PASS SUMMIT2018

Lambda Architectures in Azure

The Modern Data Warehouse

Simon Whiteley, Chief Cloud Architect, Adatis


Please silence
cell phones

Explore everything PASS has to offer

Free Online Resources

Newsletters

PASS.org

24HOURS OF PASS

Free online webinar events

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

Session evaluations

Your feedback is important and valuable.

Submit by 5pm Friday, November 16th to win prizes.

3 Ways to Access:

Go to passSummit.com

Download the GuideBook App
and search: PASS Summit 2018

Follow the QR code link displayed on session signage throughout the conference venue and in the program guide

PASS

Simon Whiteley

Chief Cloud Architect, Adatis

@MrSiWhiteley

Cloud Herder

Worked with BI for an age, but Azure Analytics for the past 3 years – it's been a bumpy ride but now we're somewhere pretty amazing!

London Cyclist

Some say we're suicidal, some say we're a menace to society. I just like cycling rather than cramming myself on the tube every day!

Other PASStimes

SQLSurrey PASS Chapter Leader & SQL London Co-Organiser. I talk at people about azure, data, analytics and crazy cloudy things…

What are we talking about?

Cloud BI

What is Lambda?

The Native Azure Approach

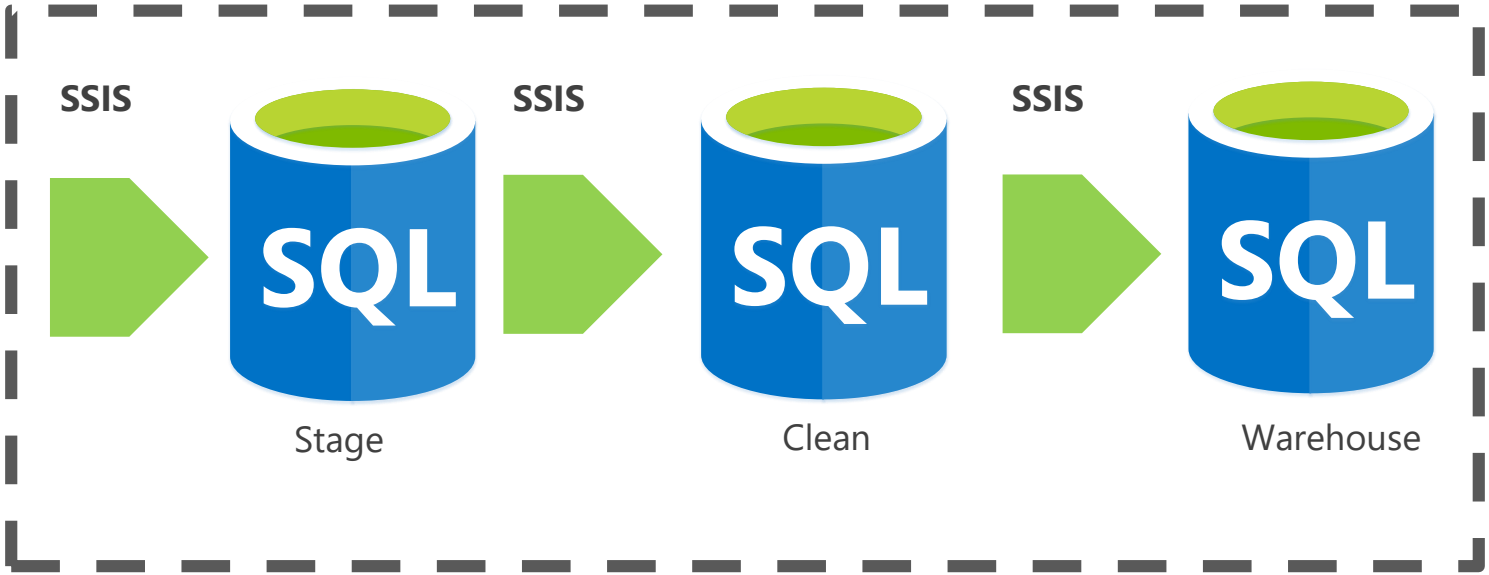
Alternative Models

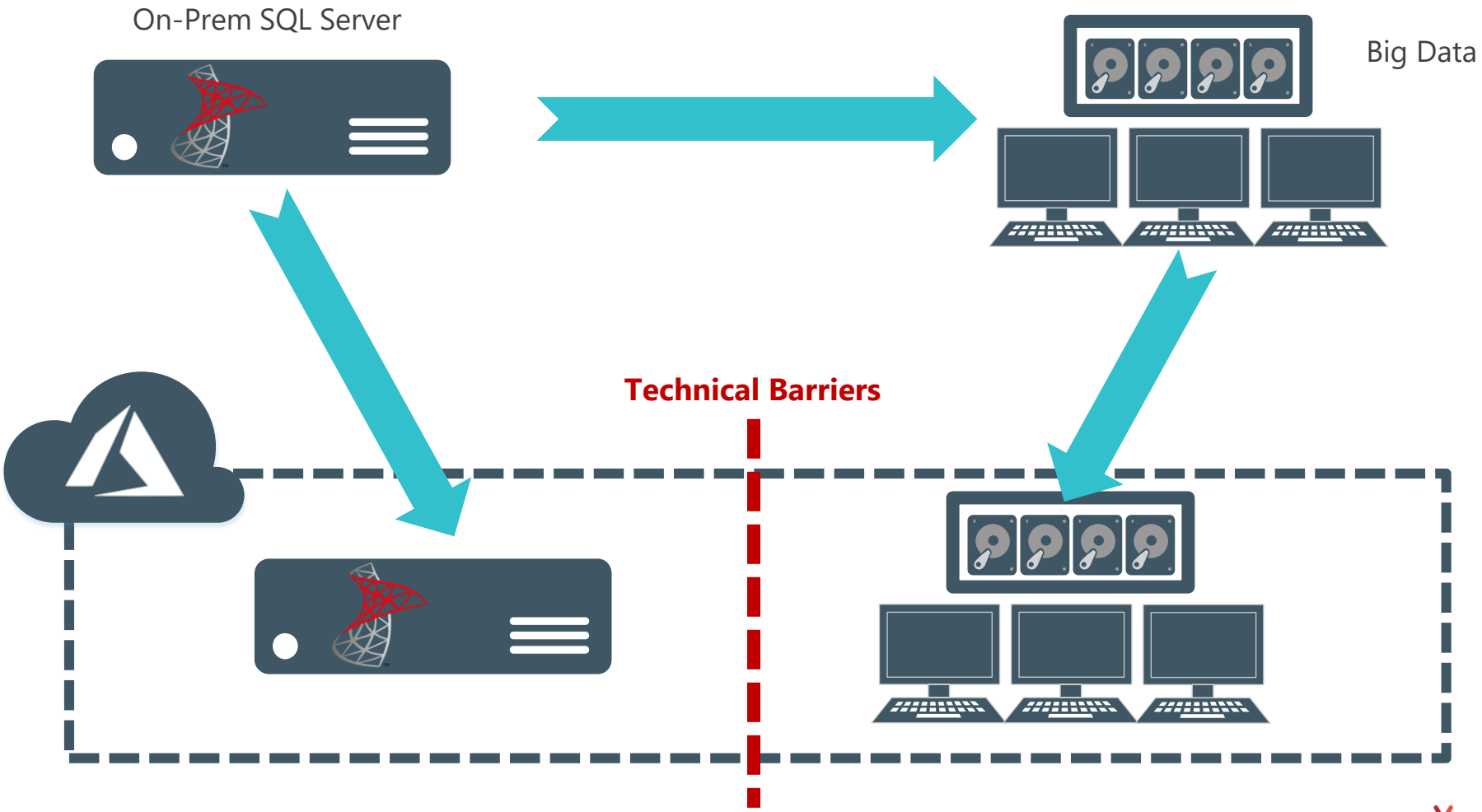
Big Data & Cloud

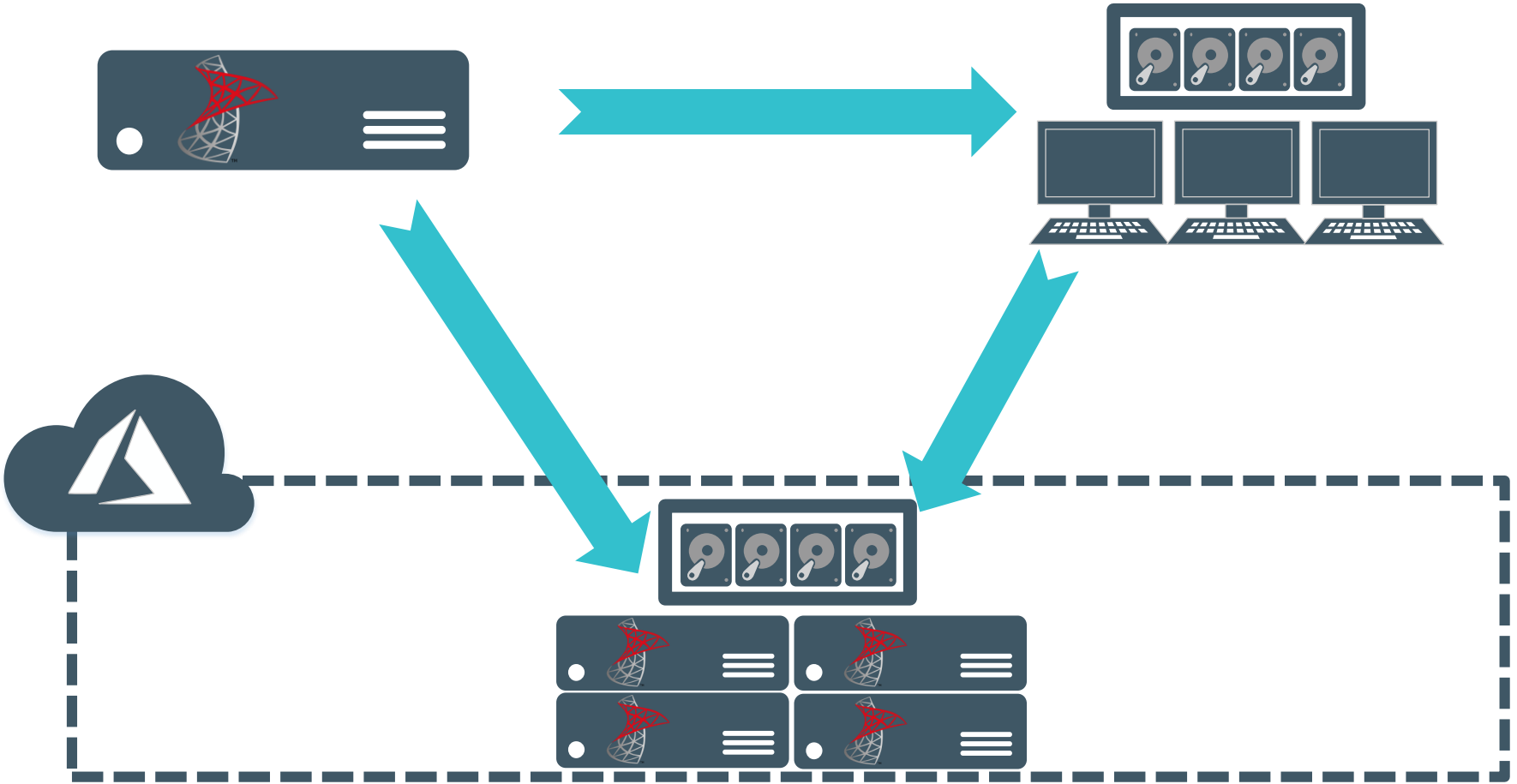
A brief history lesson

PASS

One-Box SQL BI Architecture




On-Prem SQL Server
Big Data
Technical Barriers



Modern Data Warehouse

My Life Goal:

Never to manage another Server

Modern Warehouse Wish List

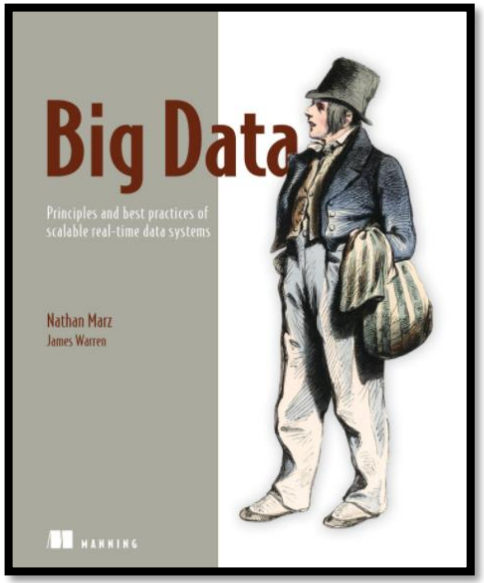
- Can Scale to huge datasets
- Linearly Scalable
- Near Real-Time
- Fault Tolerant
- Low Tech Barrier for SQL Devs
- Cost Effective

So what is LAMBDA?

PASS

Lambda Architecture

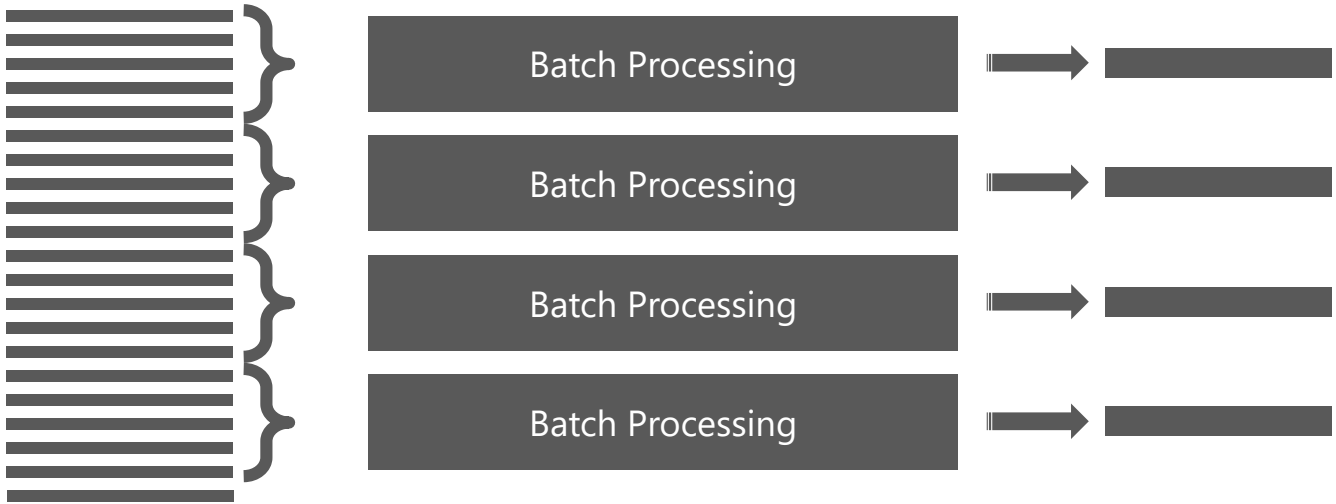
Use Batch and Stream technologies together to balance latency, throughput and fault-tolerance

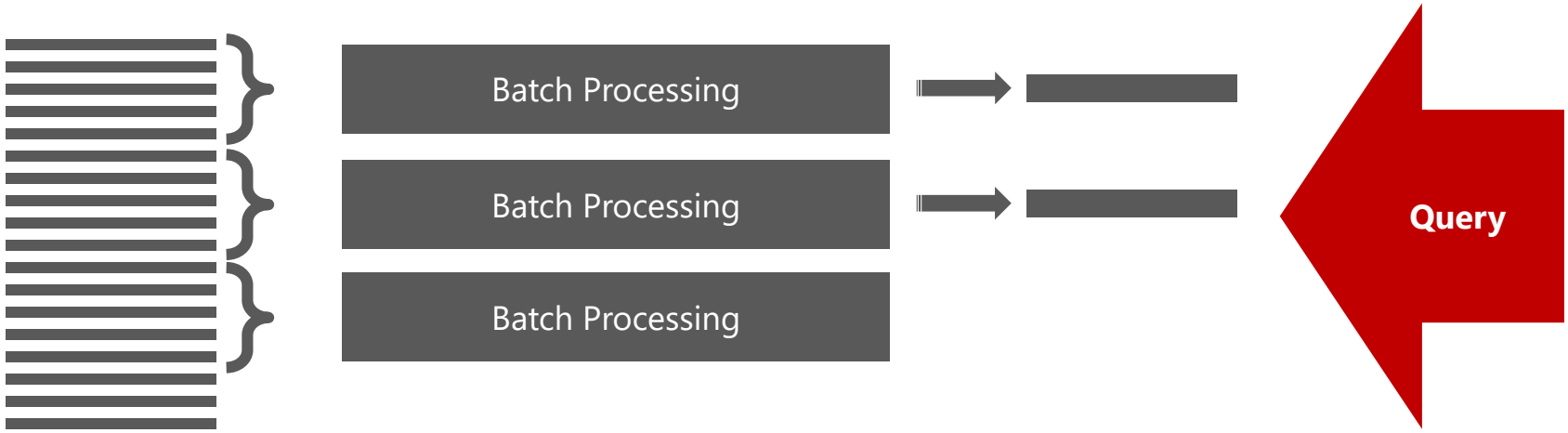


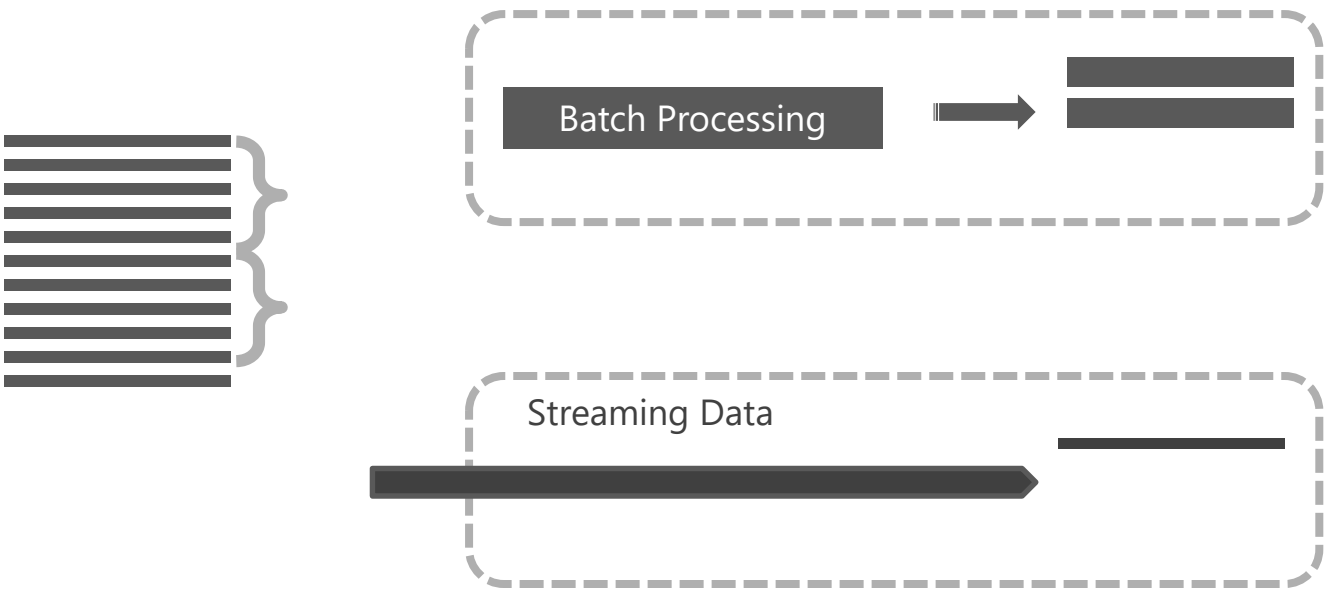
Nathan Marz &
James Warren

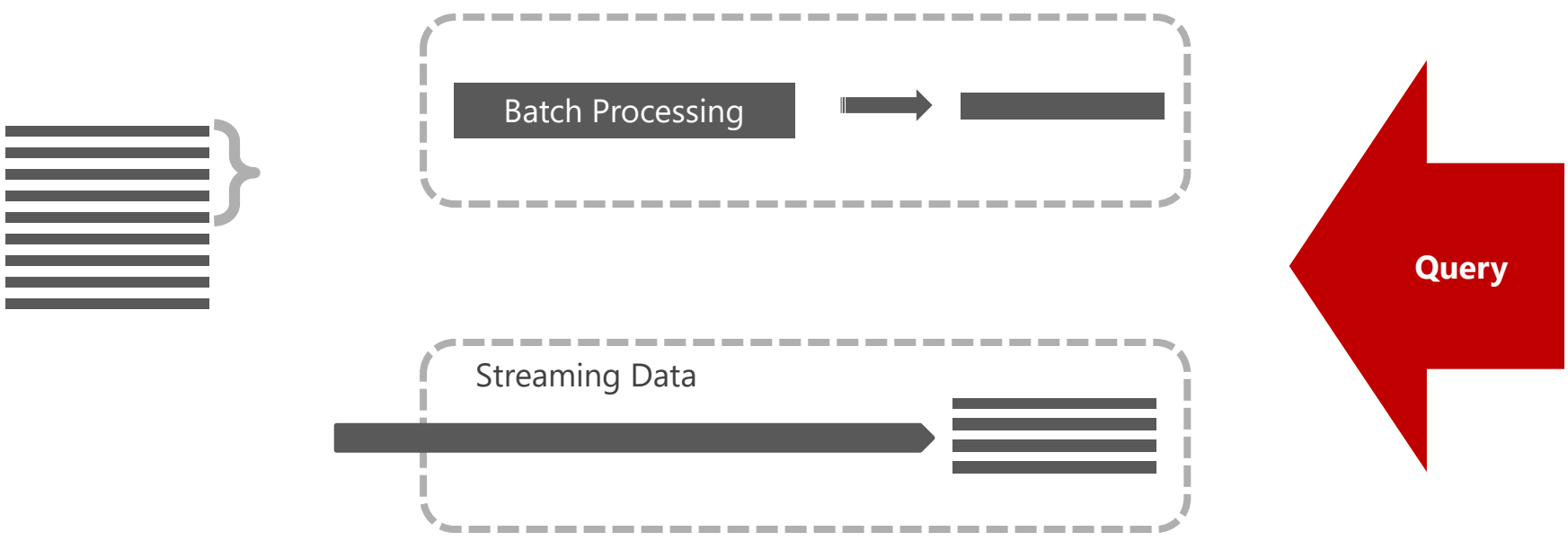
PASS

The Problem…

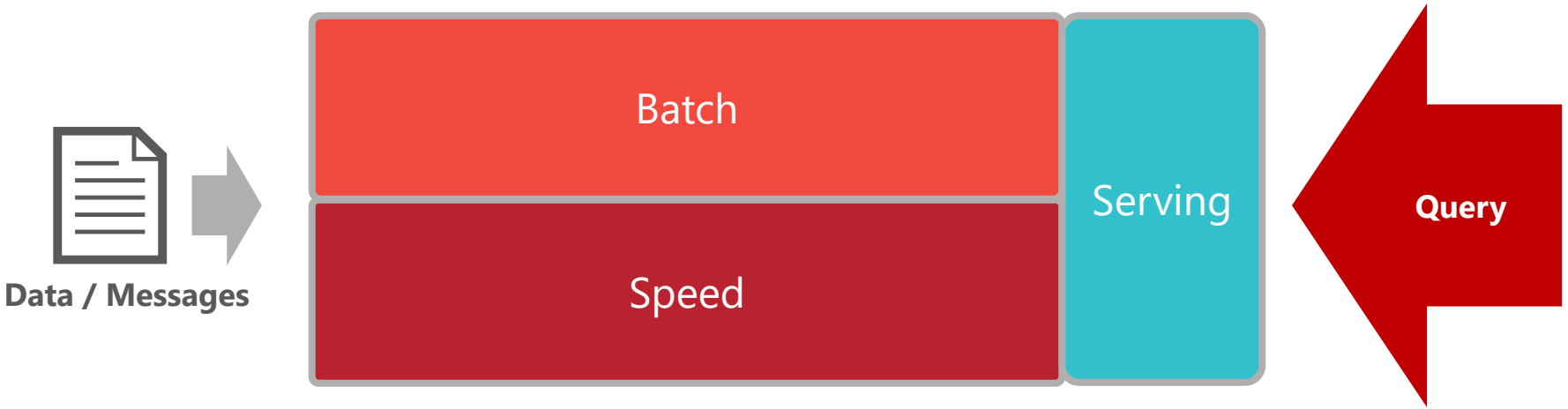

Batch Processing
Batch Processing
Batch Processing
Batch Processing


Batch Processing
Batch Processing
Batch Processing
Query

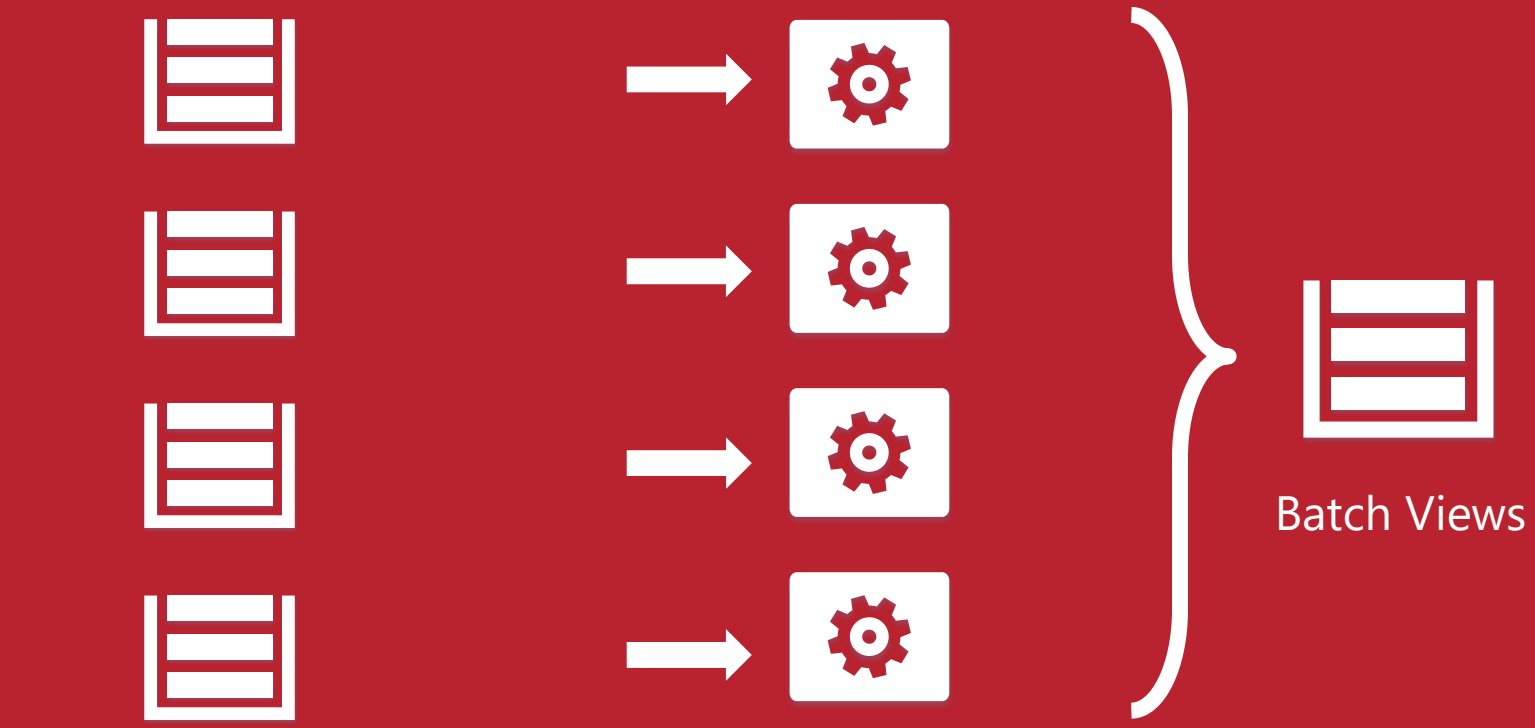

Batch Processing
Streaming Data


Batch Processing
Streaming Data
Query

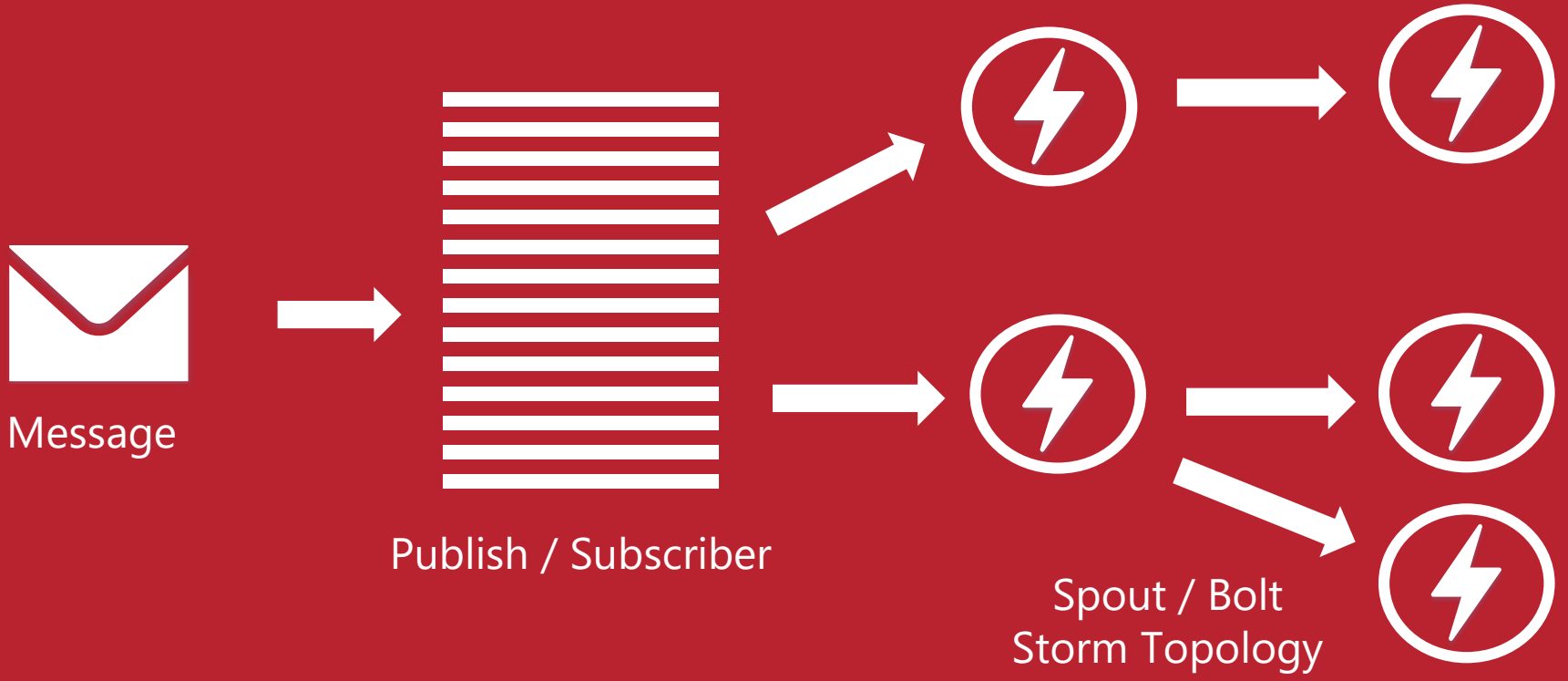
The Lambda Architecture



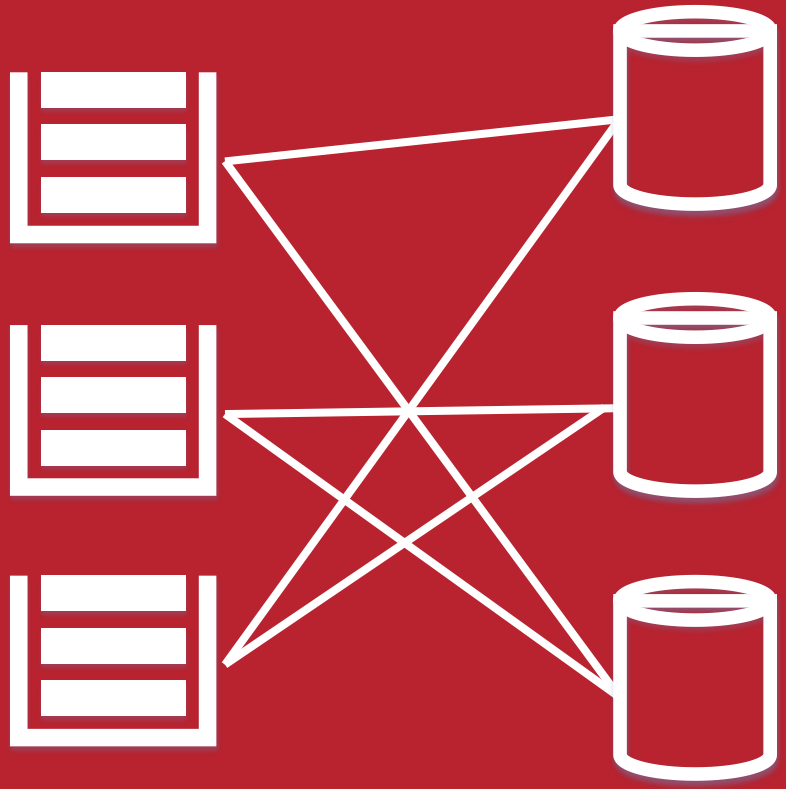
Batch Layer


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

Speed Layer



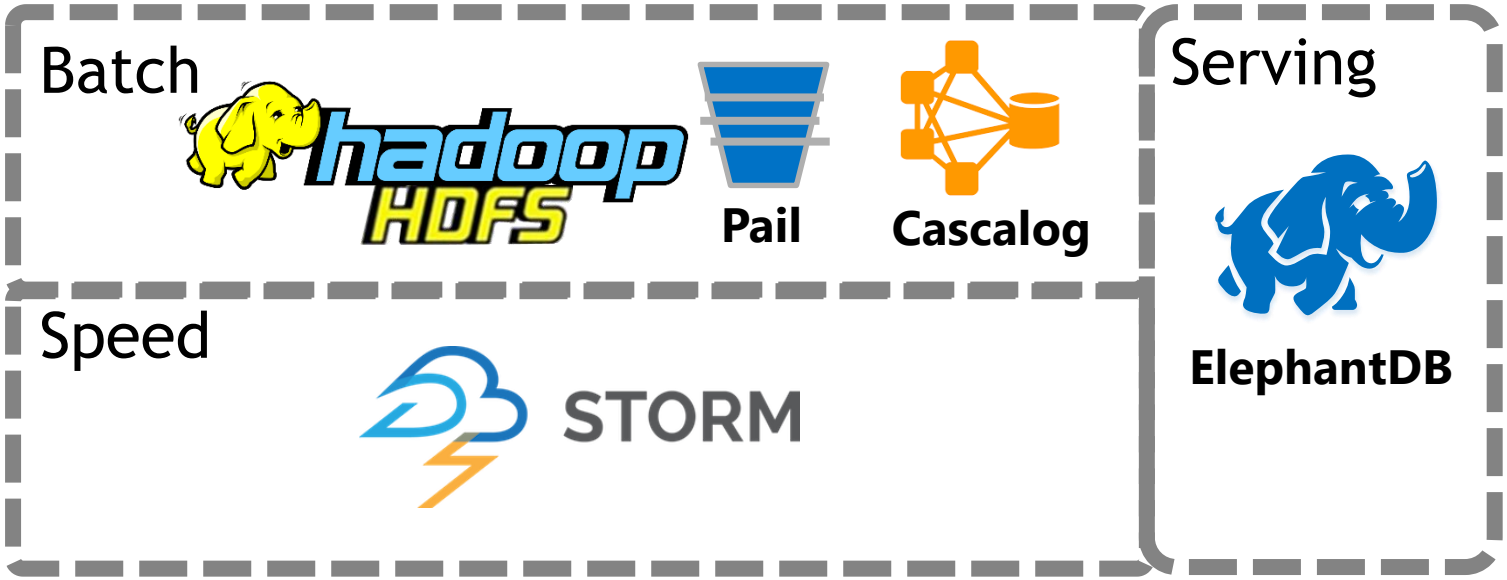
Serving Layer



Sharded Database Layer

Batch Views

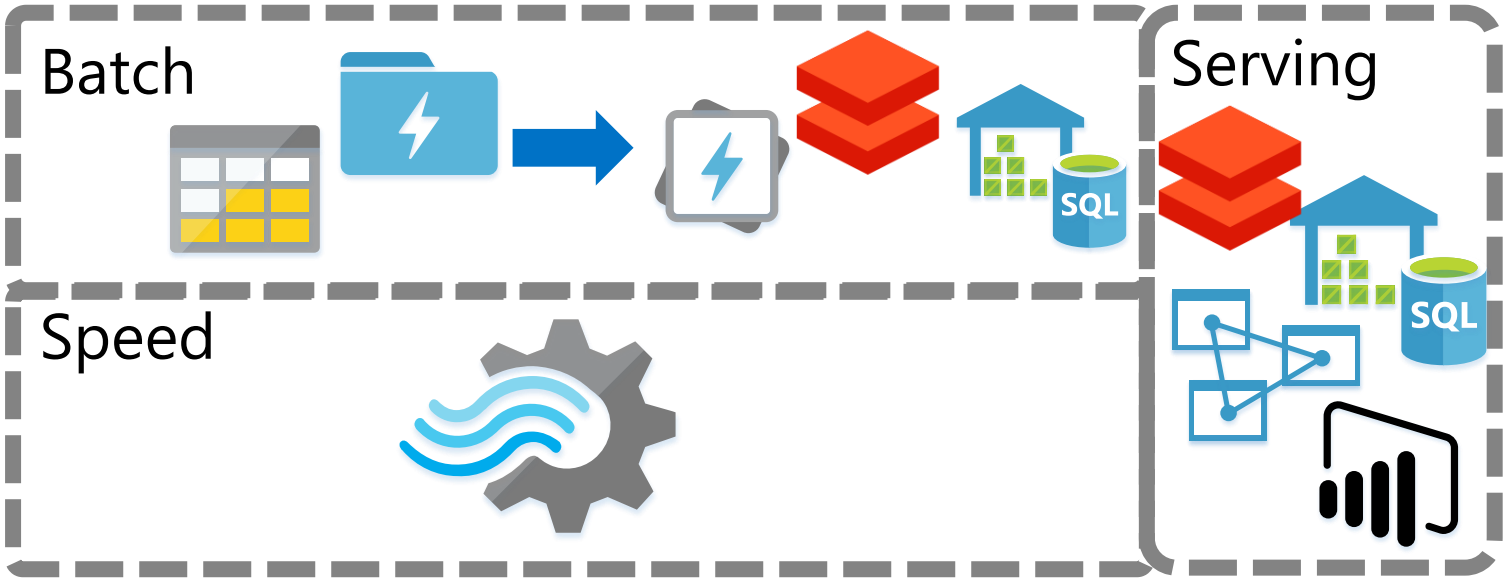
The Marz Lambda Architecture



Lambda in Azure

PASS

Applying Lambda in Azure



Batch Layer - Storage

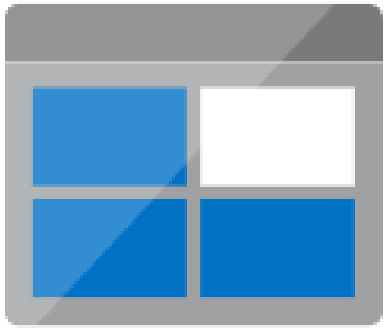
Blob Storage

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

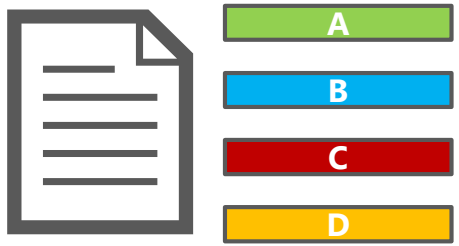
Azure Data Lake Store

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

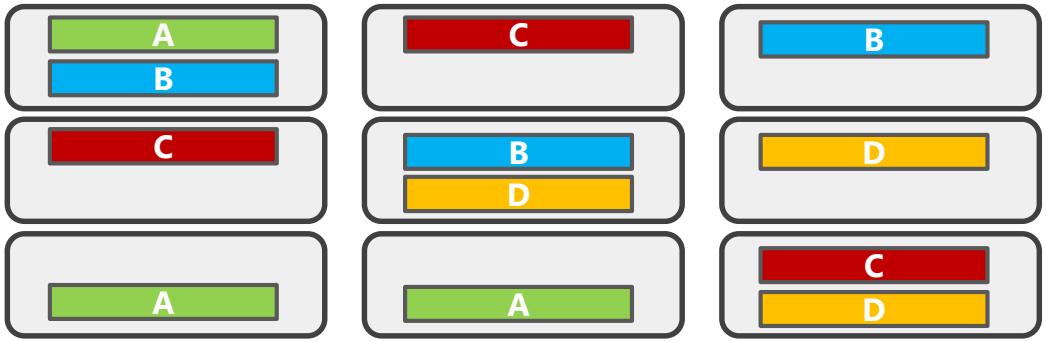
Azure Data Lake Store Gen2 (coming soon)



Data Lake Store



Parallelism



Fault Resilience

Batch Layer - Compute

Azure Data Lake Analytics

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

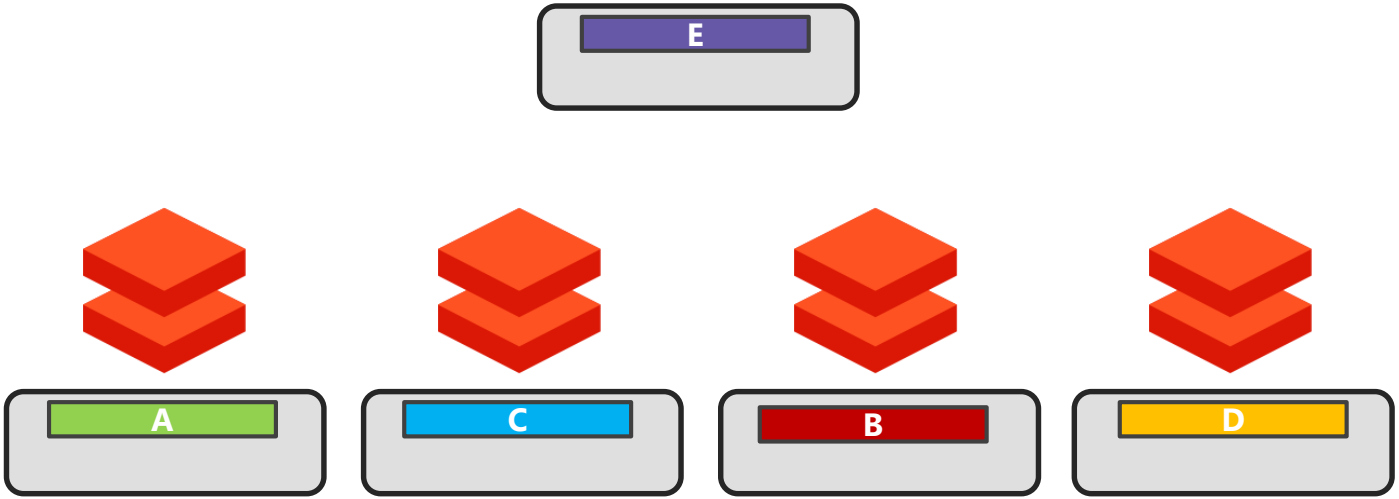
Azure DataBricks

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

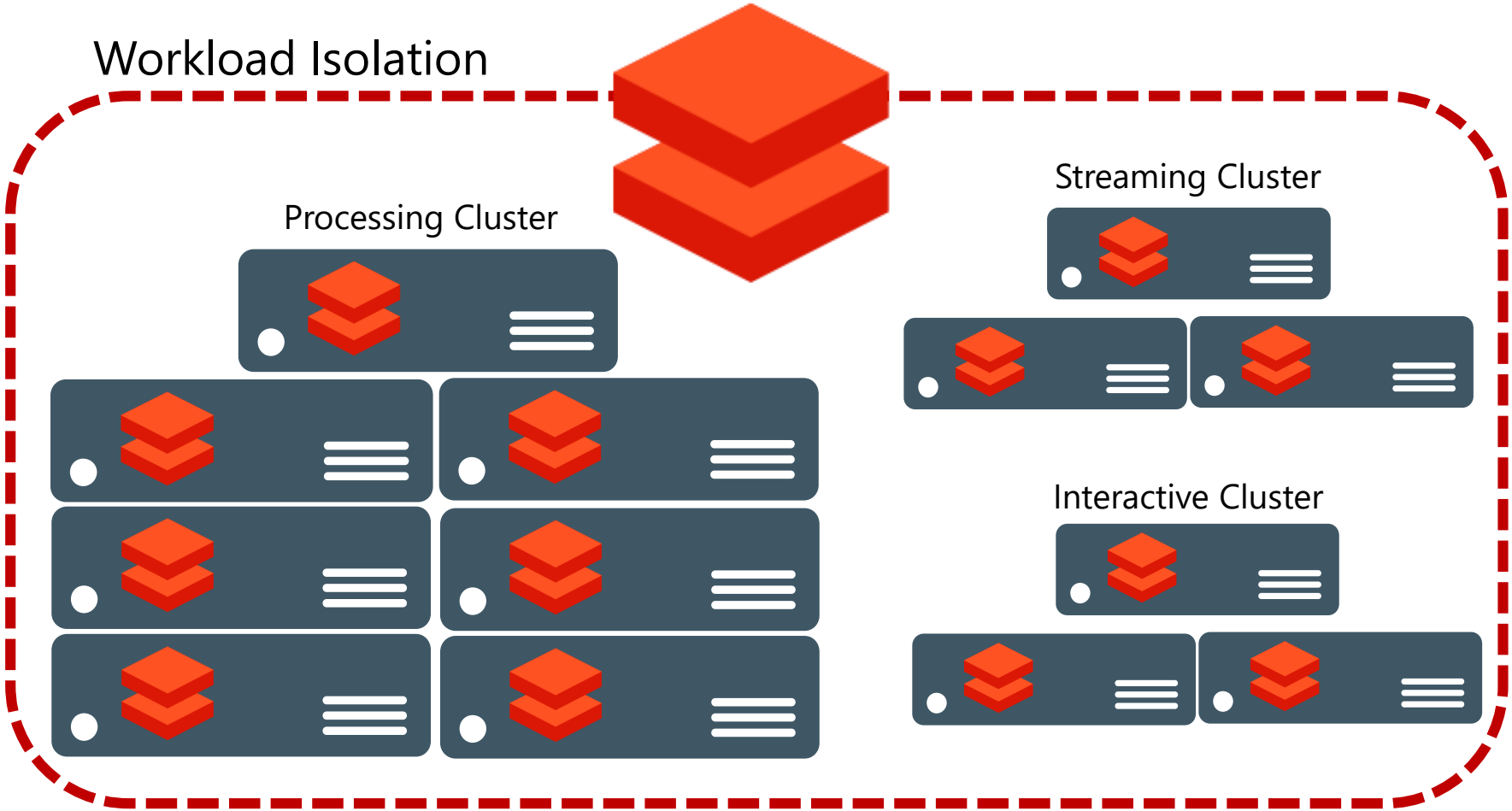
Azure SQL DataWarehouse

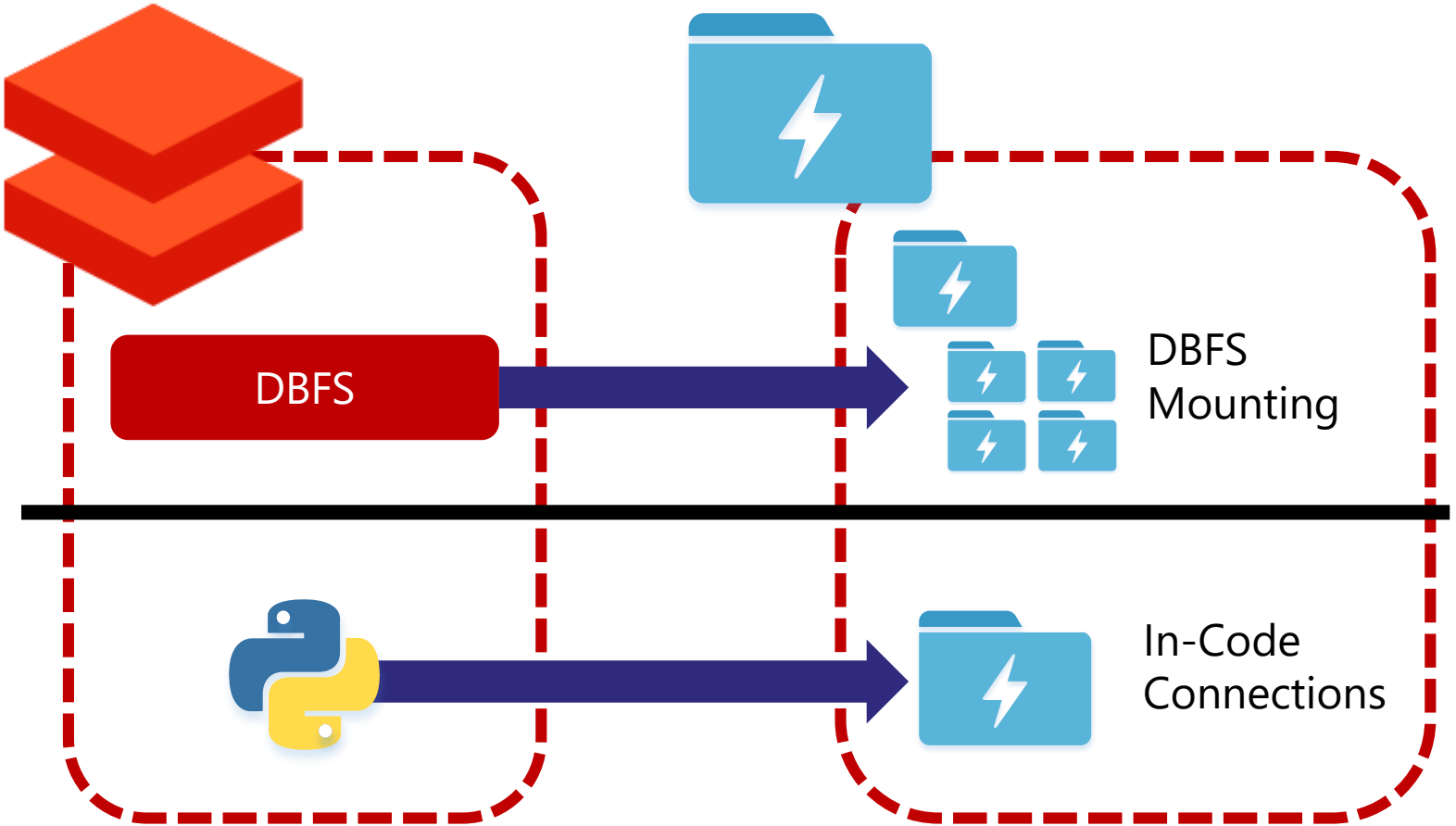
- Pay Per Hour / Node
- T-SQL
- Fully Structured
- Can use MapReduce via Polybase
- Batch or Live Query

Azure Databricks



Workload Isolation




DBFS
DBFS
Mounting
In-Code
Connections

Azure Streaming Analytics

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

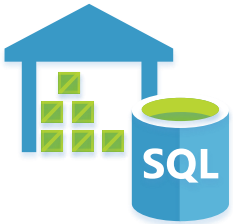
/Input/2017/06/19/0900.csv

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

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

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

Serving Layer

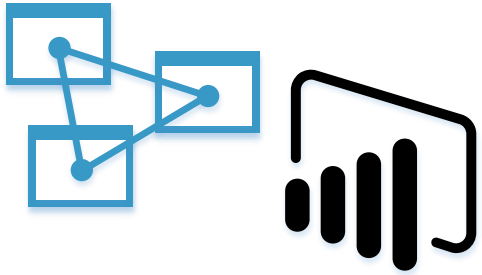


Azure SQL DataWarehouse

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

Azure DataBricks

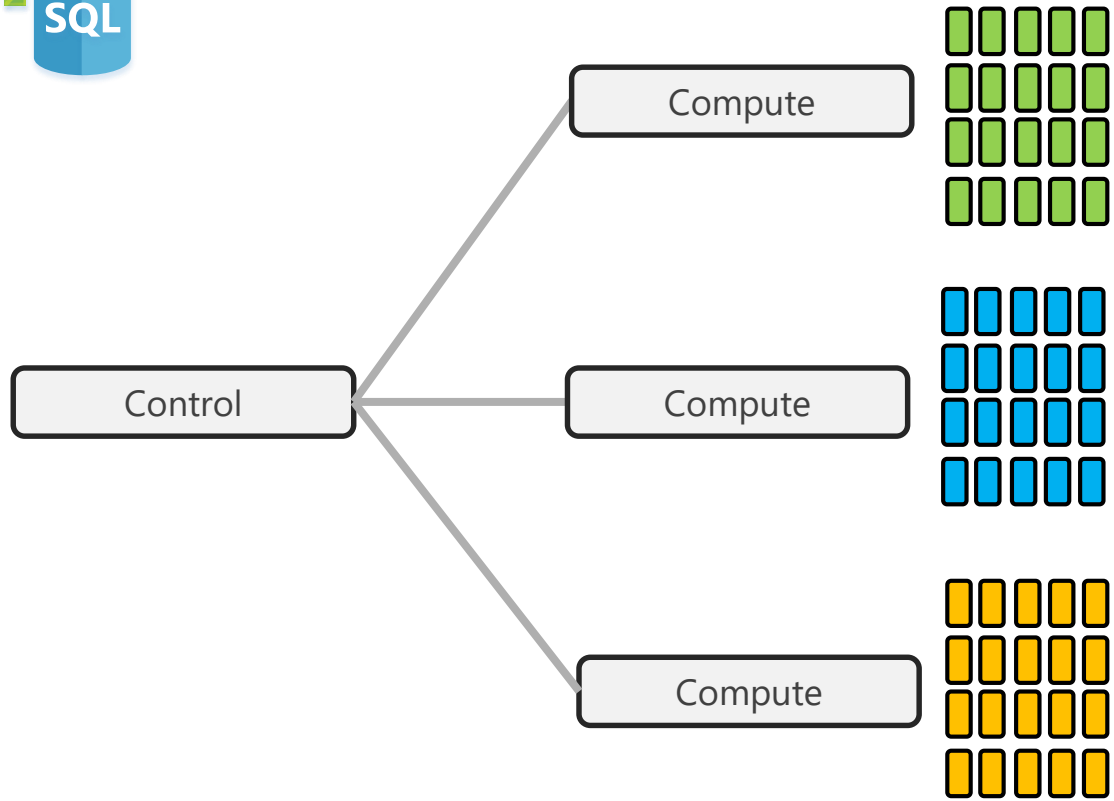
- Med Concurrency
- Direct Query over Data Lake Store
- Required Python/Scala knowledge



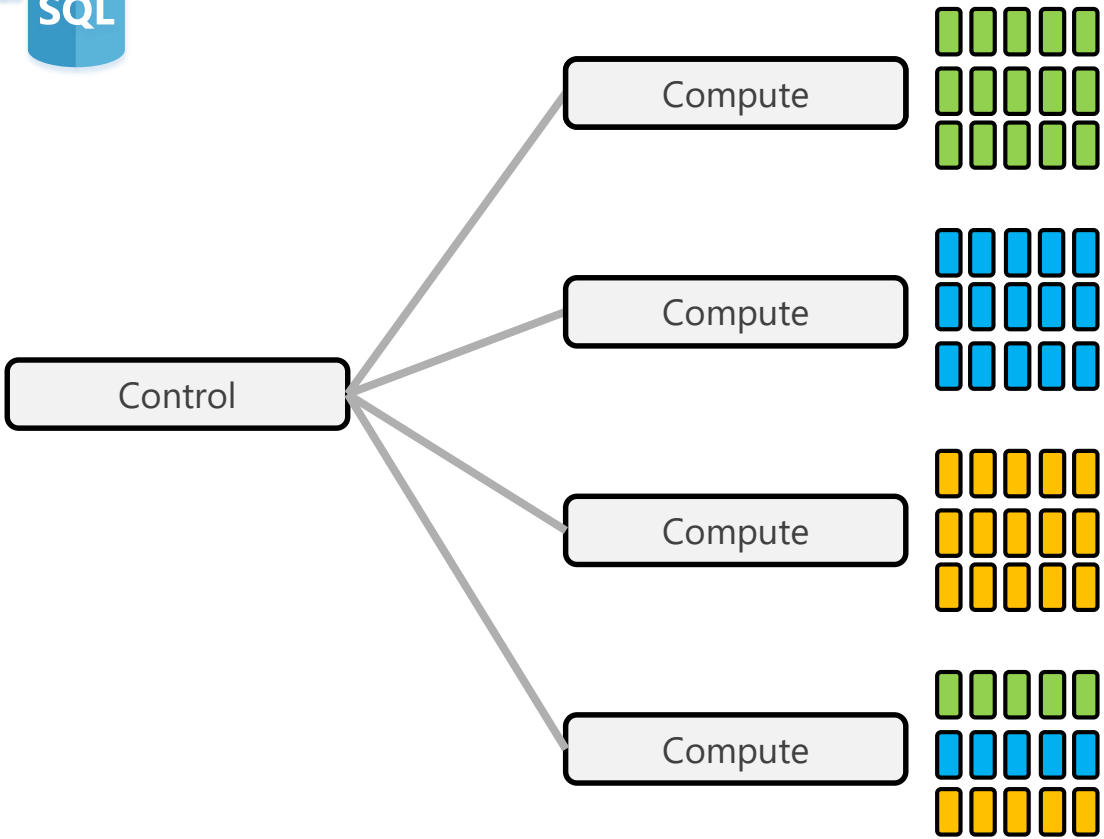
SSAS / PowerBI

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

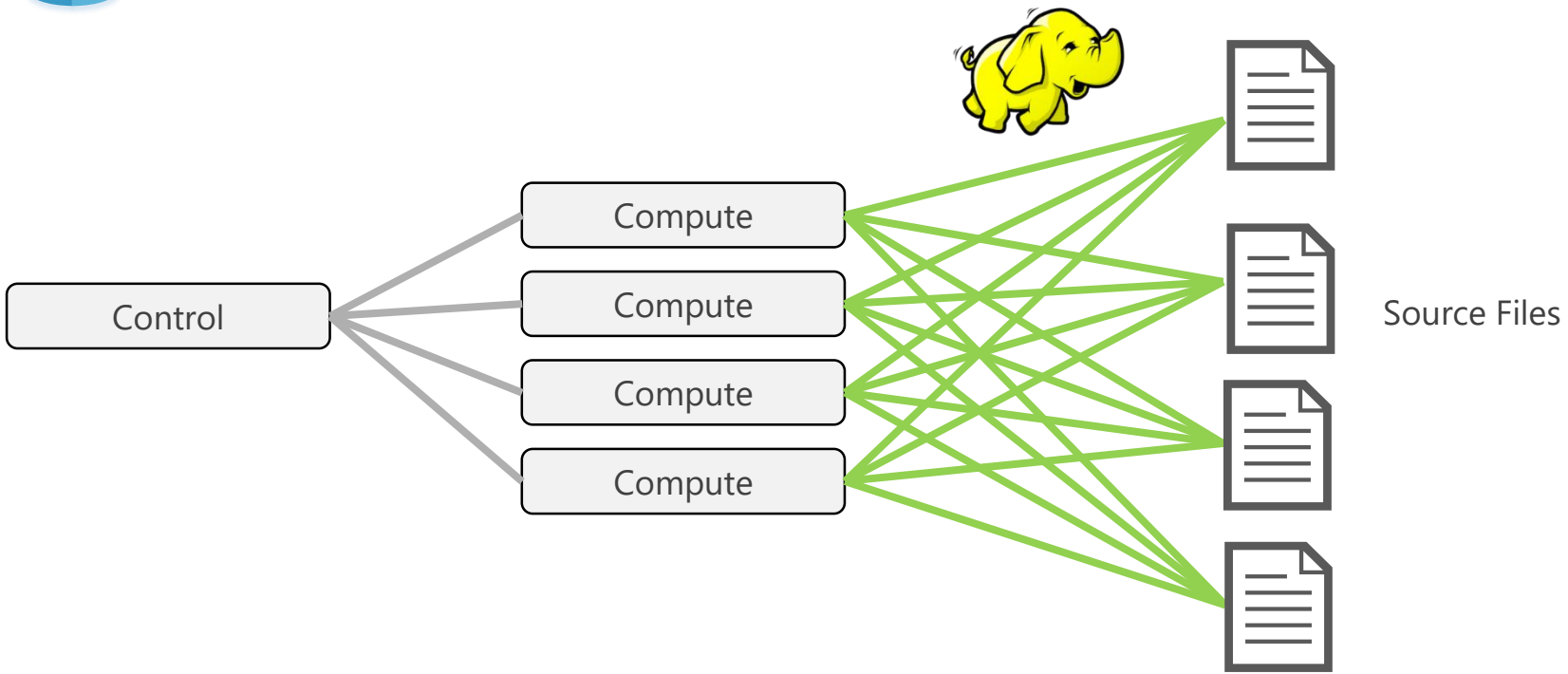
Azure SQL DataWarehouse

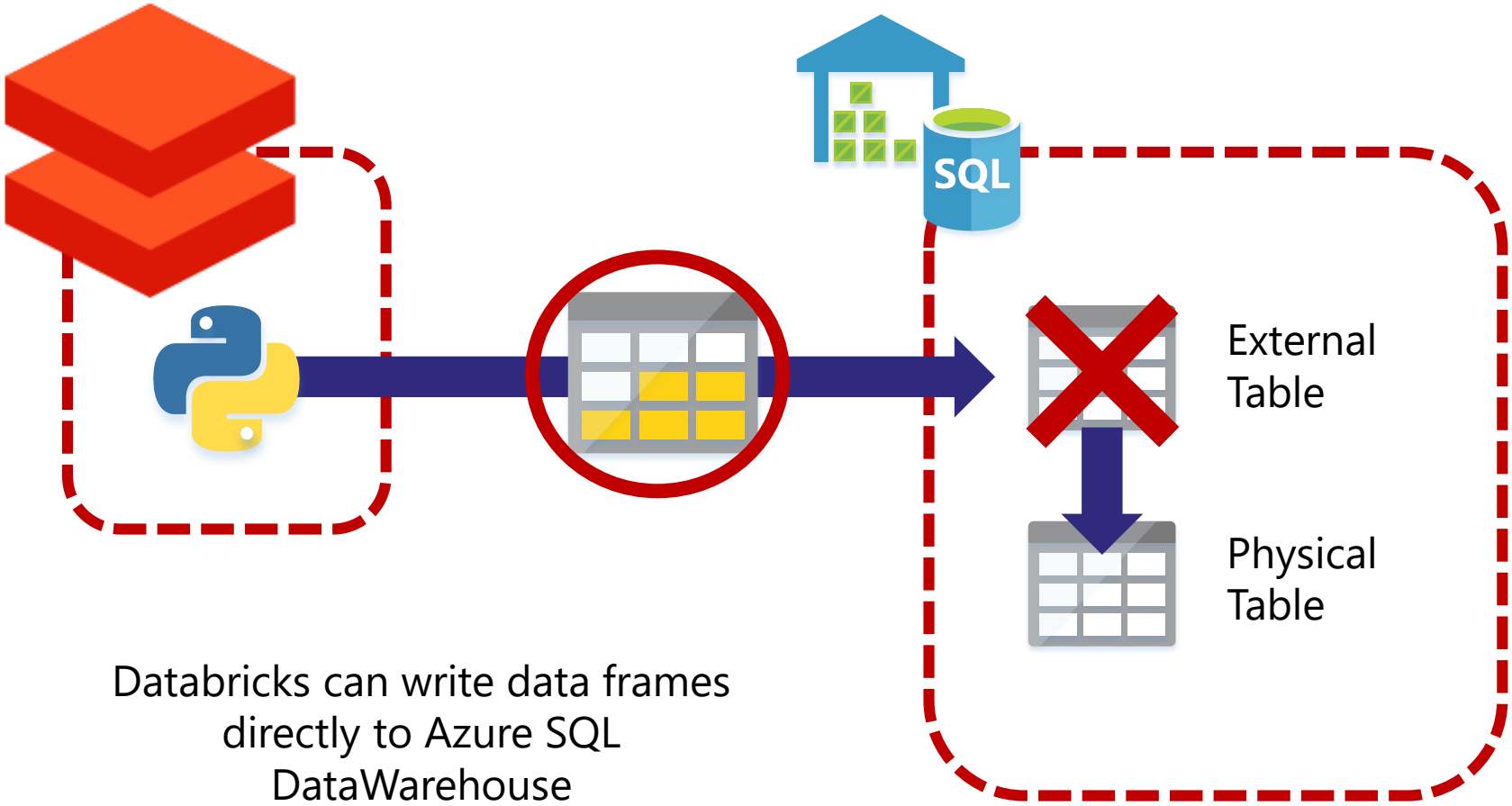


Azure SQL DataWarehouse



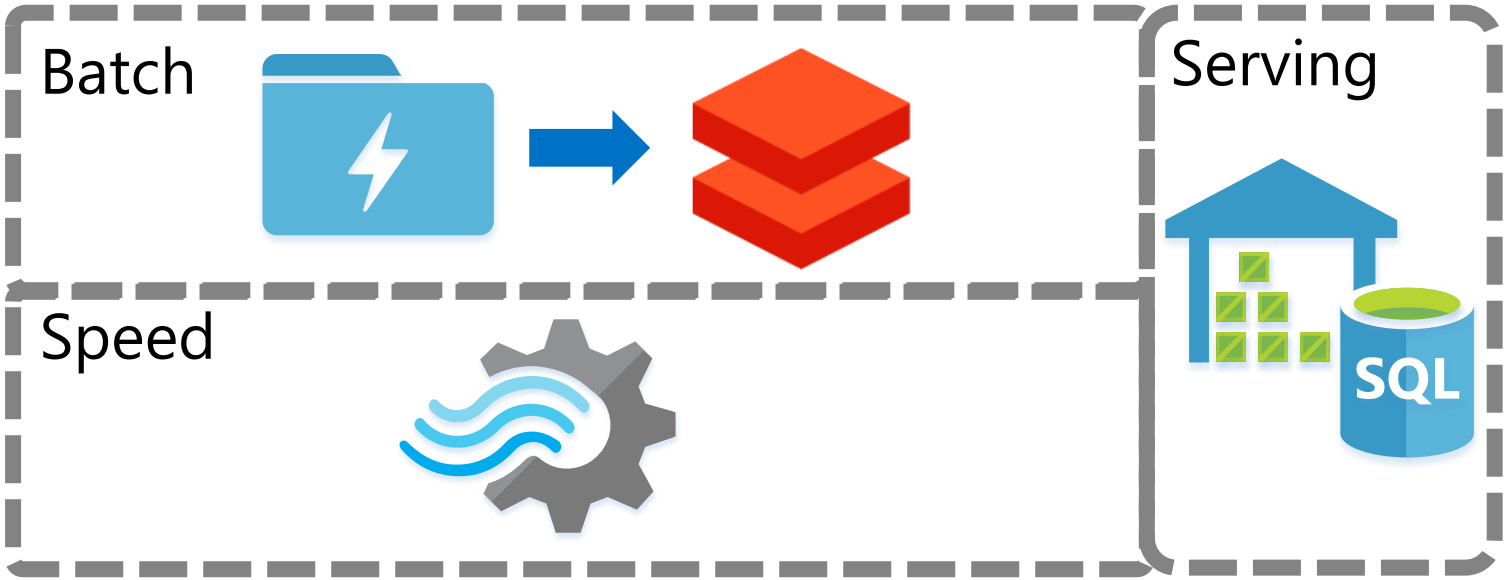
Azure SQL DataWarehouse

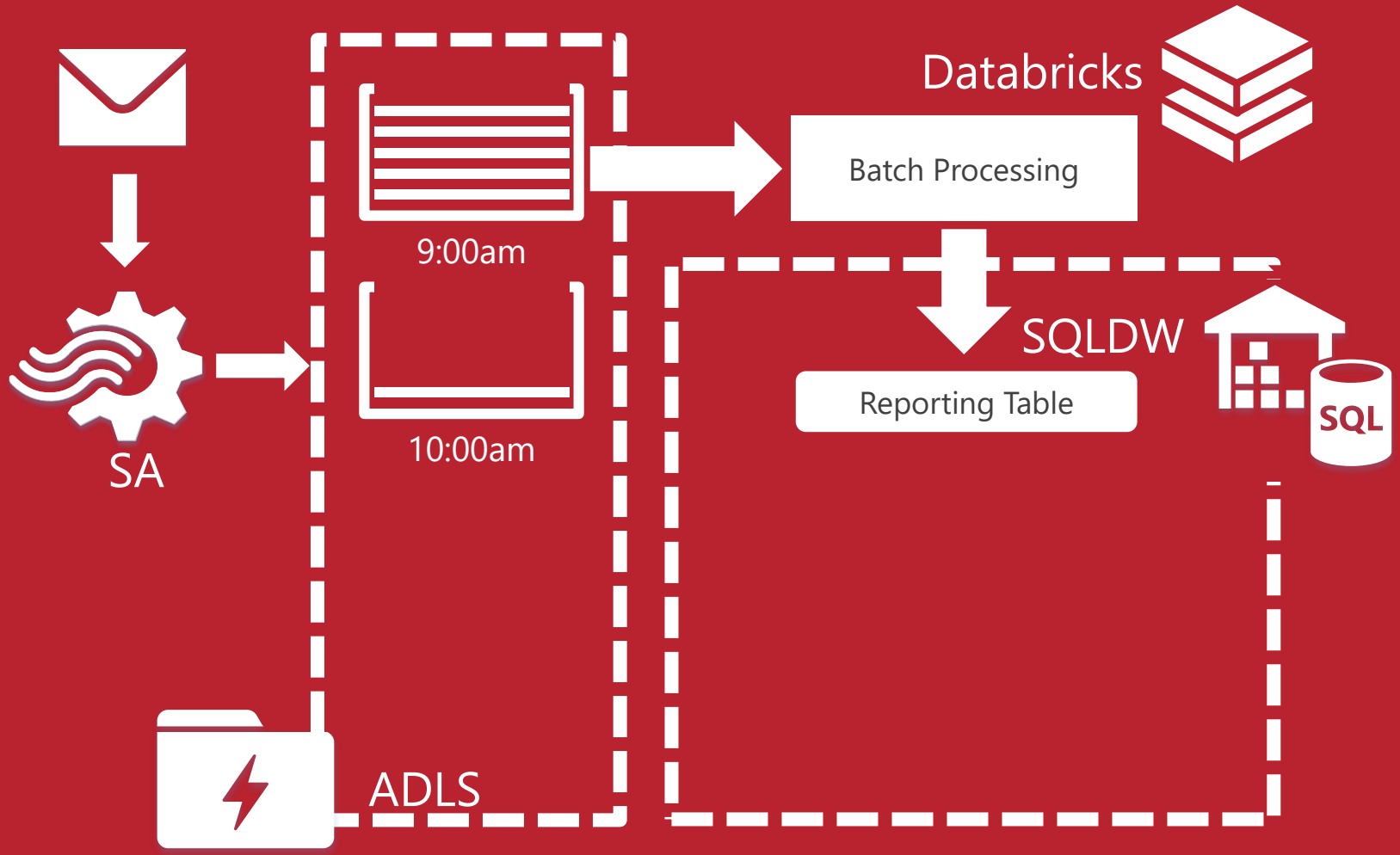


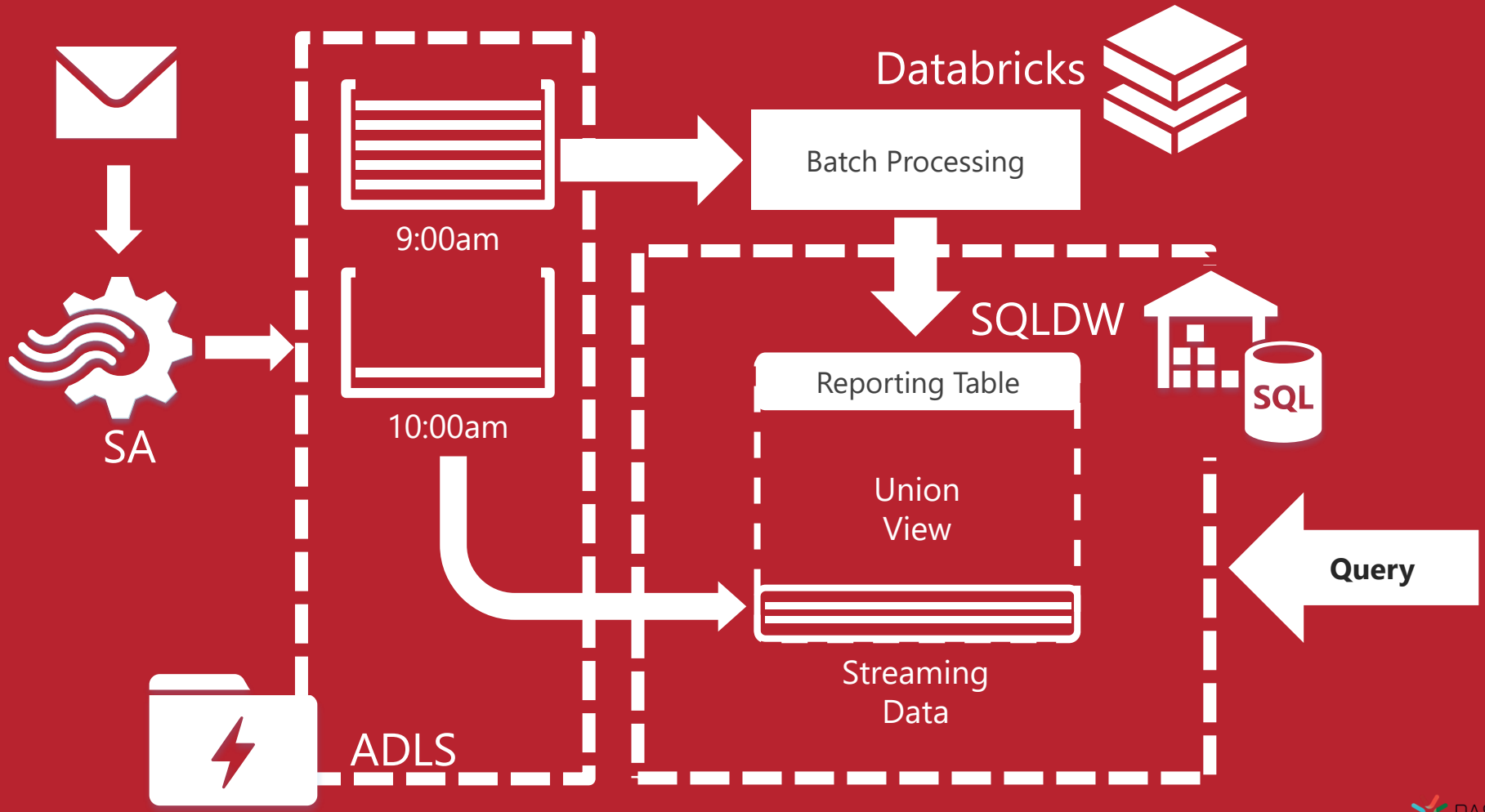


Databricks can write data frames directly to Azure SQL DataWarehouse

Applying Lambda in Azure



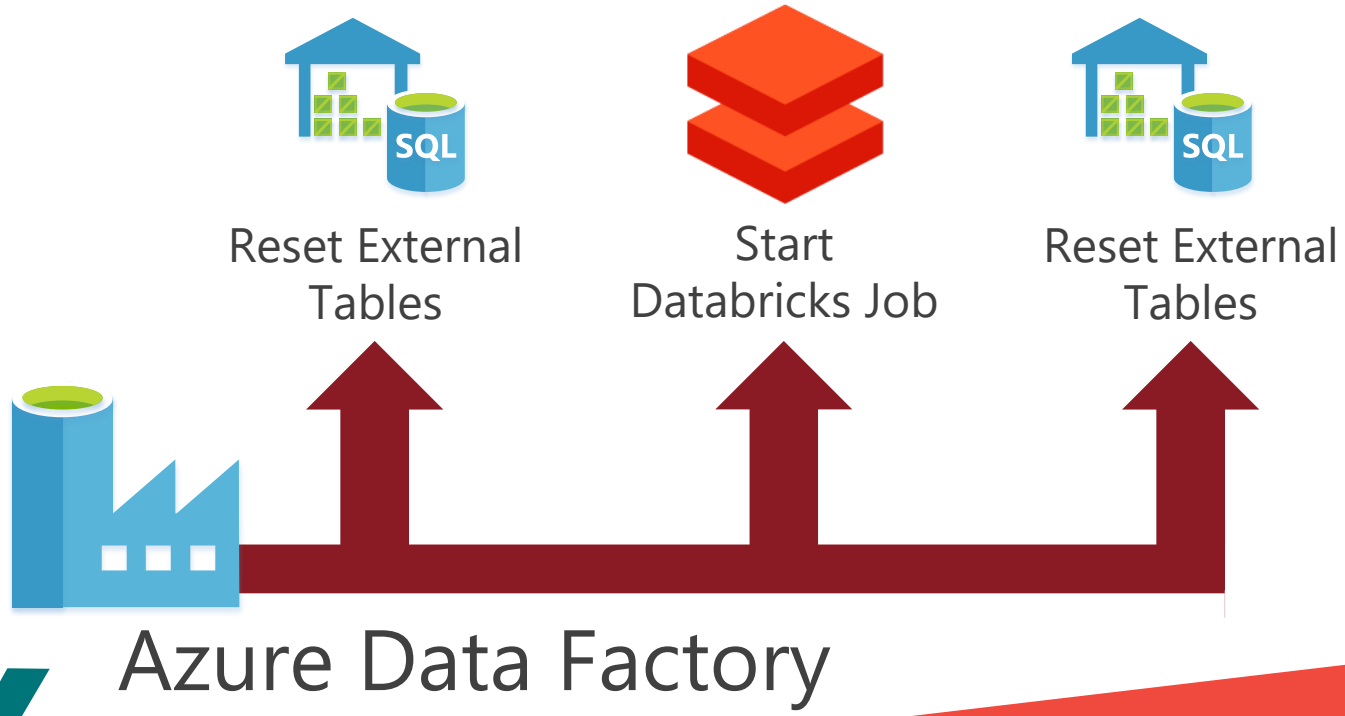

Databricks
Batch Processing
9:00am
10:00am
SA
SQLDW
Reporting Table
SQL
ADLS


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

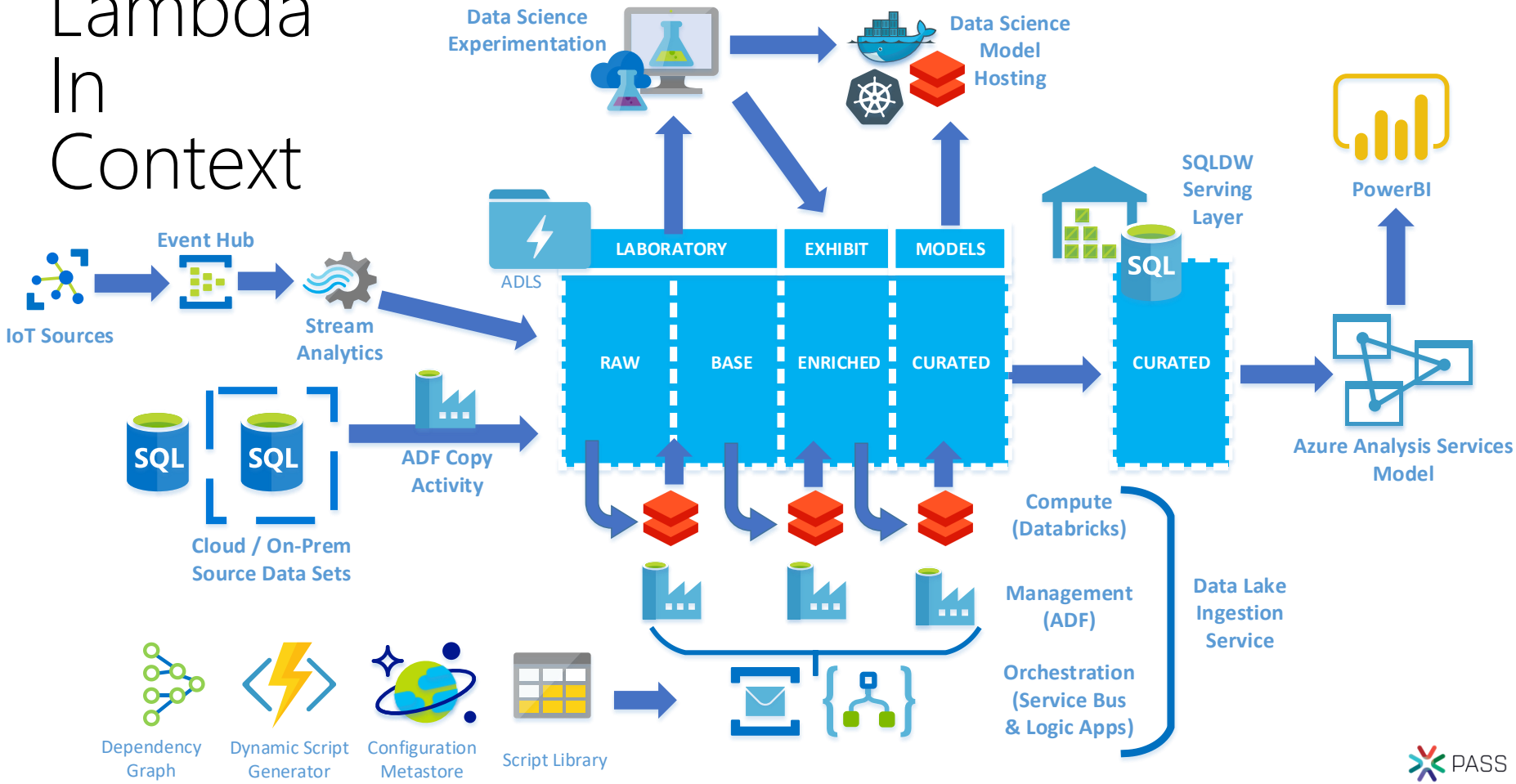
DEMO:
Lambda in Azure

PASS

Lambda Orchestration

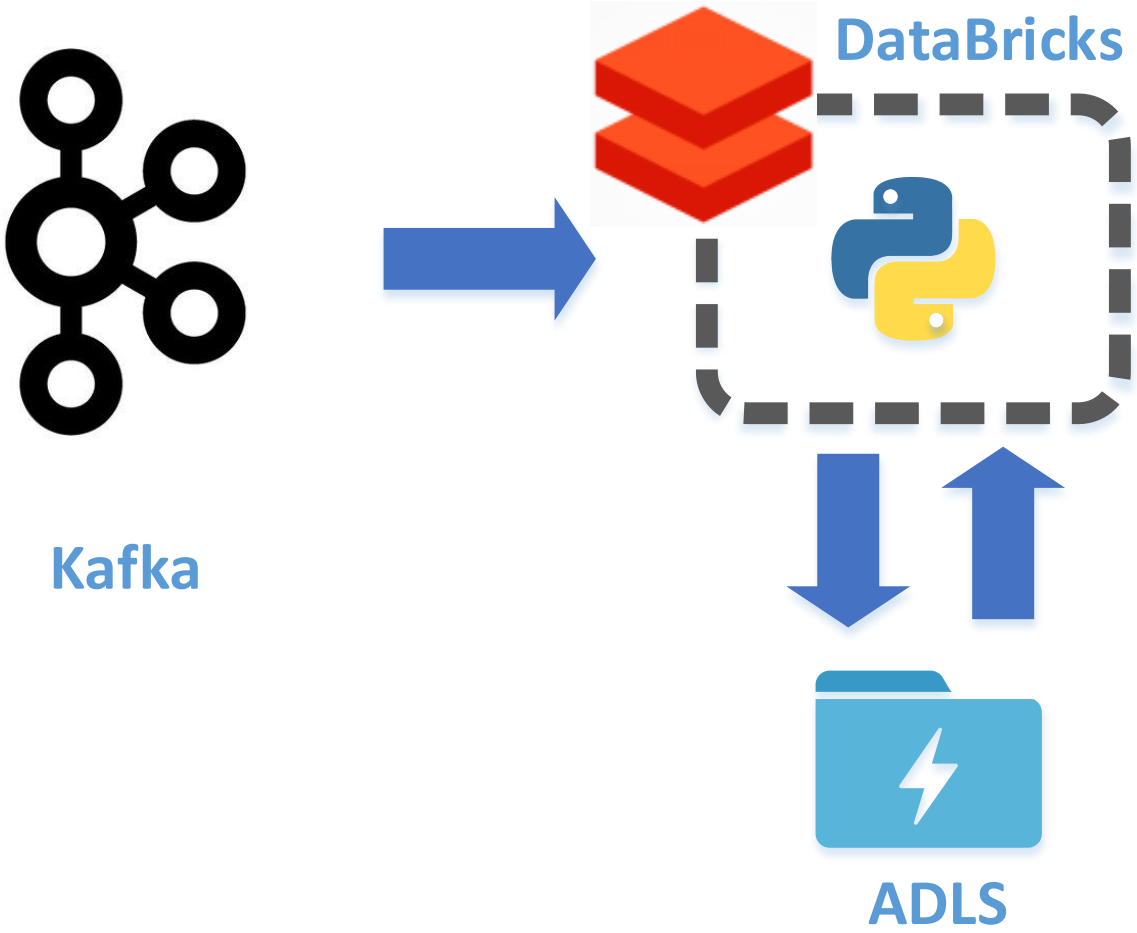


Lambda In Context

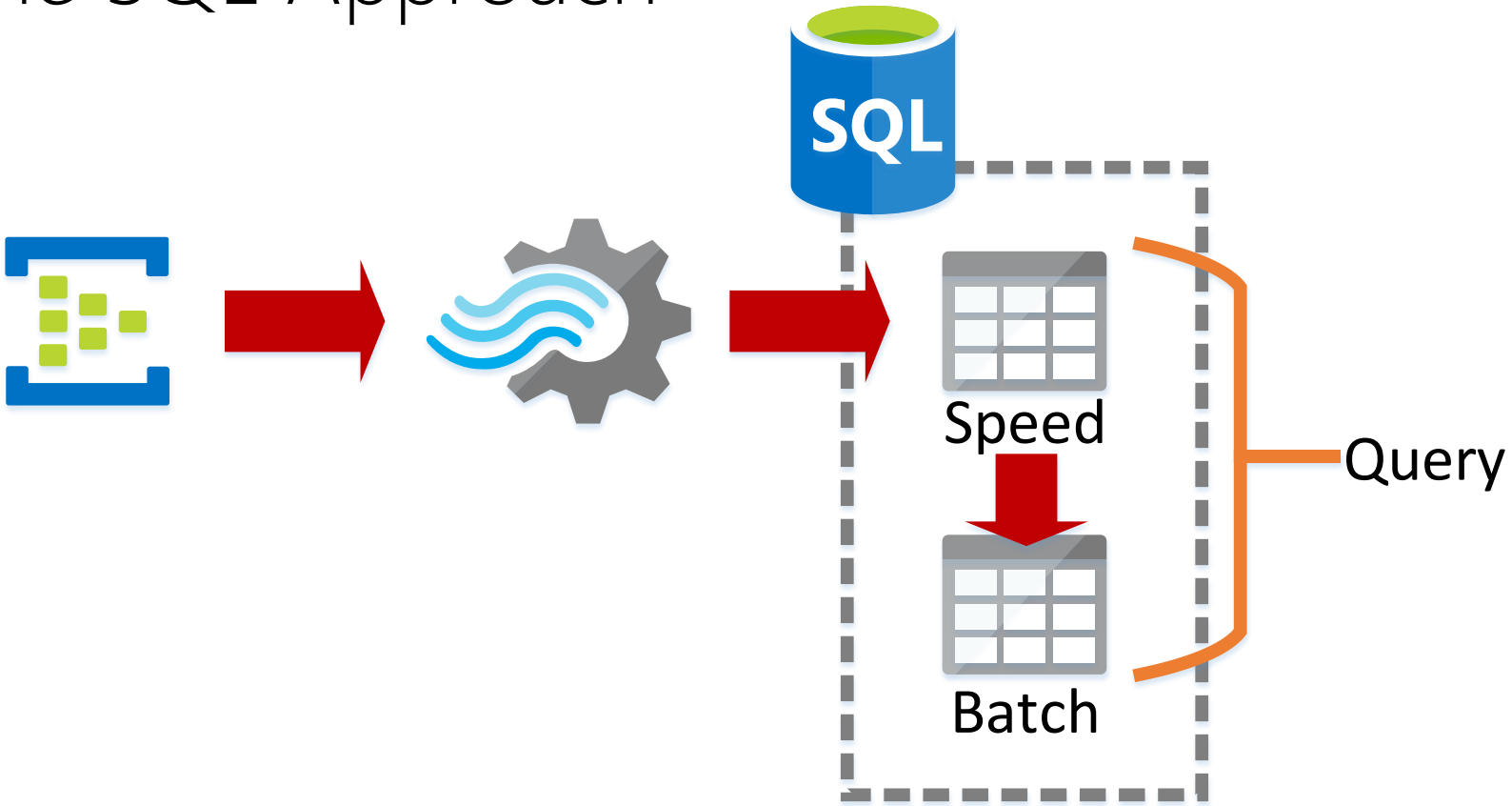


PASS

The Open-Source Approach



The SQL Approach



DEMO:
Alternative Approaches

PASS

Questions?

PASS

PASS SUMMIT2018

Thank You

Learn more from Simon Whiteley

@MrSiWhiteley

saw@adatis.co.uk