

Azure Stream Analytics

Beyond IoT Real-time Data Ingestion

Paul Andrew



altius



@MrPaulAndrew



In/MrPaulAndrew



MrPaulAndrew.tech



<https://github.com/mrpaulandrew>

CommunityEvents

Demo code, content and slides from various community events.

● C++

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

Azure Stream Analytics



Real-time data problems

What is ASA and why use it

Production Considerations

Lambda Architecture

Azure Stream Analytics



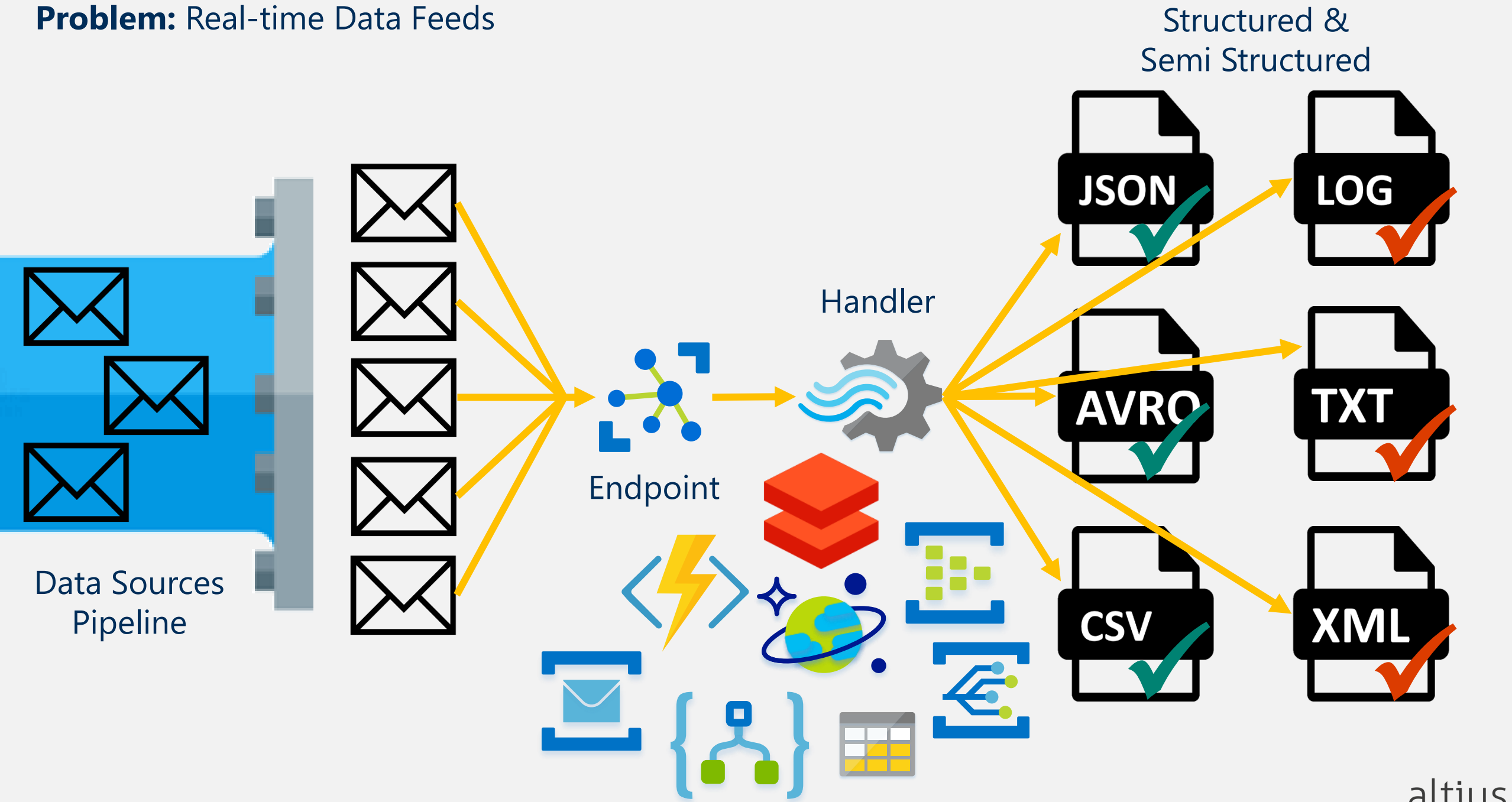
Real-time data problems

What is *ASA* and why use it

Production Considerations

Lambda Architecture

Problem: Real-time Data Feeds



Azure Stream Analytics

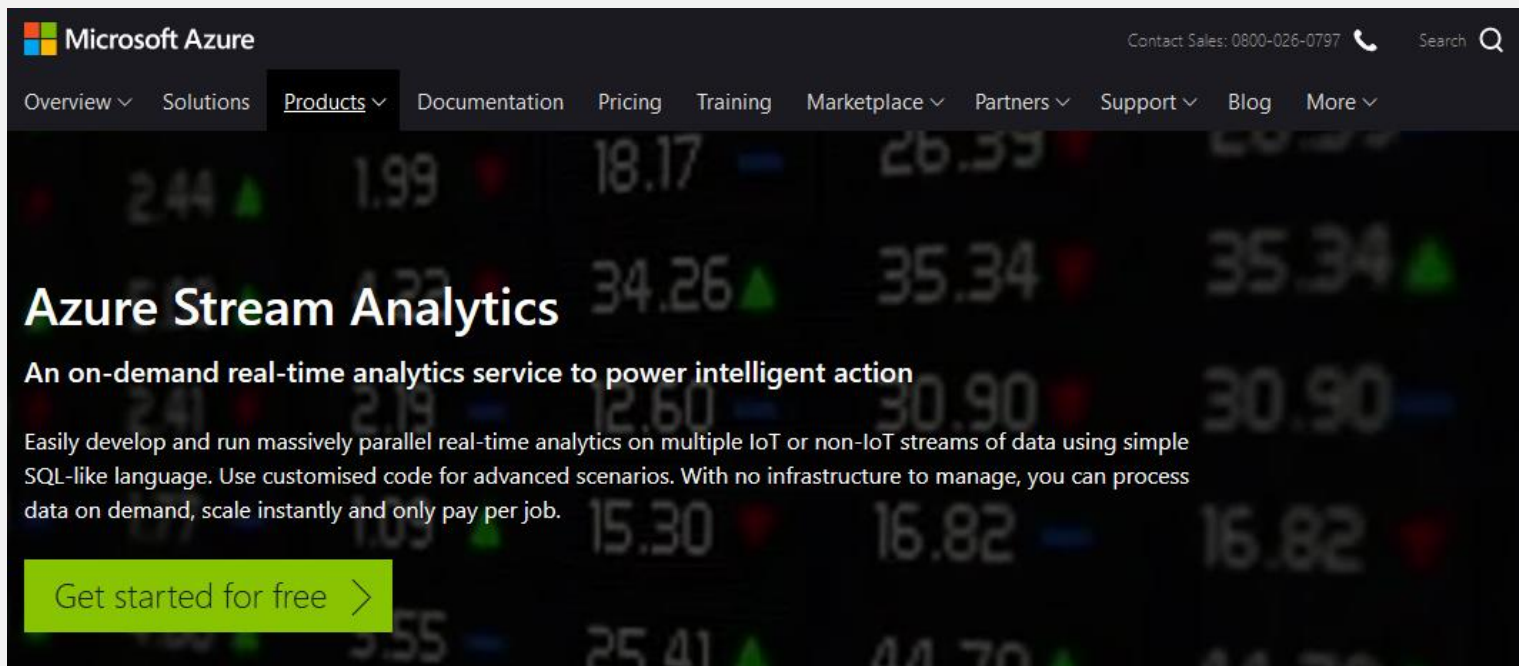


Real-time data problems

What is ASA and why use it

Production Considerations

Lambda Architecture



The screenshot shows the Microsoft Azure website's navigation bar with links for Overview, Solutions, Products, Documentation, Pricing, Training, Marketplace, Partners, Support, Blog, and More. The main heading is "Azure Stream Analytics" with the tagline "An on-demand real-time analytics service to power intelligent action". Below this, a paragraph describes the service: "Easily develop and run massively parallel real-time analytics on multiple IoT or non-IoT streams of data using simple SQL-like language. Use customised code for advanced scenarios. With no infrastructure to manage, you can process data on demand, scale instantly and only pay per job." A green button labeled "Get started for free" is at the bottom left of the screenshot.

Azure Stream Analytics



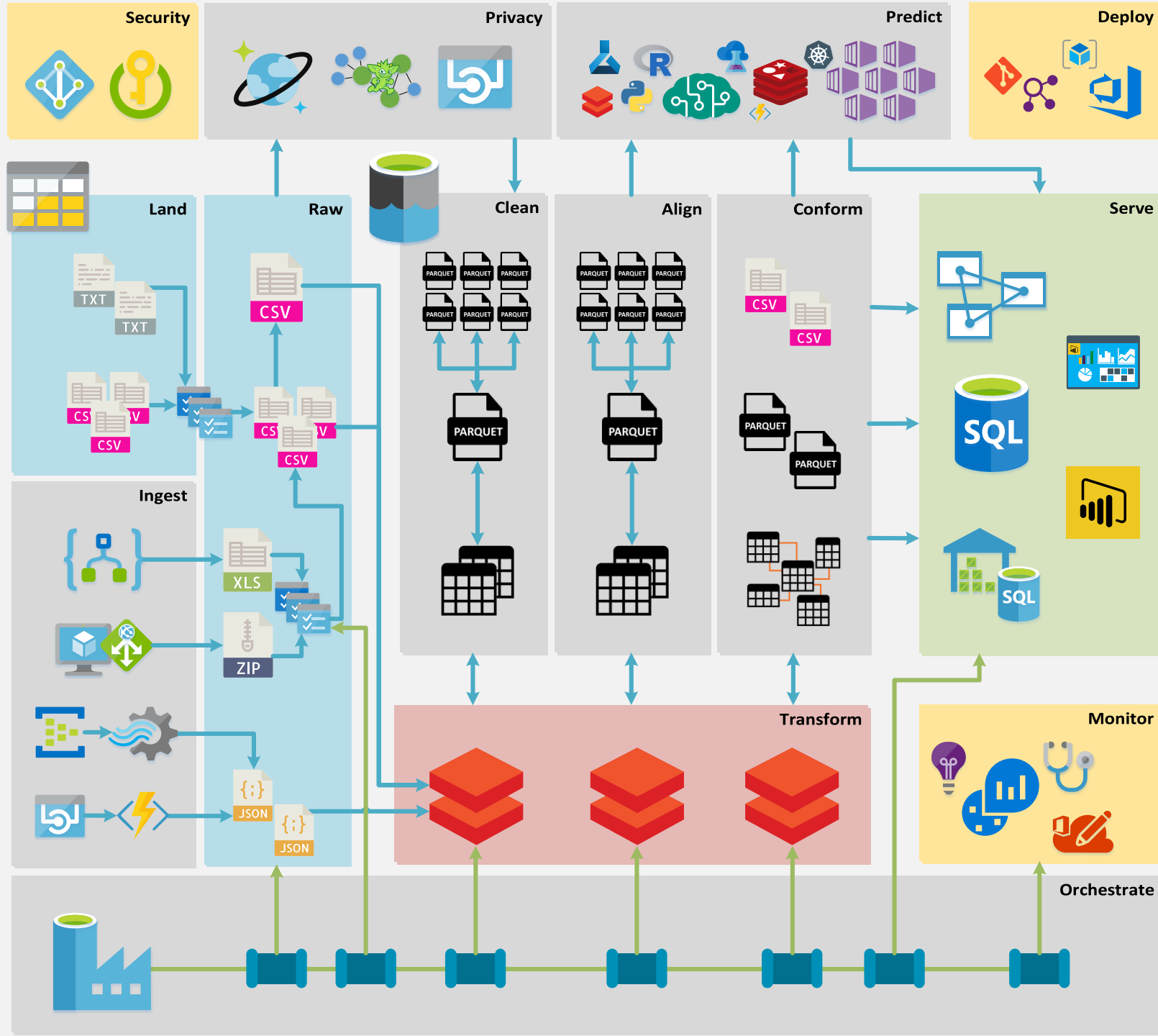
<https://azure.microsoft.com/en-gb/services/stream-analytics/>

Real-time data problems

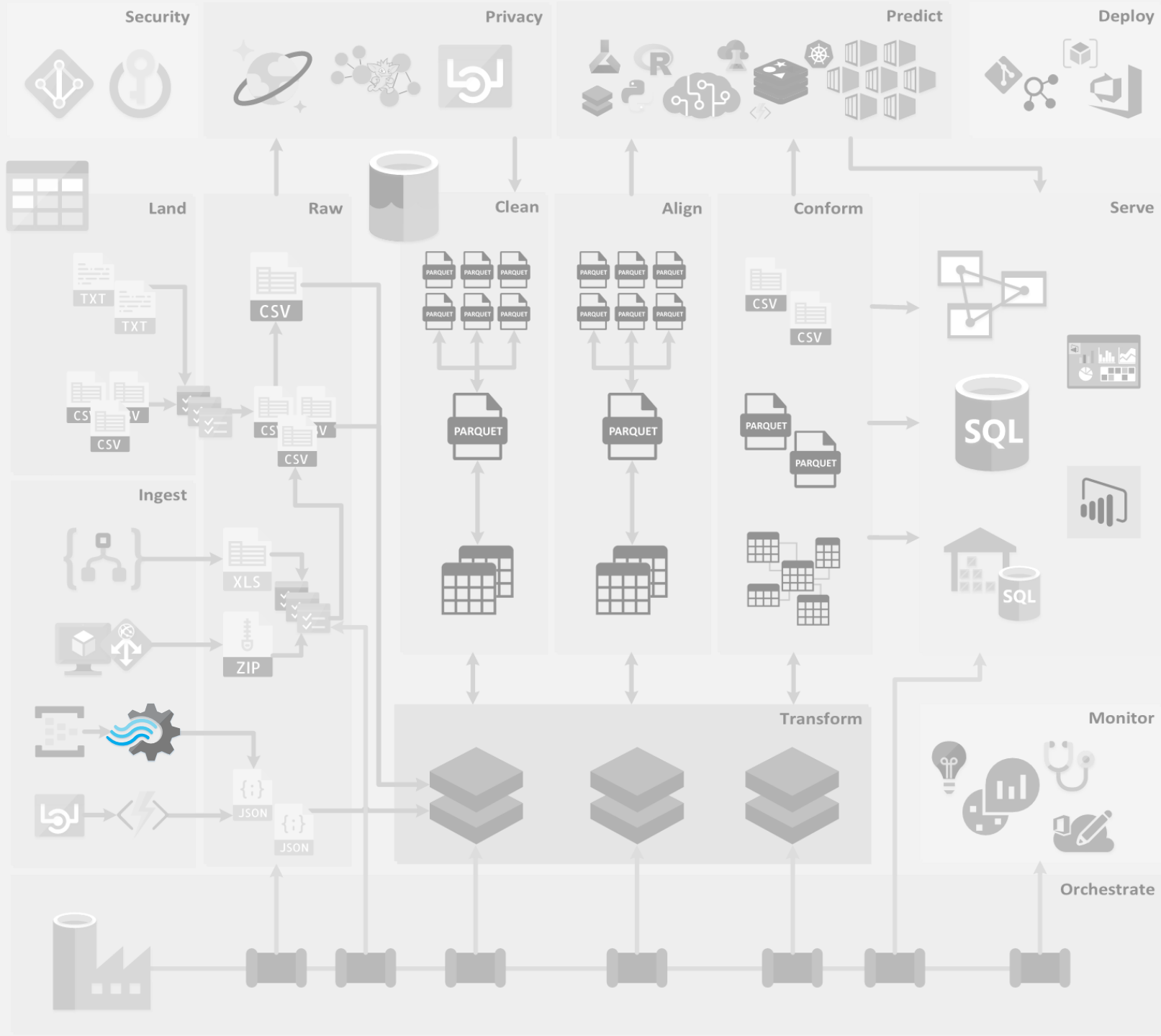
What is ASA and why use it

Production Considerations

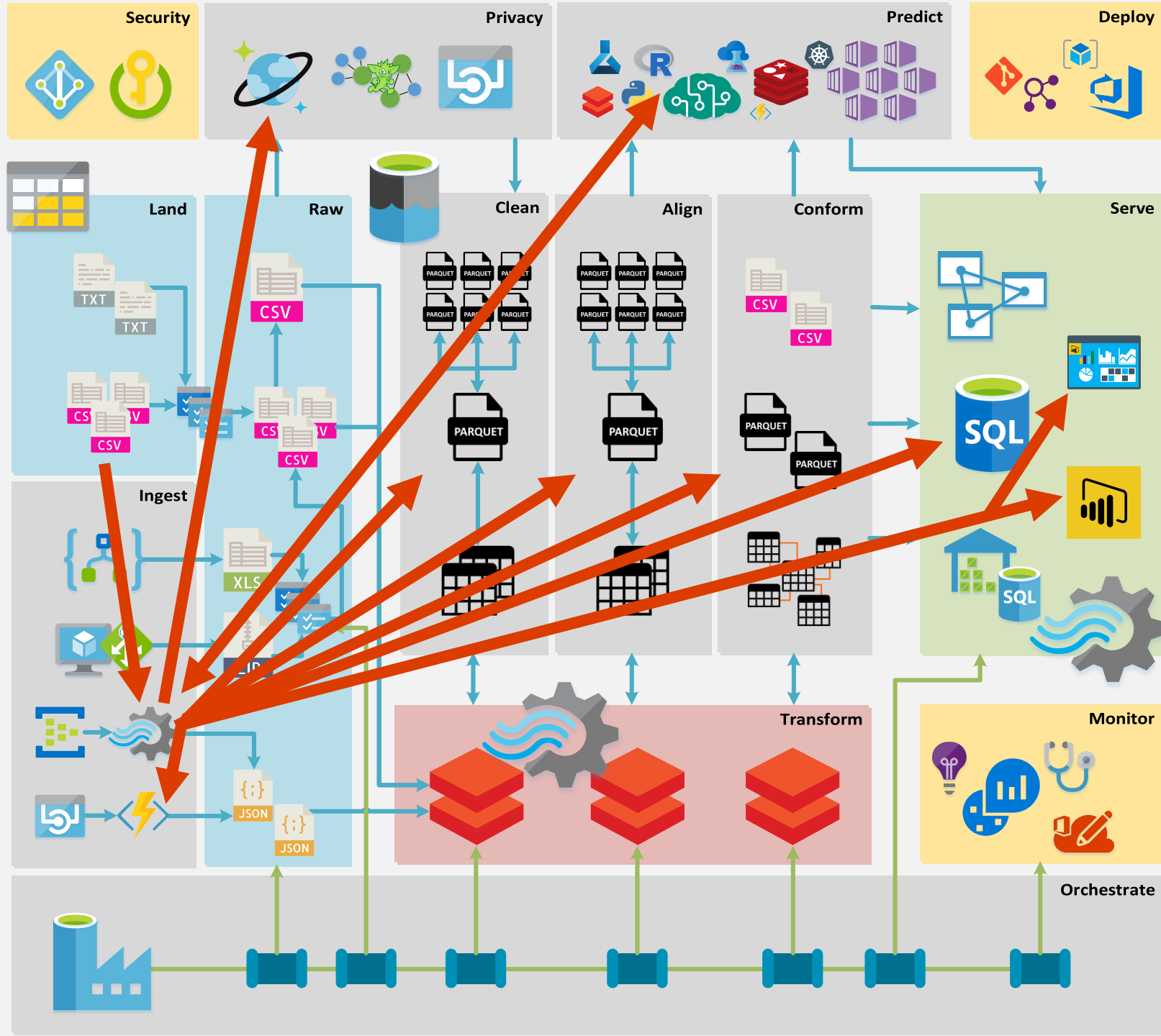
Lambda Architecture



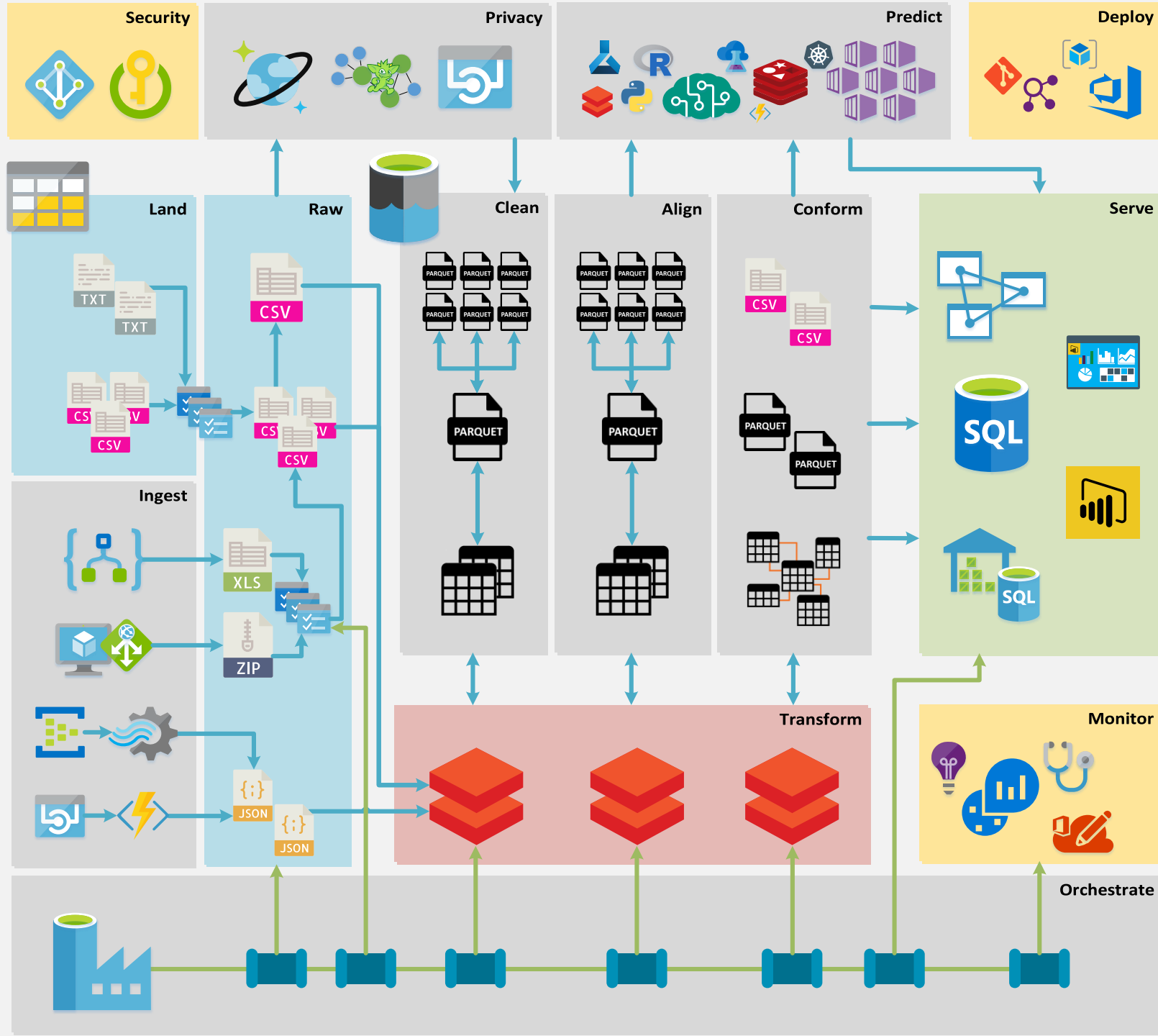
What is ASA and why use it



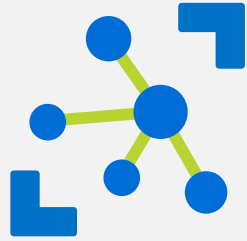
What is ASA and why use it



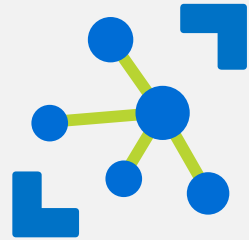
What is ASA and why use it



What is ASA and why use it

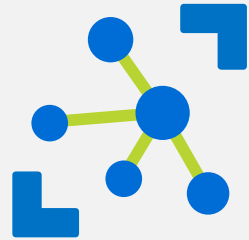


Azure IoT Hub vs Azure Event Hub

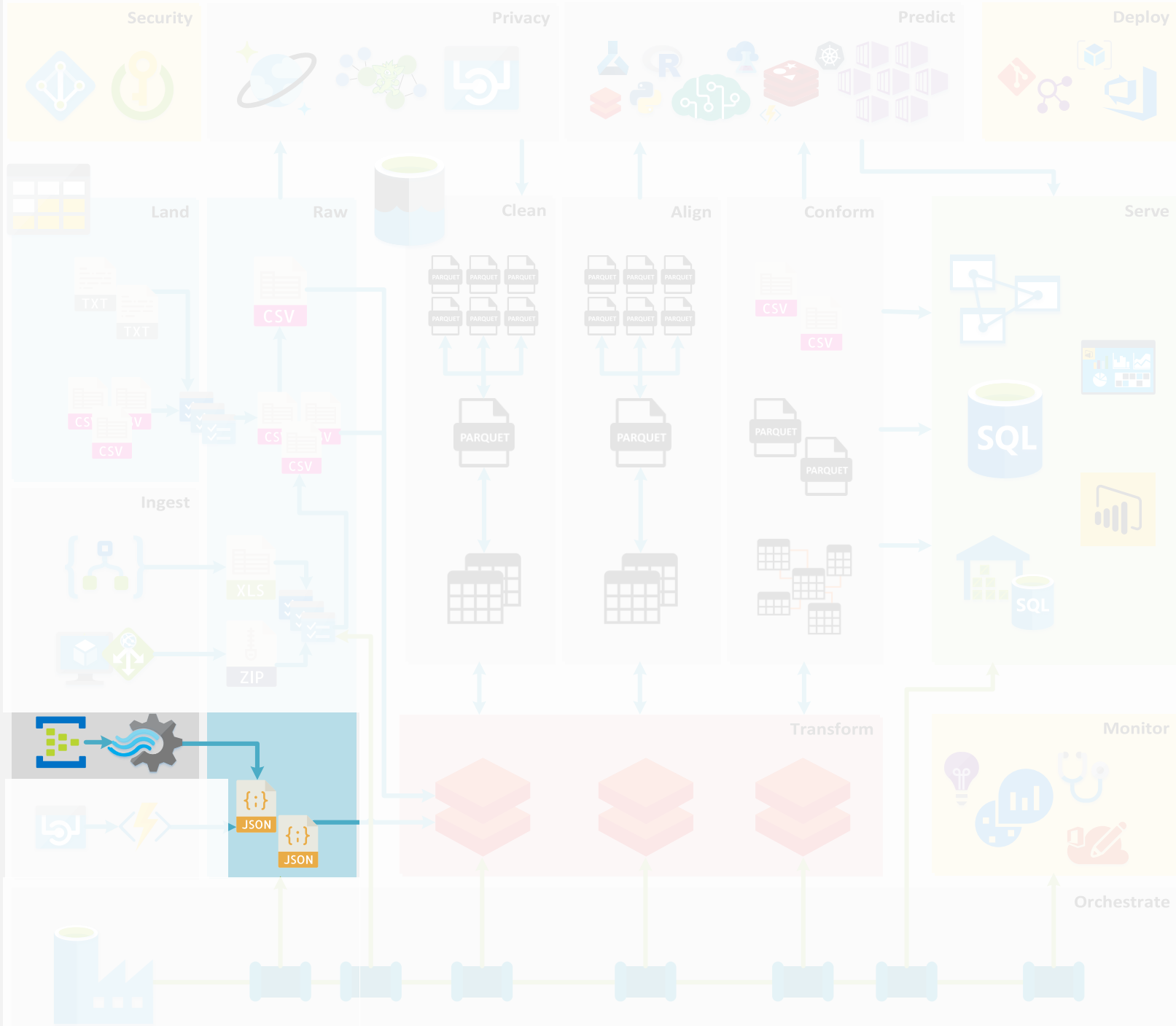


Feature	Azure IoT Hub	Azure Event Hub
Message Direction	2 Way	1 Way
Protocol Support	MQTT, AMQP, HTTP	AMQP, HTTP
Scaling	Configured	Automatic
Message Routing	Yes	No
Security	Device Level	Hub Level
Device State Support	Yes	No
Message Capturing	No	Yes
Multiple Namespaces	No	Yes
Tiers	F1/S1/S2/S3	Basic/Standard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

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)



What is ASA and why use it



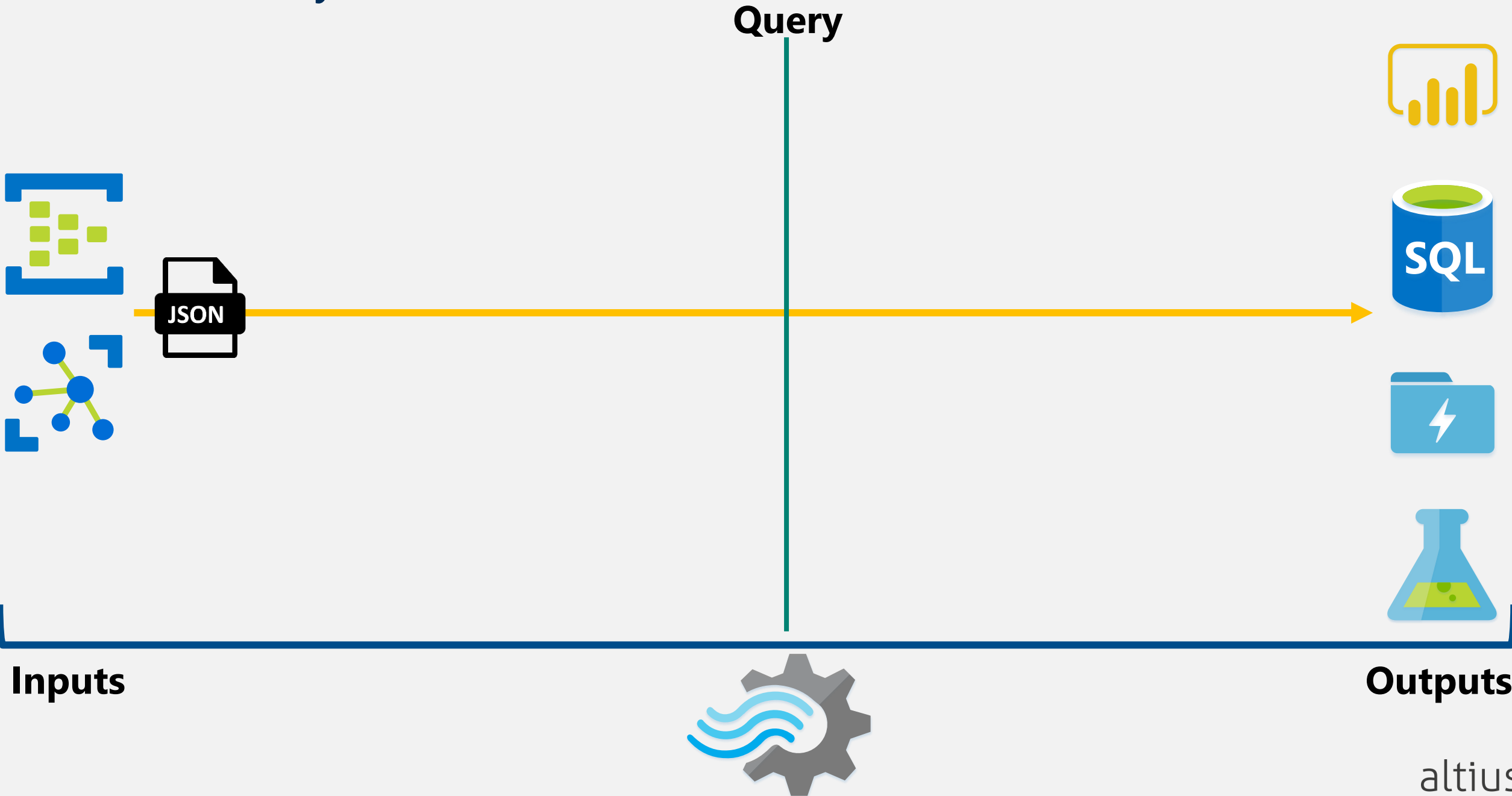
Azure Stream Analytics



Inputs **Outputs**



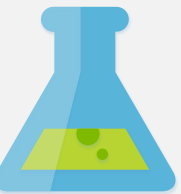
Azure Stream Analytics



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



Azure Stream Analytics

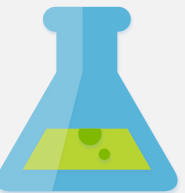
Query

```
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



Azure Stream Analytics

Query



*100MB
Limit

```
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)
```



Inputs... Source Data
Reference Data
Custom Functions



Outputs

Azure Stream Analytics

Query

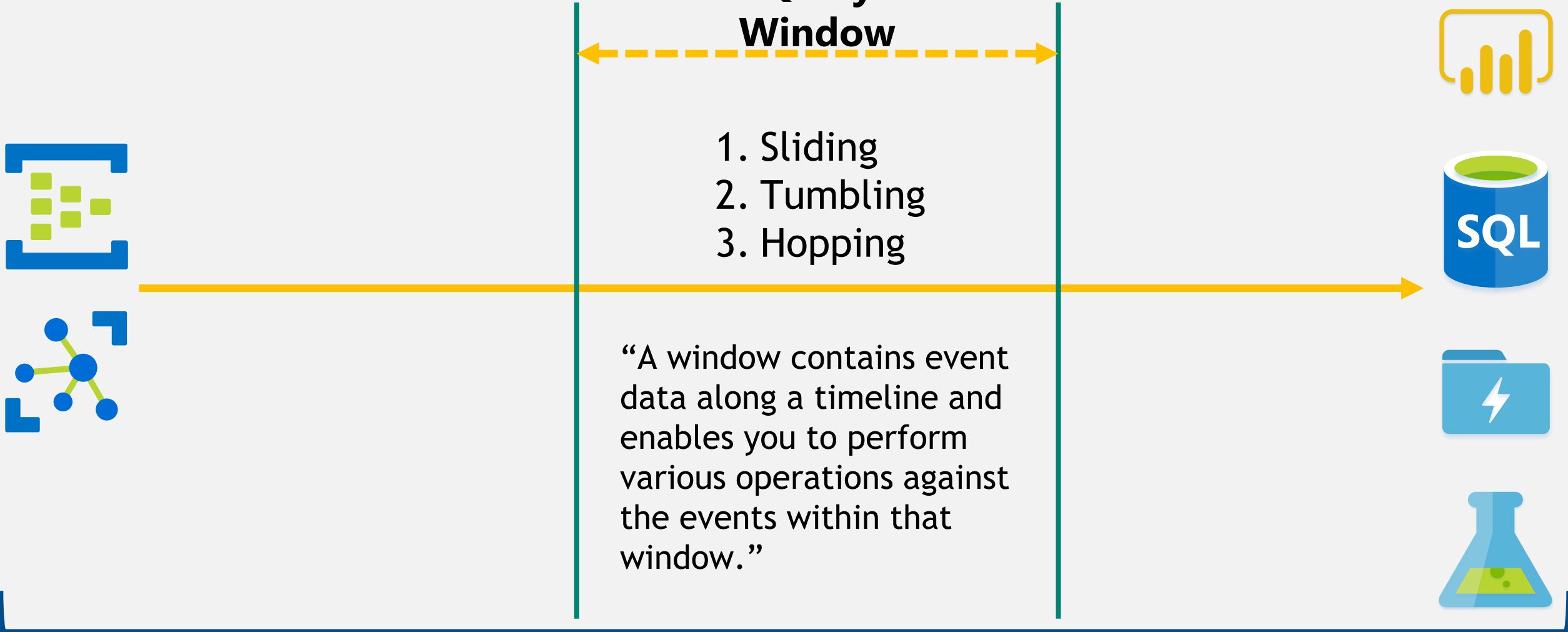


Inputs

Outputs



Azure Stream Analytics

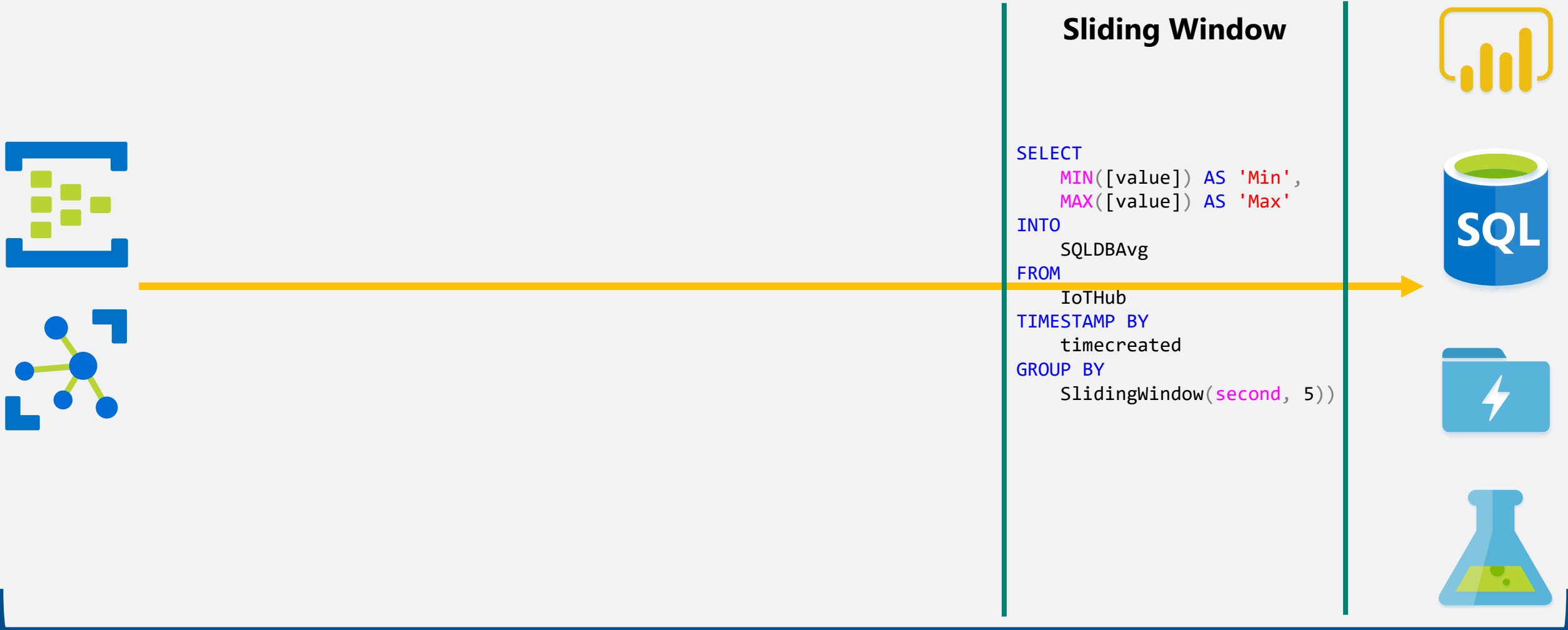


Inputs

Outputs



Azure Stream Analytics



Inputs Outputs



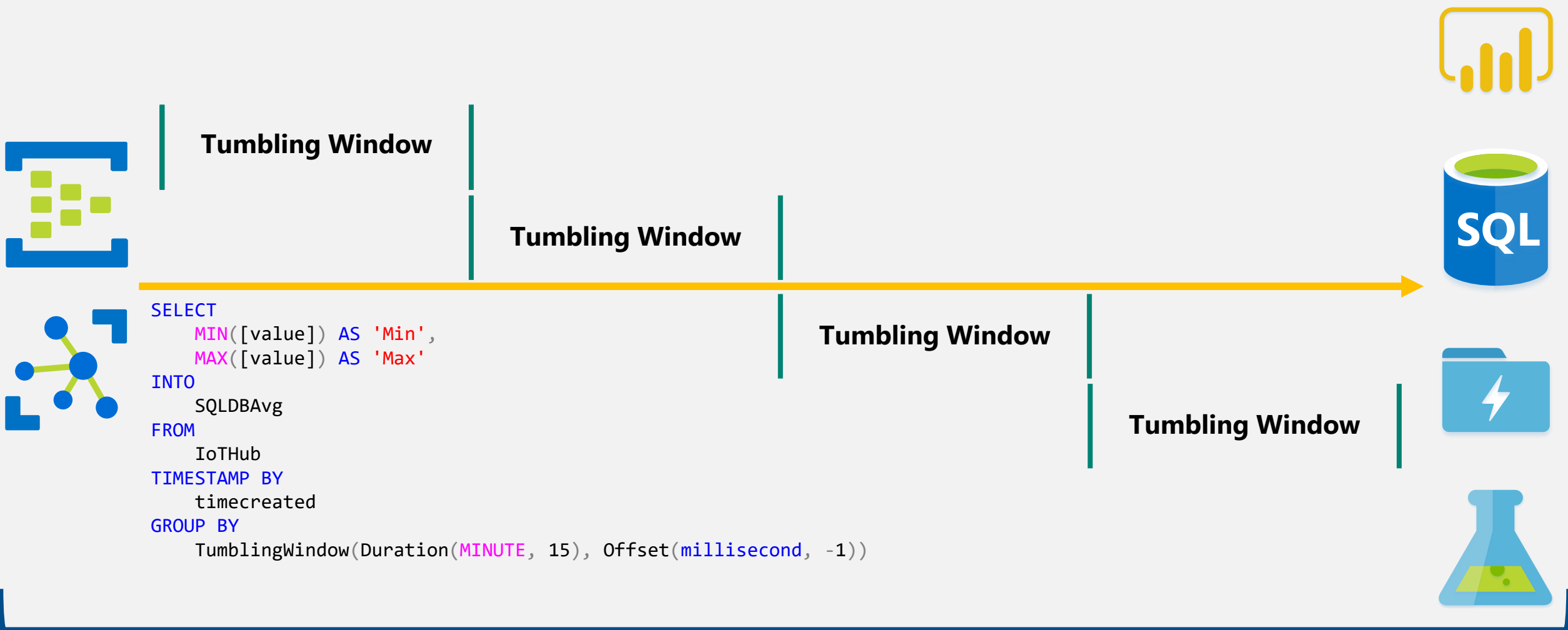
Azure Stream Analytics



Inputs **Outputs**



Azure Stream Analytics



Inputs

Outputs



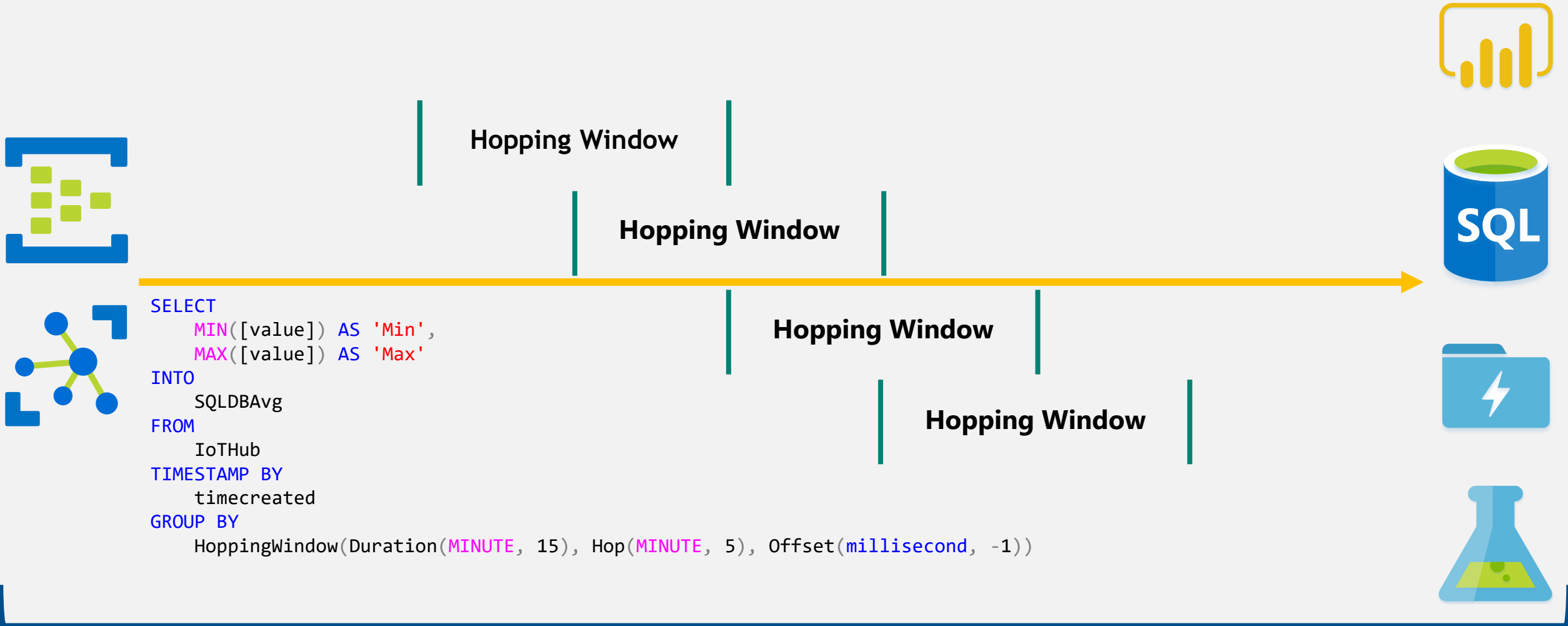
Azure Stream Analytics



Inputs **Outputs**



Azure Stream Analytics

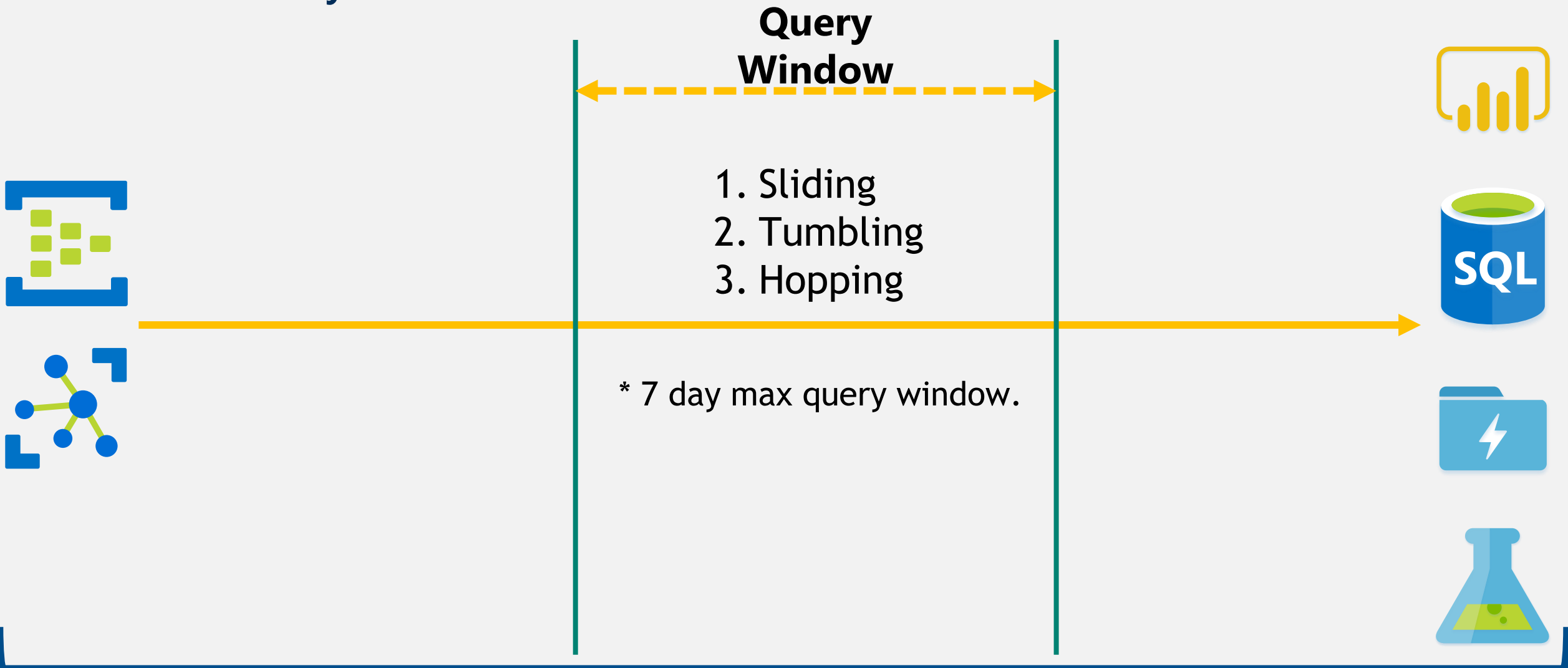


```
SELECT
    MIN([value]) AS 'Min',
    MAX([value]) AS 'Max'
INTO
    SQLDBAvg
FROM
    IoTHub
TIMESTAMP BY
    timecreated
GROUP BY
    HoppingWindow(Duration(MINUTE, 15), Hop(MINUTE, 5), Offset(milliseconds, -1))
```

Inputs

Outputs

Azure Stream Analytics



Inputs

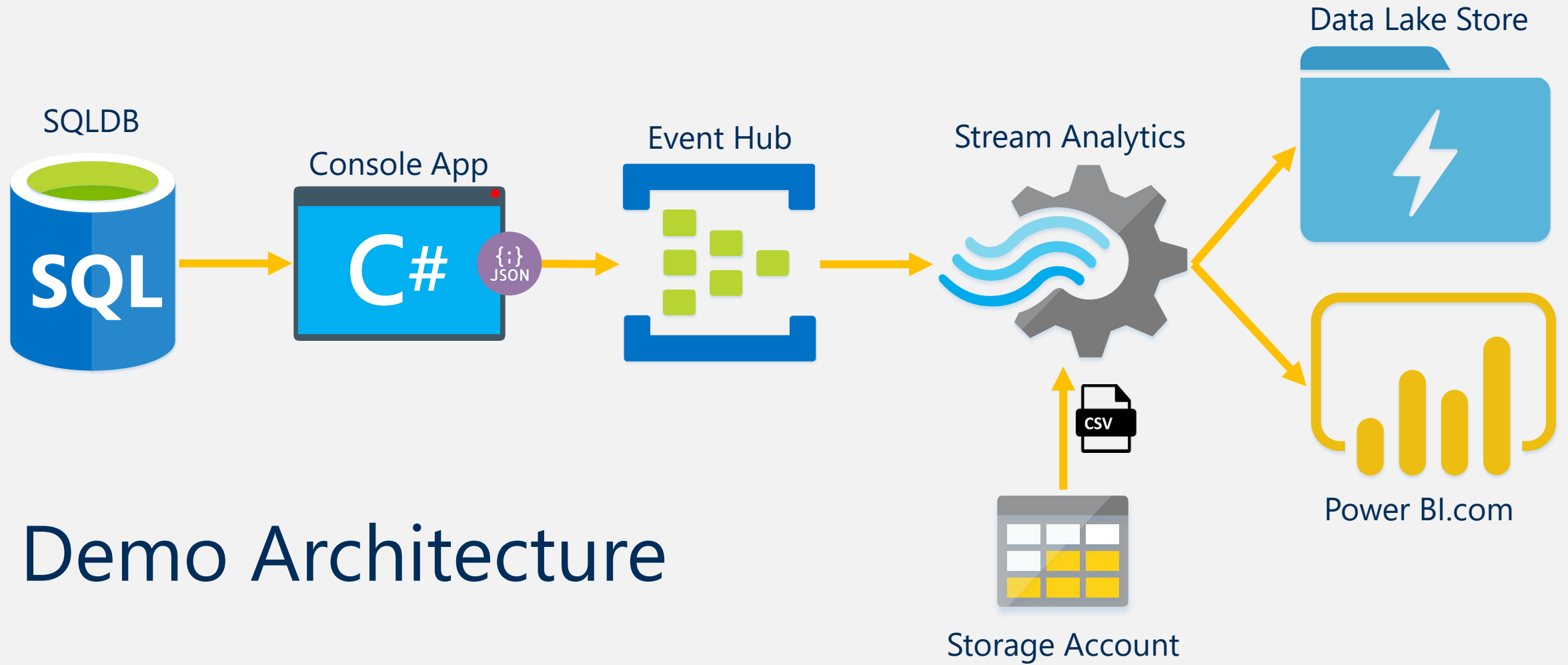
<https://msdn.microsoft.com/en-us/library/azure/dn835019.aspx>



Outputs

Demo

altius



Azure Stream Analytics



Real-time data problems

What is *ASA* and why use it

Production Considerations

Lambda Architecture

Production Considerations



On
Premises



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.



On
Premises



VPN

Express
Route



VNet



SQL



Production Considerations



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



On Premises



Express Route



VNet



3. Azure Stream Analytics doesn't currently support Service Endpoints. Coming at some point.



Production Considerations



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



On Premises



Express Route



VNet



3. Azure Stream Analytics doesn't currently support Service Endpoints. Coming at some point.



VNet



VNet

4. What other services in your solution need to use Service Endpoints?

Production Considerations



On
Premises



Capture



VNet



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



Production Considerations



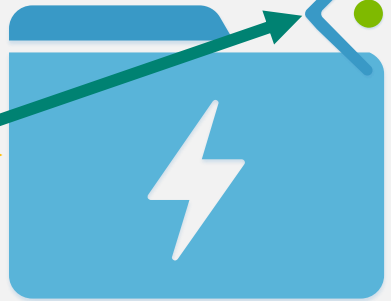
On
Premises



kafka



SQL



6. Peering gets removed if the Databricks cluster gets deallocated.

9. All private network connections with VNet peering.

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.

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

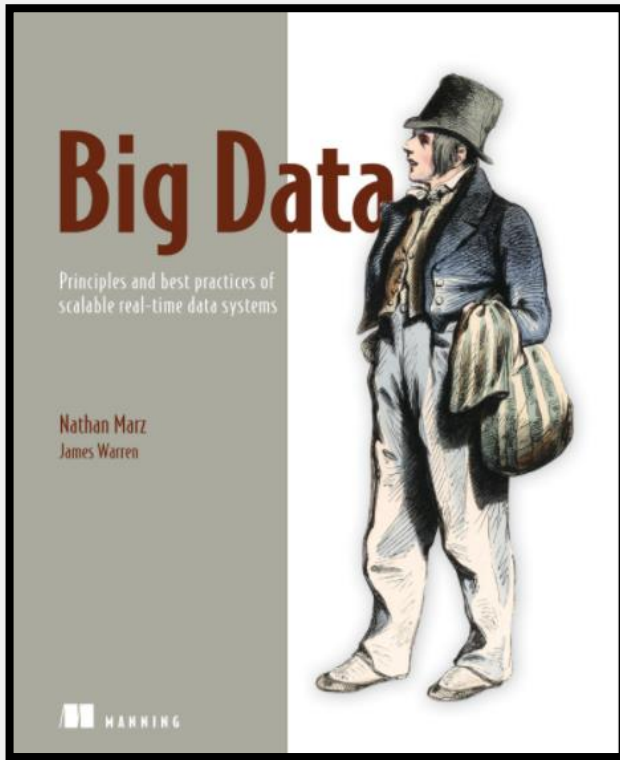
What is *ASA* and why use it

Production Considerations

Lambda Architecture

Lambda Architecture

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

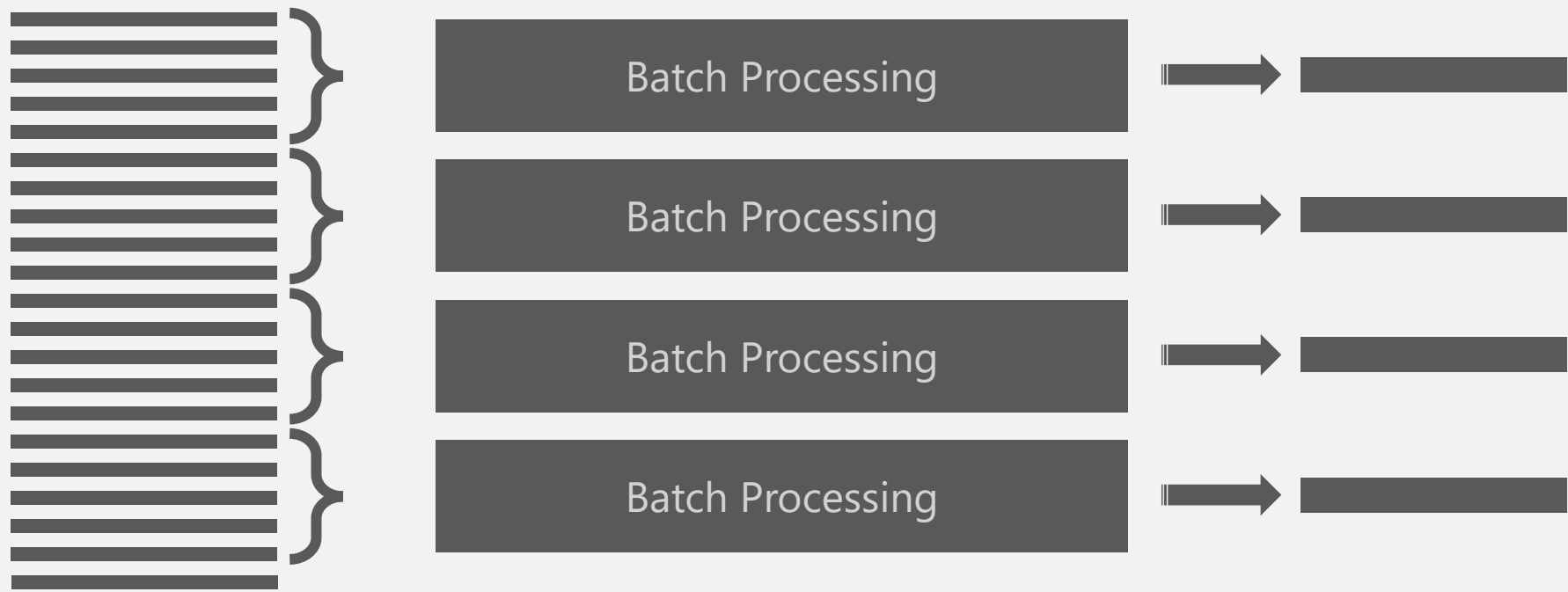


Nathan Marz
& James Warren

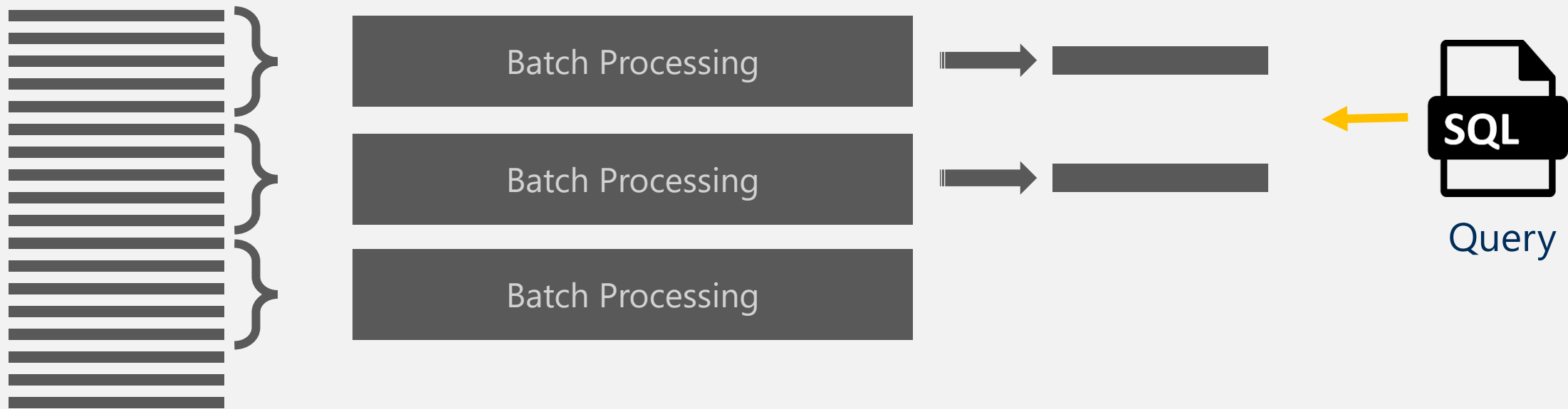


* Pages 14 to 20

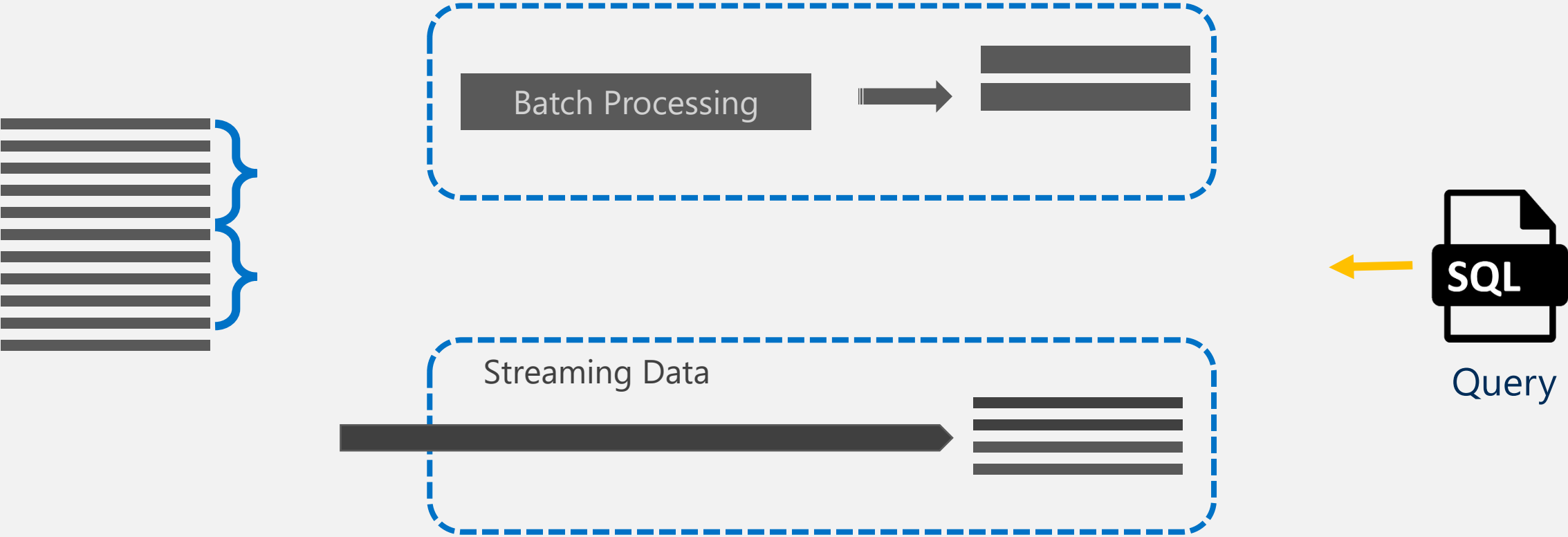
Problem: Timely Data Insights



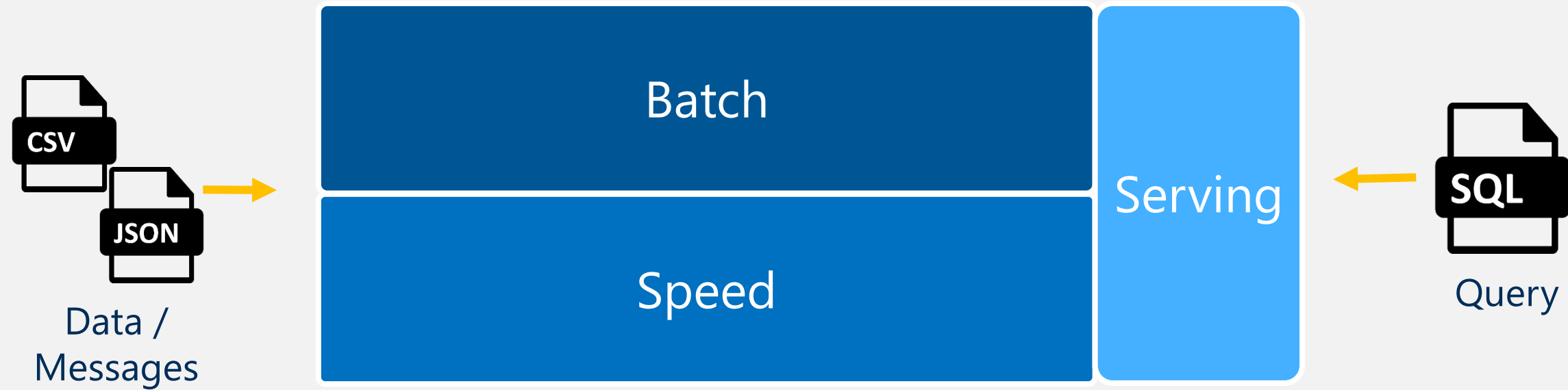
Problem: Timely Data Insights



Solution



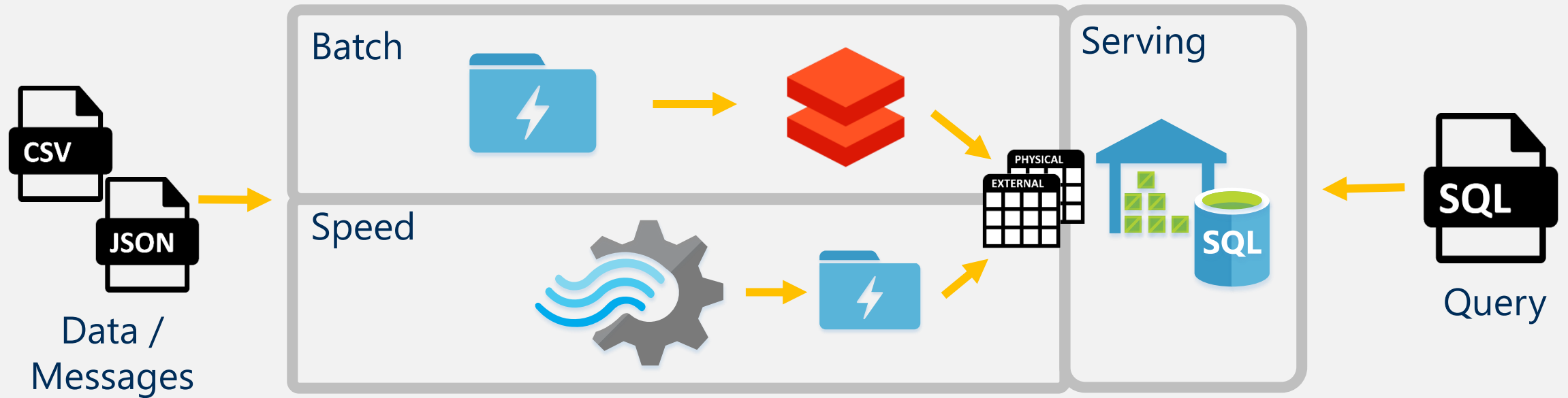
Lambda Architecture



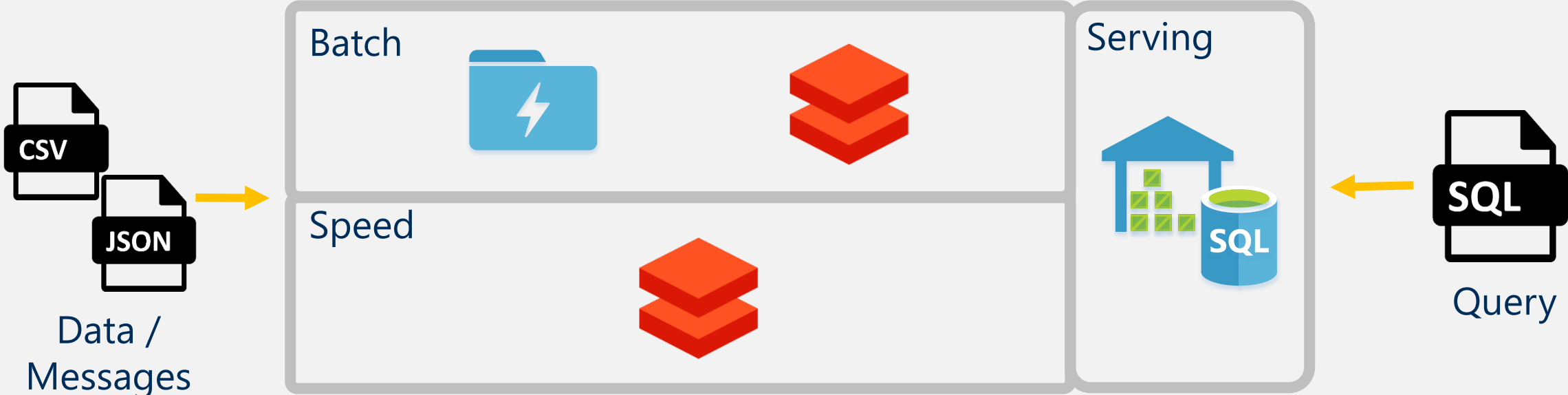
The Marz Lambda Architecture



Applying a Lambda Architecture in Azure

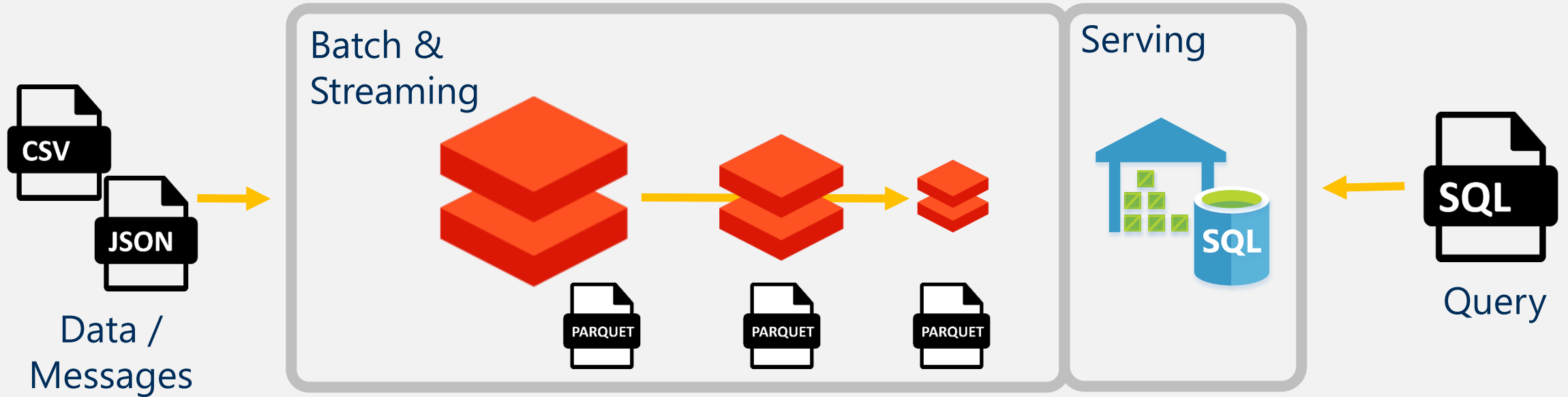


Applying a Lambda Architecture in Azure

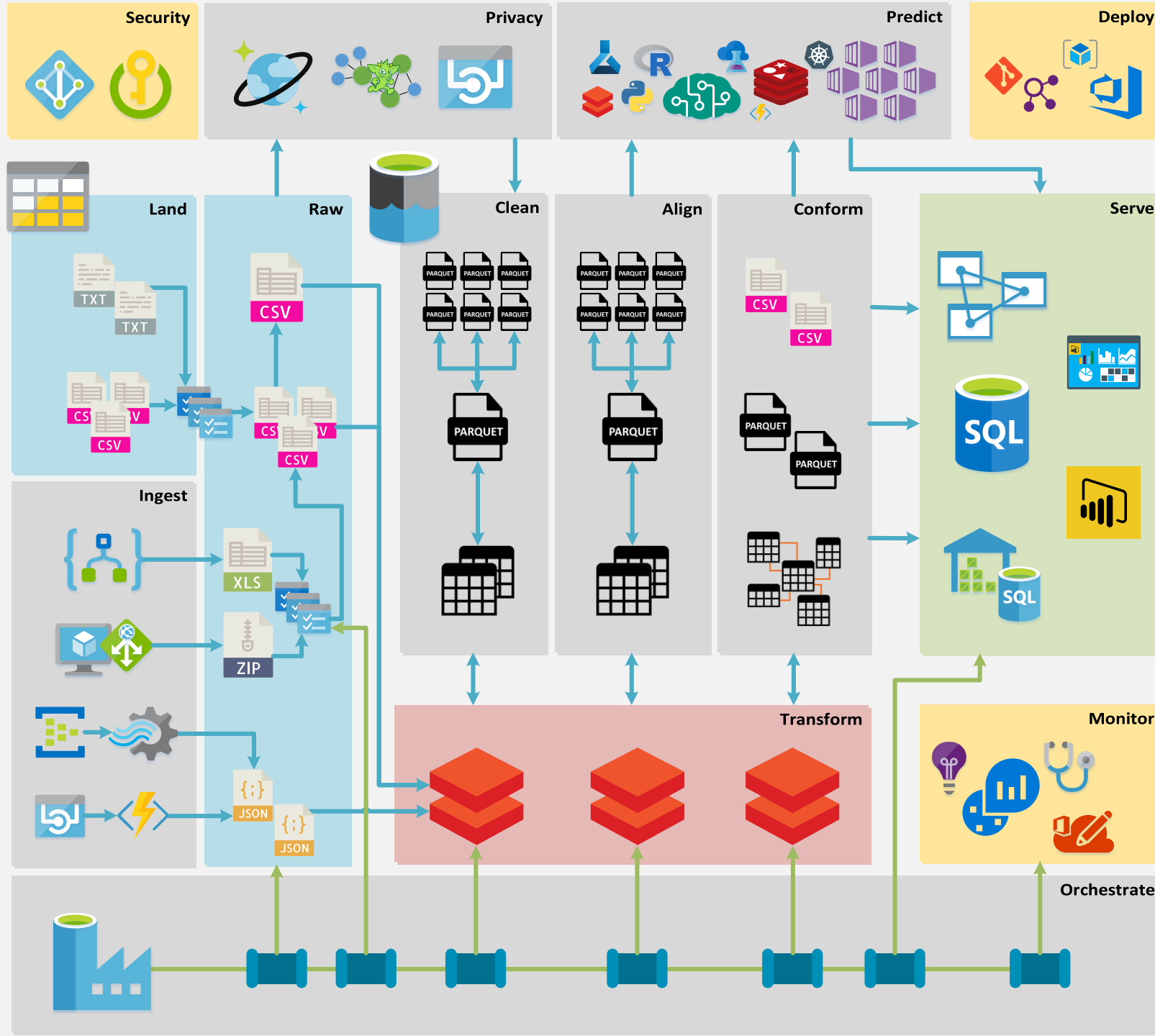


Applying a ~~Lambda~~ Architecture in Azure

^ Databricks Delta
(AKA Delta Lake)



Modern Data Warehouse



Azure Stream Analytics



Real-time data problems ✓

What is ASA and why use it ✓

Production Considerations ✓

Lambda Architecture ✓

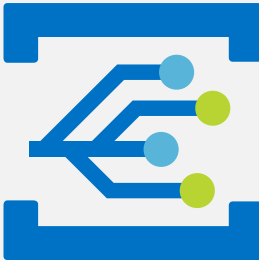
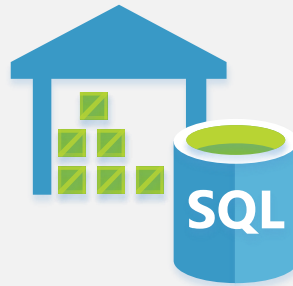
Other Thoughts...

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



2016

2019



Thanks for Listening

Paul Andrew

 @MrPaulAndrew



altius



Email: paul@mrpaulandrew.com

Blog: mrpaulandrew.com

GitHub: github.com/mrpaulandrew

<<< Slides

PDF

