LAMBDA

Simon Whiteley
@MrSiWhiteley

REALTIME &
BIG DATA IN AZURE

THE LAMBDA APPROACH TO AZURE BI

MODERN DATA WAREHOUSING

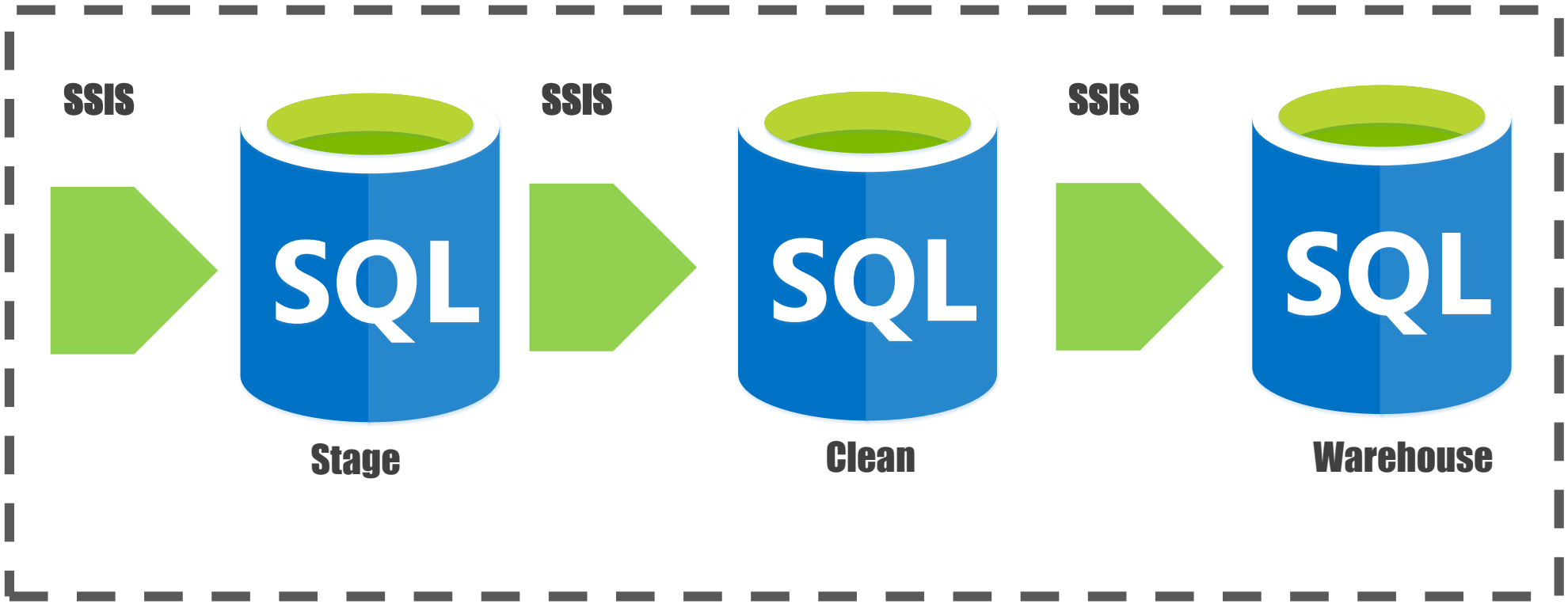
ADVANCING ANALYTICS
Data Science | AI | DataOps | Engineering

Simon Whiteley
@MrSiWhiteley

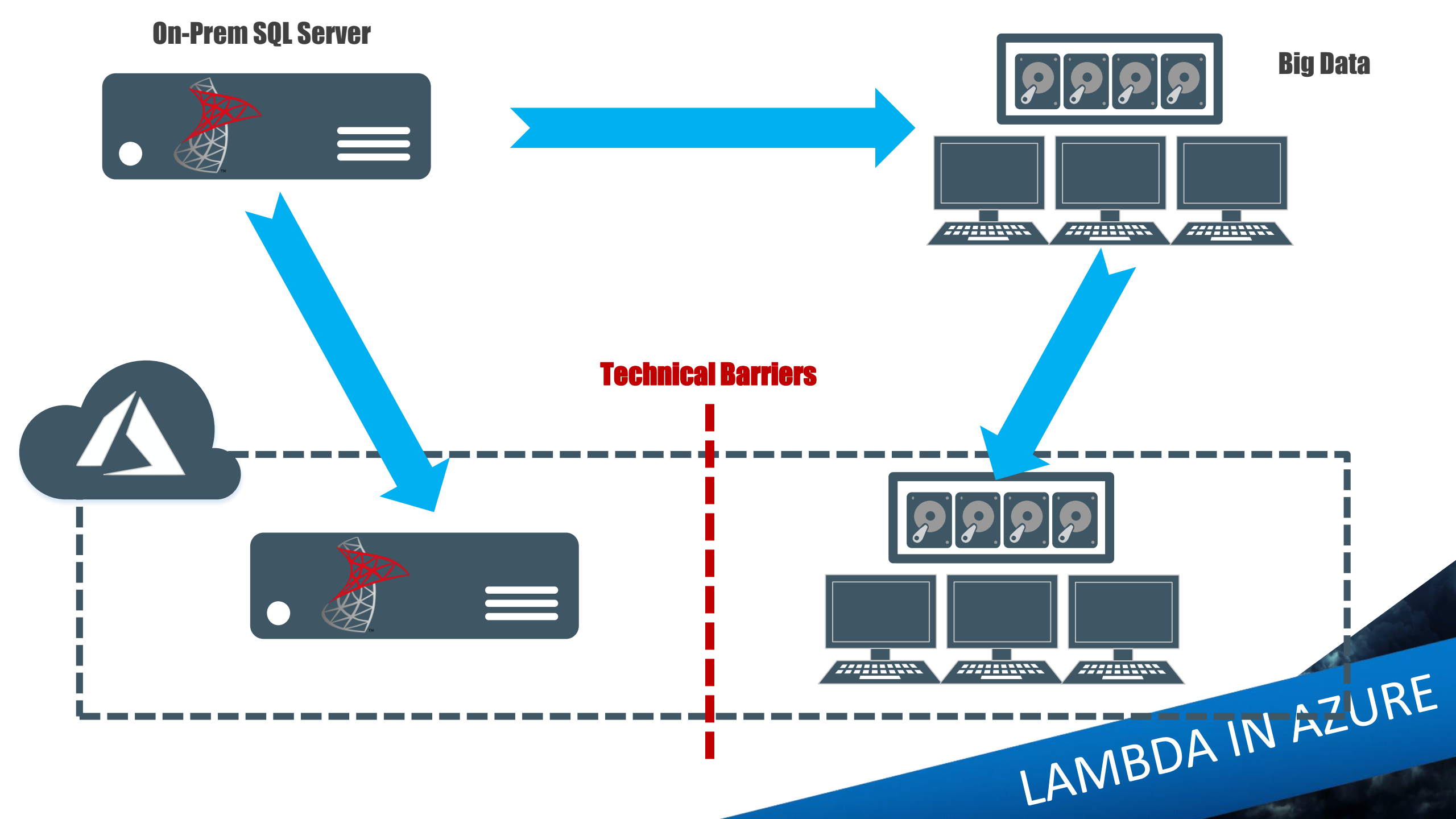
WHAT ARE WE TALKING ABOUT?

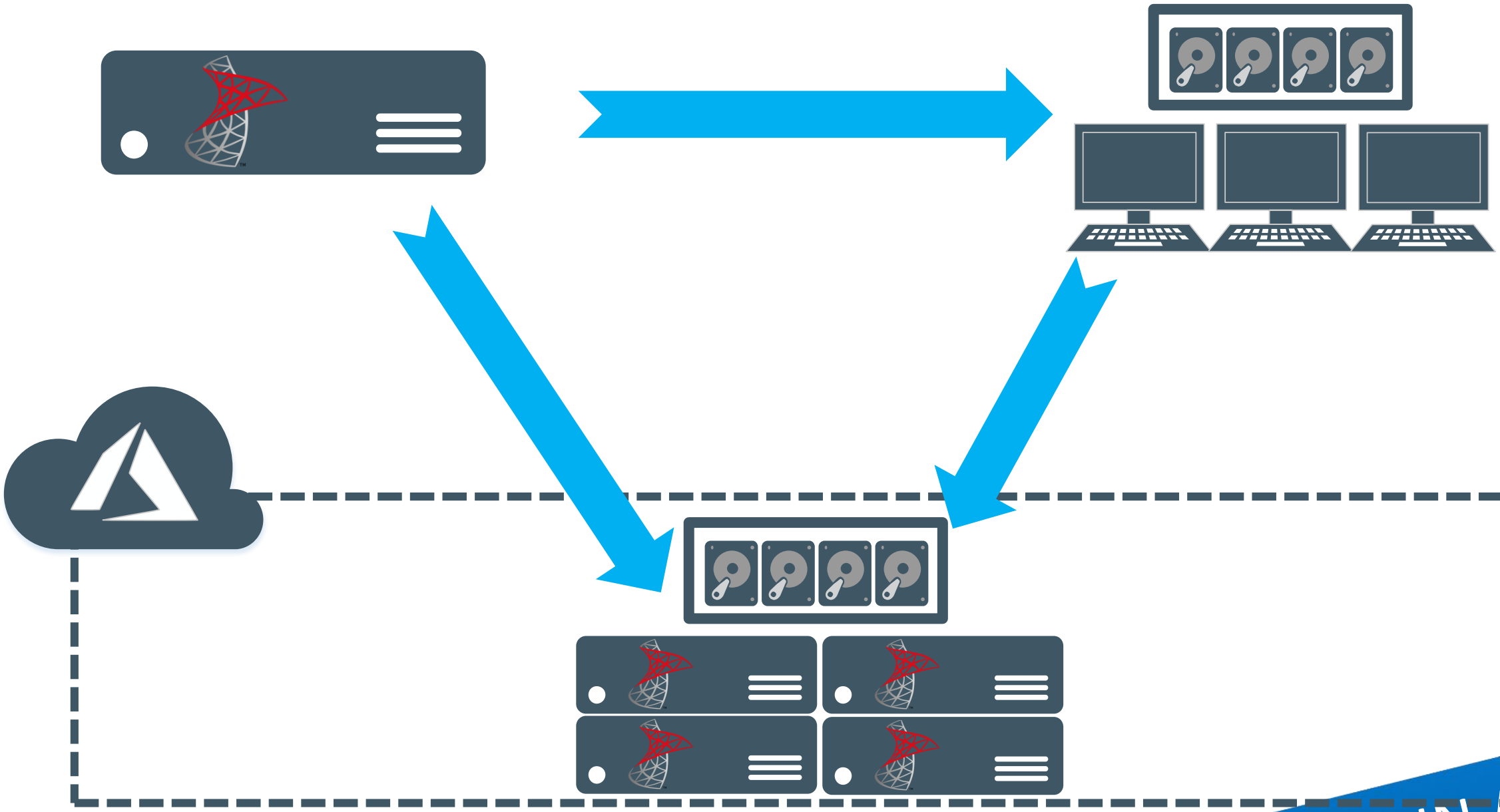
- WHY SHOULD WE LISTED TO BIG DATA?
- WHAT IS LAMBDA?
- BUT HOW DO WE DO LAMBDA IN AZURE?
- IS THAT THE ONLY WAY?
- QUESTIONS!


BIG DATA & CLOUD
A BRIEF HISTORY LESSON



LAMBDA IN AZURE


On-Prem SQL Server
Big Data
Technical Barriers
LAMBDA IN AZURE



LAMBDA IN AZURE

My Life Goal:

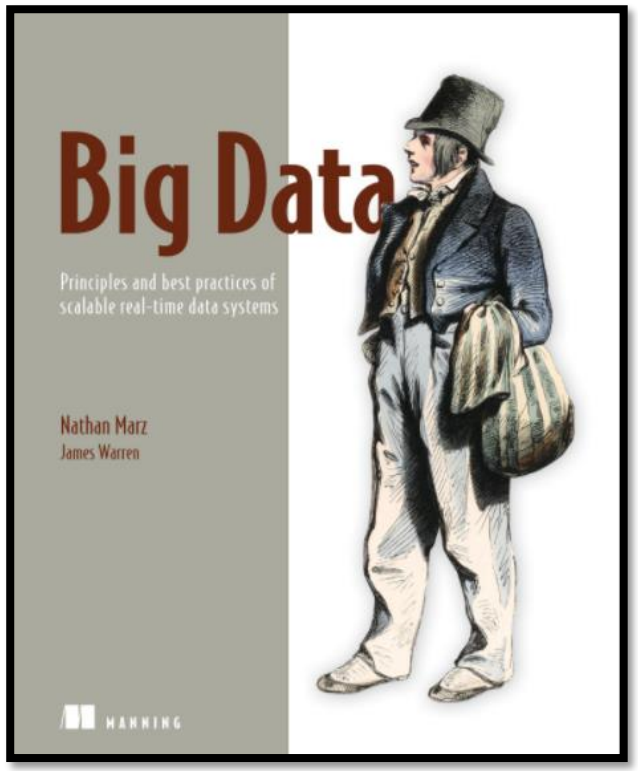
Never to manage another Server

LAMBDA IN AZURE


SO WHAT IS LAMBDA?

LAMBDA ARCHITECTURE

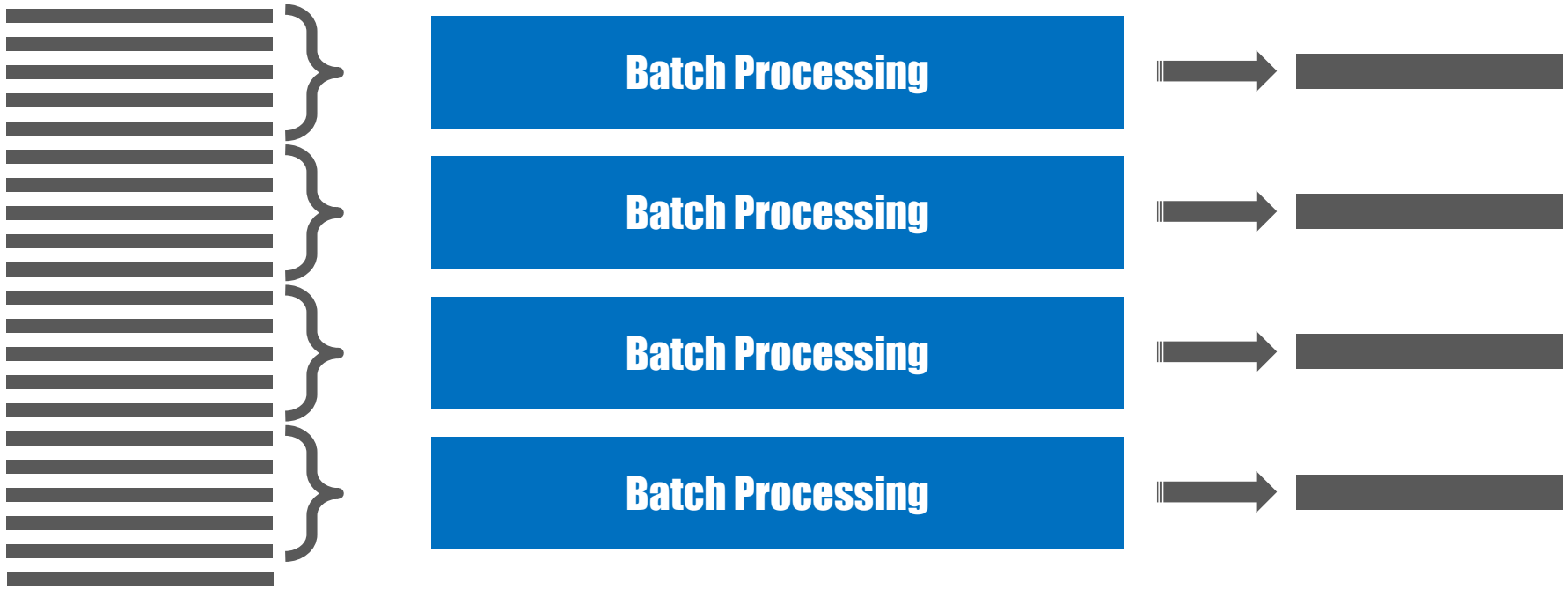
Use Batch and Stream technologies together to balance latency, throughput and fault-tolerance

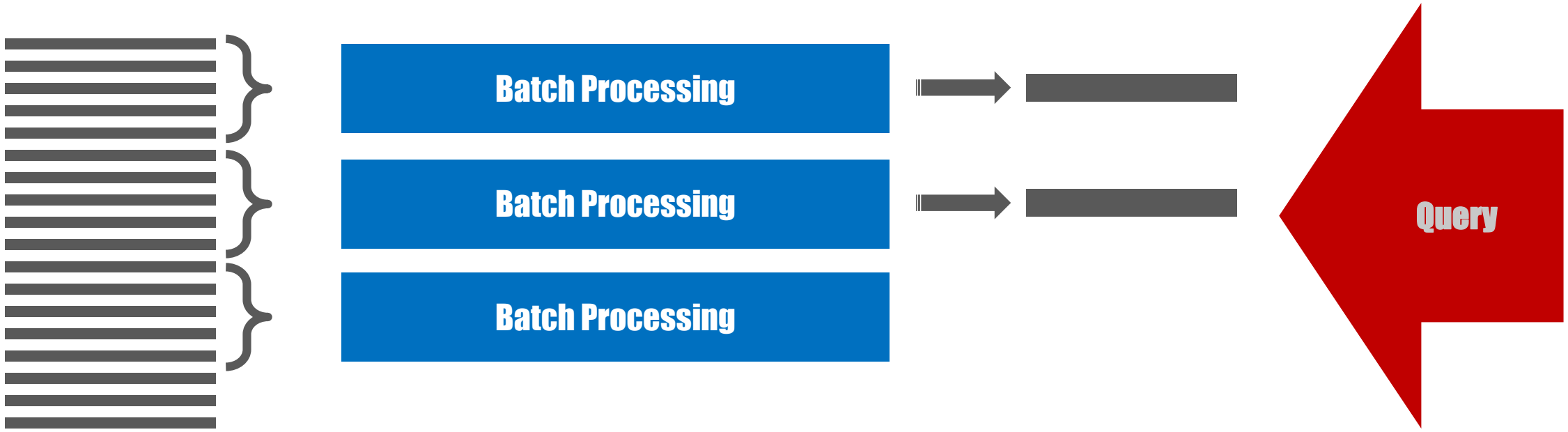


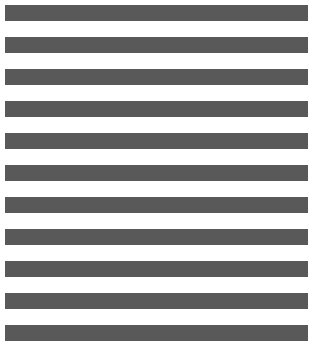
Nathan Marz &
James Warren

LAMBDA IN AZURE

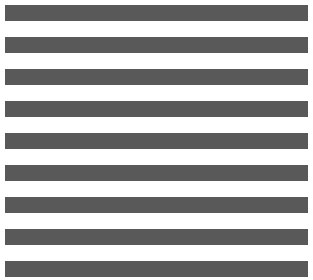
THE PROBLEM...

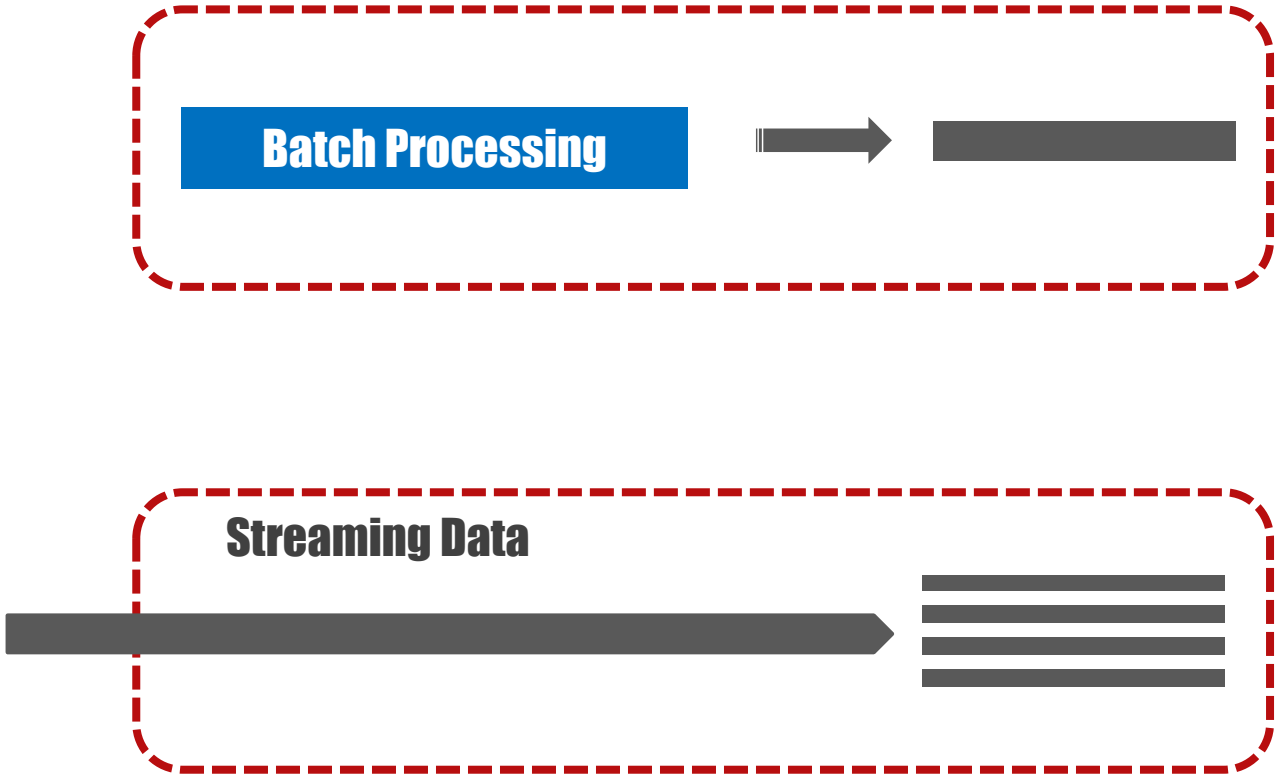

Batch Processing
Batch Processing
Batch Processing
Batch Processing


Batch Processing
Batch Processing
Batch Processing
Query



LAMBDA IN AZURE



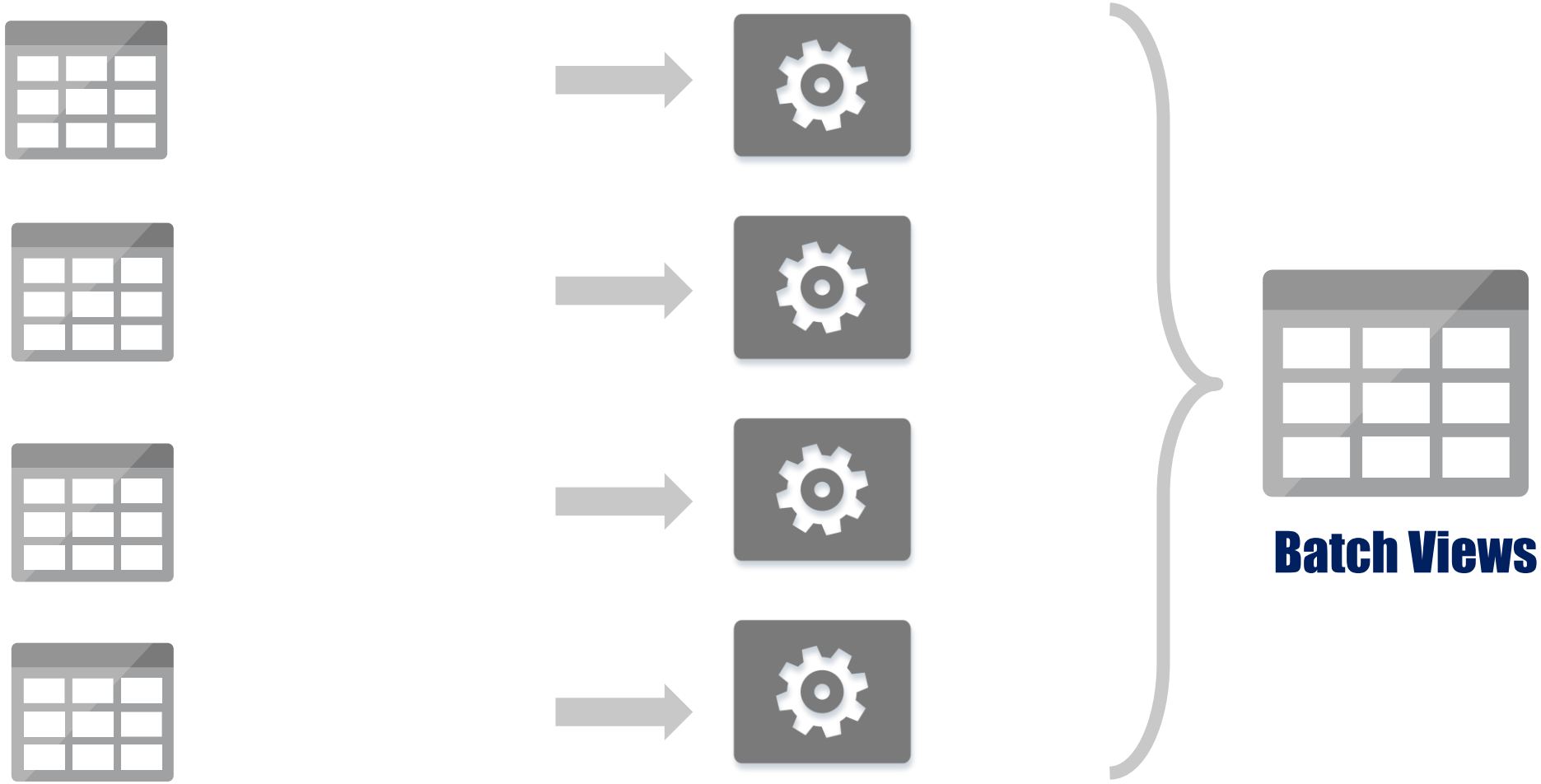


Query

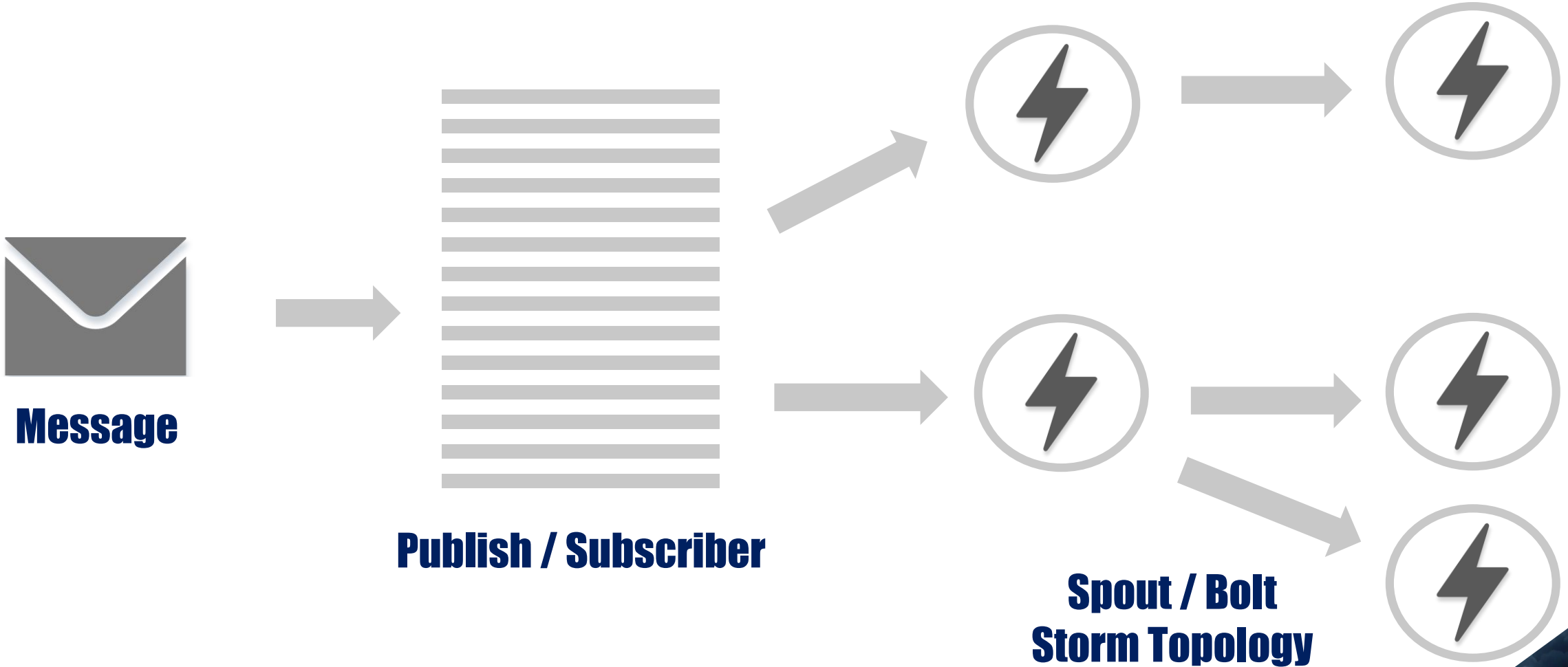
LAMBDA IN AZURE

THE LAMBDA ARCHITECTURE

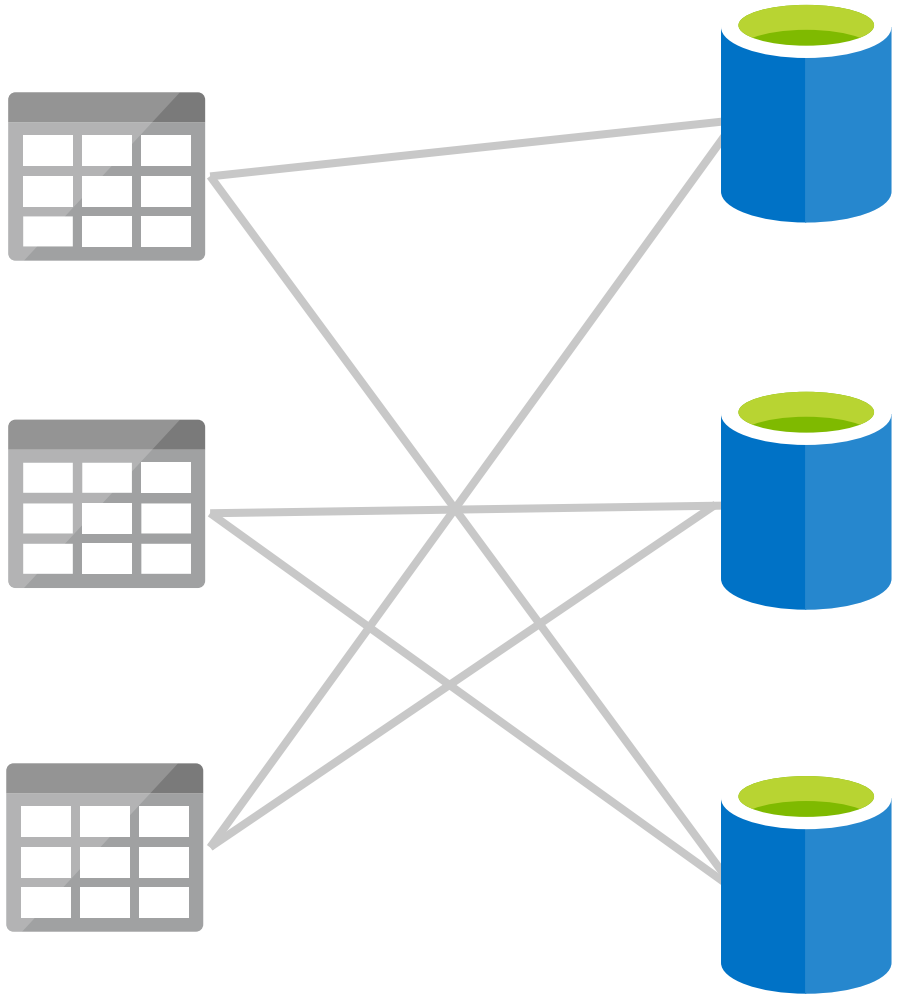
Batch Layer


Distributed Storage
(HDFS)
Parallel Processing
(MapReduce)
Batch Views

Speed Layer



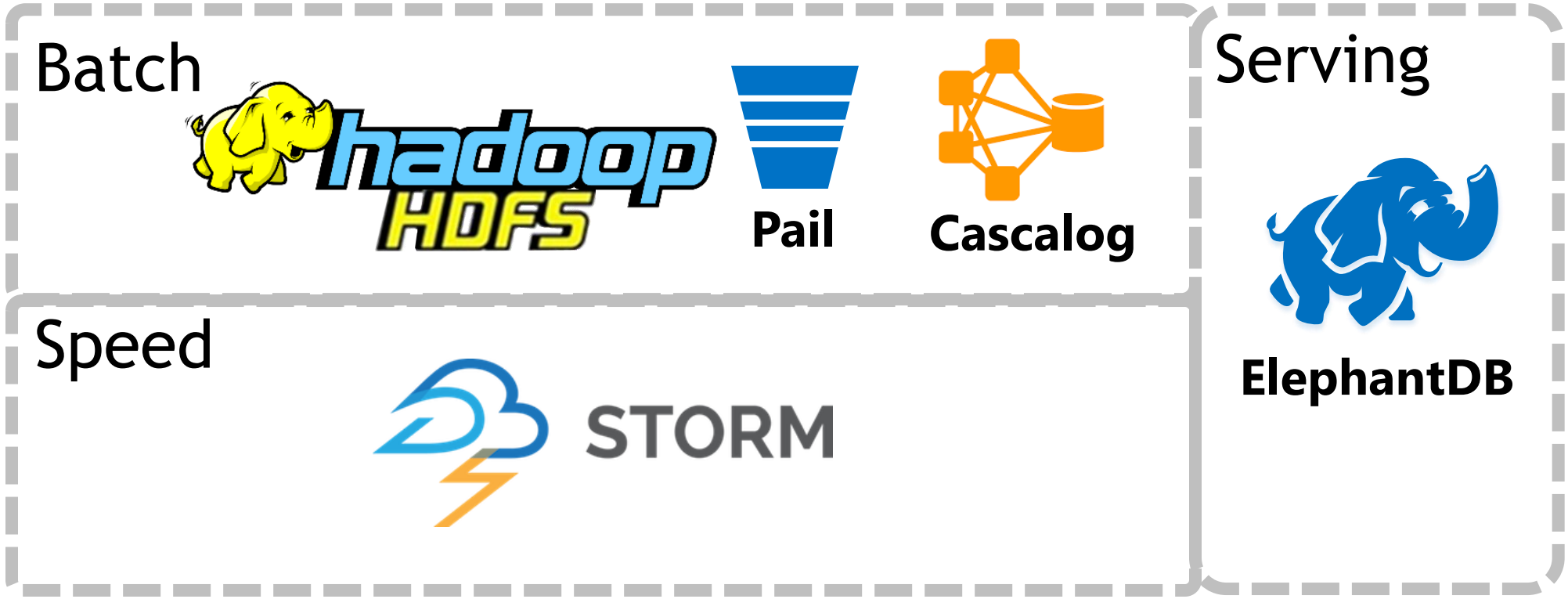
Serving Layer



Sharded Database Layer

Batch Views

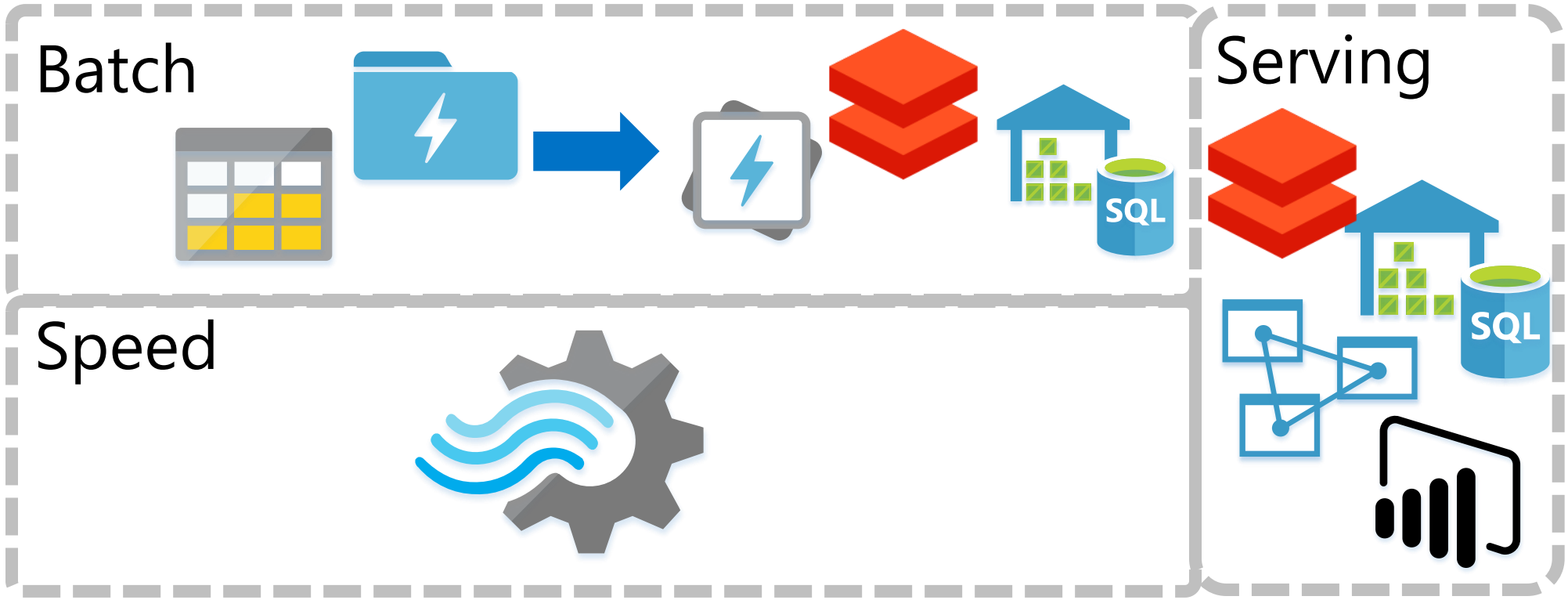
THE MARZ LAMBDA ARCHITECTURE



LAMBDA IN AZURE


LAMBDA IN AZURE

APPLYING LAMBDA IN AZURE



BATCH LAYER - STORAGE

Blob Storage

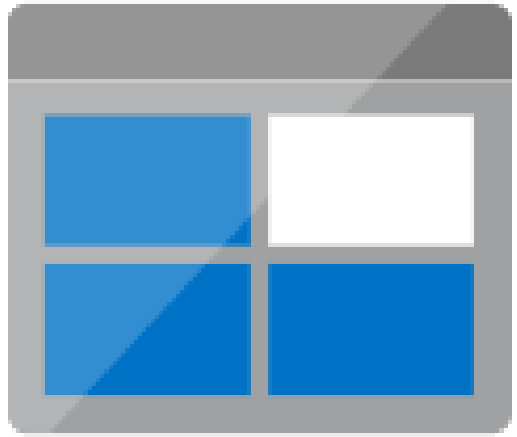
- HDFS
- Hot/Cold Storage Tiers
- Limited Security
- File Size Limitations
- Widely Compatible / Available

Azure Data Lake Store

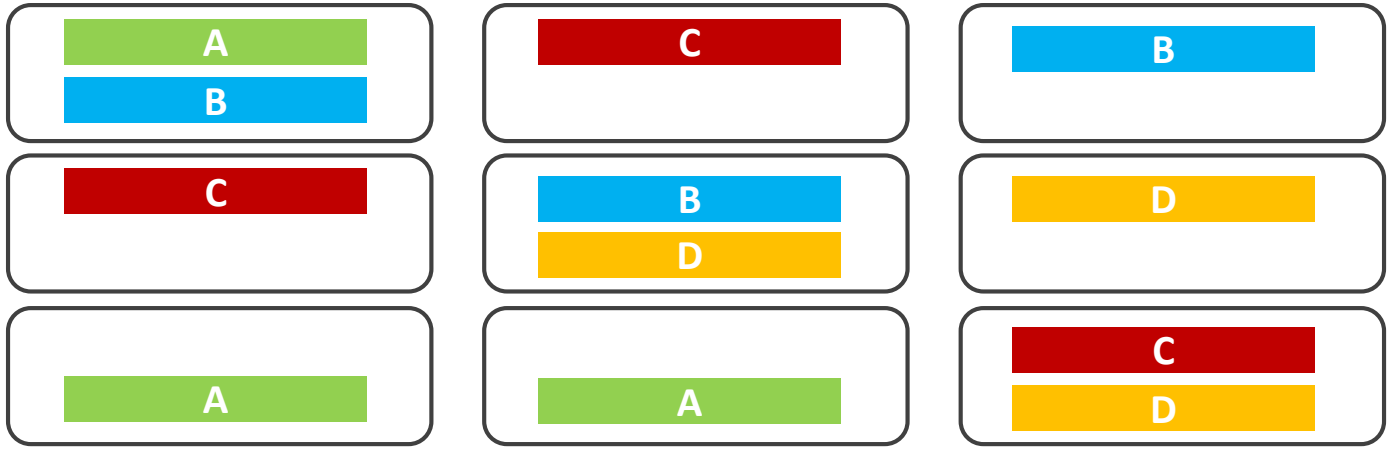
- WHDFS
- Single Pricing Model
- AAD-Integrated Security
- No Limitations
- Still Maturing

AZURE DATA LAKE STORE GEN2

- HDFS
- Hot/Cold Storage Tiers
- Full Security
- File Size Limitations
- Little Bit Broken Still...



HDFS



Parallelism

Fault Resilience

LAMBDA IN AZURE

BATCH LAYER - COMPUTE

Azure Data Lake Analytics

- Pay Per Query / Unit
- U-SQL
- Outputs Structured/Unstructured
- Uses MapReduce-style processing
- Batch Mode

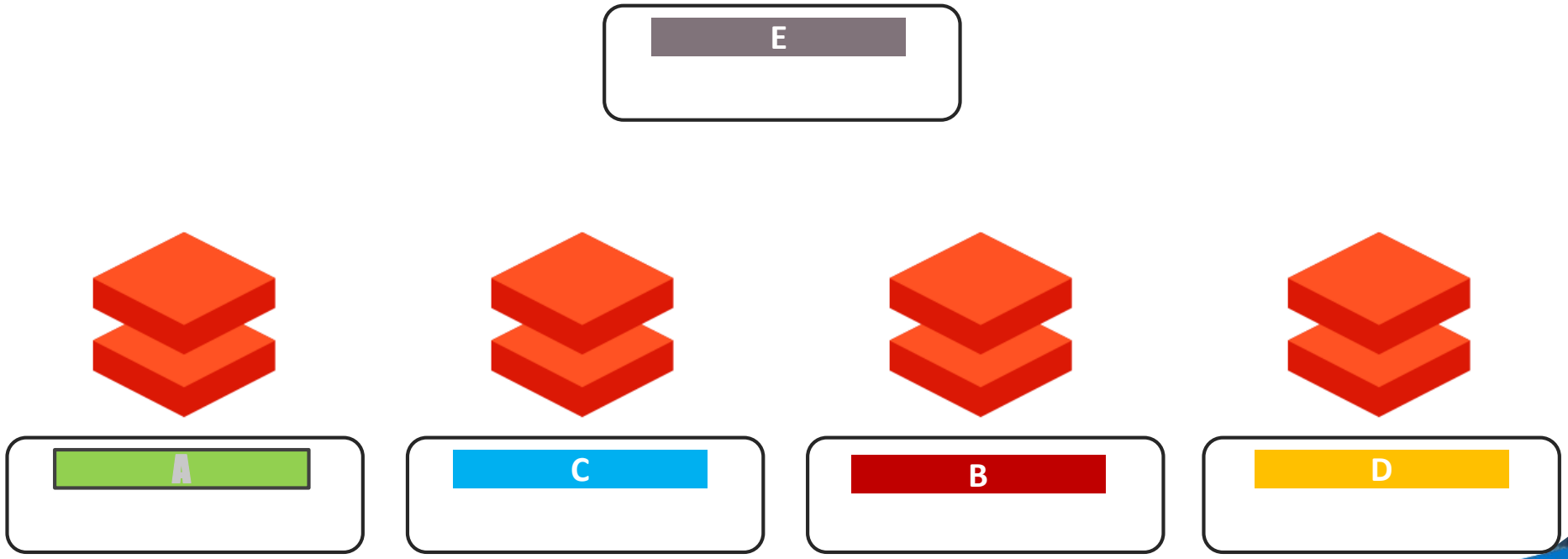
Azure DataBricks

- Pay Per second / Node
- Python/Scala/SQL
- Structured/Unstructured
- Uses in-memory Spark processing
- Batch or Live Query

Azure SQL DataWarehouse

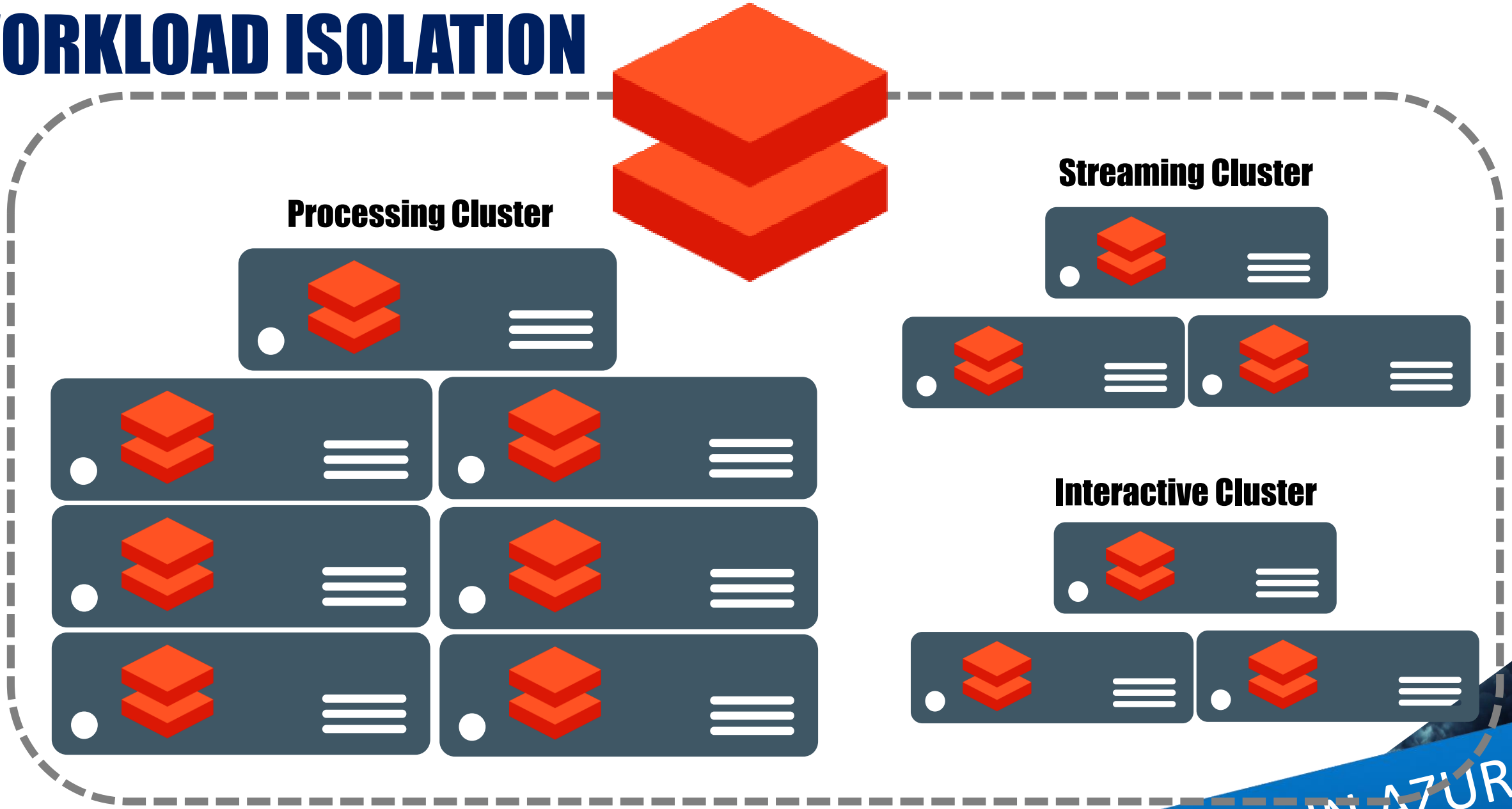
- Pay Per Hour / Node
- T-SQL
- Fully Structured
- Can use MapReduce via Polybase
- Batch or Live Query

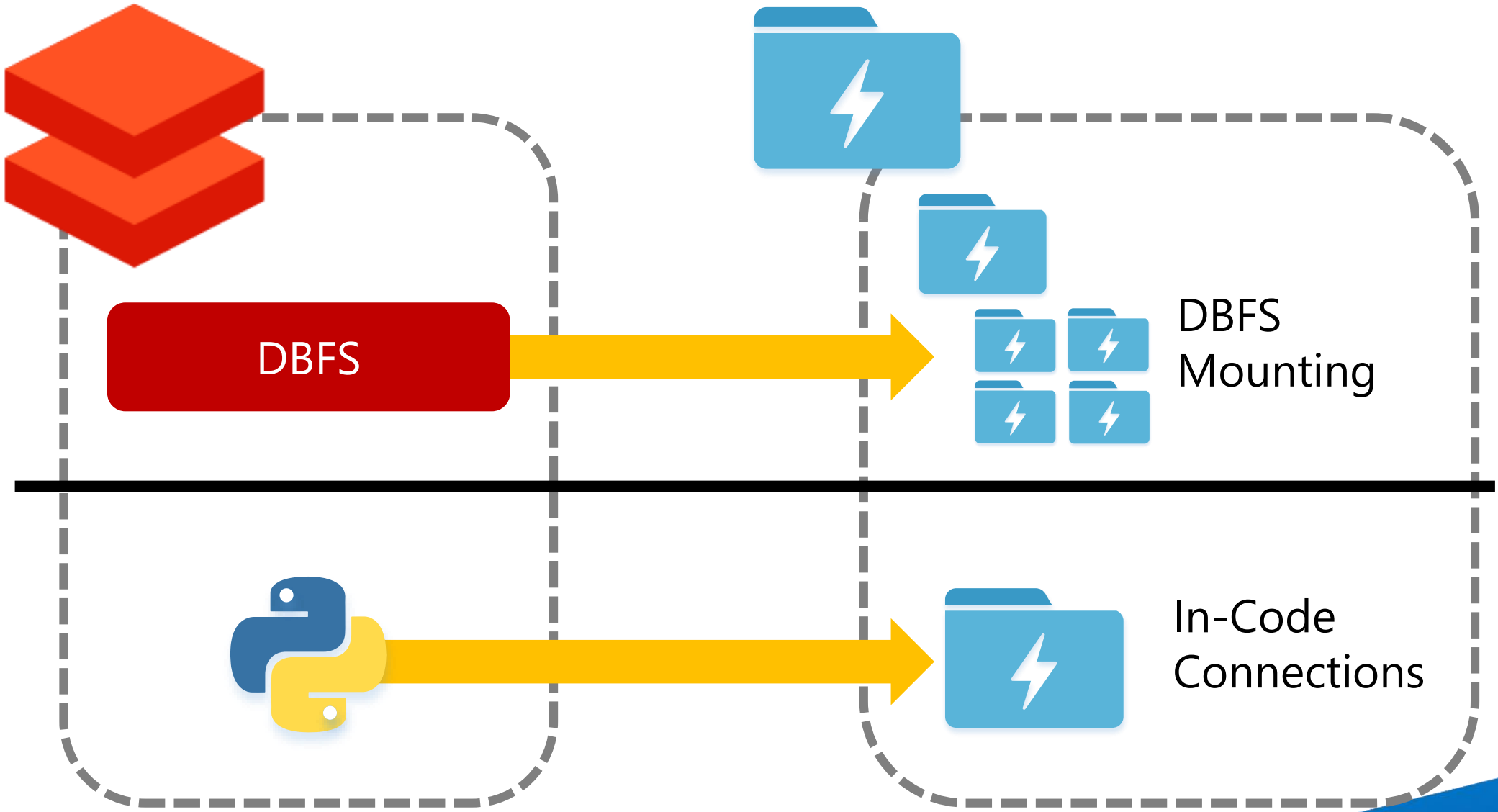
Azure Databricks



LAMBDA IN AZURE

WORKLOAD ISOLATION





LAMBDA IN AZURE

Azure Streaming Analytics

- Only PaaS Native Offering
- Uses SQL Language
- Built-in Azure Integrations
- Can Vertically Partition Files
- Can Write to Multiple Outputs

/Input/2017/06/19/0900.csv

/Input/2017/06/19/1000.csv

/Input/2017/06/19/1100.csv

/Input/2017/06/19/1200.csv

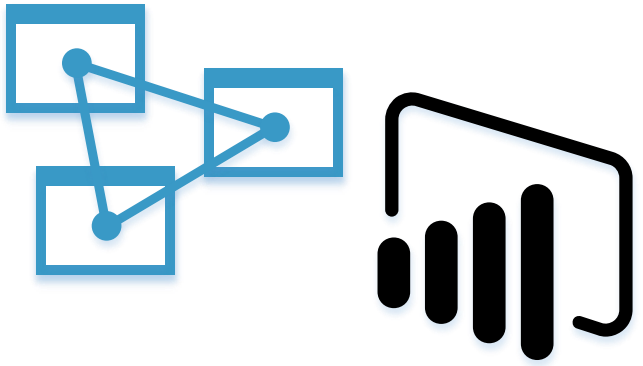
SERVING LAYER

Azure SQL DataWarehouse

- Low Concurrency (32!)
- Direct Query via Polybase
- Huge data capacity

Azure DataBricks

- Med Concurrency
- Direct Query over Data Lake Store
- Required Python/Scala knowledge

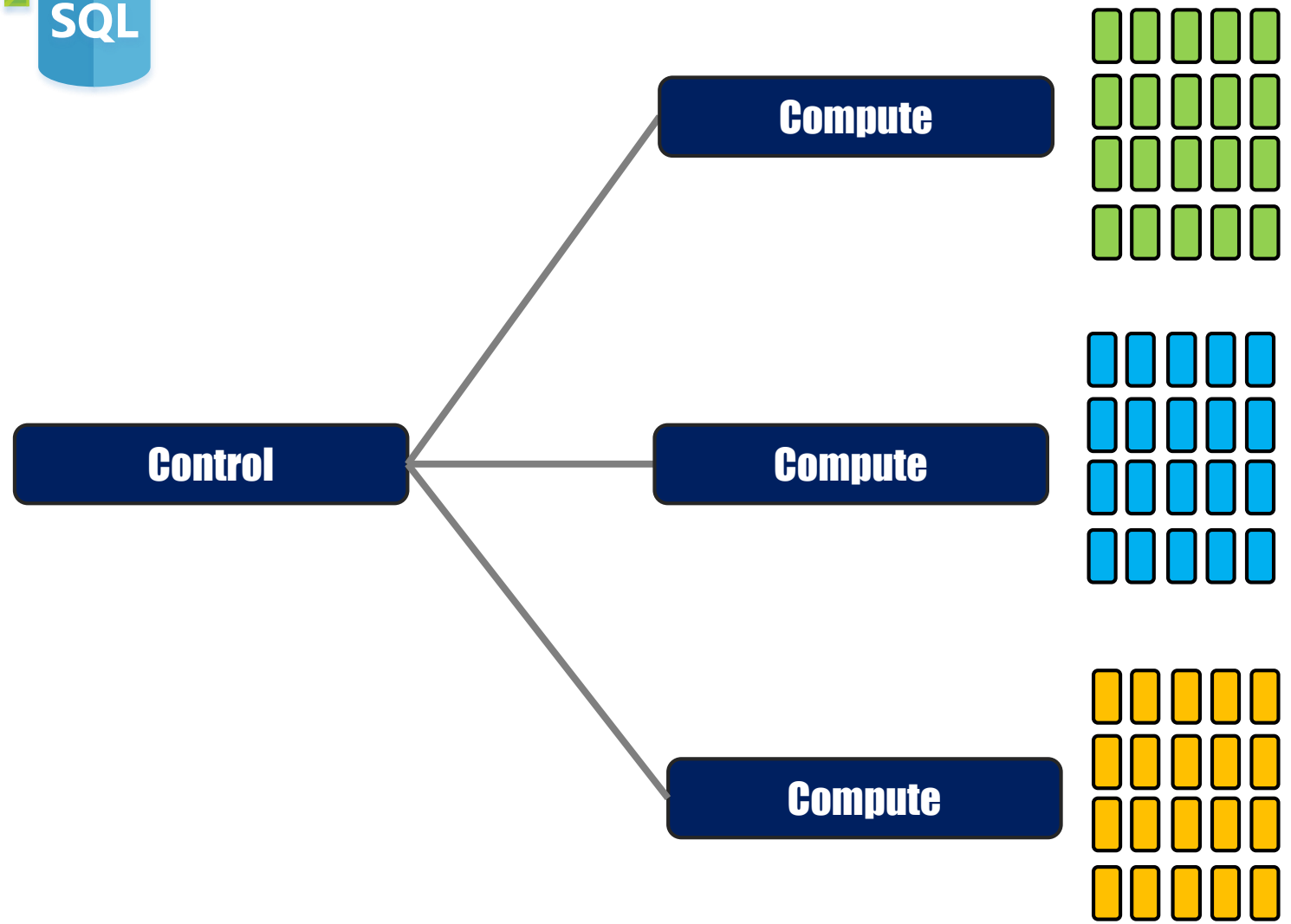


SSAS / PowerBI

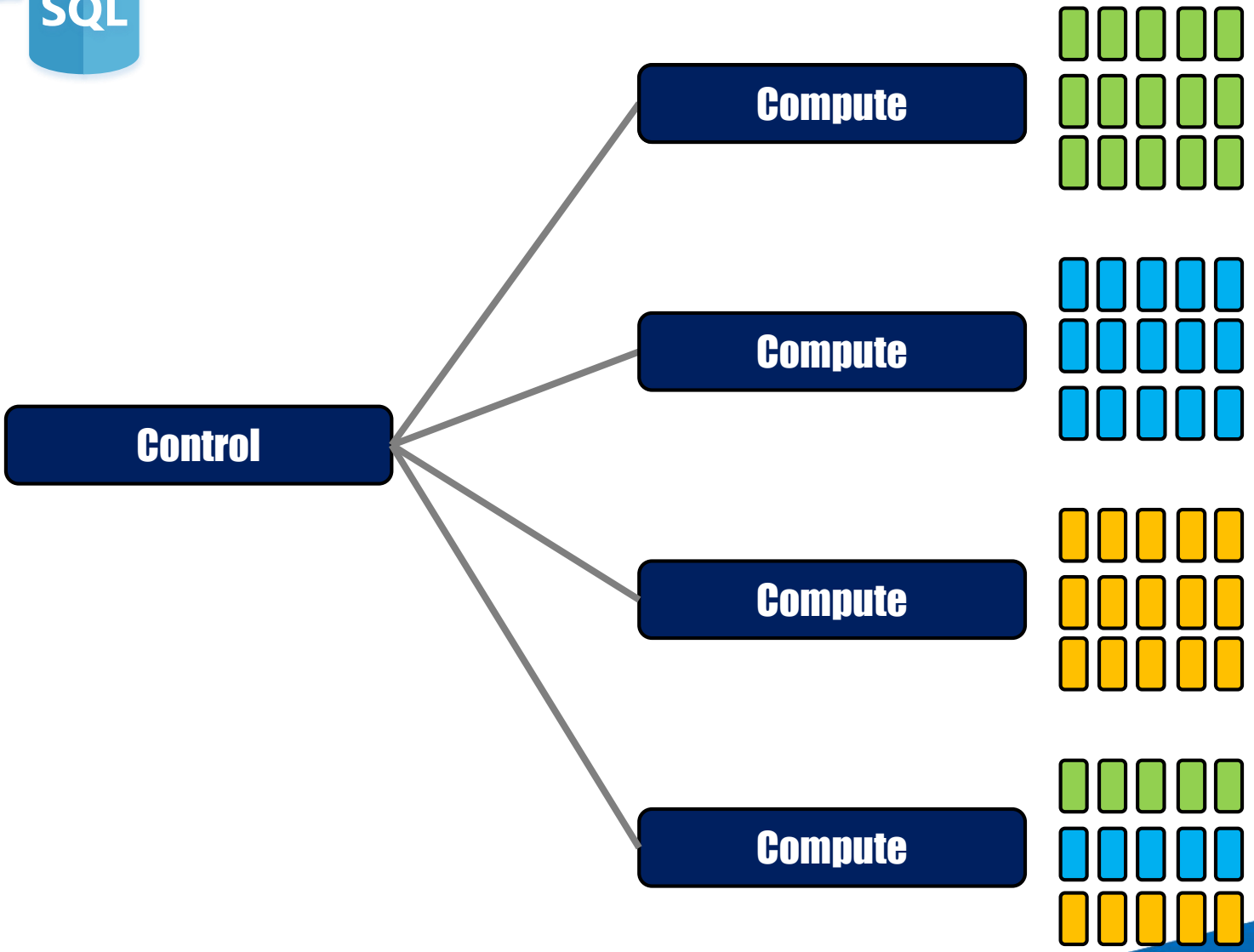
- High Concurrency
- Scheduled Refresh / Direct over DBs
- Model Size Limits

LAMBDA IN AZURE

Azure SQL DataWarehouse

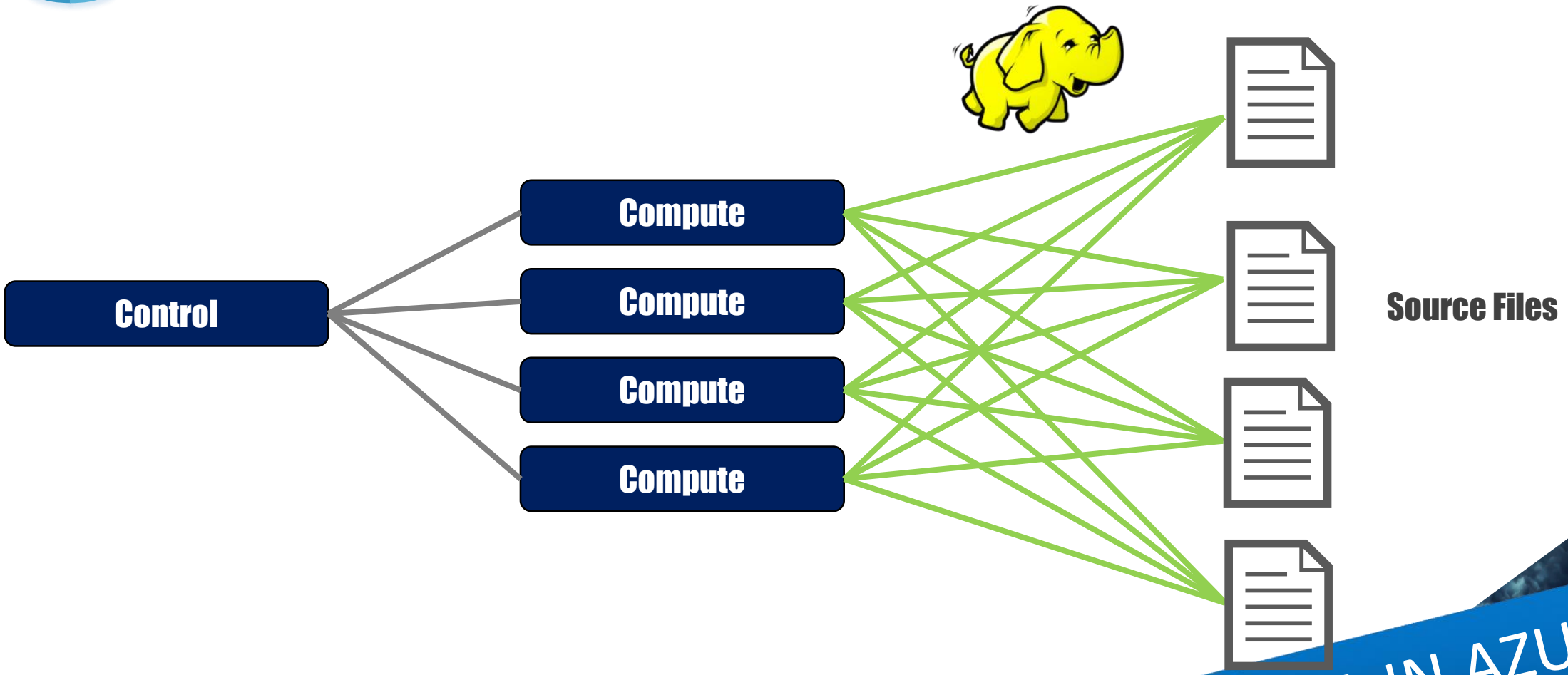


Azure SQL DataWarehouse

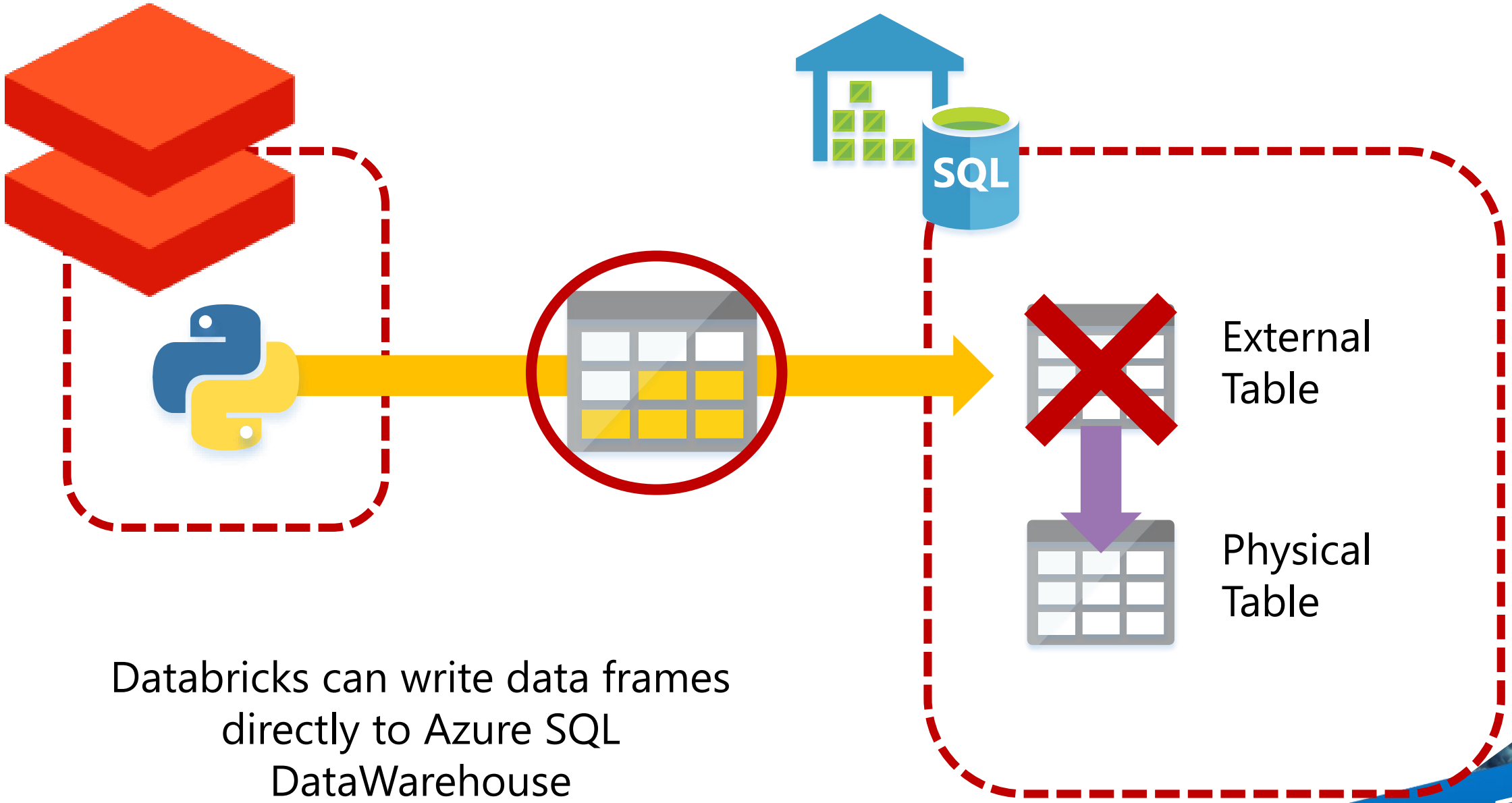


Azure SQL DataWarehouse

PolyBase

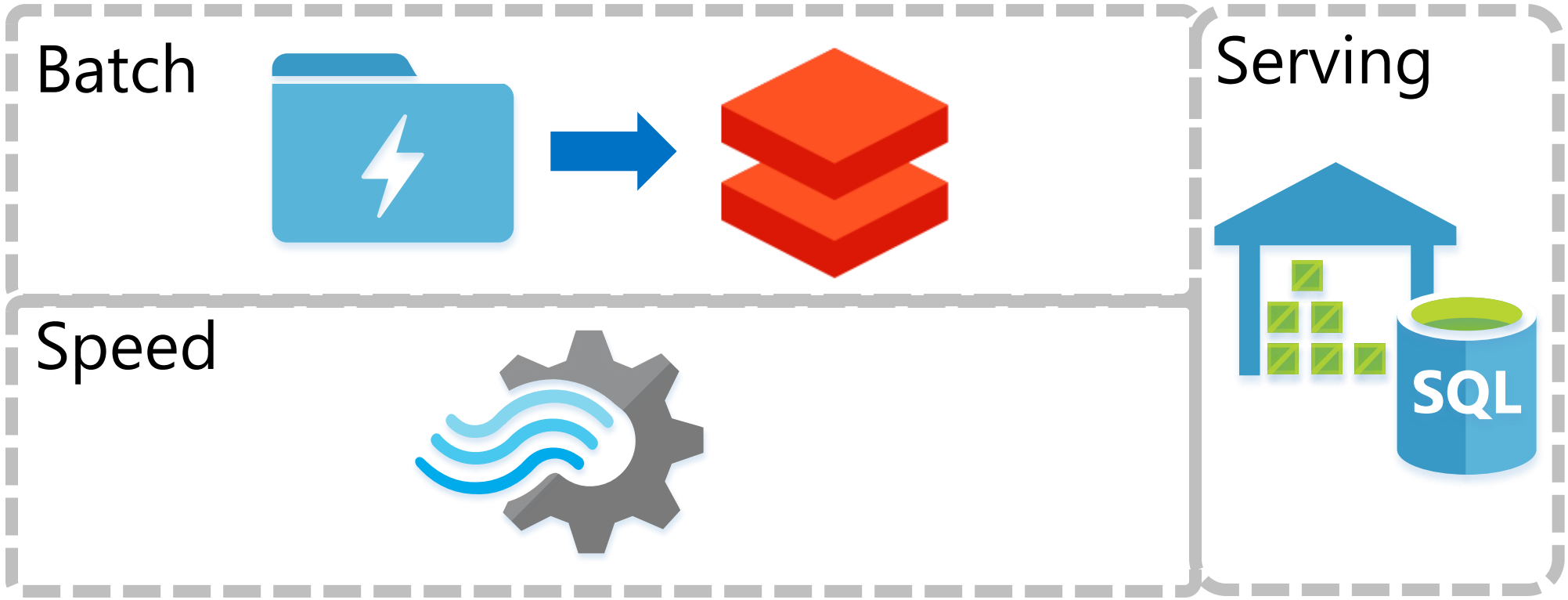


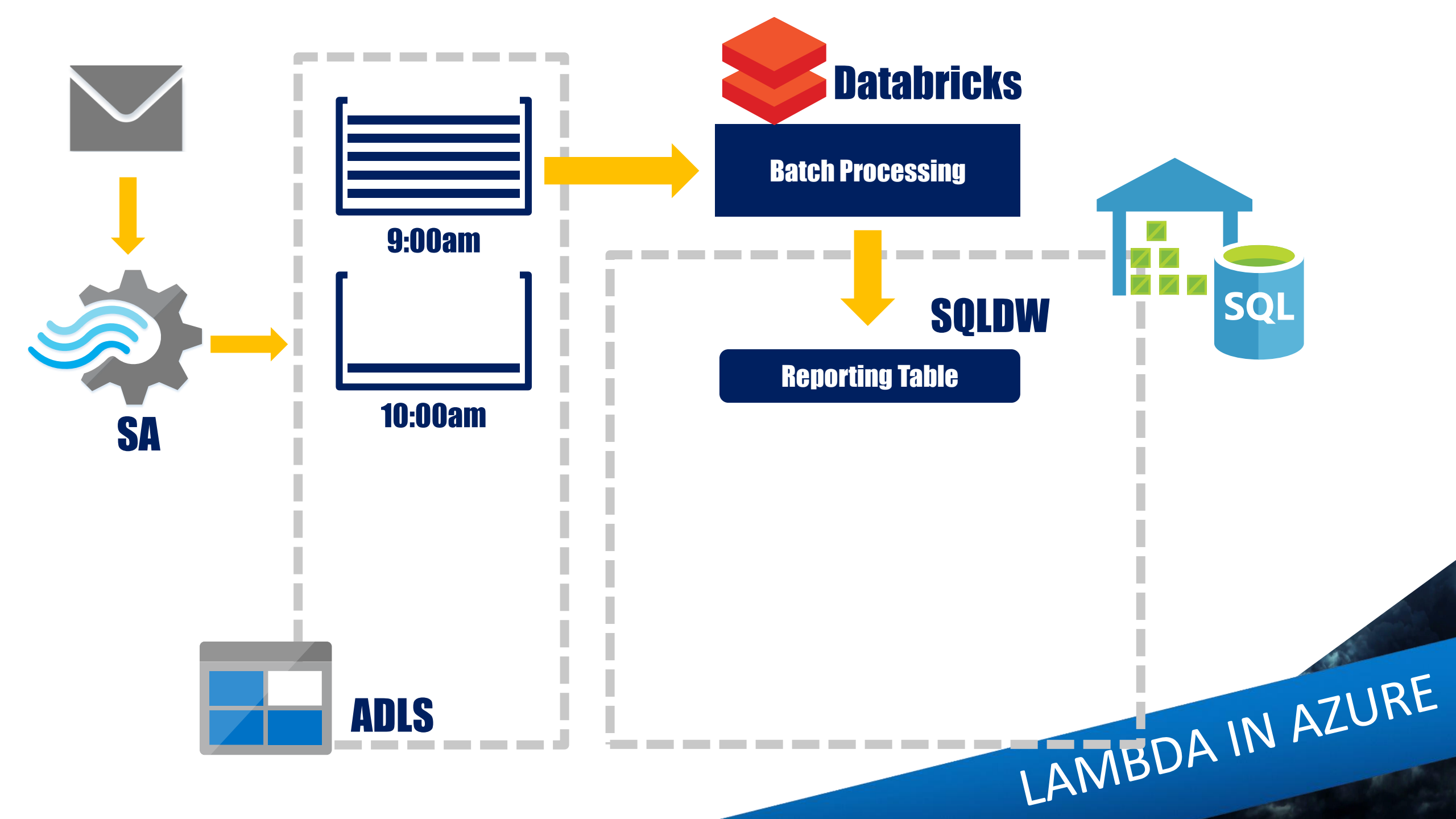
LAMBDA IN AZURE

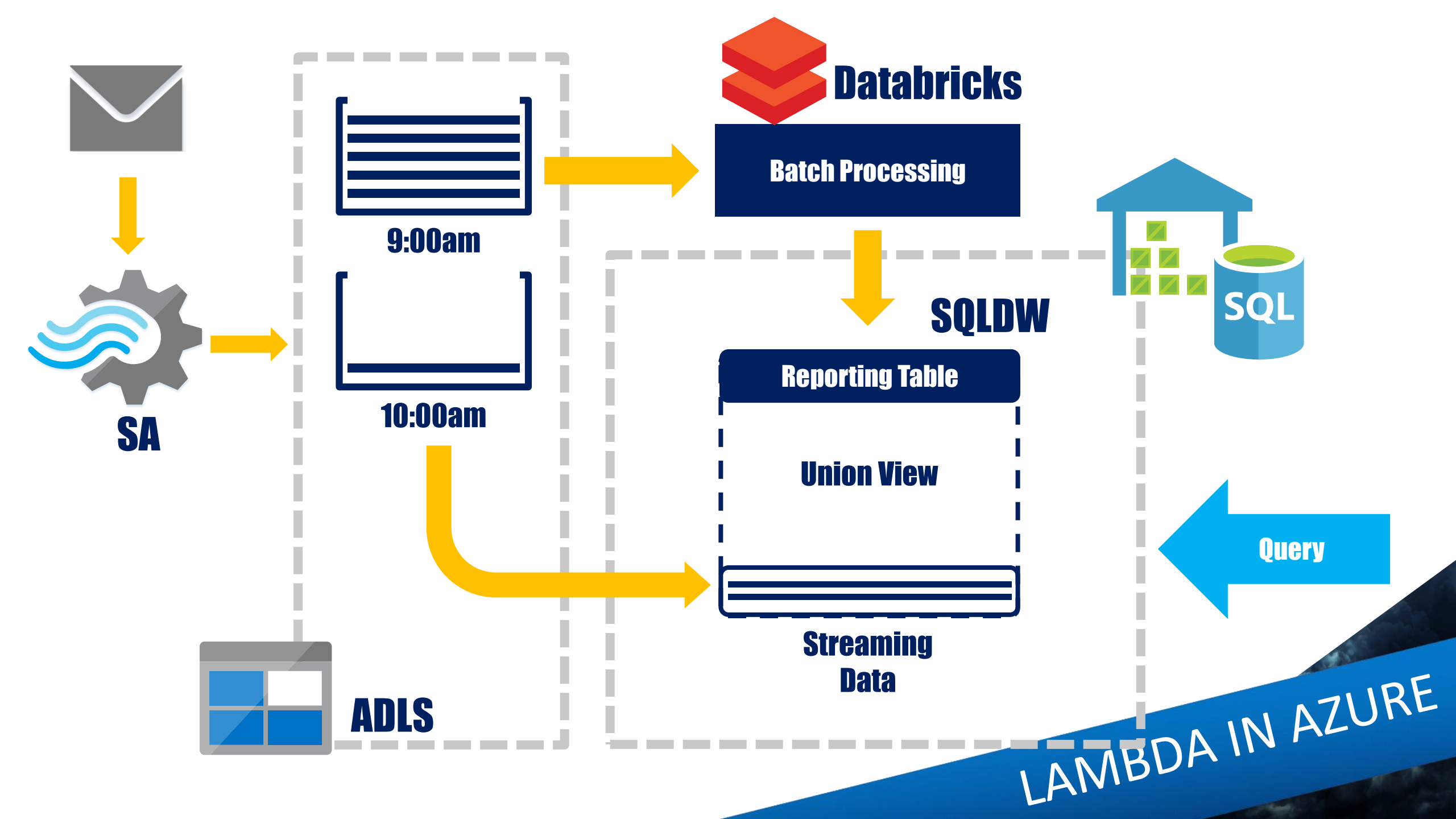


Databricks can write data frames directly to Azure SQL DataWarehouse

APPLYING LAMBDA IN AZURE




Databricks
Batch Processing
9:00am
10:00am
SA
ADLS
SQLDW
Reporting Table
SQL
LAMBDA IN AZURE

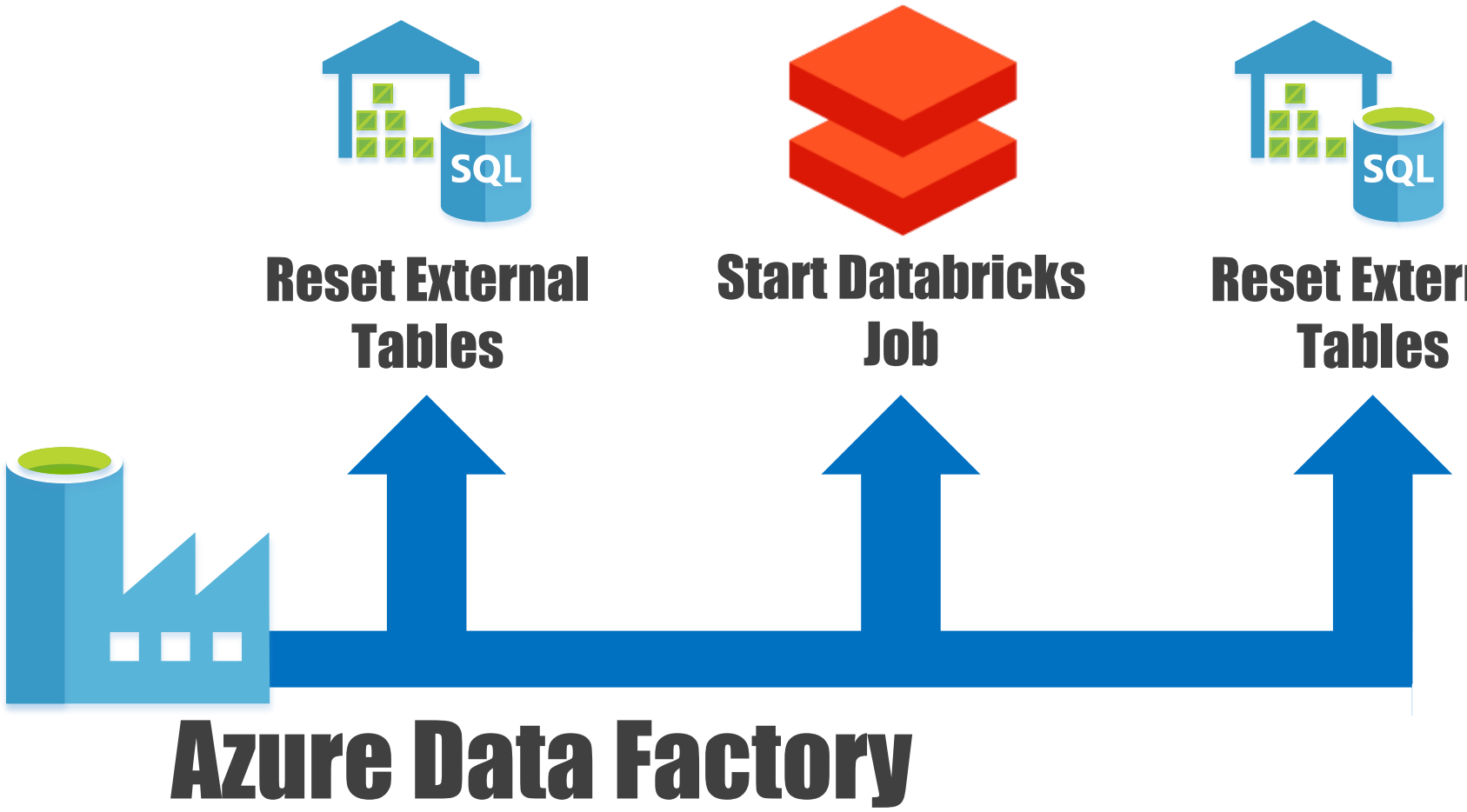

Databricks
Batch Processing
9:00am
10:00am
SA
ADLS
SQLDW
SQL
Reporting Table
Union View
Streaming
Data
Query
LAMBDA IN AZURE

DEMO:

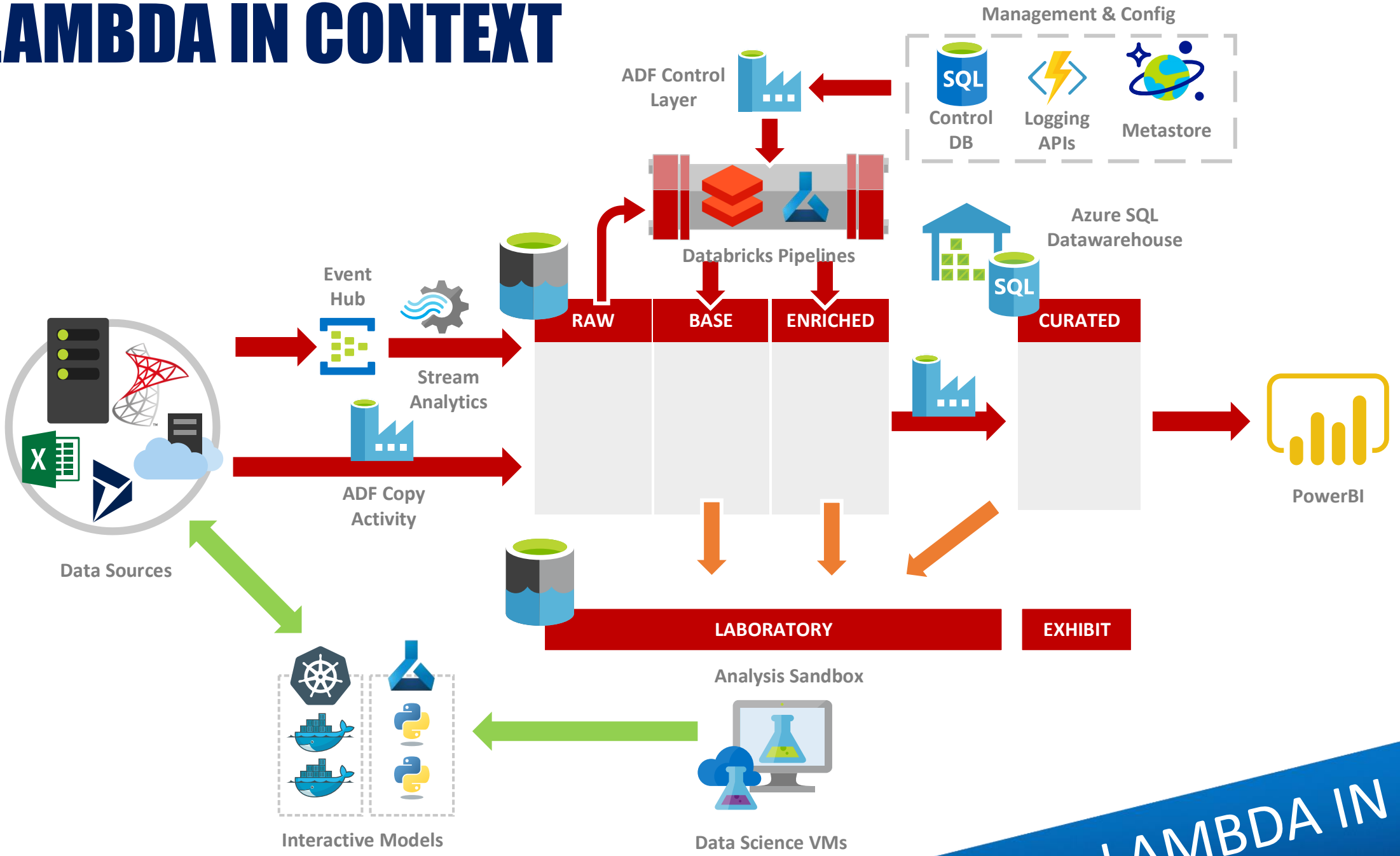
LAMBDA IN AZURE

- Streaming Analytics
- Event Hubs
- Data Lake Store
- Azure Databricks
- Azure SQL Datawarehouse

LAMBDA ORCHESTRATION

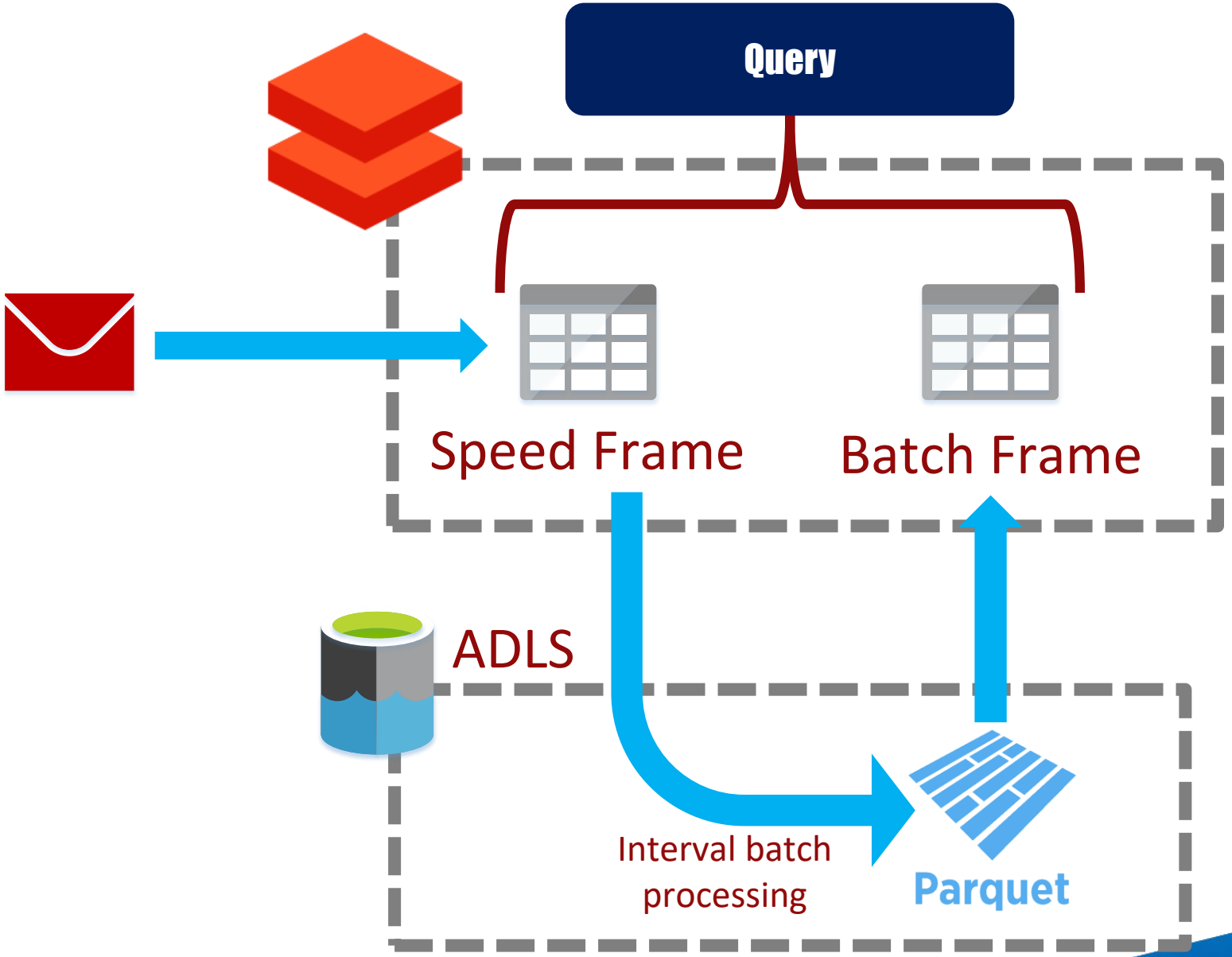


LAMBDA IN CONTEXT



LAMBDA IN AZURE

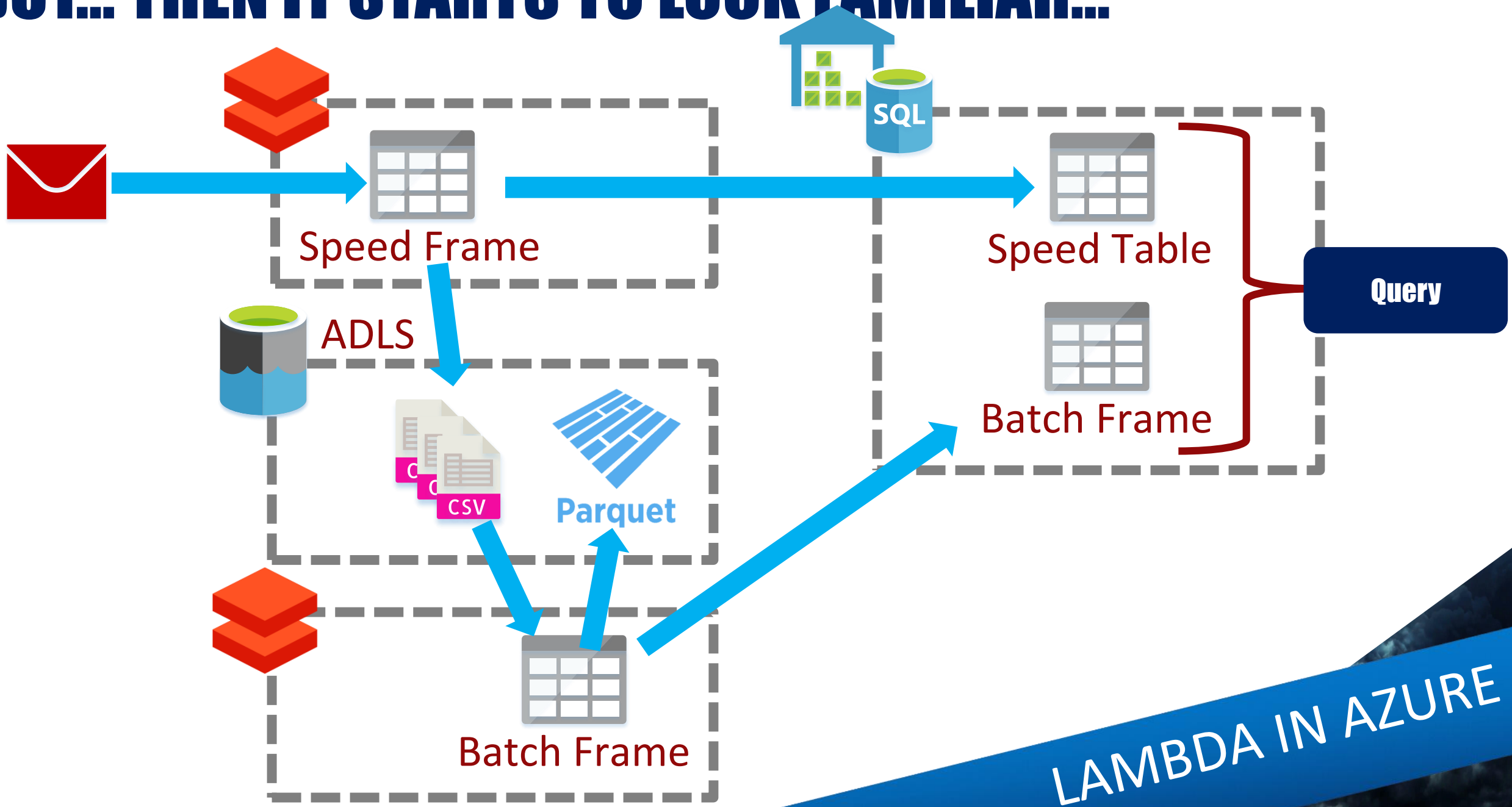
THE OPEN-SOURCE APPROACH



It's possible to do everything inside Databricks…

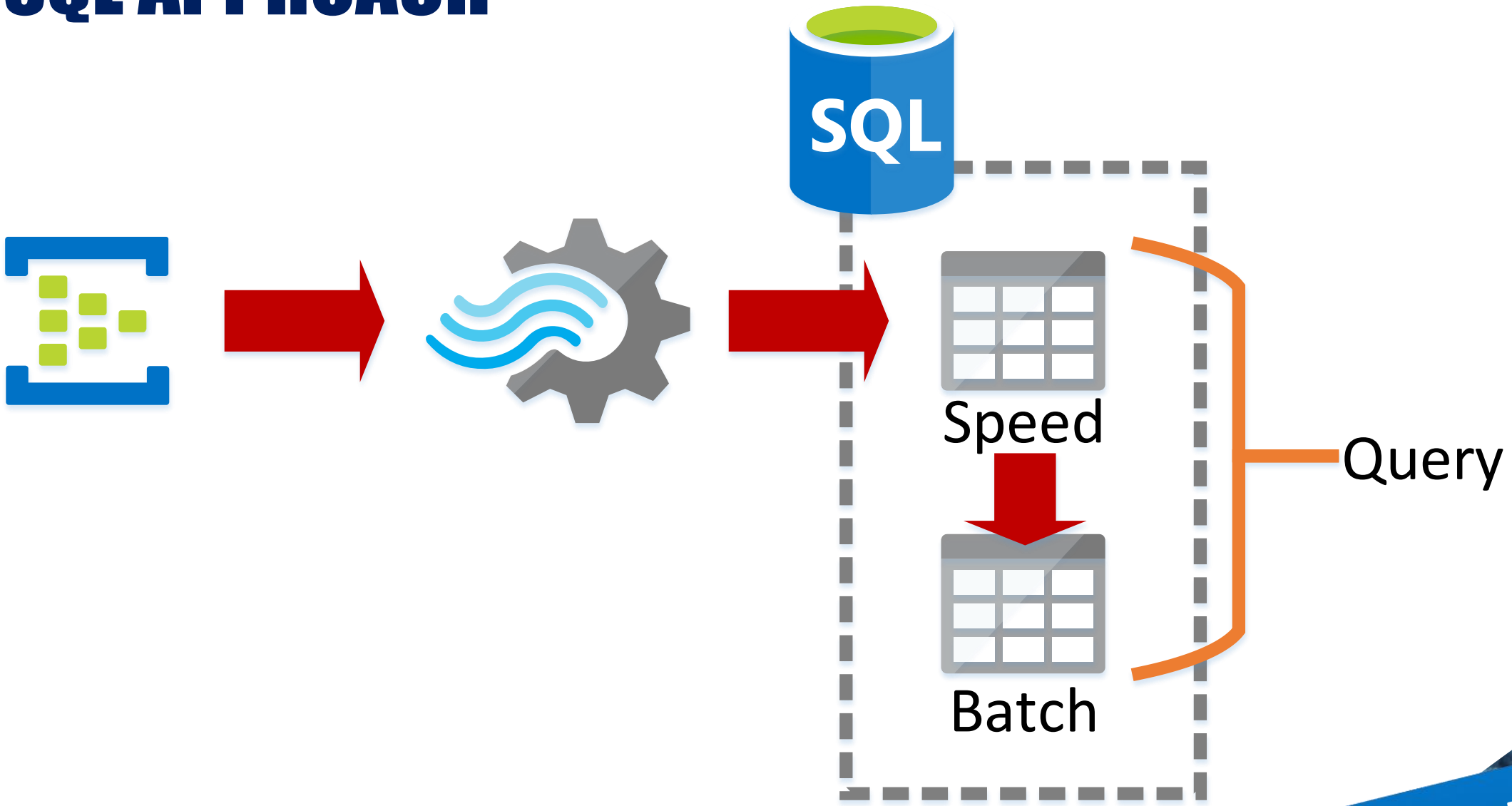
But the cluster can be expensive to leave running!

BUT... THEN IT STARTS TO LOOK FAMILIAR...

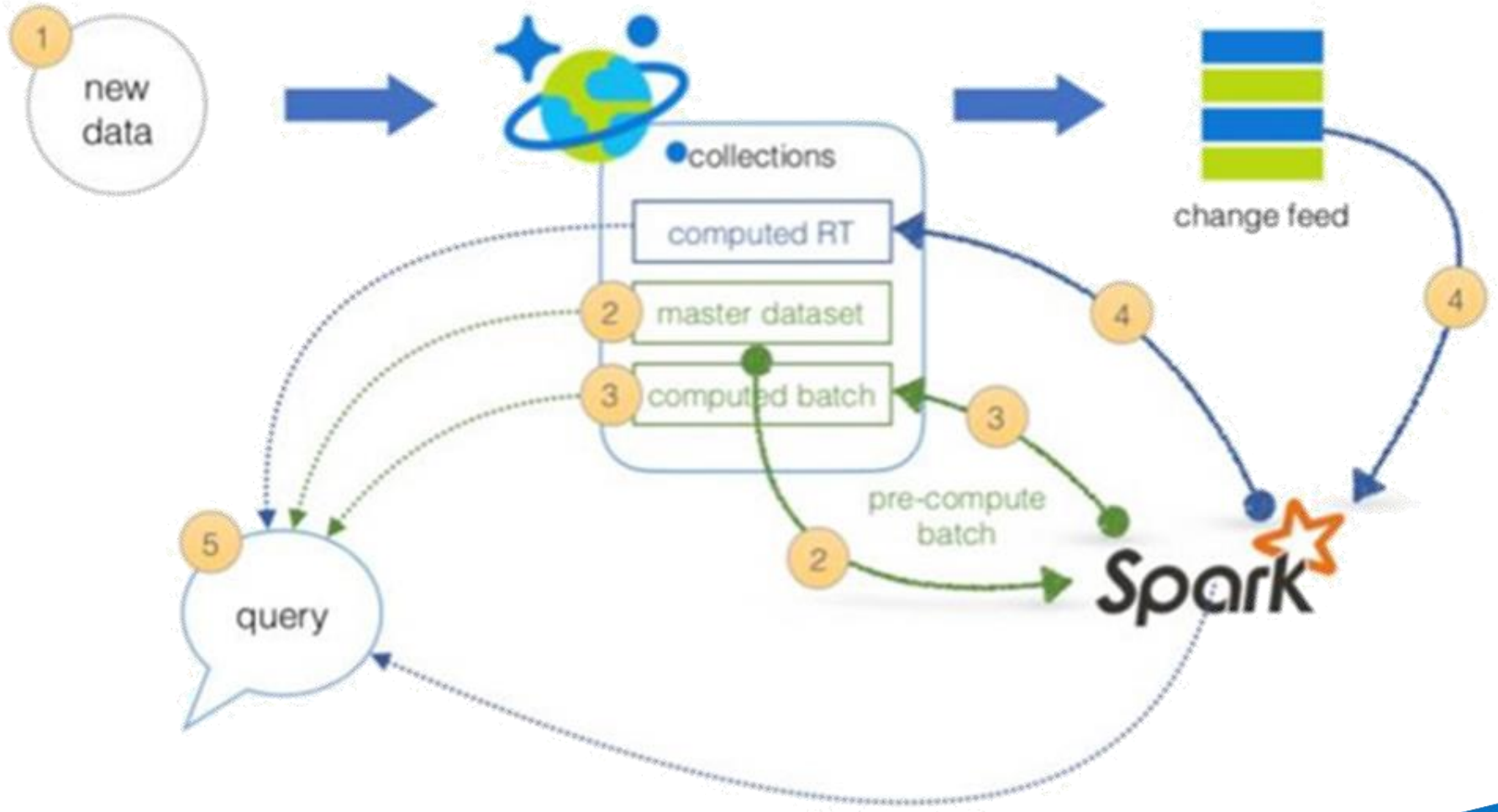


LAMBDA IN AZURE

THE SQL APPROACH



COSMOS DB



LAMBDA IN AZURE

QUESTIONS?

Simon Whiteley
@MrSiWhiteley

ADVANCING ANALYTICS
Data Science | AI | DataOps | Engineering

@AdvancingAnalyticsUK
www.advancinganalytics.co.uk

DATA SCIENCE IN PRODUCTION PODCAST