

Data Engineering in the Cloud

Work with data streams using Azure Stream Analytics

Xuemao Zhang
East Stroudsburg University

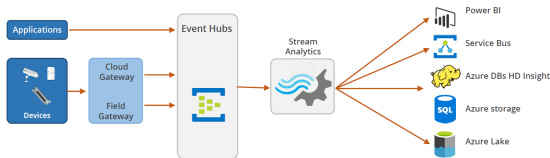
January 18, 2025

Outline

- Overview of Azure Event Hub
- Overview of Batch Processing
- Introduction to Real Time Processing
- Sending and Receiving Events
- Enabling Reliable Messaging Using Azure Event Hubs
- Working with Data Streams Using Azure Stream Analytics

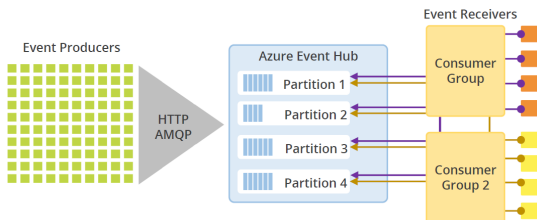
Overview of Azure Event Hub

- Business Scenario: You are hired as an Azure Developer. Your organization wants to capture streaming data for real time reporting and analysis of valuable insights. Your manager wants you to plan and execute the concepts of Azure Event Hub in the most efficient and effective manner.
- Event Hub in Microsoft Azure is a big data streaming platform and a service that does the ingestion for applications, which produces many events.
 - ▶ The events can be published in batches or individually.



Overview of Azure Event Hub

- In an Azure Event Hub, the events are transferred to partitions by making a connection using partition id or key.
- A consumer group consists of a group of producers representing the state of the entire hub.

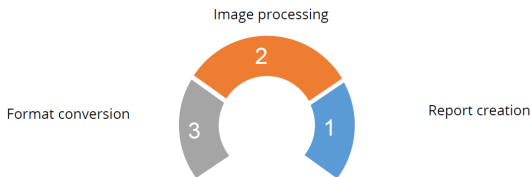


Overview of Azure Event Hub

- Throughput Units (TU): The capacity of the event hub in Azure is measured using the throughput units, the number of events, or the size of memory hub that may be taken in (Ingress) or out (Egress).
- Ingress: Taking in capacity. Each TU allows for:
 - ▶ Up to 1,000 events per second or
 - ▶ Up to 1 MB of data per second
- Egress: Sending capacity. Each TU allows for:
 - ▶ Up to 4,096 events per second or
 - ▶ Up to 2 MB of data per second

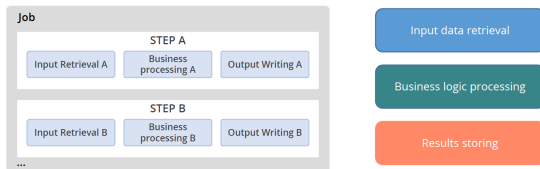
Overview of Batch Processing

- If your organization wants to send and receive events in real time via streaming to process millions of events per second, how do you plan and execute a method to achieve this in the most efficient and effective manner.
- Running batch jobs is called batch processing.



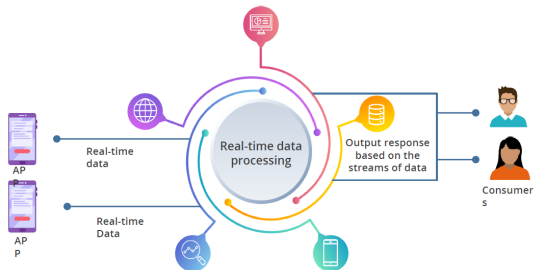
Overview of Batch Processing

- A chunk oriented step reads data from various sources, applies business logic, and stores the results in a destination.



Real Time Processing

- In real time processing, there is a rapid response from the program.
 - ▶ Google Maps is one of the famous applications where a real time processing method is used.



Real Time Processing

- Benefits of Real Time Processing



Sending and Receiving Events



Sending and Receiving Events

- **Azure Event Hub** is a fully managed, real-time data ingestion service.
 - ▶ Capable of streaming millions of events per second.

Setting Up Azure Event Hub:

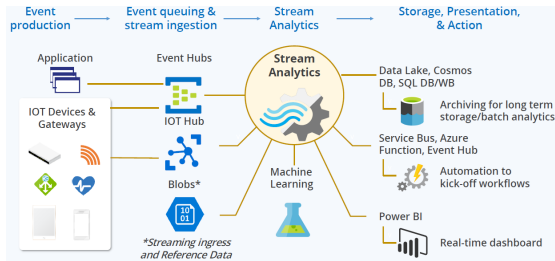
- Create Event Hub Namespace:
 - ▶ Use the Azure portal to create a namespace.
- Create Event Hub:
 - ▶ Define the Event Hub within the namespace.
- Configure Consumer Groups:
 - ▶ Allow multiple consumers to read the same event stream

Azure Stream Analytics

- Azure Stream Analytics is an engine that processes events. It includes two main parts:
 - ▶ Input source
 - ▶ Output source
- Input source is the source from which streaming data is produced.
- Output source is the destination which is configured to store results.

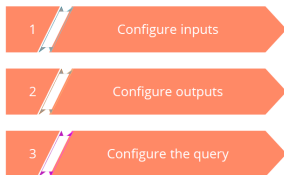
Azure Stream Analytics

- Steps for the processing of streaming data:



Azure Stream Analytics

- Steps to create a Stream Analytics job:

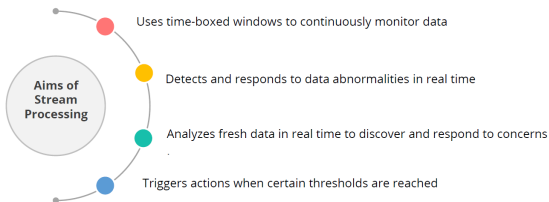


```
SELECT
    messageId,
    deviceId,
    pressure,
    pointInfo,
    EventEnqueuedUtcTime,
    EventProcessedUtcTime,
    PartitionId

INTO
    outputs
FROM
    inputs
```

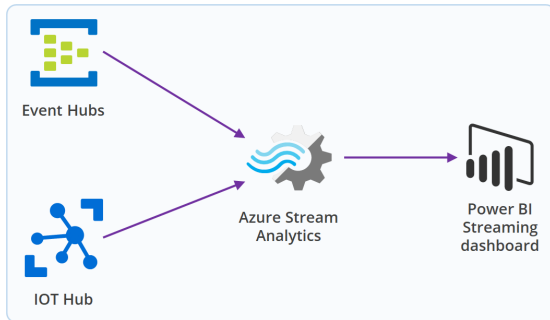
Azure Stream Analytics

- **Stream processing** is a mechanism for ingestion, manipulation, and data stream analysis. It is defined by the applications, sensors, and many other sources in real time for determining actionable insights.



Azure Stream Analytics

- Azure Stream Analytics can be used to ingest streaming data from Azure Event Hubs or Azure IoT Hub and then output the processed data to a Power BI streaming dashboard. This allows real-time data visualization and analytics in Power BI.



Azure Stream Analytics

Advantages of Azure Stream Analytics:

① Integrated Solutions

- Anomaly Detection: Built-in support for anomaly detection, allowing you to identify unusual patterns or outliers in your data streams in real-time.

② Native Azure Input and Output Adapters

• Seamless Integration:

- ▶ Native adapters for easy integration with various Azure services such as Event Hubs, IoT Hub, Blob Storage, SQL Database, Cosmos DB, Service Bus, and Power BI.
- ▶ Simplifies the process of setting up data pipelines and connecting different services.

Azure Stream Analytics

Advantages of Azure Stream Analytics:

3 Multiple Time Windows in the Same Query

- Flexibility in Time-based Analysis:

- ▶ Ability to use multiple time windows (e.g., tumbling, hopping, sliding windows) in a single query.
- ▶ Enables complex time-based analysis and aggregations.

4 Combining Multiple Temporal Operators

- Advanced Temporal Processing:

- ▶ Combine multiple temporal operators in arbitrary sequences to perform sophisticated stream processing tasks.
- ▶ Example: chaining windowed aggregates, temporal joins, and other functions to achieve complex data transformations.

Azure Stream Analytics

Advantages of Azure Stream Analytics:

5 Built-in Temporal Operators

- Comprehensive Set of Functions:

- ▶ Windowed Aggregates: Perform aggregations over specified time windows (e.g., AVG, SUM, COUNT).
- ▶ Temporal Joins: Join multiple data streams based on temporal conditions.
- ▶ Temporal Analytic Functions: Use built-in functions for advanced analytics, such as moving averages, lag, lead, etc.

6 Support for Lookup Tables and Geospatial Data

- Enhanced Contextual Analysis:

- ▶ Lookup Tables: Integrate lookup tables to enrich streaming data with reference data.
- ▶ Geospatial Reference Data: Merge streaming data with geospatial reference data for applications like geofencing and location-based analytics.
- ▶ Example: Tracking vehicle positions and generating alerts when vehicles enter or exit specific geofenced areas.

License



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](#).