PASS MARATHON

Azure Stream Analytics

Beyond IoT Real-time Data Ingestion

Paul Andrew, Adatis Consulting Ltd

Presenting Sponsor

Quest

Technical Assistance

If you require assistance during the session, type your inquiry into the question pane on the right side.

Maximize your screen with the zoom button on the top of the presentation window.

Please fill in the short evaluation following the session. It will appear in your web browser.

PASS MARATHON

Thank you to our Presenting Sponsor

PASS MARATHON

Explore everything PASS has to offer

Free Online Resources

Newsletters

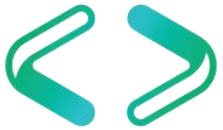
PASS.org

PASS SUMMIT

The largest conference for technical professionals leveraging the Microsoft Data Platform.

PASS LOCAL GROUPS

Local user groups around the world



PASS SQLSATURDAY

Free 1-day local training events

PASS VIRTUAL GROUPS

Online special interest user groups

PASS MARATHON

Business analytics training

PASS VOLUNTEERS

Get involved

Paul Andrew

Microsoft Most Valuable Professional

Senior Data Analytics Consultant
Adatis

mrpaulandrew.com

/mrpaulandrew

@mrpaulandrew

github.com/mrpaulandrew

paul@mrpaulandrew.com

Microsoft Data Platform MVP

First awarded in July 2017 having been an active member of the data platform community for many years. Has delivered talks at PASS Summit, SQL Bits, Data Relay, various SQL Saturdays, Data Minds, SQL Day and Big Data London.

Azure Data Engineer & Architect

5+ years experience designing and building modern data warehouse solutions using the Microsoft cloud platform. Before that 10+ years developing BI systems using the complete on premises SQL Server stack.

Father, Husband, Blood Donor

Star Wars fan. Happy to sit playing with Lego for hours. Enjoys swimming, cycling and generally being outdoors.
Well done for reading all of this ☺

PASS MARATHON

Azure Stream Analytics

Beyond IoT Real-time Data Ingestion

Paul Andrew, Adatis Consulting Ltd

Presenting Sponsor

Quest

Azure Stream Analytics 

Real-time data problems

What is ASA and why use it

Production Considerations

Lambda Architecture

Azure Stream Analytics 

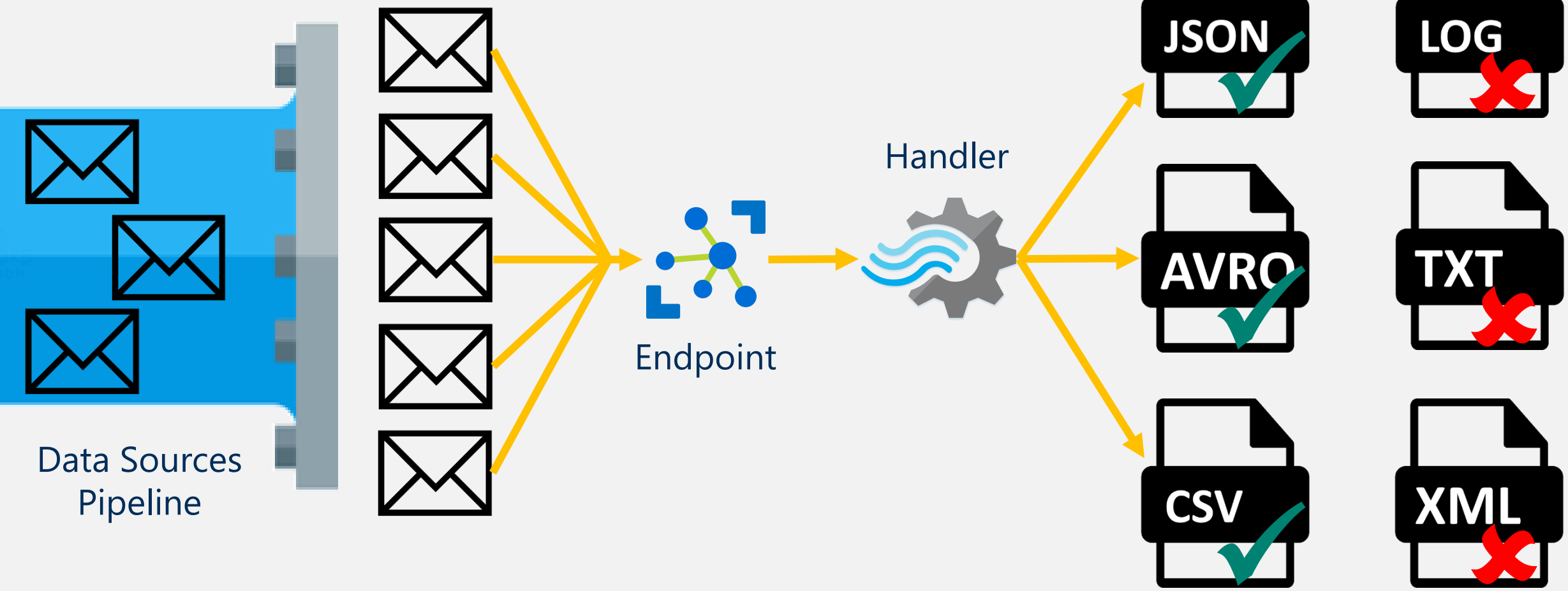
Real-time data problems

What is ASA and why use it

Production Considerations

Lambda Architecture

Problem: Real-time Data Feeds



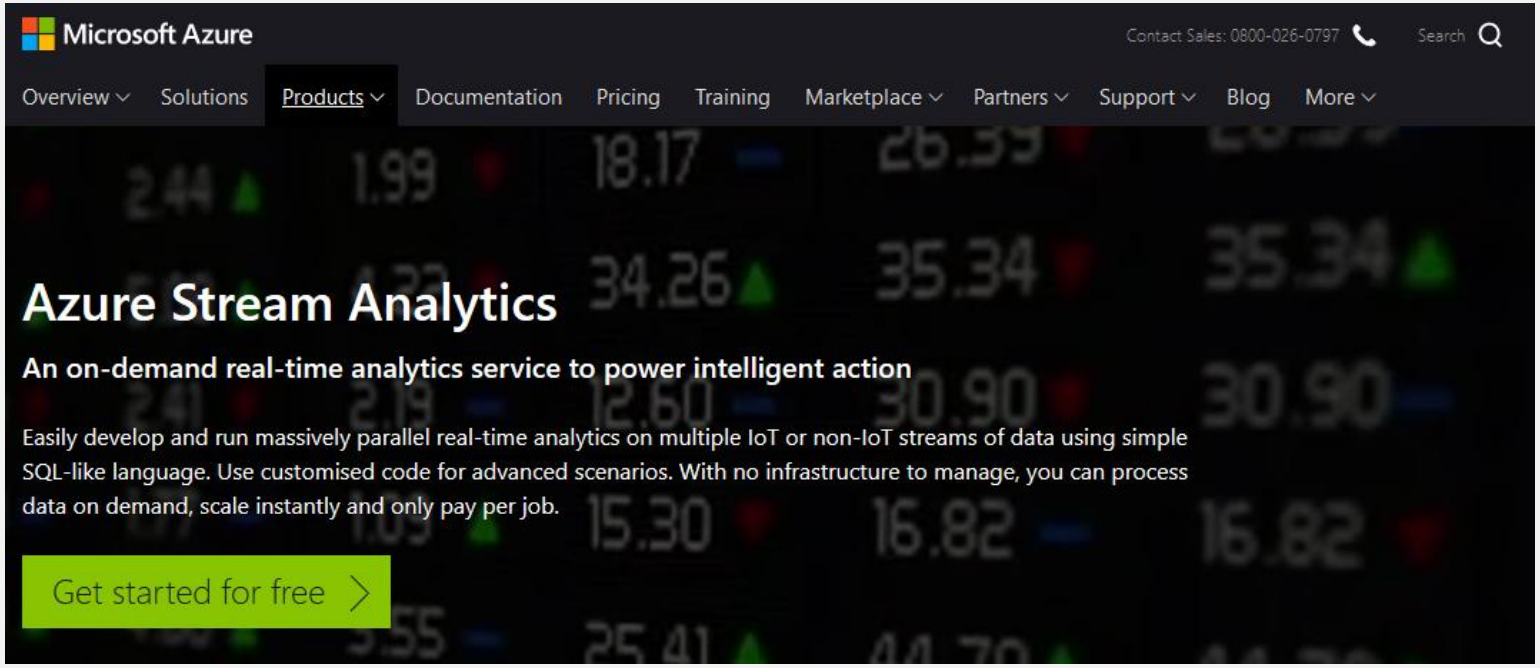
Azure Stream Analytics 

Real-time data problems

What is ASA and why use it

Production Considerations

Lambda Architecture



Azure Stream Analytics

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

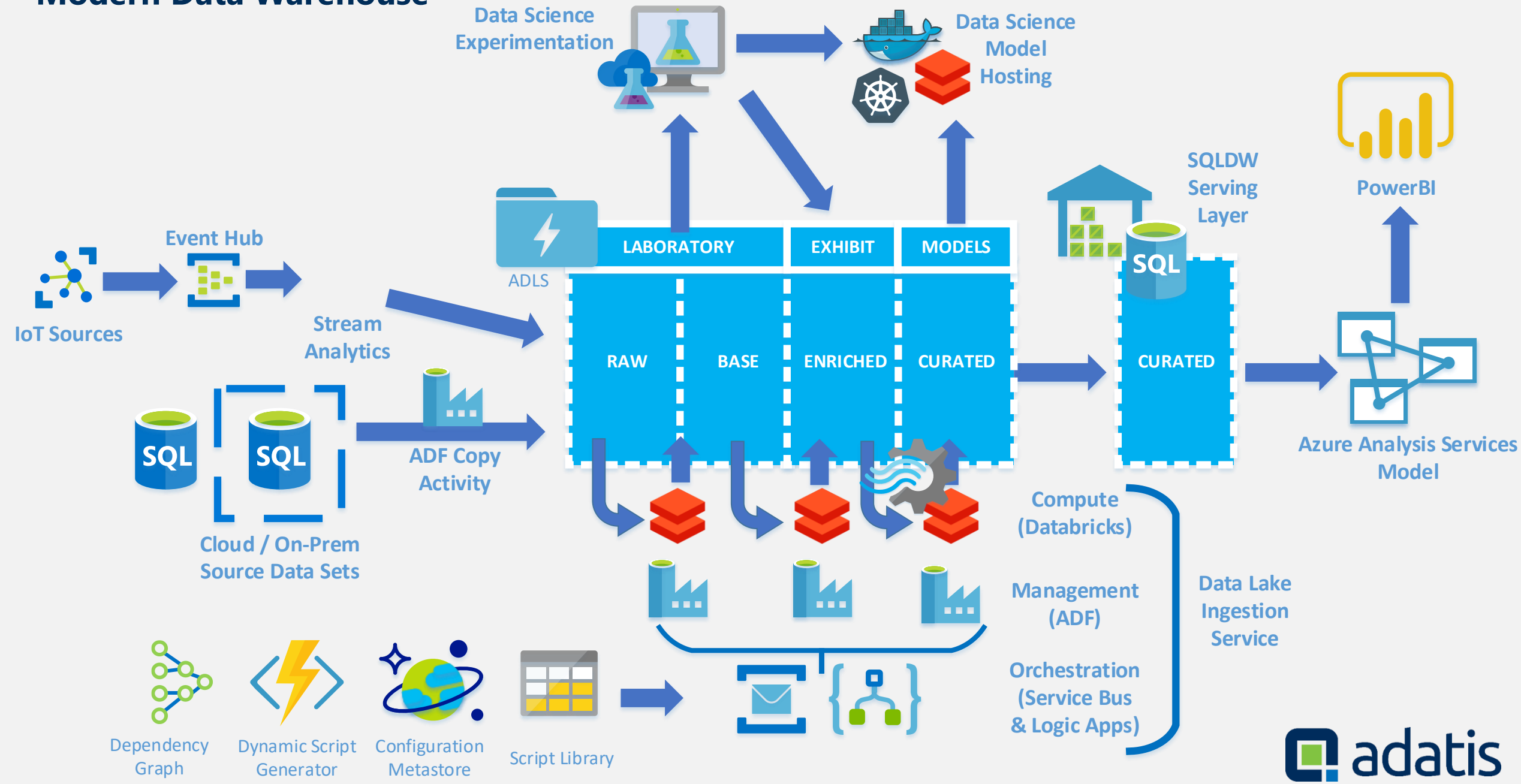
Real-time data problems

What is ASA and why use it

Production Considerations

Lambda Architecture

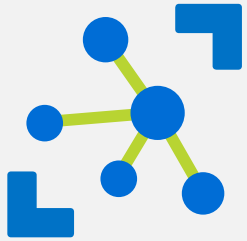
Modern Data Warehouse



adatis

Modern Data Warehouse

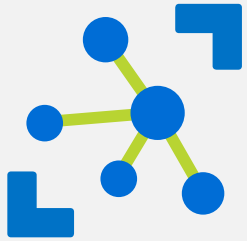
Azure IoT Hub vs Azure Event Hub



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

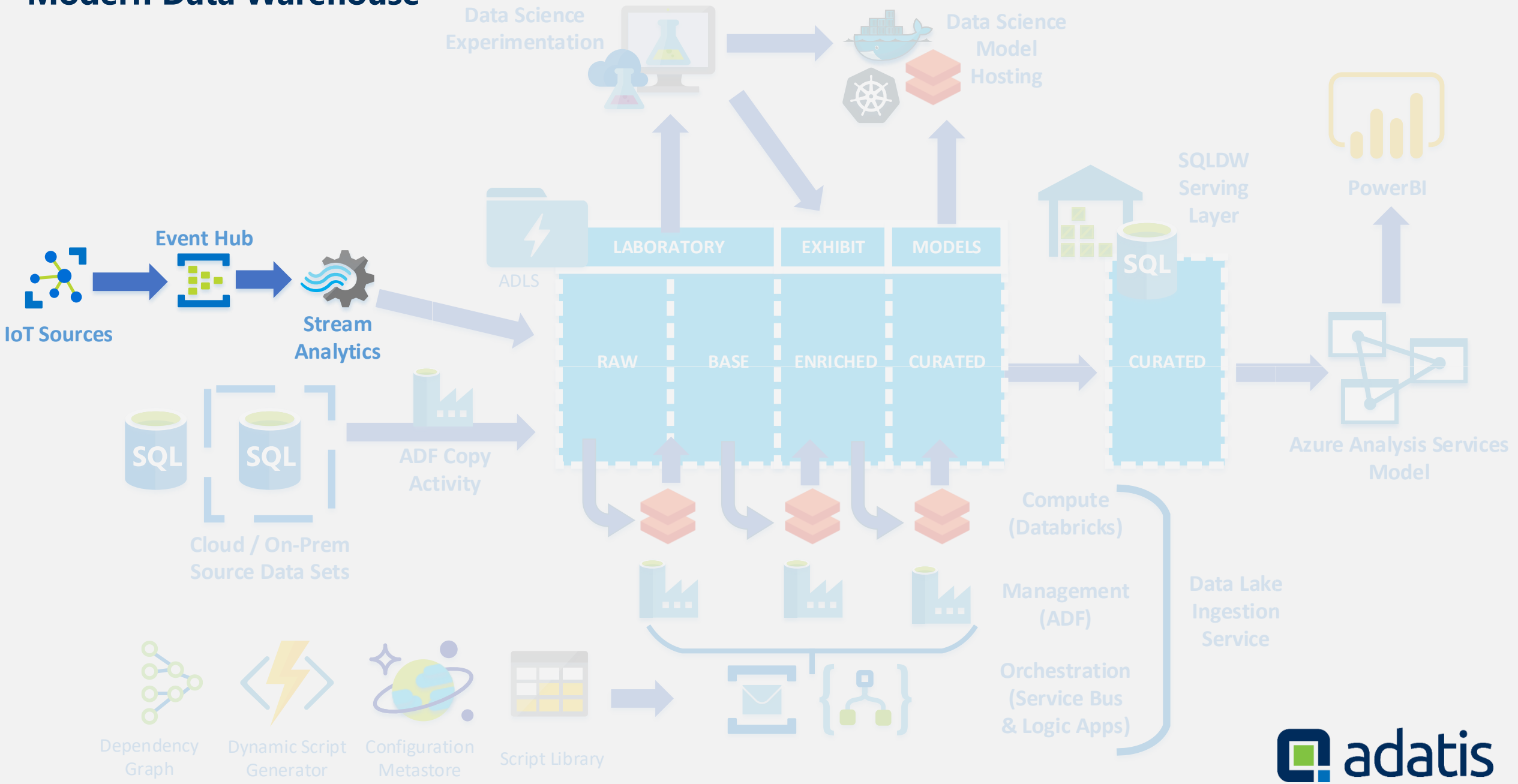
Azure IoT Hub vs Azure Event Hub



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)

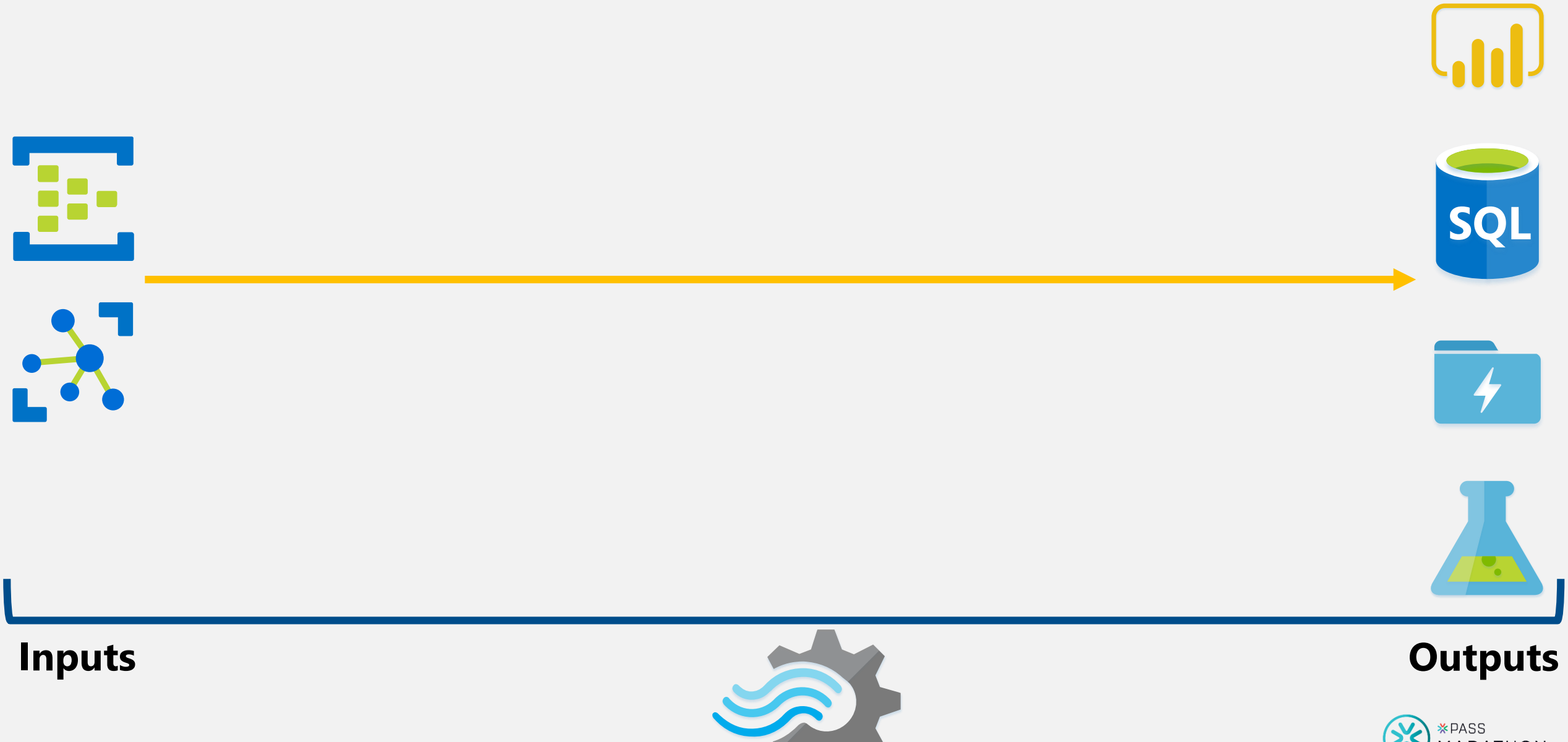
https://docs.microsoft.com/en-us/azure/iot-hub/iot-hub-compare-event-hubs

Modern Data Warehouse

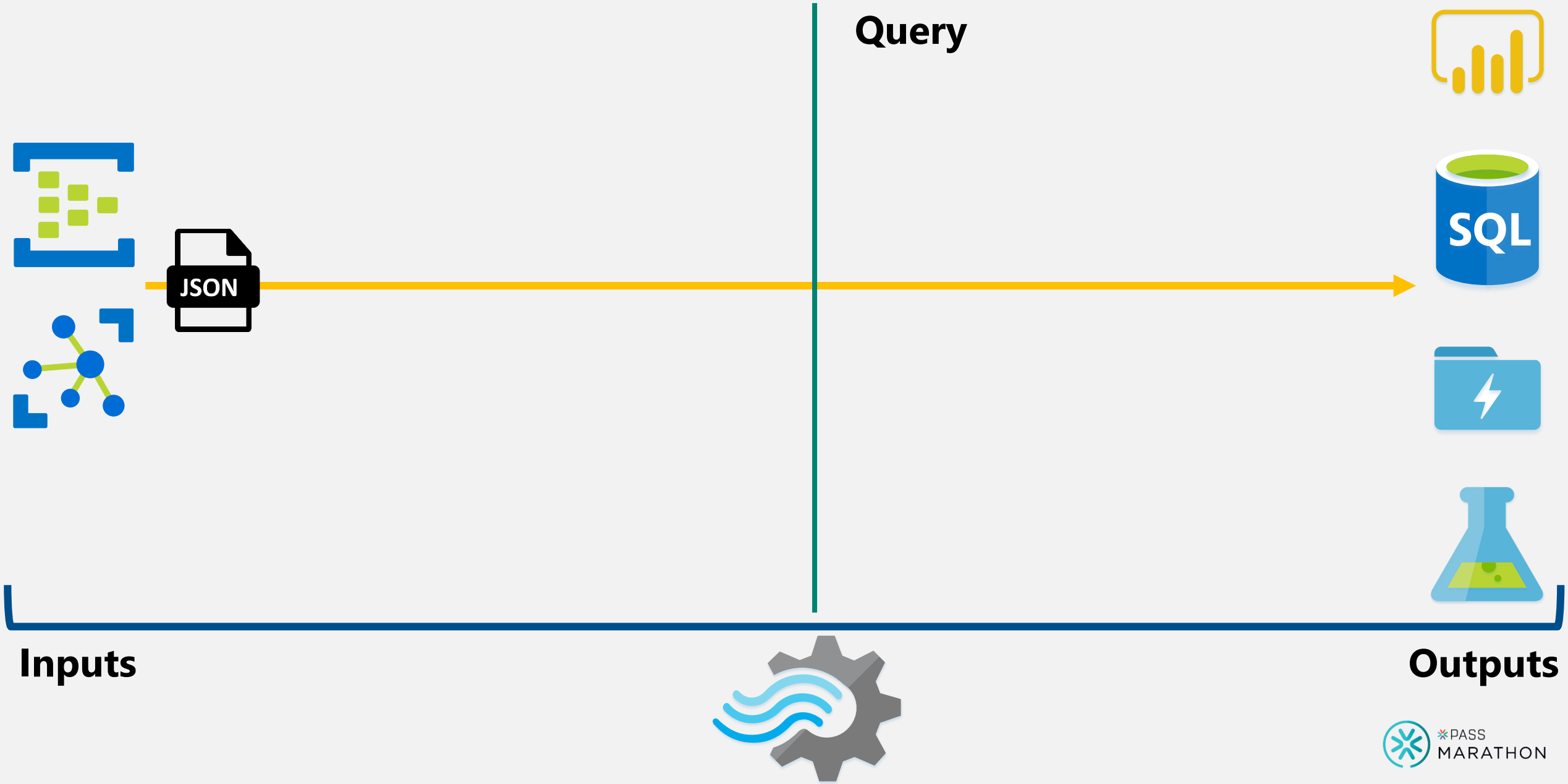


adatis

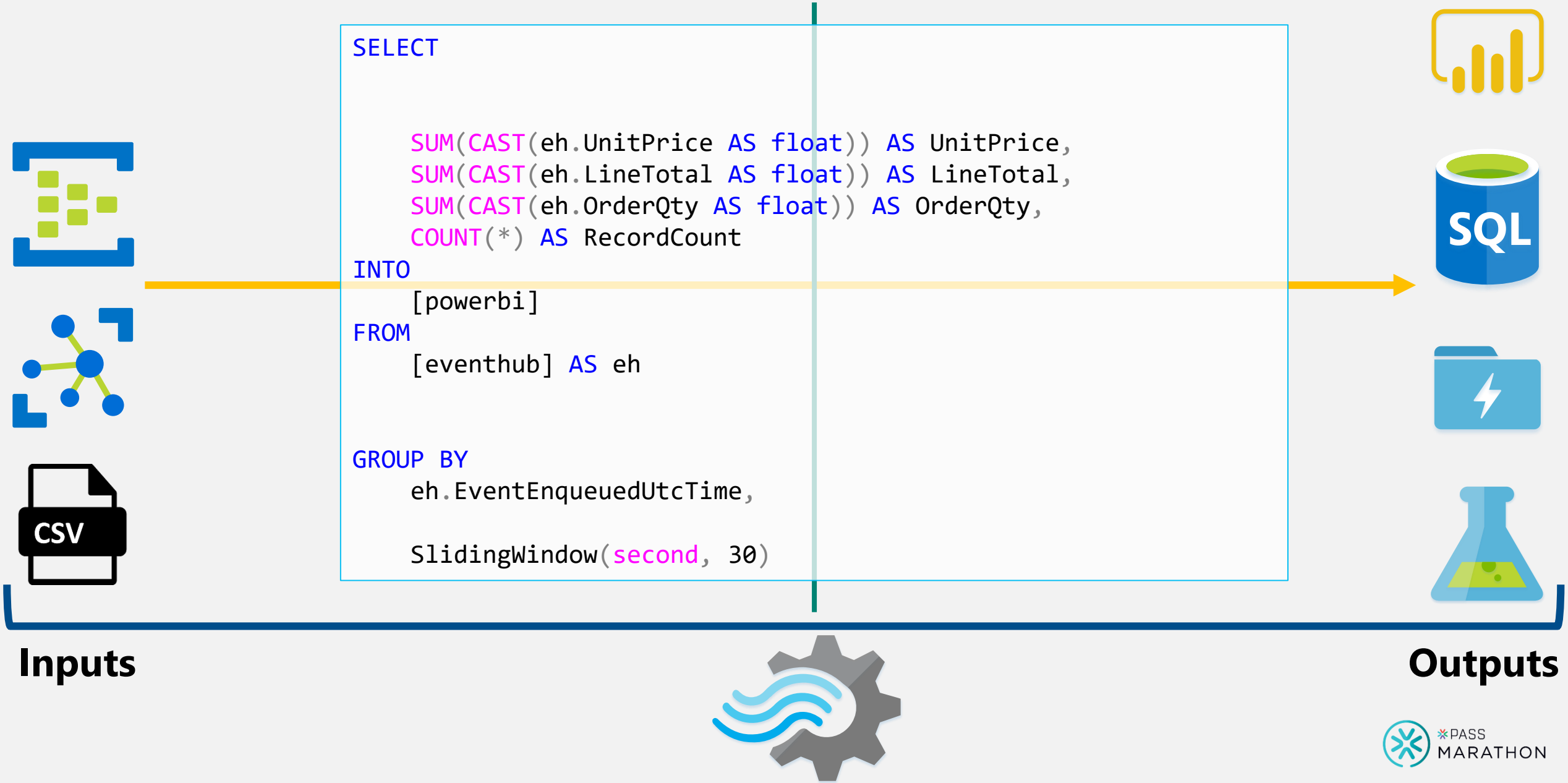
Azure Stream Analytics



Azure Stream Analytics

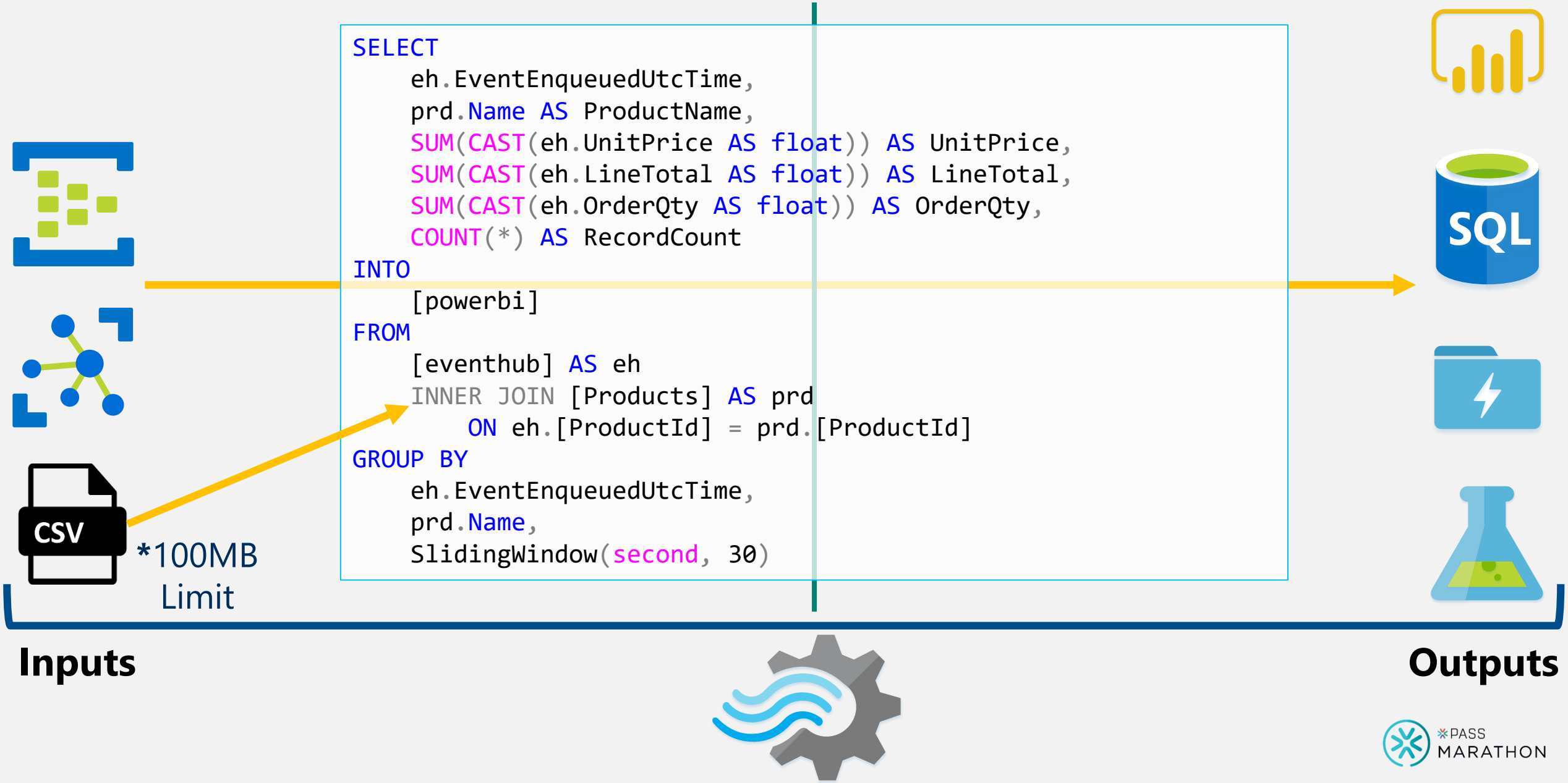


Azure Stream Analytics



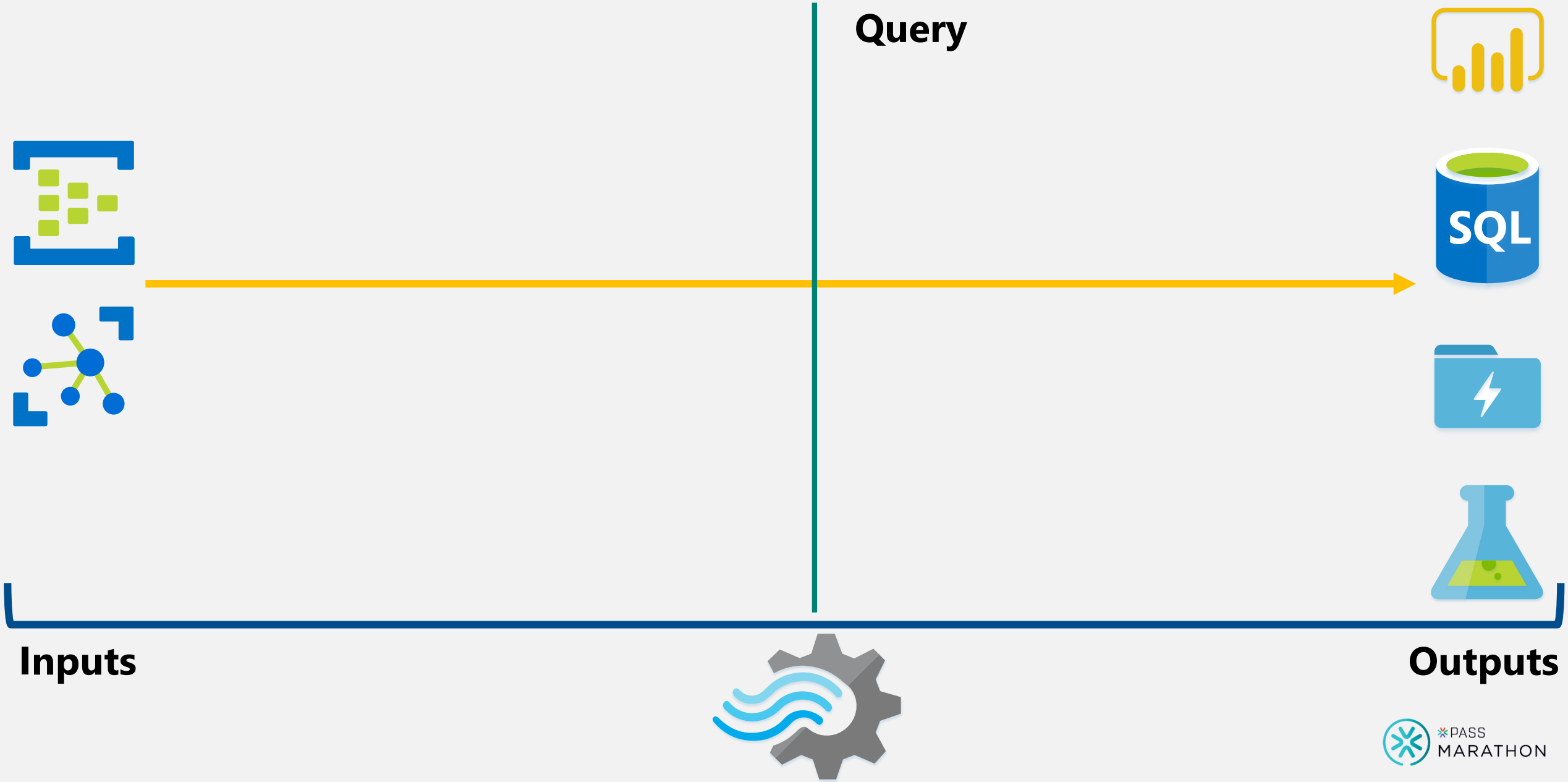
PASS MARATHON

Azure Stream Analytics

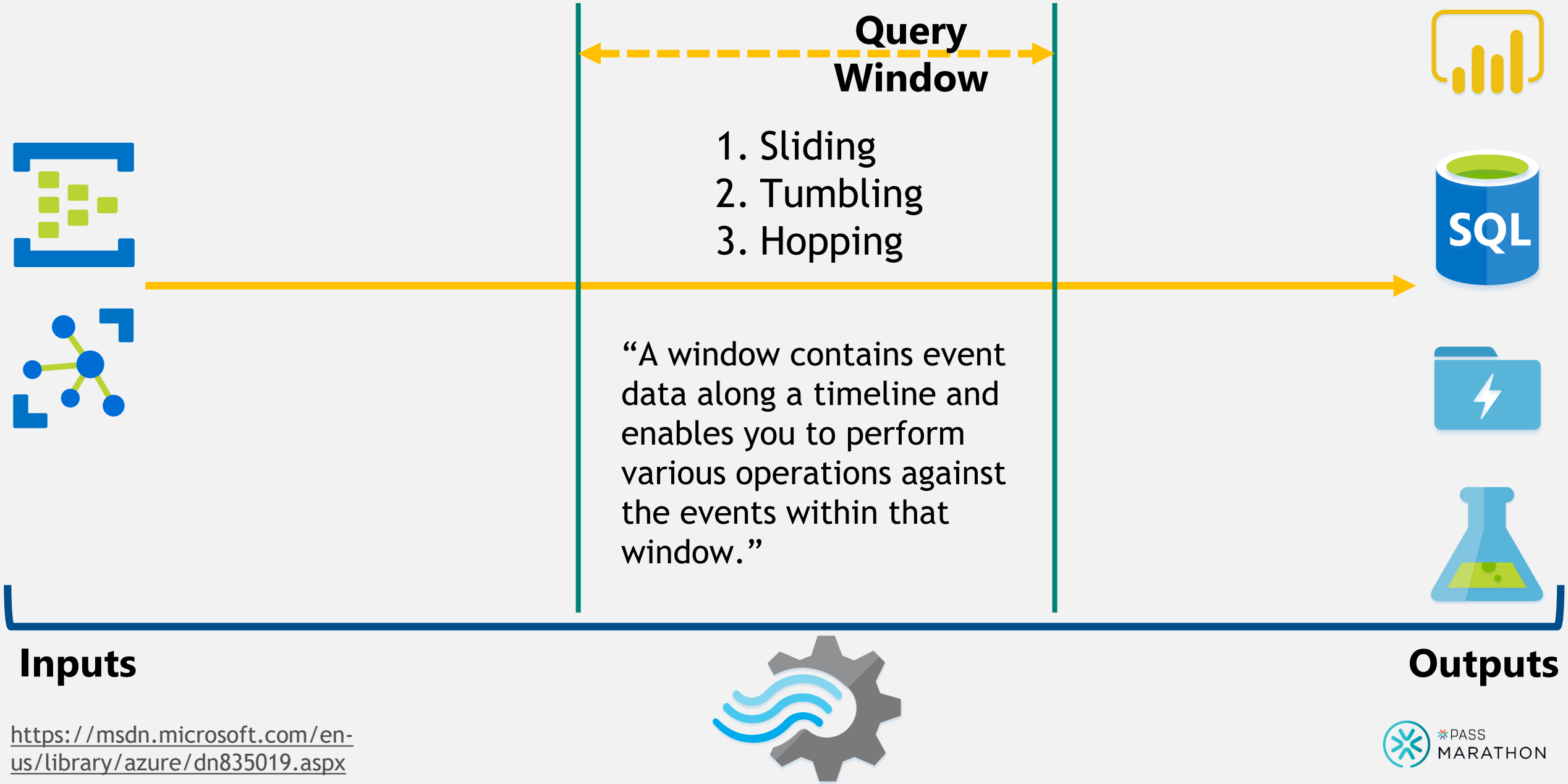


PASS MARATHON

Azure Stream Analytics



Azure Stream Analytics



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

Azure Stream Analytics

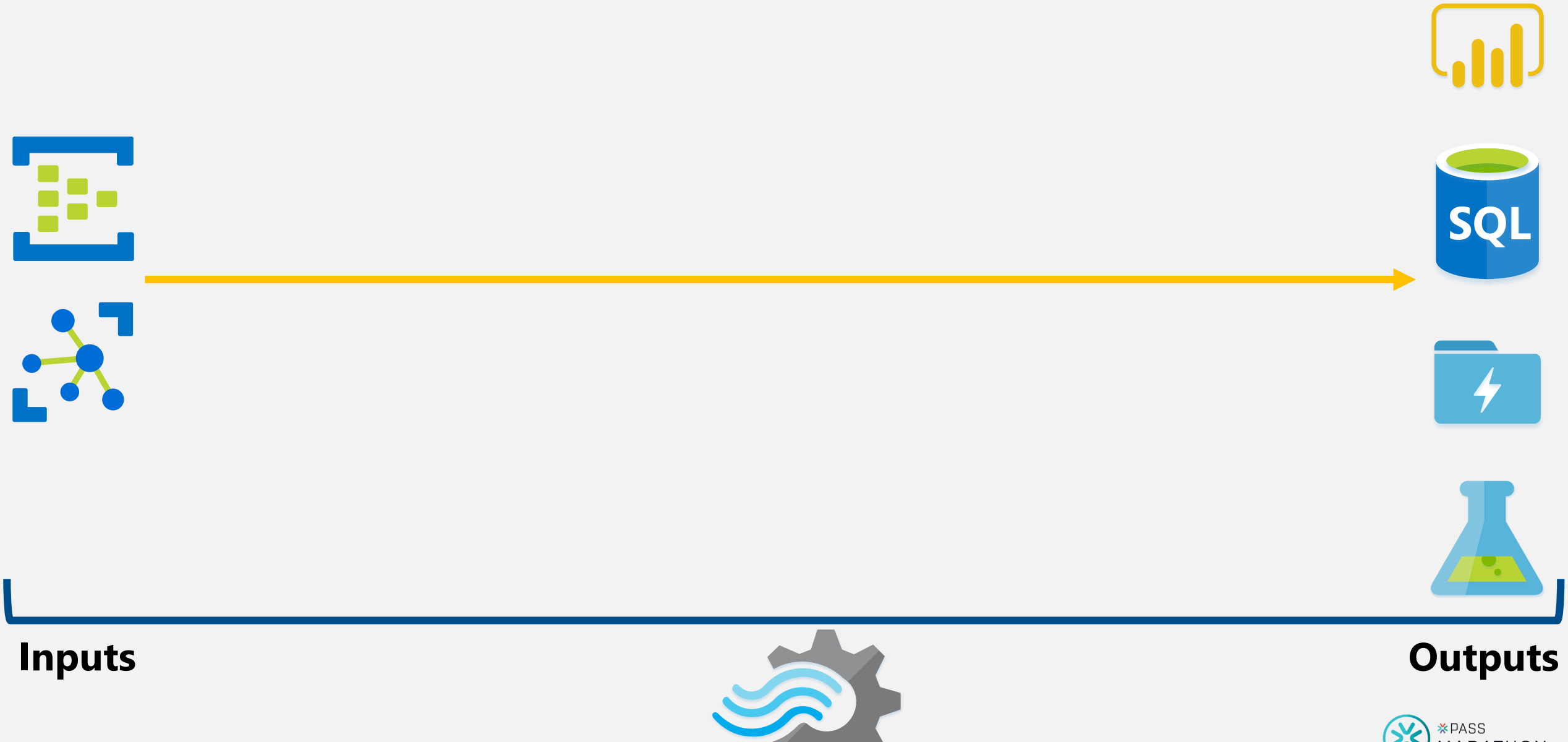
Sliding Window

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

Inputs

Outputs

Azure Stream Analytics



Azure Stream Analytics

Tumbling Window

Tumbling Window

Tumbling Window

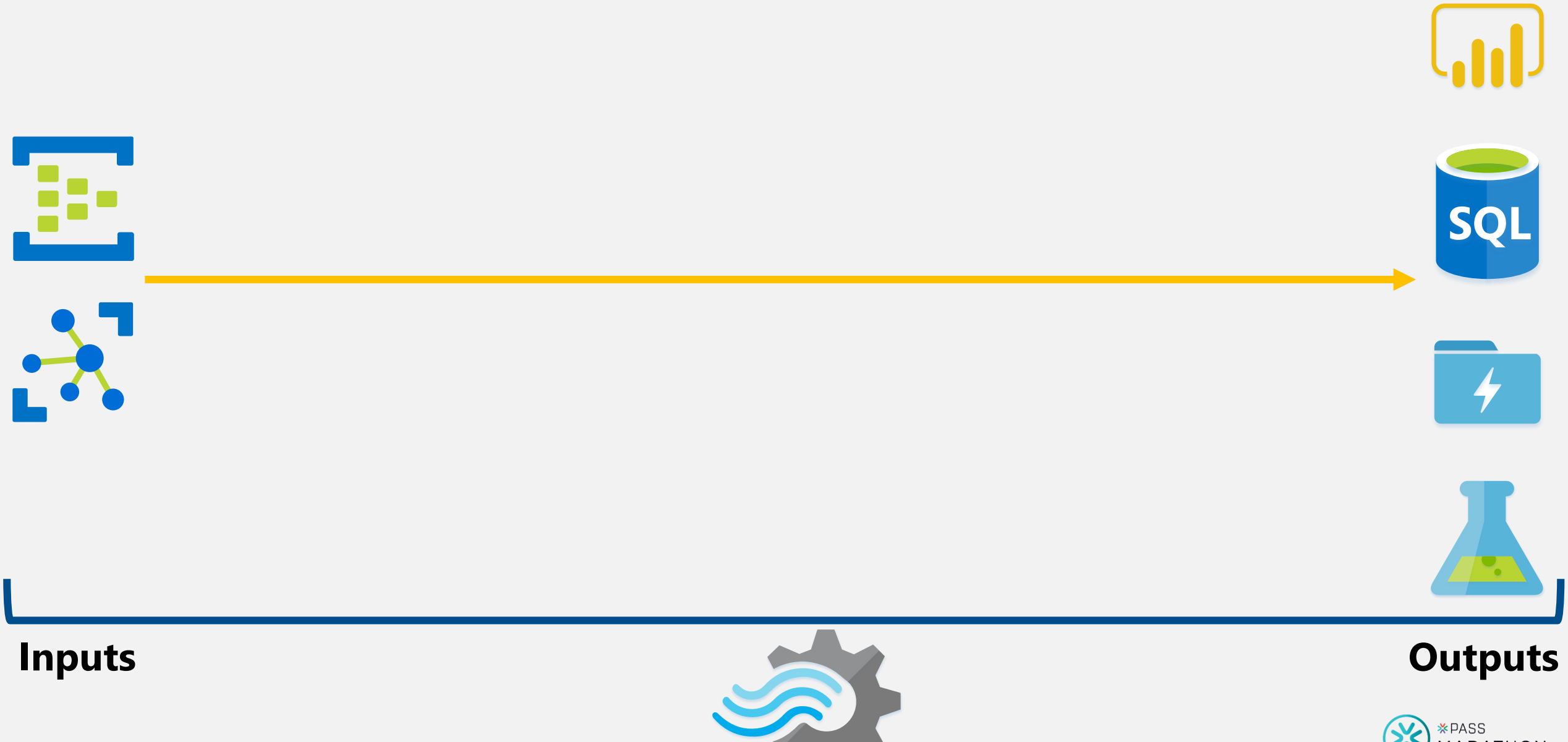
Tumbling Window

```
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

Azure Stream Analytics



Azure Stream Analytics

Hopping Window

Hopping Window

Hopping Window

Hopping Window

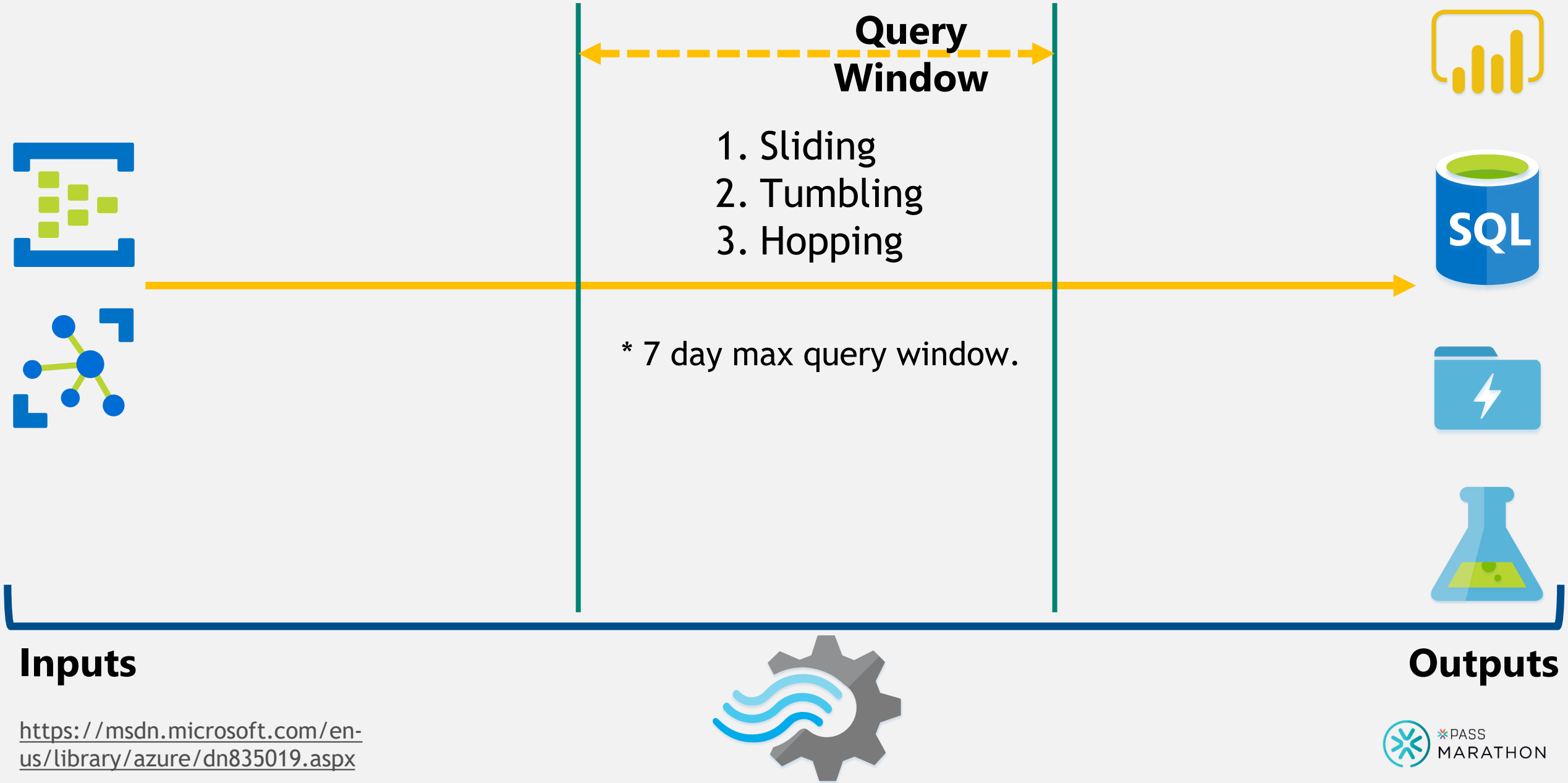
```
SELECT
    MIN([value]) AS 'Min',
    MAX([value]) AS 'Max'
INTO
    SQLDBAvg
FROM
    IoTHub
TIMESTAMP BY
    timecreated
GROUP BY
    HoppingWindow(Duration(MINUTE, 15), Hop(MINUTE, 5), Offset(millisecond, -1))
```

Inputs

Outputs

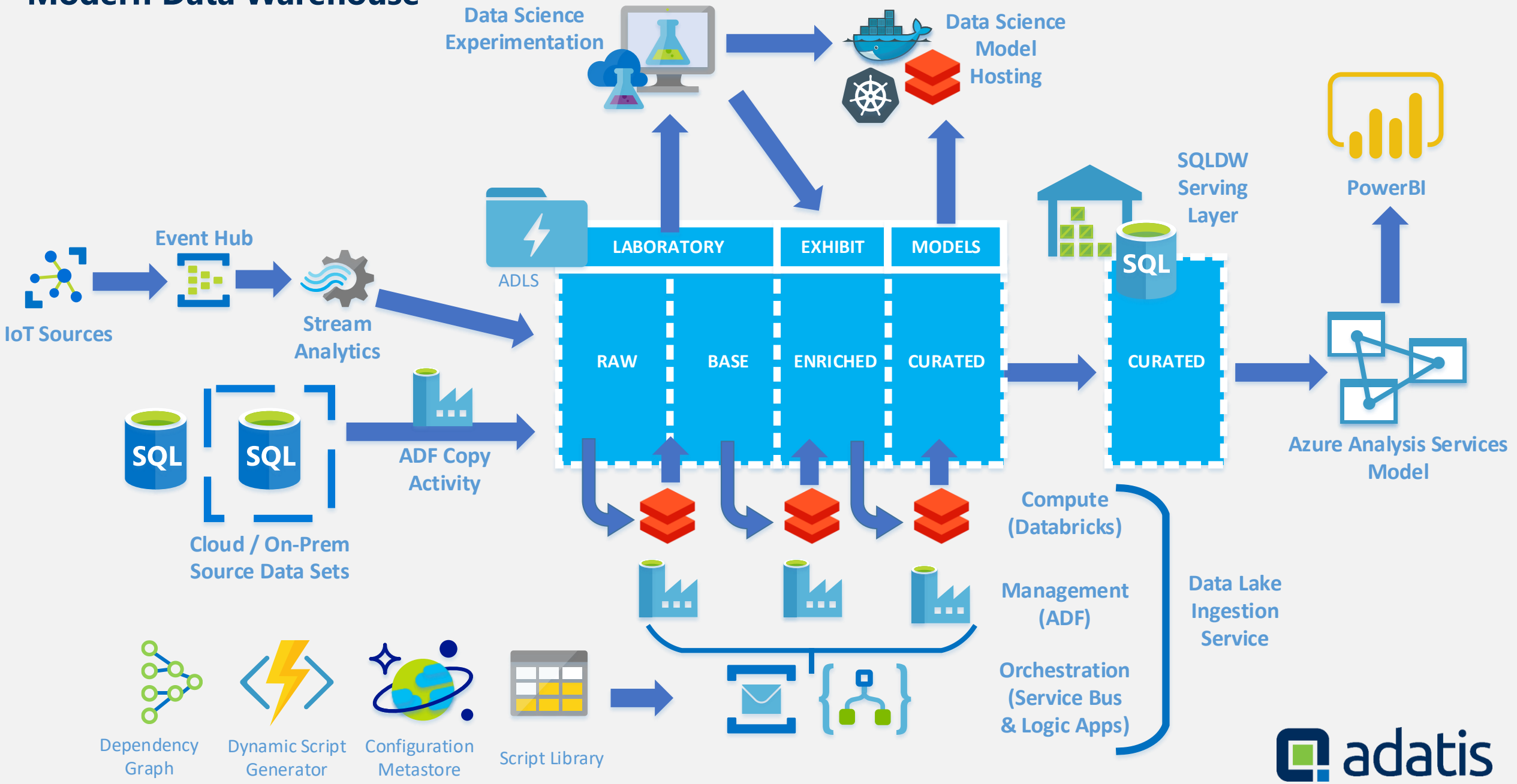
PASS MARATHON

Azure Stream Analytics



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

Modern Data Warehouse



adatis

Demo

Azure Stream Analytics 

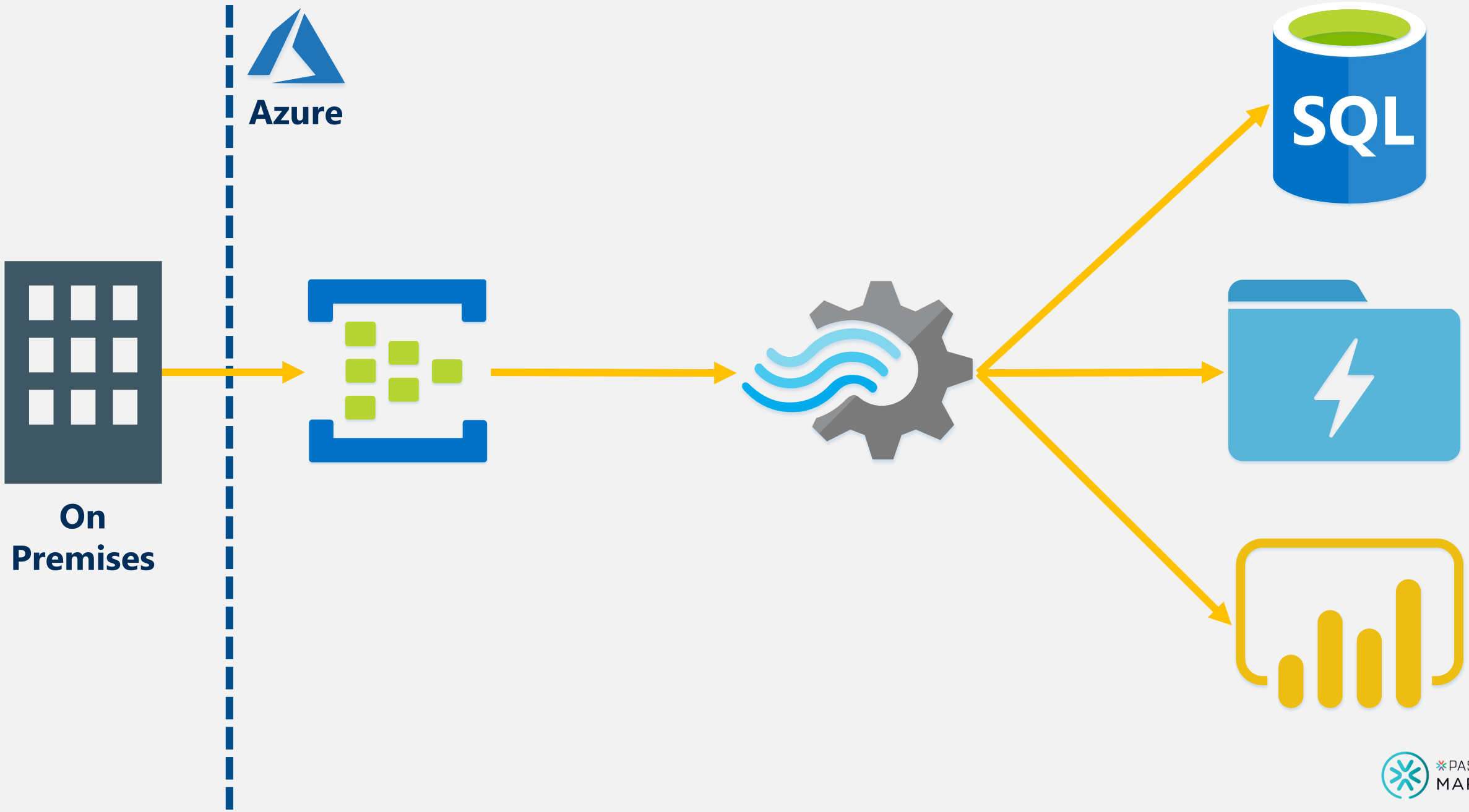
Real-time data problems

What is ASA and why use it

Production Considerations

Lambda Architecture

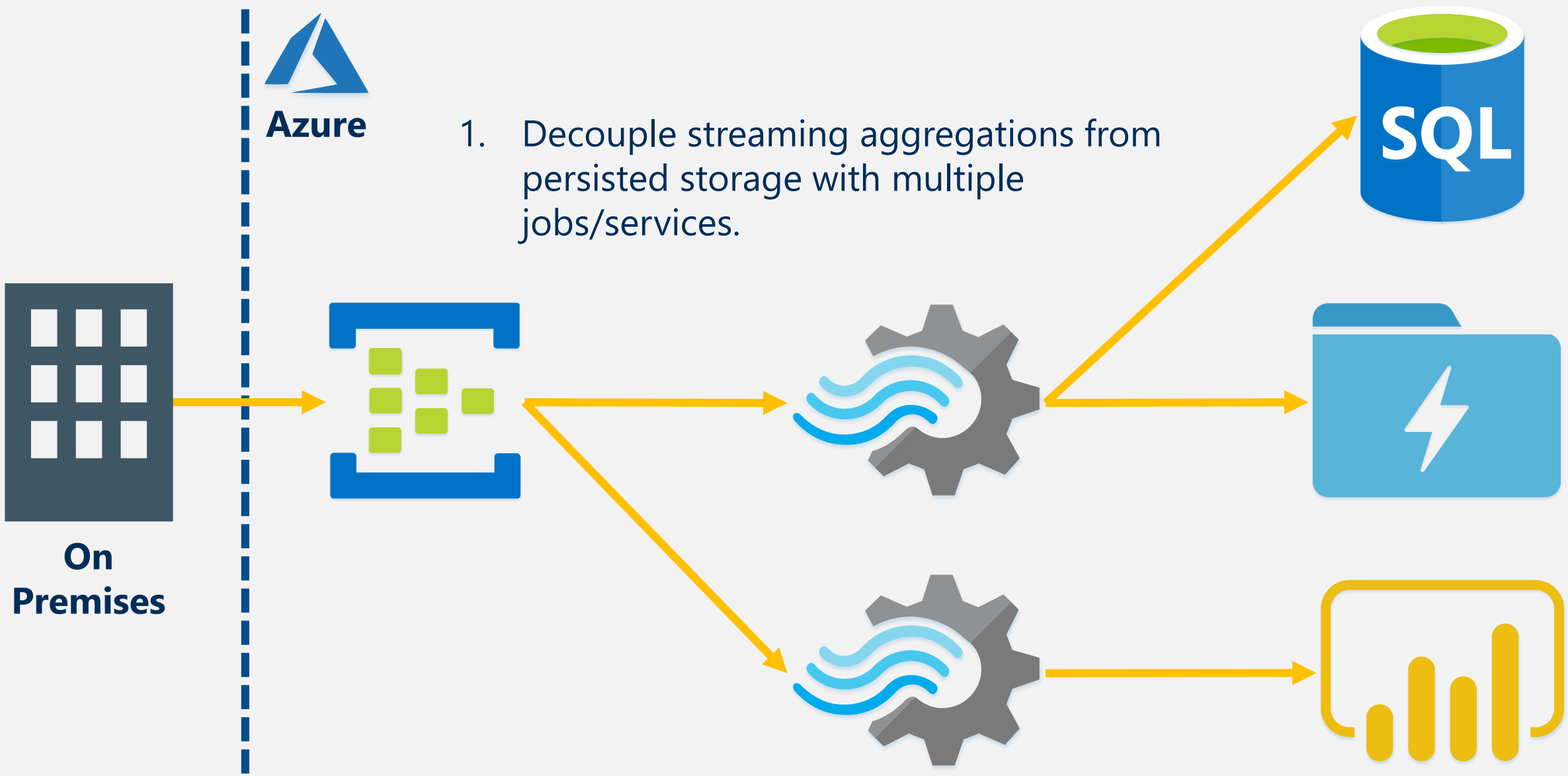
Production Considerations



Production Considerations

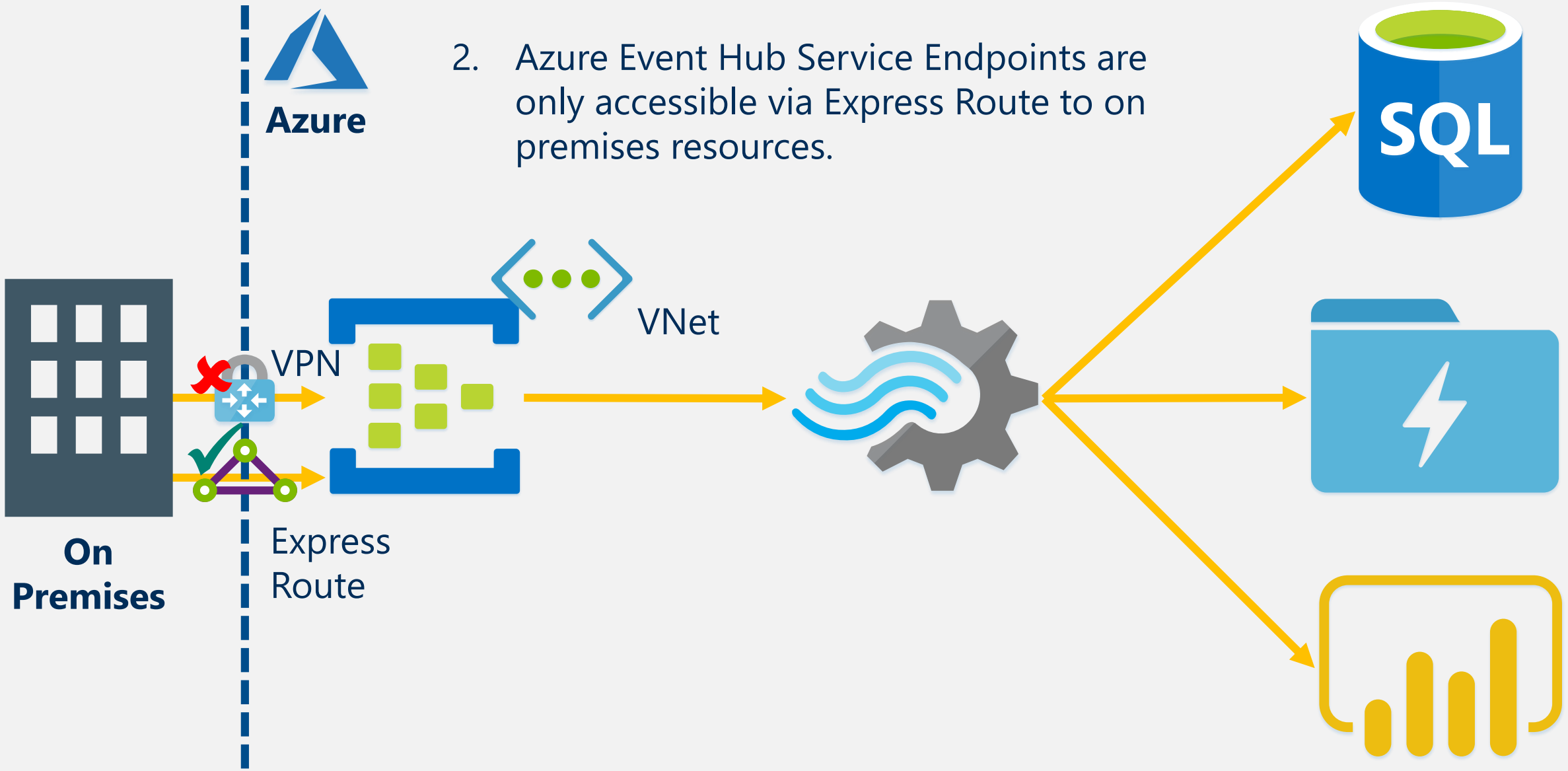
Azure

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

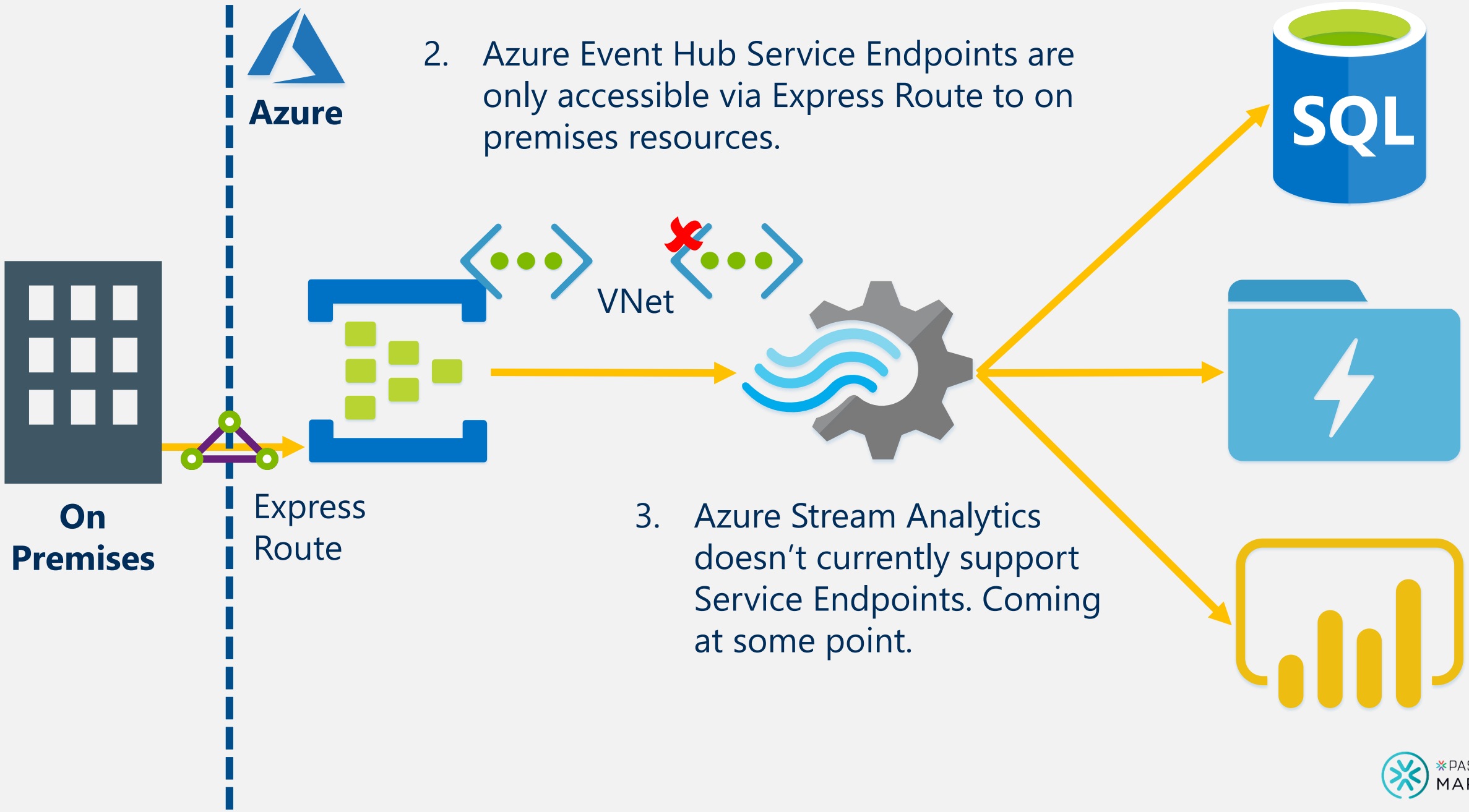


Production Considerations

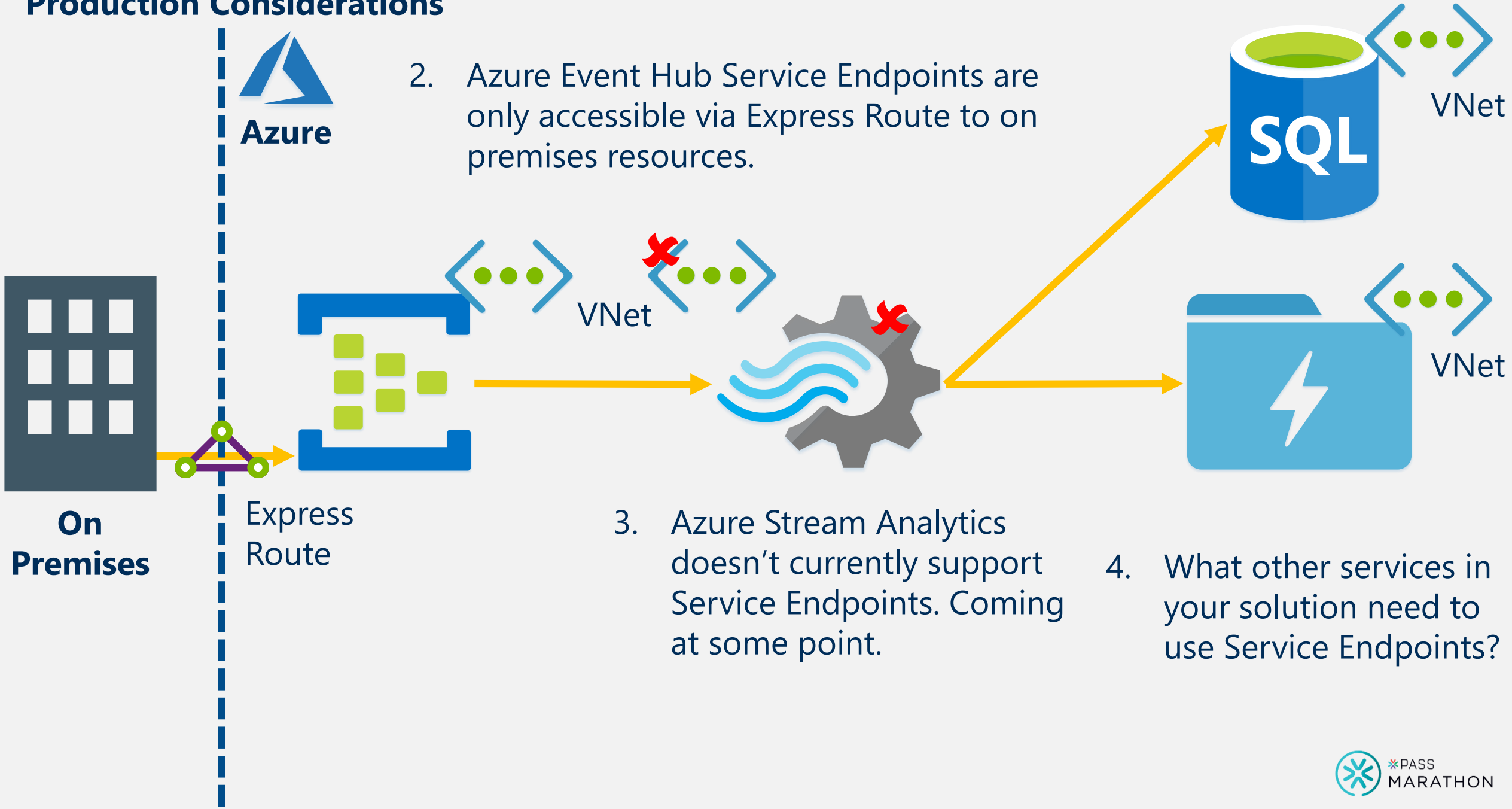
2. Azure Event Hub Service Endpoints are only accessible via Express Route to on premises resources.



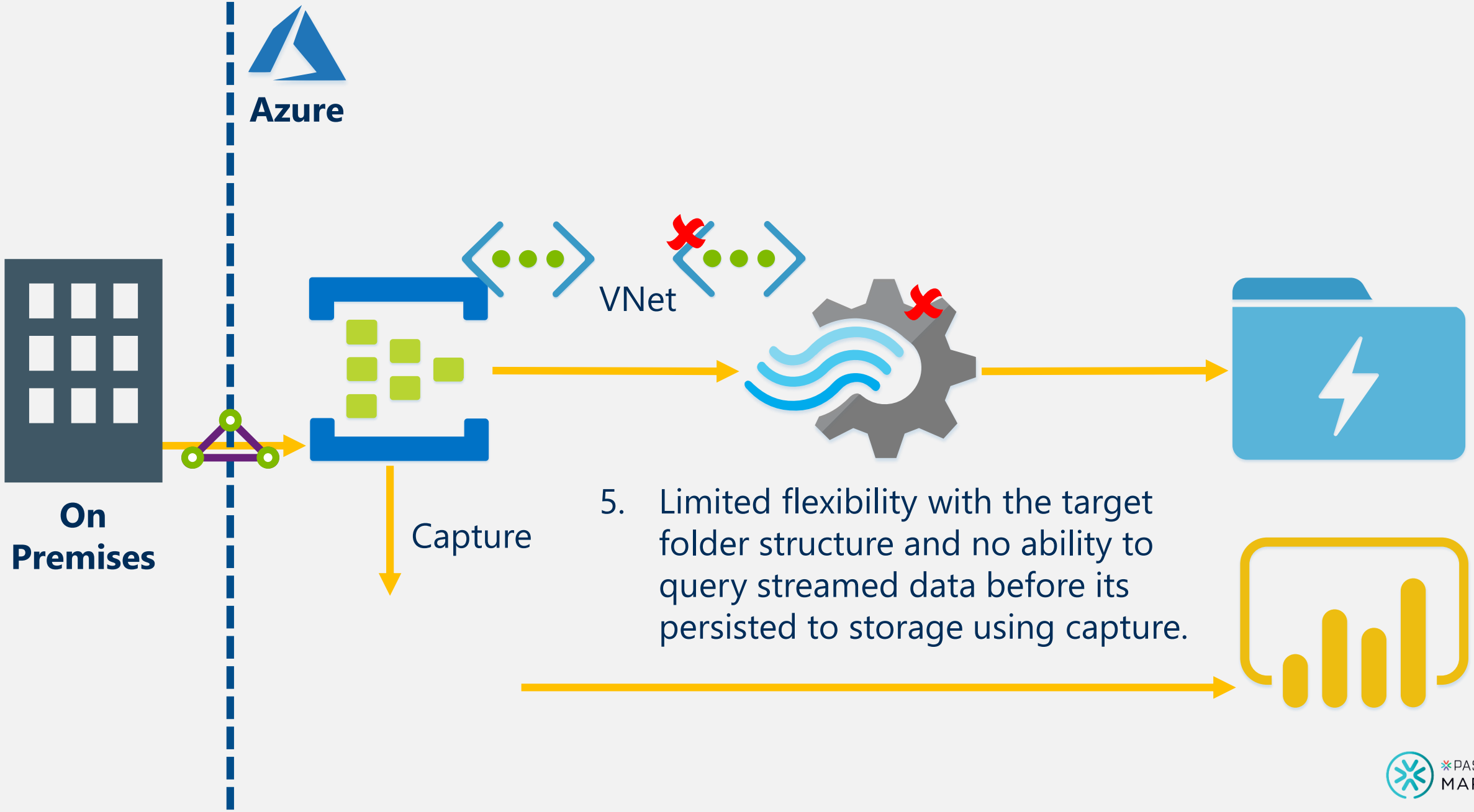
Production Considerations



Production Considerations

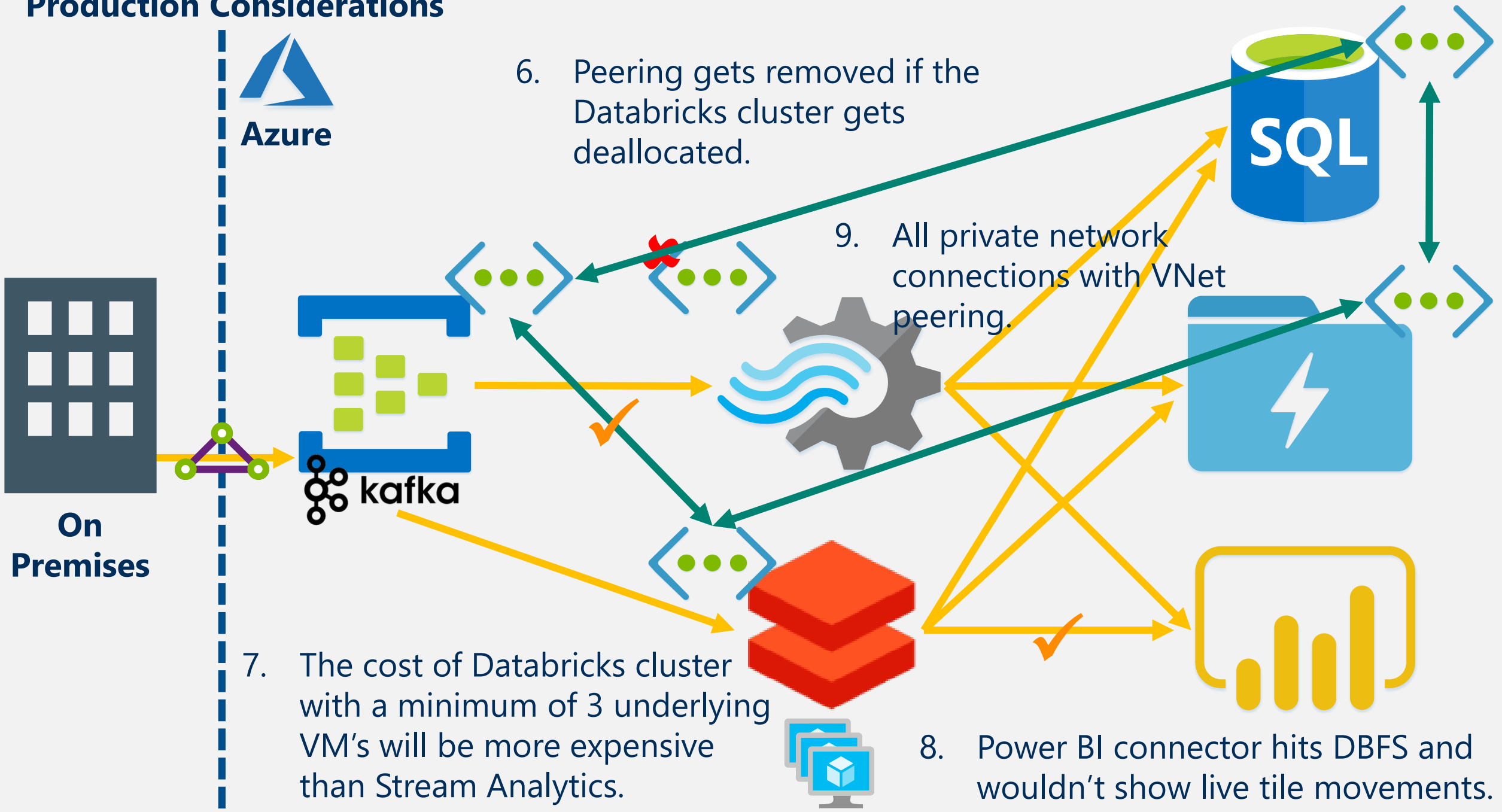


Production Considerations



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

Production Considerations



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.

Azure Stream Analytics 

Real-time data problems

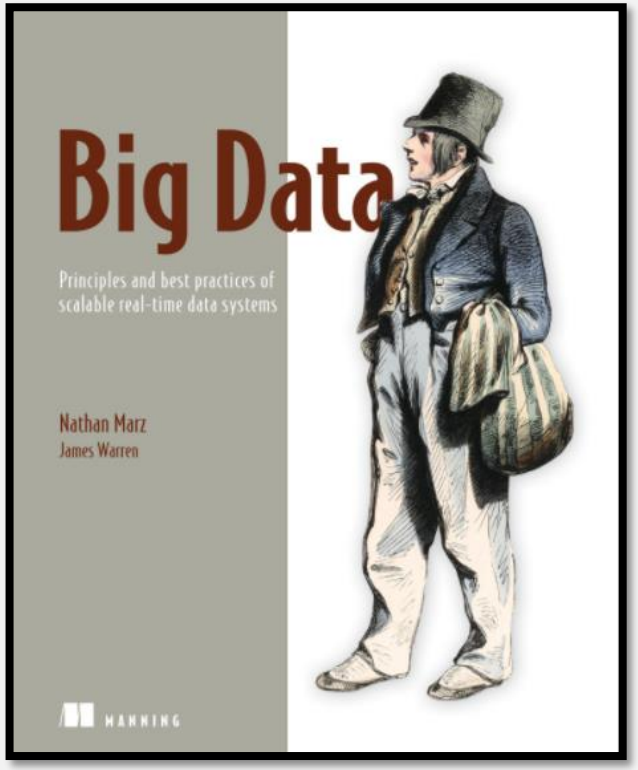
What is ASA and why use it

Production Considerations

Lambda Architecture

Lambda Architecture

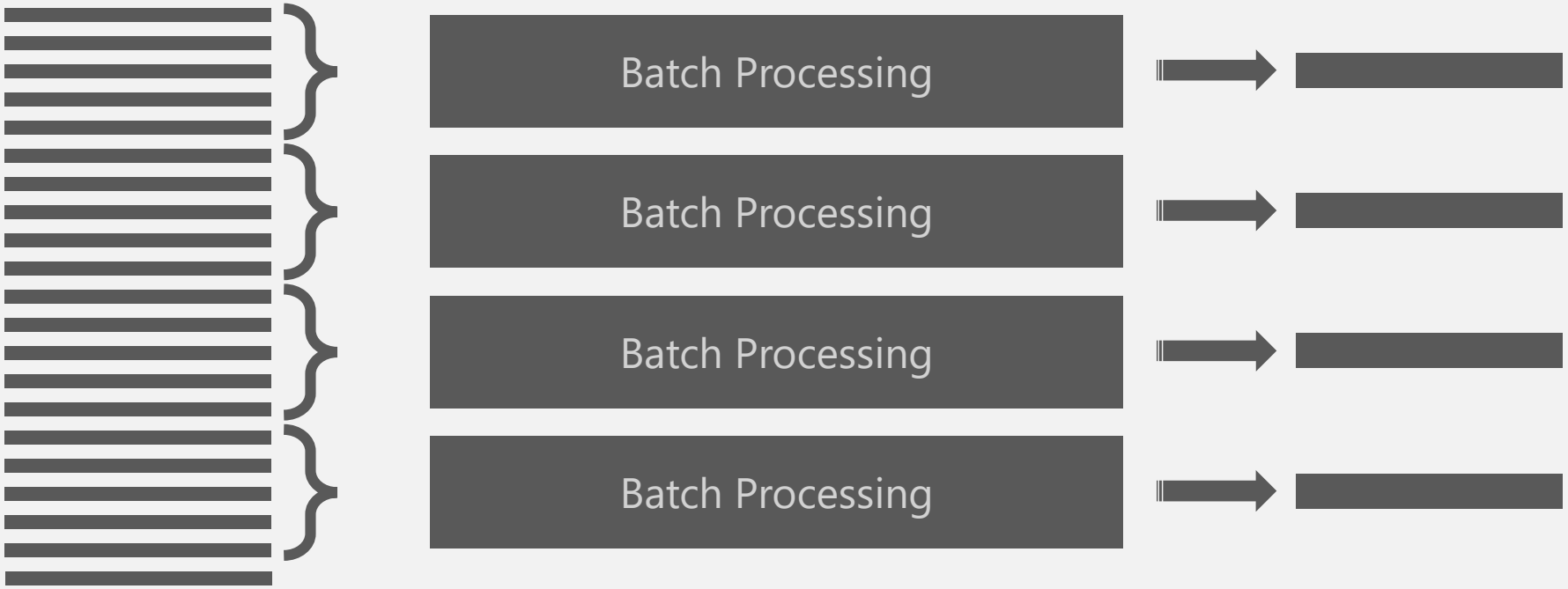
Use Batch and Stream technologies together to balance latency, throughput and fault-tolerance



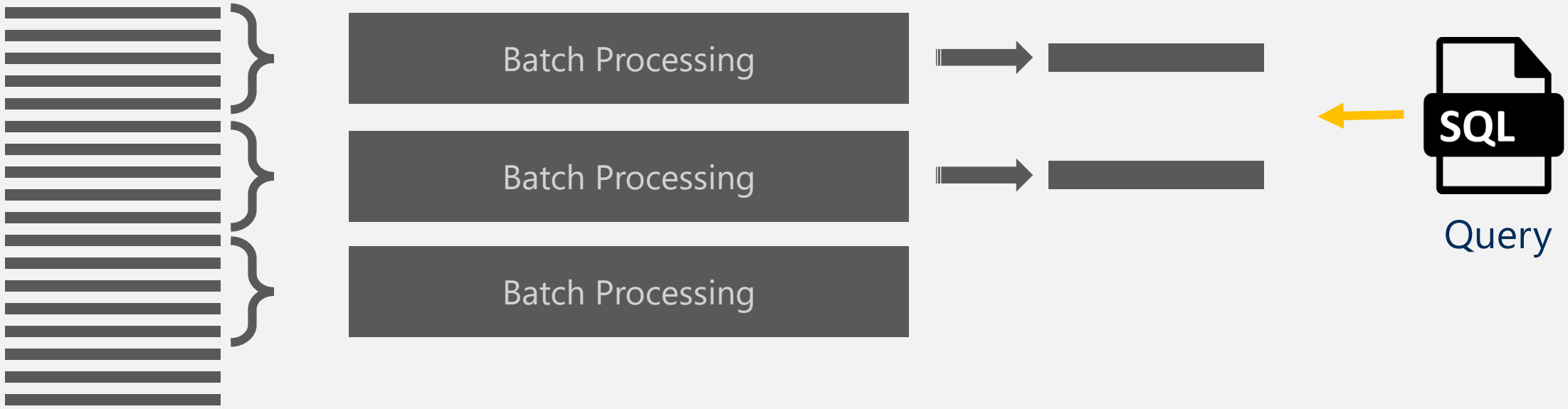
Nathan Marz
& James Warren

* Pages 14 to 20

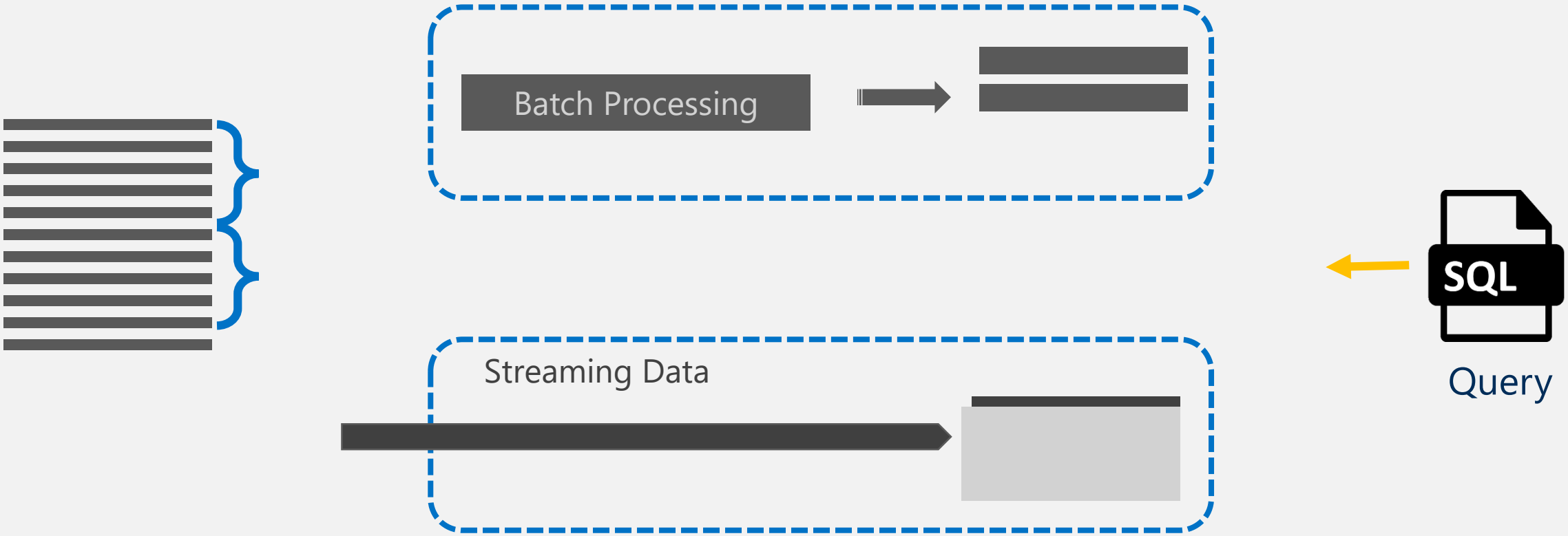
Problem: Timely Data Insights



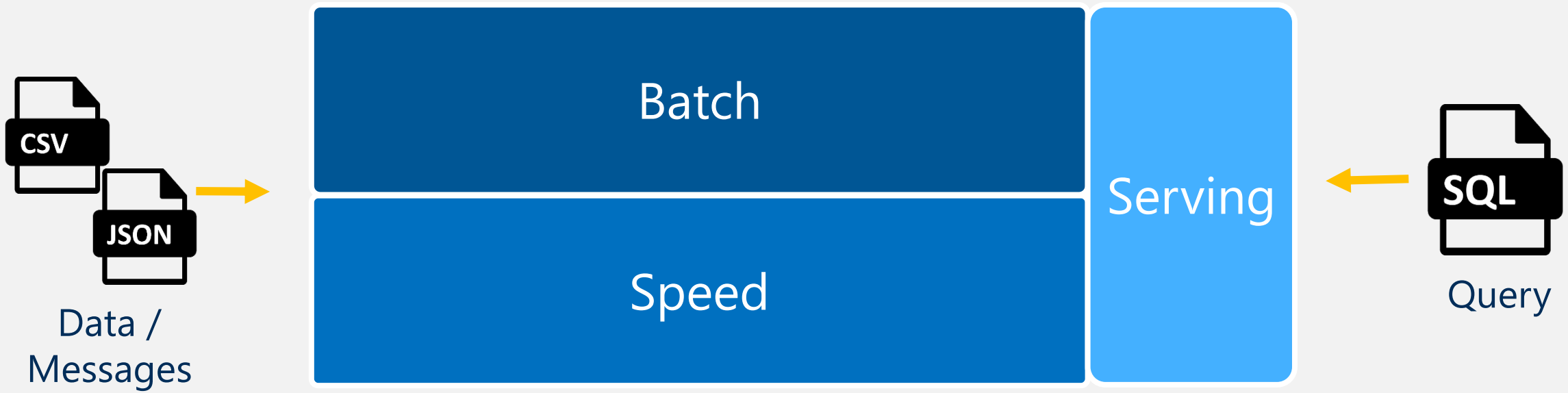
Problem: Timely Data Insights



Solution

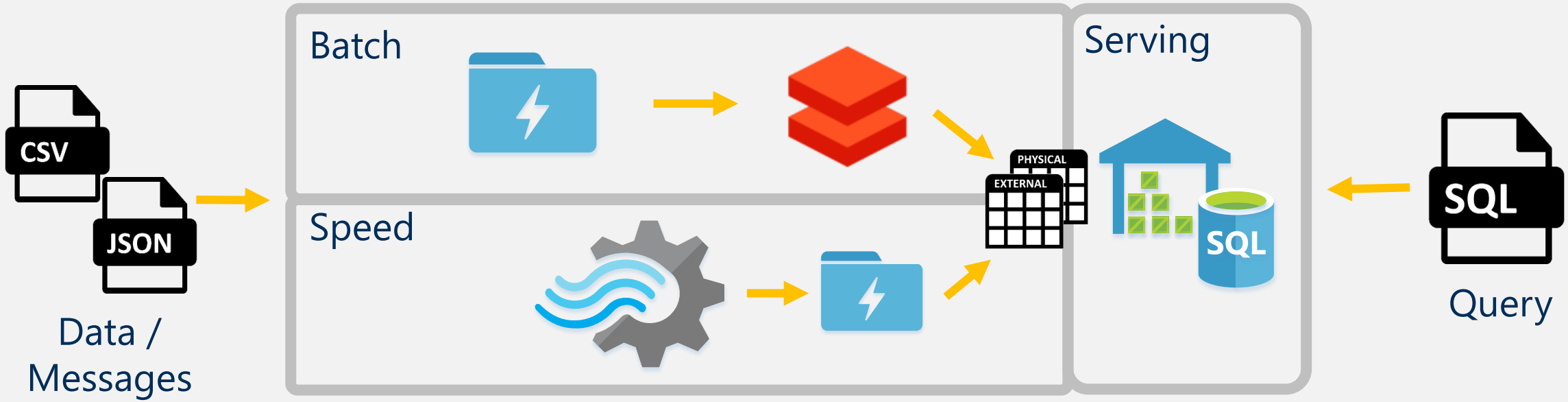


Lambda Architecture



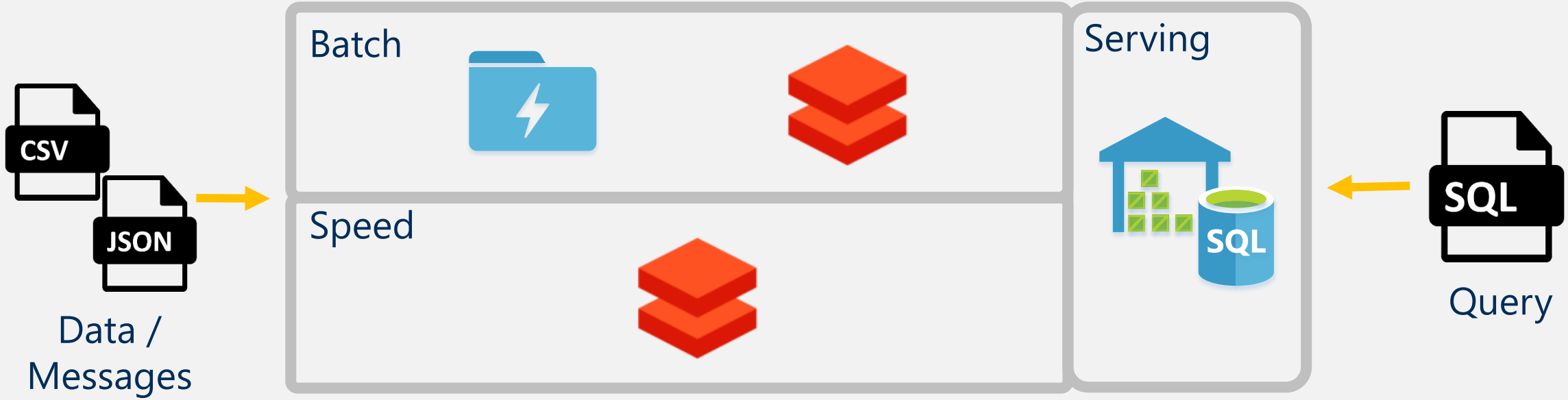
The Marz Lambda Architecture

Applying a Lambda Architecture in Azure



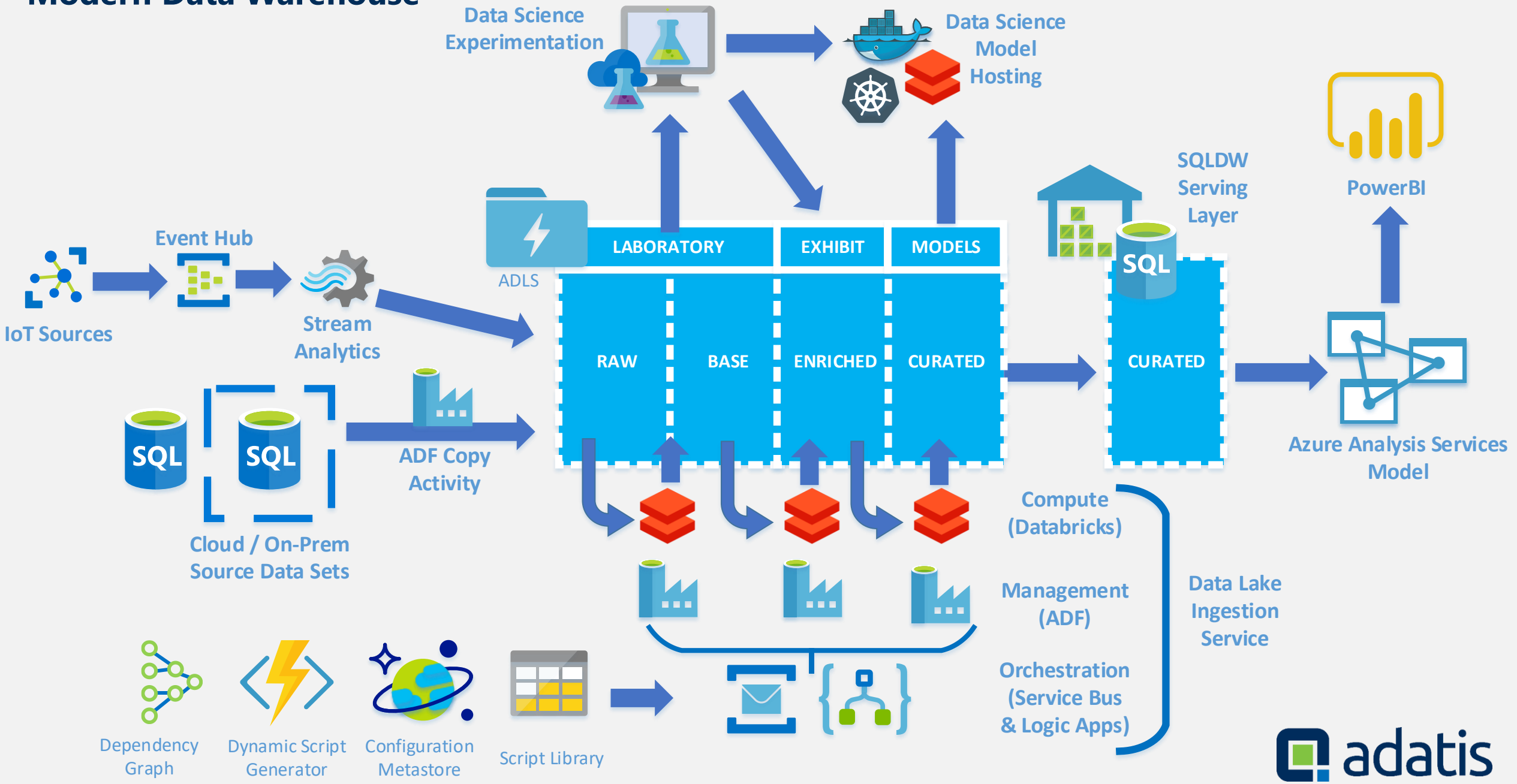
Simon Whiteley – PASS Summit 2018:
https://www.pass.org/summit/2018/Sessions/Details.aspx?name=modern-data-warehousing-the-lambda-approach-to-azure-bi

Applying a Lambda Architecture in Azure



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

Modern Data Warehouse



adatis

Azure Stream Analytics

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

PASS MARATHON

Thank you for attending

Learn more from Paul Andrew:

 mrpaulandrew.com

 github.com/mrpaulandrew

 /mrpaulandrew

 paul@mrpaulandrew.com

 @mrpaulandrew

 @sqlpass #sqlpass