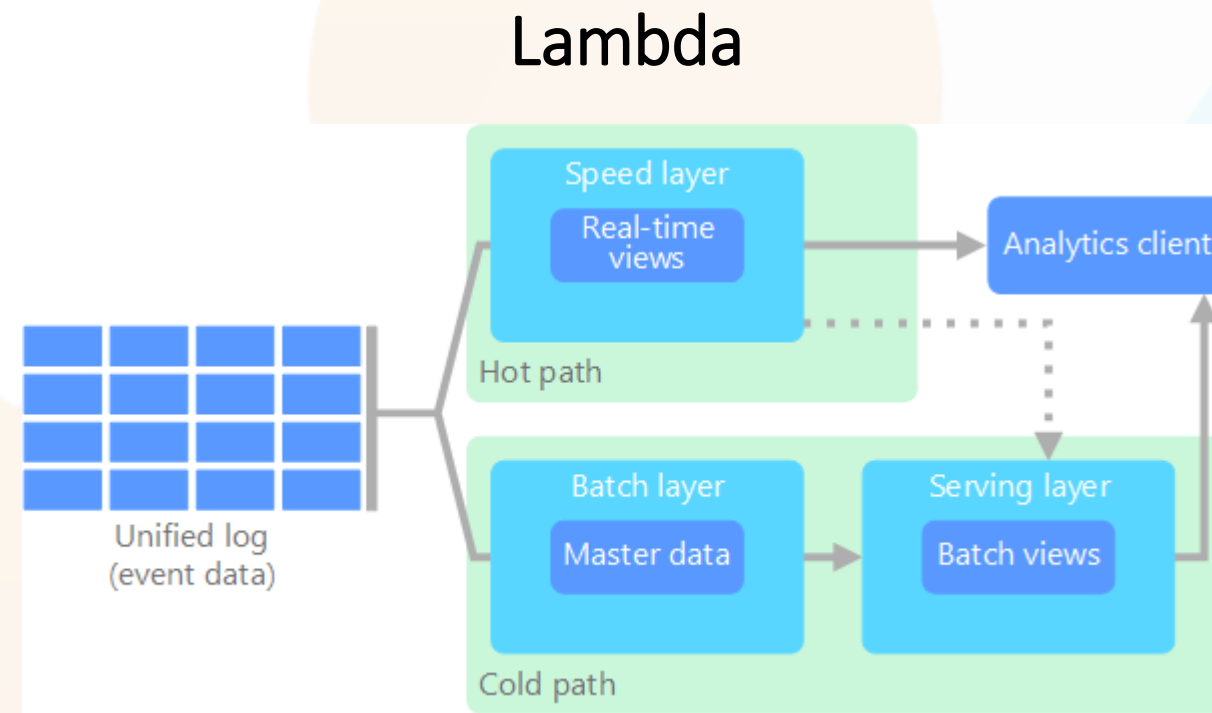
Lambda & Kappa Architectures



Cloud Formations - Knowledge Transfer & Training



Lambda & Kappa Architectures

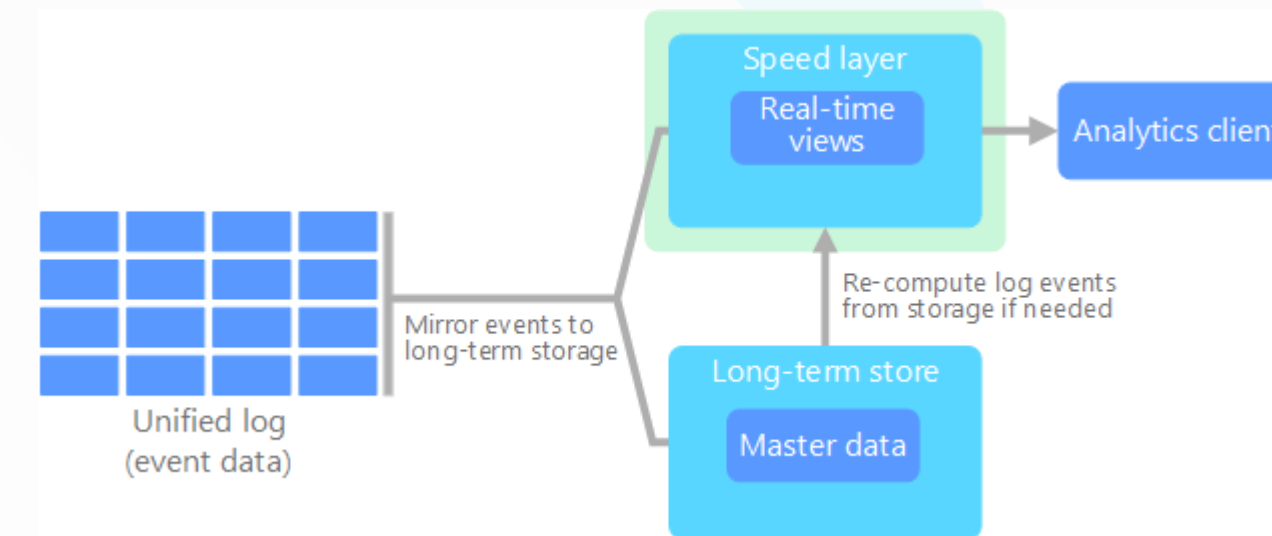


“The **lambda architecture**, first proposed by Nathan Marz, addresses this problem by creating two paths for data flow. All data coming into the system goes through these two paths:

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.”

Kappa



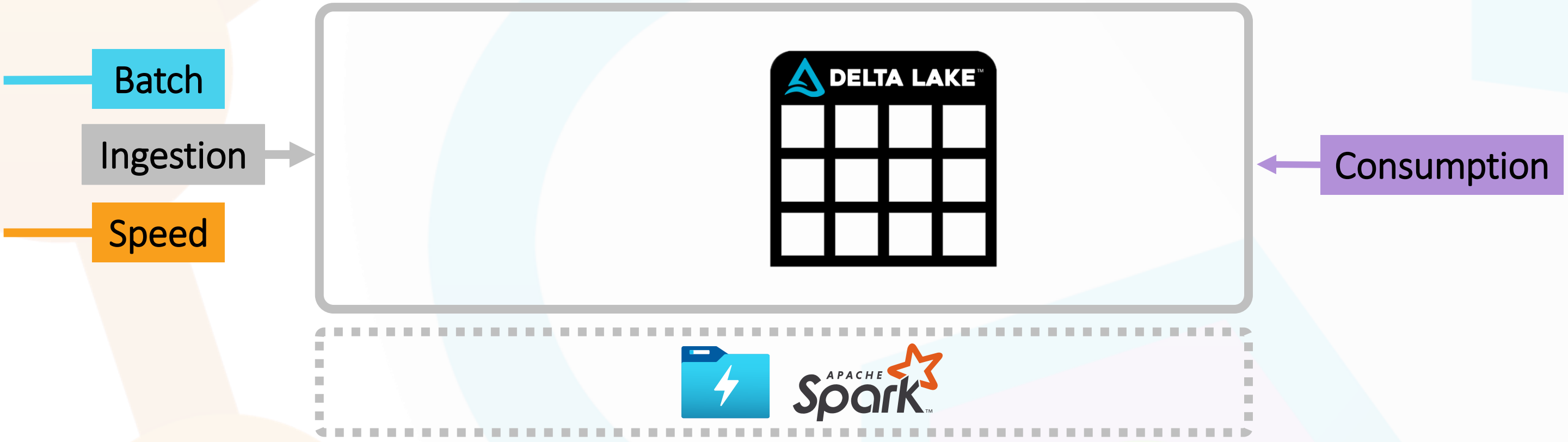
“A drawback to the lambda architecture is its **complexity**. **Processing logic appears in two different places** — the cold and hot paths — using different frameworks. This leads to duplicate computation logic and the complexity of managing the architecture for both paths.

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.”

Lambda & Kappa Architectures



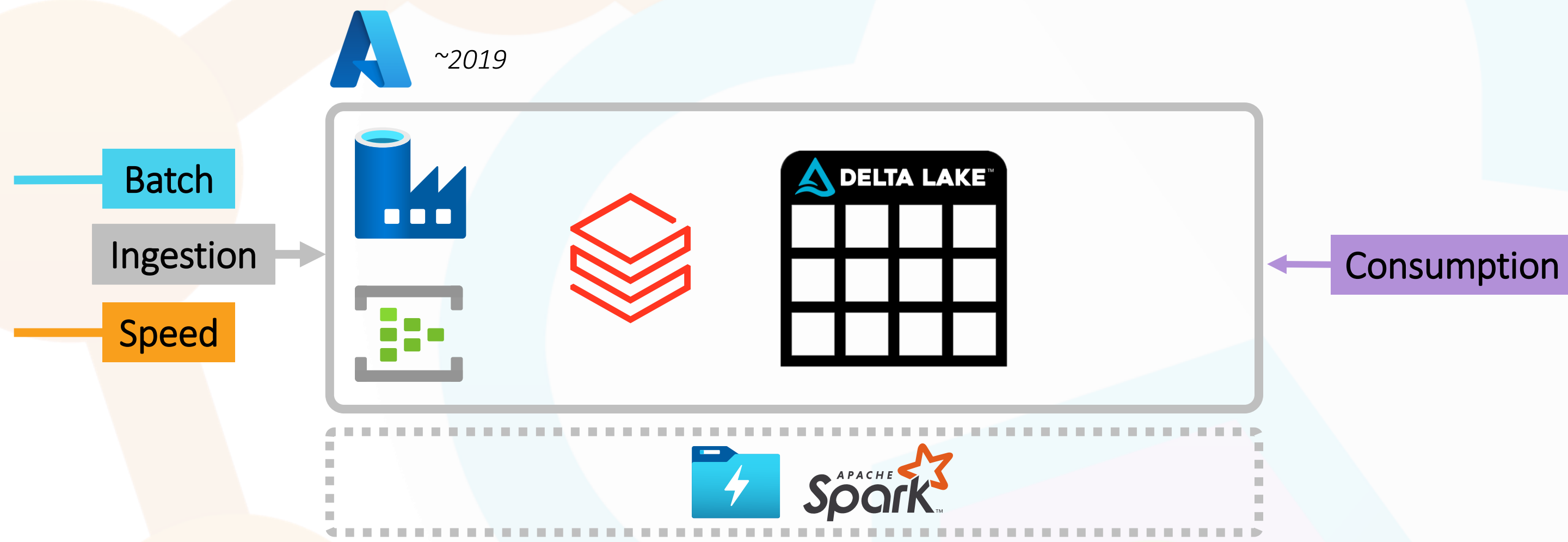
Cloud Formations - Knowledge Transfer & Training



Lambda & Kappa Architectures



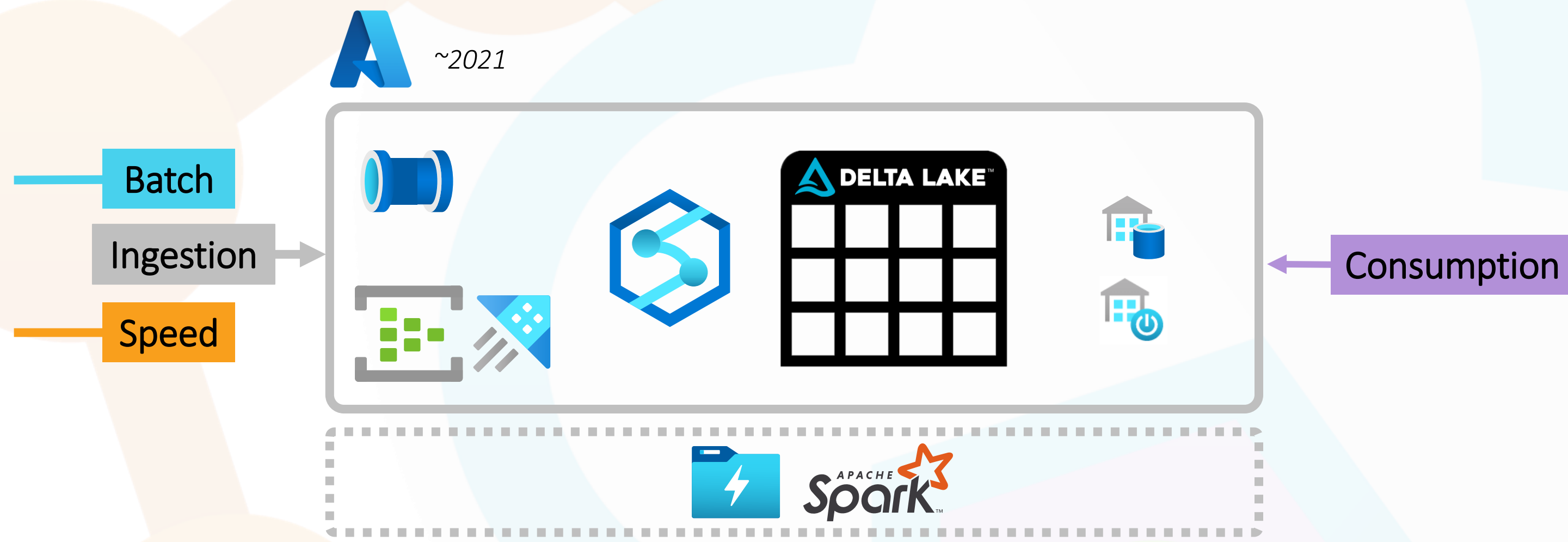
Cloud Formations - Knowledge Transfer & Training



Lambda & Kappa Architectures



Cloud Formations - Knowledge Transfer & Training

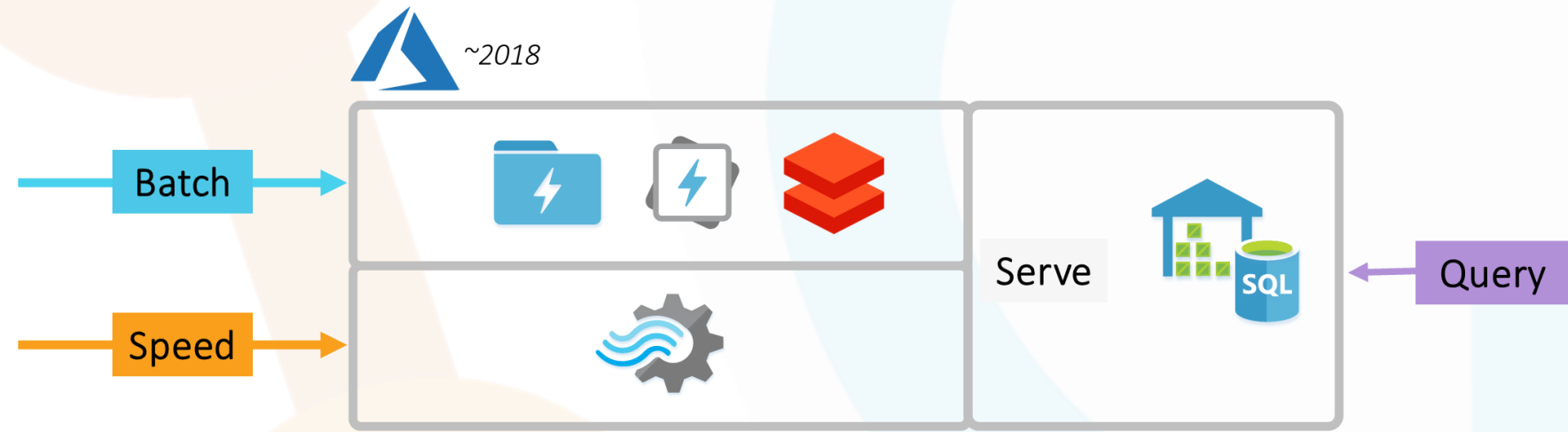


Lambda & Kappa Architectures vs Technology

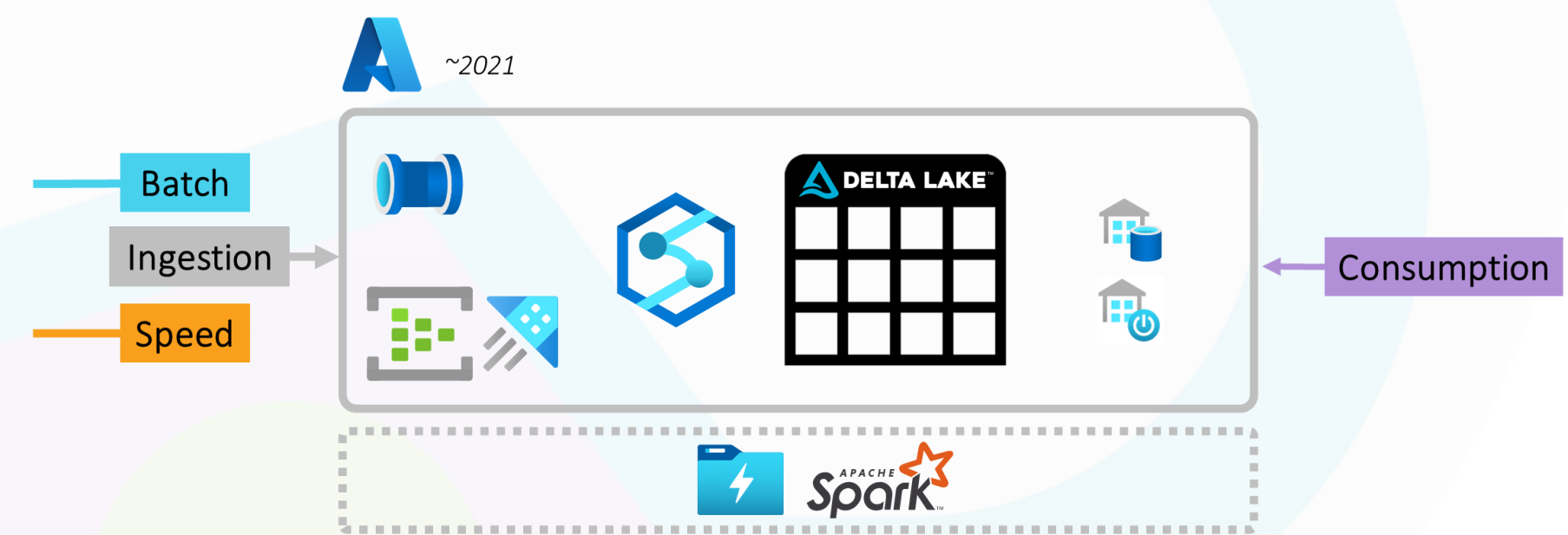
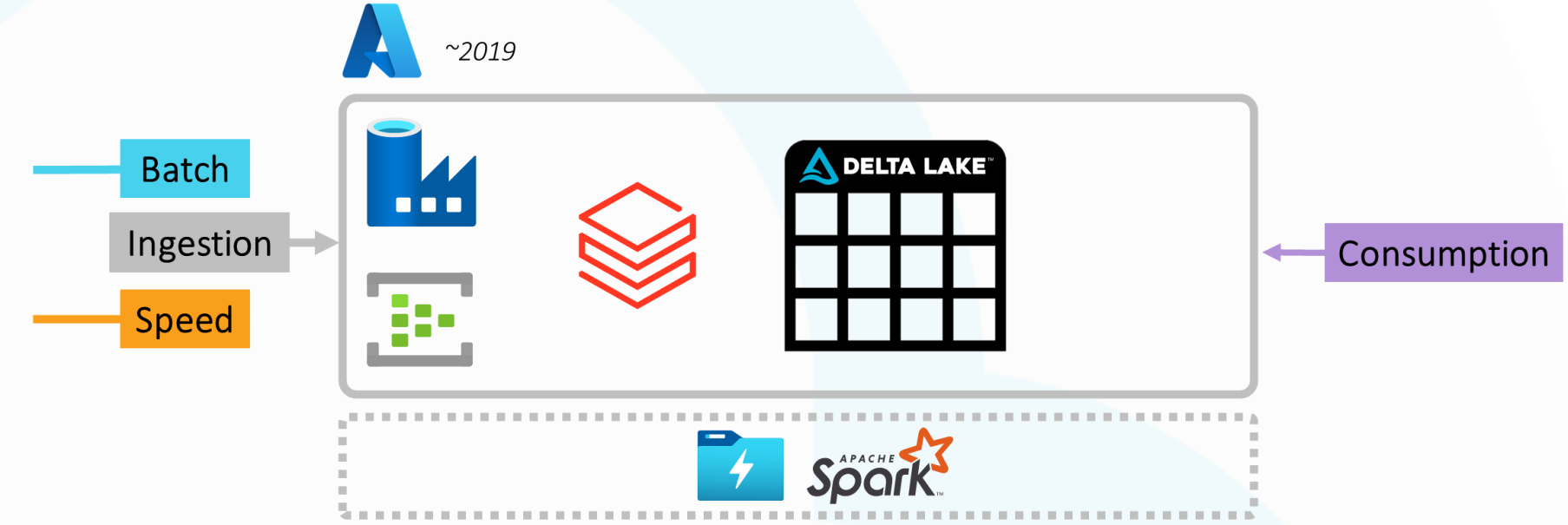


Cloud Formations - Knowledge Transfer & Training

Lambda



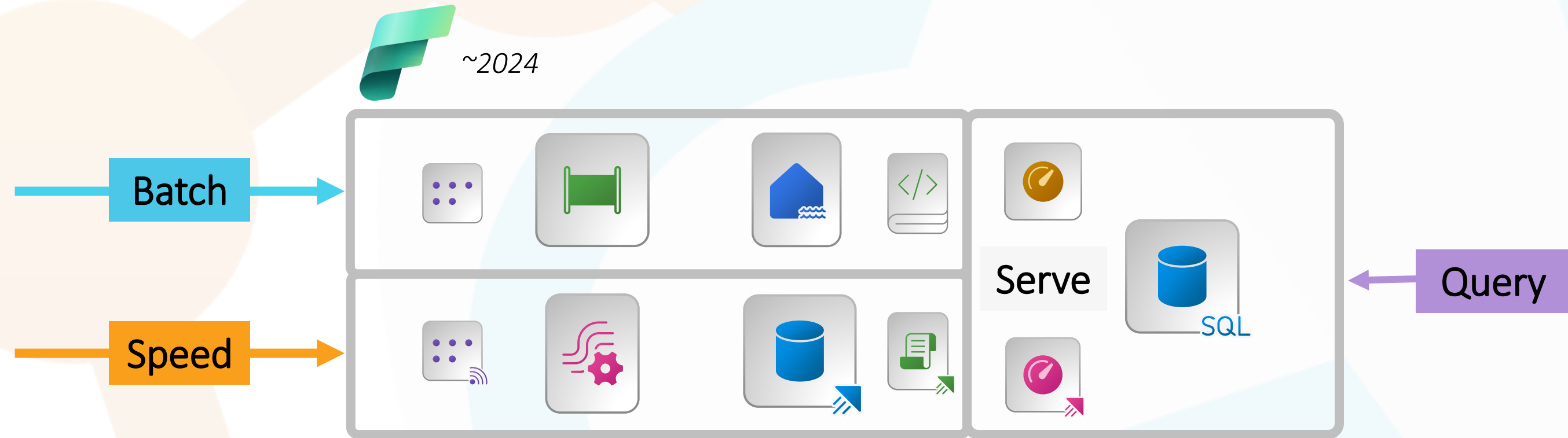
Kappa



Microsoft Fabric vs a Kappa Architecture



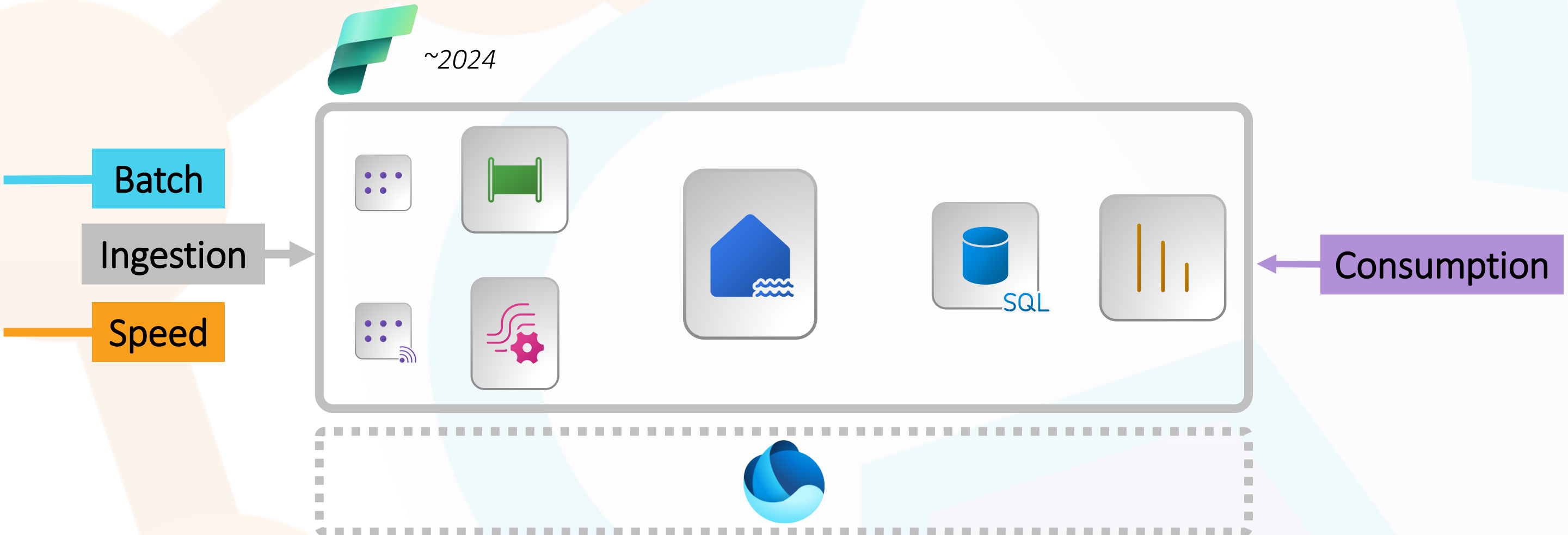
Cloud Formations - Knowledge Transfer & Training



Microsoft Fabric vs a Kappa Architecture



Cloud Formations - Knowledge Transfer & Training

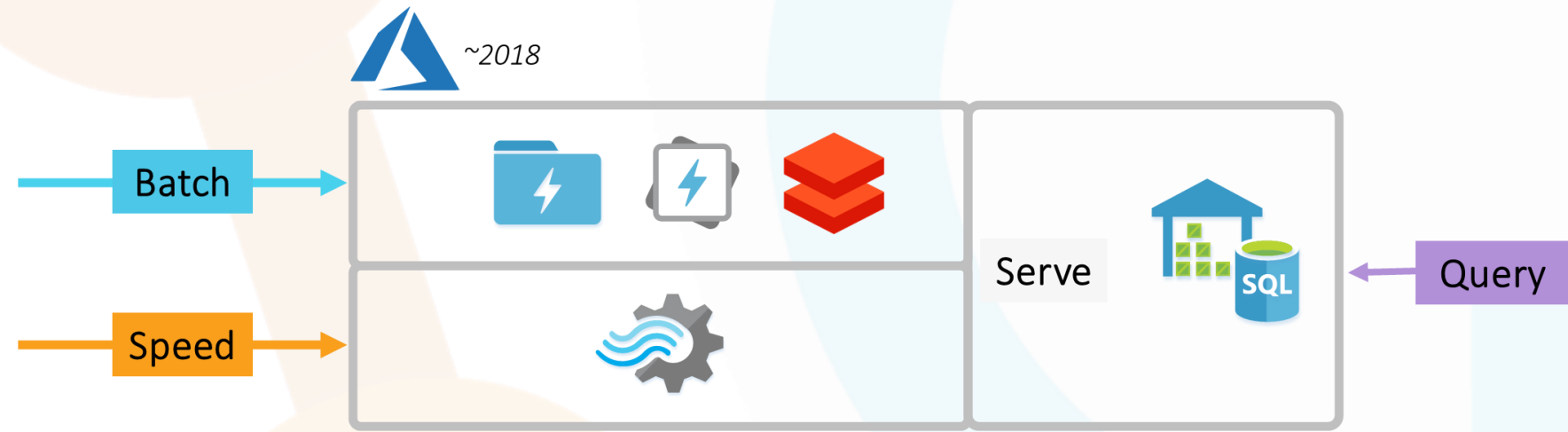


Lambda & Kappa Architectures vs Technology

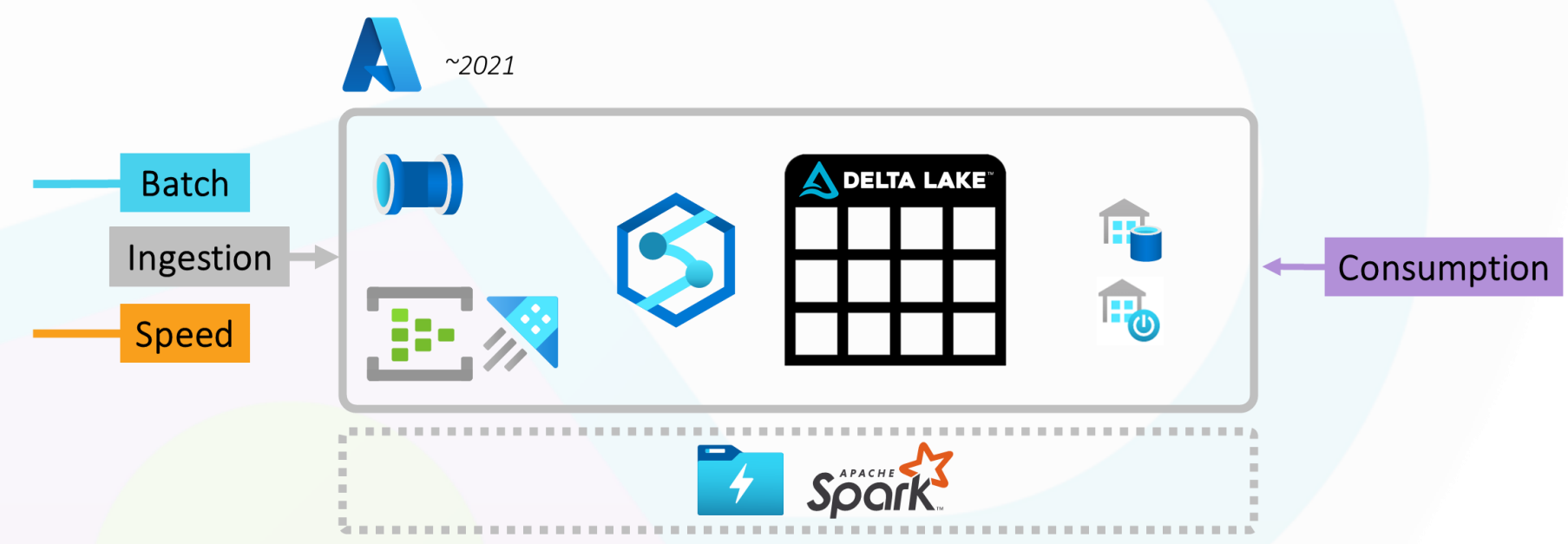
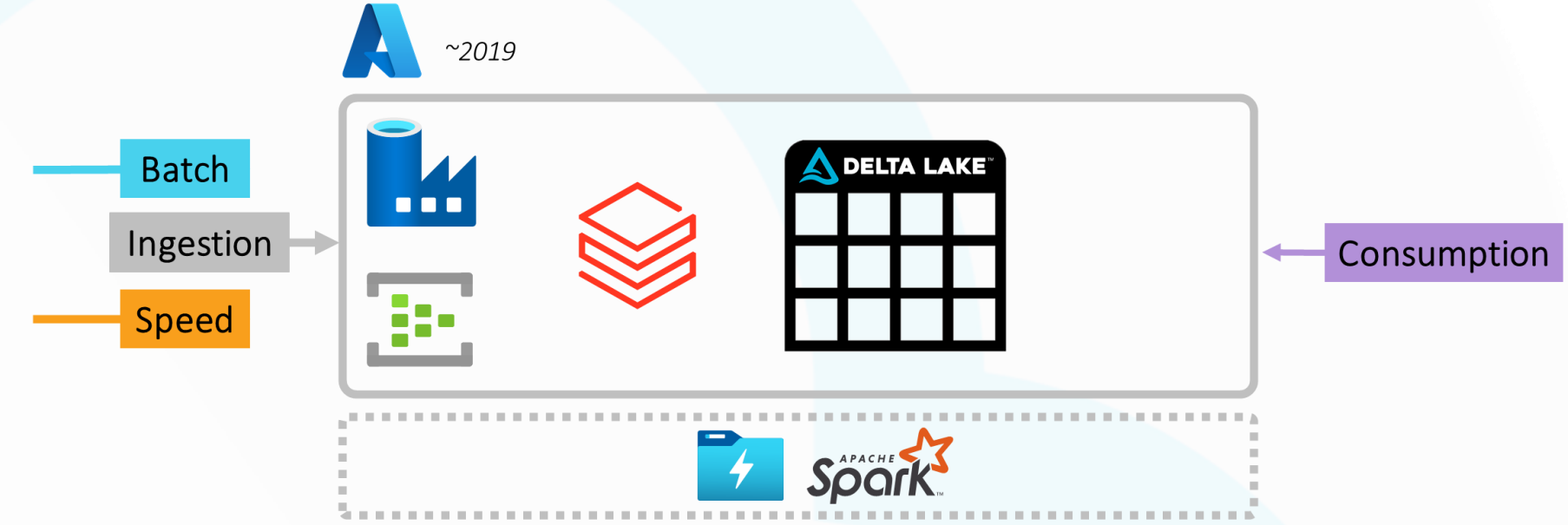


Cloud Formations - Knowledge Transfer & Training

Lambda



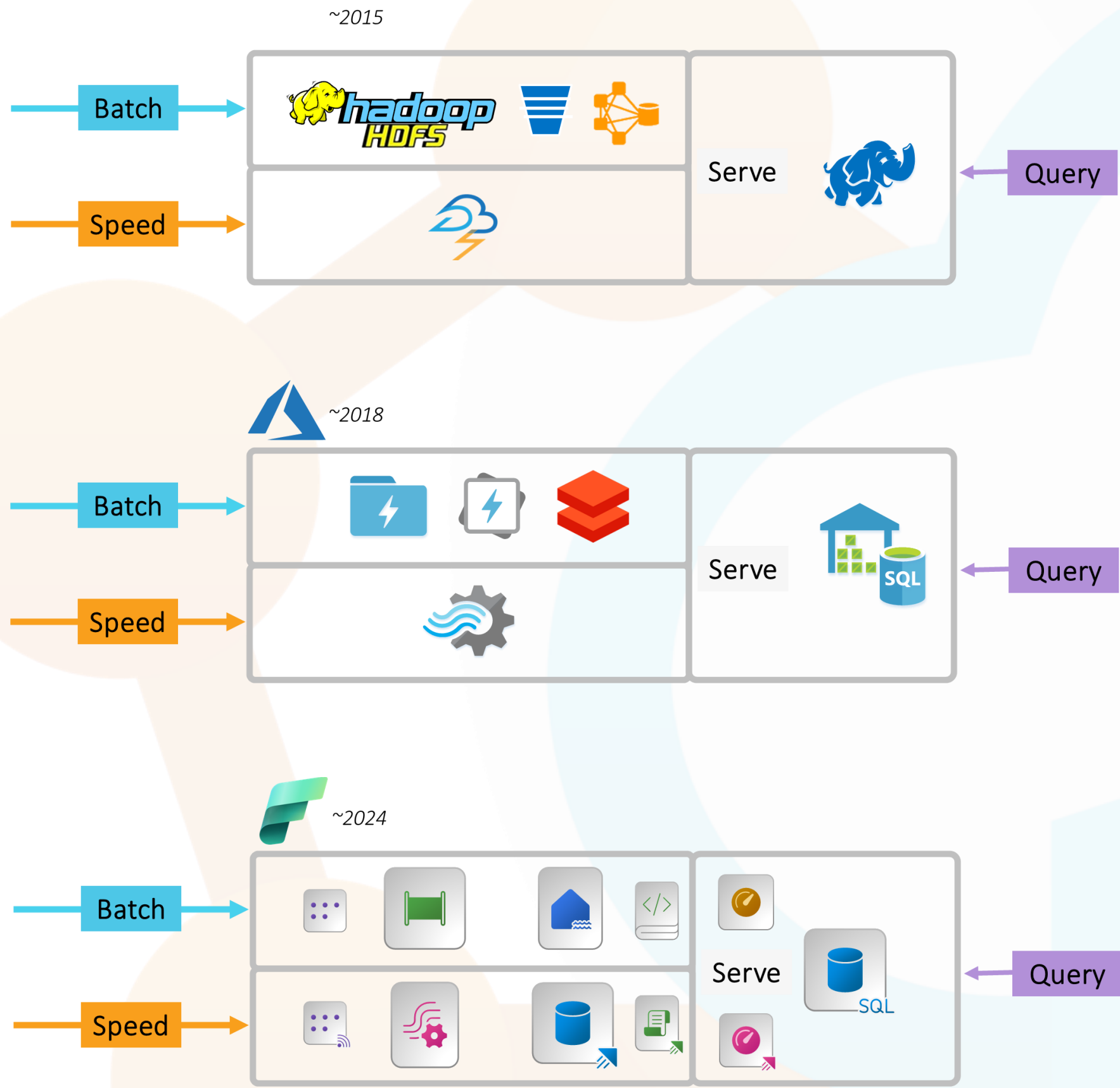
Kappa



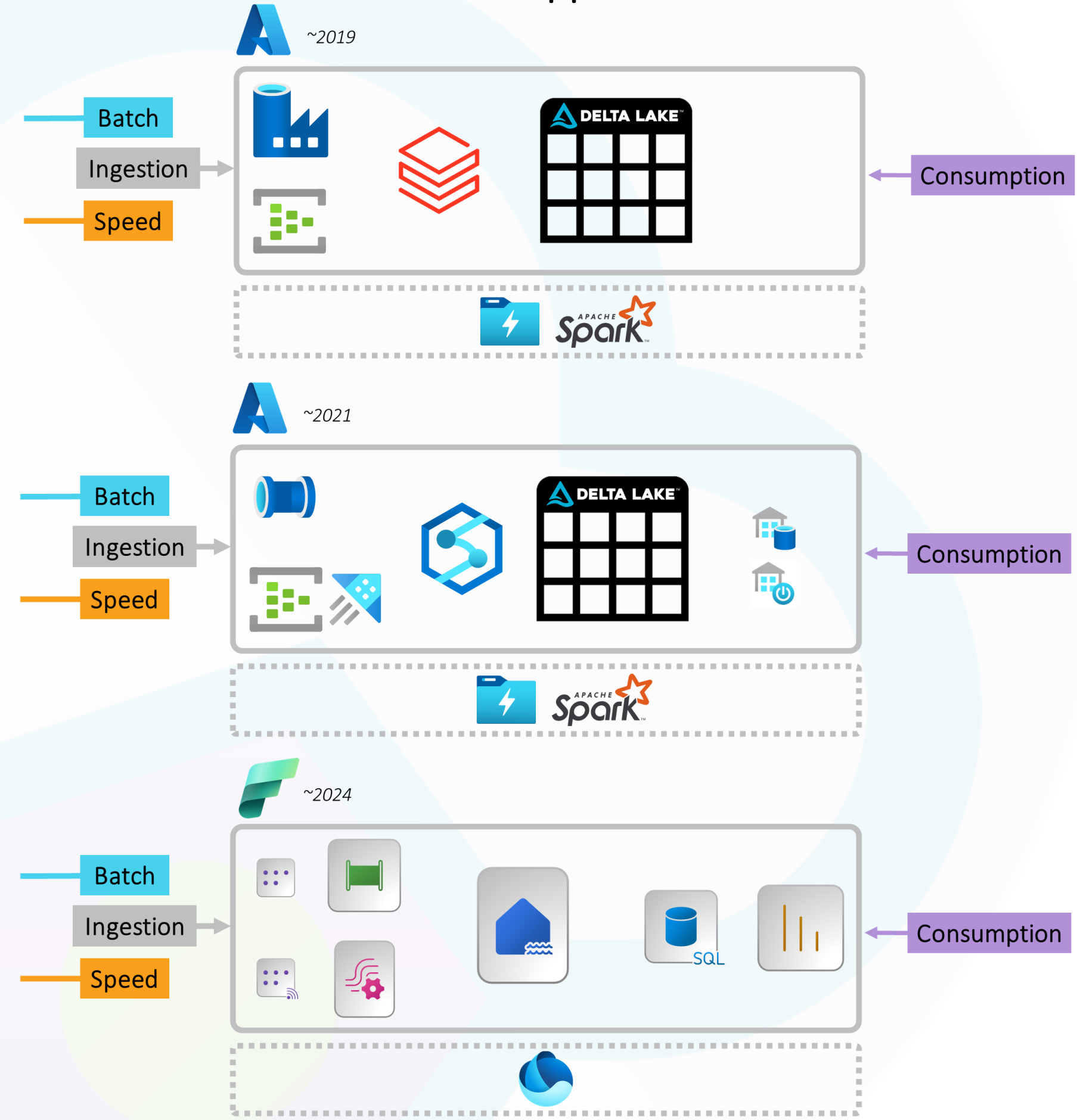
Lambda & Kappa Architectures vs Technology



Lambda



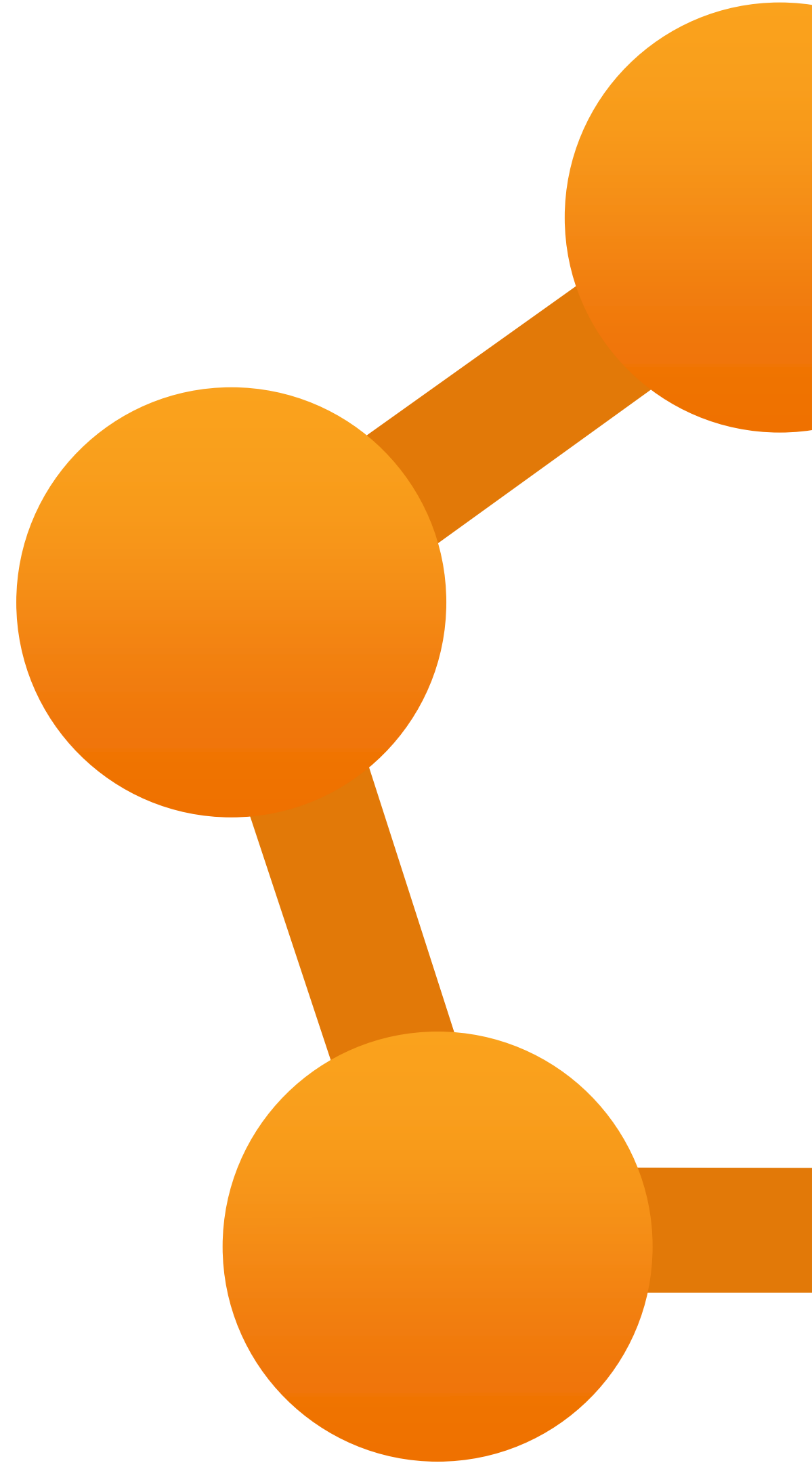
Kappa



Cloud Formations - Knowledge Transfer & Training

Let's Build Something!

Cloud Formations



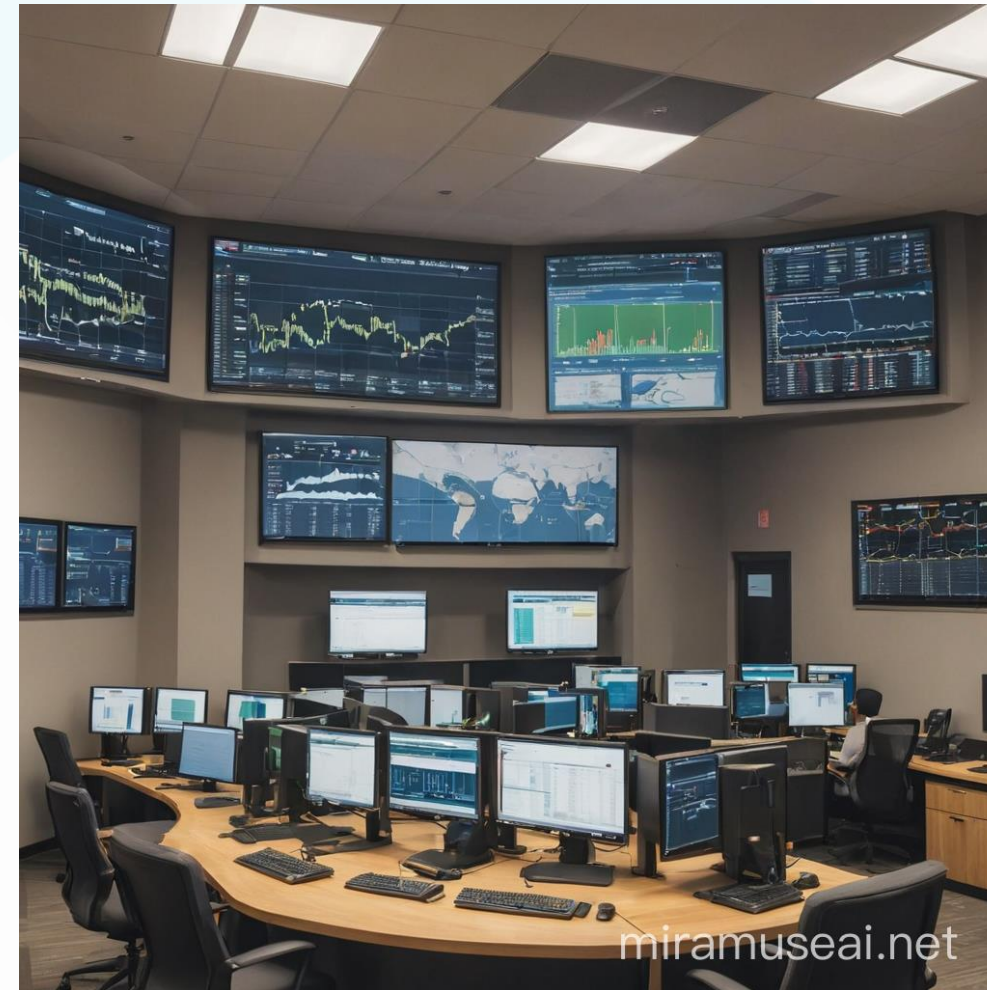
Our Use Case



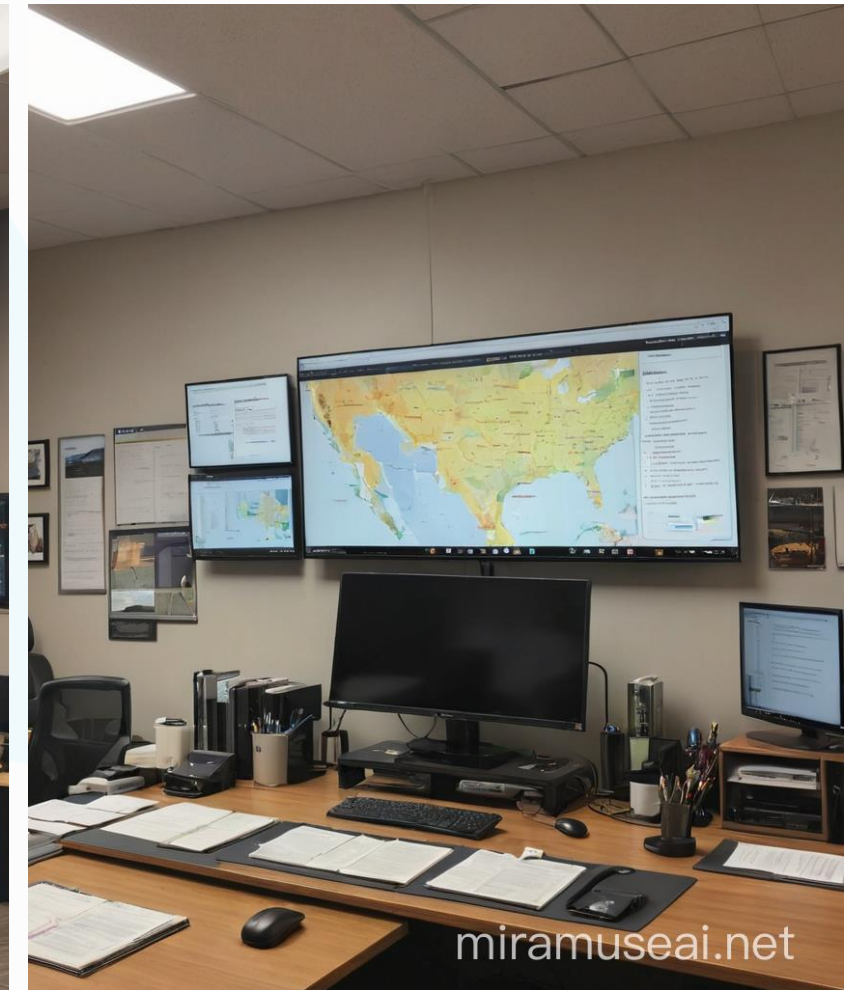
miramuseai.net



miramuseai.net



miramuseai.net



miramuseai.net

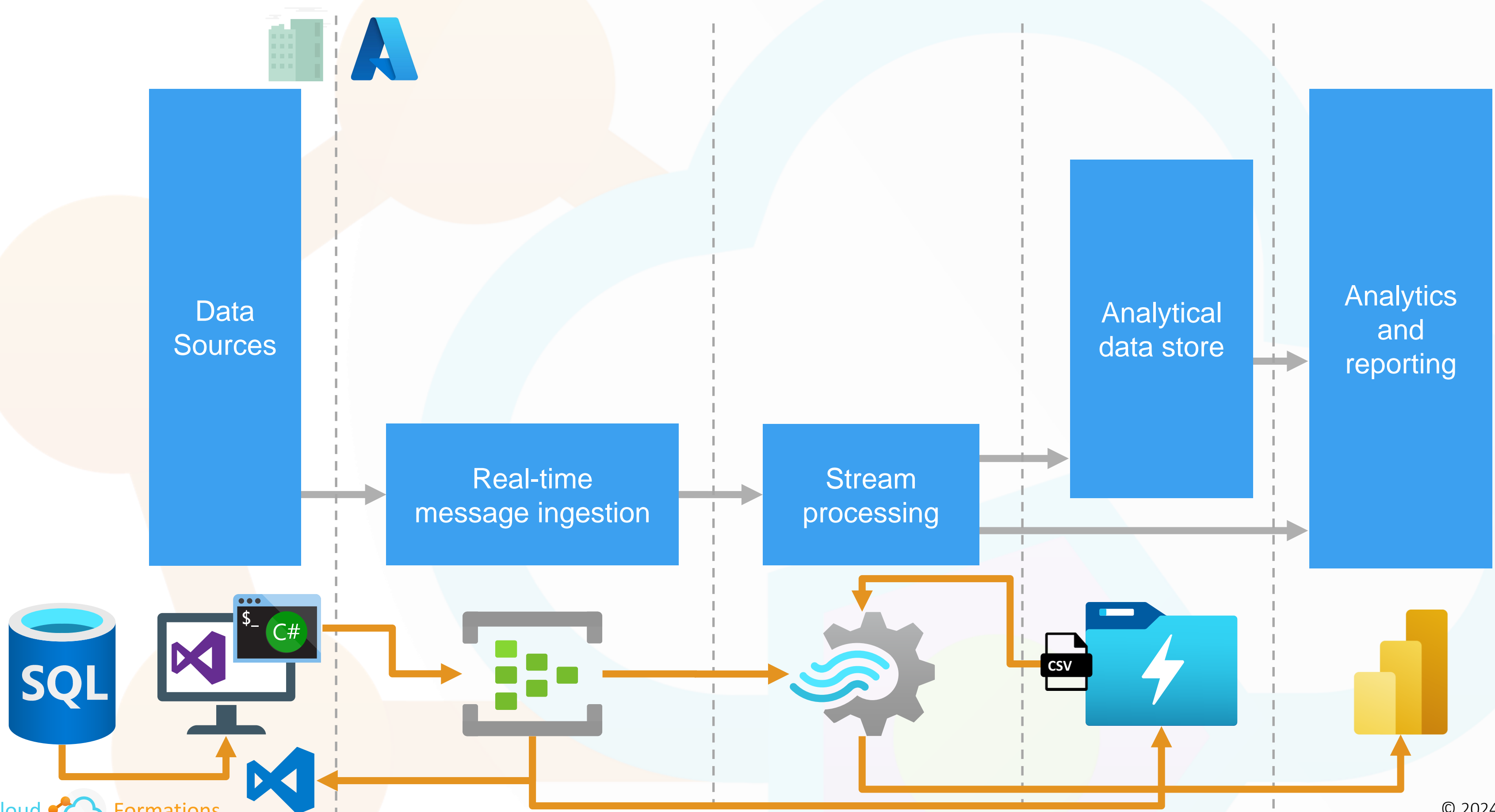
A bike shop front with large glass windows showing bike related products on the shelves inside. The shop front has a sign with the name Adventure Works.

A bike shop called Adventure Works, standing inside looking at the cashier which includes a modern point of sale till system. On the wall behind the cashier are bike related products.

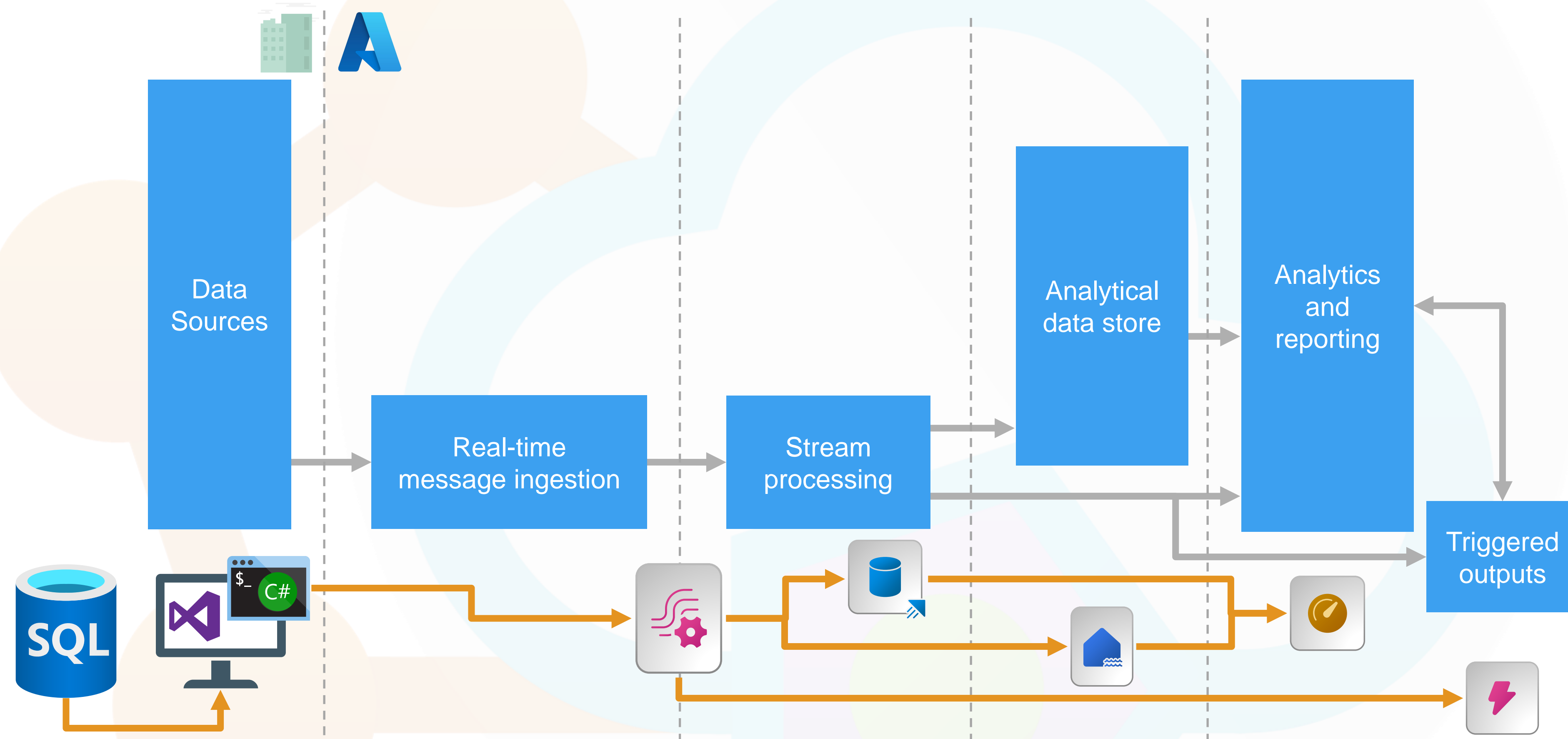
At the corporate head offices of Adventure Works standing in the main operations room. On the wall are large televisions showing a range of data analytics dashboards with charts and technical information.

Inside the office of the CEO at the company Adventure Works, on the desk is a business plan to role out more retail stores across the country based on targeted growth on an analytics dashboard visible on a large TV.

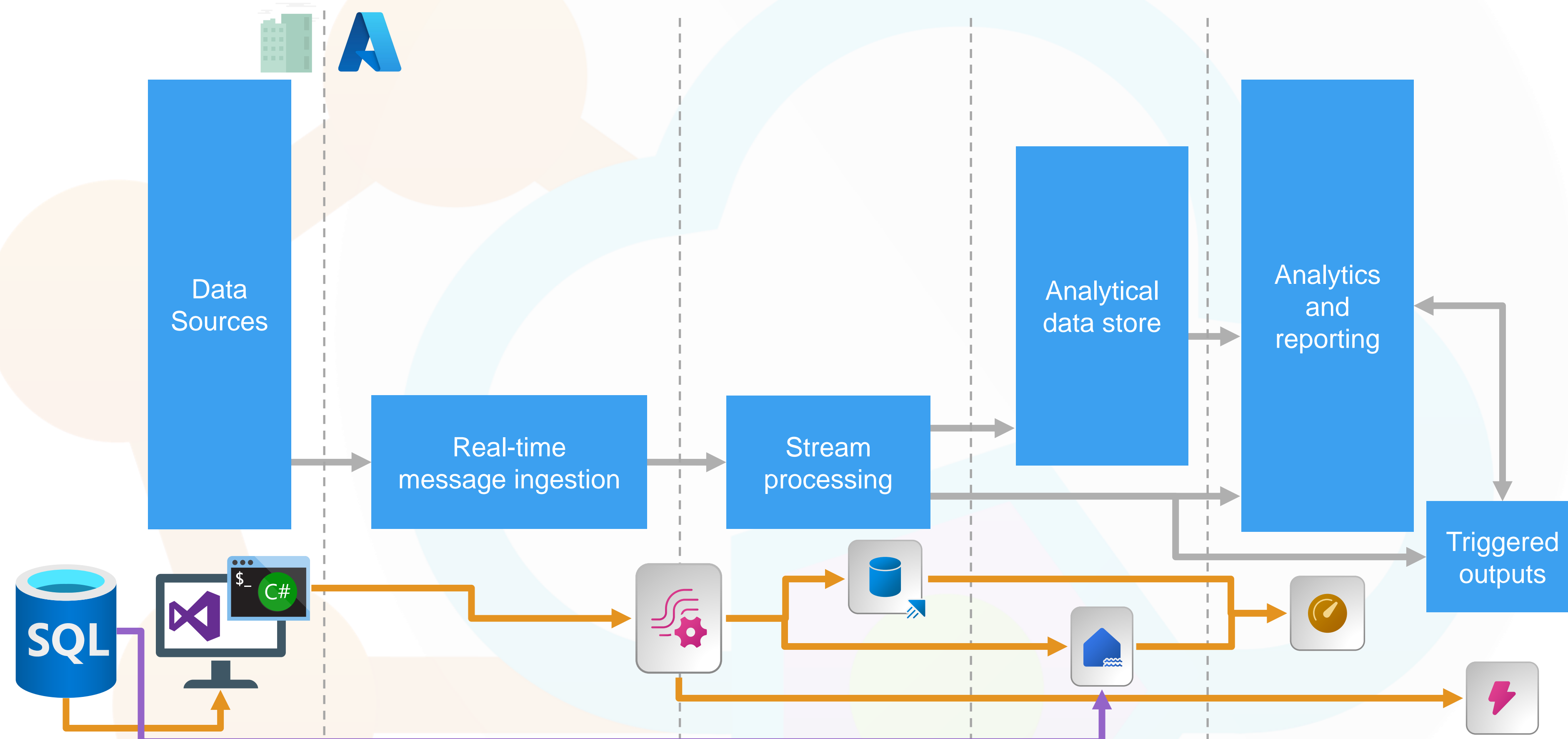
Azure Real-Time Data Handling



Fabric Real-Time Data Handling

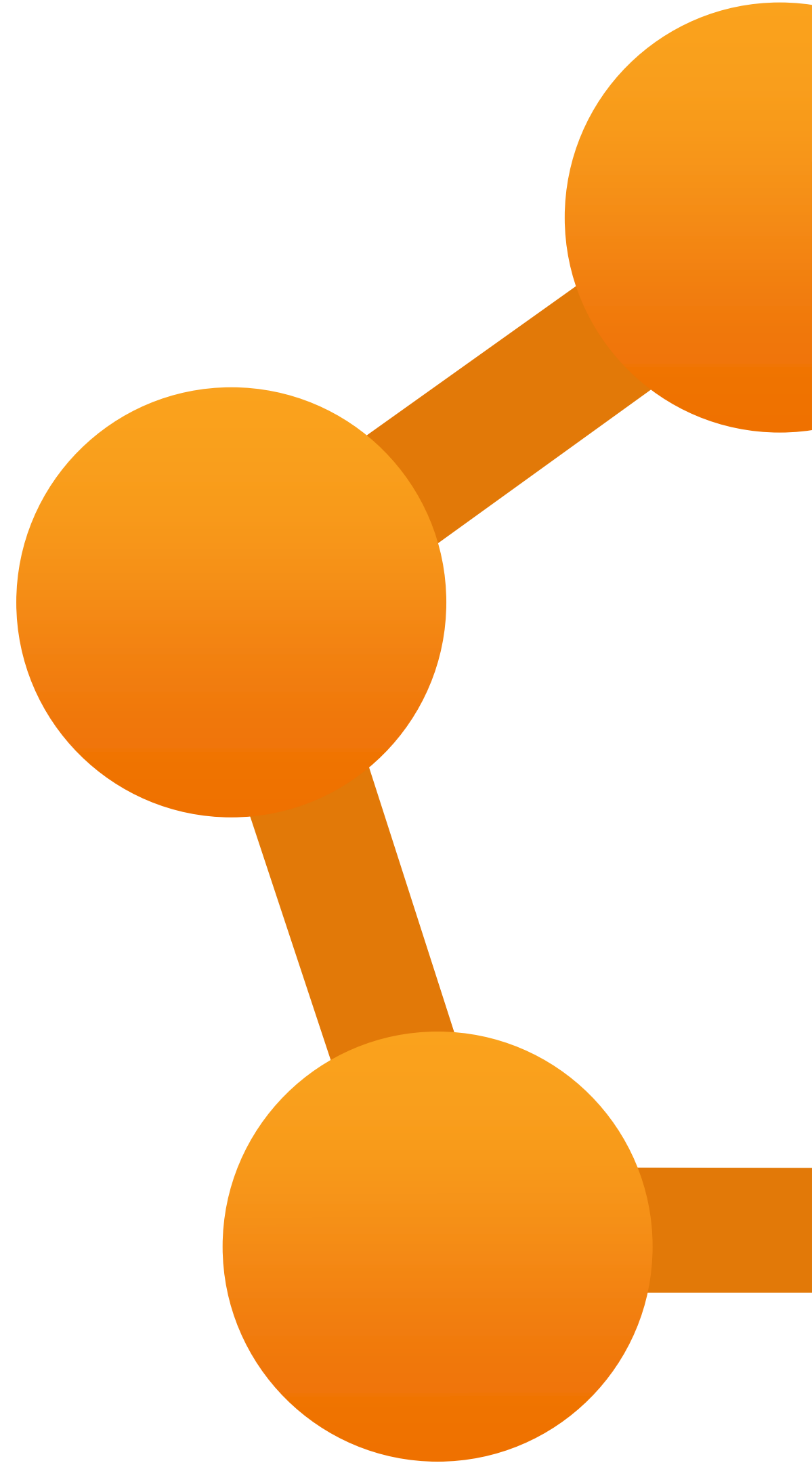


Fabric Real-Time Data Handling (Mirroring)



Conclusions

Cloud Formations



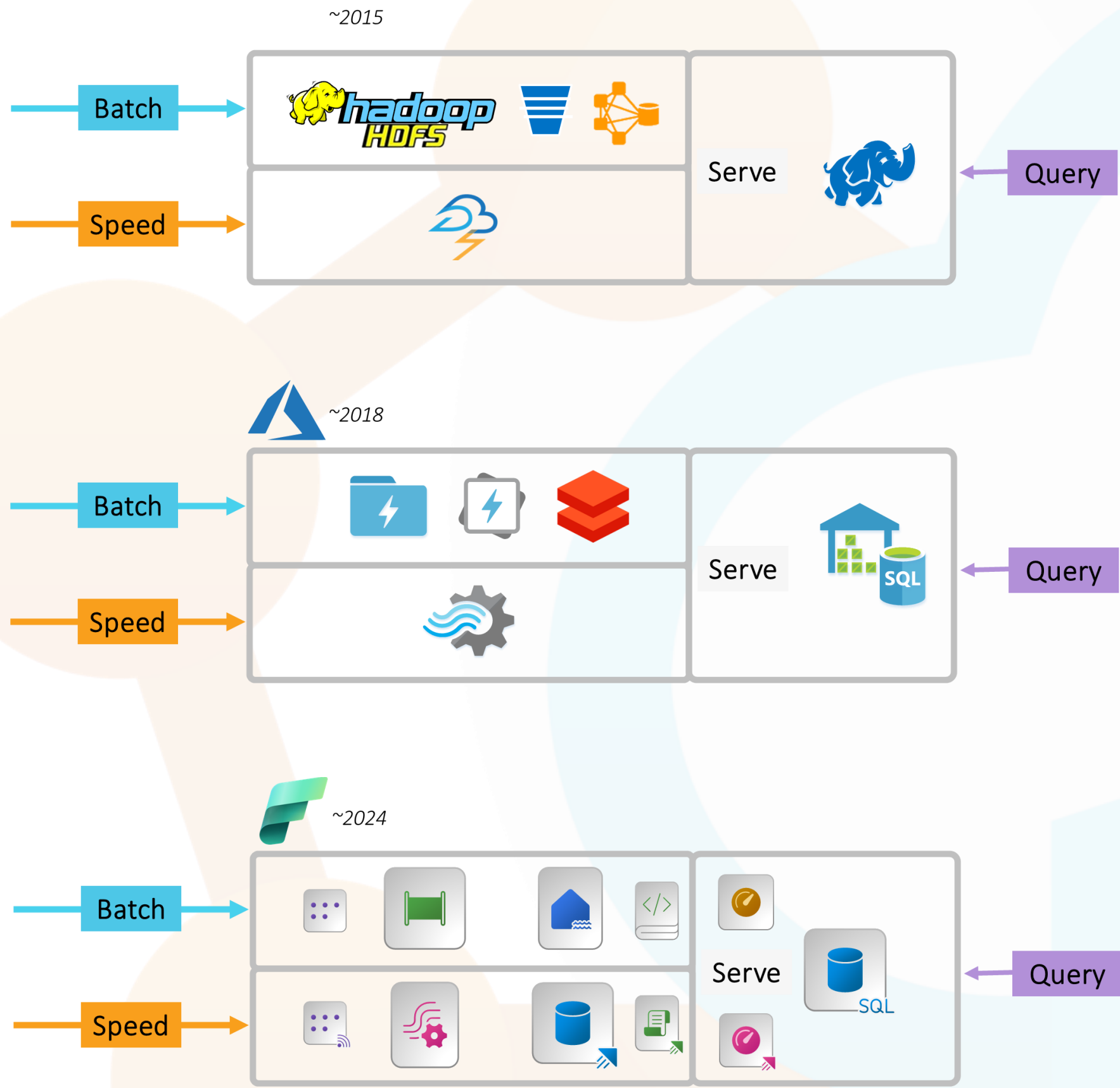
Glossary

Term	Definition
Big Data	Any data that you cannot process in the time that you have/want using the technology you have.
Real-time Data	Delivering data from the producer to consumer as fast as possible using the technology you have.
Near Real-time Data	Delivering data from the producer to consumer within 1 minute of it being created.
Data Stream	Data that is constantly flowing from producer to consumer in near real-time.

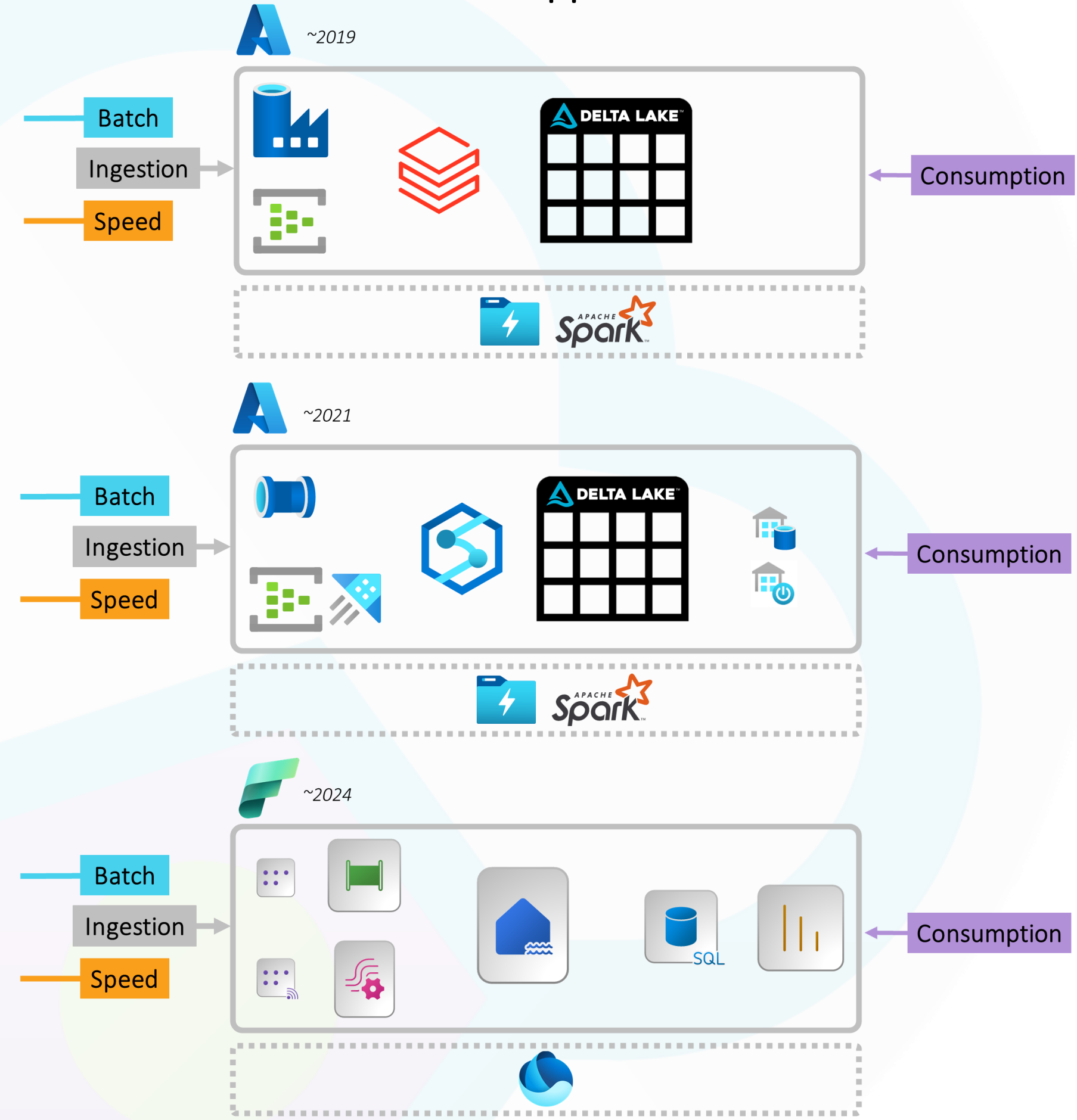
Lambda & Kappa Architectures vs Technology



Lambda



Kappa

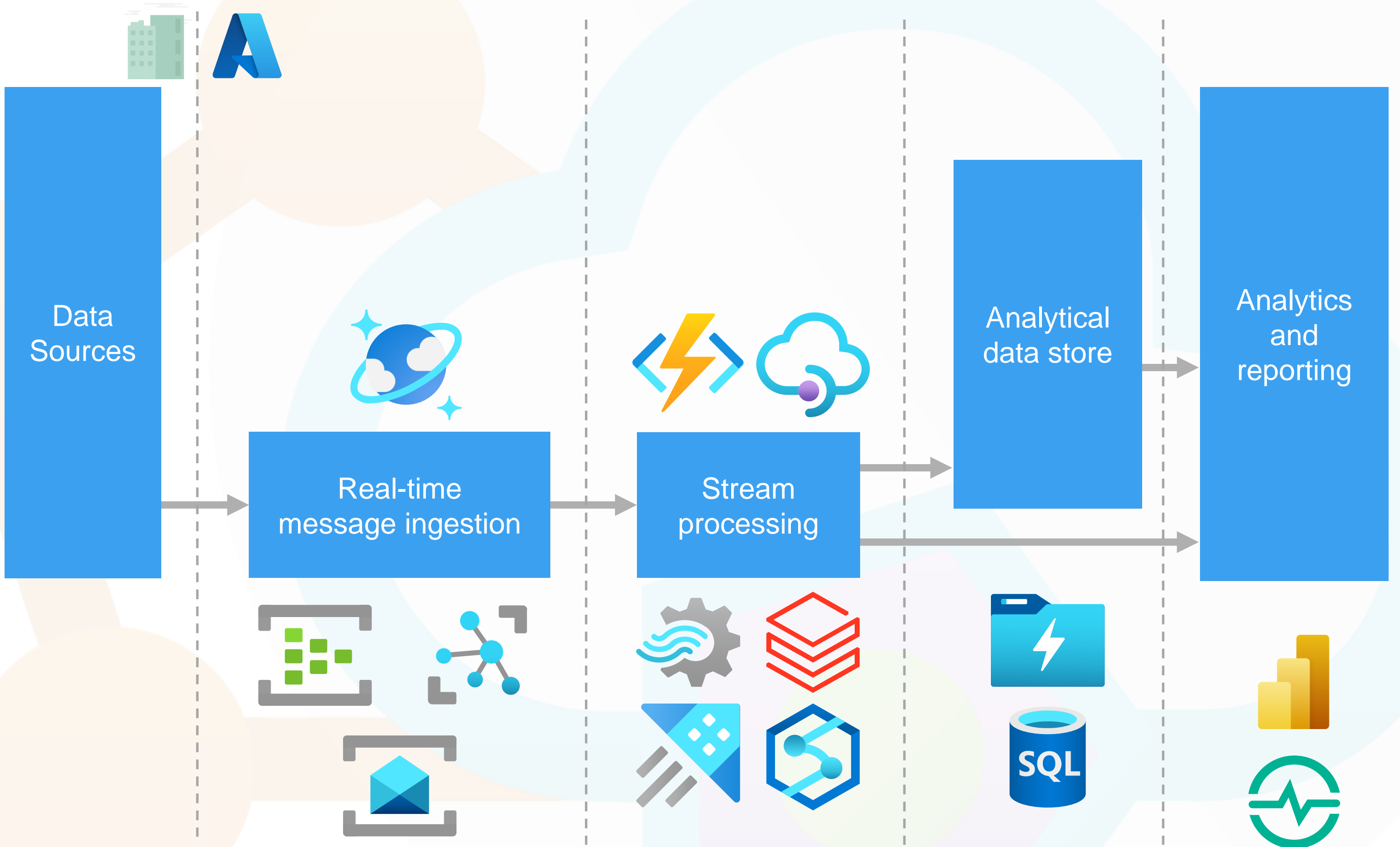


Cloud Formations - Knowledge Transfer & Training

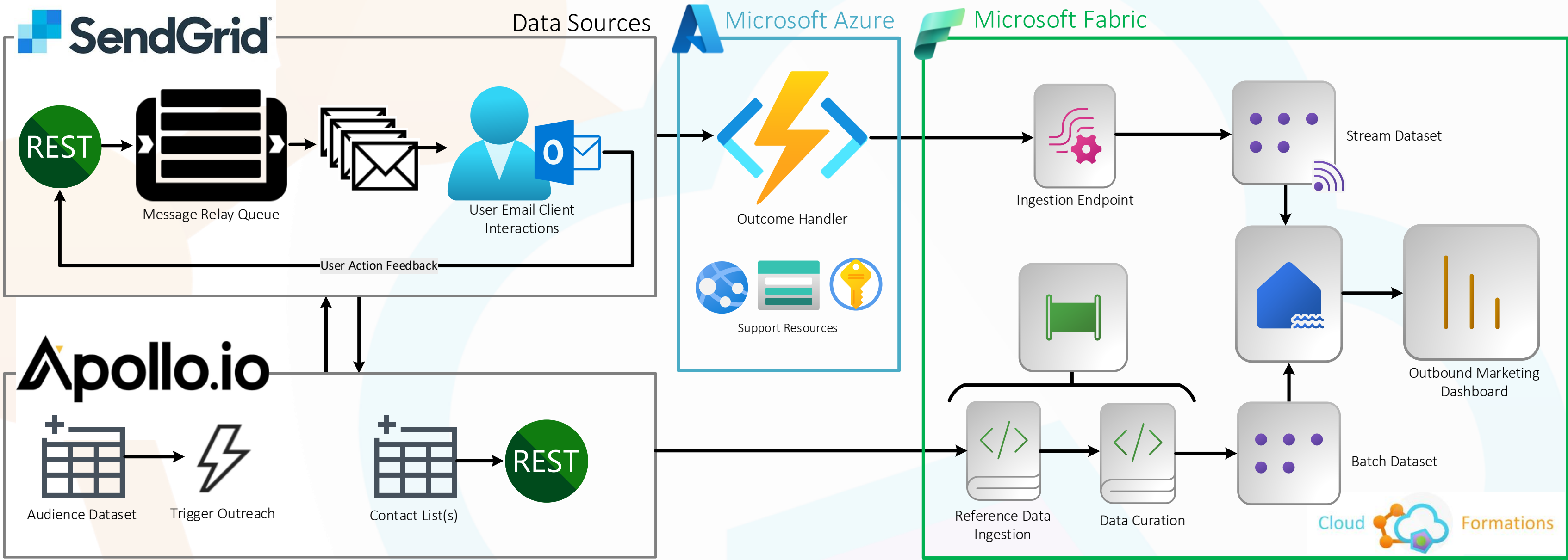
Azure Tooling – My Favourites



Cloud Formations - Knowledge Transfer & Training

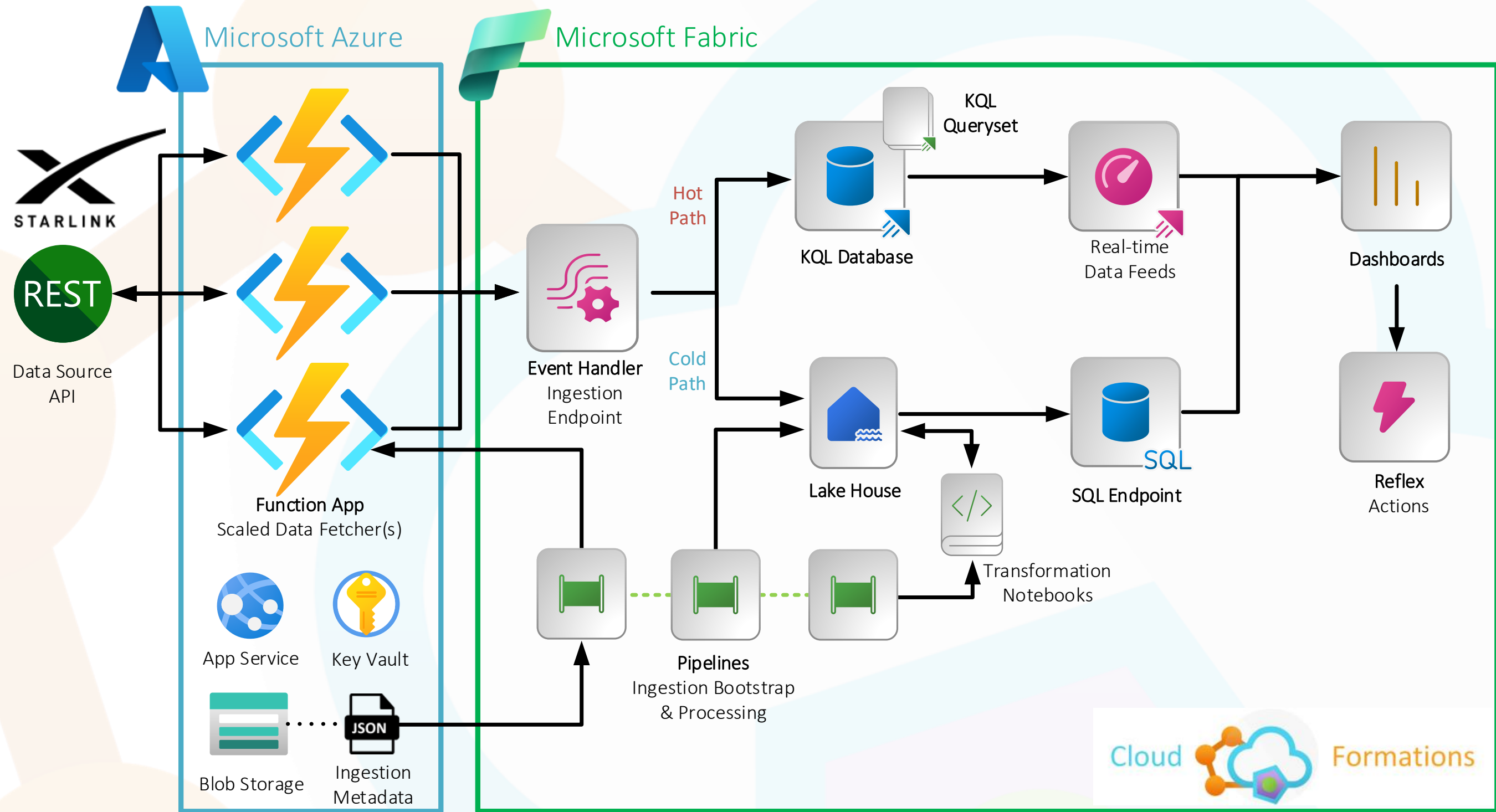


Case Study 1



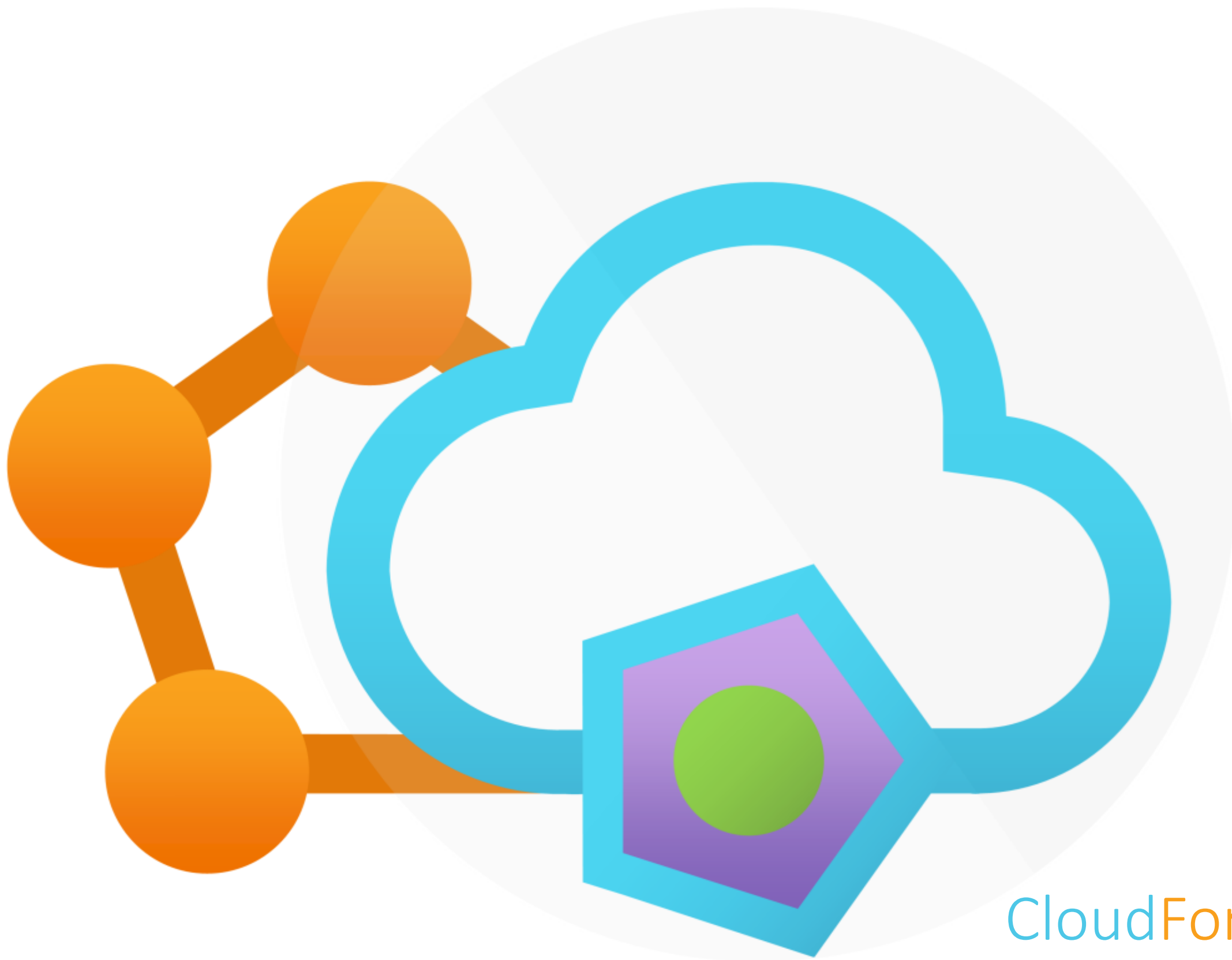
Near real-time streaming of marketing outreach data into Power BI dashboards to provide up to the minute analytics on the performance of campaigns to inform future content and audience enrolment.

Case Study 2





The greenfield implementation follows a microservices approach to data handling, leveraging Azure Functions Apps to ingest telemetry from the Starlink API at scale and metadata driven, feeding into a Microsoft Fabric Event Stream. Bootstrapped by Integration Pipelines, data feeds into Realtime Analytics Kusto Databases and Lakehouse structures for reporting using Power BI Dashboards.

Thank You



 mrpaulandrew.com

 paul@mrpaulandrew.com

 [In/mrpaulandrew](https://www.linkedin.com/in/mrpaulandrew)

 [@mrpaulandrew](https://twitter.com/mrpaulandrew)

Paul Andrew
Technical Architect



CloudFormations.org/Community-Content

