Analyzing Data on-the-fly with Stream Analytics

Warner Chaves
MS DATA PLATFORM MVP

@warchav sqlturbo.com

What's in this module?

Stream Analytics Fundamentals

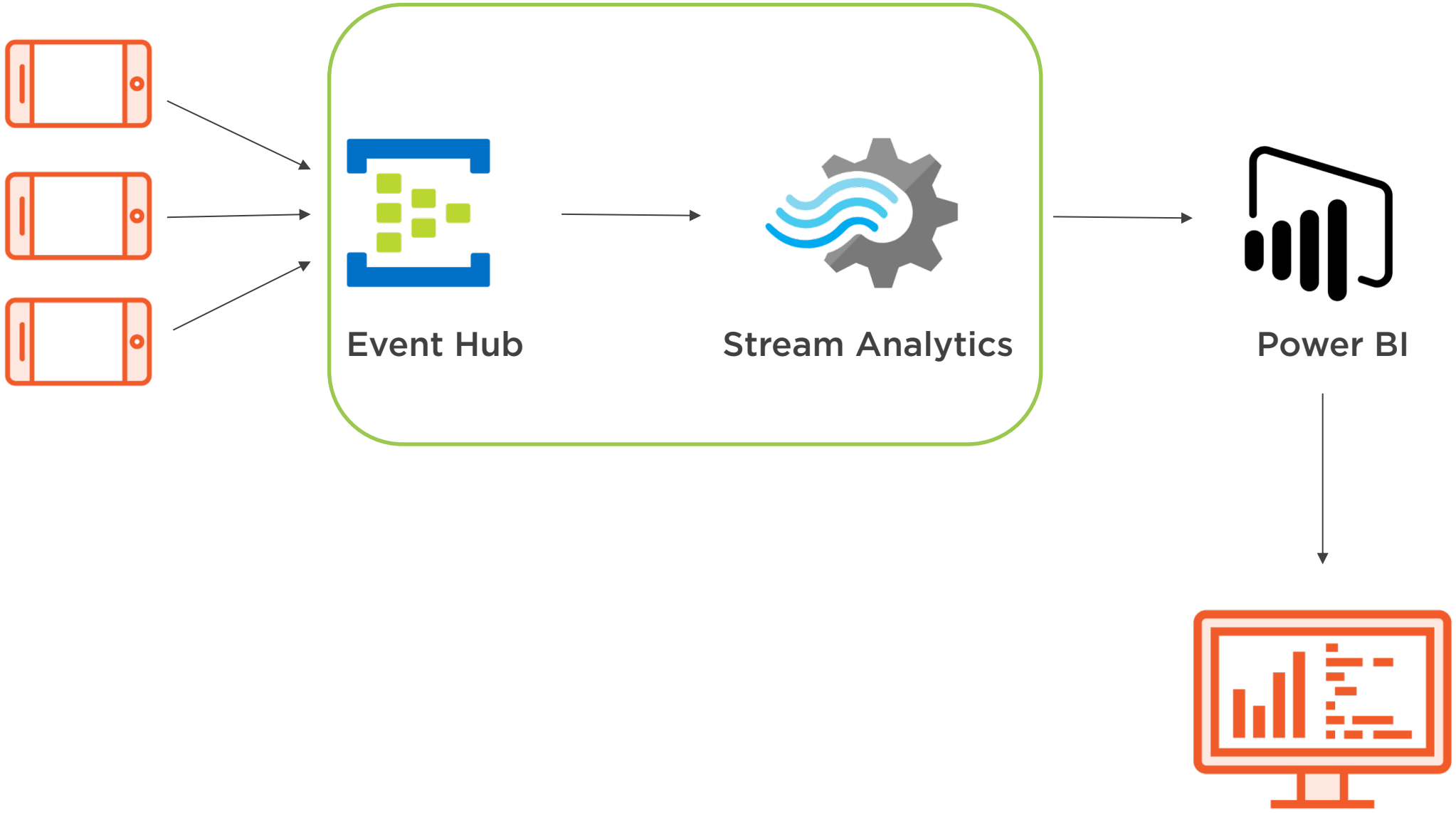
Creating a Stream Analytics Job

Stream Analytics Queries

Processing the machine's temperature Event Hub data with Stream Analytics

2. Azure Stream Analytics

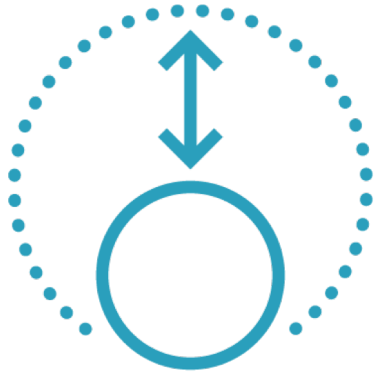
Analyze your data on-the-fly


Event Hub
Stream Analytics
Power BI

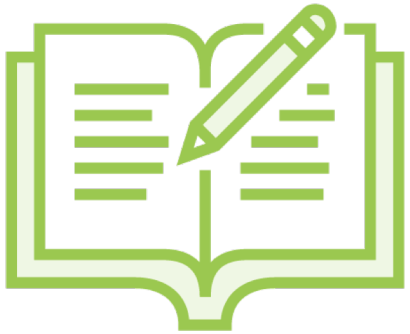
When do you need a Stream Analytics job?

Real-time data and Stream Analytics

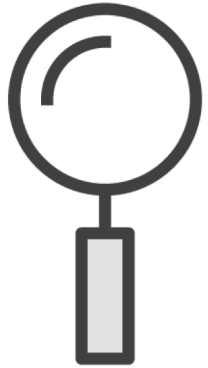
Aggregation



Transform



Reference



Analyze

How do we get started?

Stream Analytics Job

Pick a Hosting Environment

Pick a Region

Pick Streaming Units

Streaming Units

SUs represent the computing resources of a job

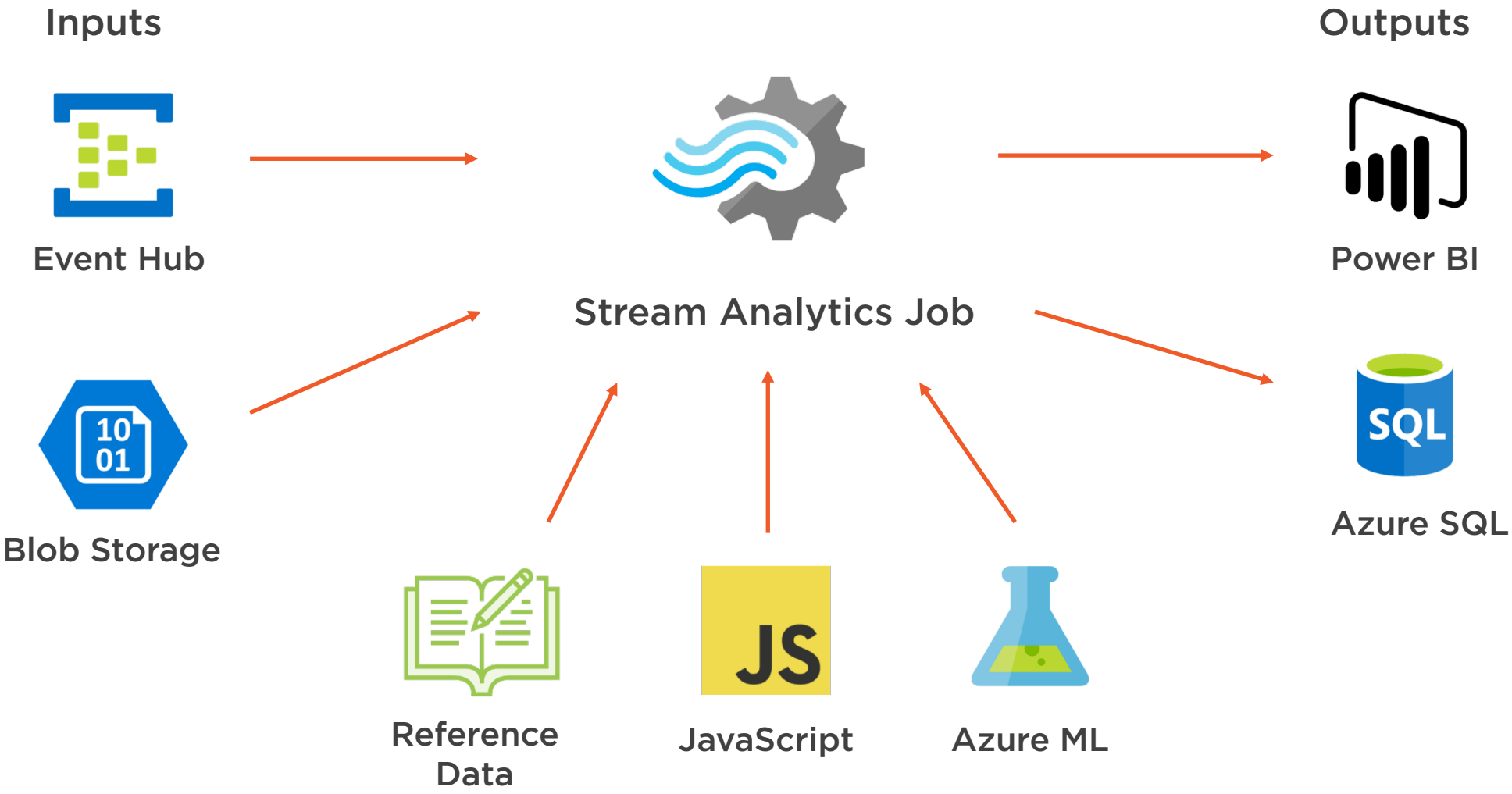
It is a blended measure of CPU, memory, and read and write rates

Number of events, query complexity and UDFs impact SU consumption

You can monitor the usage to do capacity planning

Unlike an Event Hub Namespace, a Stream Analytics job can be stopped and you will not be billed until it is started again.

Job Components



Data Formats

Stream Analytics supports reading data from the stream in these three formats: JSON, AVRO and CSV.

Demo

Creating a Stream Analytics job through the Azure Portal

How do we query the stream?

SQL style queries with optional JavaScript

Stream data is easy to manipulate based on time windows

Time-aware SQL constructs

```
SELECT temperature, currentMetric, currentDirection,
machineName, temperatureReadTimestamp

FROM

    FactoryEventHub TIMESTAMP BY temperatureReadTimestamp
```

TIMESTAMP BY

Defines the application timestamp field

```
SELECT MAX(temperatureReadTimestamp) AS WindowEnd,
AVG(temperature) AS averageTemperature,
machineName
FROM
    FactoryEventHub TIMESTAMP BY temperatureReadTimestamp
    GROUP BY TUMBLINGWINDOW(s, 1), machineName
```

TUMBLINGWINDOW

Defines non overlapping windows of time in the stream

```
WITH LastInWindow AS

(...)

SELECT *

FROM

    FactoryEventHub TIMESTAMP BY temperaturereadtimestamp INNER JOIN LastInWindow

    ON DATEDIFF(second, FactoryEventHub, LastInWindow) BETWEEN 0 AND 1

    AND FactoryEventHub.temperaturereadtimestamp = LastInWindow.LastEventTime
```

JOIN ON DATEDIFF

Defines a time boundary to limit the amount of records to evaluate for the join

If you want to do a deeper dive into Stream Analytics, refer to the course: “Understanding Azure Stream Analytics” by Alan Smith.

Let’s continue building our solution!

Machinery Temperature Monitoring

We stream the machine's temperature readings multiple times per second to Azure

We want to aggregate the temperature readings and add reference information to it

Average Temperature

Maximum Temperature

Minimum Temperature

Standard Deviation of the Temperature

Last Temperature Direction

Demo

Querying Event Hub data with Stream Analytics

Reference Data

Data that enriches the incoming stream

The data must be a file or set of files in Blob Storage

Useful when the data origin is not aware of this reference data

You refer to it as simply another table in your Stream Analytics SQL

Reference Data

We will augment the stream by adding information about which energy panel is connected to each machine.

Demo

Adding reference data with Stream Analytics

Summary

Stream Analytics allow augmenting, aggregating, and analyzing data on-the-fly

SUs represent the compute resources for the job

Jobs are defined in a SQL query language

At this point in our solution, we have the Stream Analytics query that we will use to push data forward to Power BI

Next Module:

Ingesting Streaming Data into Power BI