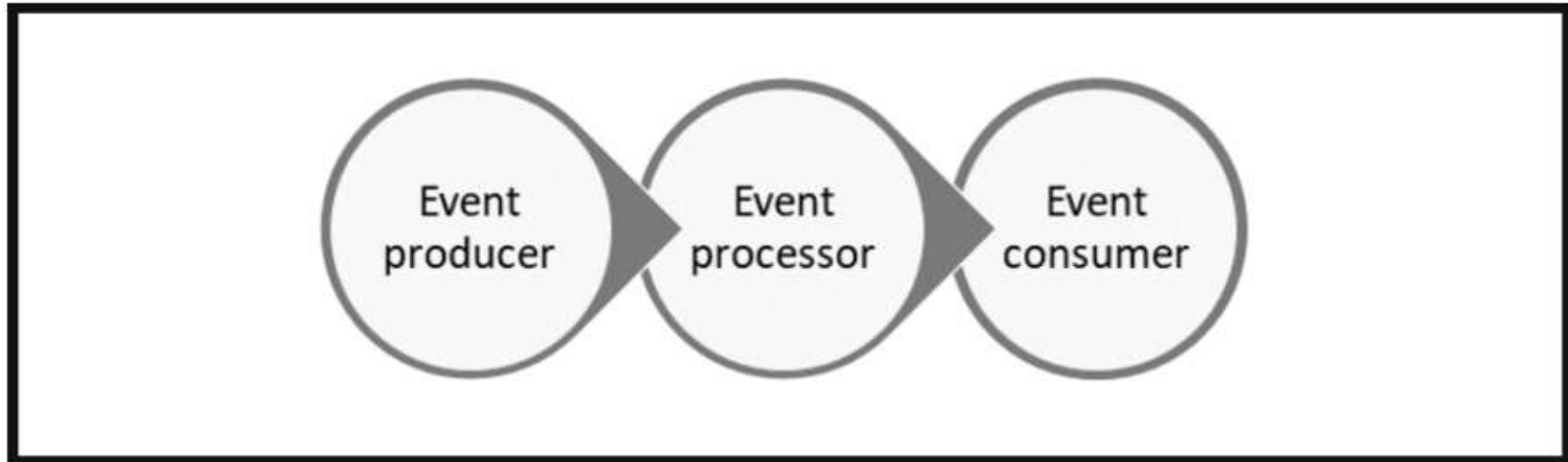


Azure Stream Analytics

Event Processing



Event Producer – Process that generate data continuously

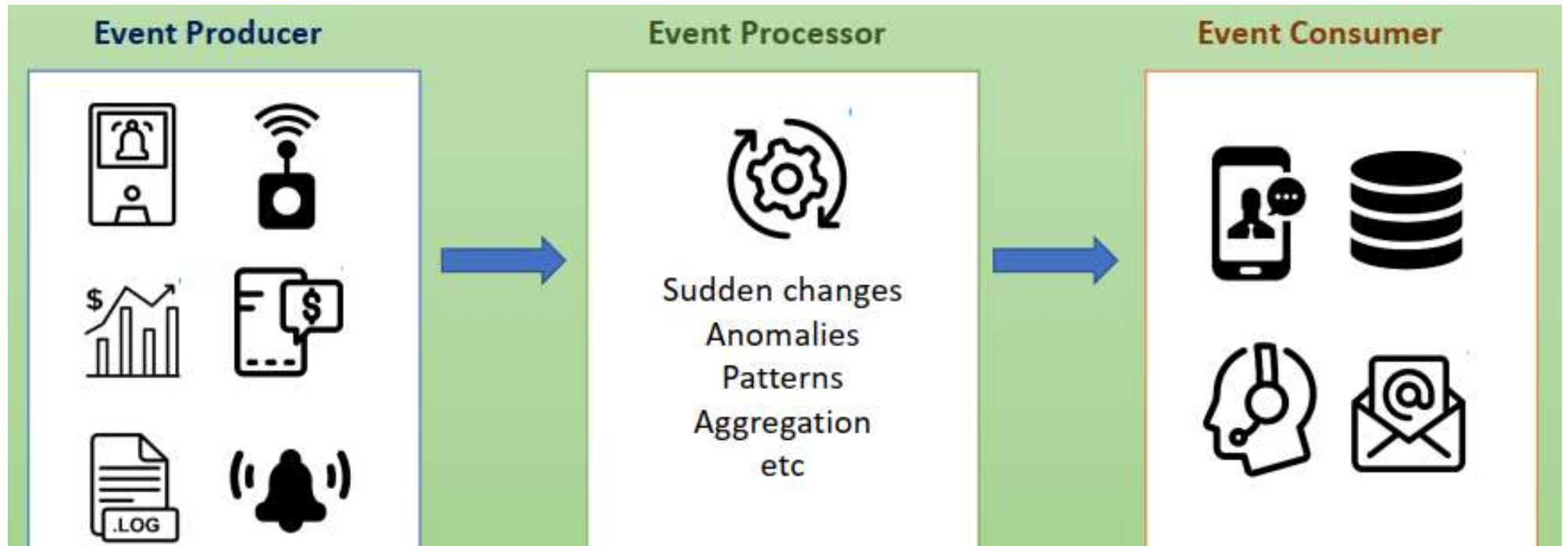


Event Processor - An engine to consume event data streams and derive insights from them. -



Event Consumer- An application that consumes the data and takes specific action based on the insights

Event Processing

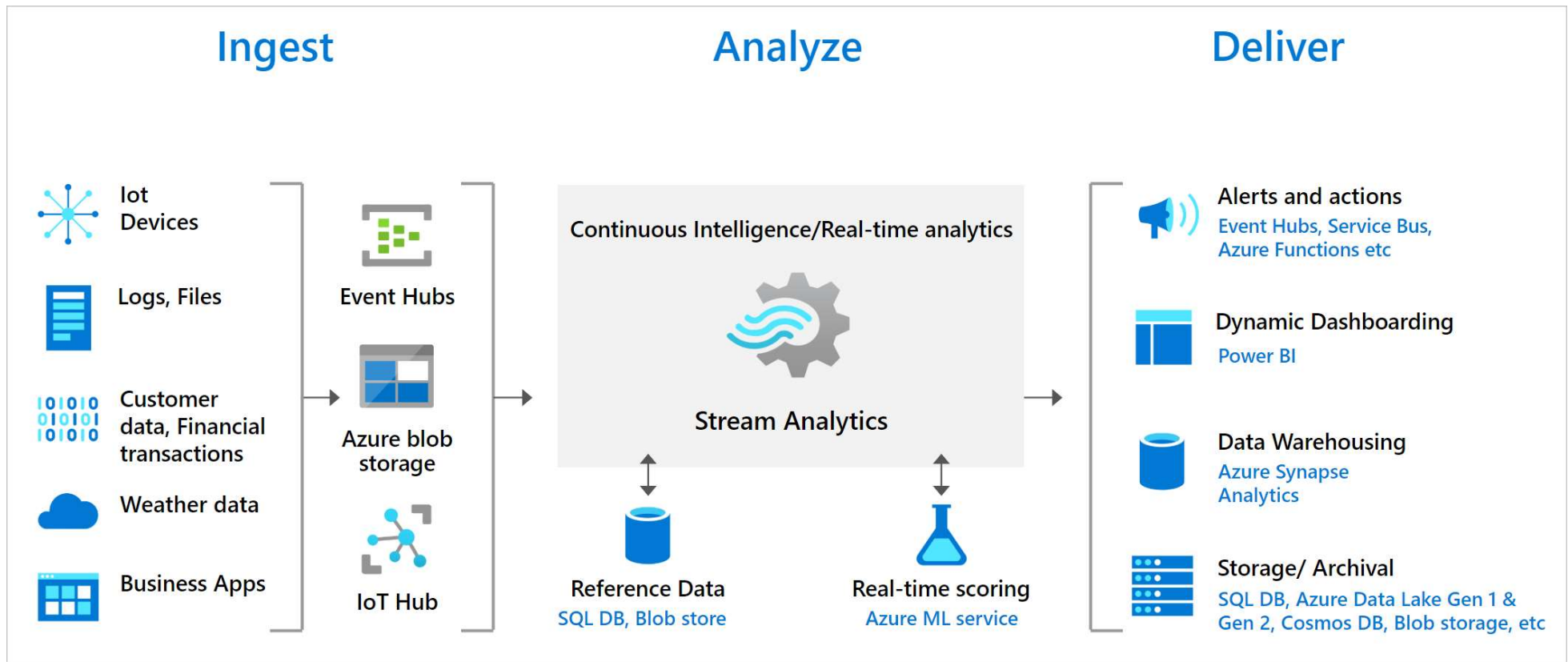


Azure Stream Analytics

- A real-time analytics and complex event-processing engine
- Designed to analyze and process high volumes of fast streaming data
- Can process data from multiple sources simultaneously



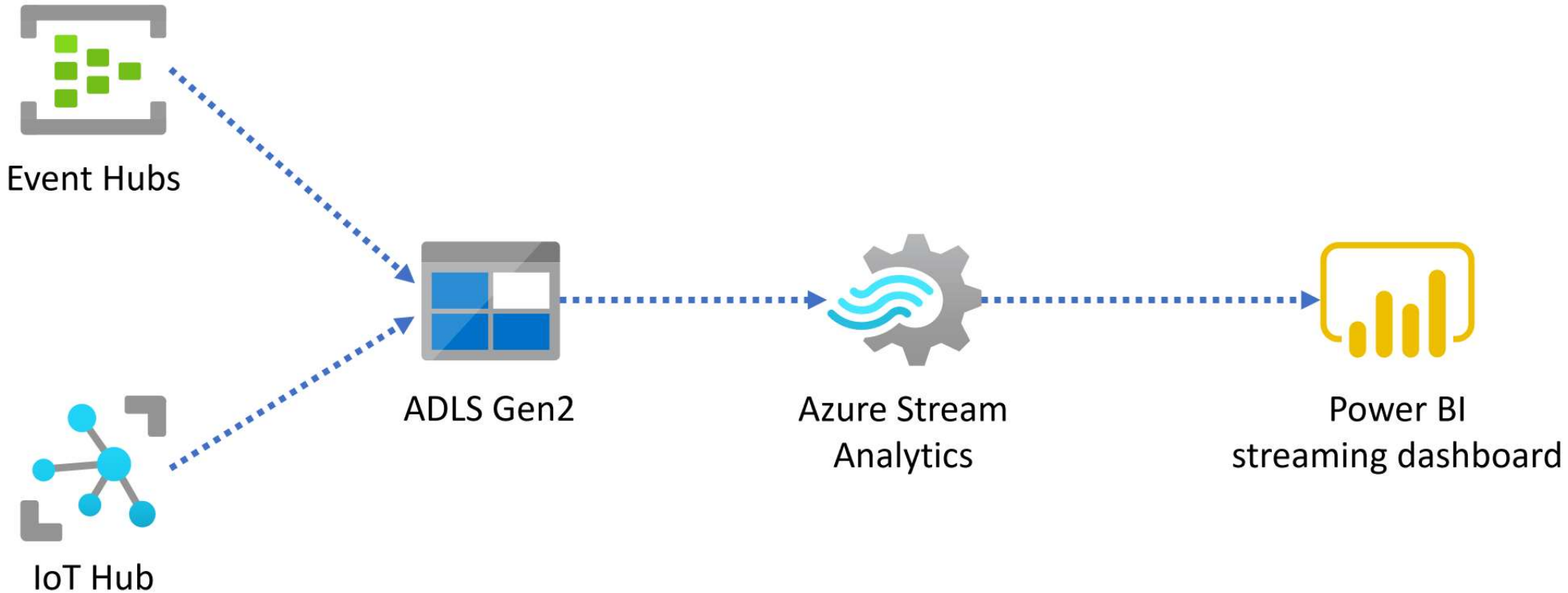
How does Stream Analytics work?



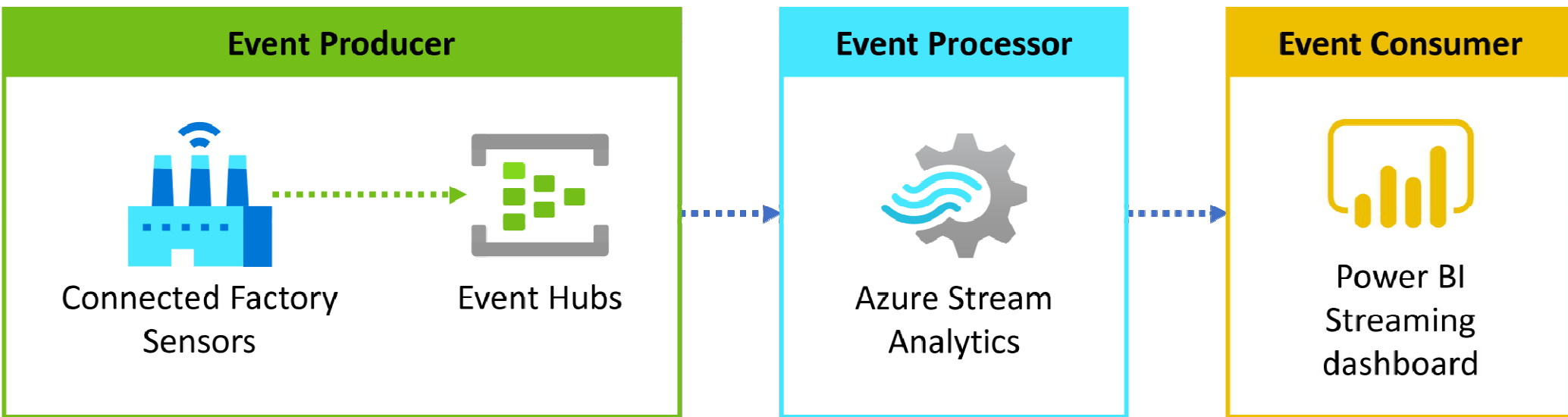
Approaches to data stream processing



Approaches to data stream processing



Understand event processing



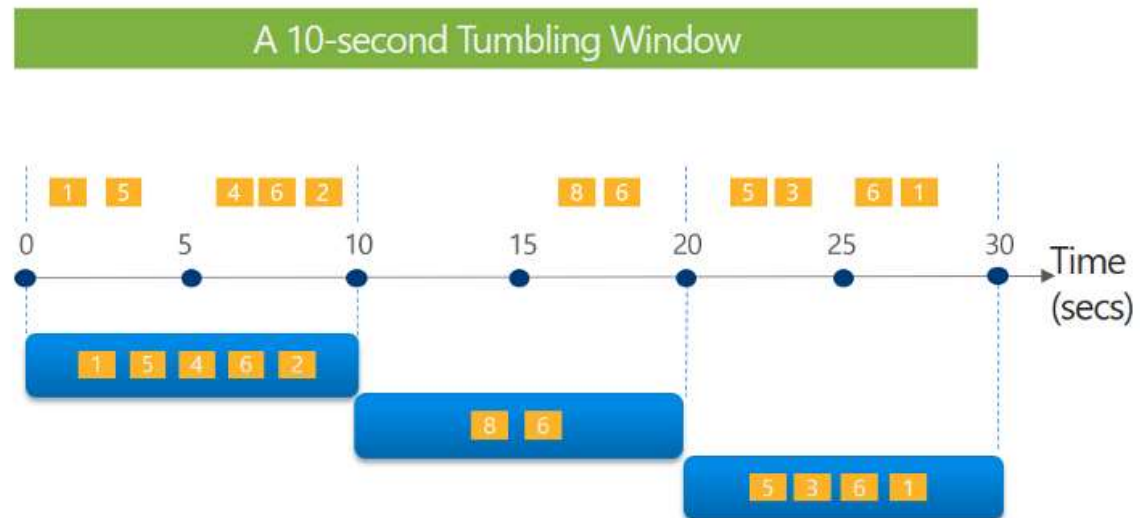
Windowing functions

Azure Stream Analytics Windowing

- Each data event has a timestamp
- There is an need to perform an operation (e.g. Count) on events falling in the same time window.
- Azure Stream Analytics achieve this through windows
- Types of window functions
 - Tumbling window
 - Hopping window
 - Sliding window
 - Session window

Tumbling Window

Tell me the count of tweets per time zone every 10 seconds

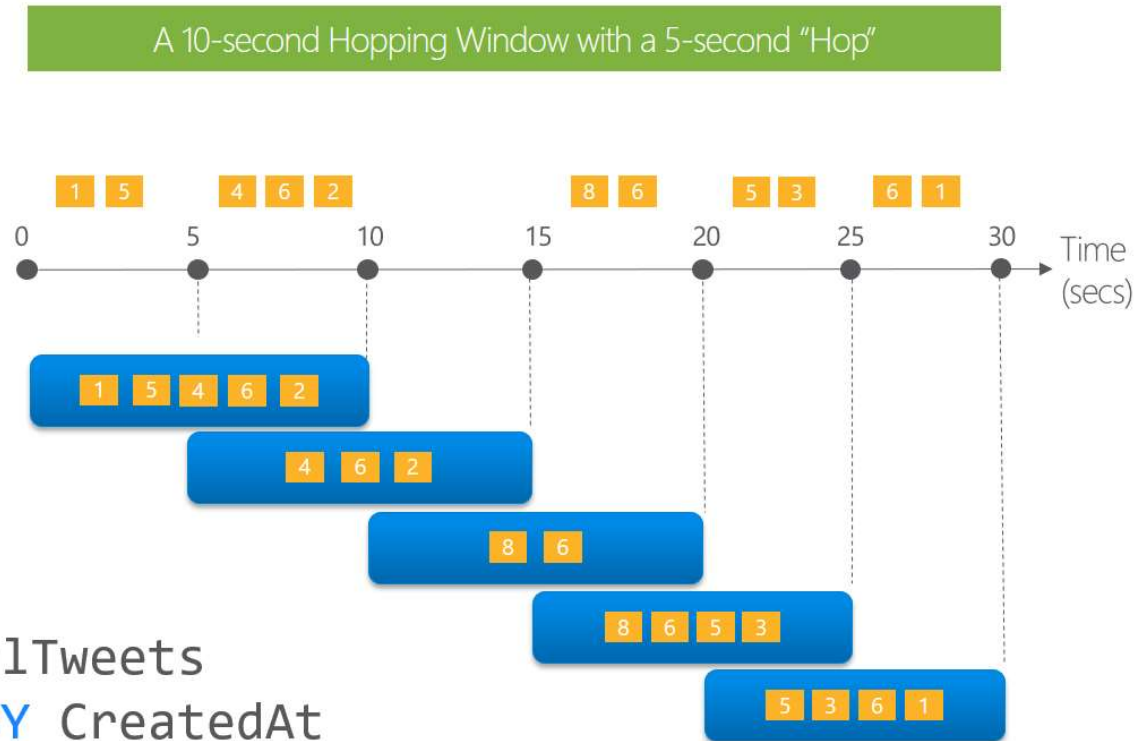


```
SELECT TimeZone, COUNT(*) AS Count
FROM TwitterStream TIMESTAMP BY CreatedAt
GROUP BY TimeZone, TumblingWindow(second,10)
```

Hopping Window

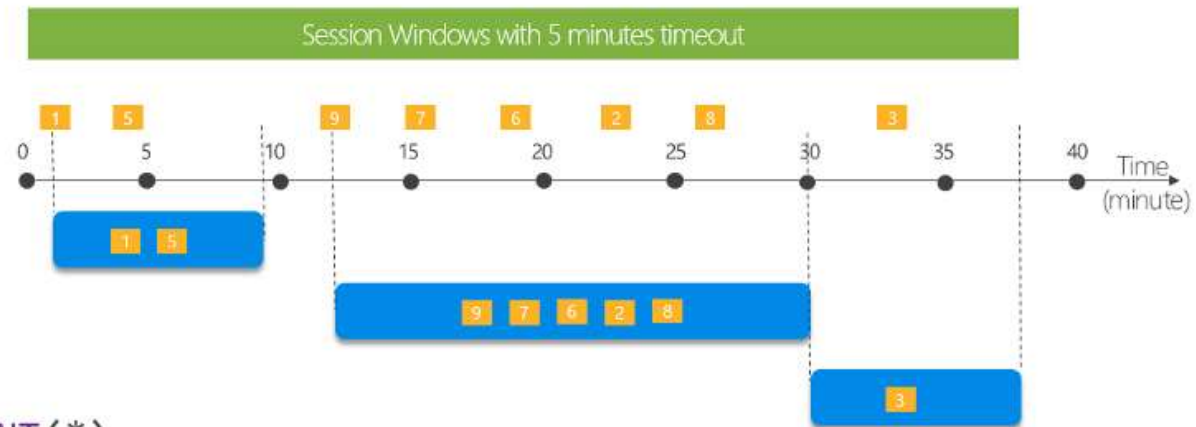
Every 5 seconds give me the count of tweets over the last 10 seconds

```
SELECT Topic, COUNT(*) AS TotalTweets
FROM TwitterStream TIMESTAMP BY CreatedAt
GROUP BY Topic, HoppingWindow(second, 10 , 5)
```



Session window

Tell me the count of tweets that occur within 5 minutes to each other.



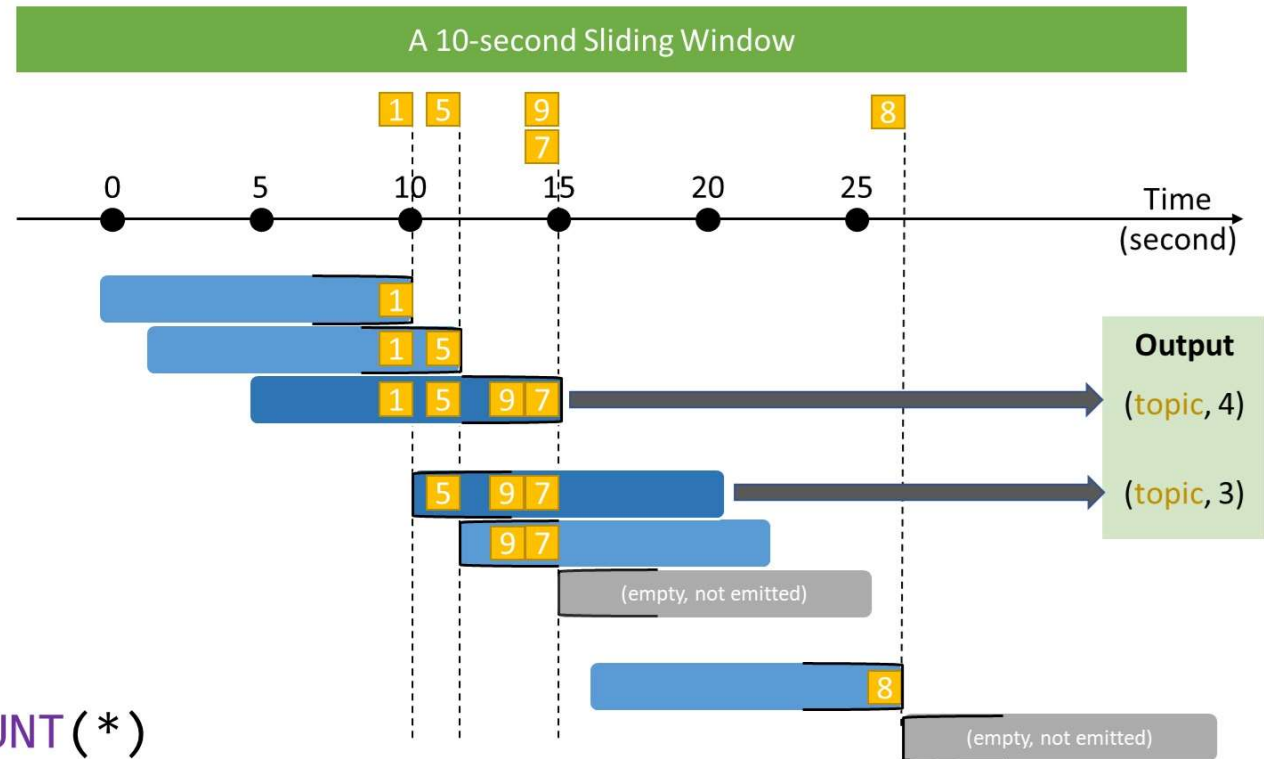
```
SELECT Topic, COUNT(*)  
FROM TwitterStream TIMESTAMP BY CreatedAt  
GROUP BY Topic, SessionWindow(minute, 5, 10)
```

Sliding Window

Alert me whenever a topic is mentioned more than 3 times in under 10 seconds

Note:

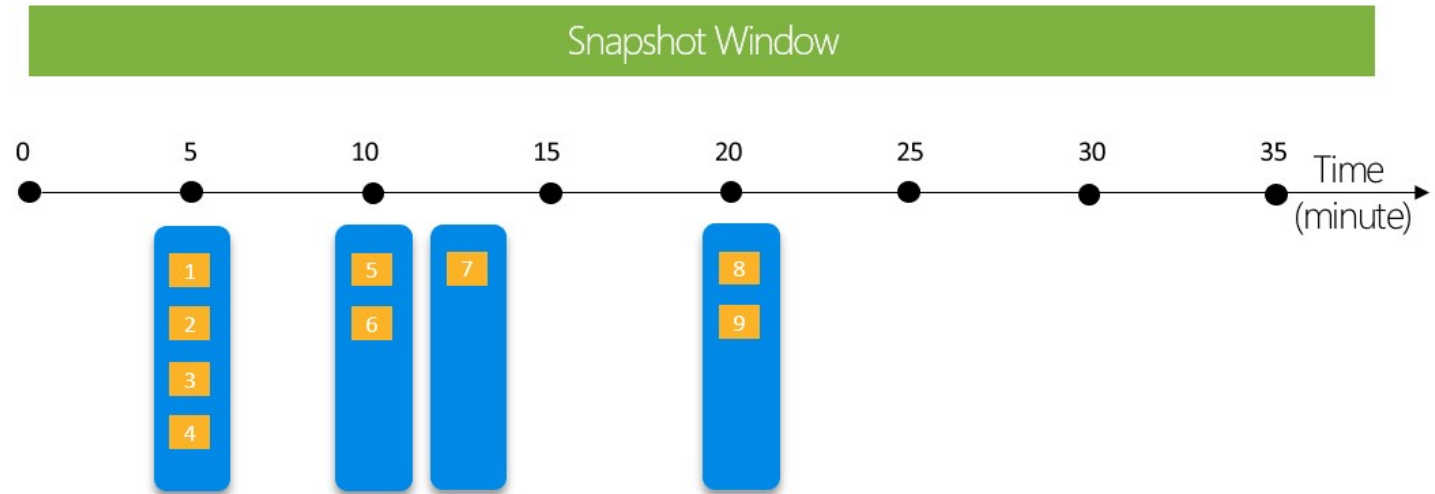
- all tweets on the diagram belong to the same topic



```
SELECT Topic, COUNT(*)  
FROM TwitterStream Timesteamp BY CreatedAt  
GROUP BY Topic, SlidingWindow(second, 10)  
HAVING COUNT(*) >= 3
```

Snapshot window

Give me the count of tweets with the same topic type that occur at exactly the same time



```
SELECT Topic, COUNT(*)  
FROM TwitterStream TIMESTAMP BY CreatedAt  
GROUP BY Topic, System.Timestamp()
```

Hands-On: Stream Analytics

Overview



Azure Blob
Storage
Introduction to blob storage

INPUT

OUTPUT

INPUT

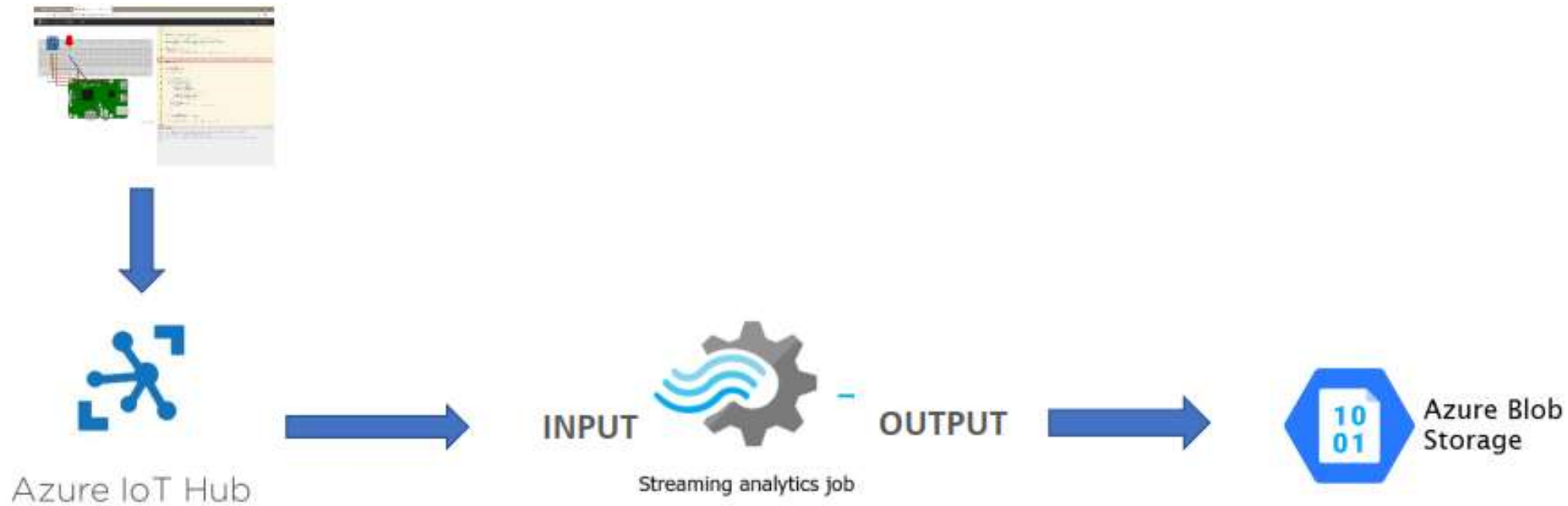
OUTPUT



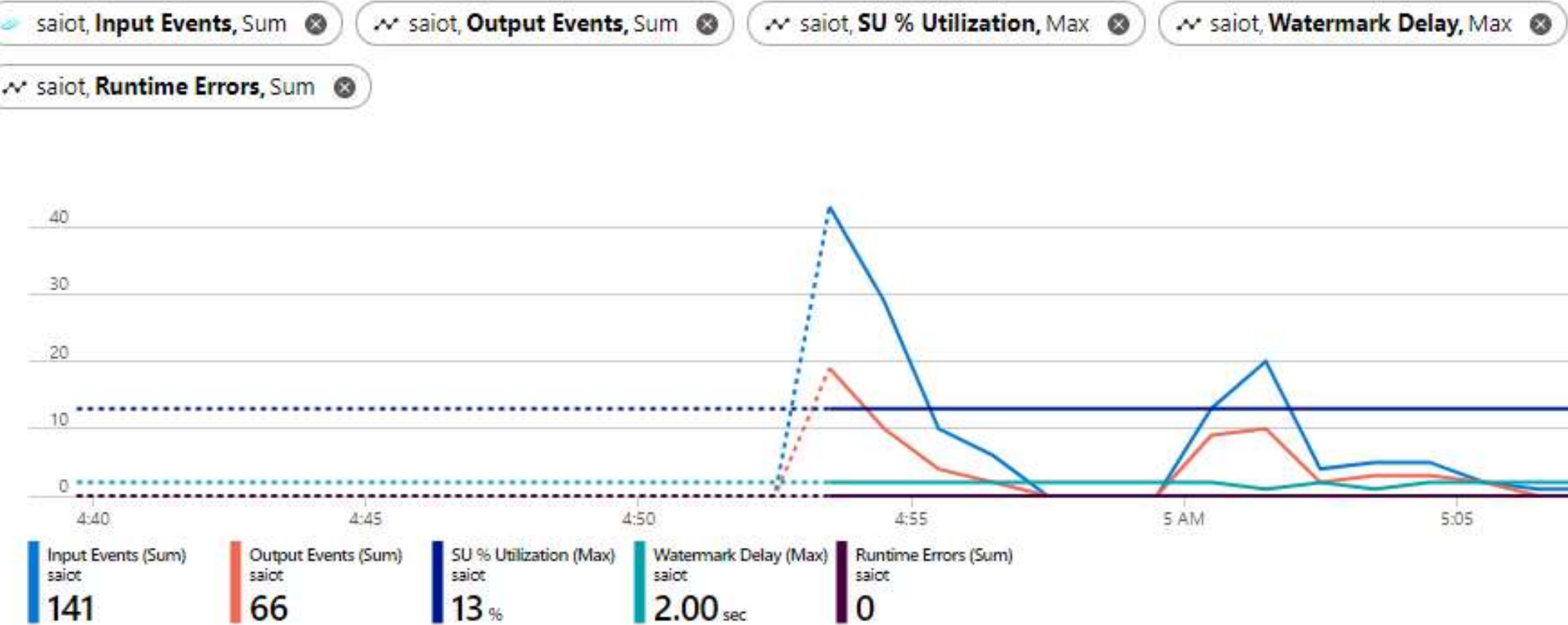
Streaming analytics job

Query (Processing Logic)

Overview



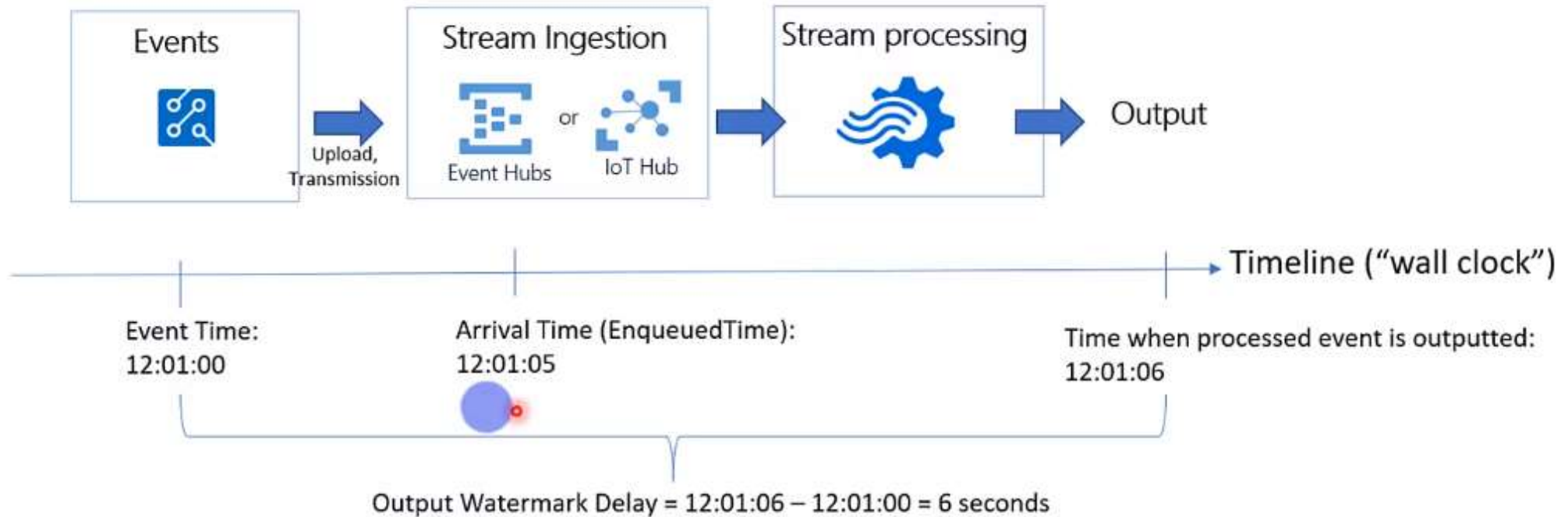
Stream Analytics Chart



Watermark delay metrics

Simple case: no time window, late arrival and out-of-order policy set to 10 seconds

```
SELECT *  
FROM input TIMESTAMP BY eventTime
```



TIMESTAMP BY (Azure Stream Analytics)

- All data stream events have a timestamp associated with them
- Events from IoT Hub are timestamped based on when the event was received
- Many streaming applications require using the exact timestamp that an event occurred, rather than the arrival time
- For example
 - In PoS application need event timestamps
- For these cases, the TIMESTAMP BY clause allows specifying custom timestamp values

TIMESTAMP BY (Azure Stream Analytics)

- SELECT deviceid, AVG(humidity) as avg_humidity, AVG(temperature) as avg_temperature
- INTO outputiot1
- FROM inputiothub TIMESTAMP BY EntryTime
- GROUP BY deviceid, HoppingWindow(second, 20, 2)

Thanks