Real-time Analytics with Azure Stream Analytics

Michael Johnson

@MikeJohnsonZA

What I'd like to share with you today

Introduction to streaming data

Overview of Azure Steam Analytics

Demonstrate a simple streaming data solution

What is streaming data

What is streaming data?

Batch Data

What is streaming data?

Streaming data

What is streaming data?

Example

What is streaming data?

Sources of streaming data

Azure Stream Analytics

Azure Stream Analytics

What is Azure Stream Analytics?

Real-time event processing engine

Azure Stream Analytics

What Microsoft like you to know.

Easy to use

Scalable

Reliable

Low Cost

Architecture

Architecture

Input

Output

Architecture

Inputs

Azure Event hub

Azure IOT hub
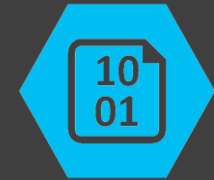
Azure Blob storage

Architecture
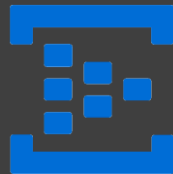
Outputs

Azure SQL database

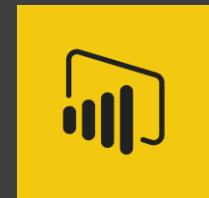Azure document DB

Azure Blob storage

Azure data lake

Azure Service bus

Azure Event hub

Azure table storage

Power BI

Query language

Query language

What does a query look like

```
SELECT
        Make,
        System.TimeStamp as Time,
        count(*) [NumOfCars]
INTO
        [Output]
FROM
        [Input] TIMESTAMP BY Time
GROUP BY
        Make,
        TumblingWindow(Second,10)
HAVING
        [Count] >= 3
```

Query language

Query language overview

DML Statements
- SELECT
- FROM
- WHERE
- GROUP BY
- HAVING
- CASE
- JOIN
- UNION

Date and time
- DATEPART
- DAY
- MONTH
- YEAR
- DATEDIFF
- DATEADD

Aggregation functions
- SUM
- COUNT
- AVG
- MIN
- MAX

String functions
- LEN
- CONCAT
- CHARINDEX

Scaling functions
- WITH
- PARTITION BY

Windowing functions
- TumblingWindow
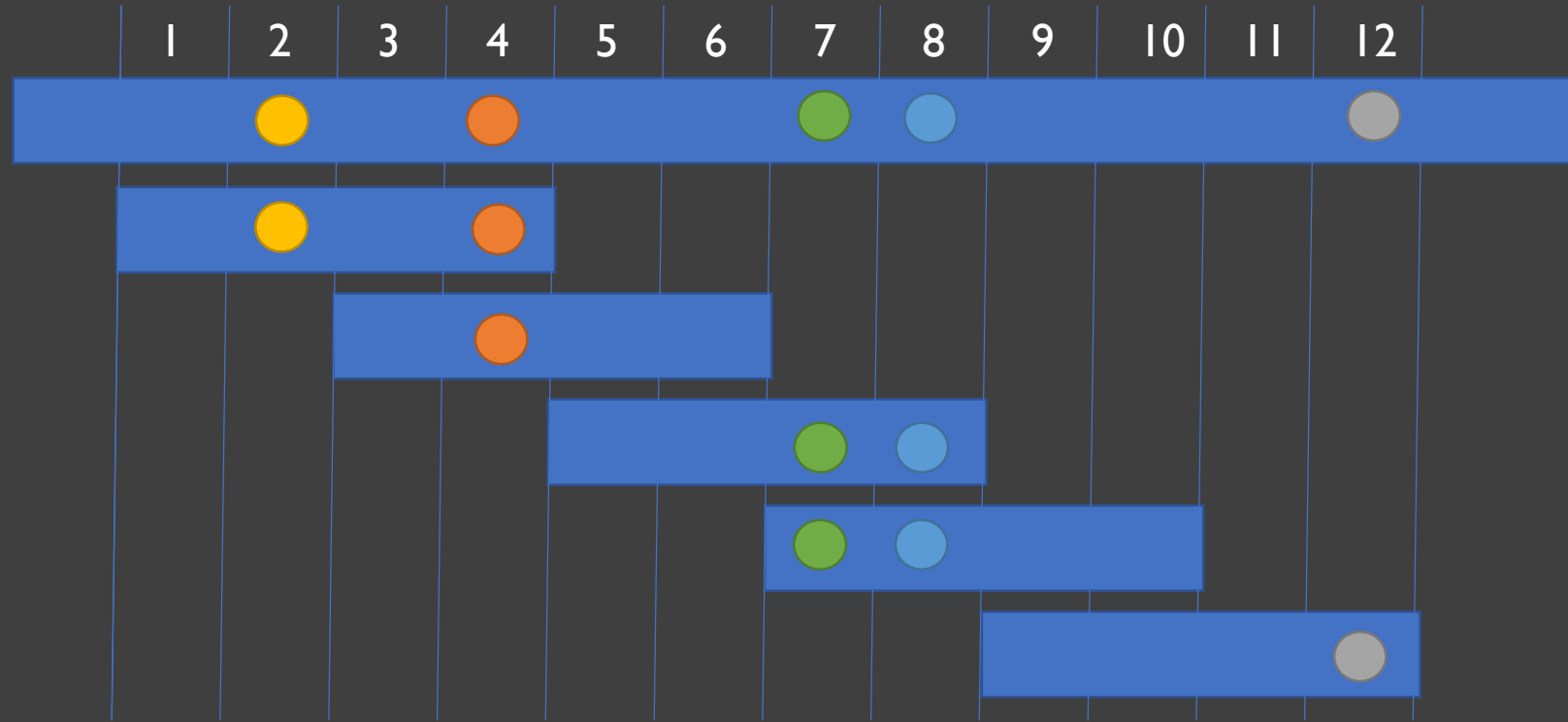- HoppingWindow
- SlidingWindow

Query language

Tumbling Window
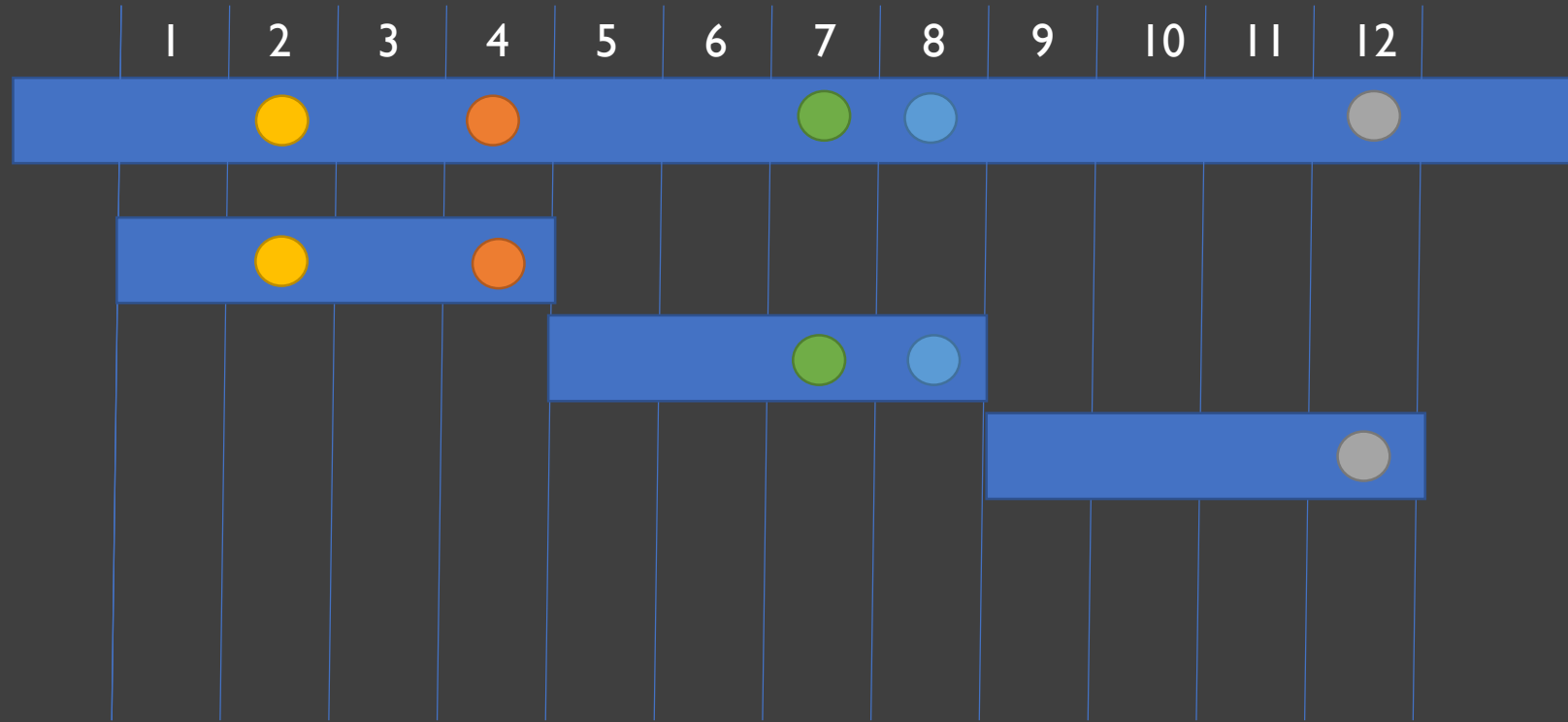
```
TUMBLINGWINDOW ( timeunit , windowsize,[offsetsize] )
TUMBLINGWINDOW ( SECOND , 4)
```

Query language

HOPPINGWINDOW (timeunit , windowsize , hopsize, [offsetsize])
HOPPINGWINDOW (SECOND , 4 , 2)

Query language

Sliding Window
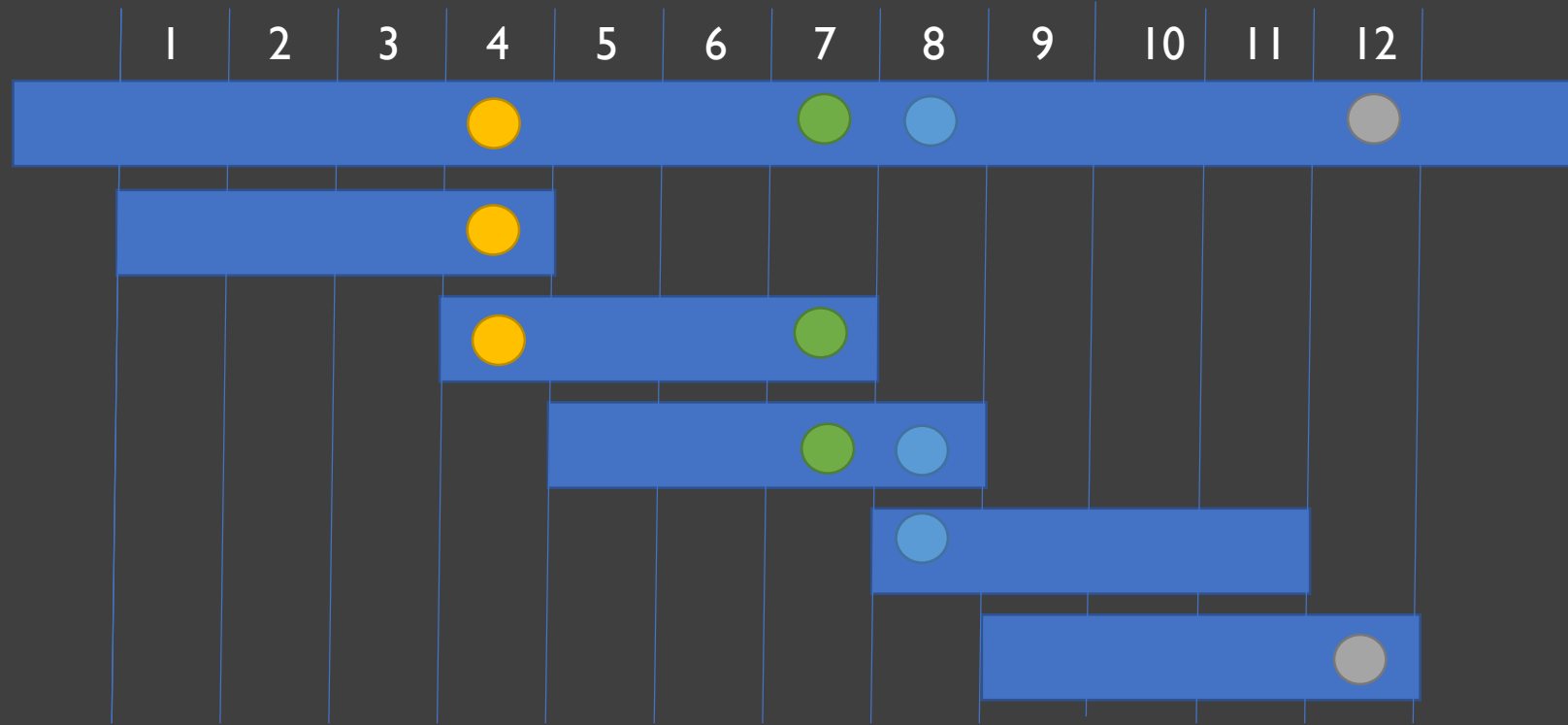
SLIDINGWINDOW (timeunit , windowsize)
SLIDINGWINDOW (SECOND,4)

Query language

- Limitations
  - Stateless
  - Compute only
  - No external connectivity

JavaScript UDF integration

```
Code
// Convert Hex value to integer.
function main(hexValue) {
    return parseInt(hexValue, 16);
}
```

```
Code
SELECT
    time,
    UDF.hex2Int(offset) AS IntOffset
INTO
    output
FROM
    InputStream
```

- Why
  - JavaScript libraries
    - REGEX
    - Math

Query language

- Data types
  - Point
  - Line
  - Polygon

Geospatial queries

- Functions
  - ST_DISTANCE
  - ST_OVERLAP
  - ST_INTERSECTS
  - ST_WITHIN

SELECT   c.Location, s.Location

FROM Cars c

JOIN Station s

ON ST_DISTANCE(c.location,s.location) < 10000

Query language

Azure machine learning

Input → Output
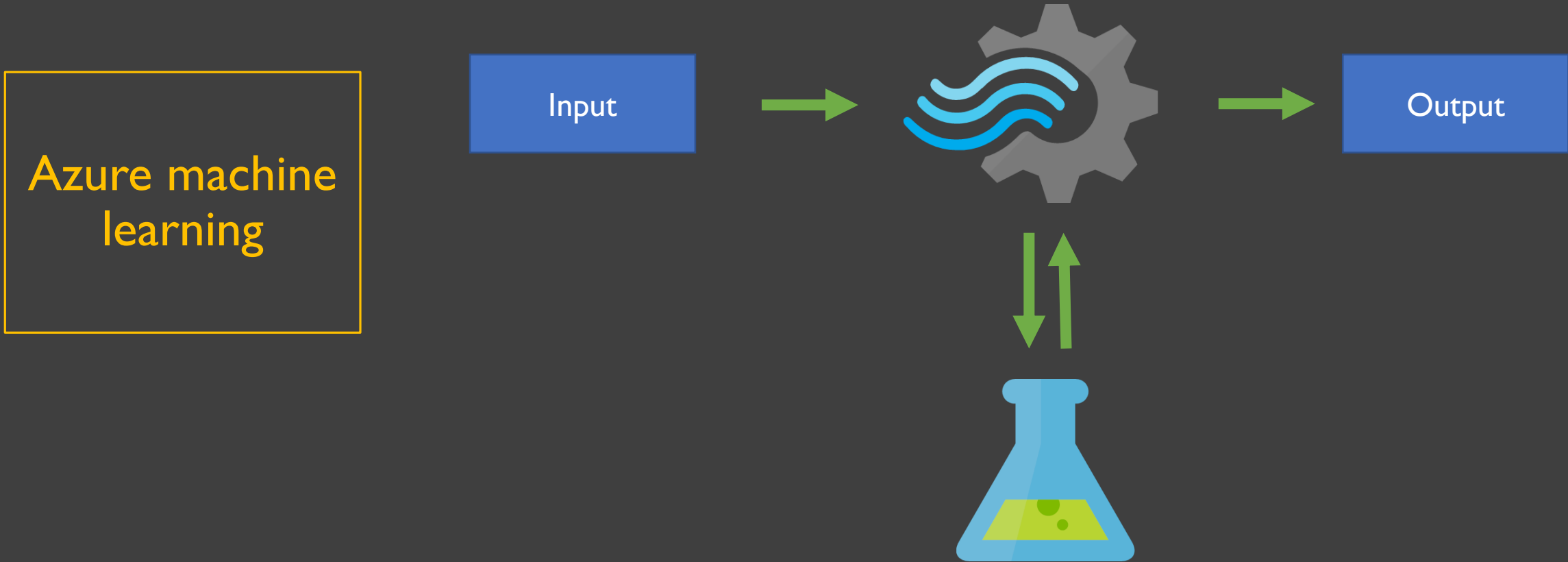
Azure Steam analytics on Azure IoT Edge

Preview
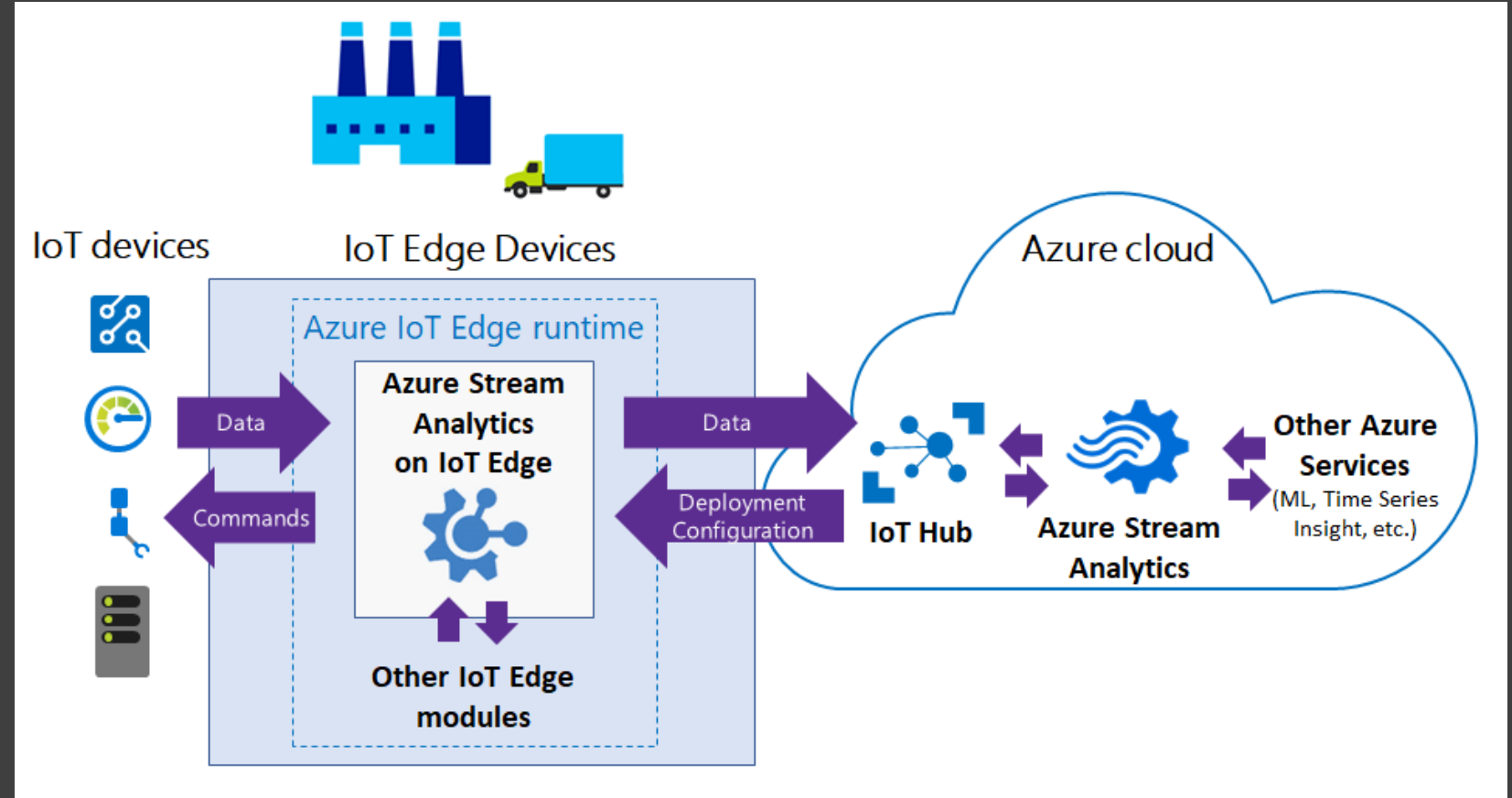
Anomaly Detection

Preview

ANOMALYDETECTION(<scalar_expression>)

OVER ([PARTITION BY <partition key>]

LIMIT DURATION(<unit>, <length>) [WHEN boolean_expression])

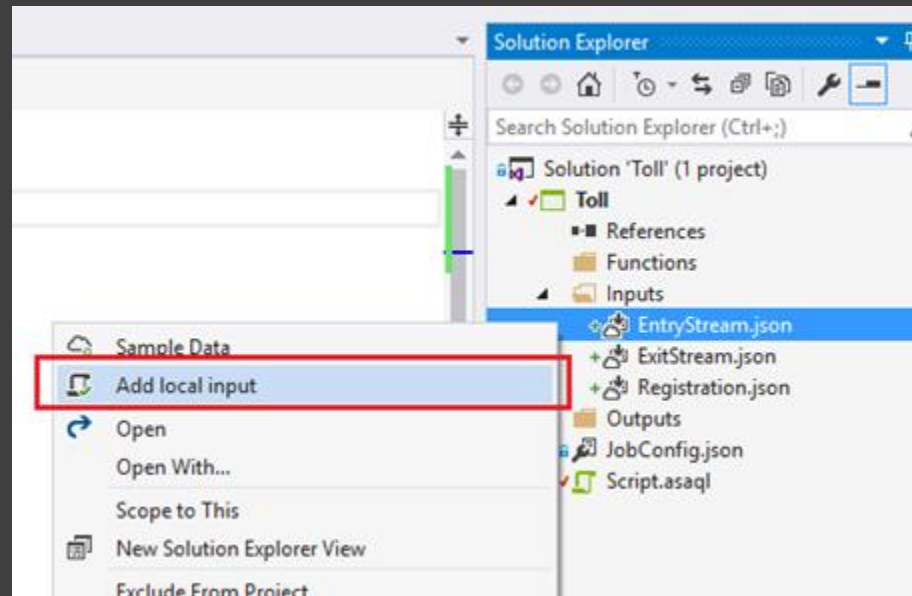# Development and deployment options

Development and deployment options

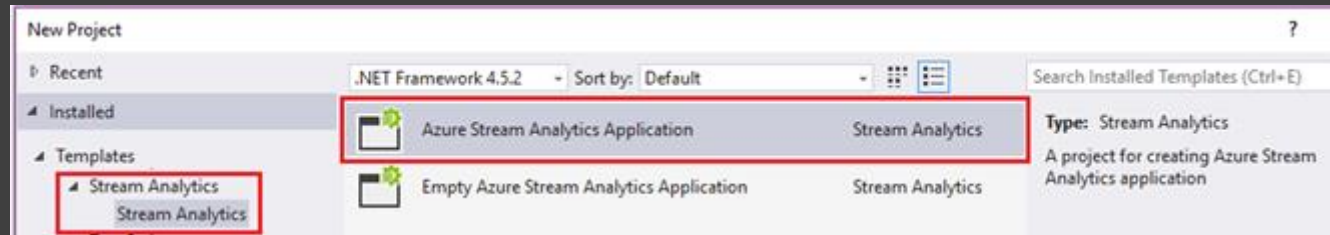Azure Portal

Development and deployment options

Visual Studio
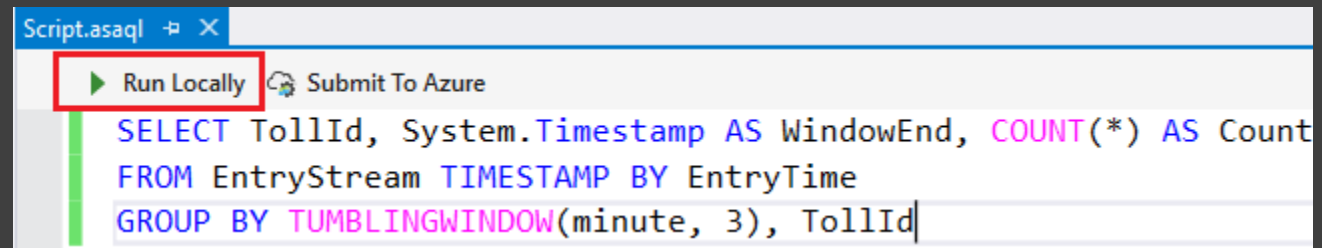
Development and deployment options

PowerShell

- Streaming unit
  - 1 streaming unit = 1MB/Second
    - 1 x 60 x 60 x 24 = 84.375 GB/day
  - Data volume + streaming unit cost
    - Data volume = $ 0.001 per GB
    - Streaming unit = $ 0.12 per hour
  - If run at max through put for a month
    - Total data ingested = 2.47 TB
    - Data = $ 2.50
    - Streaming Unit = $ 86
    - Total cost = $ 88.50

Streaming unit

Demo

DEMO

Create an Event Hub

Create event generator

Connect Event Hub to Stream Analytics

Persist output events to a database

Create a Power BI output for real-time monitoring

Questions

Just like Jimi Hendrix …

We love to get feedback

Please complete the session feedback forms

It's all about the community...

SQL BITS

Please visit Community Corner, we are trying this

year to get more people to learn about the SQL

Community, equally if you would be happy to visit

the community corner we'd really appreciate it.

Thank you