Building Streaming Data Pipelines in Microsoft Azure

AZURE STREAM ANALYTICS OVERVIEW

Reza Salehi
CLOUD CONSULTANT

@zaalion linkedin.com/in/rezasalehi2008

Overview

Live data processing and use cases

Live data processing systems

- Challenges
- Technology choices

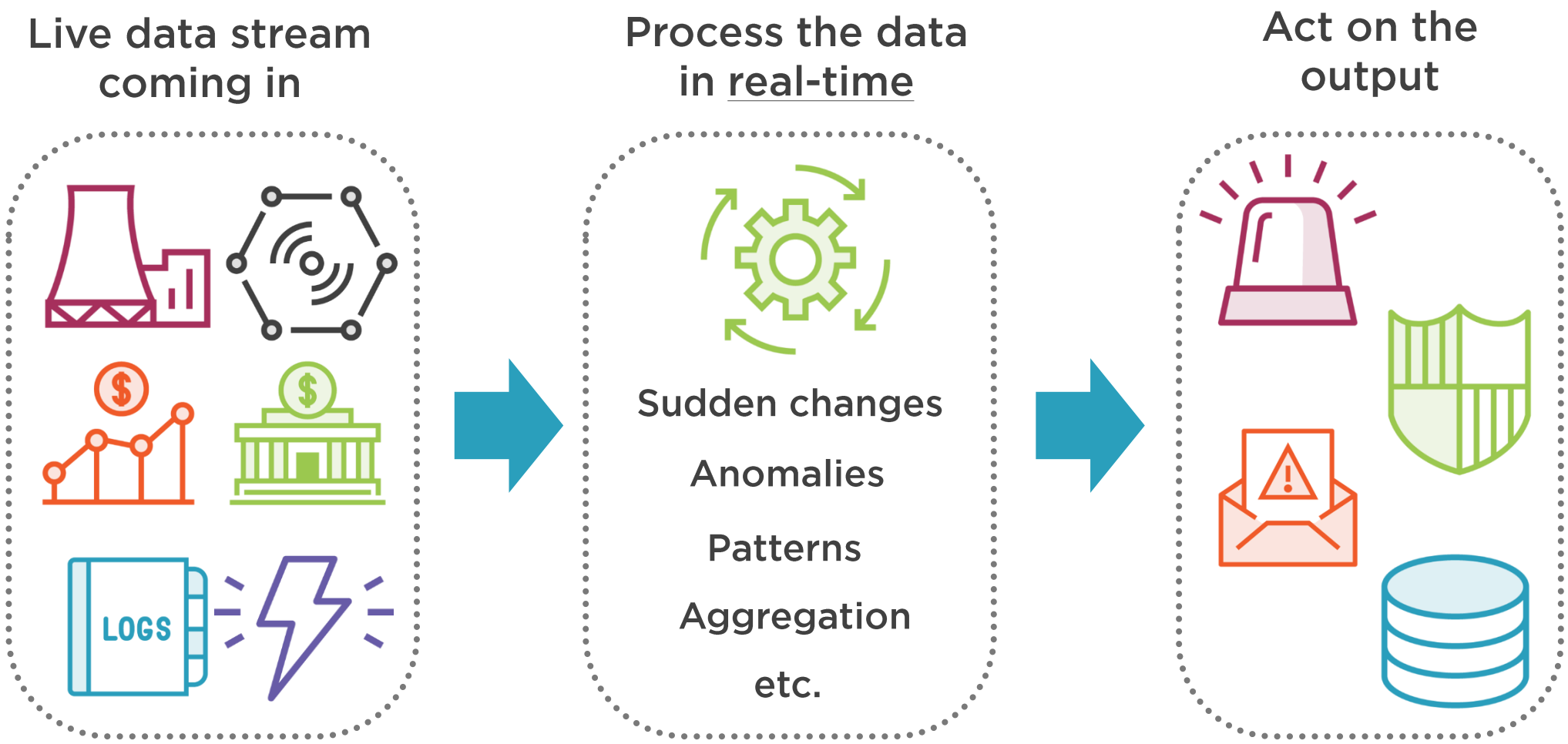
Introducing Azure Stream Analytics

- Supported inputs and outputs/sinks
- Windowing

Demo: Provisioning an Azure Stream Analytics instance

Live Data Processing

Live Data Processing Use Cases



Live Data Processing Use Cases

Getting alerts when power plant core temperature suddenly changes

Block malicious online banking user before it's too late

Automatically buy or sell stocks in real-time based on patterns

Prevent data security breaches in real-time

Data ingestion, processing and output should happen in real-time.

Live Data Processing Challenges

Should support high volume data ingestion

Enough processing power, so data ingestion is not interrupted

The output storage should have high bandwidth

Should quickly act on the output of processing

Azure Options for Live Data Processing

HDInsight with Spark Streaming

Apache Spark in Azure Databricks

WebJobs

HDInsight with Storm

Azure Functions

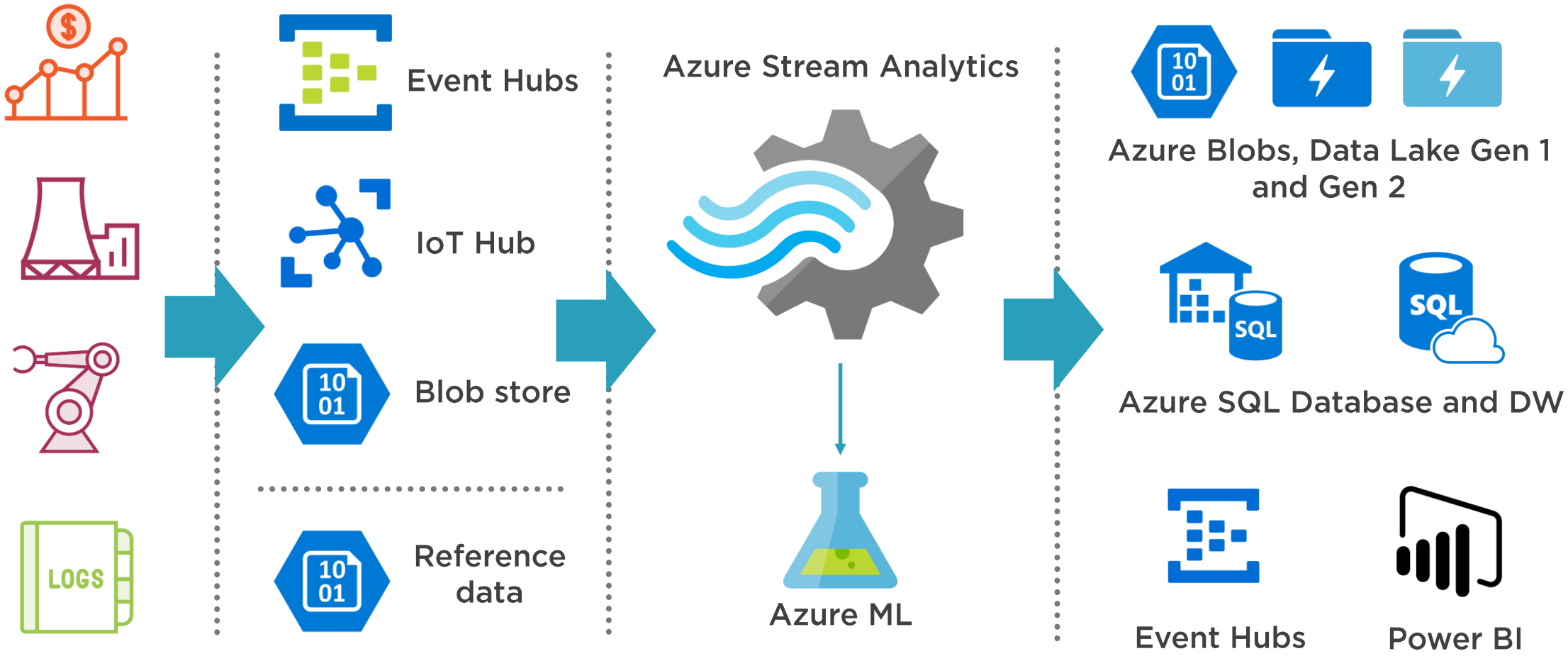
Azure Stream Analytics

Azure Stream Analytics

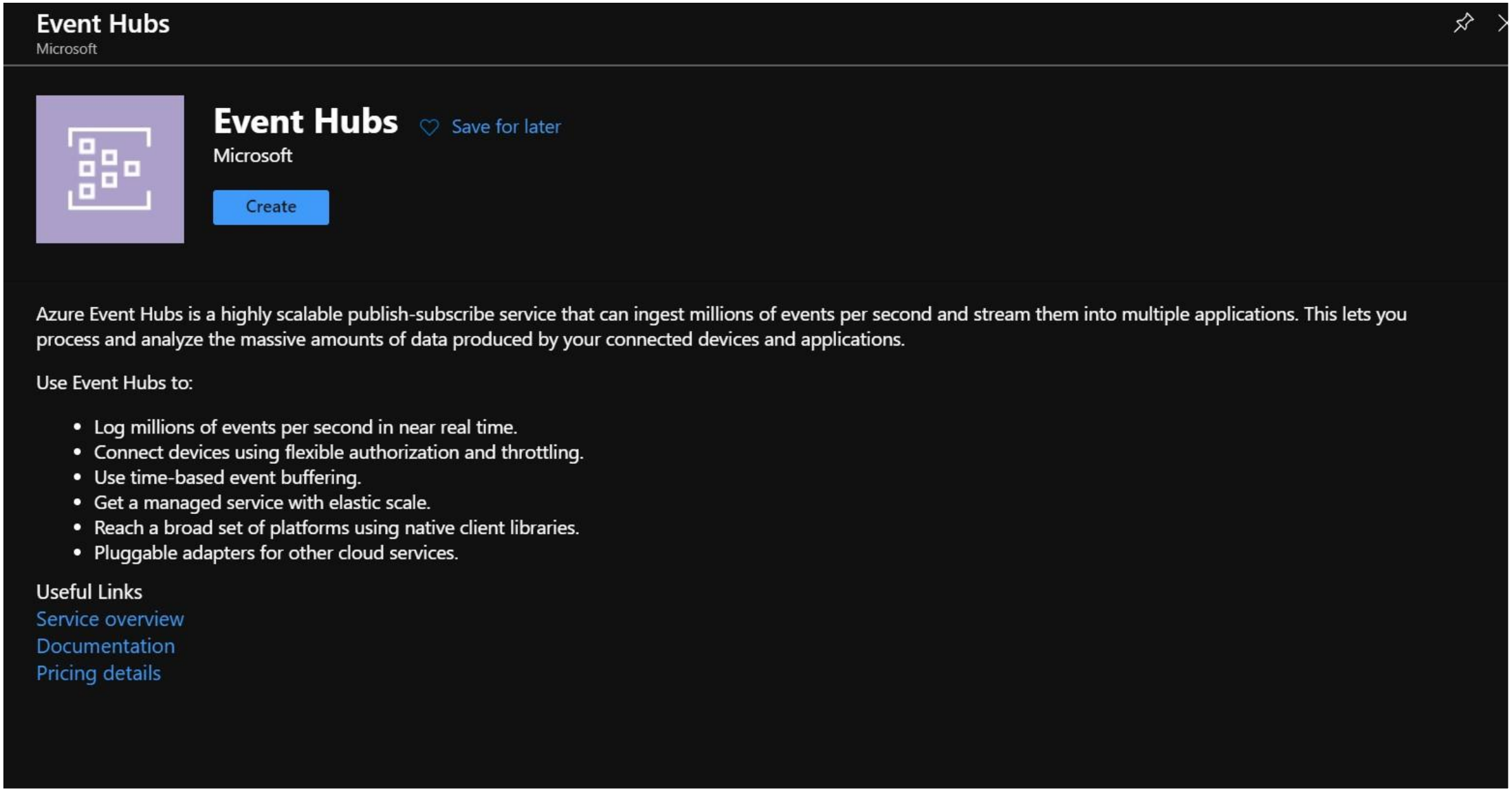
Azure Stream Analytics

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

Azure Stream Analytics Data Flow



Create a Stream Analytics Job



Create a Stream Analytics Job

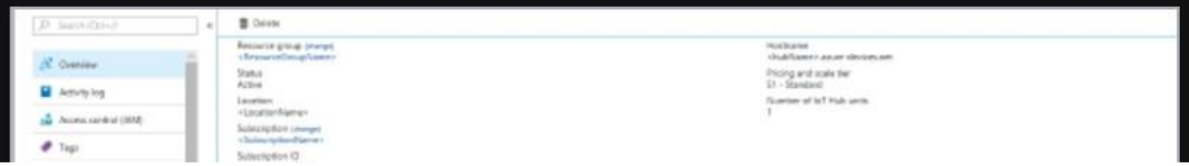
IoT Hub
Microsoft

IoT Hub ♡ Save for later
Microsoft

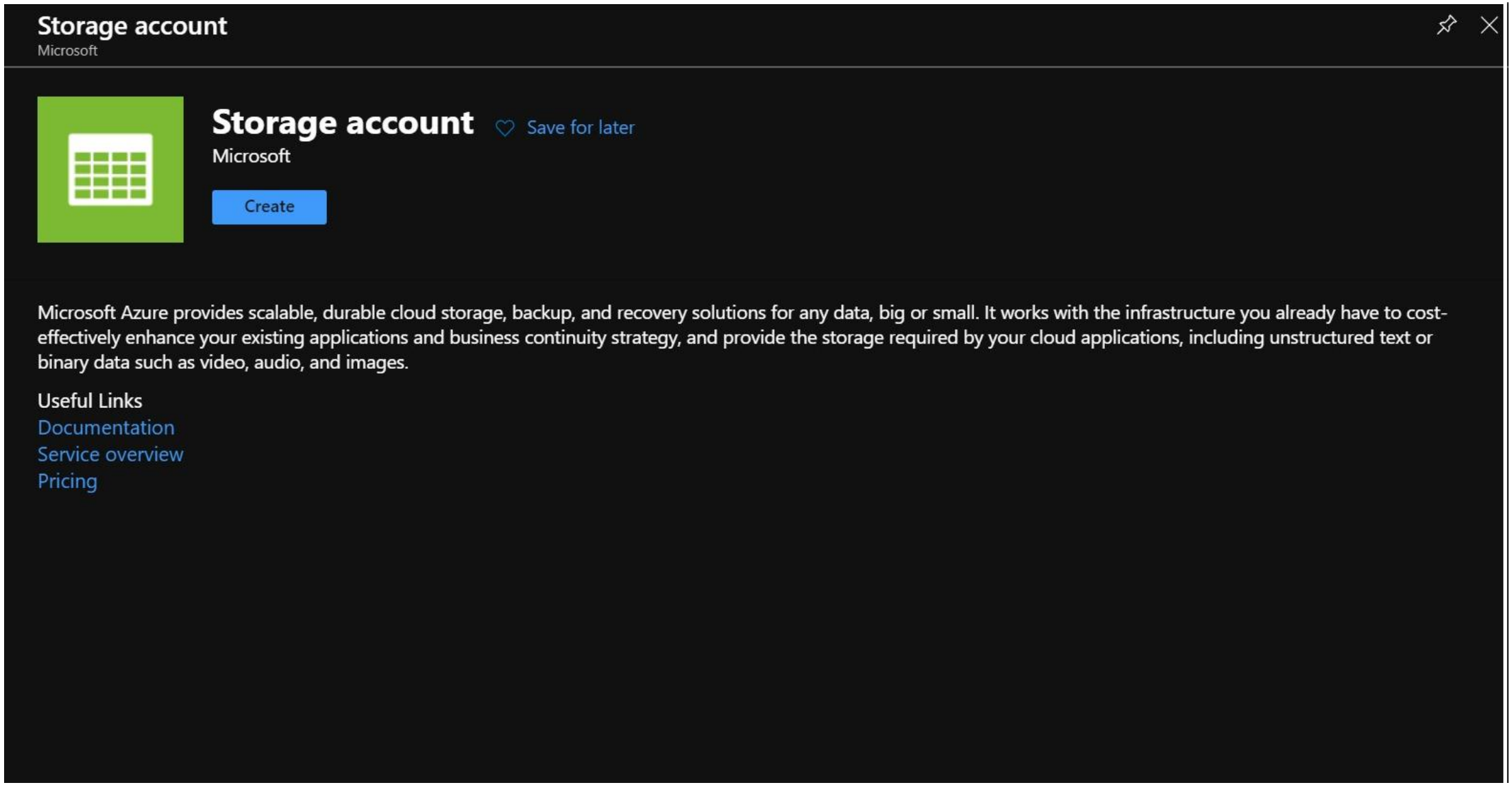
Create

Simultaneously support millions of connected devices—whether they run Windows, Linux, or real-time operating systems. Then monitor performance and send commands to accelerate your digital transformation.

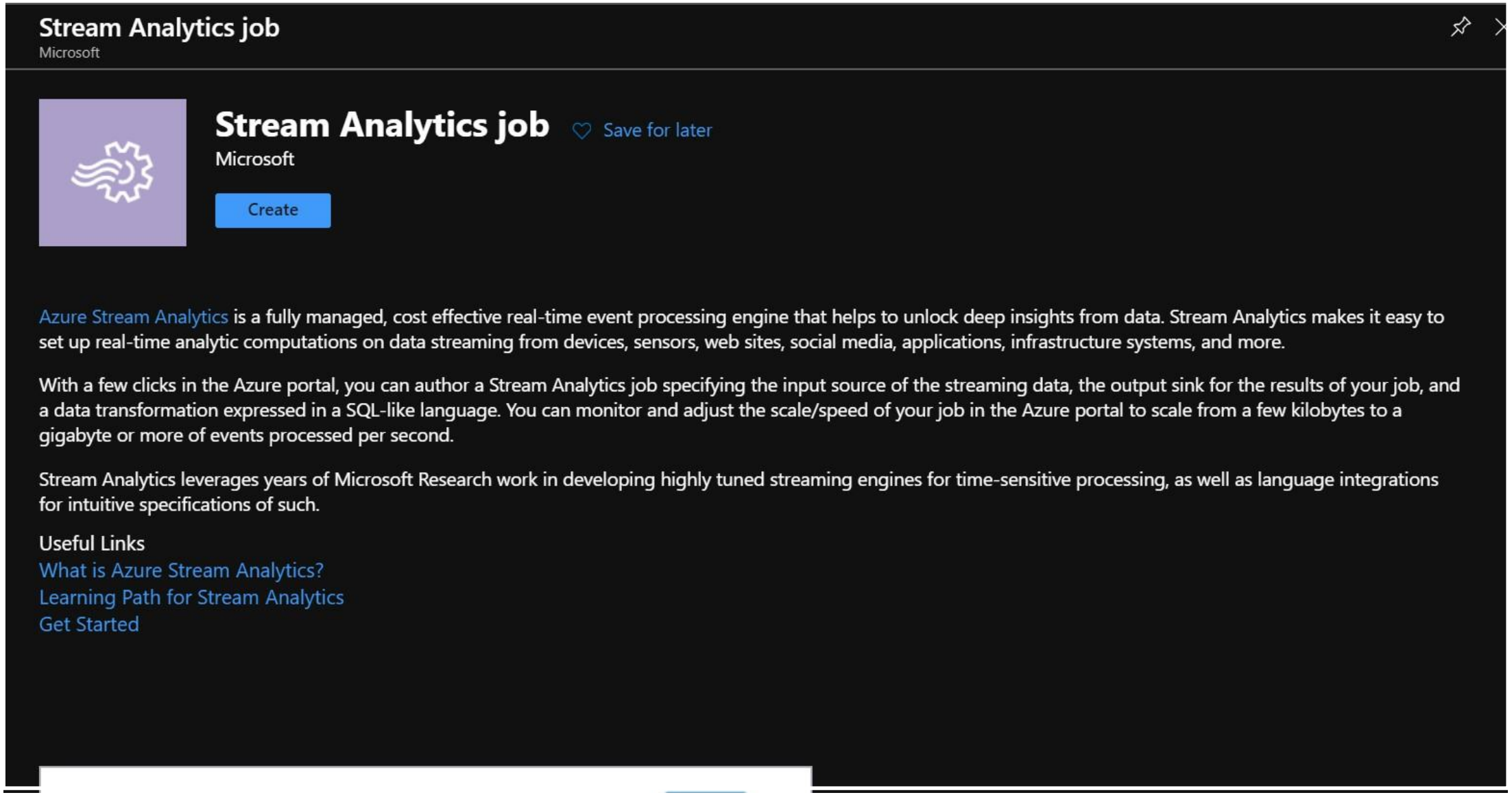
Useful Links
Documentation
Device management
Service overview
Pricing and scale details
Learn more about Azure IoT Hub



Create a Stream Analytics Job



Create a Stream Analytics Job



Time Windowing

Name	Date	TimeOff
John	Mon	1.5
Mary	Tue	3.0
Reza	Tue	2.0
John	Fri	3.0

Give me total hourly time off per employee

```
SELECT Name, SUM(TimeOff)
FROM Employee
GROUP BY (Name)
```

John	4.5
Mary	3.0
Reza	2.0

Timestamp	Sensor	Temperature
10:12:34:00	SEN01	97.5
10:12:34:05	SEN01	99.2
10:12:34:10	SEN01	120.4
10:12:34:15	SEN01	170.6
10:12:34:20	SEN01	180.9
10:12:34:25	SEN01	195.0
10:12:34:30	SEN01	200.5

Give me the average SEN01 temperature for the past 10 seconds

SEN01	109.37
SEN01	117.17
SEN01	191.97

Azure Stream Analytics Windowing

Each data event has a timestamp

There is a need to perform an operation on events falling in the same time window (e.g. COUNT)

An easy method is needed to find these subsets of data events

Azure Stream Analytics achieves this through windows

Azure Stream Analytics Windowing

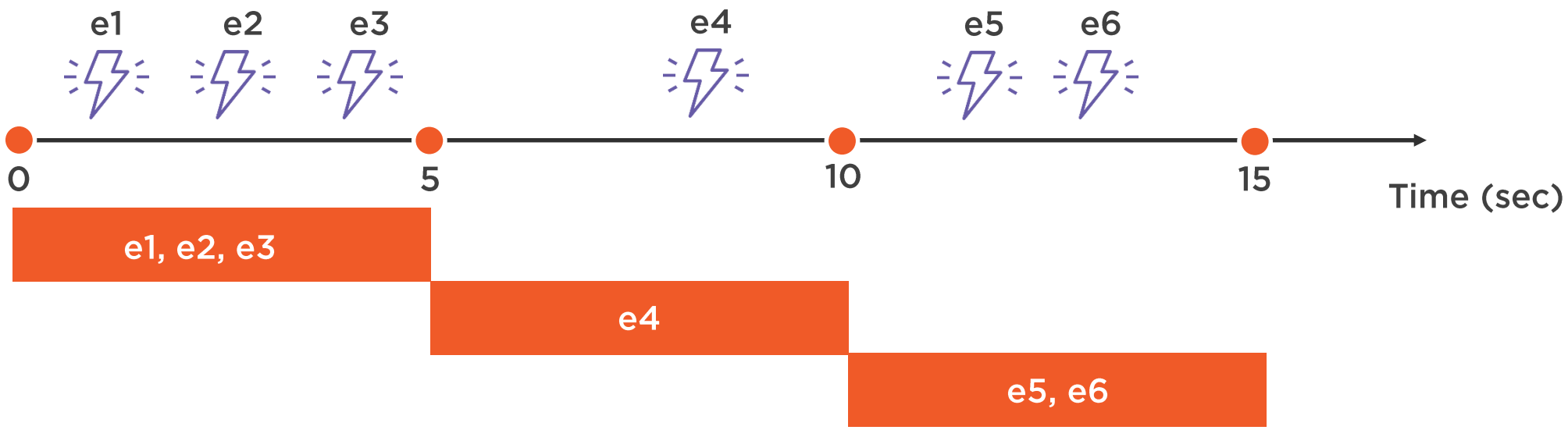
Tumbling Window

Hopping Window

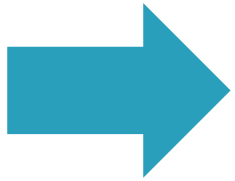
Sliding Window

Session window

Tumbling Window

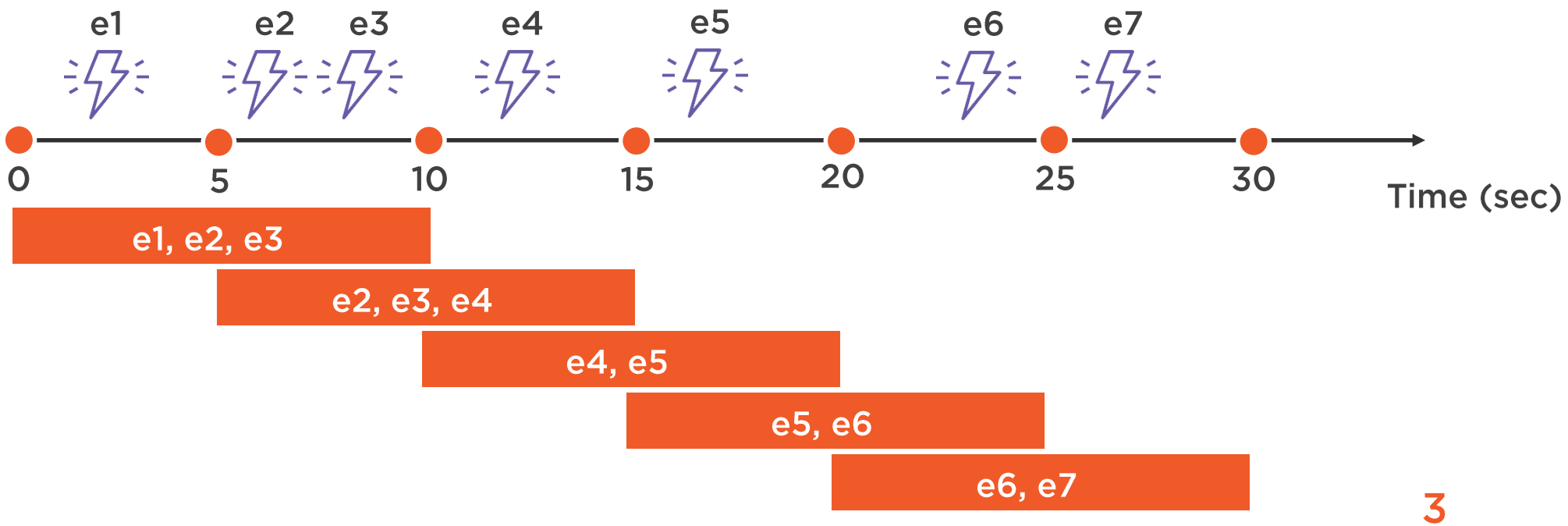



```
SELECT COUNT(*)
FROM Input
GROUP BY TumblingWindow(second, 5)
```



3
1
2

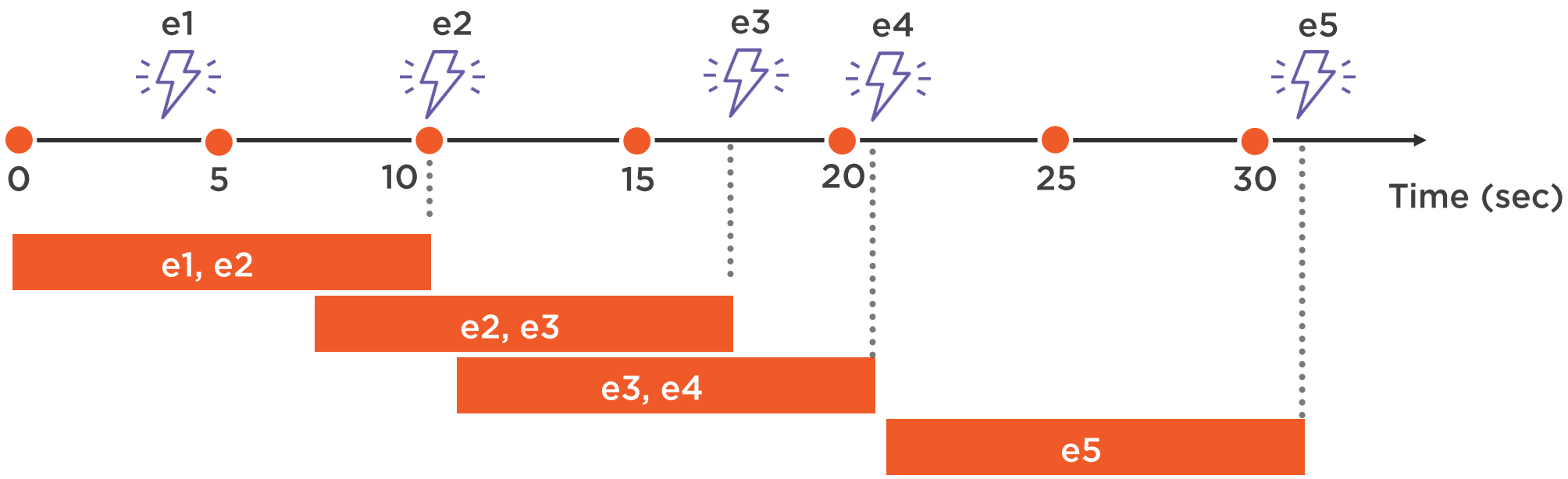
Hopping Window




```
SELECT COUNT(*)
FROM Input
GROUP BY HoppingWindow(second, 10, 5)
```

3
3
2
2
2

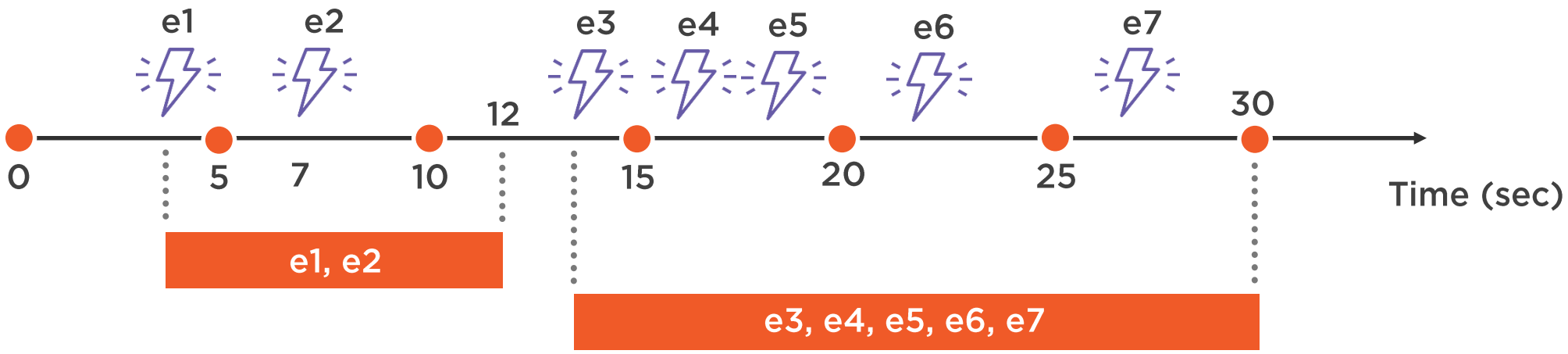
Sliding Window



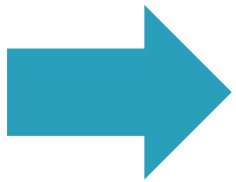

```
SELECT COUNT(*)
FROM Input
GROUP BY SlidingWindow(second, 10)
```

2
2
2
1

Session Window




```
SELECT COUNT(*)
FROM Input
GROUP BY SessionWindow(second, 5, 10)
```



2
5

Demo

Provision a new Azure Stream Analytics instance

Process blob storage input

Summary

Live stream analytics and use cases

Live data processing systems

- Architecture and challenges
- Technology choices

Introducing Azure Stream Analytics

- Inputs and outputs/sinks
- Windowing
- Provisioning steps

Demo: Provision a new Azure Stream Analytics instance