Azure Streaming Data
Services, Architectures, Patterns

Rolf Tesmer
Cloud Solution Architect (Data | AI)
Melbourne, Australia
Linked In: https://www.linkedin.com/in/rolftesmer/
Blog: https://mrfoxsql.com.au/

Rivaaj Jumna
Cloud Solution Architect (Data | AI)
Perth, Australia
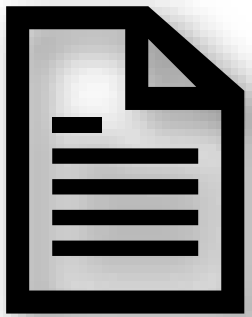**Linked In:** https://www.linkedin.com/in/rivaaj/

Our Five Assumptions About *YOU*...

1. Have **awareness** of **Azure Public Cloud**

2. Have a **working background** with **data**

3. Are **curious** about **Streaming Systems**

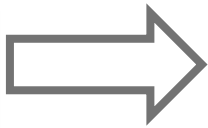4. Love **solving data problems**

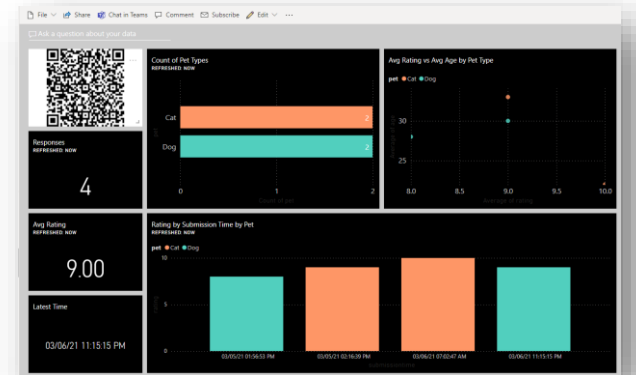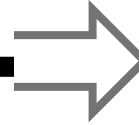5. **You will rate our session 10 / 10** ☺

INTRODUCTION

Let's take a quick poll...

Microsoft Forms

Power Automate

Power BI Dashboard

What do we mean by **real-time** or **latency** (5ms, 5sec, 5mins, 5hours, ?)

What exactly is "Streaming Data" anyway?

- Different definitions **depending on which vendor you talk to**
- **Microsoft** has **no formal definition** per se
- *But...* a couple of definitions that **we like**...

"***data streams*** *are formed when* ***multiple individual 'fit for purpose' services*** *are aligned in* ***sequences*** *that perform a set of specific* ***targeted actions*** *on* **unbounded data** *that is* ***typically in transit*** *- for the key outcome of* ***making a business decision***."

"*a* ***data streaming pipeline*** *is a set of* ***data processing elements connected in series*** *where the* ***output of one*** *element is the* ***input of the next*** *one. Elements of a* ***data stream*** *are often* ***executed in parallel*** *or in* ***time-sliced fashion***"

"*Life doesn't happen in* ***batches****...it happens in* ***heartbeats***"

Real time processing - Azure Architecture Center | Microsoft Docs

2

AZURE SERVCES FOR DATA STREAMING

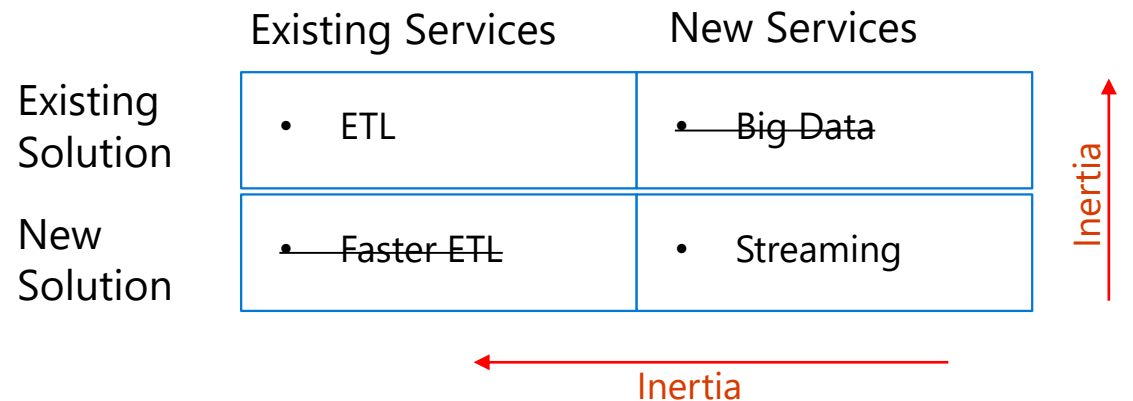https://imgflip.com/memetemplate/36656000/Drinking-from-Fire-hose

Streaming Solution Observations

- A streaming solution is a layering of architectures

- Not faster ETL

- Optimise locally, measure globally

Performance	Cost	Change Management
Data Integration	App Integration	Process Integration
New development	Enhancements	Research spikes

- -

	Existing Services	New Services
Existing Solution	• ETL	• ~~Big Data~~
New Solution	• ~~Faster ETL~~	• Streaming

Inertia

Inertia

Streaming Blueprint
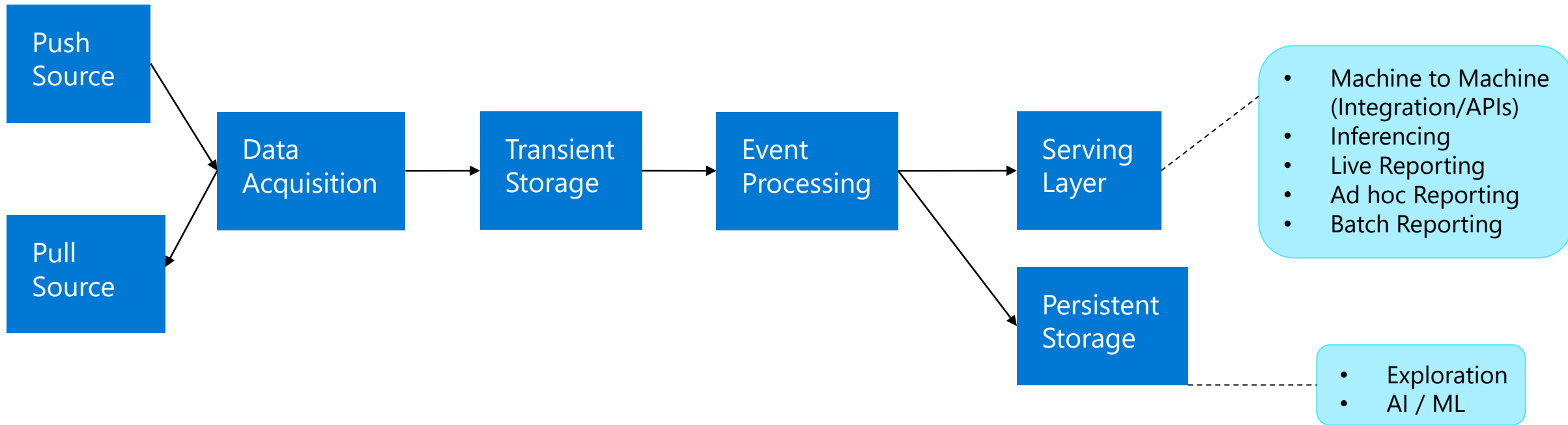
Data Acquisition

Data Acquisition (What)
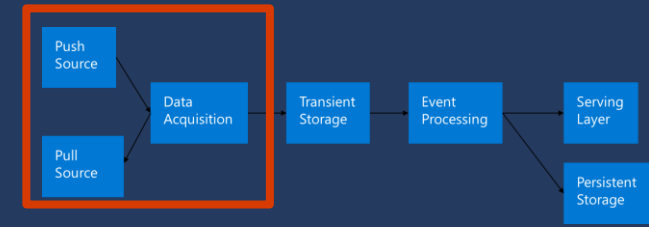

- Gaining access and transporting data

- Often intangible, viewed as a property of the data source

- Real world constraints

  - Potential to significantly alter architecture and implementation

- Movement of data is well understood, orchestration less so

Data Acquisition (How)

- Push
  - Source triggered

  - By Data source
    - User interactions
    - Devices (IoT)
    - Storage engine

  - By intermediary
    - Data Pump
    - ETL tools
    - Protocol translation

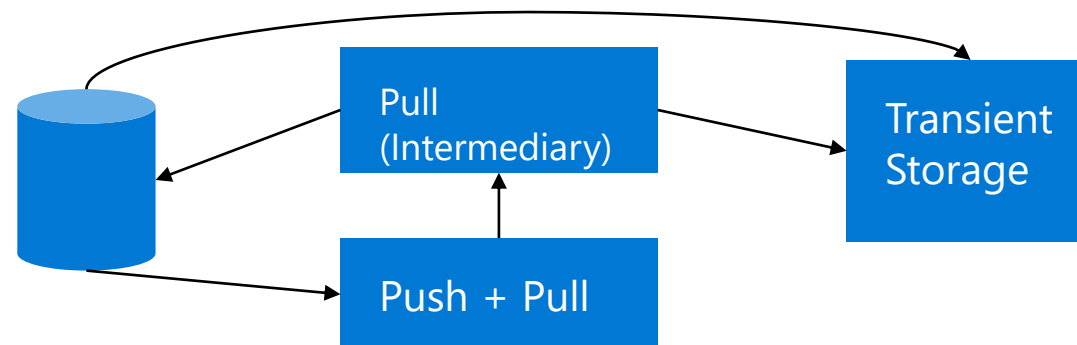| Data Acquisition	→	Transient Storage
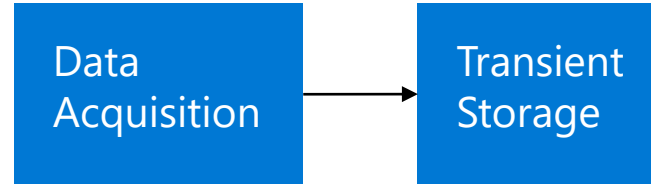
- Push + Pull
  - Event by source
  - Lookup by intermediary

- Pull
  - External trigger

  - By intermediary
    - Polling
    - ETL tools

Services for Data Acquisition

- Push
  - Event Grid (Message delivery)
  - Cosmos Db (Change Feed)
  - Azure IoT Edge (Protocol translation)
  - Logic Apps (Trigger and publish)
  - Azure Functions (Trigger and output)

- Pull
  - CDC (Azure SQL Database, PostgreSQL)
  - Logic Apps (Connect and poll)
  - Data Factory (Connect and poll)
  - Azure IoT Edge (Edge compute)

- Push + Pull
  - Azure Storage (Pass by reference)
  - Event Grid (Event publishing)
  - Azure Durable Functions (Trigger and long running)

Transient Storage

Transient Storage (What)
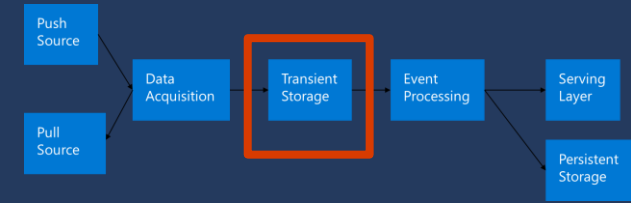
- Enable delivery of messages from producers to consumers
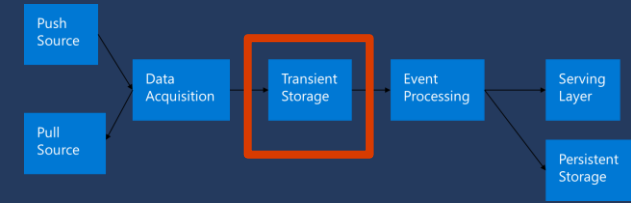- Producers append, processors consume in-order
- Decouple rate of data arrival and rate of consumption
- Scalable, transient storage or buffer
- Not always a single service

Transient Storage (How)

| Producer | Data Acquisition | → | Transient Storage | → | Event Processing | Consumer |

- ## Producer Considerations
  - Supported protocols
  - Throttling
  - Payload size (Event vs Message)
  - Quality of service
  - Offline scenarios

- ## Consumer Considerations
  - Delivery guarantee
  - Ordering
  - Poison messages
  - Delivery patterns

Services for Transient Storage

- Typically Azure messaging services

Service	Common Use	When to use
Event Hubs / IoT Hub	Partitioned distributed commit log, big data scale	Telemetry, distributed data streaming, no client expectation
Service Bus	Message broker	Message processing, client expectation
Storage Queues	Work backlog	Worker queue
Event Grid	Event distribution	React to service changes

Services for Transient Storage

Event Hubs

Event Grids

Service Bus

Message Queue with Messages

Topic with three Subscriptions with Messages
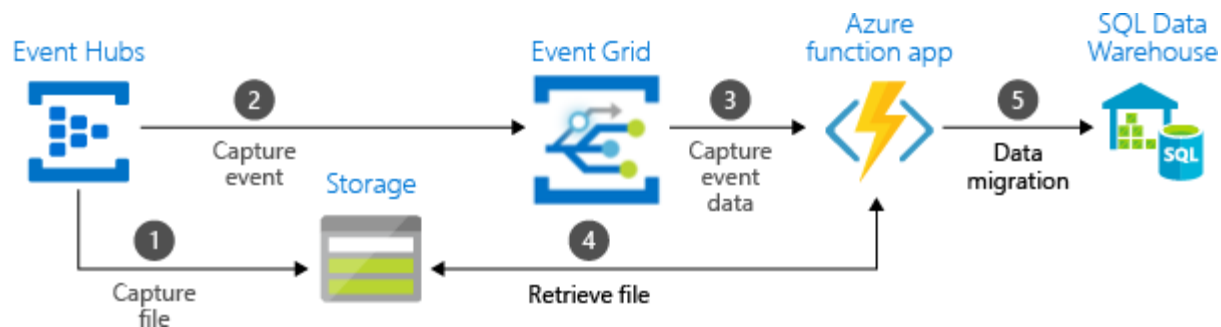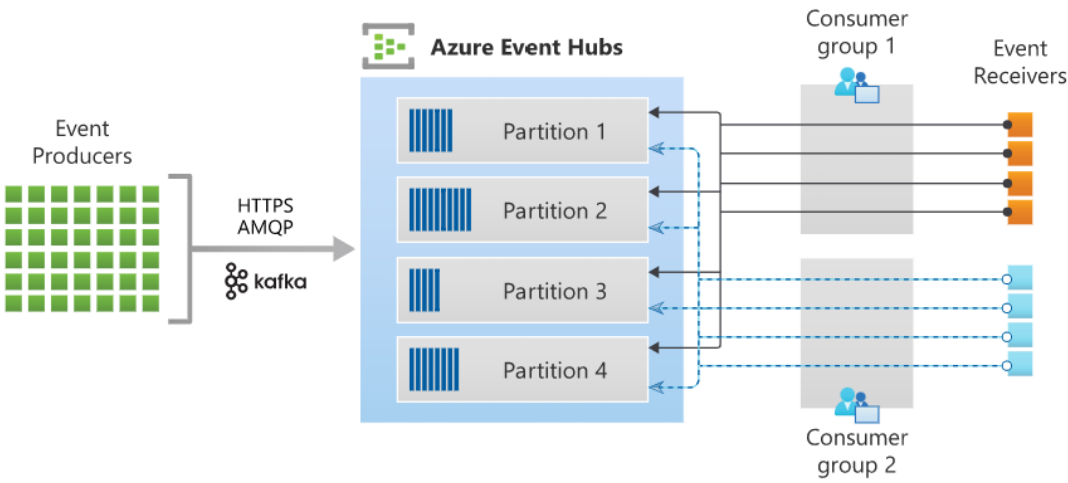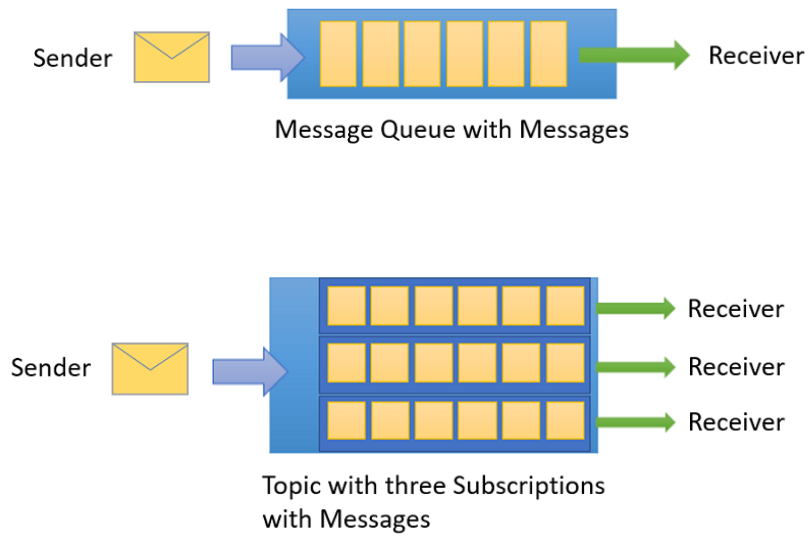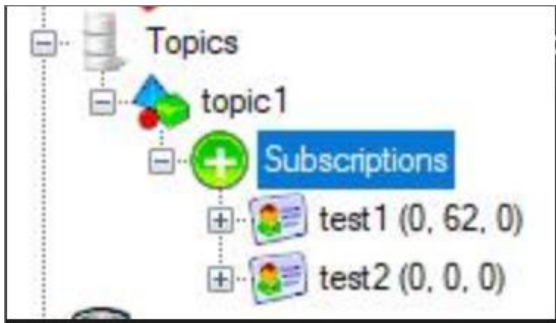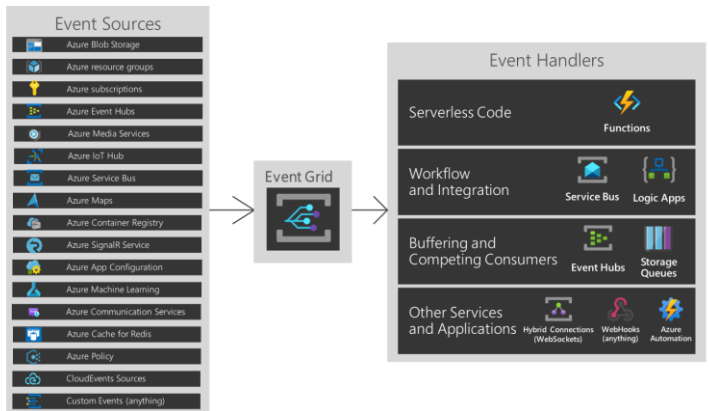
Event Processing

What is Event Processing

Push Source
Pull Source
Data Acquisition
Transient Storage
Event Processing
Serving Layer
Persistent Storage
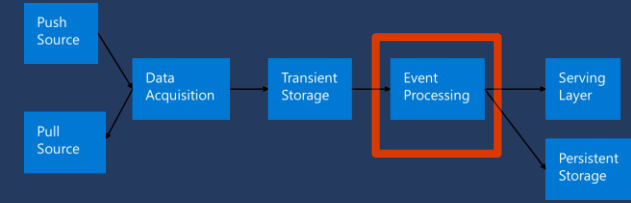
- **Consume** individual messages
- **Execute** business logic per message to produce business events
- **Integrate** business events into serving layer
- **Persist** messages and business events to long terms storage
- Not always a big data problem

Event Processing (How)

Transient Storage → **1** Messages → Event Processing → **2** **3** Integrating Business Event → Serving Layer

Event Processing → Messages Business Events → **4** → Persistent Storage

1 - Consuming Messages
- Consumption pattern (Competing consumers, Pub-Sub, In-order)
- Push or pull
- Client or server side cursors

3 - Integrating with Serving Layer
- Integration logic
  - App integration (REST API, etc)
  - Data Integration (Data Model)

2 - Event Processing
- Creating the business event(s)
  - Business logic ordering expectation
  - Idempotency
  - Workload Type
  - Skillsets

4 - Long Term Retention
- What will be retained, messages, events?
- Latency expectations
- Schema on read / write

Event Processing Services

- Provisioned or serverless
- Team skills and polyglot programming
- Distributed compute
- Azure Services
  - Stream Analytics
  - Apache Spark family (Synapse Spark Pools, Azure Databricks, HD Insight, Azure Kubernetes Service)
  - Azure Functions, Durable Functions
  - In engine features (Azure Data Explorer, Synapse)

Services for Event Processing

- Distribution of compute, can a single instance process the event?
- Computation, how is the logic implemented
- Delivery of outputs

- Many independent tasks
  - Stream Analytics
  - Azure Functions
  - Azure Batch
  - Azure Kubernetes Services (Inferencing)
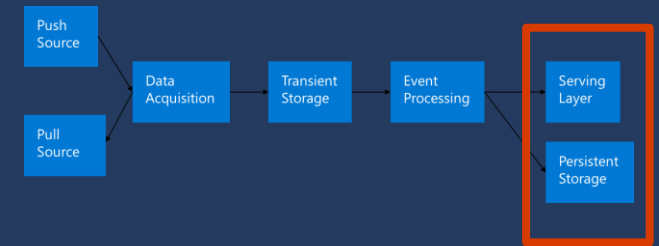 - Azure VM

- Interrelated tasks
  - Apache Spark (Synapse Spark Pools, Azure Databricks)
  - Logic Apps
  - Durable Functions
  - Azure Kubernetes Service (Micro services)
  - In engine processing
    - Cosmos Db
    - Azure Data Explorer
    - Synapse Analytics
    - Azure SQL DB

Serving and Long term retention

Serving and Persisting

- Delivering on the value proposition
- End user preferences and change management
- Not always a single service, workload matters

Workload Type	Azure Services
Analytical Workloads	Synpase Workspaces Power BI Azure Data Explorer
Transactional Workloads	Cosmos Db Azure SQL DB OSS (Azure Database for PosgresSQL, MariaDB)
Data Lake (file based) and Unstructured Workloads	Azure Data Lake Store Gen 2
Application Workloads	Azure Functions, Azure API Management

Building Blocks

3
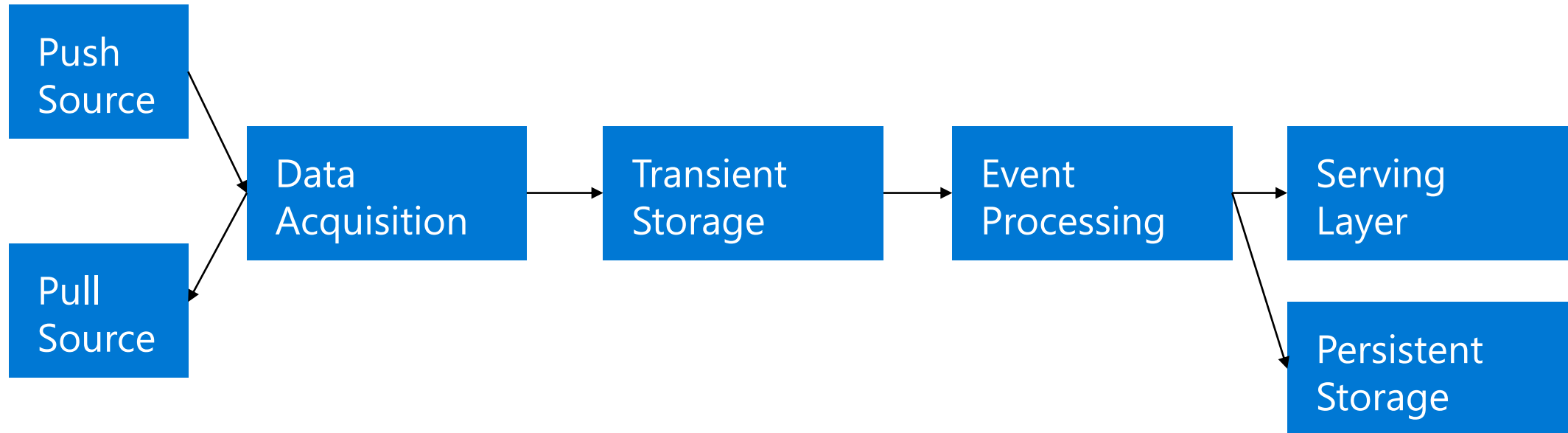
COMMON ENTERPRISE REFERENCE ARCHITECTURE PATTERNS HOSTED IN AZURE
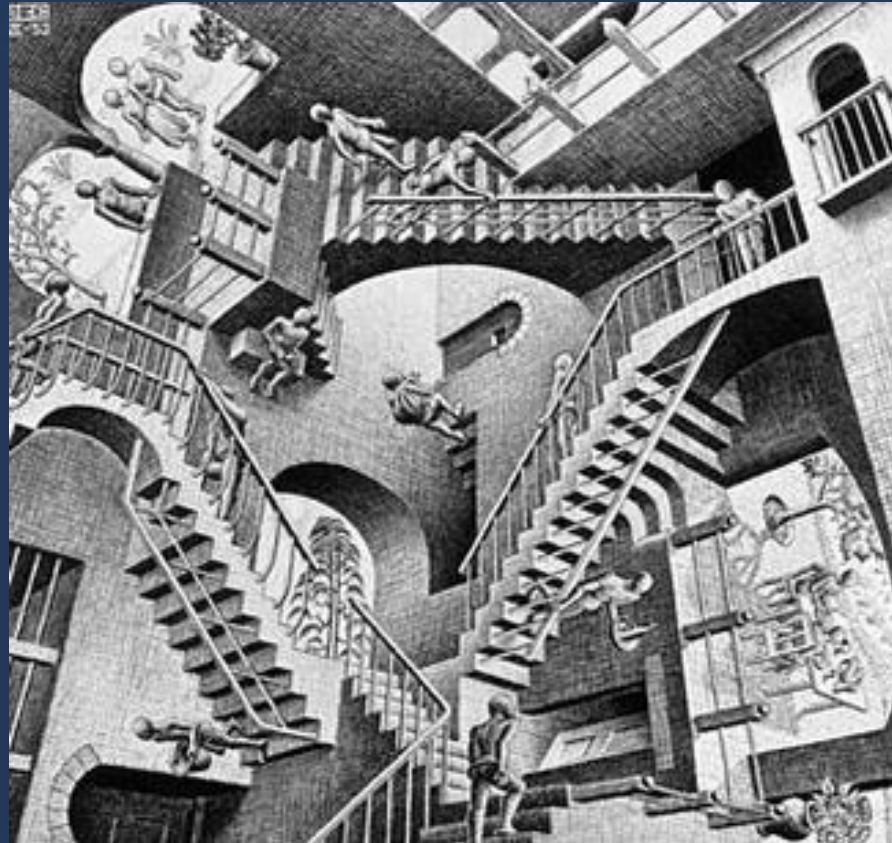
https://en.wikipedia.org/wiki/Relativity_(M._C._Escher)

Introduction

Common Azure hosted architecture patterns;

- Lambda Architecture
  - Example 1: Azure Data Explorer and Azure Synapse

- Kappa Architecture
  - Example 1: Azure Databricks and Azure CosmosDB
  - Example 2: Open Source (OSS) Implementation

Reference: What is the LAMBDA architecture?

*"The Objective of **Lambda Architecture** is to leverage the combined power of both **batch** & **real-time** processing to address the business scenarios where it requires both **historic view of the data** as well as getting insight into the **data in real-time** as business happens."*
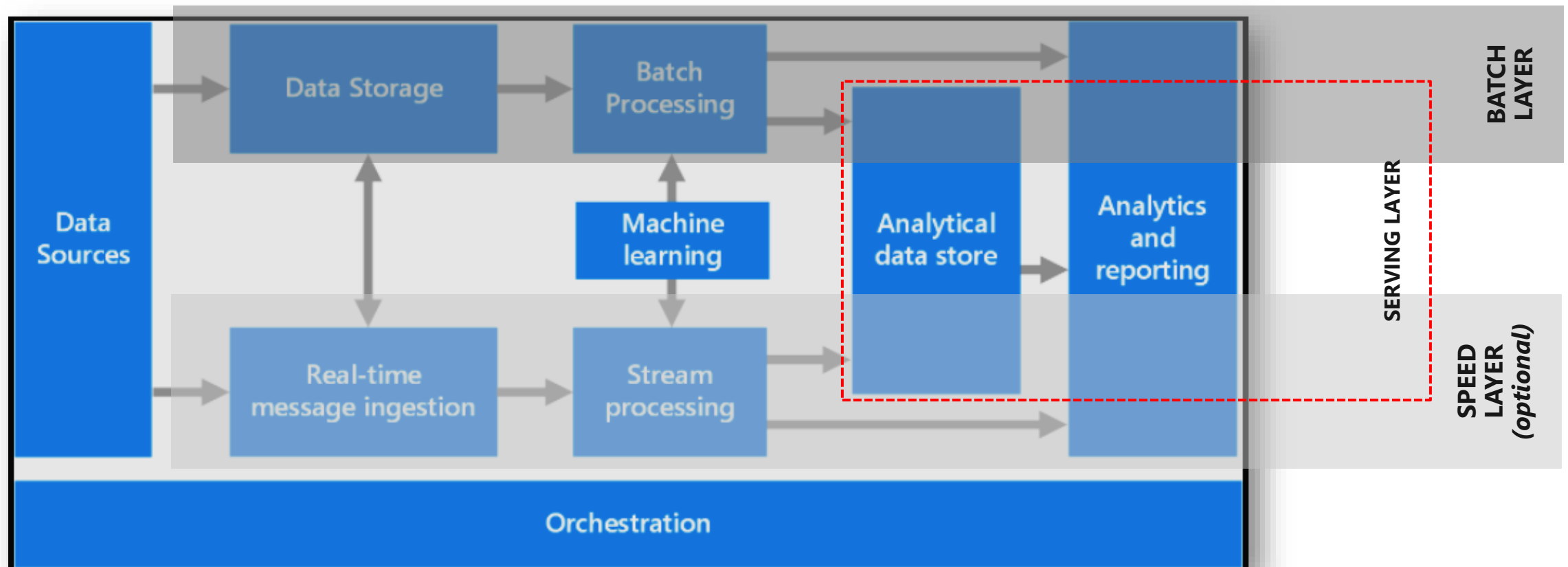
Example 1 – LAMBDA Deployed in Azure Data Explorer and Synapse
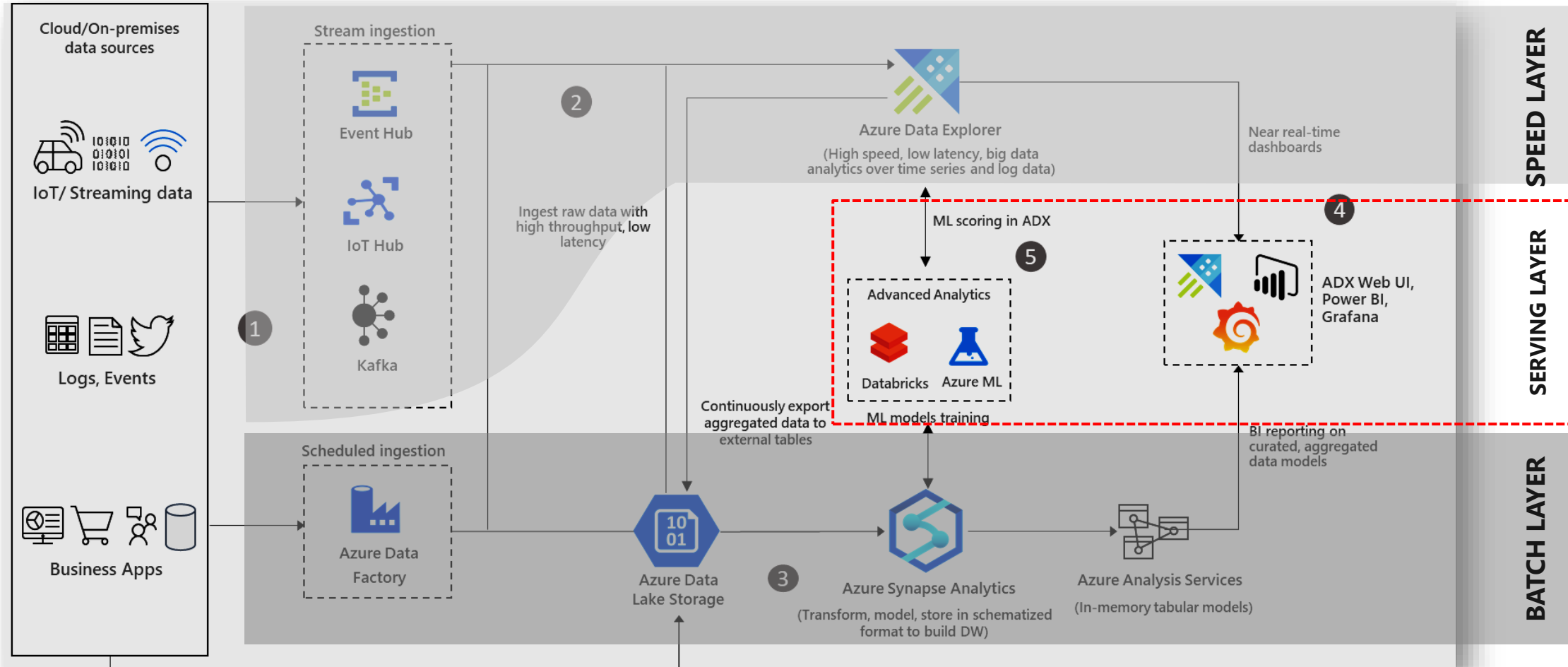
Reference: What is the KAPPA architecture?

*"The Objective of **Kappa Architecture** is similar to Lambda, however all data flows through a single path, using only a stream processing system. (ie everything is considered a stream)"*
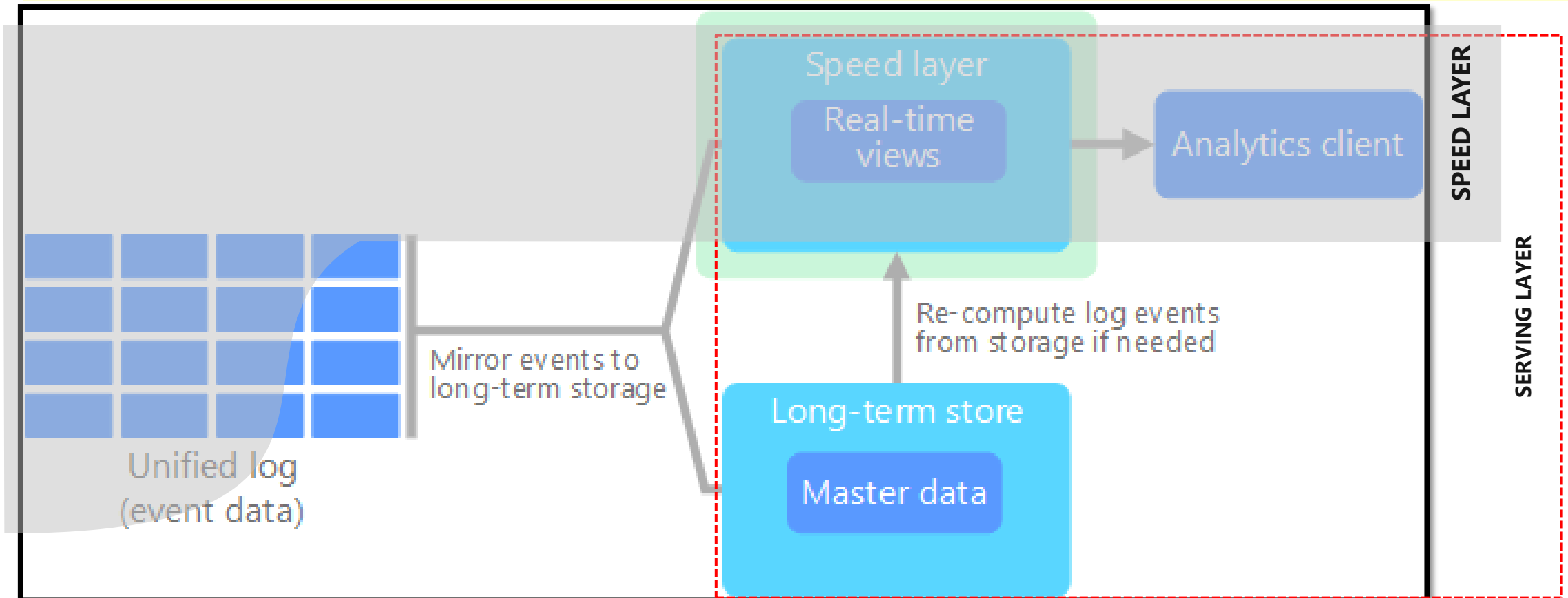
Example 1 – KAPPA Deployed in Azure Databricks and CosmosDB

https://docs.microsoft.com/en-us/azure/architecture/reference-architectures/data/stream-processing-databricks
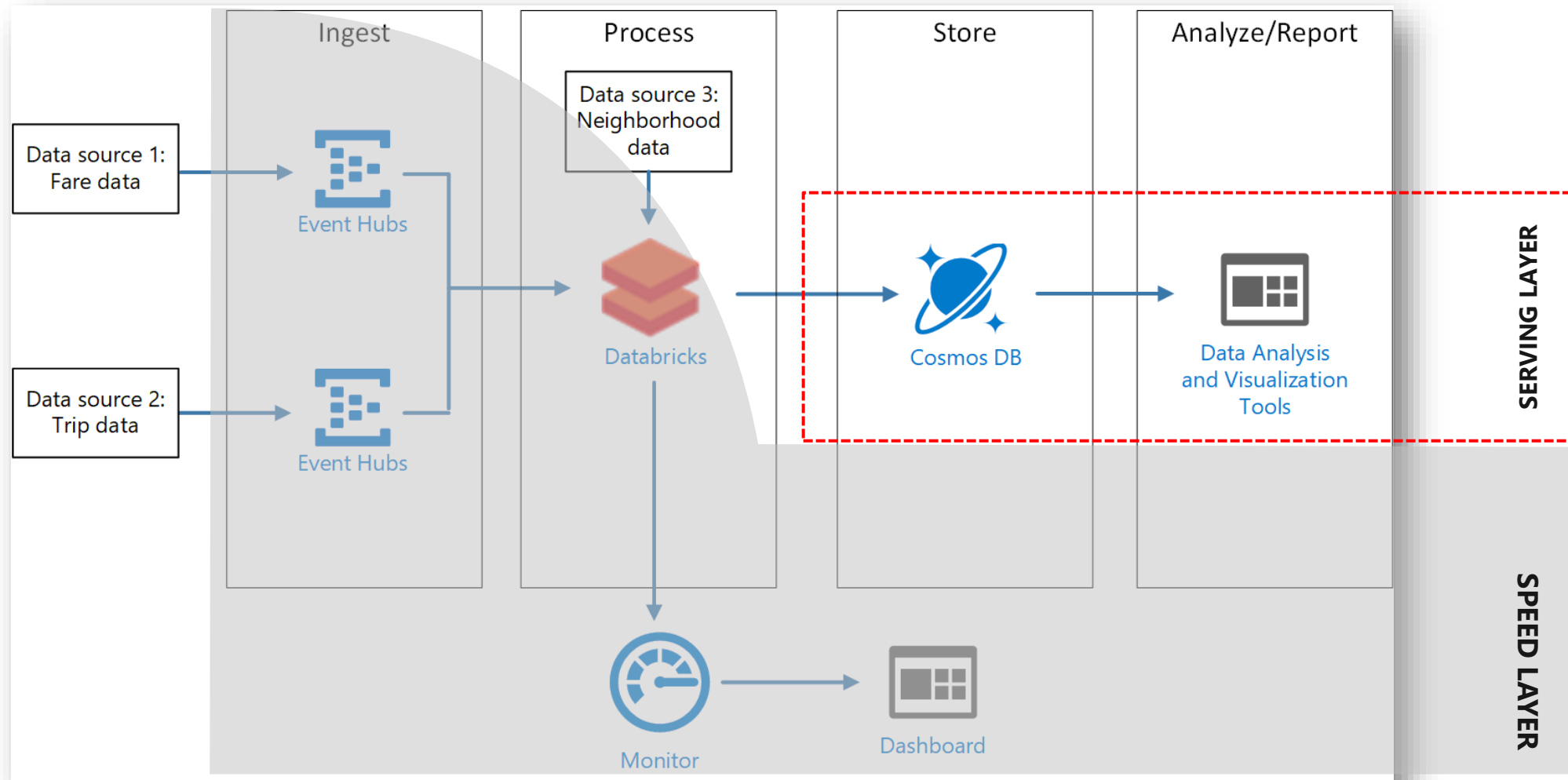
Example 2 – KAPPA Deployed in Azure Open Source (OSS) Implementation
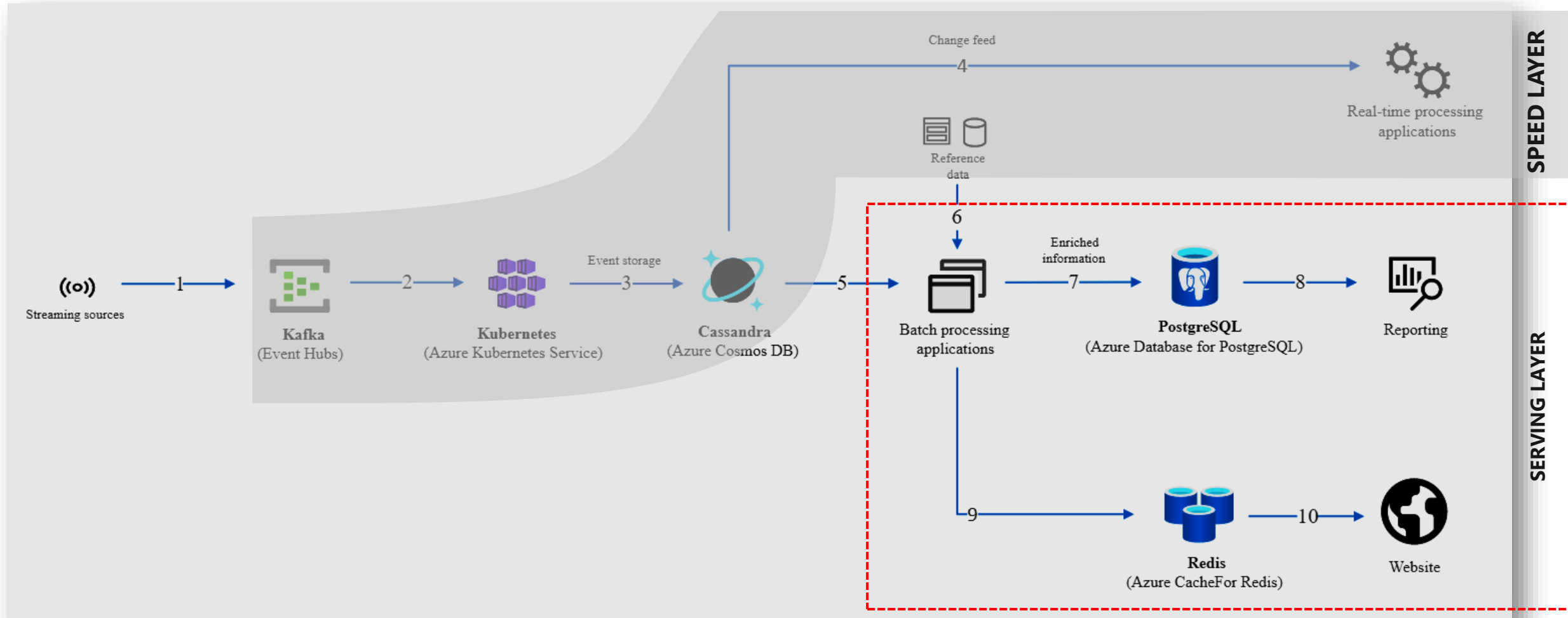
WALKTHOUGH CUSTOMER EXAMPLES - AZURE STREAMING ARCHITECTURES

4
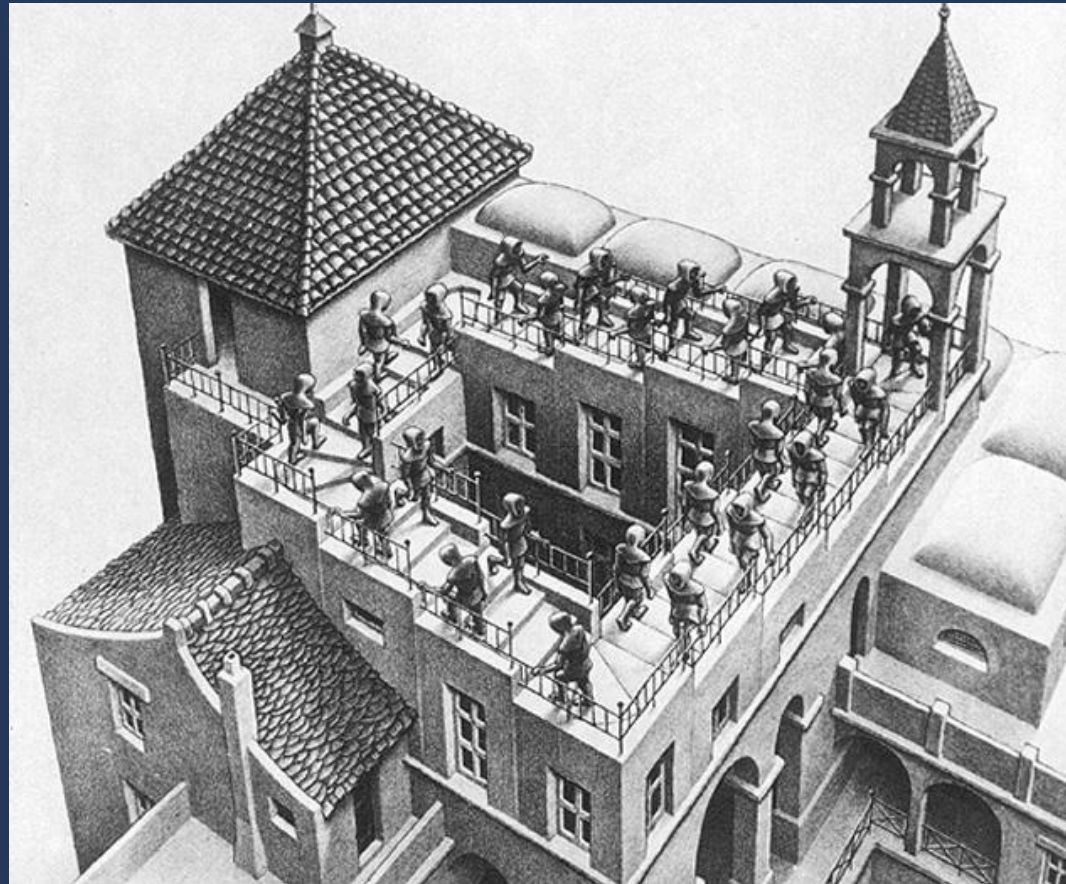
https://en.wikipedia.org/wiki/Ascending_and_Descending

Introduction

Example Architectures Deployed by Customers in Azure...

1. Real-Time "Price Drop" Detection
2. Virtual Power Plant (VPP) IoT
3. Real-Time Web Search Telemetry Capture and Query
4. Enterprise Messaging via Service Bus
5. Running IIOT in a Layered Architecture

Ex 1 - Real-Time "Price Drop" Detection
Multiple Event Buffering for Scalable Asynchronous Processing
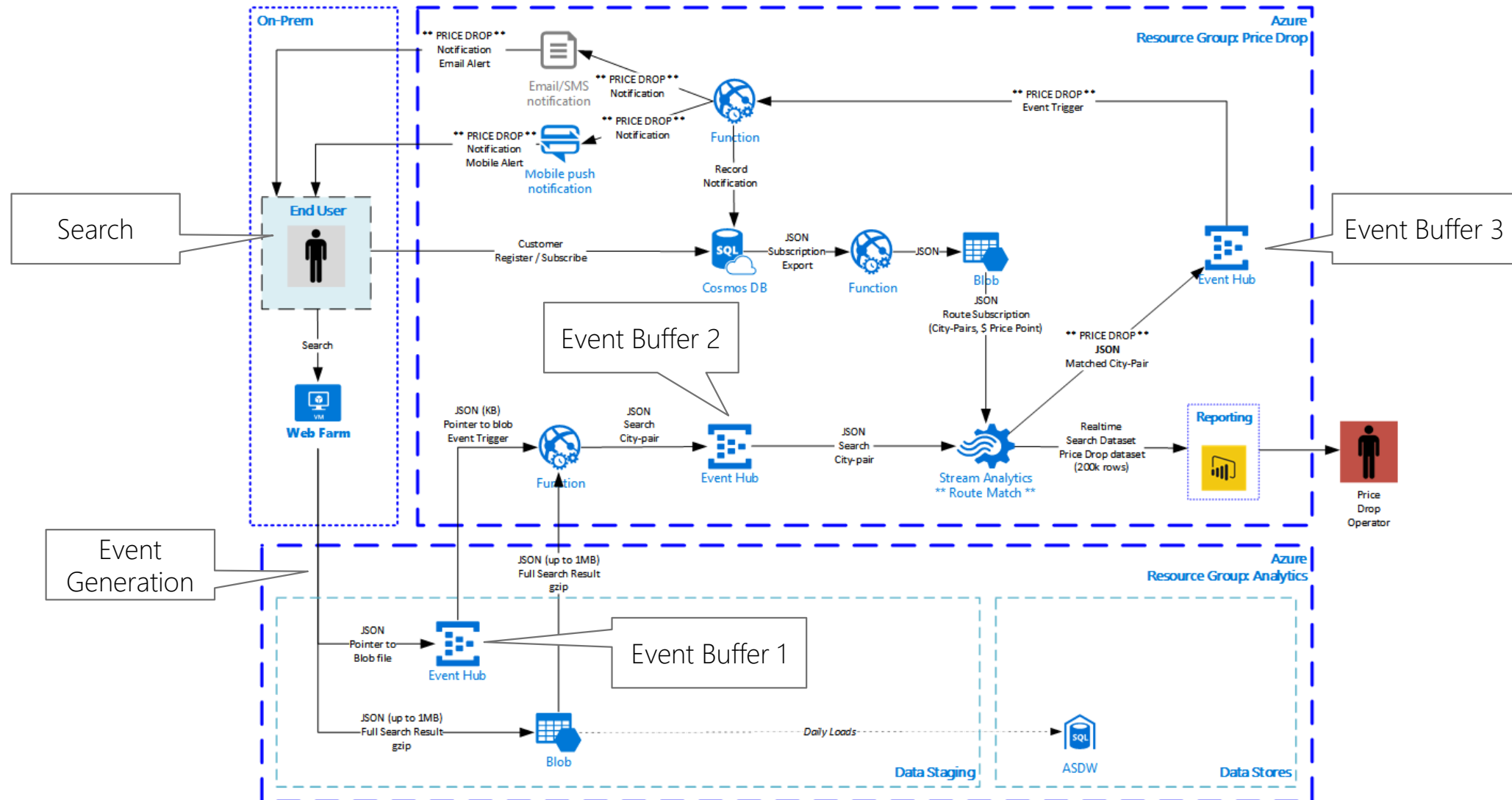
Ex 2 - Virtual Power Plant (VPP) IoT
High-Scale IoT Ingestion and Device Control / Management
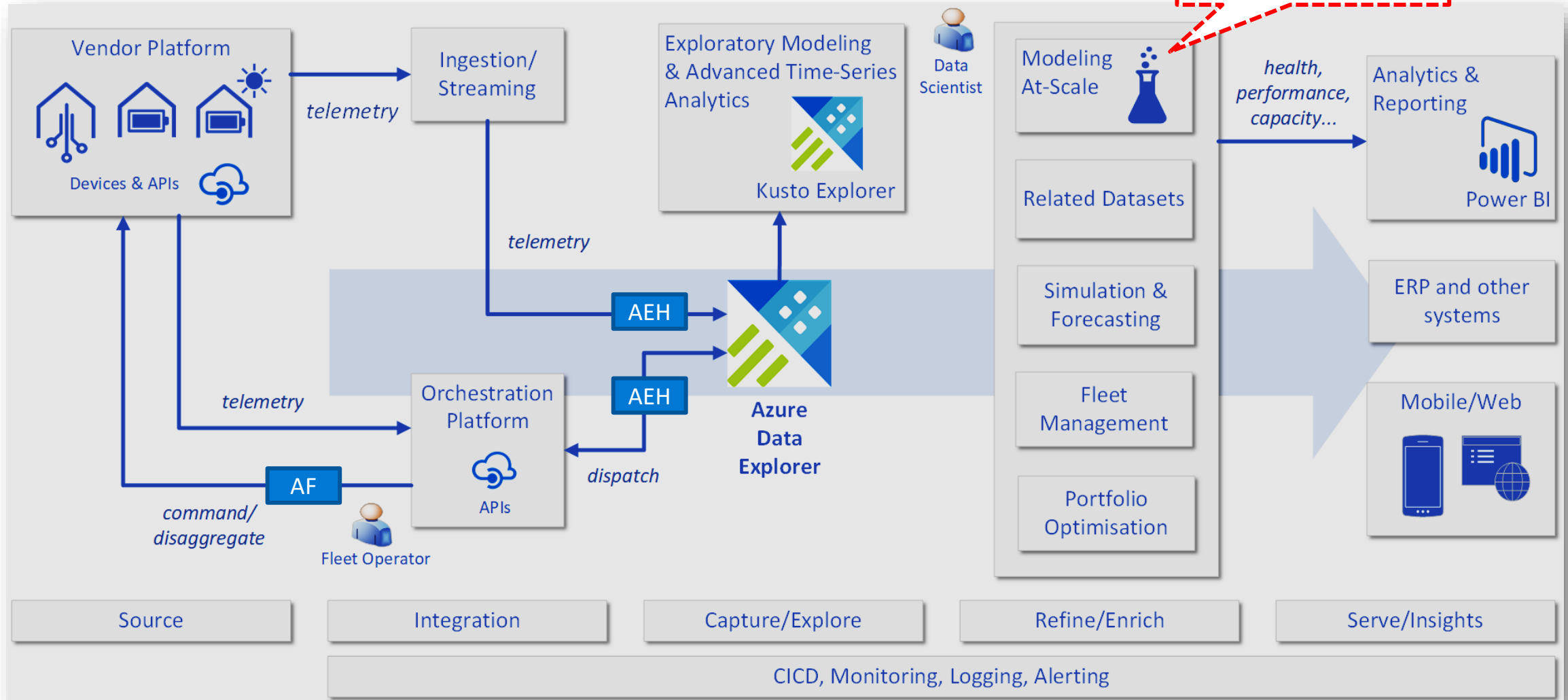
Azure ML Platform integration
Case Study Here

Vendor Platform
Devices & APIs

telemetry →

Ingestion/ Streaming

Exploratory Modeling & Advanced Time-Series Analytics
Kusto Explorer

Data Scientist

Modeling At-Scale

health, performance, capacity... →

Analytics & Reporting
Power BI

telemetry

AEH

Related Datasets

Azure Data Explorer

Simulation & Forecasting

ERP and other systems

AEH

Fleet Management

telemetry

Orchestration Platform
APIs

dispatch

Portfolio Optimisation

Mobile/Web

AF

command/ disaggregate

Fleet Operator

Source	Integration	Capture/Explore	Refine/Enrich	Serve/Insights

CICD, Monitoring, Logging, Alerting

Case Study: Microsoft Customer Story-Growing an innovative energy partnership across Australia

Ex 3 – Real-Time Web Search Telemetry Capture/Query
Large Scale Event Ingestion with Interactive Live Query

SH_Data_Streaming
(West Europe)

AVRO Event Archive
Batch
(COLD Path)

Blob Store / ADLS
SHEventStore

JSON Report

Service Bus Queue
~1/hour

Data Archive
Avg: 56GB/day

Logic App
Status Report
~1/hour

JSON
Event

alerts

Hourly

Logic App
SHLogicApp
PostEvents

ServiceBus Q
SHSBQEgress

Event Hub
Max: 3900/sec
Avg: 2300/sec

Stream Analytics
Telemetry Input
3900/sec

ref data

SQL SP

Power BI
On Demand

search

Server

Server

Server

Server

JSON
Event Type

Event Hub
SHIngress

JSON Events
Stream
(HOT Path)

Stream Analytics
SHEgressASDB
...Telemetry
...Bookings
...Agents
...Proviers

Tabular
Events

SQL DB
SHEventHistory
(Short Term Store)

Reporting
data

Historical
Reports

reports

On-Demand

Stream Analytics
Power BI
Aggregation Path
3900/Sec → 1/min

Azure SQL Database
Max: 3900/sec
Avg: 2300/sec
(5 days = 1b rows)
(1 year = 72b rows)

Power BI
1/min
(200K tumbling window)

JSON Events
Stream
(HOT Path)

Stream Analytics
SHEgressPBI

Realtime Stream
(200K rows moving window)

Real Time
Dashboards

reports

1 Min Window

Ex 3 – Real-Time Web Search Telemetry Capture/Query
Metrics: Total Events By Hour By Day (Last 7 Days)
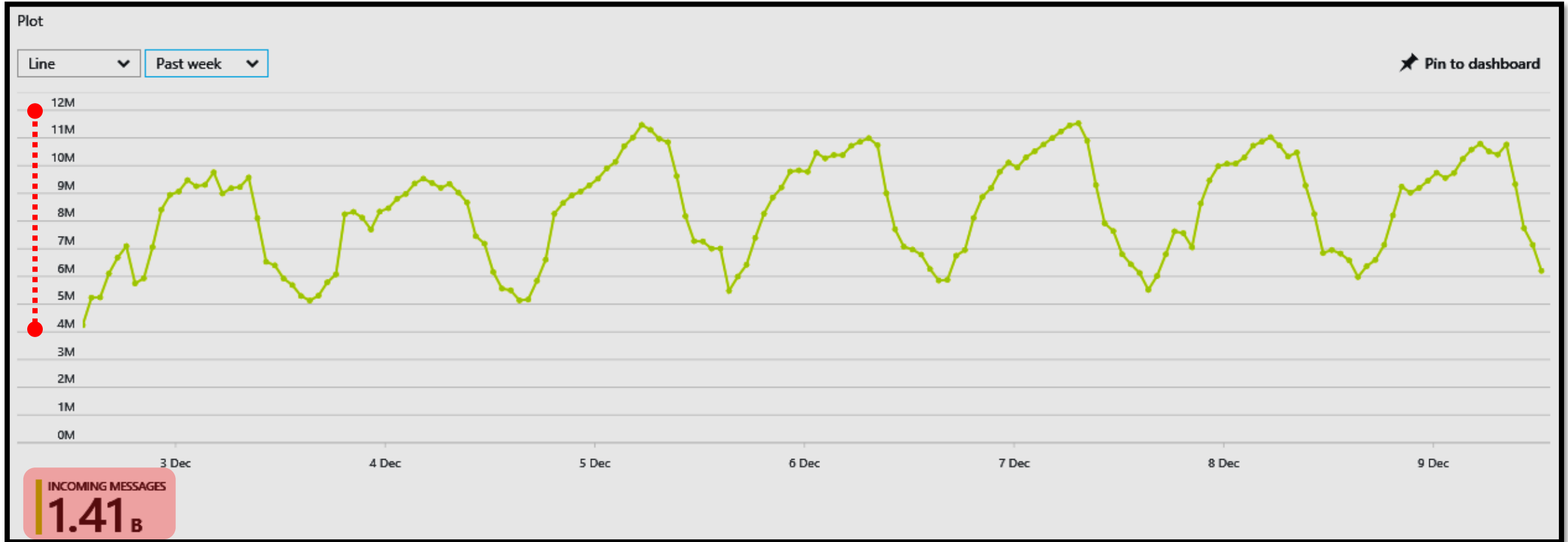
Average Load → **1,410,000,000 / week**

= **201,000,000 / day**

= **8,392,000 / hour**

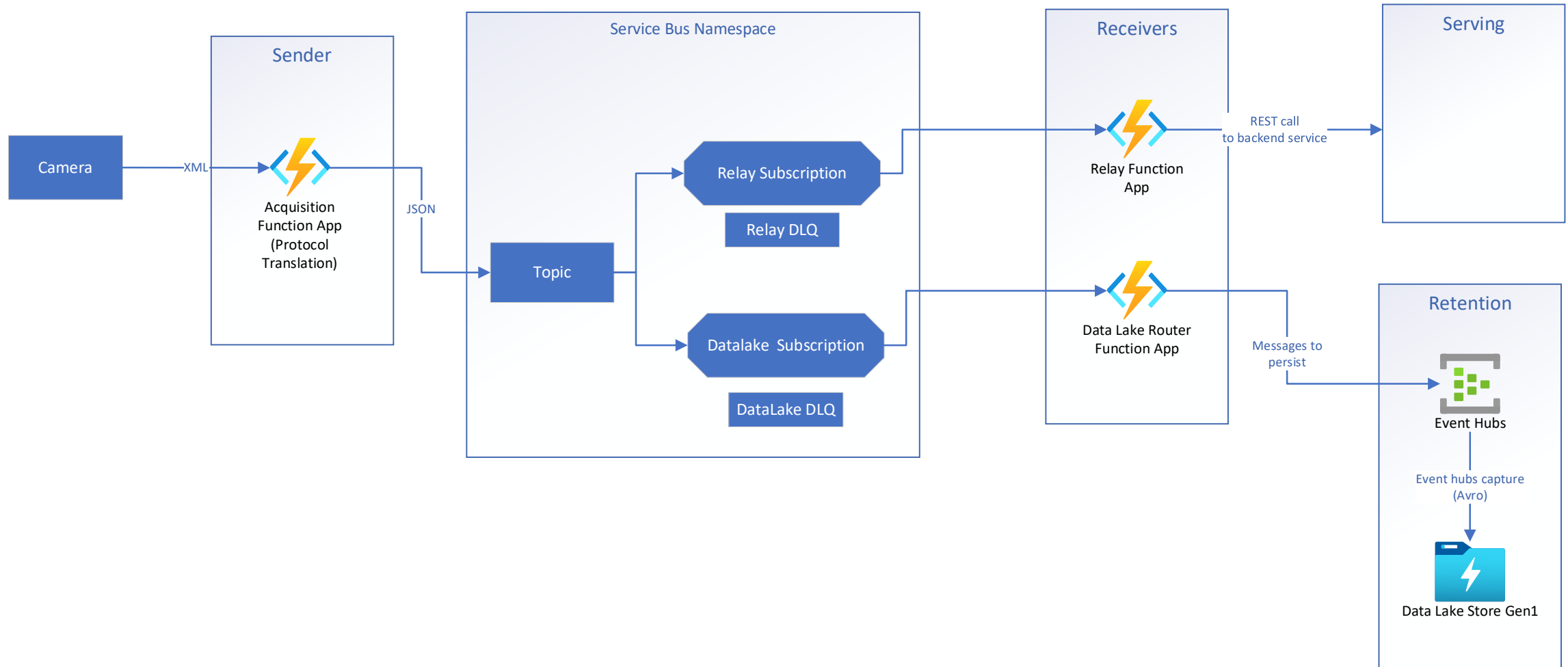= **139,000 / min** = **2,330 / sec**

Ex 3 – Real-Time Web Search Telemetry Capture/Query
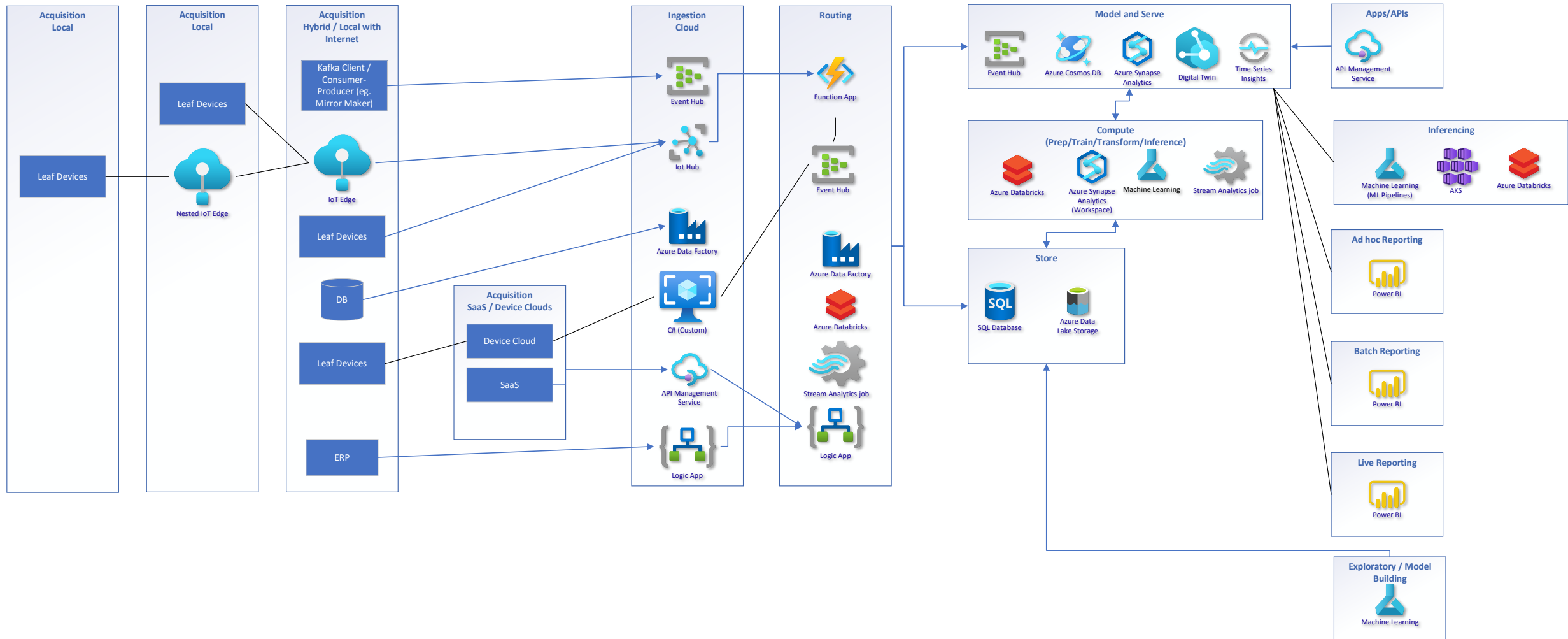## Metrics: Min | Avg | Max Events per Sec (Grouped by Hour of Day)

Service Bus

IIoT – Layered Network Model

5

DEMOS;
LOOP BACK ON THE THEORY

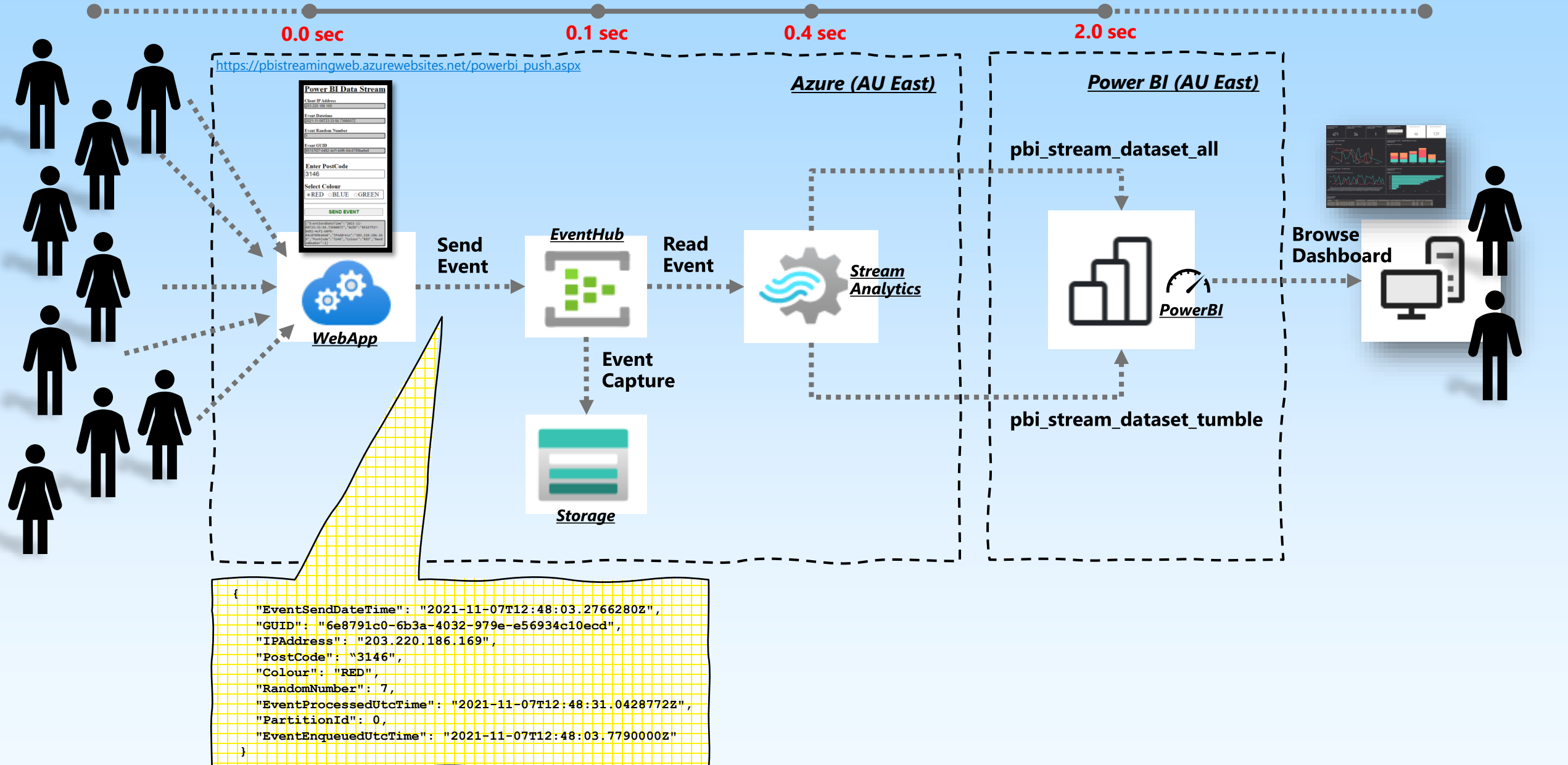https://www.pinterest.fr/mariechollet58/absurde/

Introduction

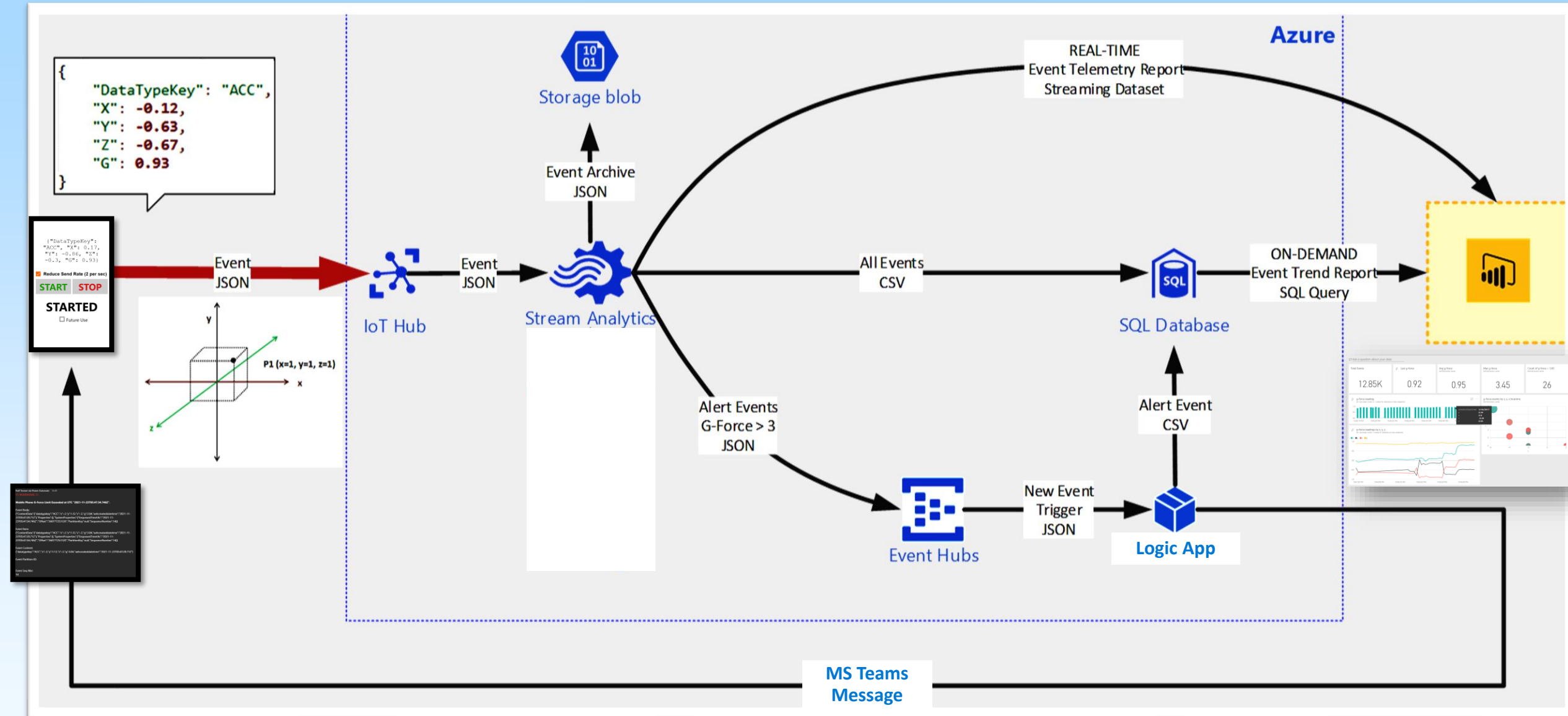**Let's bring the various architectures together via some demos!**

- Interactive Web Telemetry
- Mobile G-Force Phone Application
- Power BI Real-Time Streaming Datasets
- Stream Processing with Event Hubs and Spark Structured Streaming
- Lambda Architecture with Cosmos Db

DEMO: Interactive Web Telemetry

0.0 sec   **0.1 sec**   **0.4 sec**   **2.0 sec**

https://pbistreamingweb.azurewebsites.net/powerbi_push.aspx

Azure (AU East)

Power BI (AU East)

pbi_stream_dataset_all

WebApp

Send Event

EventHub

Read Event

Stream Analytics

Event Capture

Storage

PowerBI

pbi_stream_dataset_tumble

Browse Dashboard

```
{
    "EventSendDateTime": "2021-11-07T12:48:03.2766280Z",
    "GUID": "6e8791c0-6b3a-4032-979e-e56934c10ecd",
    "IPAddress": "203.220.186.169",
    "PostCode": "3146",
    "Colour": "RED",
    "RandomNumber": 7,
    "EventProcessedUtcTime": "2021-11-07T12:48:31.0428772Z",
    "PartitionId": 0,
    "EventEnqueuedUtcTime": "2021-11-07T12:48:03.7790000Z"
}
```
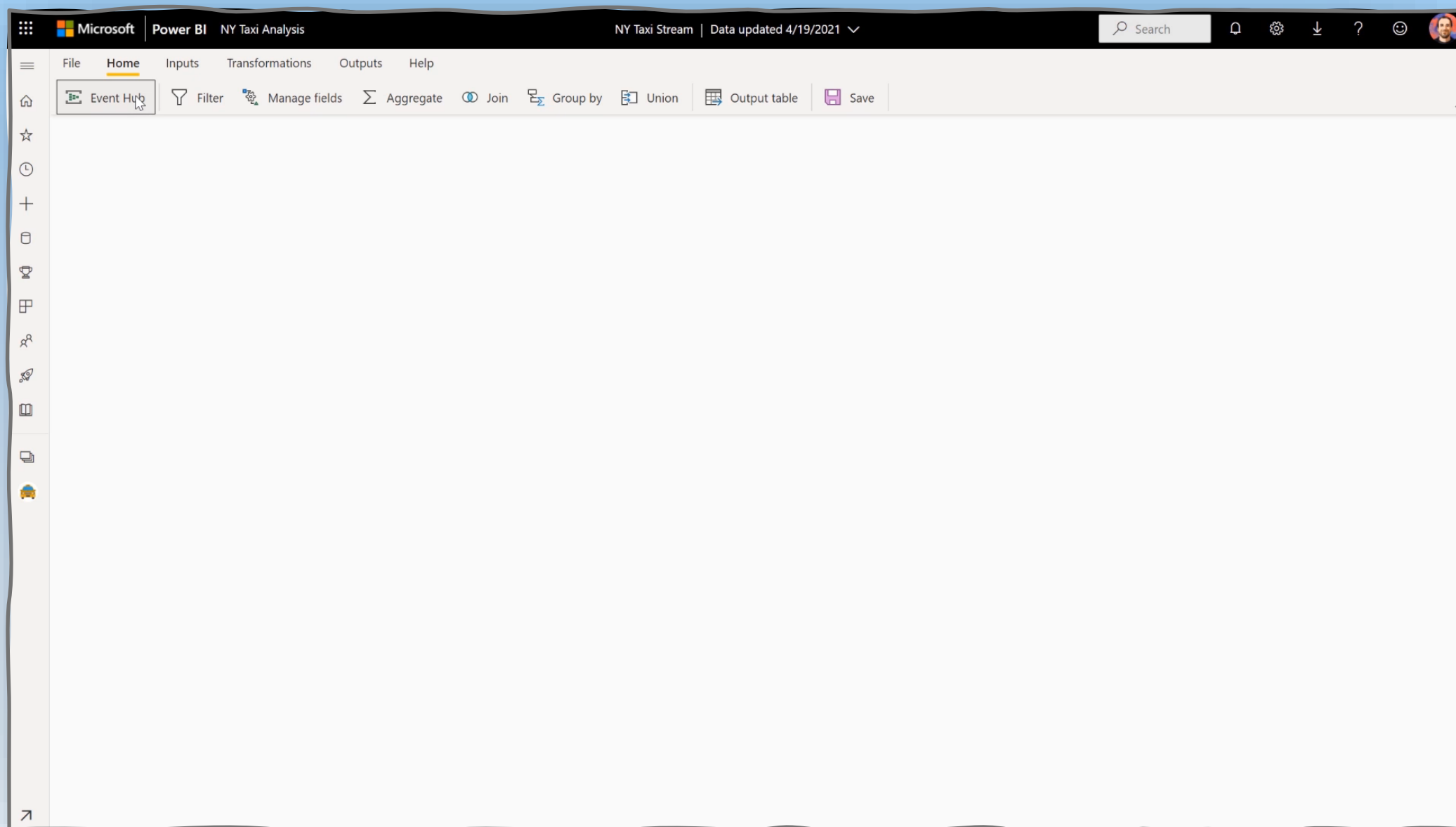
DEMO: Mobile G-Force Phone Application

DEMO: Power BI Real-Time Streaming Datasets

https://powerbi.microsoft.com/en-us/blog/enabling-intelligent-experiences-with-power-bi-for-developers-data-scientists-and-data-engineers/


- *you can have a streaming dataflow, combined with multiple other data sources where you can link the streaming data to other datasets*

- *Power BI incorporates similar functionality to that of Azure Stream Analytics but directly within the PBI Dataflows Engine*

DEMO: Digital Twin with Azure Digital Twin, IoT Hub, Event Hubs and Spark Structured Streaming
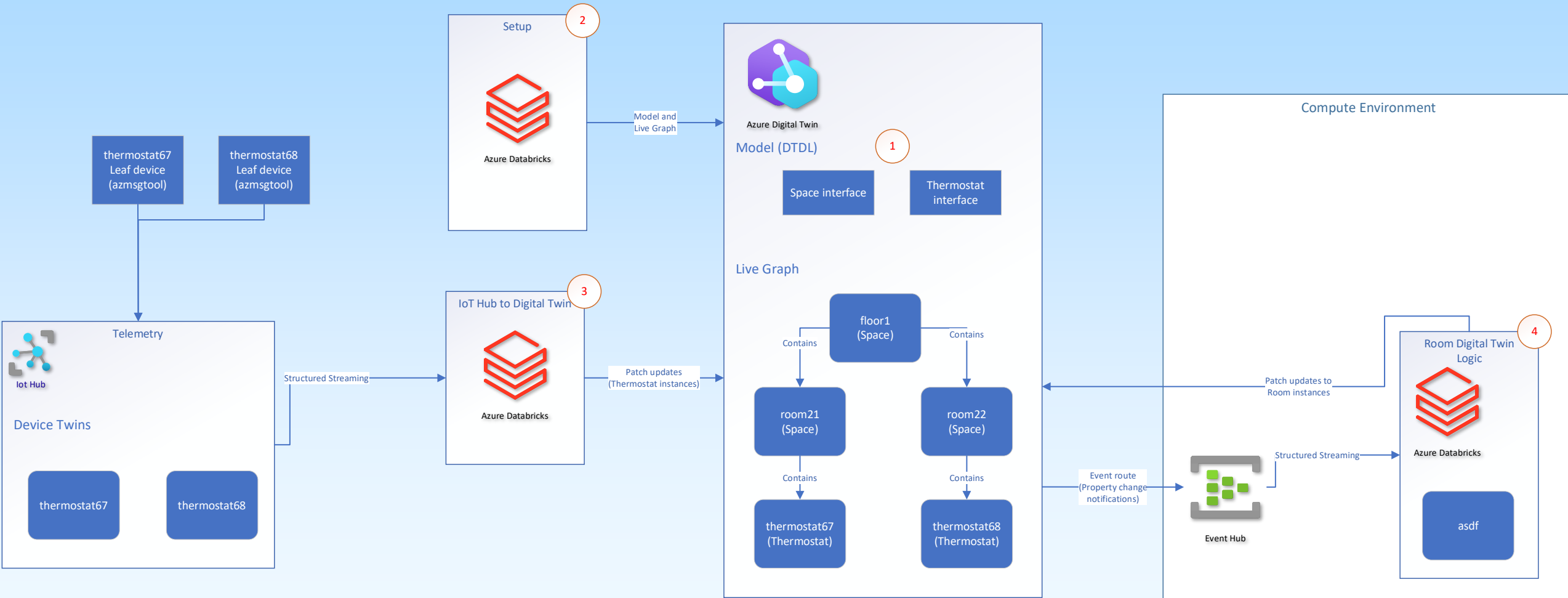
DEMO: Lambda Architecture with Cosmos Db
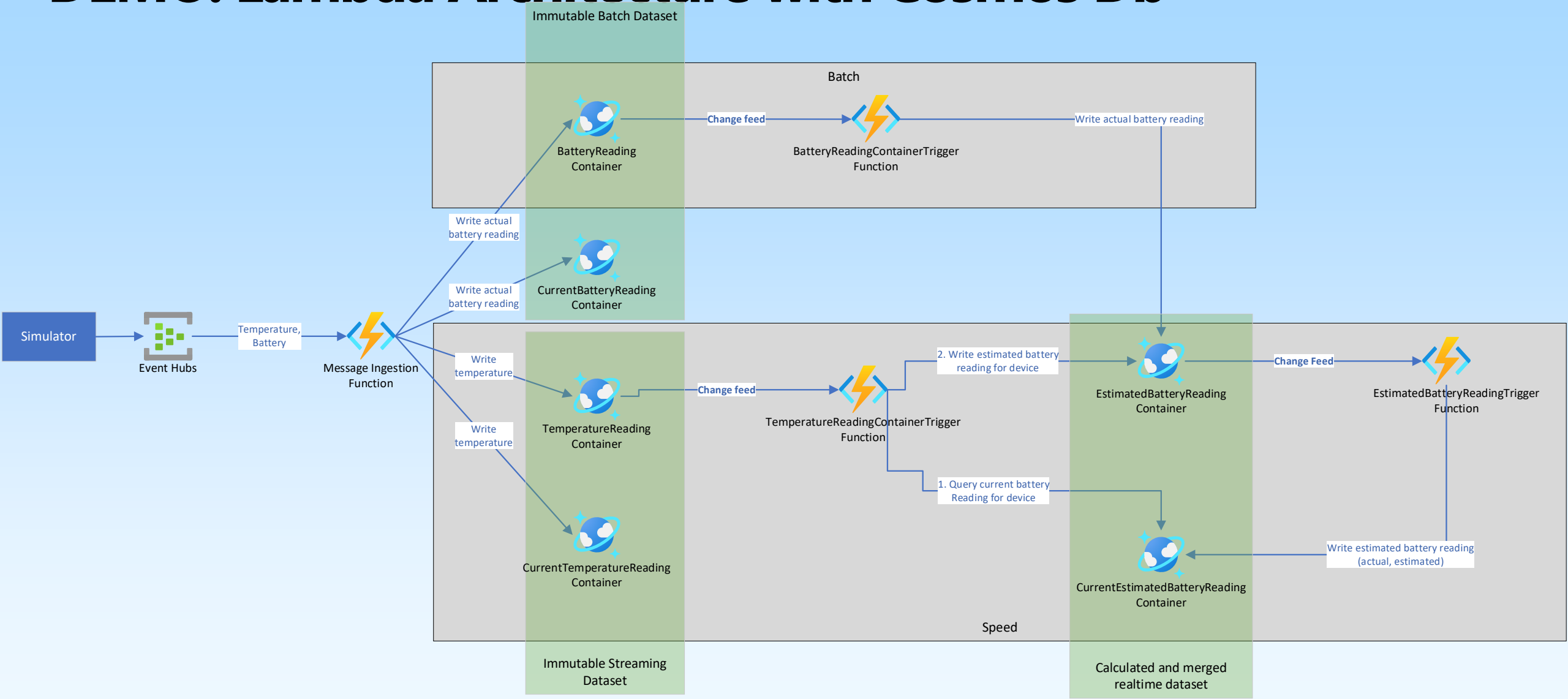
WRAP UP AND SUMMARY

6

What did we learn today?

We provided 4 key focus areas for today...

1. Azure Services normally used for Streaming Systems, Pro/Con

2. Common architecture patterns which accommodate Streaming

3. Actual examples of deployed Streaming Solutions at scale

4. Demos to showcase how Streaming Services work

Where can you go to learn more?

MS Docs Architecture Centre – for Streaming Solutions on Azure

- Analytics architecture design - Azure Architecture Center

Azure Courses on MS Learn – for Streaming Solutions on Azure

- Browse all - Learn | Microsoft Docs

Tutorials – for Azure Stream Analytics + Azure DataBricks

- Tutorial - Analyze fraudulent call data with Azure Stream Analytics and visualize results in Power BI dashboard
- Tutorial: Stream data into Azure Databricks using Event Hubs

Recommended Azure Exams / Certifications Encompassing Data/Streaming

- Microsoft Certified: Azure Fundamentals - Learn | Microsoft Docs
- Microsoft Certified: Azure Data Fundamentals - Learn | Microsoft Docs
- Microsoft Certified: Azure Data Engineer Associate - Learn | Microsoft Docs
- Microsoft Certified: Data Analyst Associate - Learn | Microsoft Docs

APPENDIX AND REFERENCES

A

Microsoft Docs References

1. [Big data architectures - Azure Architecture Center | Microsoft Docs](#)

2. [Real time processing - Azure Architecture Center | Microsoft Docs](#)

3. [Real time data ingestion Services Options](#)

4. [Choose an Azure Streaming Data Processing Service](#)

5. [Event Sourcing Pattern – Enterprise Messaging and Data Streaming](#)

6. [Browse Azure Architectures - Azure Architecture Center | Microsoft Docs](#)

7. [Push Datasets - REST API (Power BI Power BI REST APIs) | Microsoft Docs](#)

8. [Processing 100,000 events per second on Azure Functions | Azure Blog and Updates | Microsoft Azure](#)

9. [Analytics architecture design - Azure Architecture Center | Microsoft Docs](#)

10. [Azure Service Bus Explorer (SBE) Tool](#)