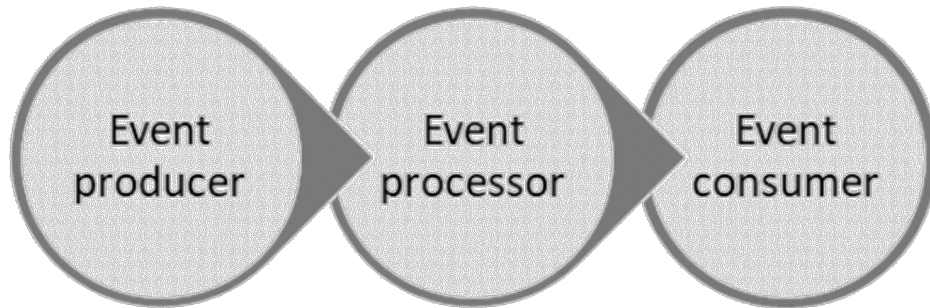




Azure Streaming Analytics

Event Processing



Event Producer – Process that generate data continuously

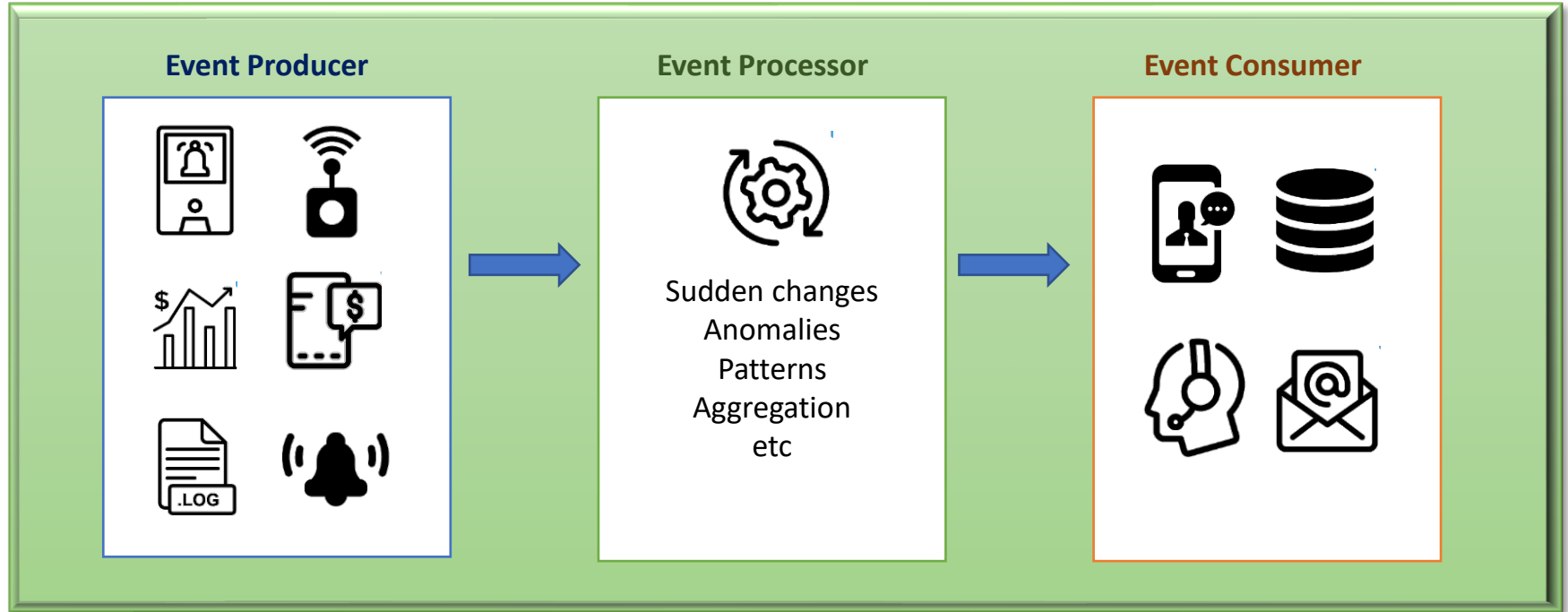


Event Processor - An engine to consume event data streams and derive insights from them. -



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

Live Event Processing

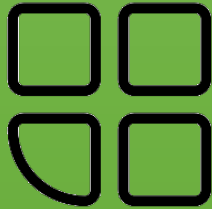




Challenges

Live Data Processing Challenges

- Data ingestion, processing and output should happen in **real-time**
- Support **high volume of data**
- Enough **processing power**
- Output storage should have **high bandwidth**
- Quick act on Output processing



* ASA FOR
CERTIFICATION

Azure options for Live Data Processing

HDInsight with Spark
Streaming

HDInsight with Storm

Apache Spark in Azure
Databricks

Azure Functions

WebJobs

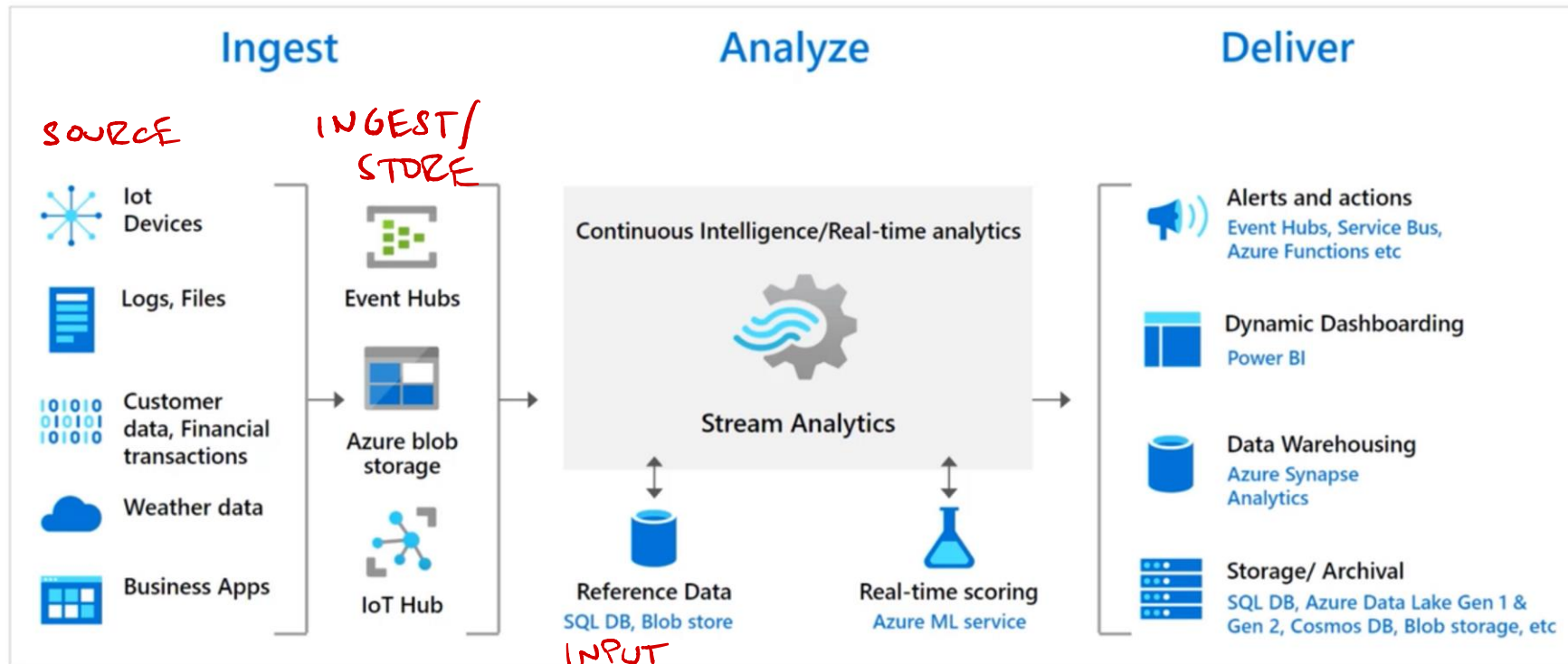
Azure Stream Analytics



Azure Stream Analytics

"A fully managed, real-time analytics service designed to process fast moving streams of data."

Azure Stream Analytics Data Flow

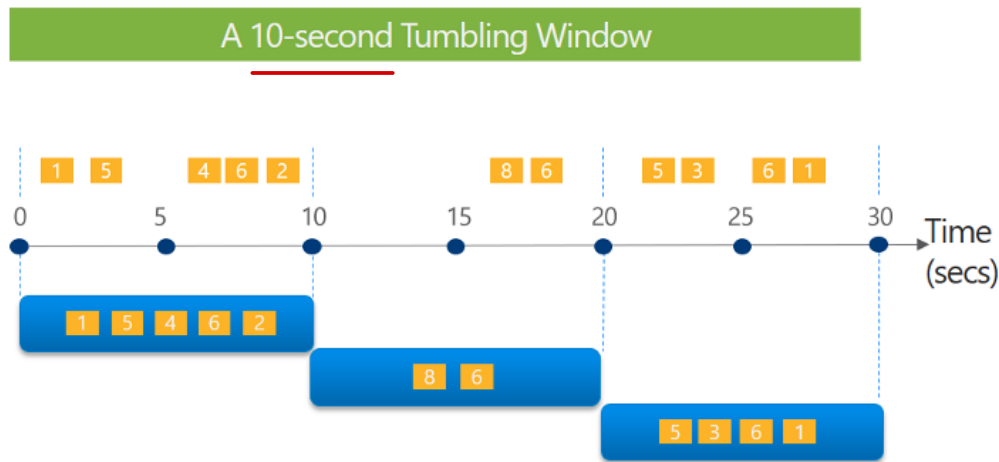


Azure Stream Analytics Windowing



- 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
- Four types of window functions
 - Tumbling window
 - Hopping window
 - Sliding window
 - Session 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)
```

TIME DIVIDED INTO
EVEN BUCKETS

TIME
UNIT

VALUE

HOPPING WINDOW

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

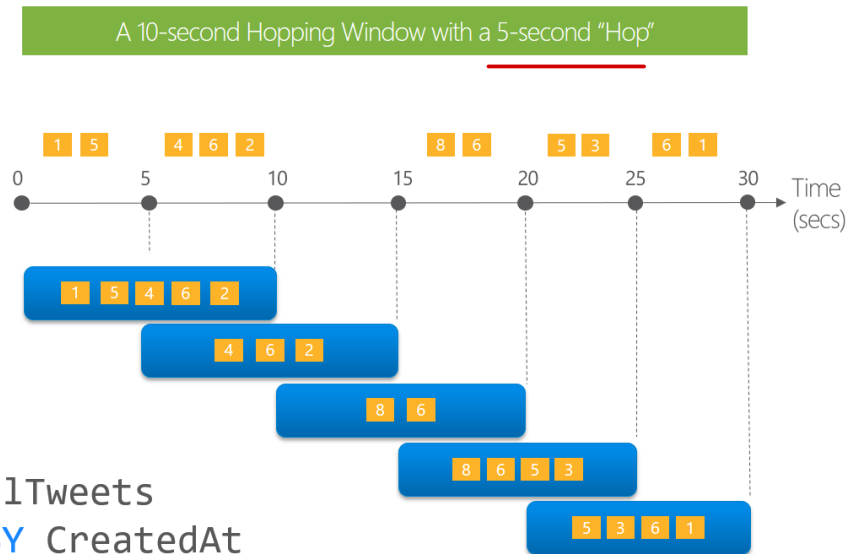
```
SELECT Topic, COUNT(*) AS TotalTweets
FROM TwitterStream TIMESTAMP BY CreatedAt
GROUP BY Topic, HoppingWindow(second, 10 , 5)
```

TIME
OVERLAPPED

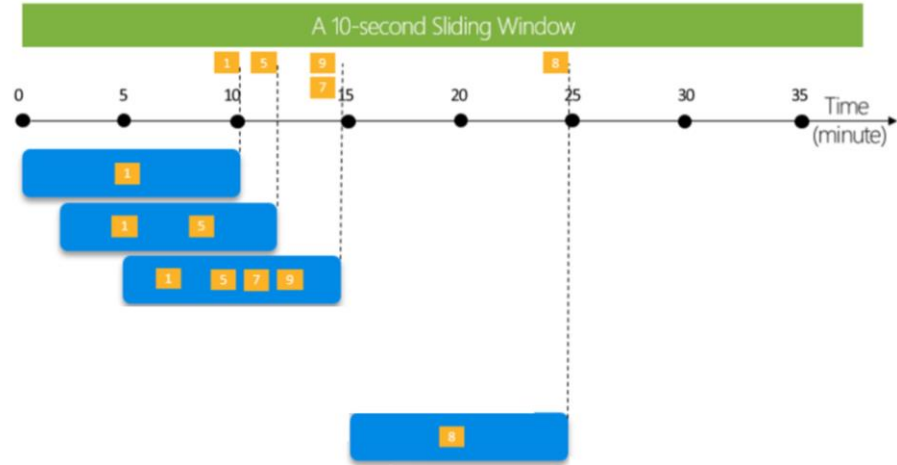
TIME
UNIT

WINDOW
LENGTH

HOP INTERVAL
(OVERLAP)



SLIDING WINDOW



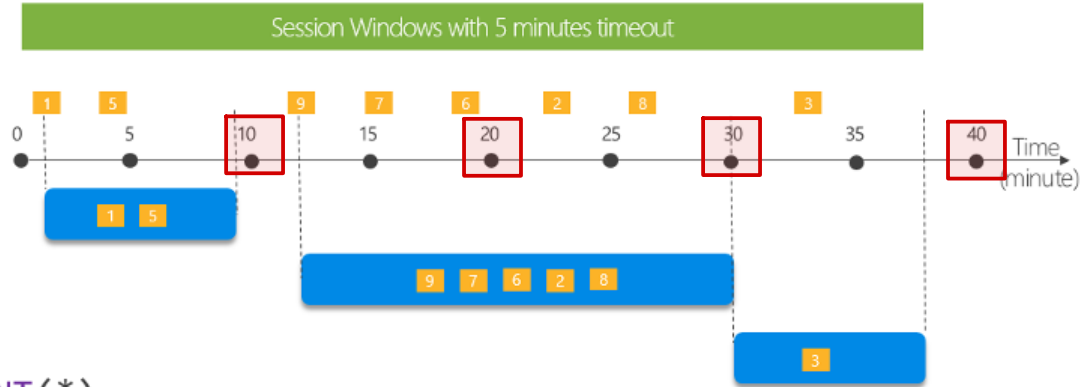
```
SELECT COUNT(*)  
FROM Input  
GROUP BY SlidingWindow(second, 10)
```

FIXED
LENGTH

WINDOW SLIDES UNTIL THERE ARE
NO EVENT IN THE INTERVAL

LENGTH NOT FIXED
WINDOWS DO NOT OVERLAP

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

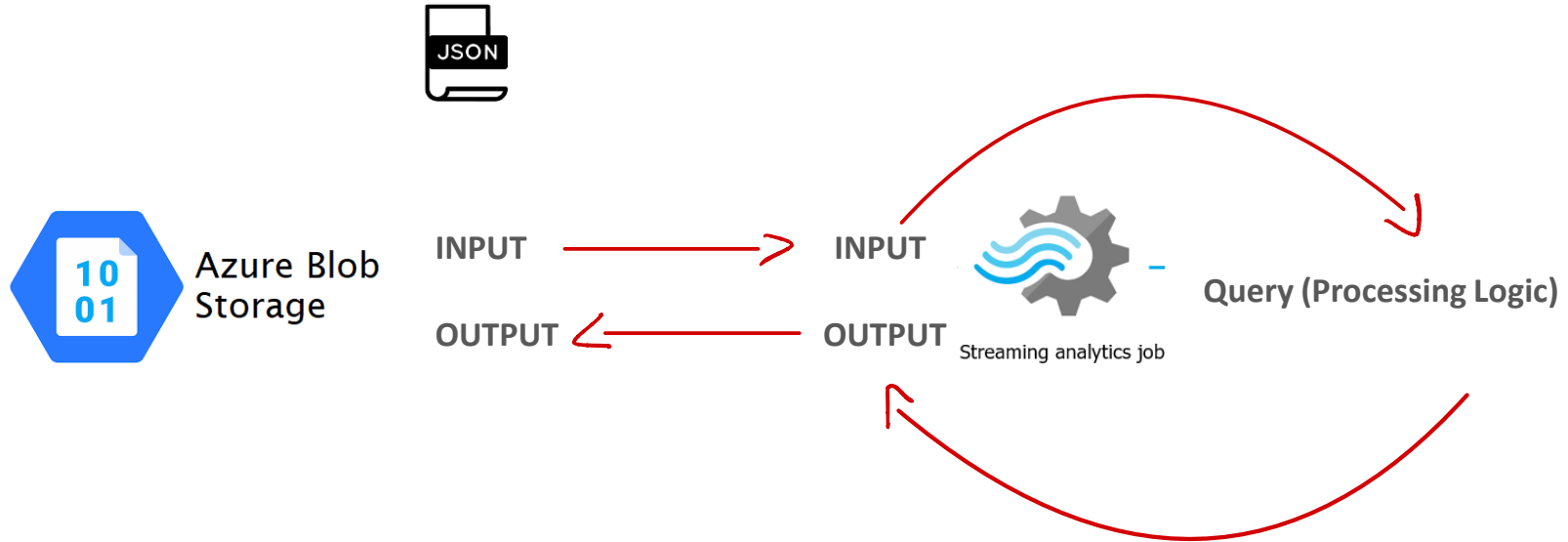


```
SELECT Topic, COUNT(*)
FROM TwitterStream TIMESTAMP BY CreatedAt
GROUP BY Topic, SessionWindow(minute, 5, 10)
```

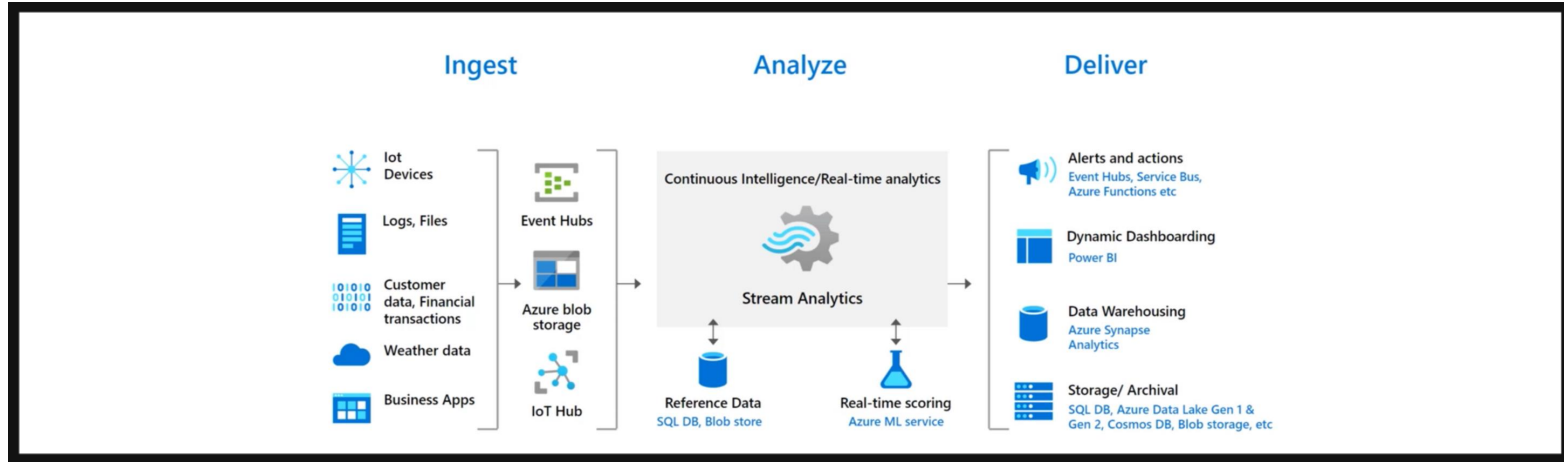
TIME UNIT TIME OUT

WINDOW TERMINATES IF THERE ARE NO EVENTS IN TIMEOUT PERIOD, OR THE 10m WINDOW REACHED. \Rightarrow

Demo Overview



Reference Data Inputs



Metadata Lookup

Device capacity, name, etc.



Acceptable thresholds

Allowed temperatures, etc.



Trusted entities

Registered devices



Any lookup or slow

Changing data



Azure Stream Analytics Stream Data Output

