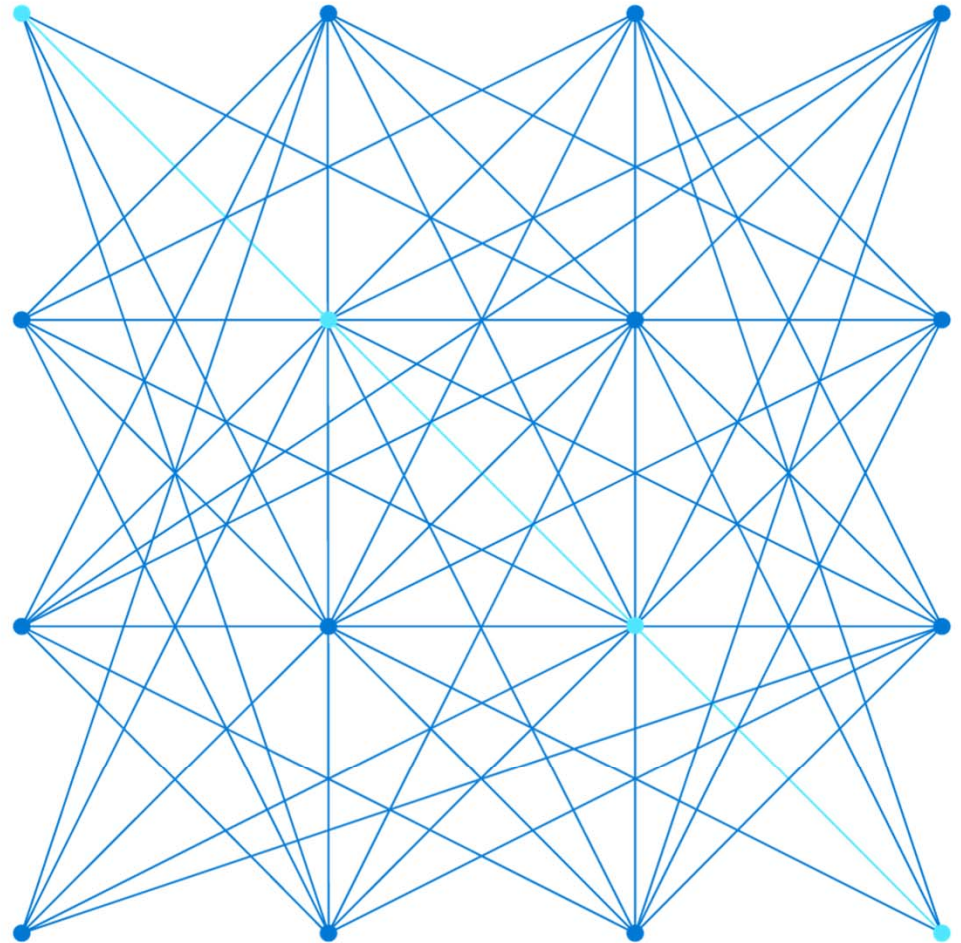


# DP-203T00: Real-time Stream Processing with Stream Analytics



# Agenda



Lesson 01 – Enable reliable messaging for Big Data applications using Azure Event Hubs

---



Lesson 02 – Work with data streams by using Azure Stream Analytics

---



Lesson 03 – Transform data by using Azure Stream Analytics

## Lesson 01: Enable reliable messaging for Big Data applications using Azure Event Hubs



# Azure Event Hubs



*Azure Event Hubs is a highly scalable publish-subscribe service that can ingest millions of events per second and stream them into multiple applications*

Home > Create a resource > Event Hubs >

## Create Namespace

Event Hubs

Basics Tags Review + create

### PROJECT DETAILS

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription \*

Resource group \*  [Create new](#)

### INSTANCE DETAILS

Enter required settings for this namespace, including a price tier and configuring the number of units (capacity).

Namespace name \*  ✓  
servicebus.windows.net

Location \*

**i** The region selected supports Availability zones. Your namespace will have Availability Zones enabled. [Learn more.](#)

Pricing tier (View full pricing details) \*

Throughput Units \*

# Create an Event Hub

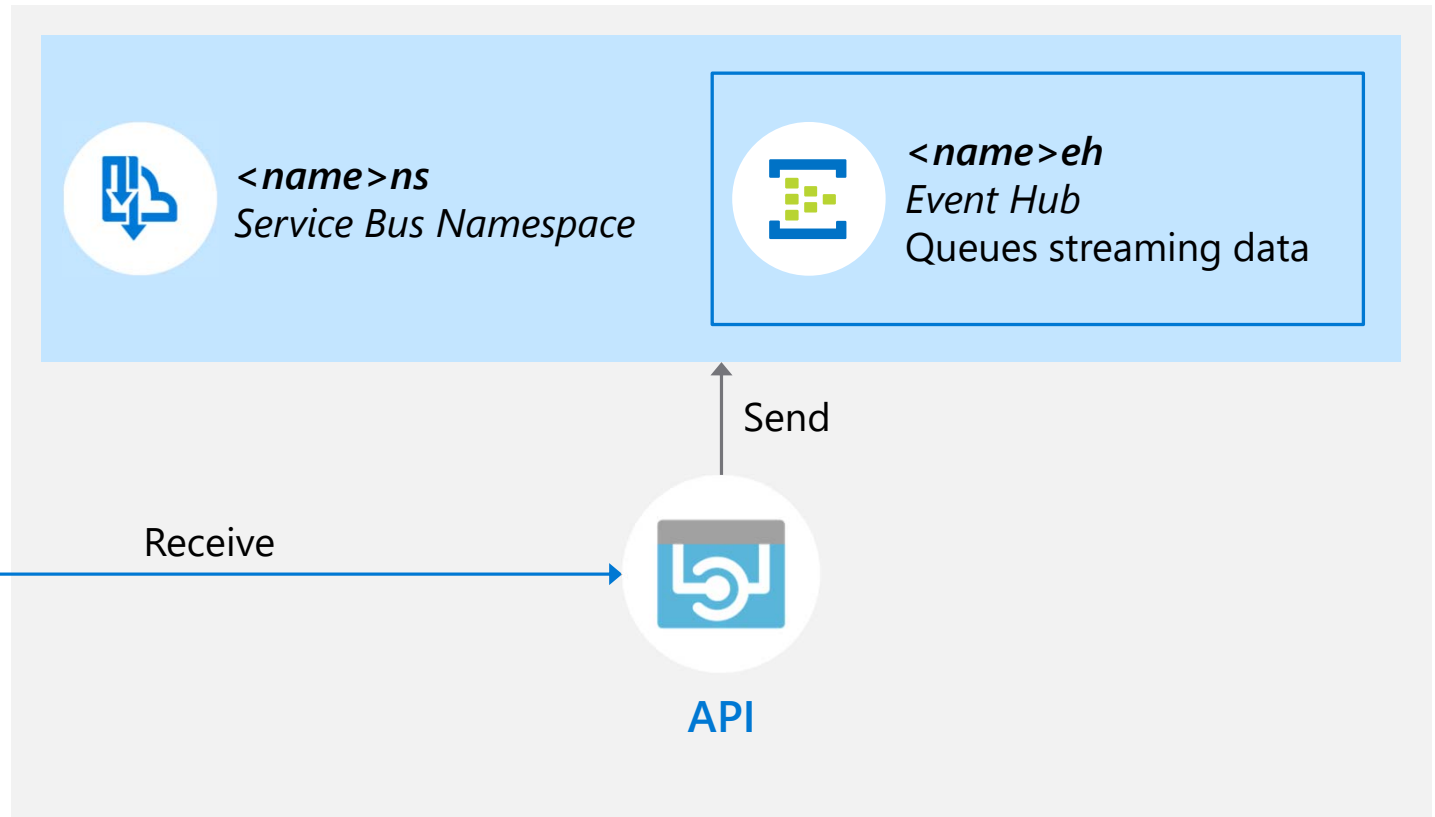
## Create an event hub namespace

1. In the [Azure portal](#), select + **Create a Resource**, type **Event Hubs**, and then select **Event Hubs** from the resulting search. Then select **Create**
2. Provide a name for the event hub, and then create a resource group. Specify **xx-name-eh** and **xx-name-rg** respectively, XX- represent your initials to ensure uniqueness of the Event Hub name and Resource Group name
3. Click the checkbox to **Pin to the dashboard**, then select the **Create** button

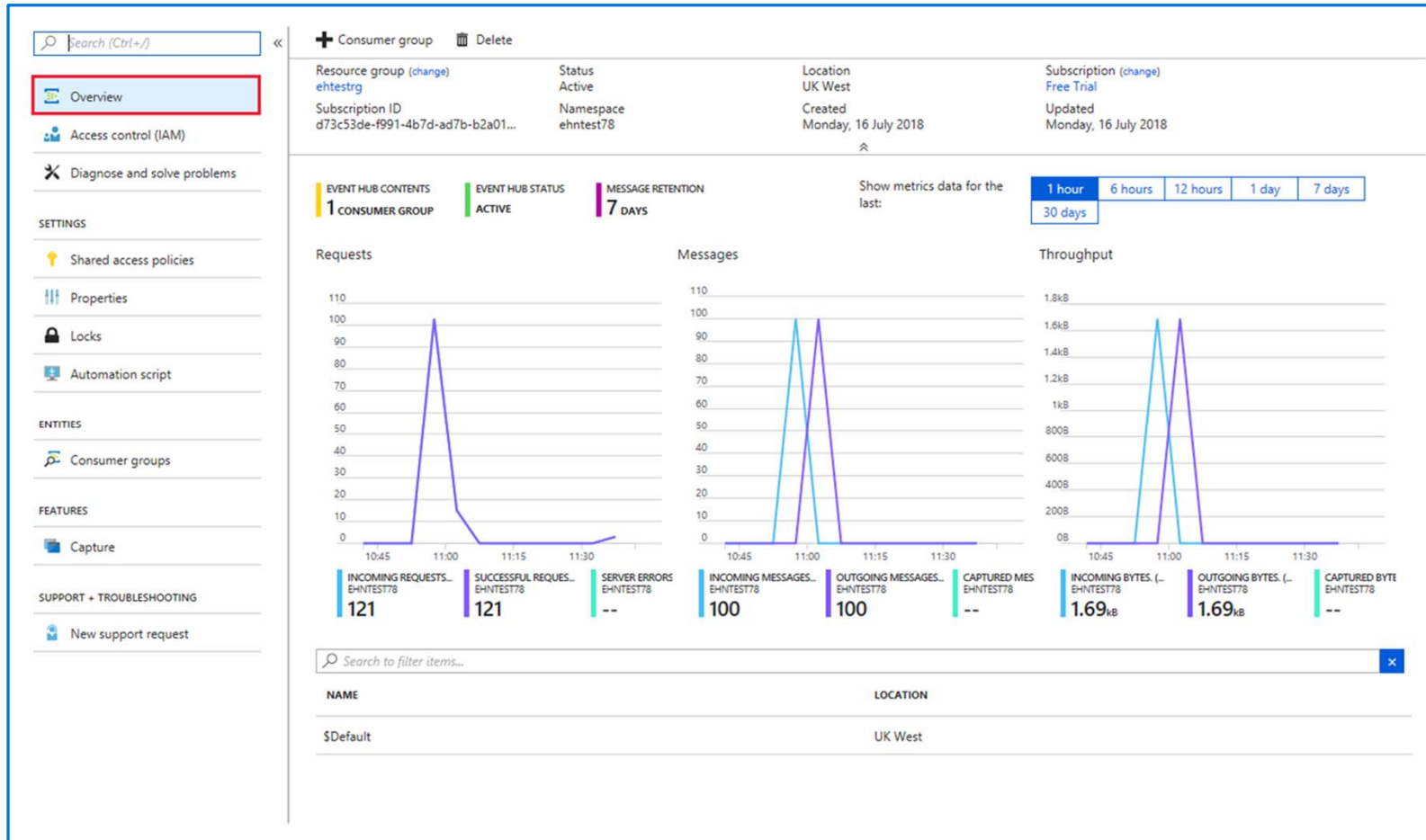
## Create an event hub

1. After the deployment is complete, click the **xx-name-eh** event hub on the dashboard
2. Then, under **Entities**, select **Event Hubs**
3. To create the event hub, select the + **Event Hub** button. Provide the name **socialstudy-eh**, and then select **Create**
4. To grant access to the event hub, we need to create a shared access policy. Select the **socialstudy-eh** event hub when it appears, and then, under **Settings**, select **Shared access policies**
5. Under **Shared access policies**, create a policy with **MANAGE** permissions by selecting + **Add**. Give the policy the name of **xx-name-eh-sap**, check **MANAGE**, and then select **Create**
6. Select your new policy after it has been created, and then select the copy button for the **CONNECTION STRING – PRIMARY KEY** entity
7. Paste the **CONNECTION STRING – PRIMARY KEY** entity into Notepad, this is needed later in the exercise
8. Leave all windows open

The diagram illustrates a data flow from a mobile application to an Event Hub. On the left, a black smartphone icon represents the **Mobile Application**. A blue arrow labeled **Receive** points from the mobile application to a circular icon representing the **API**. From the API, a grey arrow labeled **Send** points to a light blue rectangular box representing the **Service Bus Namespace**. Inside this namespace box, there is a sub-box for the **Event Hub**, labeled **<name>eh** and **Event Hub**, with the note **Queues streaming data**. The Service Bus Namespace icon is a blue cloud with a downward arrow, and the Event Hub icon is a blue square with a grid of yellow dots.



# Evaluating the performance of Event Hubs



## Lesson 02: Work with data streams by using Azure Stream Analytics





# What are data streams

## Data streams:

In the context of analytics, data streams are event data generated by sensors or other sources that can be analyzed by another technology

## Data stream processing approach:

There are two approaches. Reference data is streaming data that can be collected over time and persisted in storage as static data. In contrast, streaming data have relatively low storage requirements. And run computations in sliding windows

## Data streams are used to:

### Analyze data:

Continuously analyze data to detect issues and understand or respond to them

### Understand systems:

Understand component or system behavior under various conditions to fuel further enhancements of said system

### Trigger actions:

Trigger specific actions when certain thresholds are identified

## Event processing

The process of consuming data streams, analyzing them, and deriving actionable insights out of them is called Event Processing and has three distinct components:

<b>Event producer</b>	Examples include sensors or processes that generate data continuously such as a heart rate monitor or a highway toll lane sensor
<b>Event processor</b>	An engine to consume event data streams and deriving insights from them. Depending on the problem space, event processors either process one incoming event at a time (such as a heart rate monitor) or process multiple events at a time (such as a highway toll lane sensor)
<b>Event consumer</b>	An application which consumes the data and takes specific action based on the insights. Examples of event consumers include alert generation, dashboards, or even sending data to another event processing engine

## Processing events with Azure Stream Analytics

Microsoft Azure Stream Analytics is an event processing engine. It enables the consumption and analysis of high volumes of streaming data in real time

Source	Ingestion	Analytical engine	Destination
Sensors Systems Applications	Event Hubs IoT Hubs Azure Blob Store	Stream Analytics Query Language .NET SDK	Azure Data Lake Cosmos DB SQL Database Blob Store Power BI

## Lesson 02: Transform data by using Azure Stream Analytics



# Create Stream Analytics service

Job name

Subscription

Resource group

Location

Home > New > Stream Analytics job > New Stream Analytics job

### New Stream Analytics job


\* Job name  
cto-asa-job1 ✓

\* Subscription  
▼

\* Resource group  
cto\_rg ▼  
[Create new](#)

\* Location  
West Europe ▼

Hosting environment ⓘ  
**Cloud** Edge

Streaming units (1 to 120) ⓘ  
 6

# Create a Stream Analytics Job input

Event Hub

New input

>

\* Input alias

cto-asa-input01

✓

☐ Provide Event Hub settings manually

☒ Select Event Hub from your subscriptions

Subscription

LearnAI Training Subscription

▼

\* Event Hub namespace ⓘ

cto-eh-ns

▼

\* Event Hub name ⓘ

☐ Create new ☒ Use existing

cto-name-eh

▼

\* Event Hub policy name ⓘ

RootManageSharedAccessKey

▼

Event Hub policy key

.....

Event Hub consumer group ⓘ

\* Event serialization format ⓘ

JSON

▼

Encoding ⓘ

UTF-8

▼

Event compression type ⓘ

None

▼

# Create a Stream Analytics Job output

Home > Resource groups > cto\_rg > cto-asa-job1 > Outputs

## Outputs

**+ Add**

- Event Hub
- SQL Database
- Blob storage**
- Table storage
- Service Bus topic
- Service Bus queue
- Cosmos DB
- Power BI
- Data Lake Storage Gen1

SINK

### Blob storage

New output

\* Output alias  
cto-asa-output01 ✓

☐ Provide Blob storage settings manually  
☒ Select Blob storage from your subscriptions

Subscription  
LearnAI Training Subscription ▼

\* Storage account ⓘ  
ctoazureblob ▼

\* Storage account key ⓘ  
.....

\* Container ⓘ  
☒ Create new ☐ Use existing

\* socialmedia ✓

Path pattern ⓘ  
..... ✓

Date format  
YYYY/MM/DD ▼

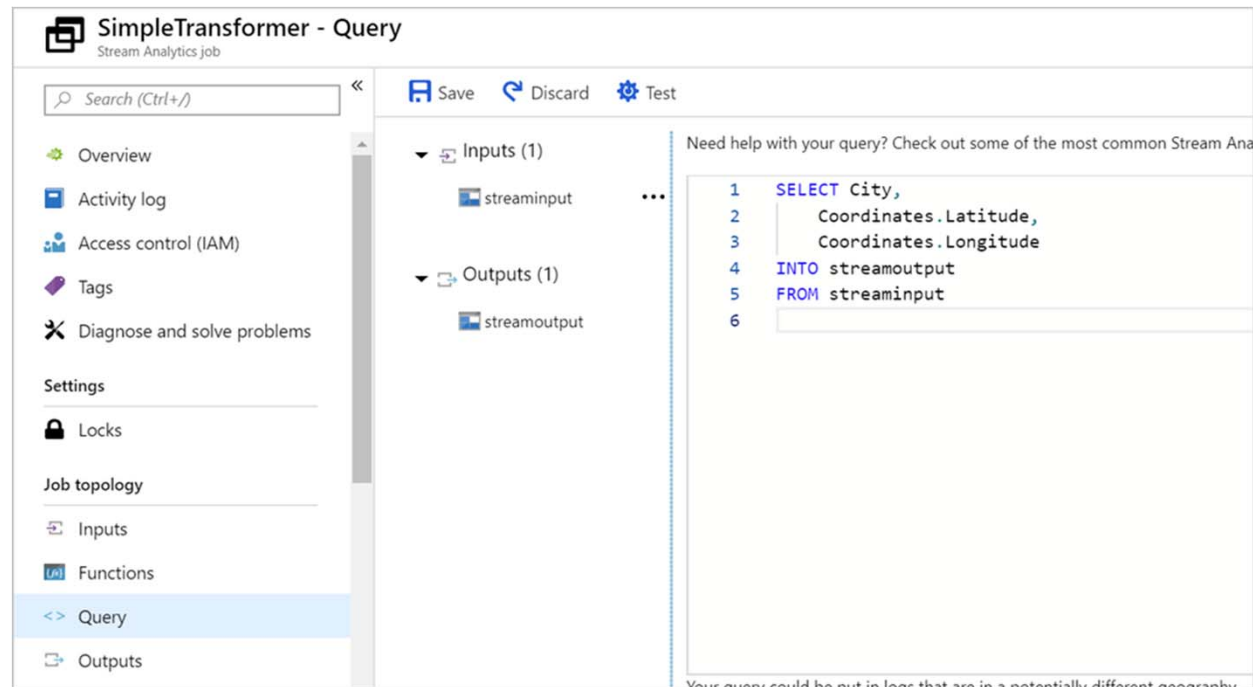
Time format  
HH ▼

\* Event serialization format ⓘ  
JSON ▼

Encoding ⓘ  
UTF-8 ▼

# Transform data by using Azure Stream Analytics

- Declarative SQL like language
- Contains testing capabilities
- Performs aggregations over windows of time





# Using Windowing functions with Azure Stream Analytics

Performing operations on the data contained in temporal windows is a common pattern. Stream Analytics has native support for windowing function including

Tumbling	Hopping	Sliding	Session
<p>Tumbling window functions are used to segment a data stream into distinct time segments and perform a function against them.</p> <p>For example, tell me the count of tweets per time zone every 10 seconds</p>	<p>Hopping window functions hop forward in time by a fixed period.</p> <p>For example, every 5 seconds, give me the count of tweets for the last 10 seconds</p>	<p>Sliding windows, output events only for points in time when the content of the window actually changes.</p> <p>For example, give a count of all tweets which are tweeted more than 10 times on a given topic</p>	<p>Session window functions group events that arrive at similar times, filtering out periods of time where there is no data.</p> <p>For example, give the count of tweets that occur within 5 minutes of each other</p>

## Windowing functions examples with Azure Stream Analytics

Tumbling	Hopping	Sliding	Session
<pre>SELECT   Timezone,   Count(*) as Count FROM   TwitterStream TIMESTAMP BY   CreatedAt GROUP BY   Timezone TUMBLINGWINDOW (second, 10)</pre>	<pre>SELECT   Timezone,   Count(*) as TotalTweet FROM   TwitterStream TIMESTAMP BY   CreatedAt GROUP BY   Topic HOPPINGWINDOW (second, 10, 5)</pre>	<pre>SELECT   Topic,   Count(*) FROM   TwitterStream TIMESTAMP BY   CreatedAt GROUP BY   Topic SLIDINGWINDOW (second, 10) HAVING COUNT(*) &gt; 10</pre>	<pre>SELECT   Timezone,   Count(*) FROM   TwitterStream TIMESTAMP BY   CreatedAt GROUP BY   Topic SESSIONWINDOW (minute 5, 10)</pre>

# Start a Stream Analytics Job

cto-asa-job1  
Stream Analytics job

Search (Ctrl+J)

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Settings

Locks

Job topology

Inputs

Functions

Query

Outputs

Configure

Storage account settings

Start

Stop

Delete

Resource group (change) : cto\_rg

Status : Created

Location : West Europe

Subscription (change) : LearnAI Training Subscription

Subscription ID : 5be49961-ea44-42ec-8021-b728be90d58c

Send feedback

Created

Started

Output water

Hosting env

Inputs

1

cto-asa-input01

Outputs

1

cto-asa-output01

Query

1 SELECT

2 \*

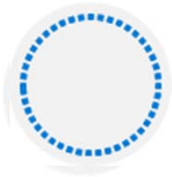
3 INTO

4 [cto-asa-output01]

5 FROM

6 [cto-asa-input01]

## Review questions



Q01 – Which technology typically provide an ingestion point for data streaming in an event processing solution that uses static data as a source?

A01 – Azure Blob storage

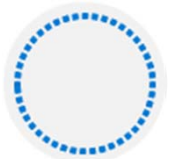
---



Q02 – Which three parts consist of an Azure Stream Analytics job pipeline

A02 – An input, a transformation query, and an output

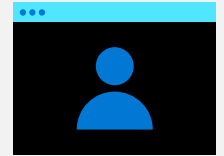
---



Q03 – What query language does Azure Stream Analytics use?

A03 – SAQL

# Lab: Real-time Stream Processing with Stream Analytics



## Lab overview

This lab teaches you how to process streaming data with Azure Stream Analytics. You will ingest vehicle telemetry data into Event Hubs, then process that data in real time, using various windowing functions in Azure Stream Analytics. You will output the data to Azure Synapse Analytics. Finally, you will learn how to scale the Stream Analytics job to increase throughput.

### Lab objectives

After completing this lab, you will be able to:

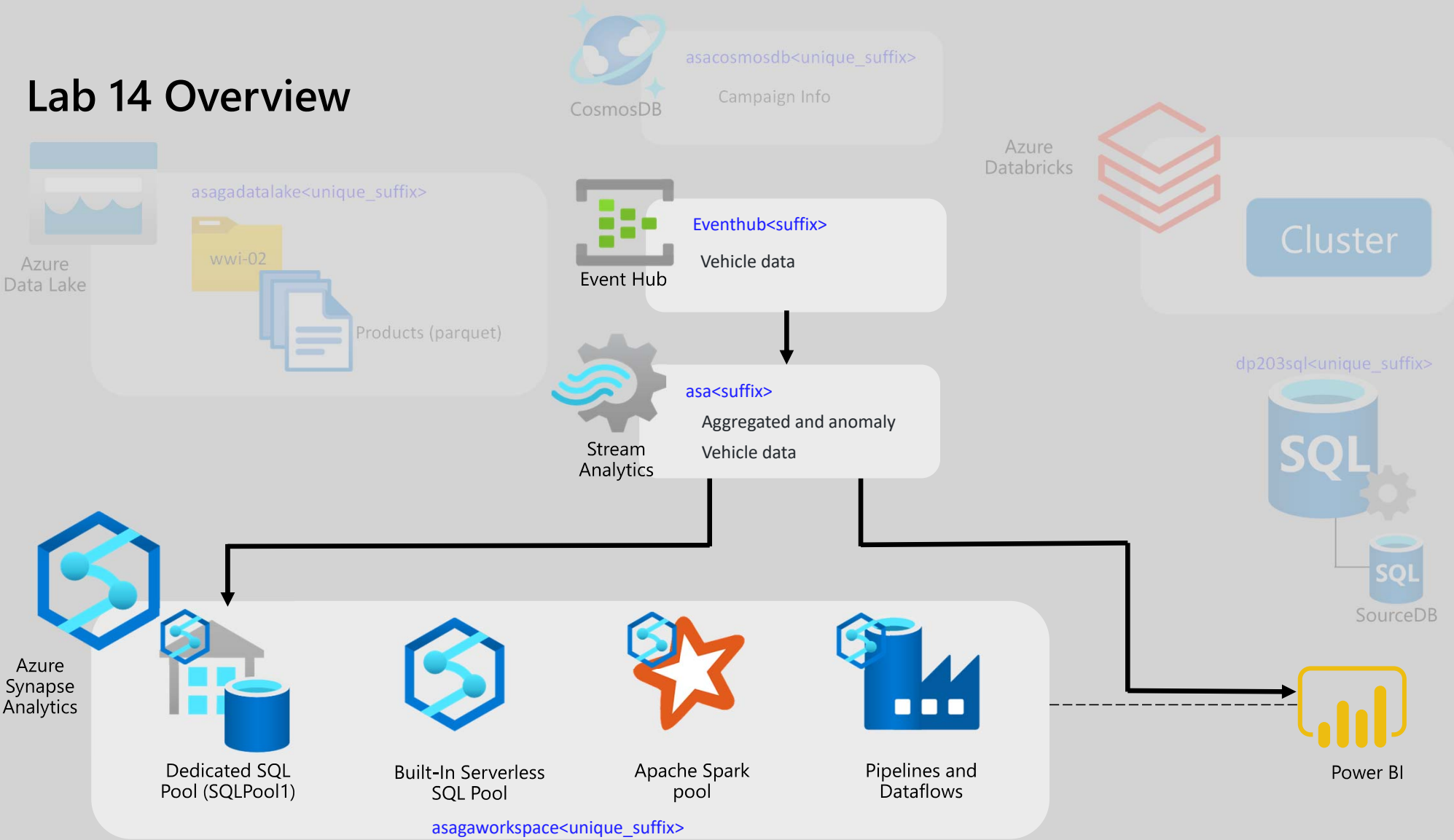
Use Stream Analytics to process real-time data from Event Hubs

Use Stream Analytics windowing functions to build aggregates and output to Synapse Analytics

Scale the Azure Stream Analytics job to increase throughput through partitioning

Repartition the stream input to optimize parallelization

# Lab 14 Overview

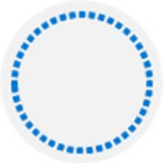


## Lab review



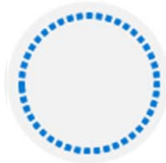
Q01 – How can you extend Azure Stream Analytics with custom code?

---



Q02 – What is Azure Event Hubs?

---



Q03 – What does the following command do: `GROUP BY TumblingWindow(Duration(second, 2))`



## Module summary

In this module, you have learned about:

Azure Stream Analytics

Azure Event Hubs

Windowing Functions

Partitioning

Azure Synapse Analytics

## Next steps

After the course, consider visiting the website that explores a [[Azure Stream Analytics solution patterns](#)] Azure Stream Analytics solution pattern and the associated documentation that goes into more depth about this architecture.

