

Building Streaming Data Pipelines in Microsoft Azure

AZURE STREAM ANALYTICS OVERVIEW



Reza Salehi

CLOUD CONSULTANT

@zaalion [linkedin.com/in/rezasalehi2008](https://www.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

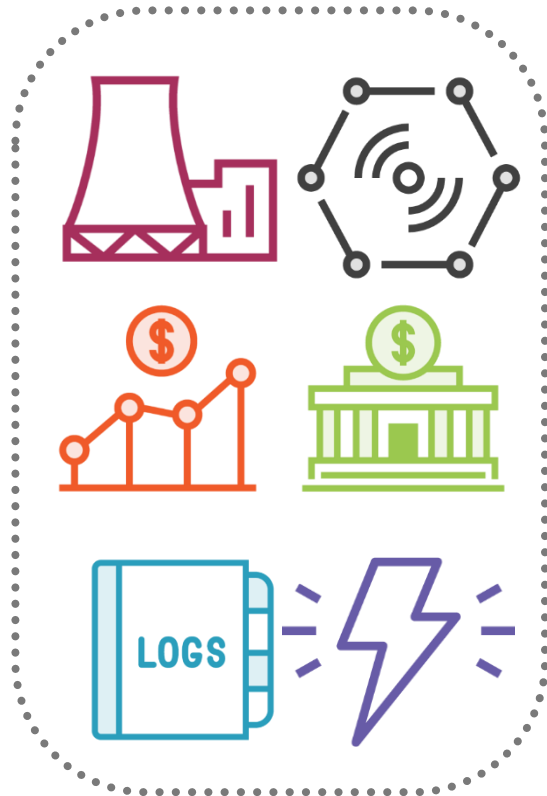


Live Data Processing

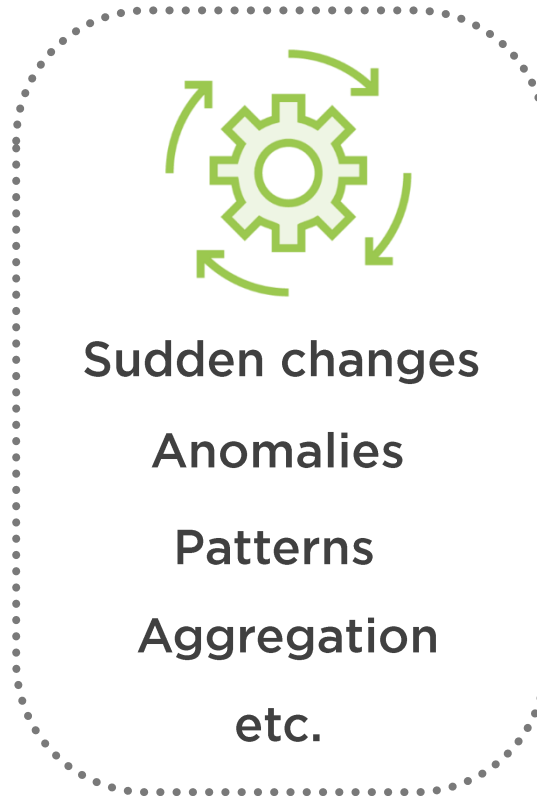


Live Data Processing Use Cases

Live data stream
coming in



Process the data
in real-time



Act on the
output



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

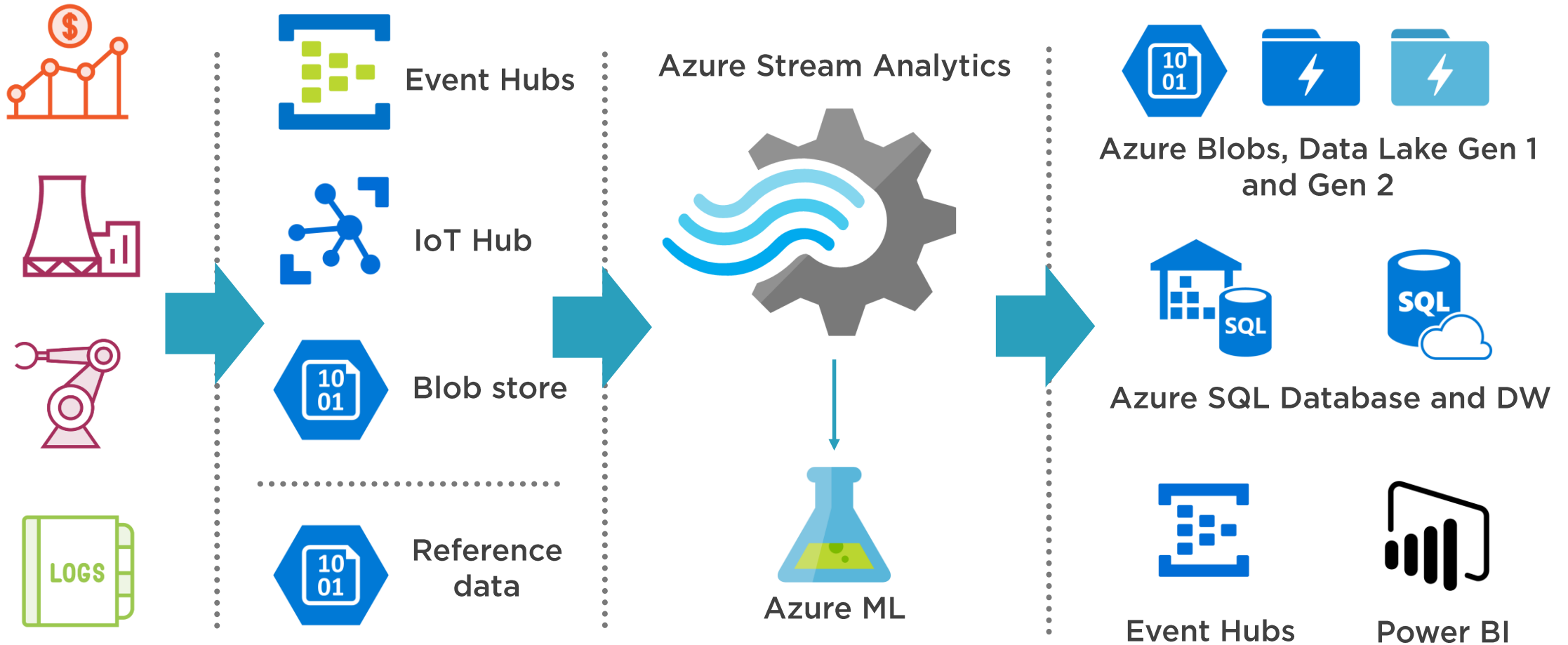


Azure Stream Analytics

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

Event Hubs
Microsoft


**Event Hubs**  [Save for later](#)
Microsoft
[Create](#)

Azure Event Hubs is a highly scalable publish-subscribe service that can ingest millions of events per second and stream them into multiple applications. This lets you process and analyze the massive amounts of data produced by your connected devices and applications.

Use Event Hubs to:


- Log millions of events per second in near real time.
- Connect devices using flexible authorization and throttling.
- Use time-based event buffering.
- Get a managed service with elastic scale.
- Reach a broad set of platforms using native client libraries.
- Pluggable adapters for other cloud services.

Useful Links
[Service overview](#)
[Documentation](#)
[Pricing details](#)



Create a Stream Analytics Job

IoT Hub
Microsoft

**IoT Hub**
Microsoft

♡ Save for later

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](#)

Search (Ctrl+J)

Overview

Activity log

Device central (BETA)

Tags

Create

Resource group (string)

ResourceGroup/Name

Status

Active

Location

Location/Region

Subscription (string)

Subscription/Name

Subscription ID

Module

hubName - azure-iot-hub

Pricing and scale tier

S1 - Standard


Number of IoT Hub units

1



Create a Stream Analytics Job

Storage account
Microsoft



Storage account

Microsoft

[Save for later](#)

Create

Microsoft Azure provides scalable, durable cloud storage, backup, and recovery solutions for any data, big or small. It works with the infrastructure you already have to cost-effectively enhance your existing applications and business continuity strategy, and provide the storage required by your cloud applications, including unstructured text or binary data such as video, audio, and images.


Useful Links

- [Documentation](#)
- [Service overview](#)
- [Pricing](#)



Create a Stream Analytics Job

Stream Analytics job
Microsoft



Stream Analytics job

Microsoft

Create

Save for later

[Azure Stream Analytics](#) is a fully managed, cost effective real-time event processing engine that helps to unlock deep insights from data. Stream Analytics makes it easy to set up real-time analytic computations on data streaming from devices, sensors, web sites, social media, applications, infrastructure systems, and more.

With a few clicks in the Azure portal, you can author a Stream Analytics job specifying the input source of the streaming data, the output sink for the results of your job, and a data transformation expressed in a SQL-like language. You can monitor and adjust the scale/speed of your job in the Azure portal to scale from a few kilobytes to a gigabyte or more of events processed per second.

Stream Analytics leverages years of Microsoft Research work in developing highly tuned streaming engines for time-sensitive processing, as well as language integrations for intuitive specifications of such.

Useful Links

- [What is Azure Stream Analytics?](#)
- [Learning Path for Stream Analytics](#)
- [Get Started](#)



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

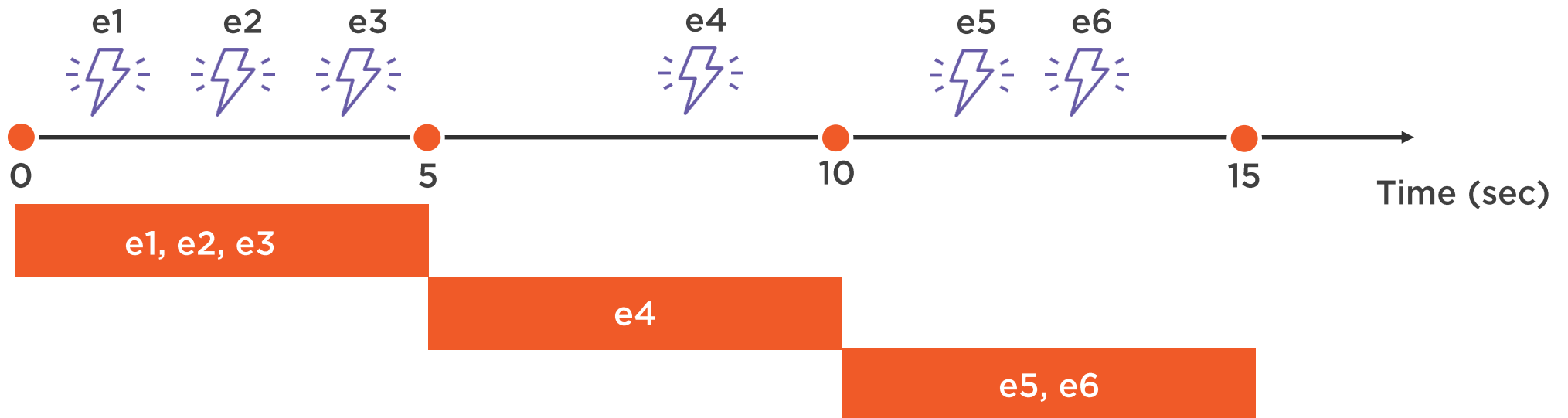
Hopping Window

Sliding Window

Session window



Tumbling Window

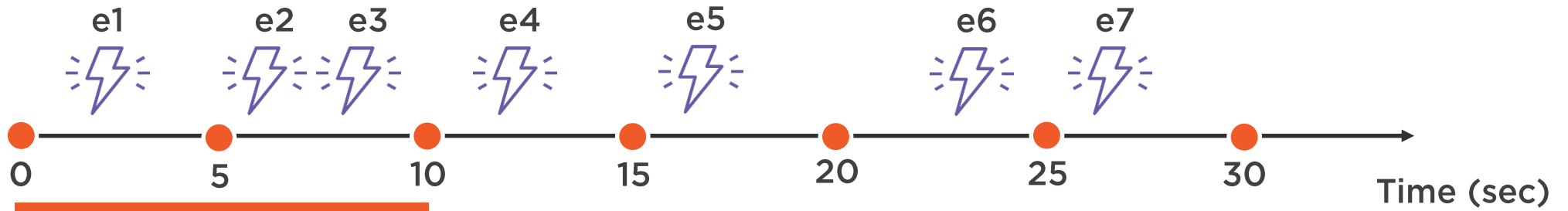


```
SELECT COUNT(*)  
FROM Input  
GROUP BY TumblingWindow(second, 5)
```

3
1
2



Hopping Window



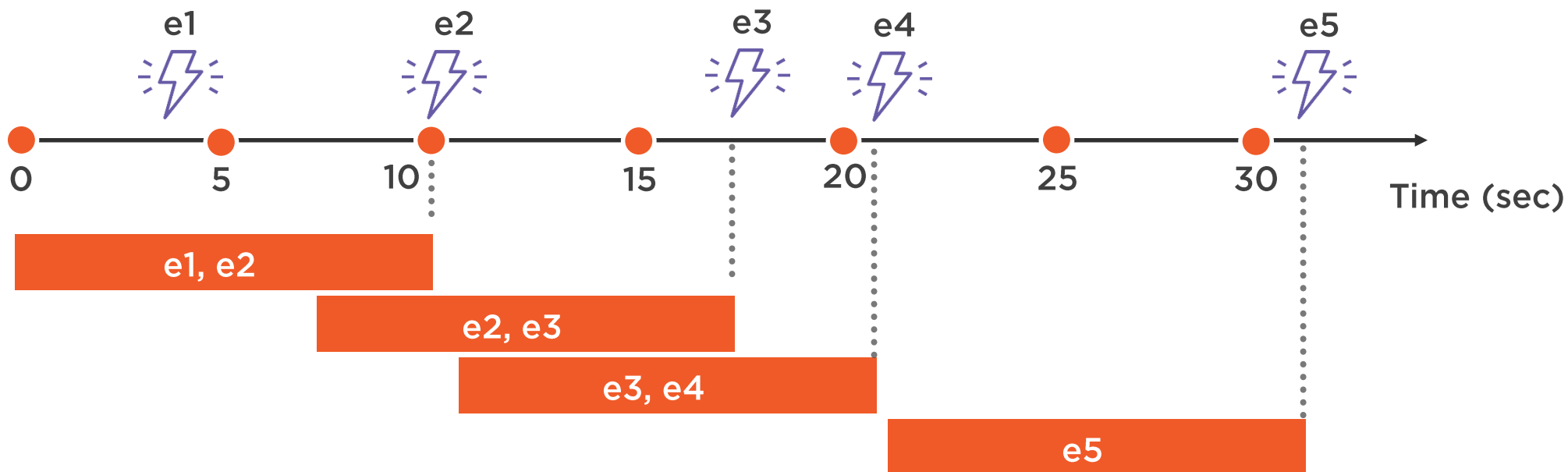
```
SELECT COUNT(*)  
FROM Input  
GROUP BY HoppingWindow(second, 10, 5)
```



3
3
2
2
2



Sliding Window

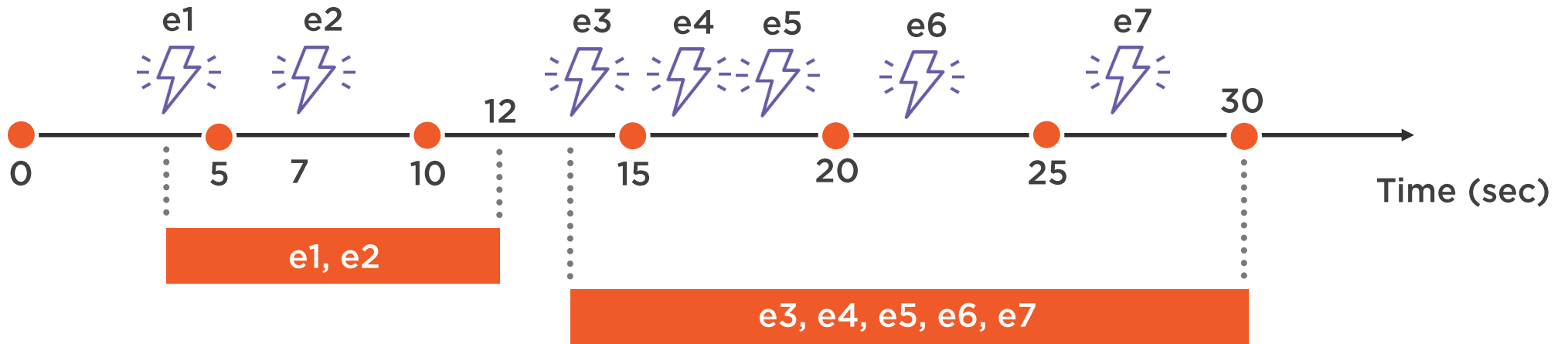


```
SELECT COUNT(*)  
FROM Input  
GROUP BY SlidingWindow(second, 10)
```

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

