

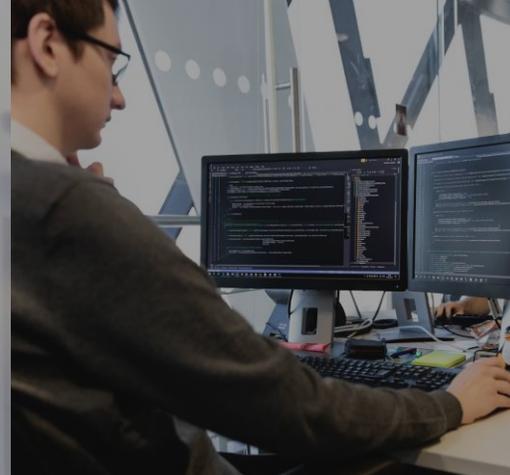


# DP-200T01: Performing Real- Time Analytics with Stream Analytics



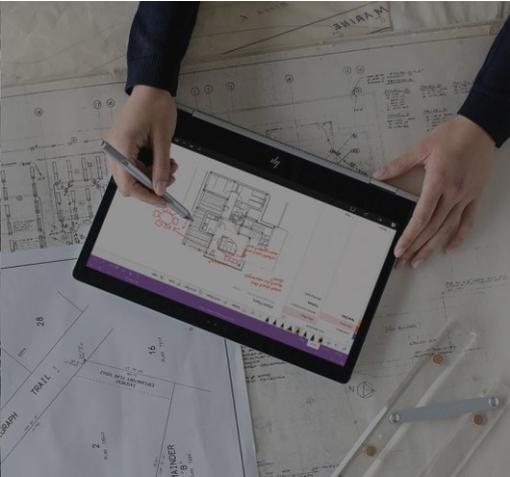
# Agenda

- L01 - Data streams and event processing
- L02 - Data ingestion with Event Hubs
- L03 - Processing data with Stream Analytics Jobs



# Lesson 01

## Data Streams and Event Processing



# Lesson Objectives

- Explain data streams
- Explain event processing
- Learn about processing events with 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:

## Event producer

Examples include sensors or processes that generate data continuously such as a heart rate monitor or a highway toll lane sensor

## Event processor

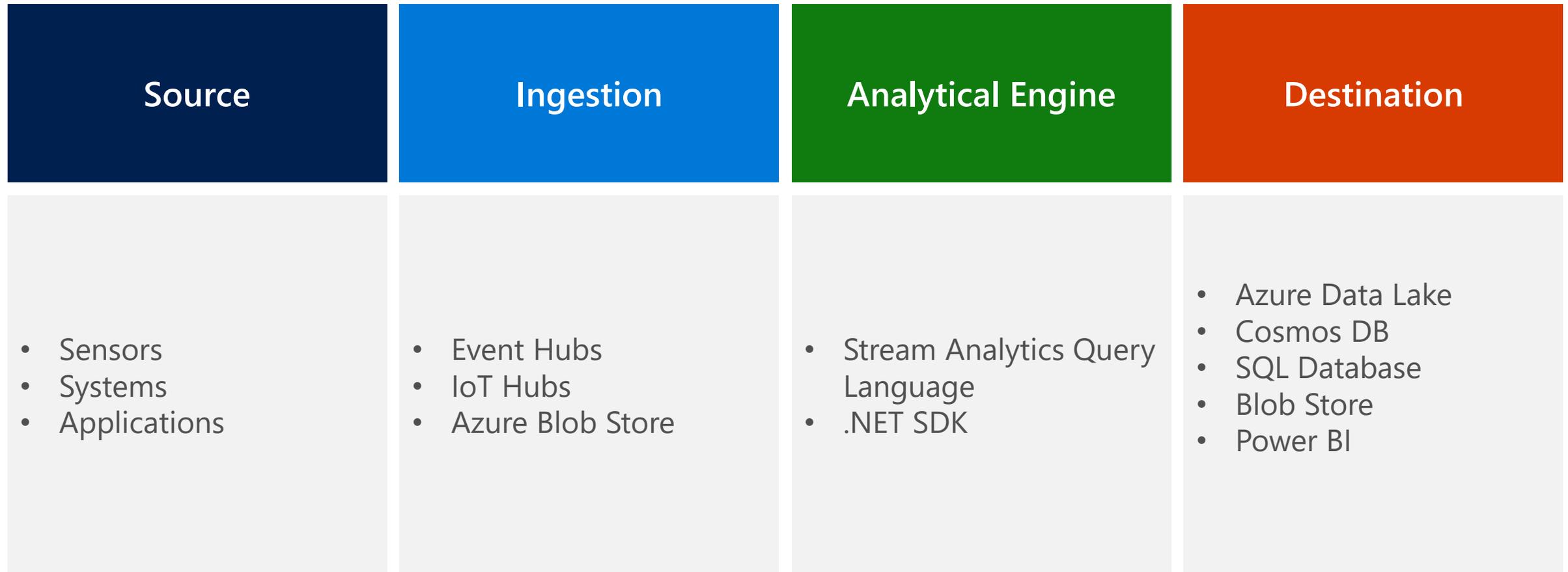
An engine to consume event data streams and deriving insights from them. Depending on the problem space, event processors either process one incoming event at a time (such as a heart rate monitor) or process multiple events at a time (such as a highway toll lane sensor)

## Event consumer

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

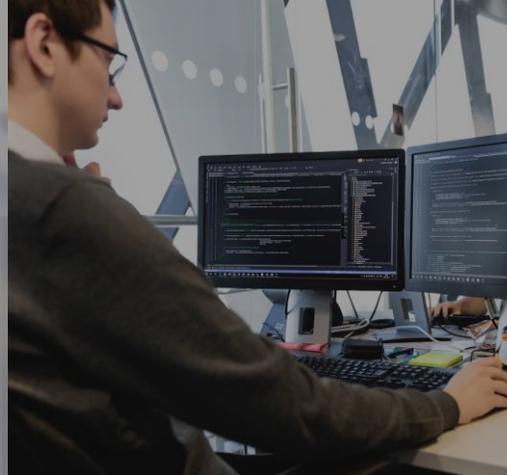
# Processing events with Azure Stream Analytics

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



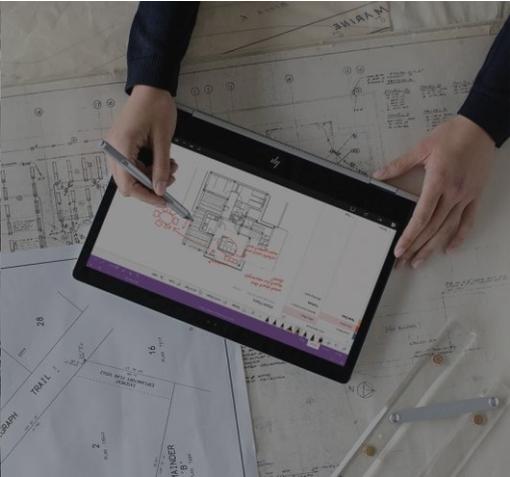
# Review Questions

- Q01 – Which of the following technologies typically provide an ingestion point for data streaming in an event processing solution that uses static data as a source?
- A01 – Azure Blob storage



# Lesson 02

## Data Ingestion with Event Hubs

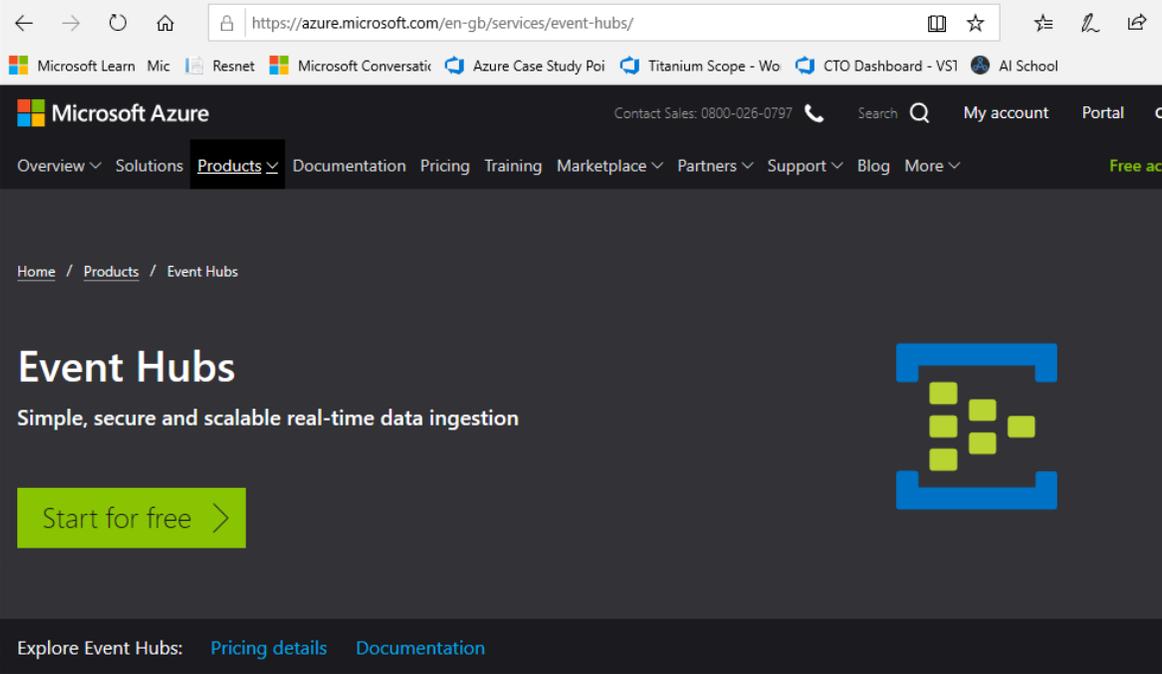


# Lesson Objectives

- Describe Azure Event Hubs
- Create an Event Hub
- Evaluate the performance of an Event Hub
- Configure applications to use an Event Hub

# 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."*

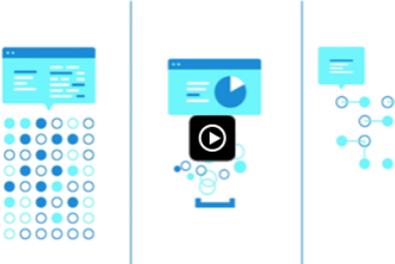


The screenshot shows the Microsoft Azure website for the Event Hubs product. The browser address bar displays <https://azure.microsoft.com/en-gb/services/event-hubs/>. The navigation menu includes Overview, Solutions, Products (highlighted), Documentation, Pricing, Training, Marketplace, Partners, Support, Blog, and More. The main content area features the heading "Event Hubs" with the subtext "Simple, secure and scalable real-time data ingestion" and a prominent "Start for free" button. Below this, there are links for "Pricing details" and "Documentation".

Event Hubs is a fully managed, real-time data ingestion service that's simple, trusted and scalable. Stream millions of events per second from any source to build dynamic data pipelines and immediately respond to business challenges. Keep processing data during emergencies using the [geo-disaster recovery](#) and [geo-replication](#) features.

Integrate seamlessly with other Azure services to unlock valuable insights. Allow existing Apache Kafka clients and applications to talk to Event Hubs without any code changes – you get a managed Kafka experience without having to manage your own clusters. Experience real-time data ingestion and microbatching on the same stream.

[Link to video >](#)



# Create an Event Hub

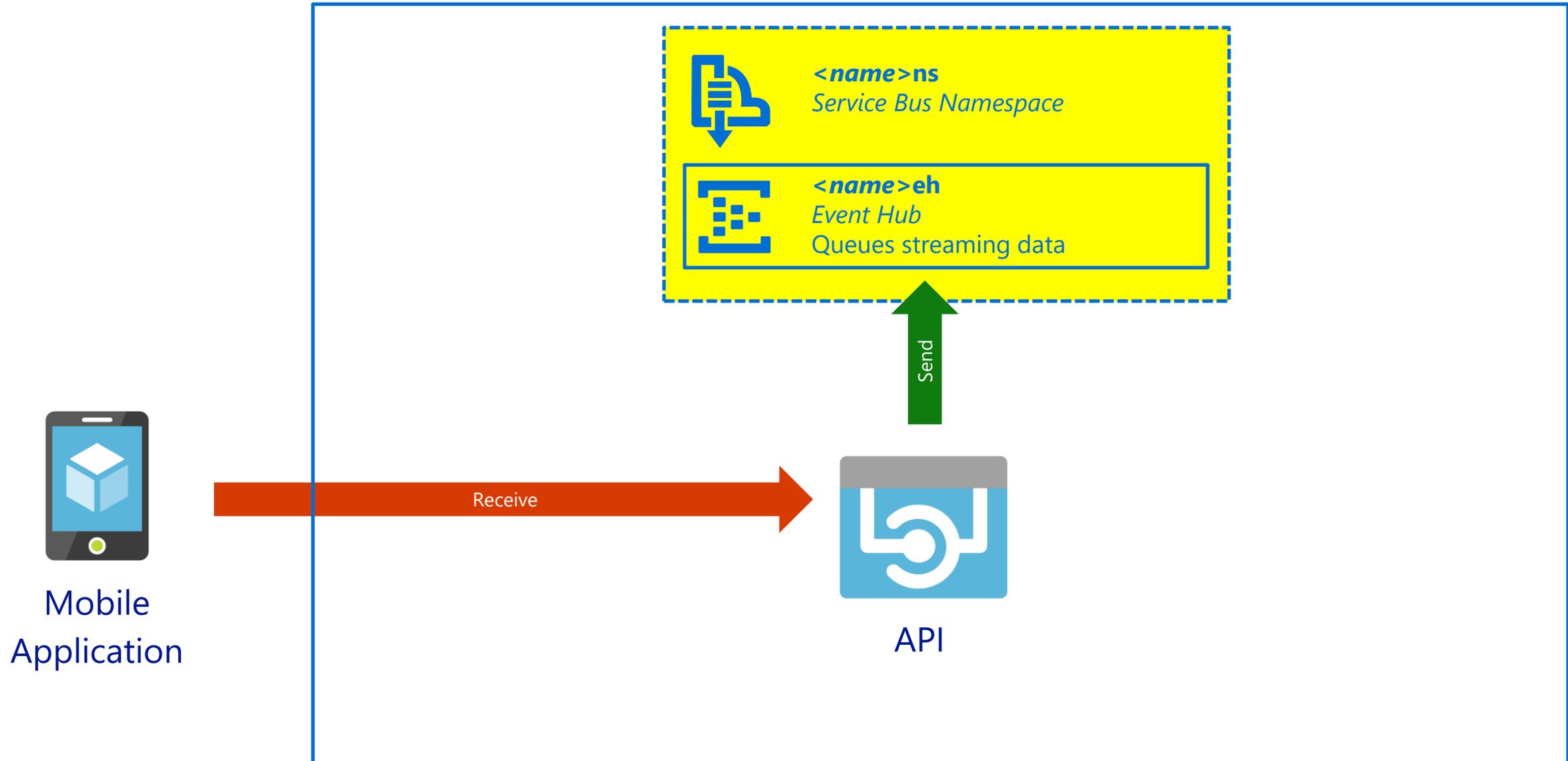
## Create an event hub namespace:

1. In the [Azure portal](#), select **NEW**, 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

# Configure Applications to use Event Hubs



# Evaluating the Performance of Event Hubs

Search (Ctrl+/)

- Overview
- Access control (IAM)
- Diagnose and solve problems

SETTINGS

- Shared access policies
- Properties
- Locks
- Automation script

ENTITIES

- Consumer groups

FEATURES

- Capture

SUPPORT + TROUBLESHOOTING

- New support request

+ Consumer group Delete

Resource group (change) **ehtestrg** Status Active Location UK West Subscription (change) Free Trial

Subscription ID d73c53de-f991-4b7d-ad7b-b2a01... Namespace ehntest78 Created Monday, 16 July 2018 Updated Monday, 16 July 2018

EVENT HUB CONTENTS 1 CONSUMER GROUP

EVENT HUB STATUS ACTIVE

MESSAGE RETENTION 7 DAYS

Show metrics data for the last: 1 hour 6 hours 12 hours 1 day 7 days 30 days

Requests

Messages

Throughput

INCOMING REQUESTS... EHNTTEST78 121

SUCCESSFUL REQUES... EHNTTEST78 121

SERVER ERRORS EHNTTEST78 --

INCOMING MESSAGES... EHNTTEST78 100

OUTGOING MESSAGES... EHNTTEST78 100

CAPTURED MES EHNTTEST78 --

INCOMING BYTES. (... EHNTTEST78 1.69kB

OUTGOING BYTES. (... EHNTTEST78 1.69kB

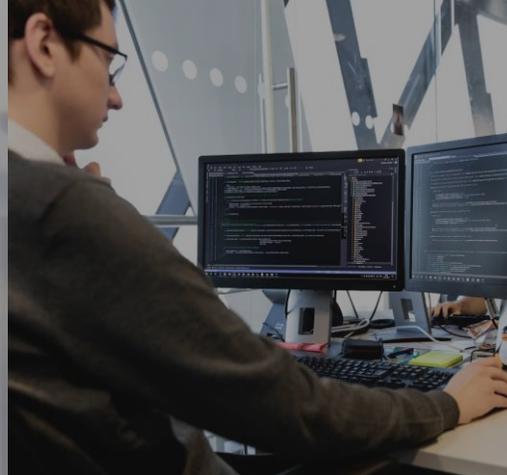
CAPTURED BYTE EHNTTEST78 --

Search to filter items...

NAME	LOCATION
\$Default	UK West

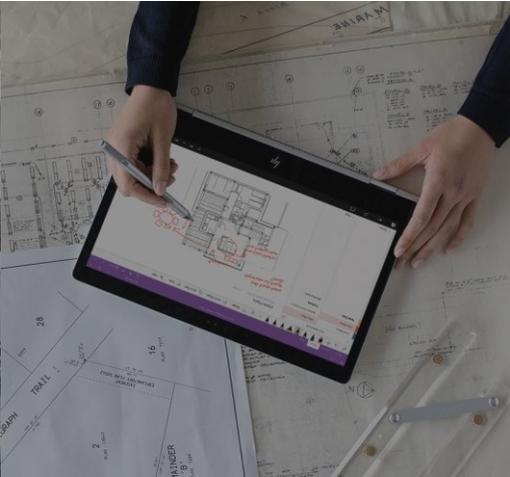
# Review Questions

- Q01 - Applications that publish messages to Azure Event Hub very frequently will get the best performance using Advanced Message Queuing Protocol (AMQP) because it establishes a persistent socket. True or False?
  - A01 – True
- Q02 - By default, how many partitions will a new Event Hub have?
  - A02 – 2
- Q02 - If an Event Hub goes offline before a consumer group can process the events it holds, those events will be lost. True or False?
  - A02 – False



# Lesson 03

## Processing Data with Stream Analytics Jobs

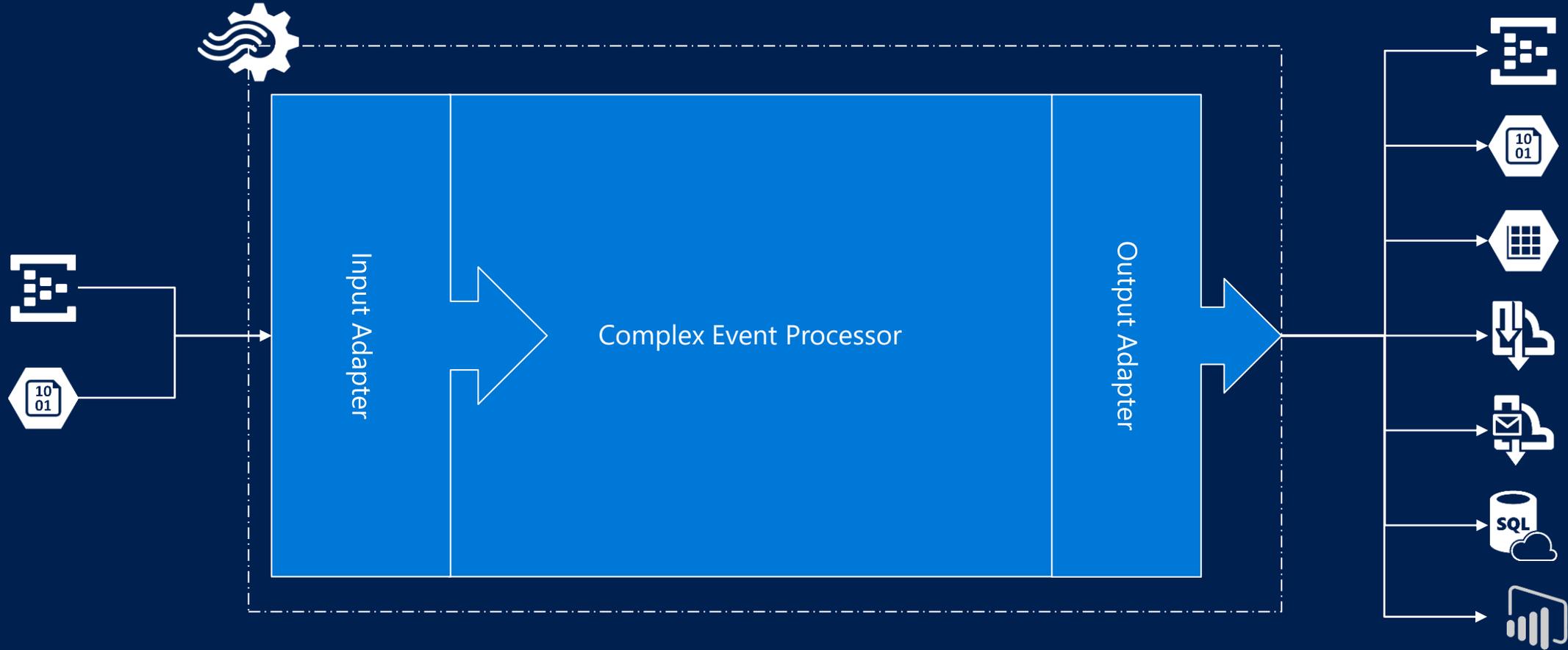


# Lesson Objectives

- Explore the Streaming Analytics workflow
- Create a Stream Analytics Job
- Configure a Stream Analytics job input
- Configure a Stream Analytics job output
- Write a transformation query
- Start a Stream Analytics job

# Azure Stream Analytics Workflow

*Complex Event Processing of Stream Data in Azure*



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

## Blob storage

New output

\* Output alias  ✓

Provide Blob storage settings manually  
 Select Blob storage from your subscriptions

Subscription  ▼

\* Storage account ⓘ  ▼

\* Storage account key

\* Container  
 Create new  Use existing

\* ⓘ  ✓

Path pattern ⓘ  ✓

Date format  ▼

Time format  ▼

\* Event serialization format ⓘ  ▼

Encoding ⓘ

# Write a transformation query

The screenshot shows the configuration page for a Stream Analytics job named 'cto-asa-job1'. The interface includes a left-hand navigation menu with options like Overview, Activity log, Access control, Tags, and Diagnose and solve problems. The main area displays job details such as Resource group (cto\_rg), Status (Created), Location (West Europe), Subscription (LearnAI Training Subscription), and Subscription ID (5be49961-ea44-42ec-8021-b728be90d58c). Below these details, there are sections for Inputs and Outputs, each with one item listed: 'cto-asa-input01' and 'cto-asa-output01'. A red box highlights the 'Query' section, which contains the following SQL query:

```
1 SELECT
2   *
3 INTO
4   [cto-asa-output01]
5 FROM
6   [cto-asa-input01]
```

# Start a Stream Analytics Job

The screenshot displays the Azure Stream Analytics job configuration interface for 'cto-asa-job1'. The left sidebar contains navigation options: Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings (Locks), Job topology (Inputs, Functions, Query, Outputs), and Configure (Storage account settings). The main area shows job details and configuration options.

**Job Details:**

- Resource group (change): [cto\\_rg](#)
- Status: Created
- Location: West Europe
- Subscription (change): [LearnAI Training Subscription](#)
- Subscription ID: 5be49961-ea44-42ec-8021-b728be90d58c

**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 job input consumes data streams from applications at low latencies and high throughput?
  - A01 – Event Hubs
- Q02 – Streaming Analytics Query Language is a subset of which query language
  - A02 – T-SQL
- Q03 – You are a Data Engineer for Contoso. You want to view key health metrics of your Stream Analytics jobs. Which tool in Streaming Analytics should you use?
  - A03 – Dashboards

# Lab: Performing Real-Time Analytics with Stream Analytics



# Lab overview

The students will be able to describe what data streams are and how event processing works and choose an appropriate data stream ingestion technology for the AdventureWorks case study. They will provision the chosen ingestion technology and integrate this with Stream Analytics to create a solution that works with streaming data.

## Lab objectives

After completing this lab, you will be able to:

1. Explain data streams and event processing
2. Data Ingestion with Event Hubs
3. Processing Data with Stream Analytics Jobs

# Lab scenario

As part of the digital transformation project, you have been tasked by the CIO to help the marketing departments become more productive with aspects of their work. Over the last few years the marketing department has been using Twitter to amplify marketing message around the bicycle products that are sold.

Whilst the department can provide reach numbers post campaign, they are unable to understand who is interacting with their campaigns in real-time, as the volumes are difficult to track manually. As a result, they would like to implement a system that can track in real-time who is responding to their campaign.

At the end of this lab, you will have:

1. Explain data streams and event processing
2. Data Ingestion with Event Hubs
3. Processing Data with Stream Analytics Jobs

# Lab review

- Exercise 1 – In event processing, what would social media sites be classed?
- Exercise 2 – What methods are available for ingesting data? How are they different?
- Exercise 3 – Is it mandatory to define a job output?

# Module Summary >

## In this module, you have learned about:

- Data streams and event processing.
- Data Ingestion with Event Hubs.
- Processing Data with Stream Analytics Jobs.

## Next steps >

After the course, consider reading the [Reference architecture for real-time event processing with Microsoft Azure Stream Analytics](#)

