Azure Stream Analytics

Beyond IoT Real-time Data Ingestion

Paul Andrew

Microsoft® MVP Most Valuable Professional

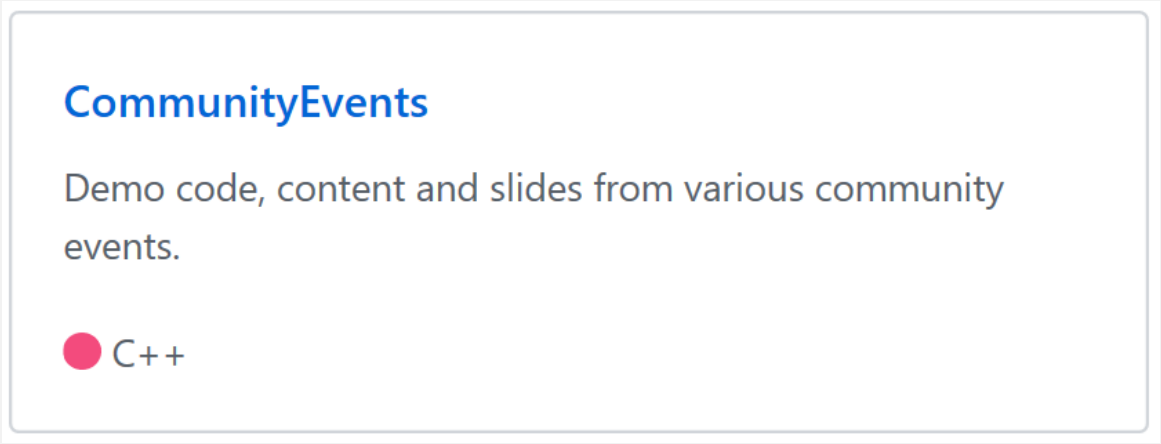
@MrPaulAndrew

In/MrPaulAndrew

MrPaulAndrew.tech

altius

https://github.com/mrpaulandrew



{Event/Location}-{Month}-{Year}

altius

Azure Stream Analytics

Real-time data problems

What is ASA and why use it

Production Considerations

Lambda Architecture

Azure Stream Analytics

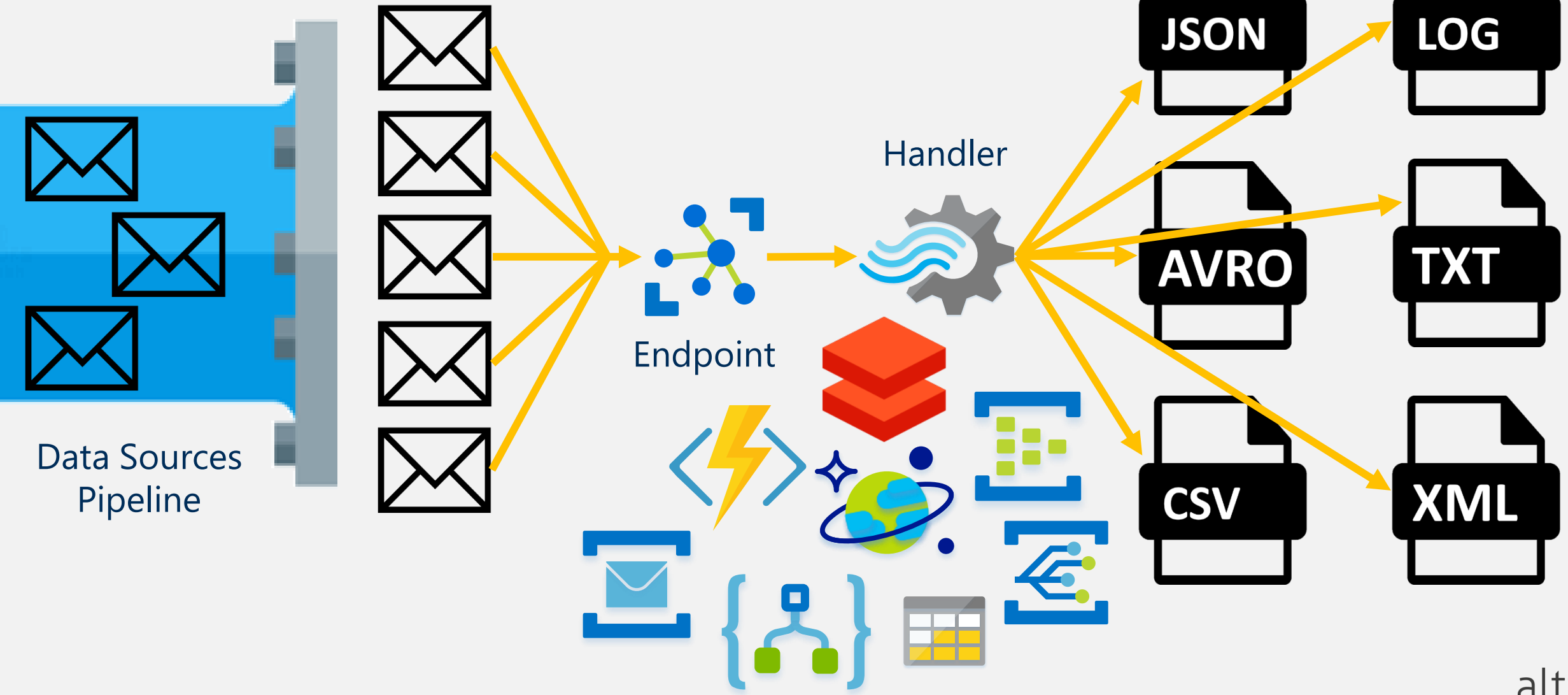
Real-time data problems

What is ASA and why use it

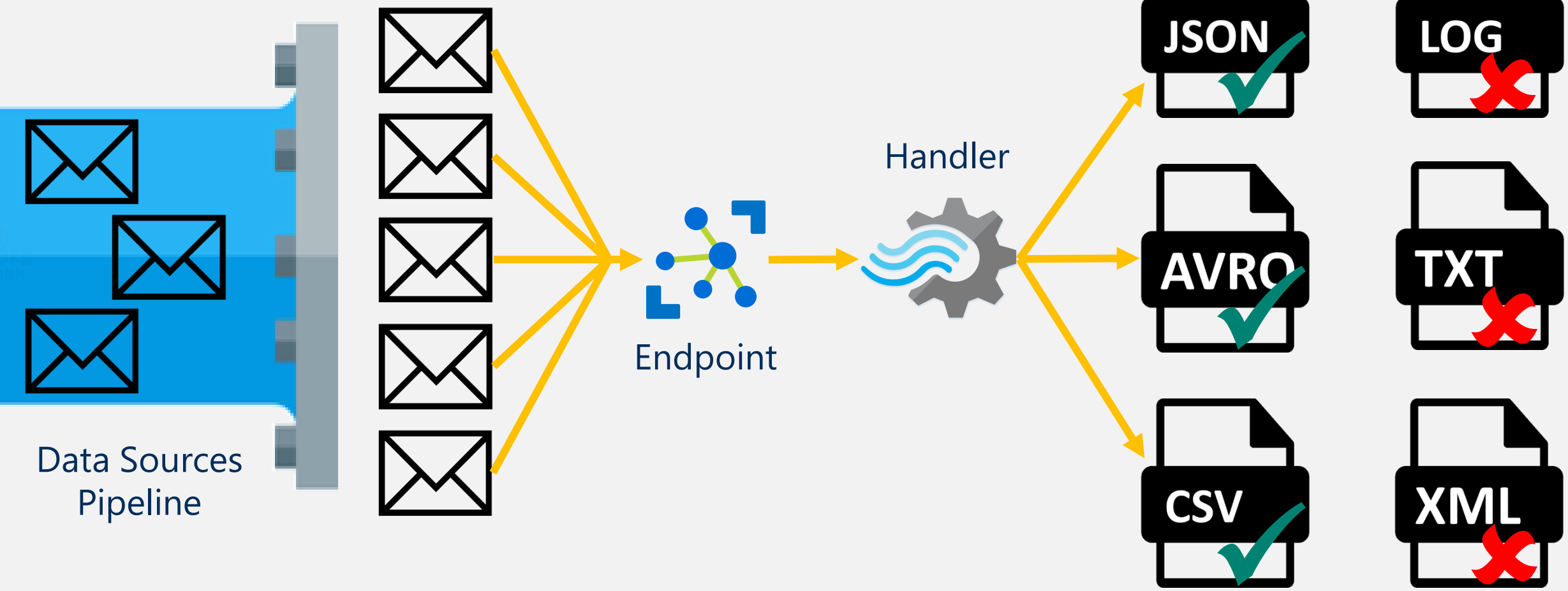
Production Considerations

Lambda Architecture

Problem: Real-time Data Feeds



Problem: Real-time Data Feeds



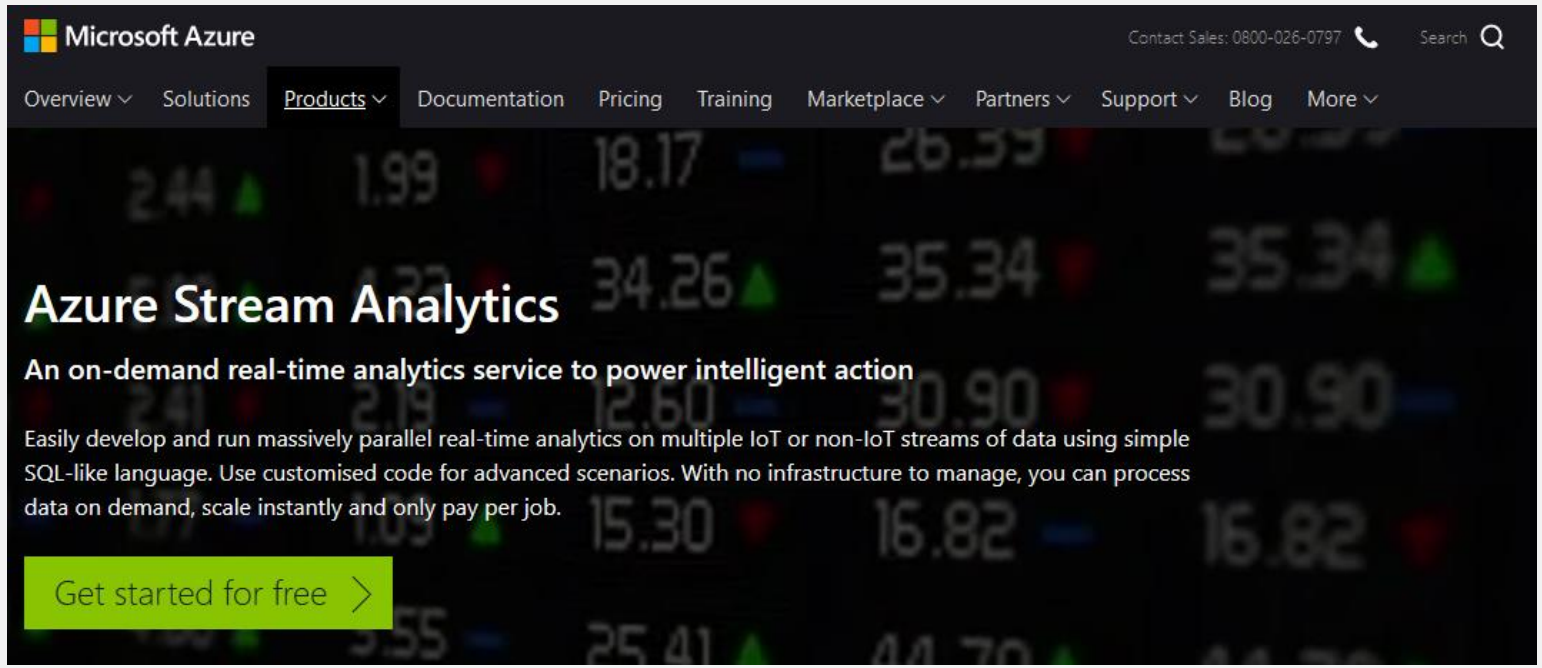
Azure Stream Analytics

Real-time data problems

What is ASA and why use it

Production Considerations

Lambda Architecture



Azure Stream Analytics

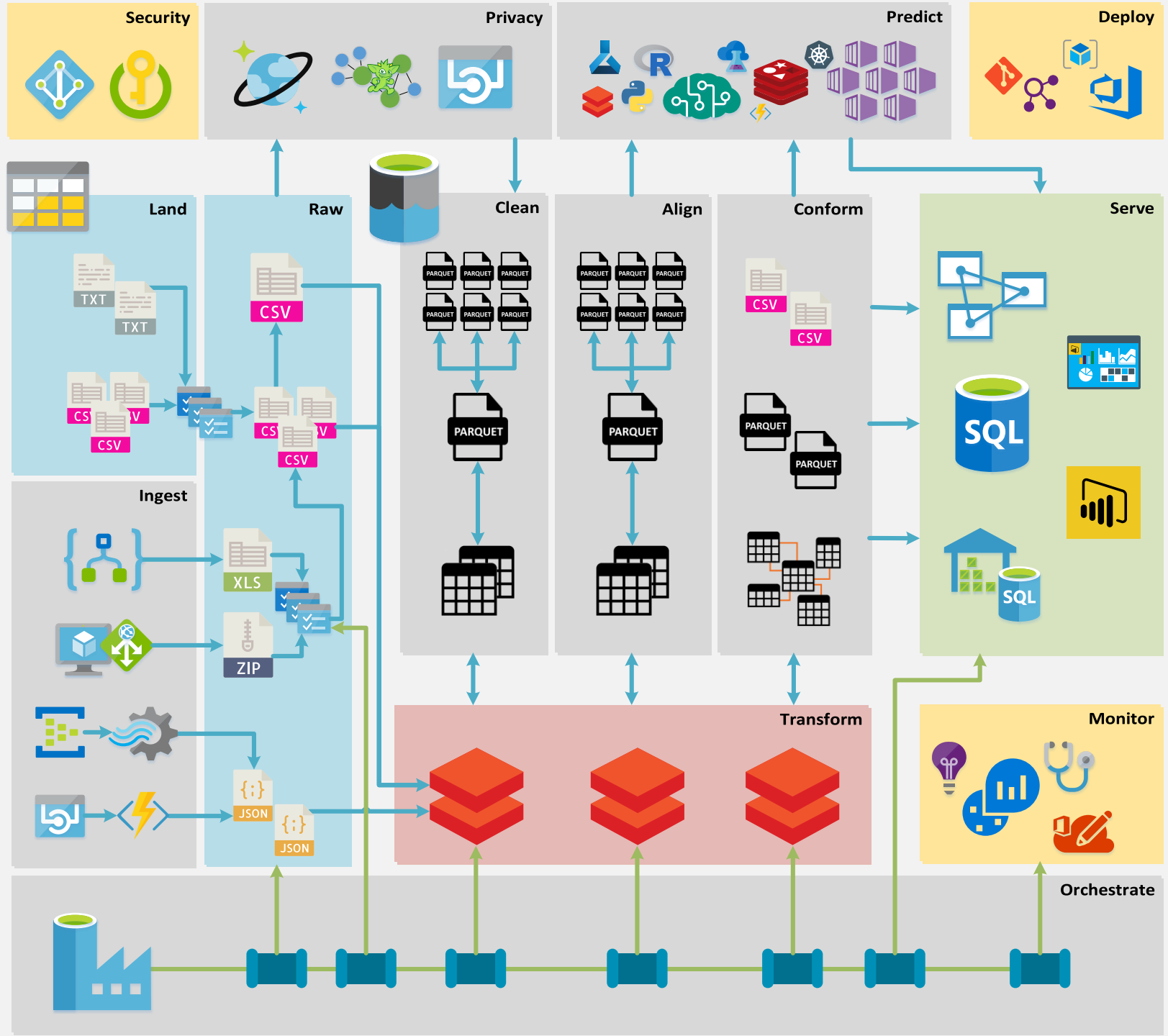
https://azure.microsoft.com/en-gb/services/stream-analytics/

Real-time data problems

What is ASA and why use it

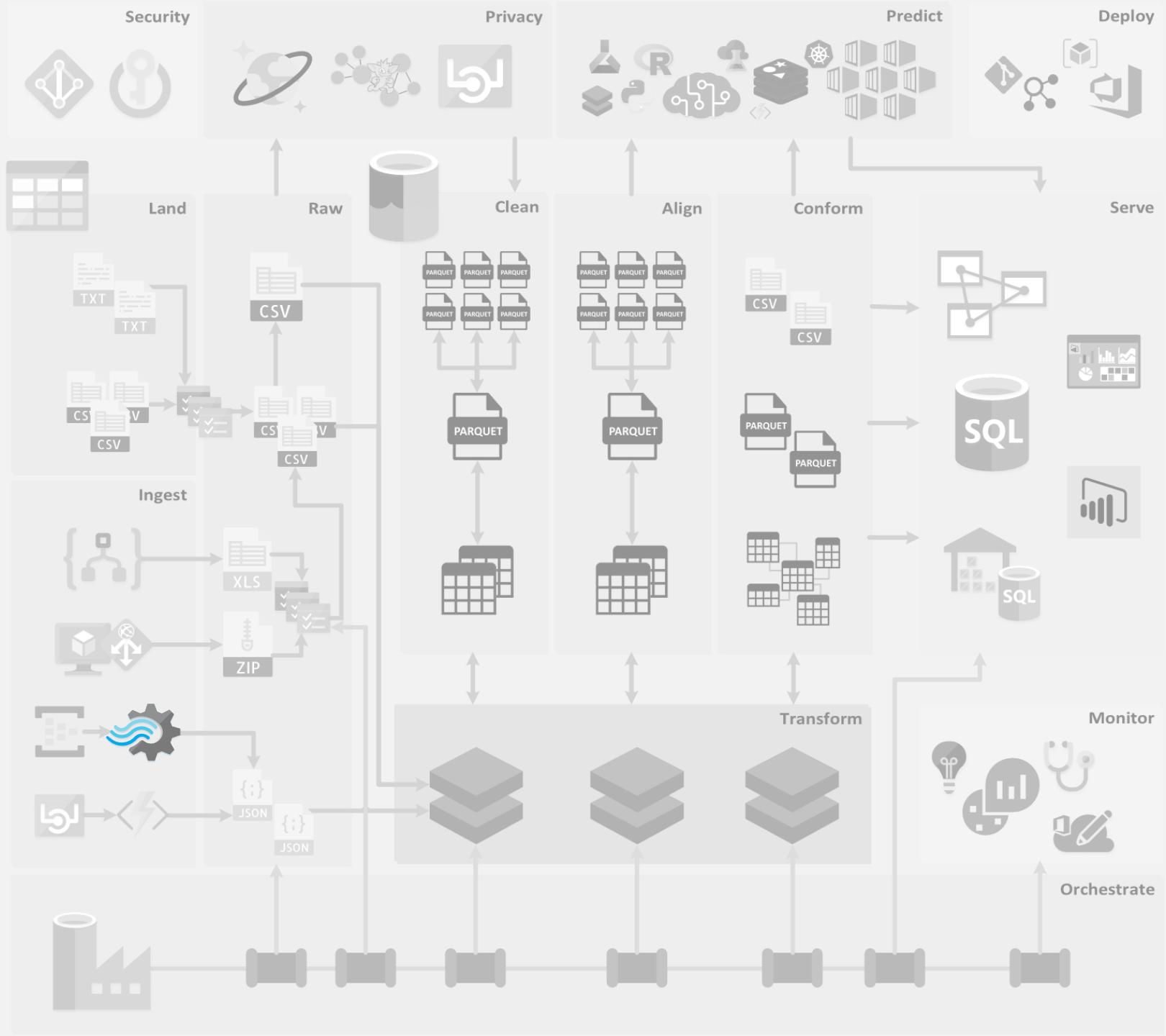
Production Considerations

Lambda Architecture



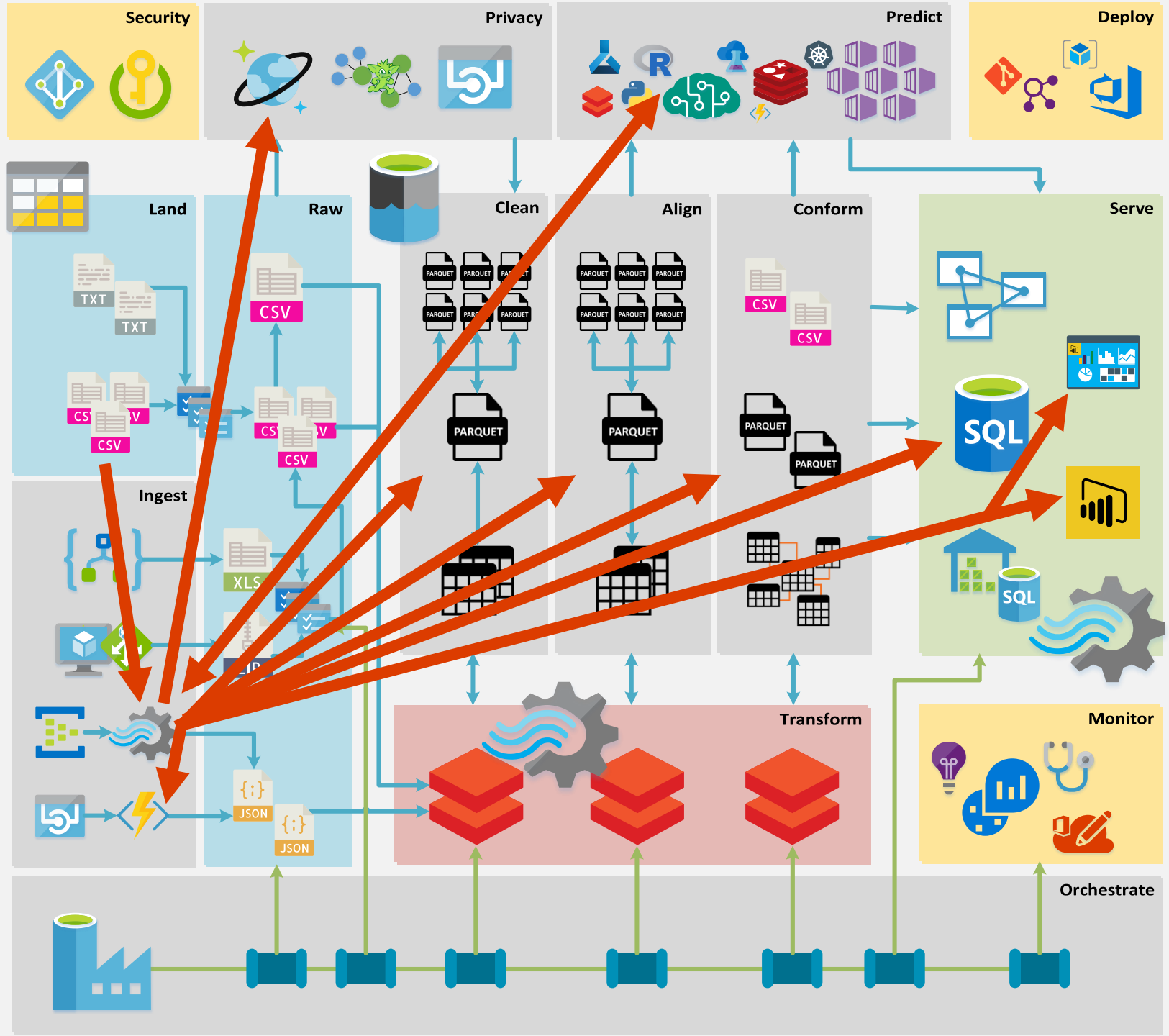
What is ASA and why use it

altius



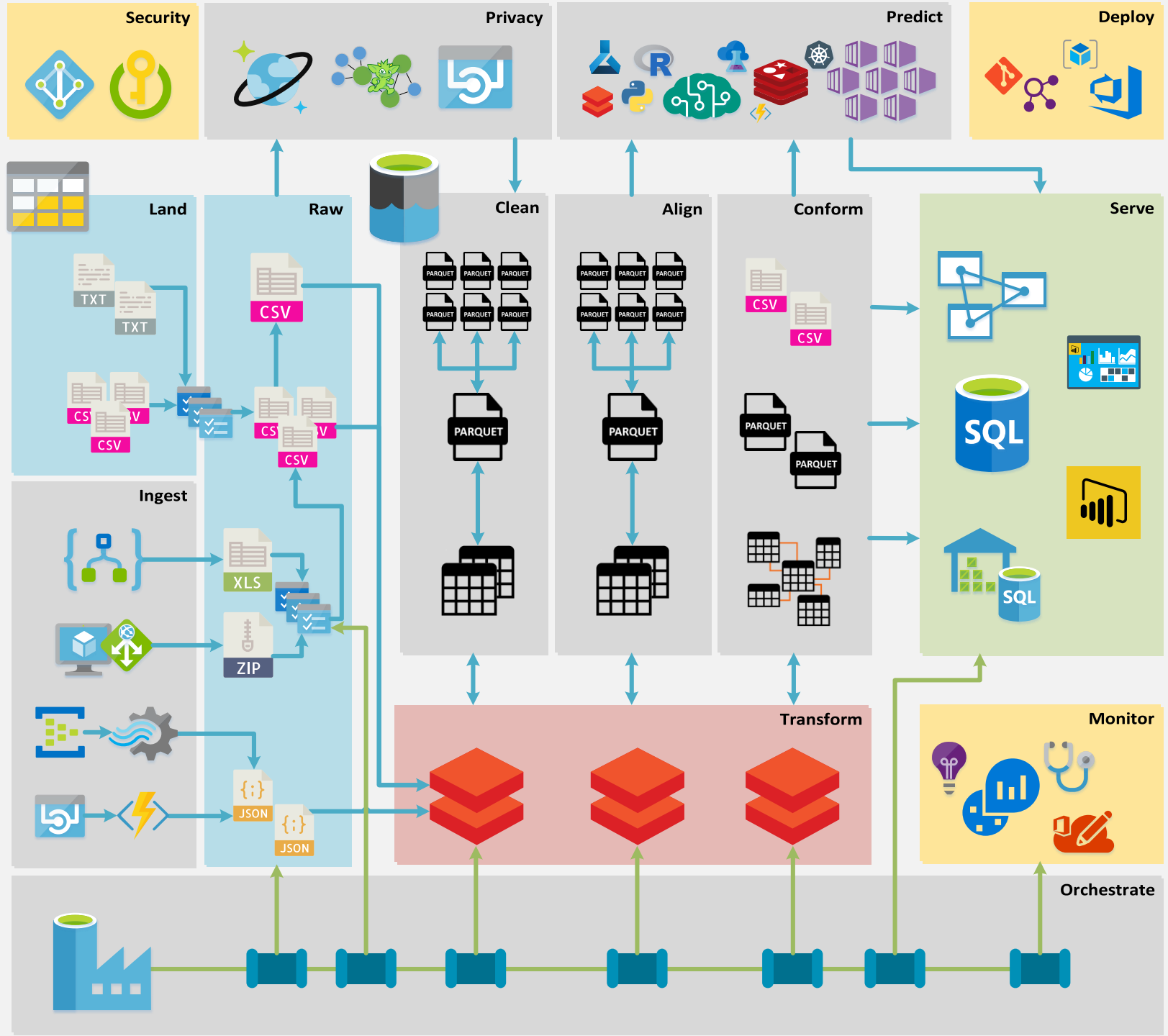
What is ASA and why use it

altius



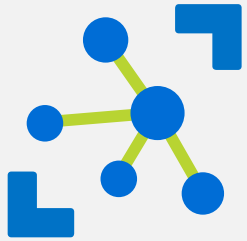
What is ASA and why use it

altius

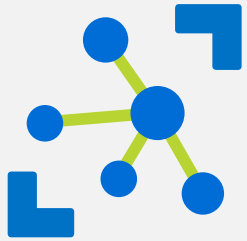


What is ASA and why use it

altius



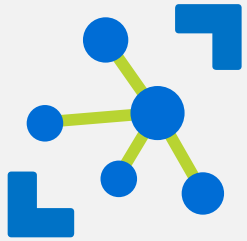
Azure IoT Hub vs Azure Event Hub



Feature	Azure IoT Hub	Azure Event Hub
Message Direction	2 Way	1 Way
Protocol Support	MQTT, AMQP, HTTP	AMQP, HTTP
Scaling	Configured	Automatic
Message Routing	Yes	No
Security	Device Level	Hub Level
Device State Support	Yes	No
Message Capturing	No	Yes
Multiple Namespaces	No	Yes
Tiers	F1/S1/S2/S3	Basic/Standard
Service Endpoint	Yes	Yes (preview)

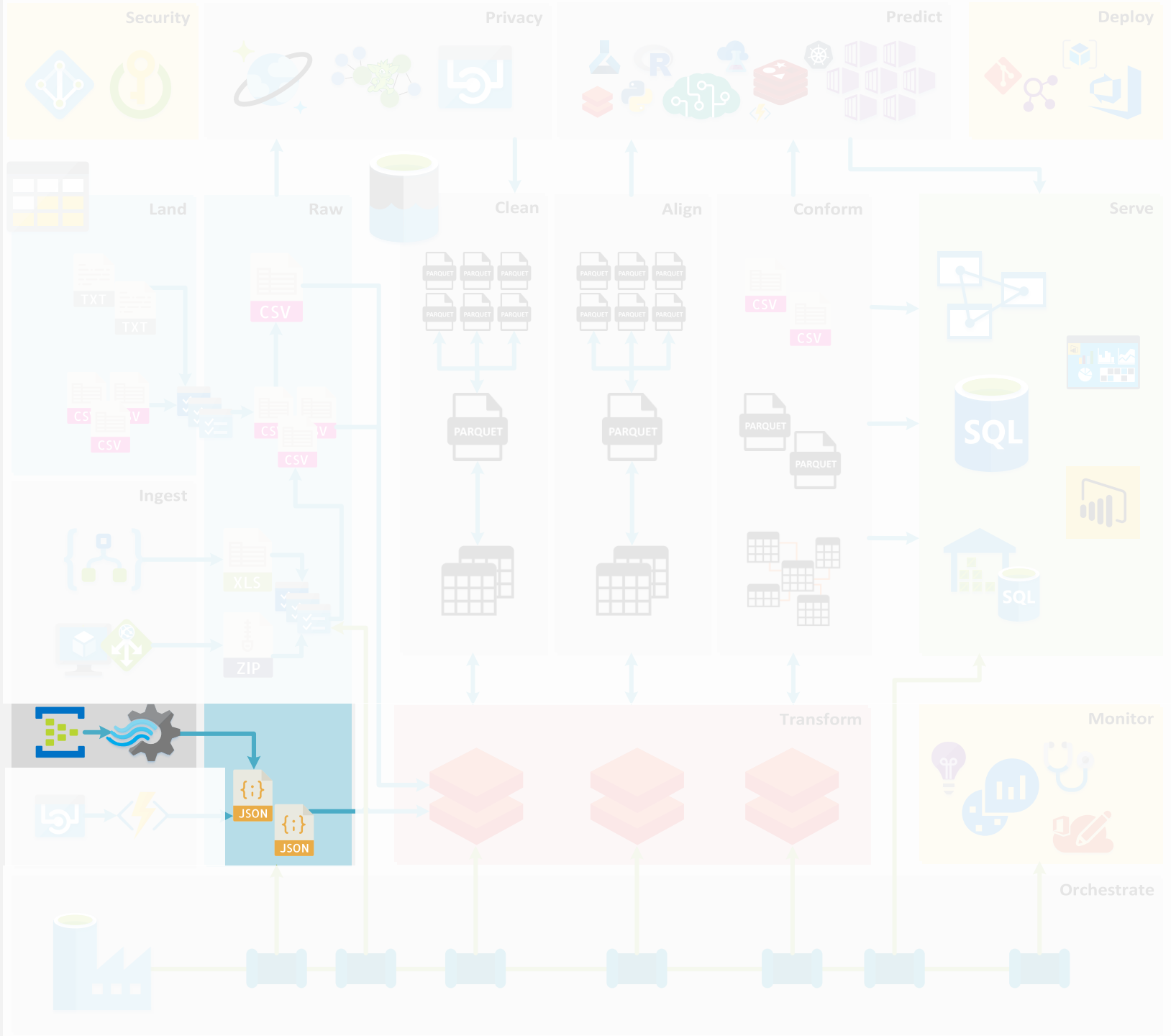
https://docs.microsoft.com/en-us/azure/iot-hub/iot-hub-compare-event-hubs

Azure IoT Hub vs Azure Event Hub



Feature	Azure IoT Hub	Azure Event Hub
Message Direction	2 Way	1 Way
Protocol Support	MQTT, AMQP, HTTP	AMQP, HTTP
Scaling	Configured	Automatic
Message Routing	Yes	No
Security	Device Level	Hub Level
Device State Support	Yes	No
Message Capturing	No	Yes
Multiple Namespaces	No	Yes
Tiers	F1/S1/S2/S3	Basic/Standard
Service Endpoint	Yes	Yes (preview)

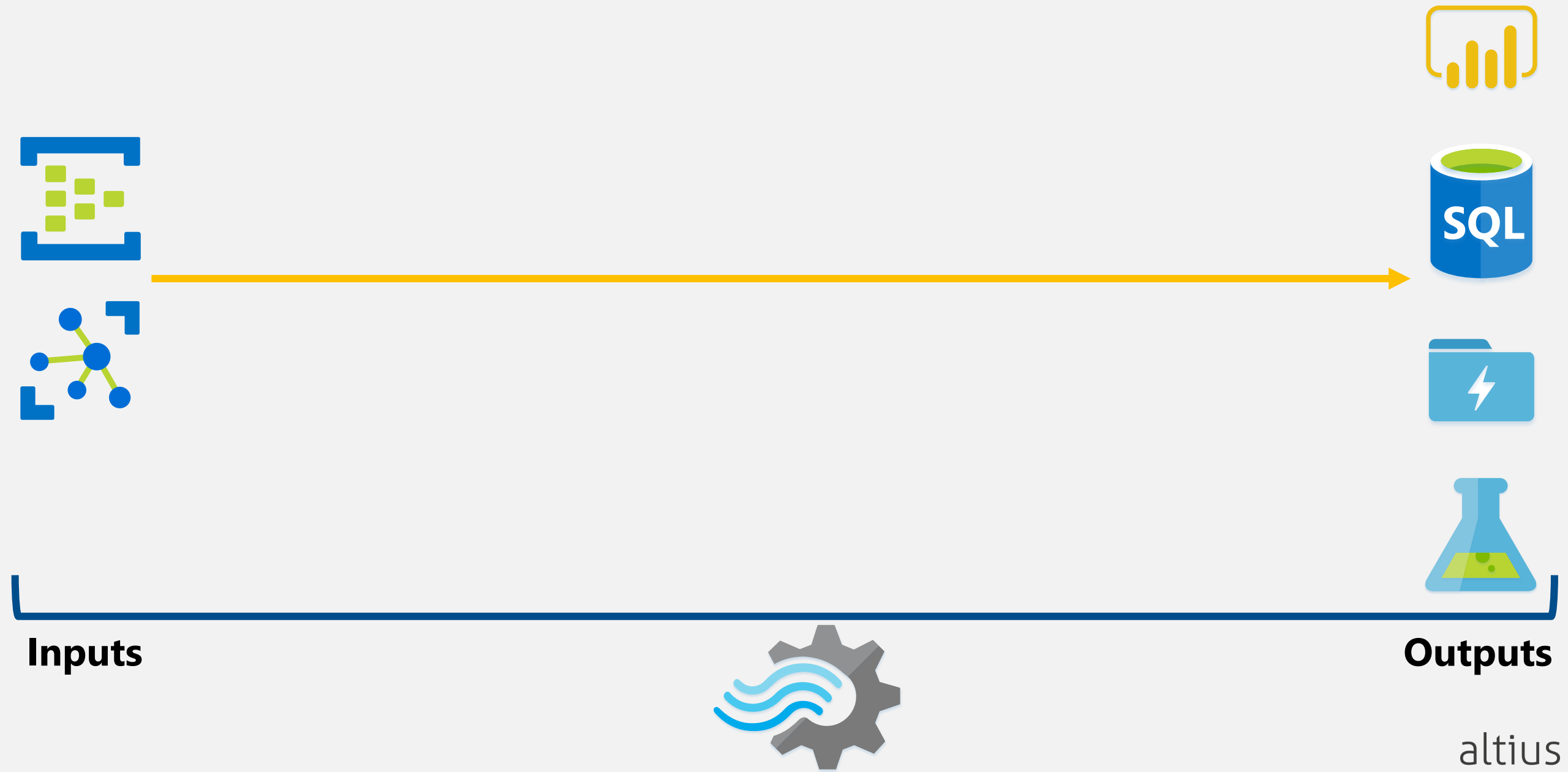
https://docs.microsoft.com/en-us/azure/iot-hub/iot-hub-compare-event-hubs



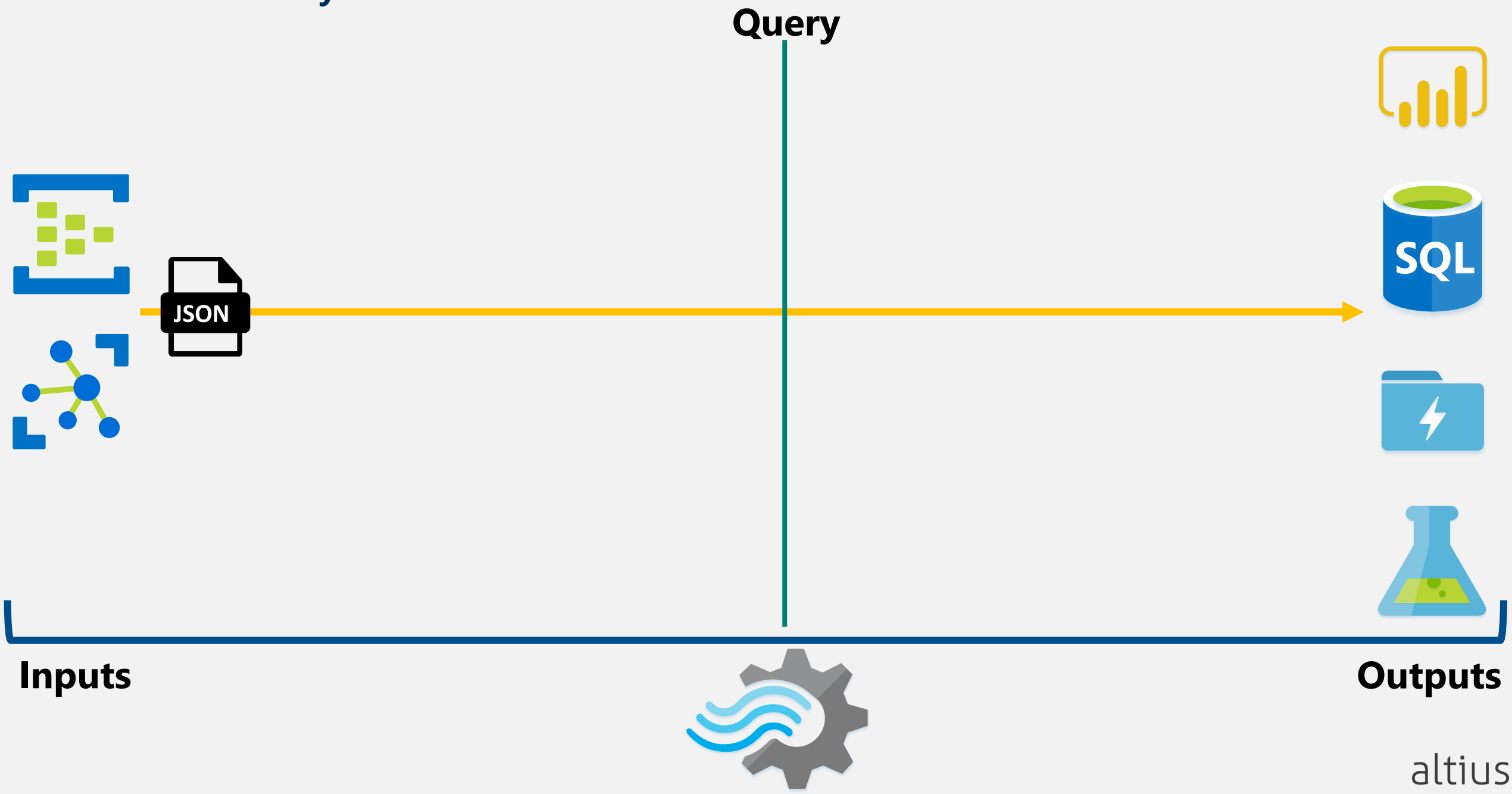
What is ASA and why use it

altius

Azure Stream Analytics



Azure Stream Analytics



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

altius

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)
```

*100MB Limit

Inputs

Outputs

altius

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

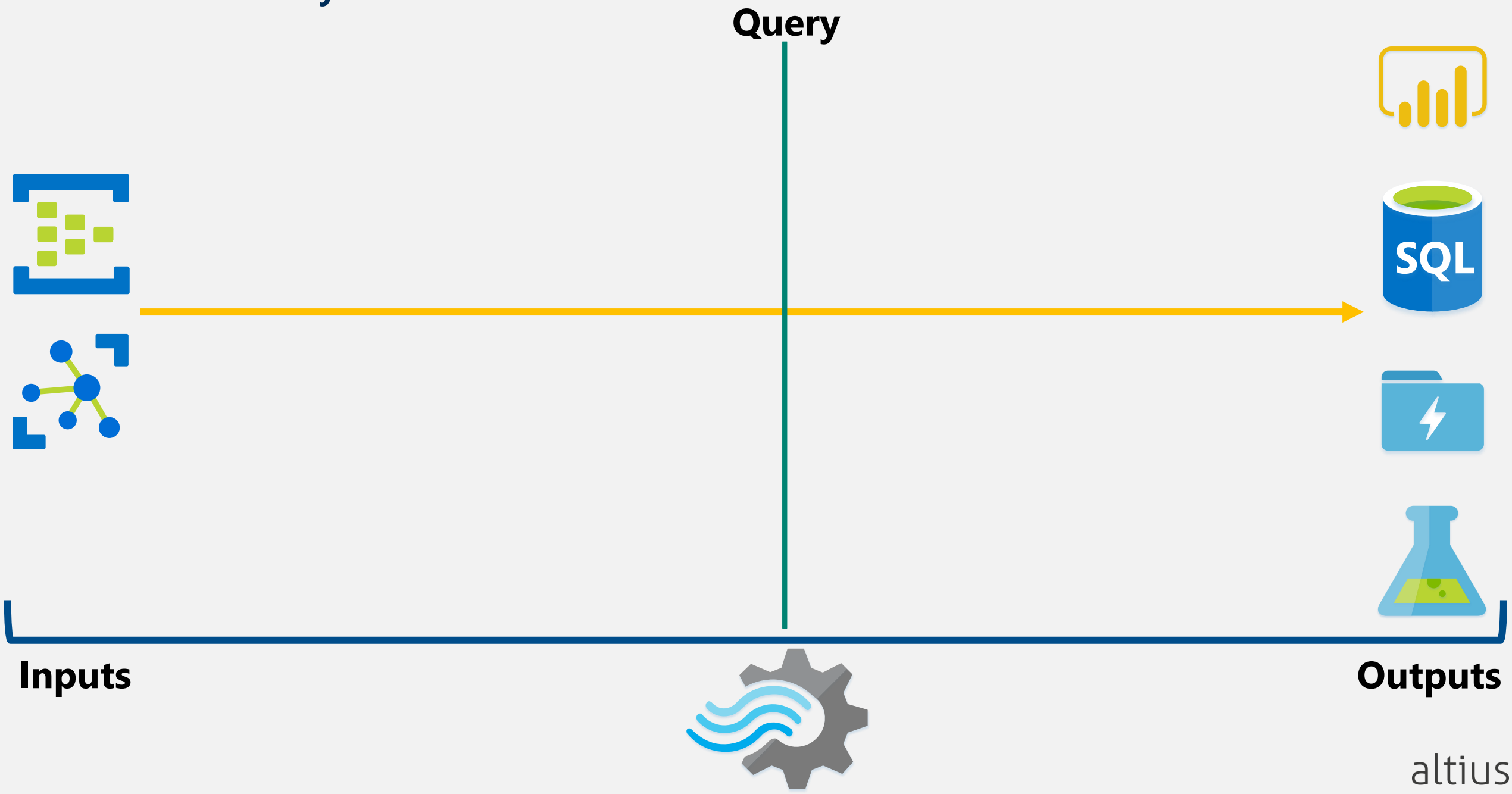
*100MB Limit

Inputs… Source Data
Reference Data
Custom Functions

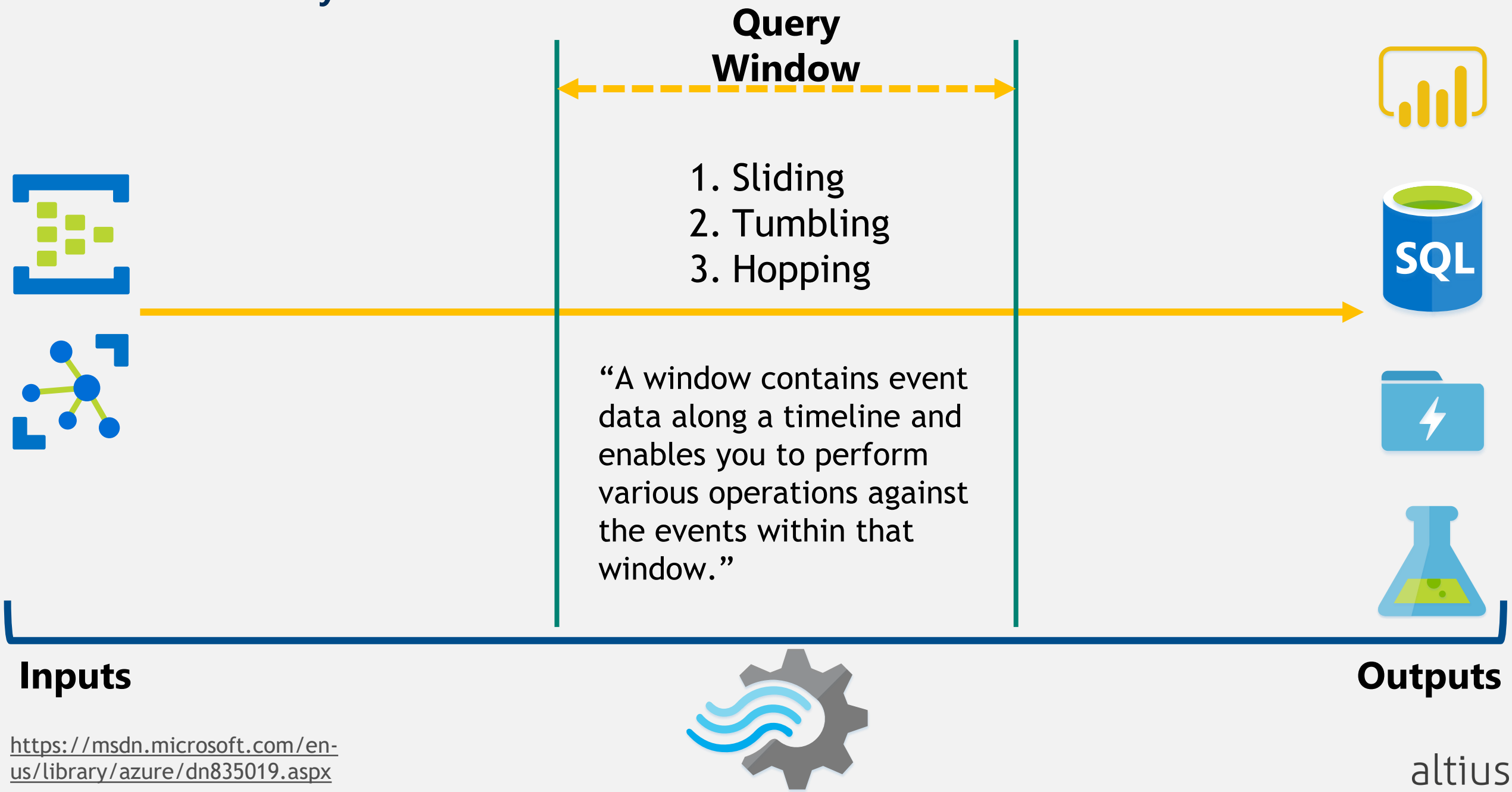
Outputs

altius

Azure Stream Analytics



Azure Stream Analytics



https://msdn.microsoft.com/en-us/library/azure/dn835019.aspx

altius

Azure Stream Analytics

Sliding Window

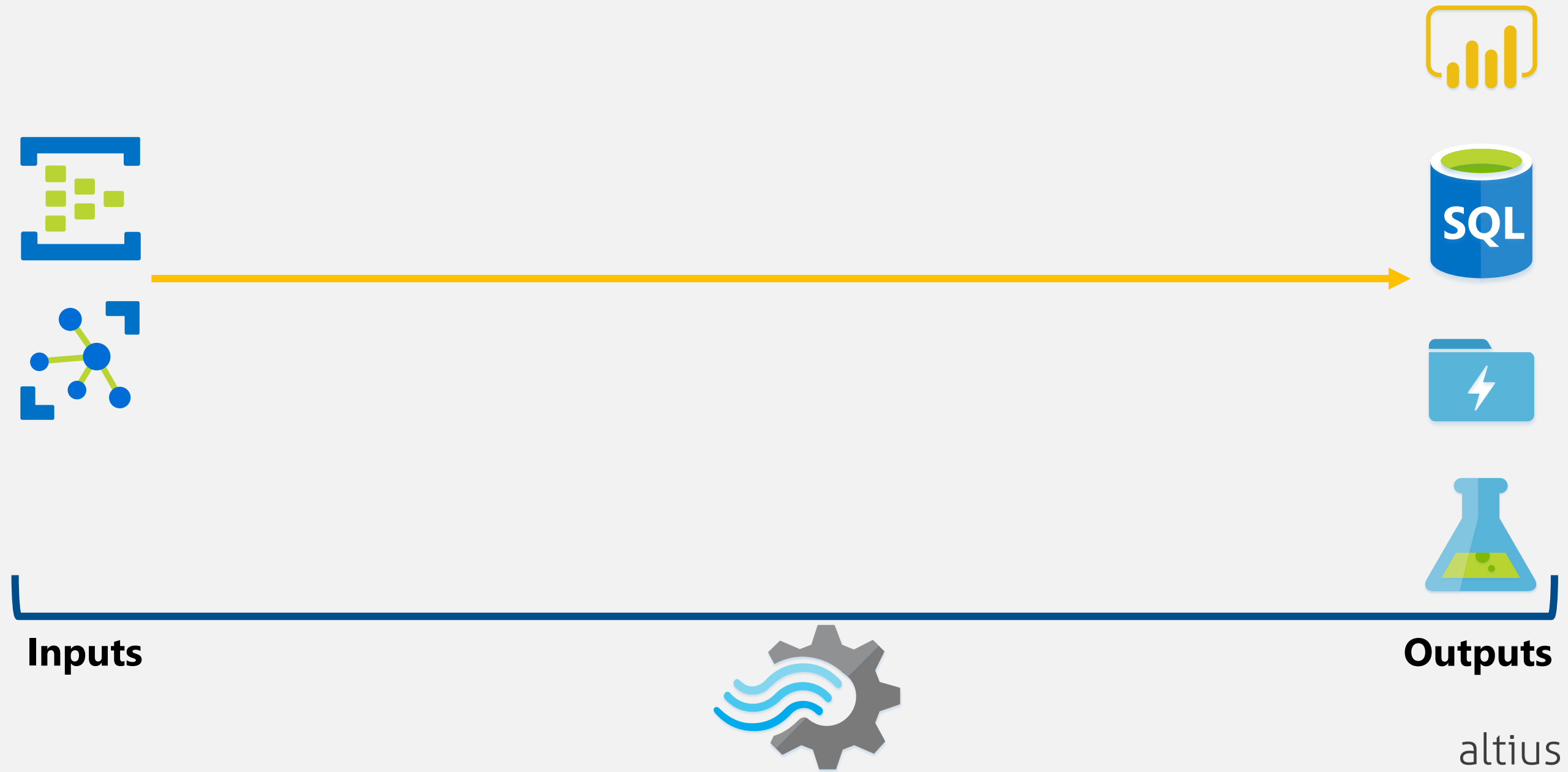
```
SELECT
    MIN([value]) AS 'Min',
    MAX([value]) AS 'Max'
INTO
    SQLDBAvg
FROM
    IoTHub
TIMESTAMP BY
    timecreated
GROUP BY
    SlidingWindow(second, 5))
```

Inputs

Outputs

altius

Azure Stream Analytics



Azure Stream Analytics

Tumbling Window

Tumbling Window

Tumbling Window

Tumbling Window

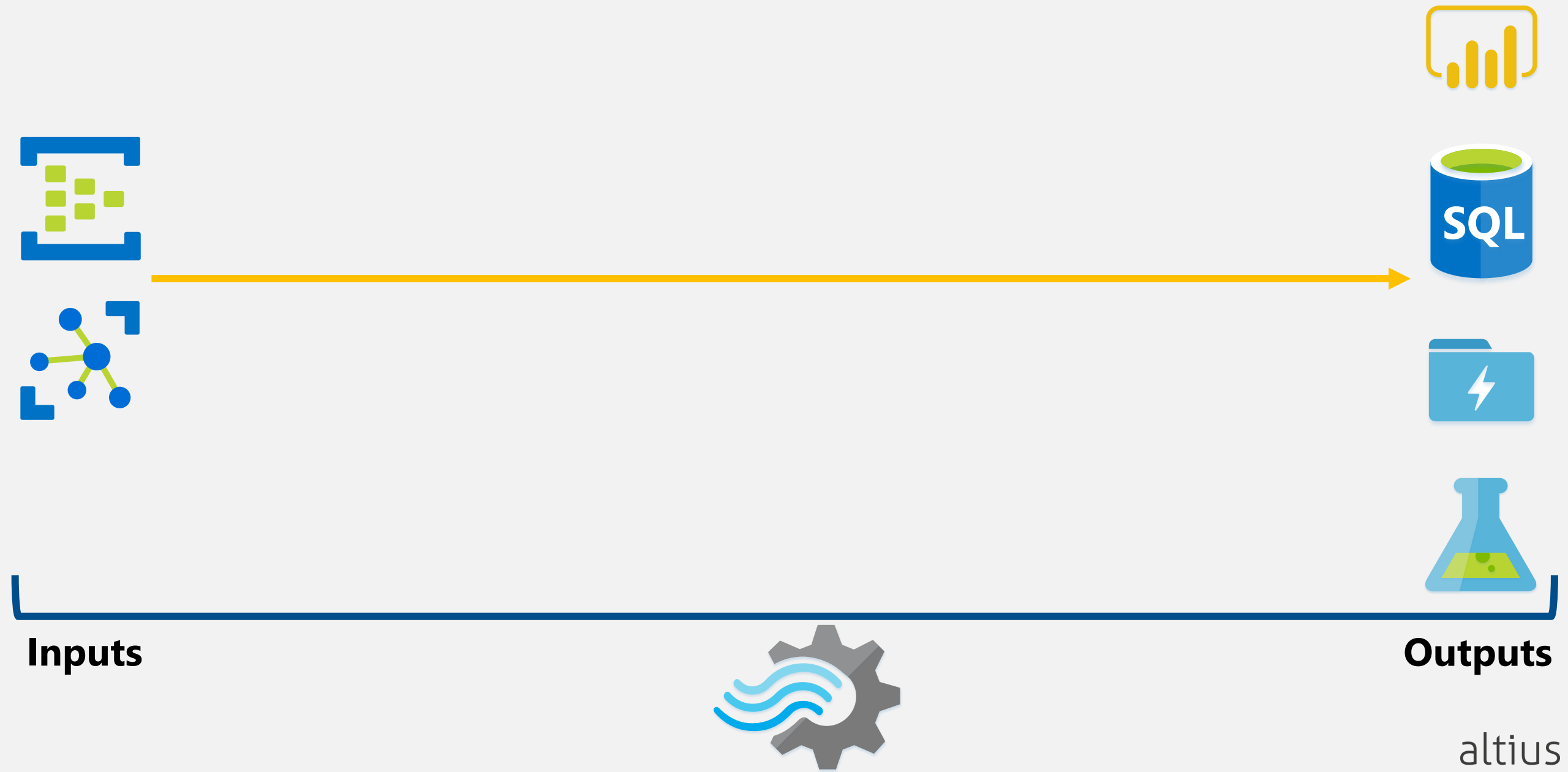
```
SELECT
    MIN([value]) AS 'Min',
    MAX([value]) AS 'Max'
INTO
    SQLDBAvg
FROM
    IoTHub
TIMESTAMP BY
    timecreated
GROUP BY
    TumblingWindow(Duration(MINUTE, 15), Offset(millisecond, -1))
```

Inputs

Outputs

altius

Azure Stream Analytics



Azure Stream Analytics

Hopping Window

Hopping Window

Hopping Window

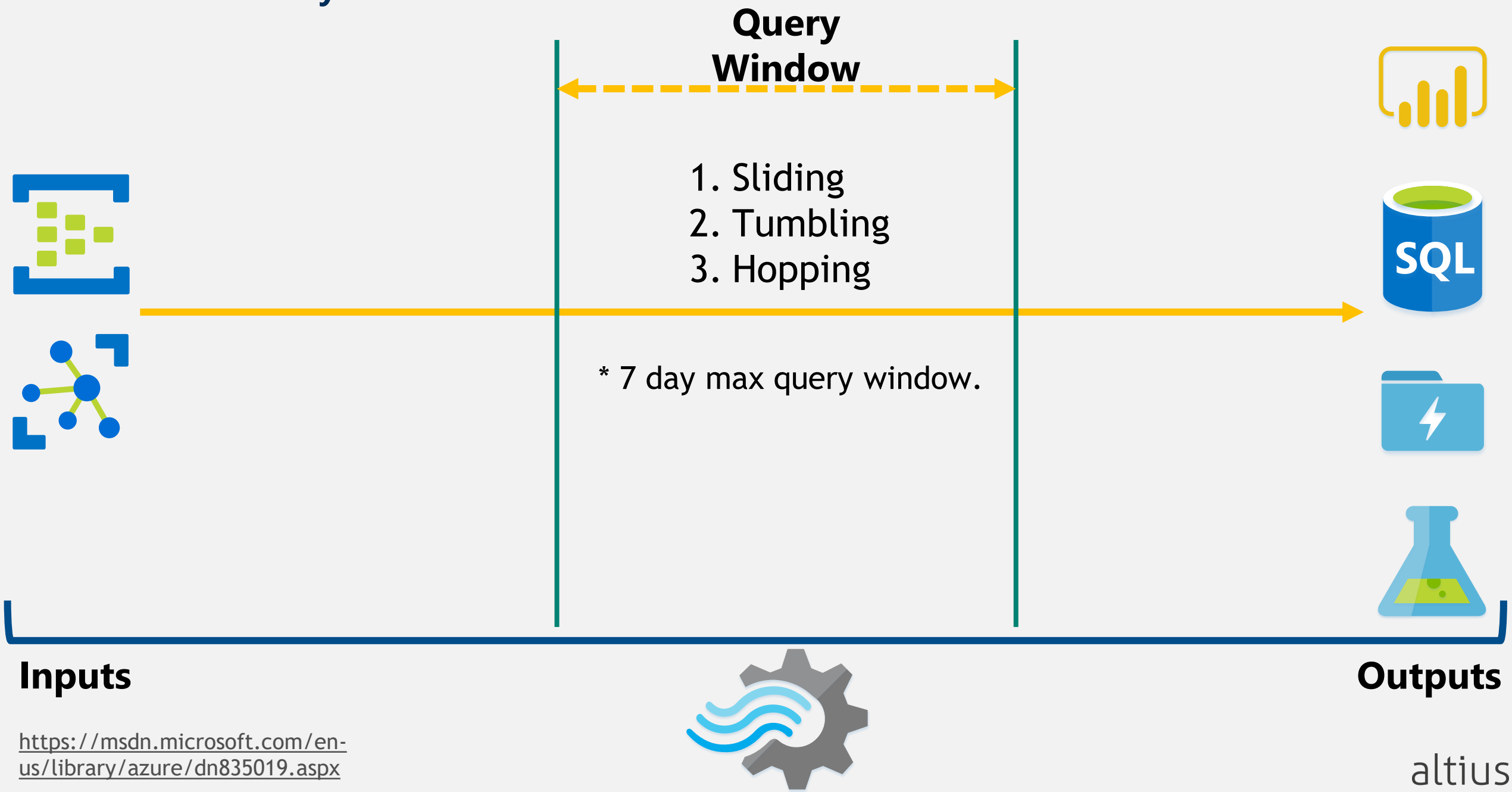
Hopping Window

```
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

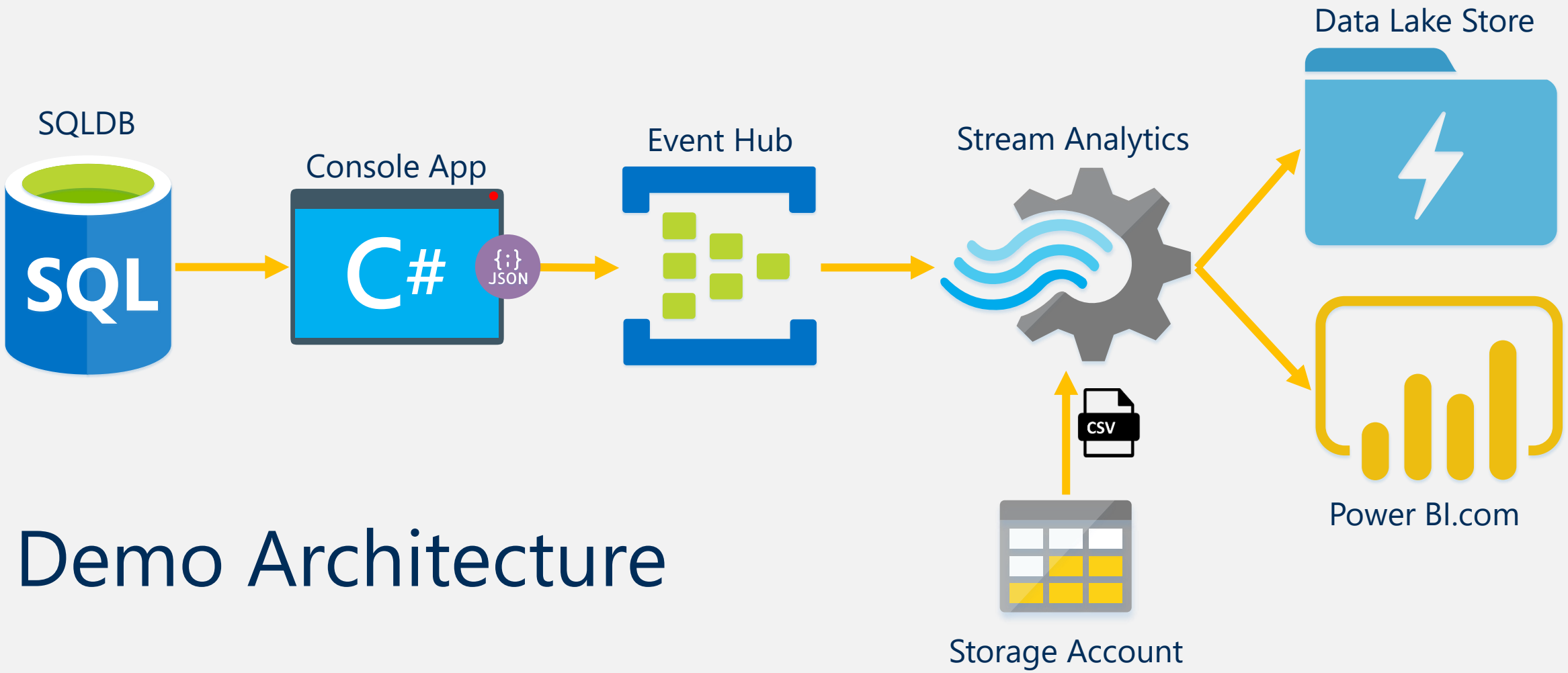
Azure Stream Analytics



https://msdn.microsoft.com/en-us/library/azure/dn835019.aspx

Demo

altius



Demo Architecture

altius

Azure Stream Analytics

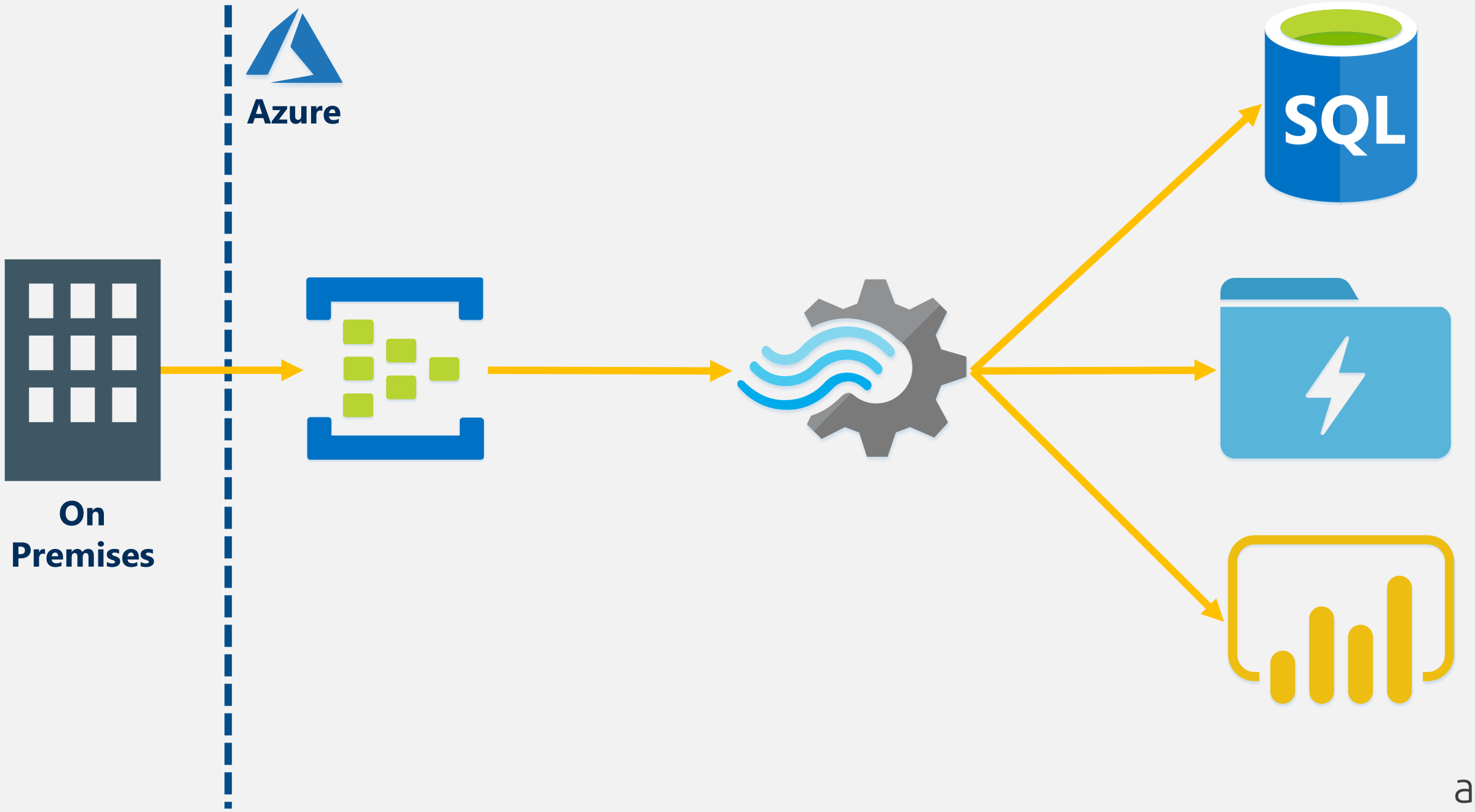
Real-time data problems

What is ASA and why use it

Production Considerations

Lambda Architecture

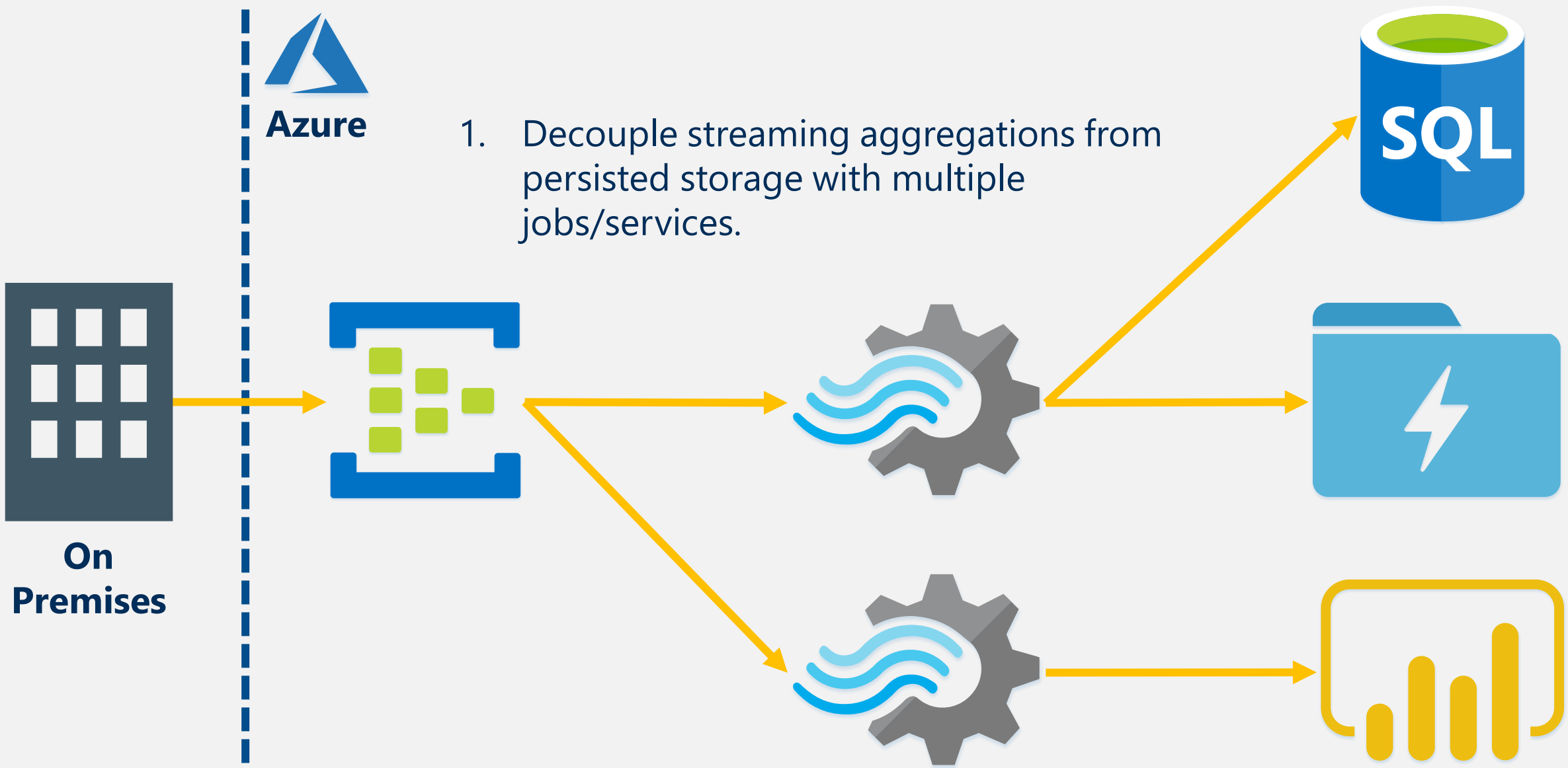
Production Considerations



Production Considerations

Azure

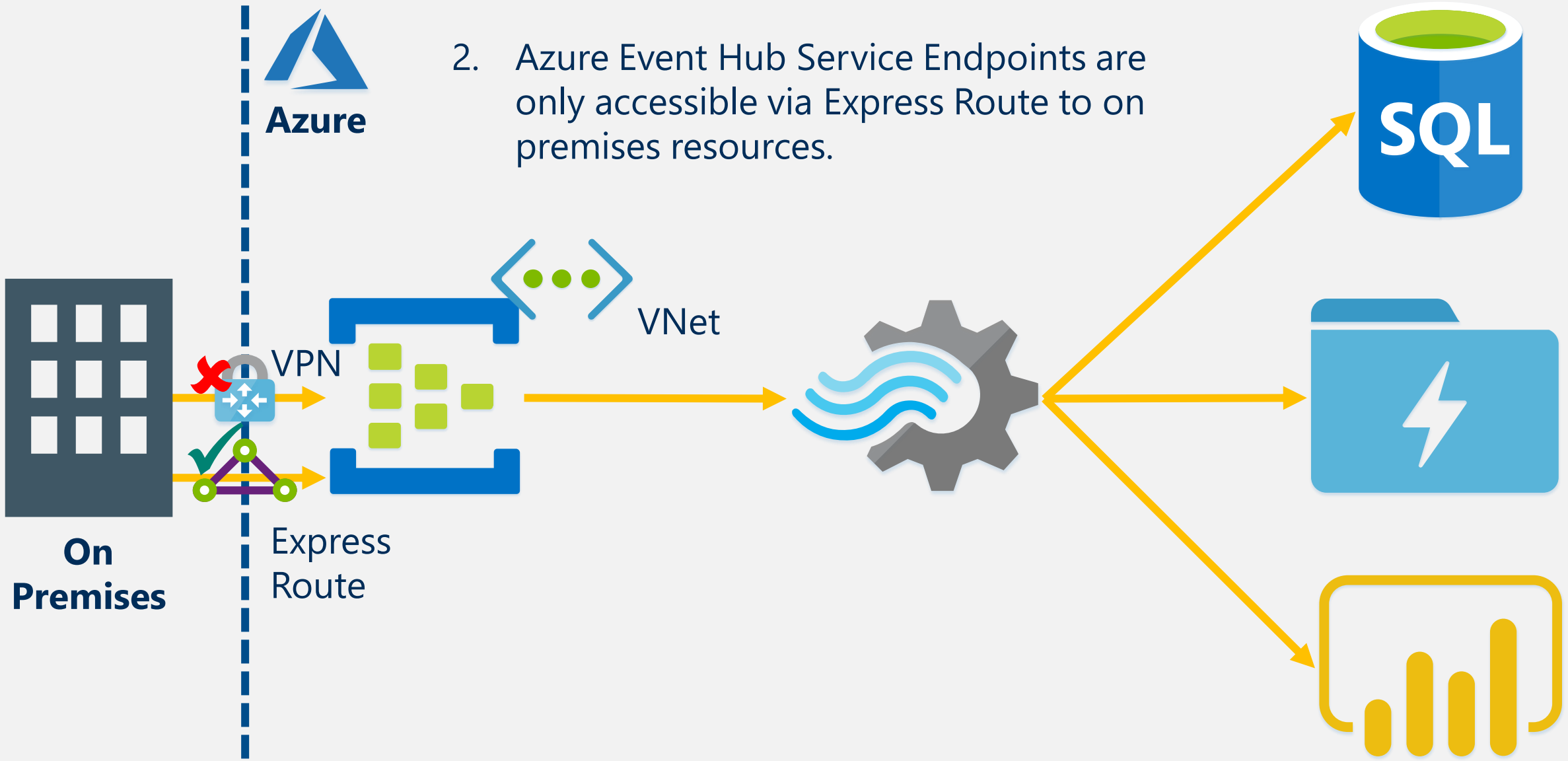
1. Decouple streaming aggregations from persisted storage with multiple jobs/services.



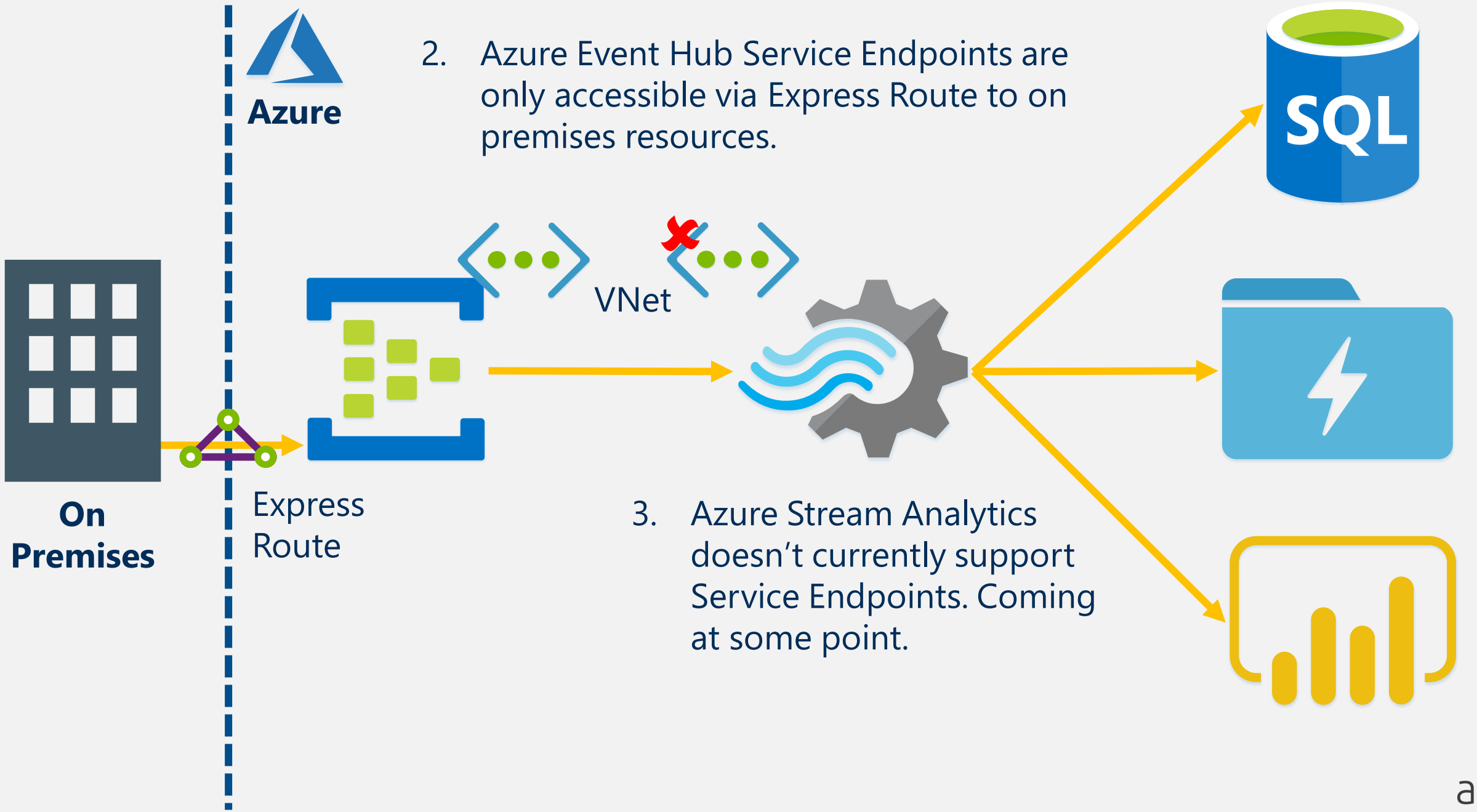
On Premises

Production Considerations

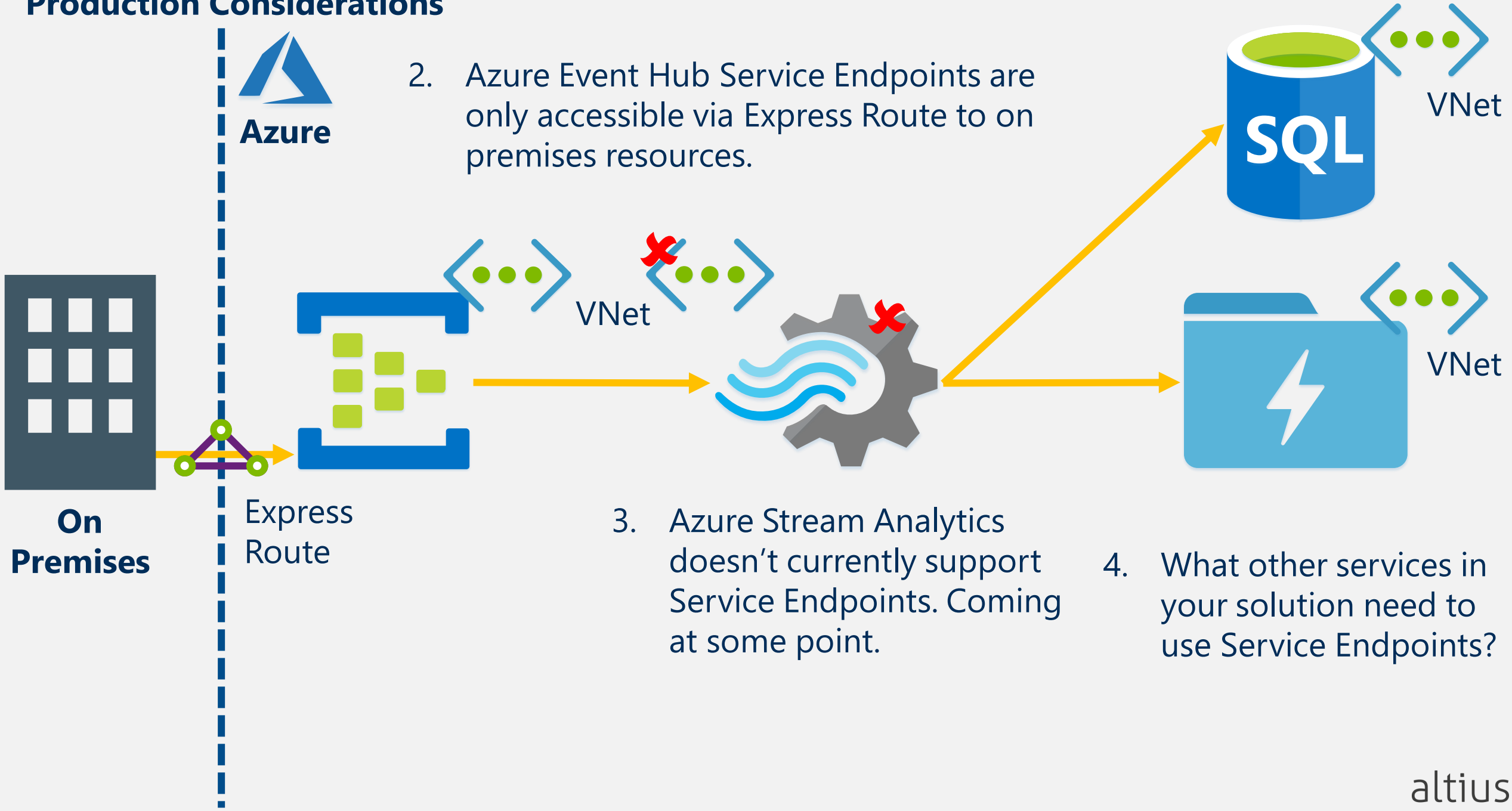
2. Azure Event Hub Service Endpoints are only accessible via Express Route to on premises resources.



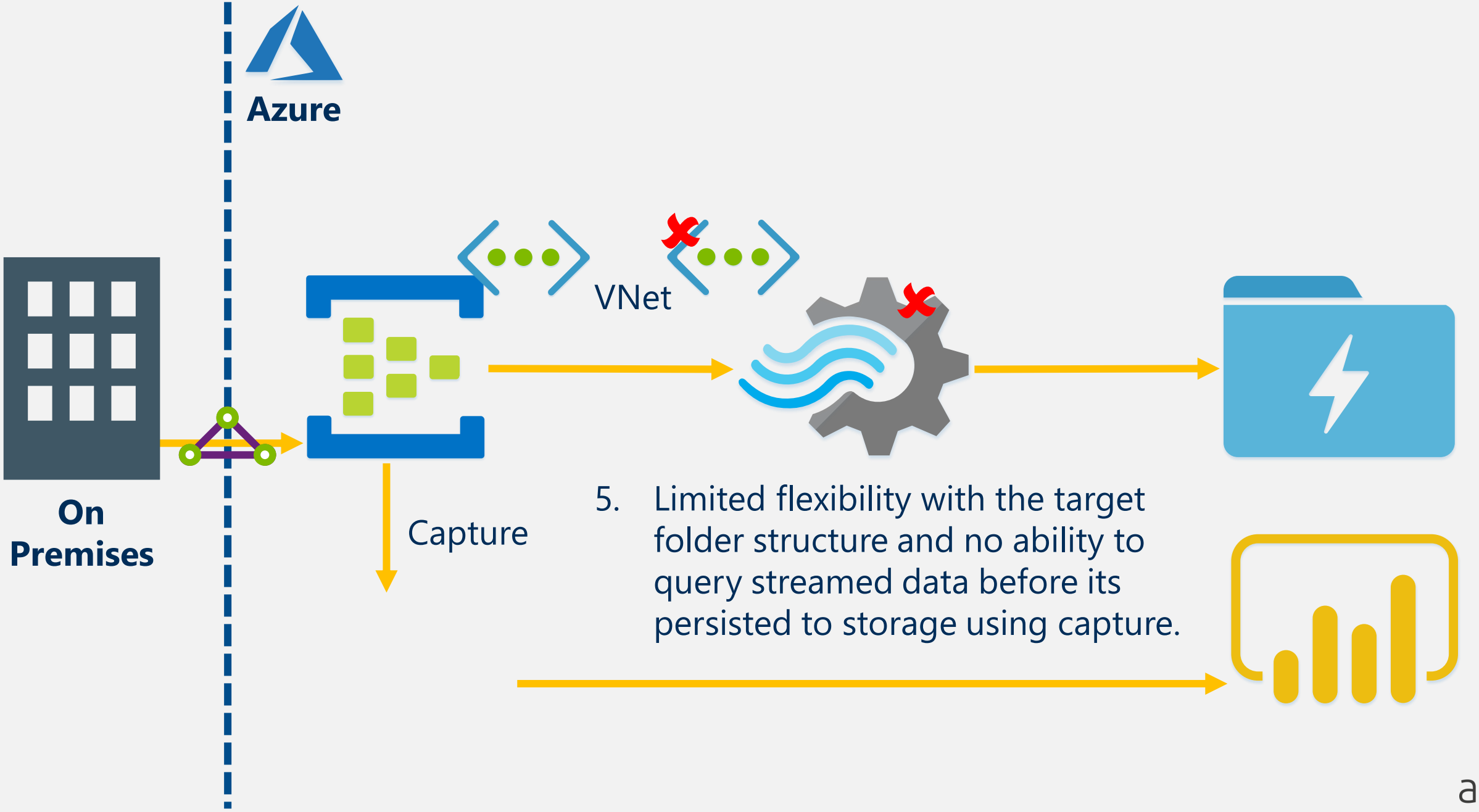
Production Considerations



Production Considerations

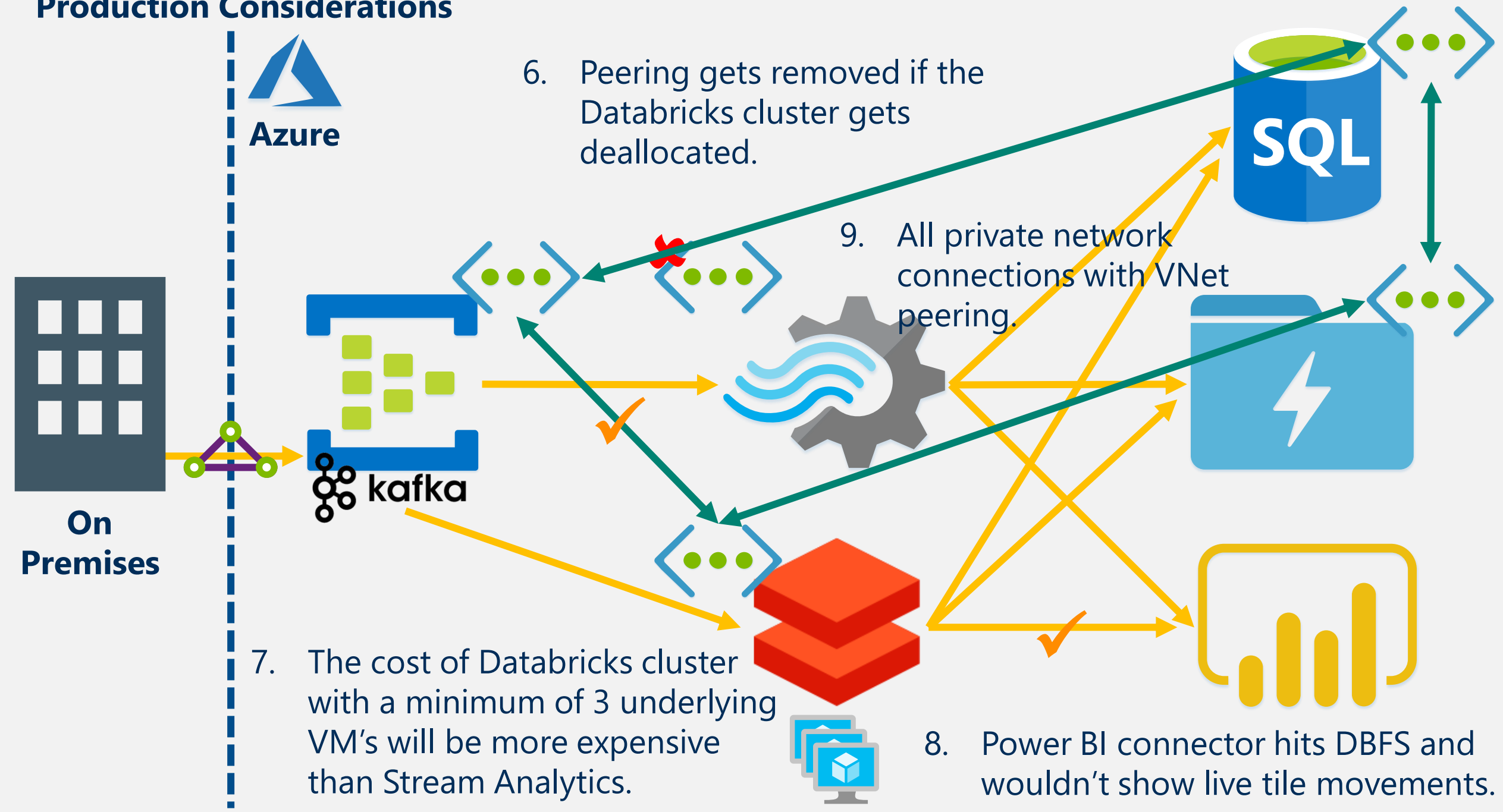


Production Considerations



5. Limited flexibility with the target folder structure and no ability to query streamed data before its persisted to storage using capture.

Production Considerations



Production Considerations Summary

1. Decouple streaming aggregations from persisted storage with multiple jobs/services.
2. Azure Event Hub Service Endpoints are only accessible via Express Route to on premises resources.
3. Azure Stream Analytics doesn't currently support Service Endpoints. Coming at some point.
4. What other services in your solution need to use Service Endpoints?
5. Limited flexibility with the target folder structure and no ability to query streamed data before its persisted to storage using capture.
6. Peering gets removed if the Databricks cluster gets deallocated.
7. The cost of Databricks cluster with a minimum of 3 underlying VM's will be more expensive than Stream Analytics.
8. Power BI connector hits DBFS and wouldn't show live tile movements.
9. All private network connections with VNet peering.

https://docs.microsoft.com/en-us/azure/event-hubs/event-hubs-service-endpoints

Azure Stream Analytics

Real-time data problems

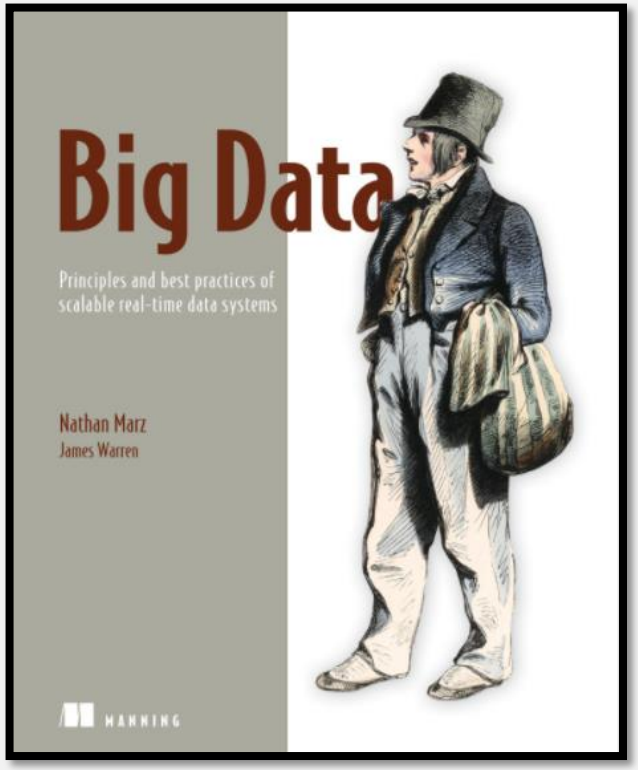
What is ASA and why use it

Production Considerations

Lambda Architecture

Lambda Architecture

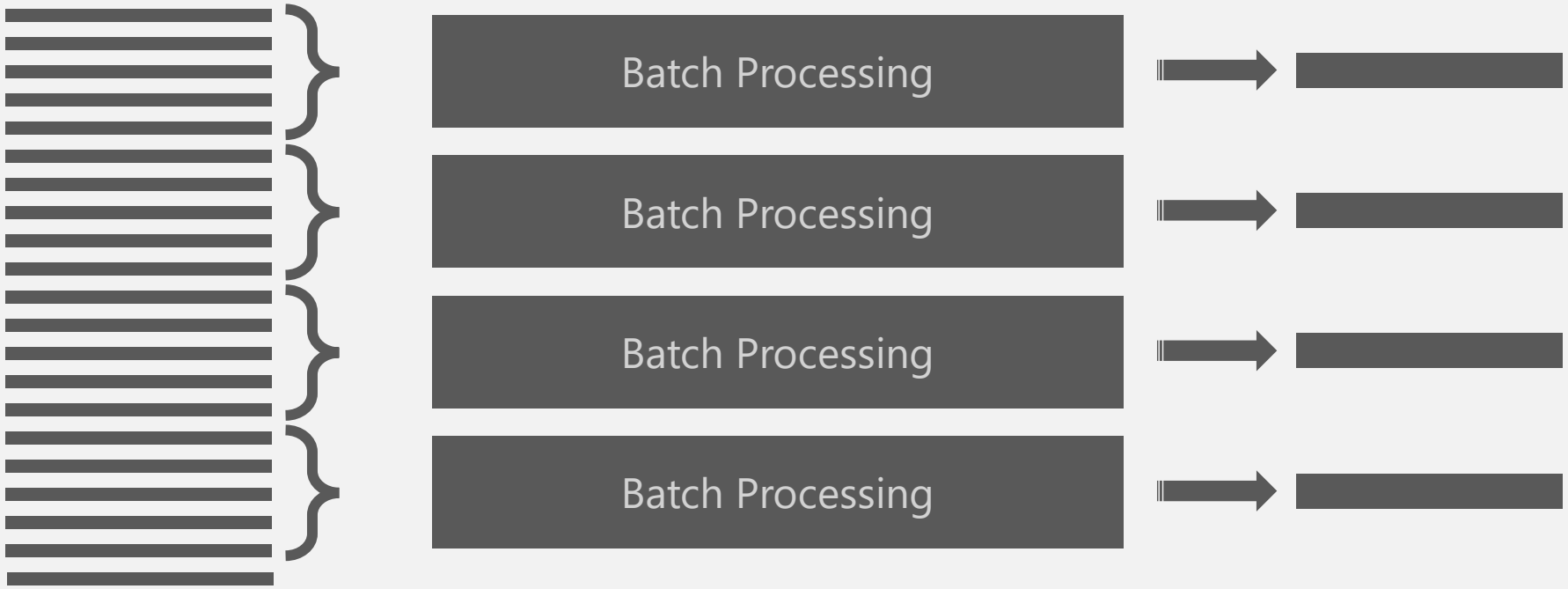
Use Batch and Stream technologies together to balance latency, throughput and fault-tolerance



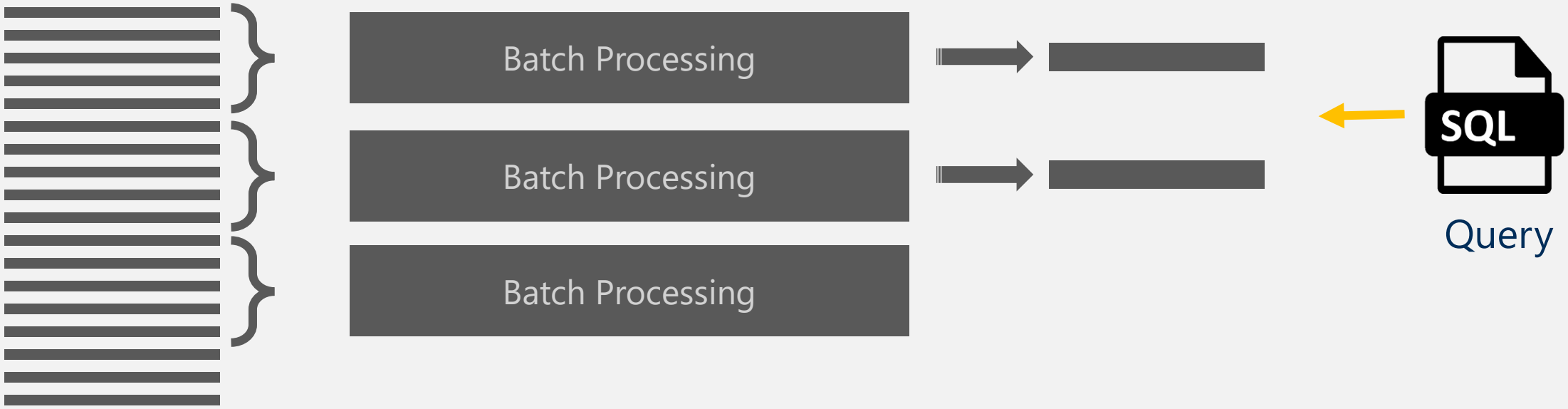
Nathan Marz
& James Warren

* Pages 14 to 20

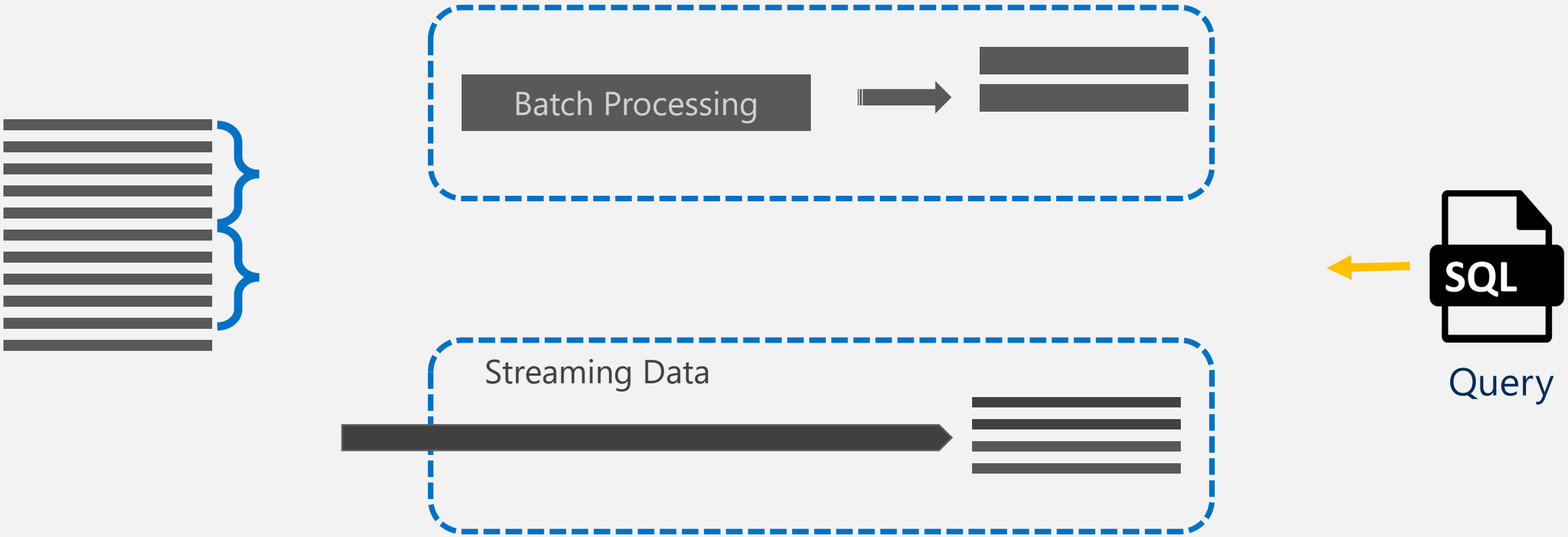
Problem: Timely Data Insights



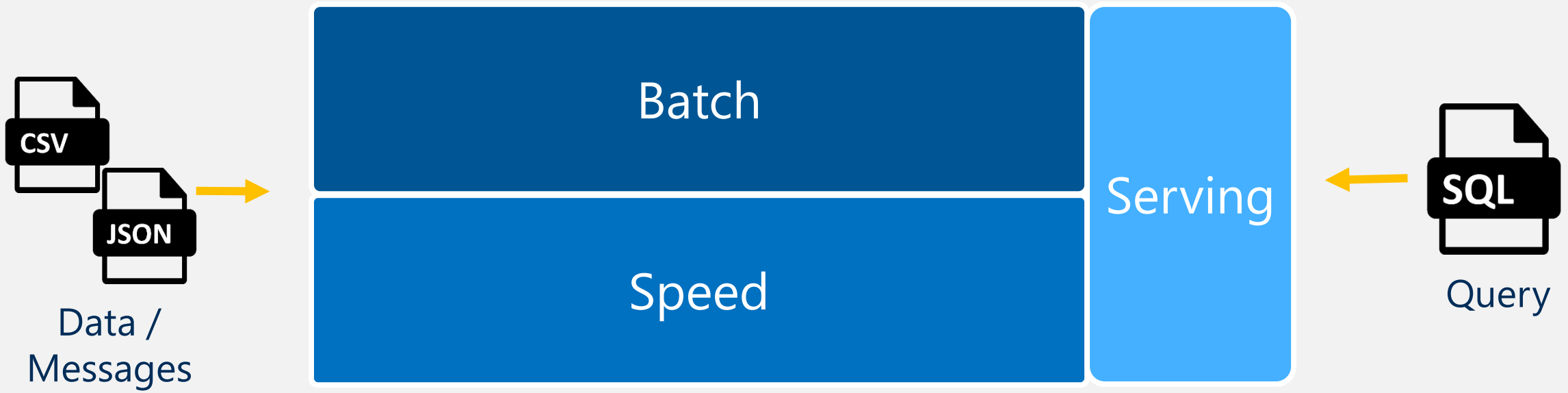
Problem: Timely Data Insights



Solution

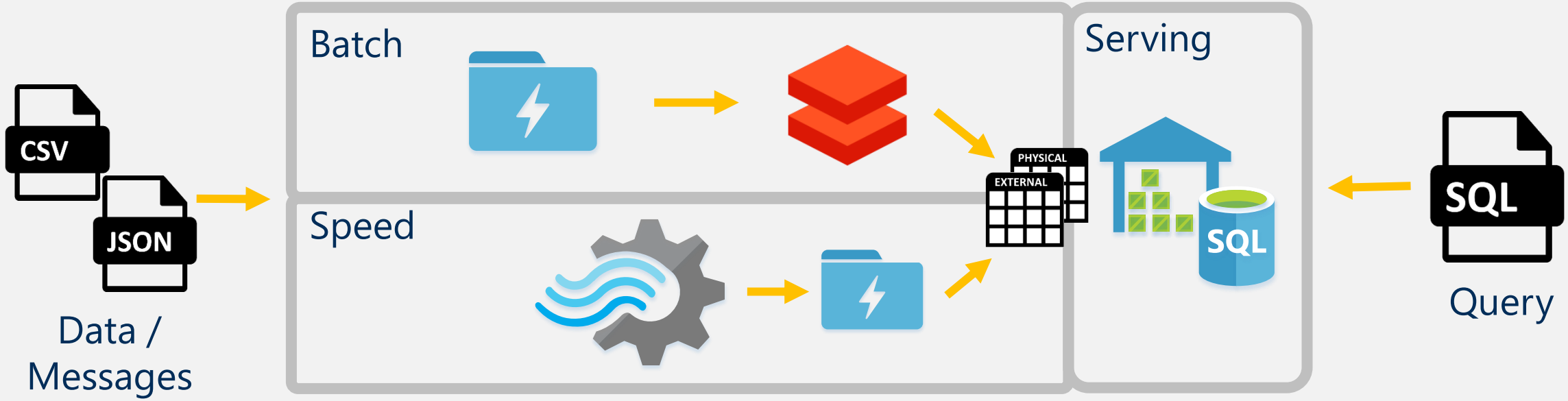


Lambda Architecture



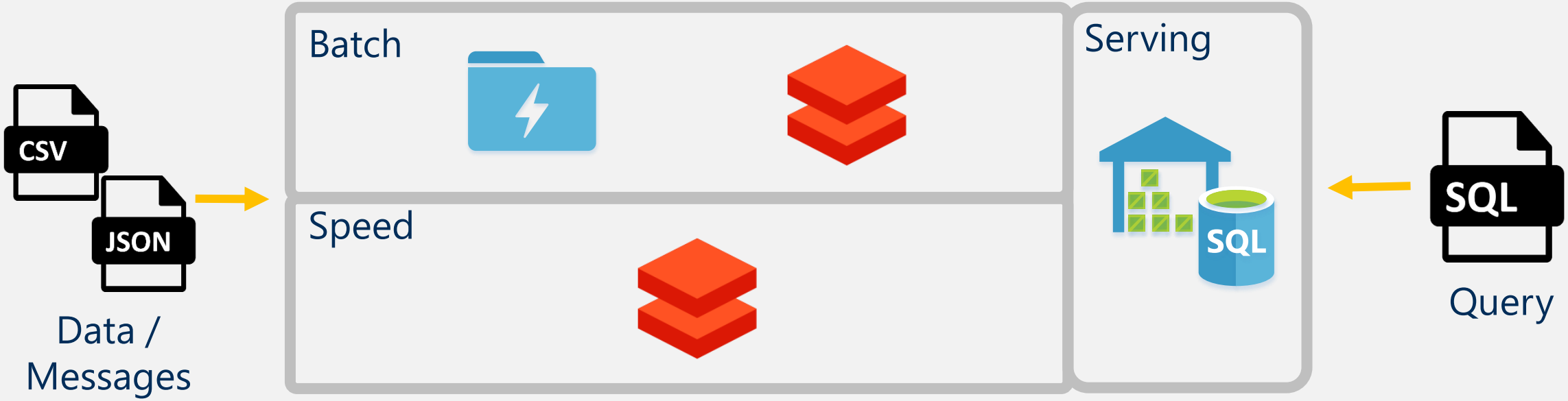
The Marz Lambda Architecture

Applying a Lambda Architecture in Azure

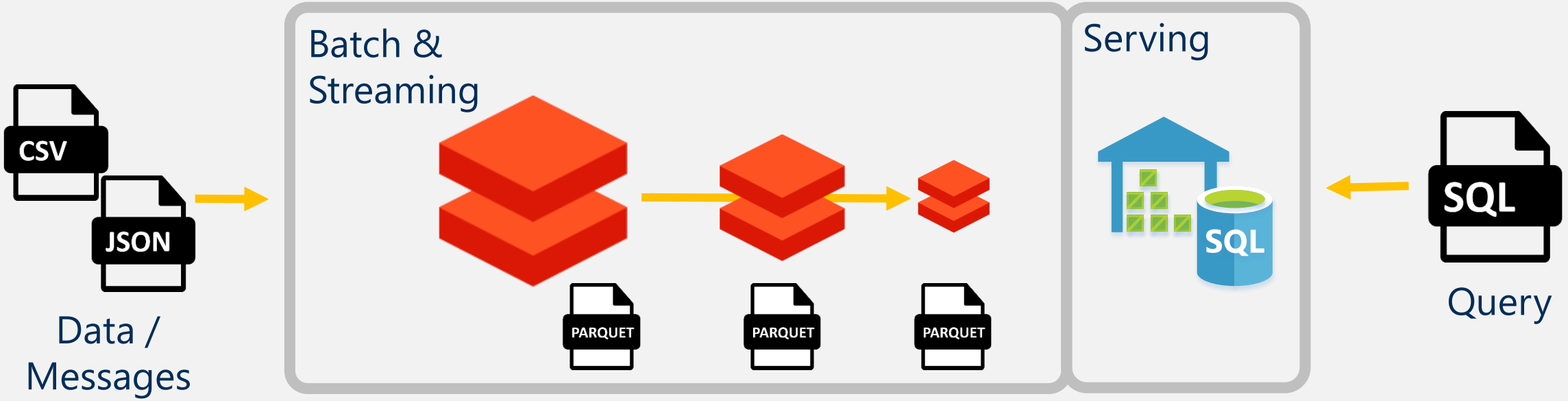


Simon Whiteley – PASS Summit 2018:
https://www.pass.org/summit/2018/Sessions/Details.aspx?name=modern-data-warehousing-the-lambda-approach-to-azure-bi

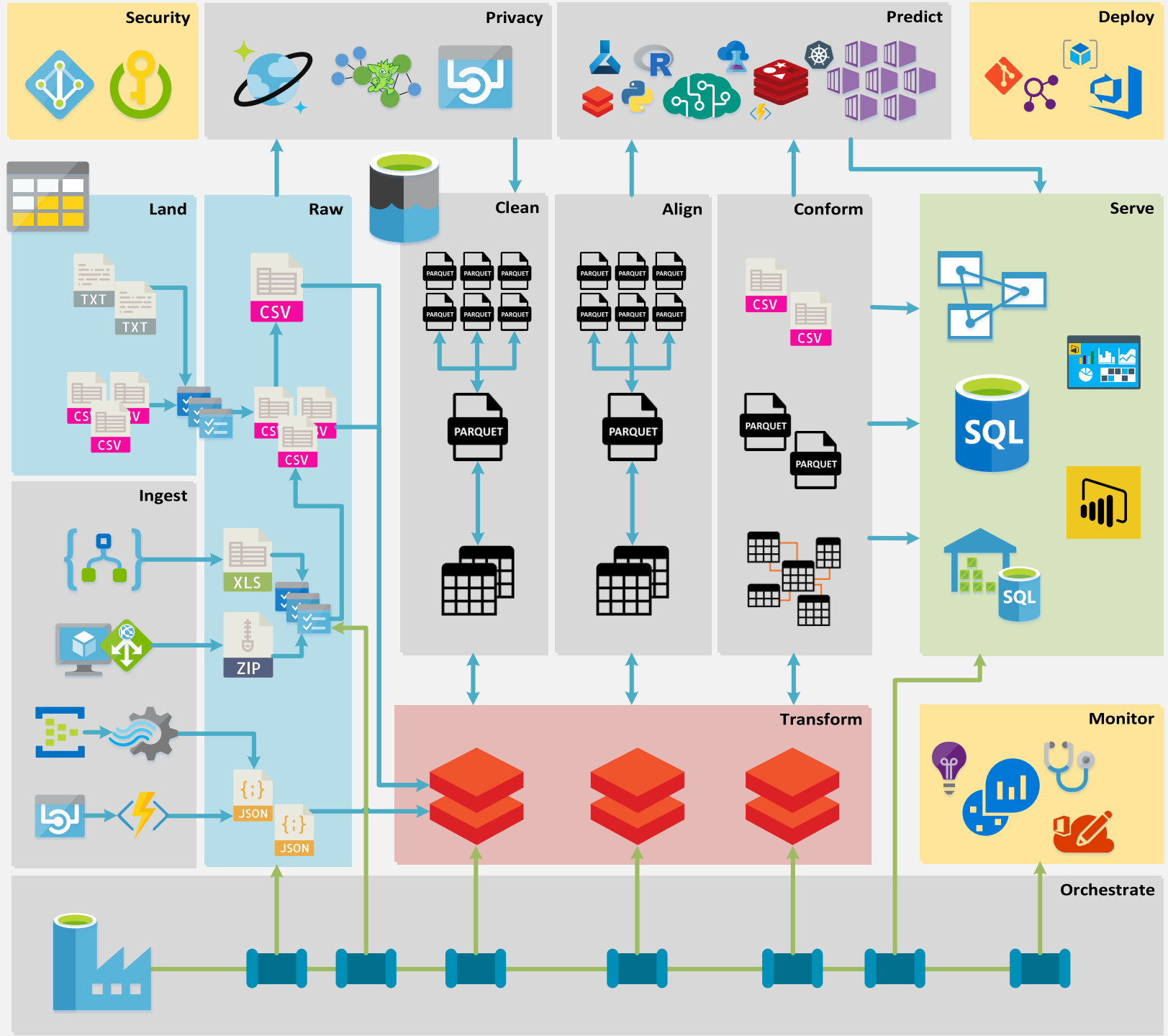
Applying a Lambda Architecture in Azure



Applying a ~~Lambda~~ Architecture in Azure
^ Databricks Delta
(AKA Delta Lake)



Modern Data Warehouse



altius

Azure Stream Analytics

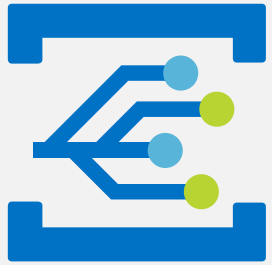
- **Real-time data problems** ✓
- **What is ASA and why use it** ✓
- **Production Considerations** ✓
- **Lambda Architecture** ✓

Other Thoughts…

Are there any other real-time tools and techniques we should consider?

2016

2019



Thanks for Listening

Paul Andrew

@MrPaulAndrew

Microsoft® Most Valuable Professional

altius

Email: paul@mrpaulandrew.com
Blog: mrpaulandrew.com
GitHub: github.com/mrpaulandrew

<<< Slides PDF