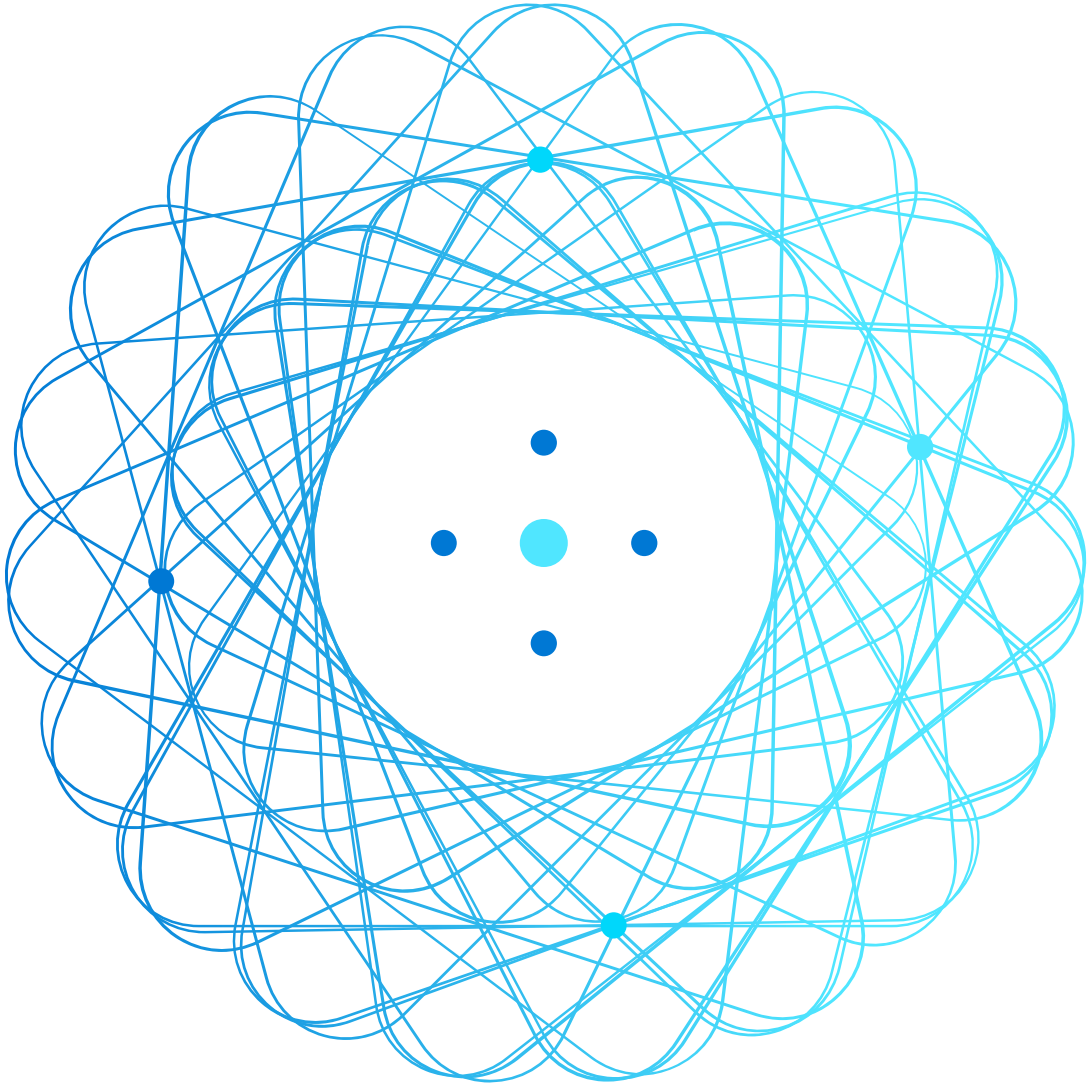
Microsoft Azure

Implement a data streaming solution with Azure Stream Analytics



Agenda

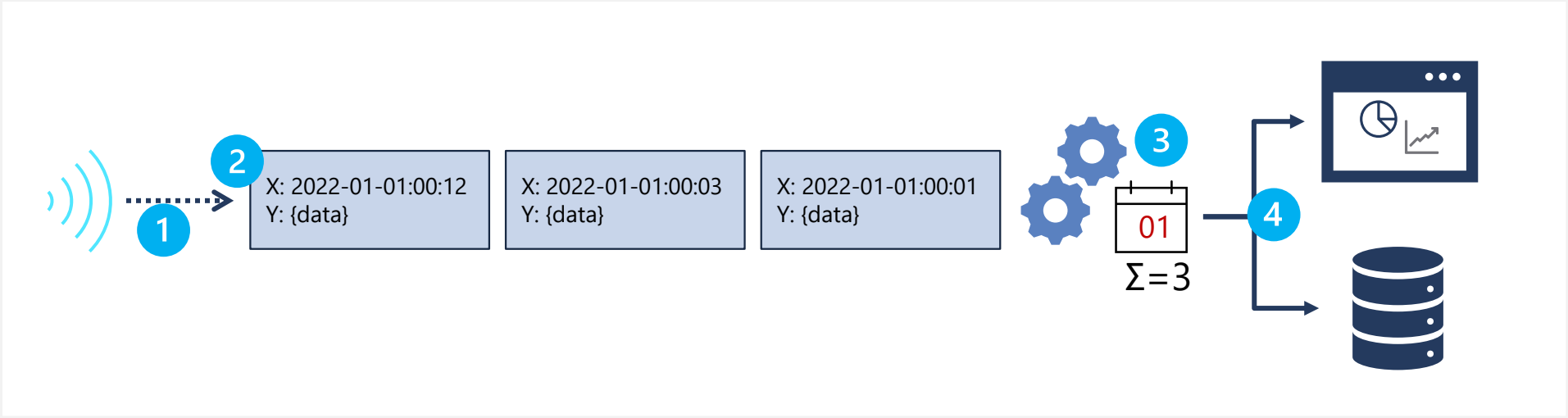
Get started with Azure Stream Analytics

Ingest streaming data using Azure Stream Analytics and Azure Synapse Analytics

Visualize real-time data with Azure Stream Analytics and Power BI

Get started with Azure Stream Analytics

Introduction to data streams



1. Unbounded data source – records added perpetually
2. Each data record represents an event at a specific time
3. Data values can be aggregated over temporal (time-based) windows
4. Results are typically used to support real-time visualization or ingested into an analytical store for historic analysis

Event processing with Azure Stream Analytics

1. Ingest data from an ***input***

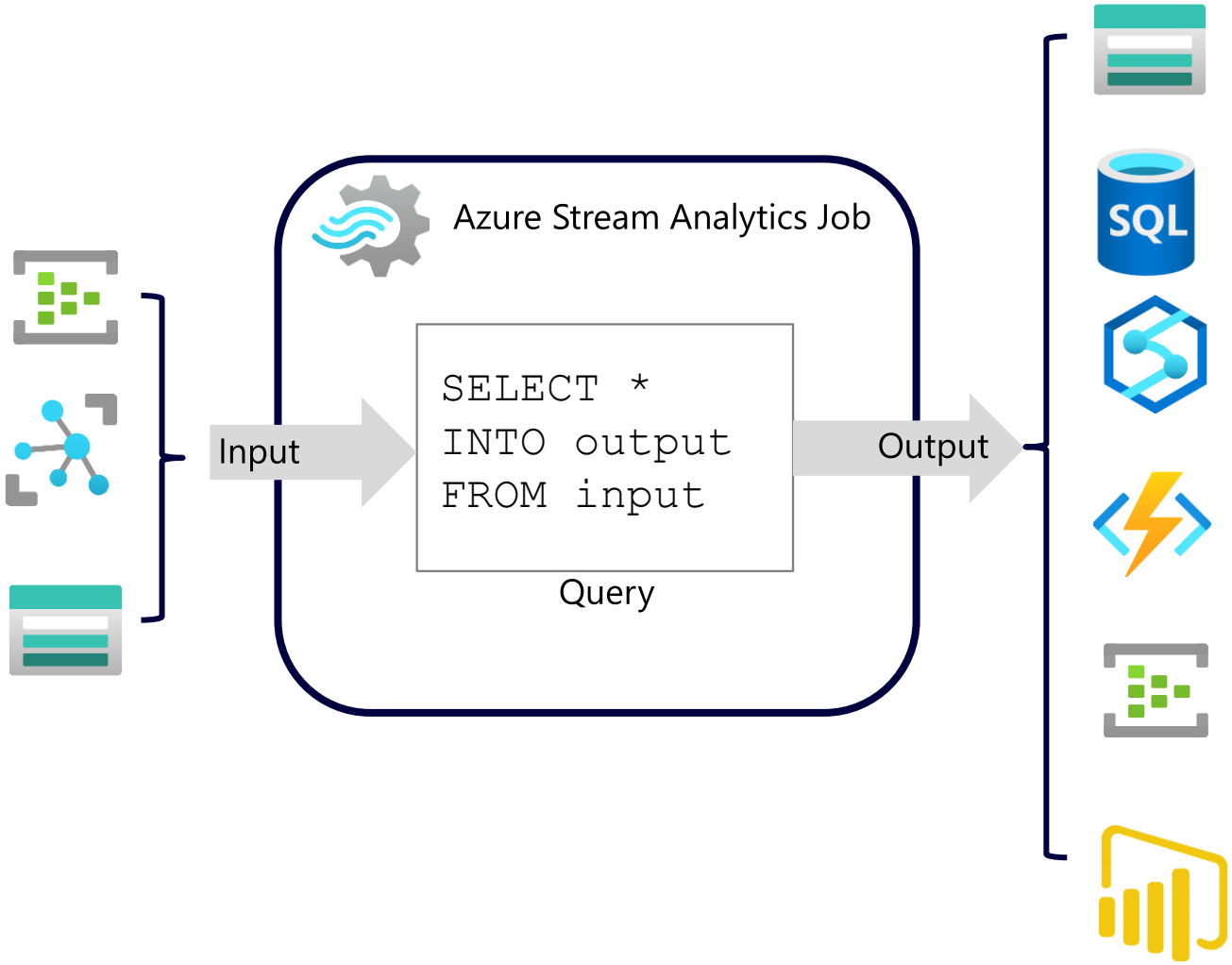
   Azure Event Hubs, Azure IoT Hub, or Azure Storage blob container

2. Process the data by using a ***query***

   Select, filter, and aggregate data values.

3. Write the results to an ***output***

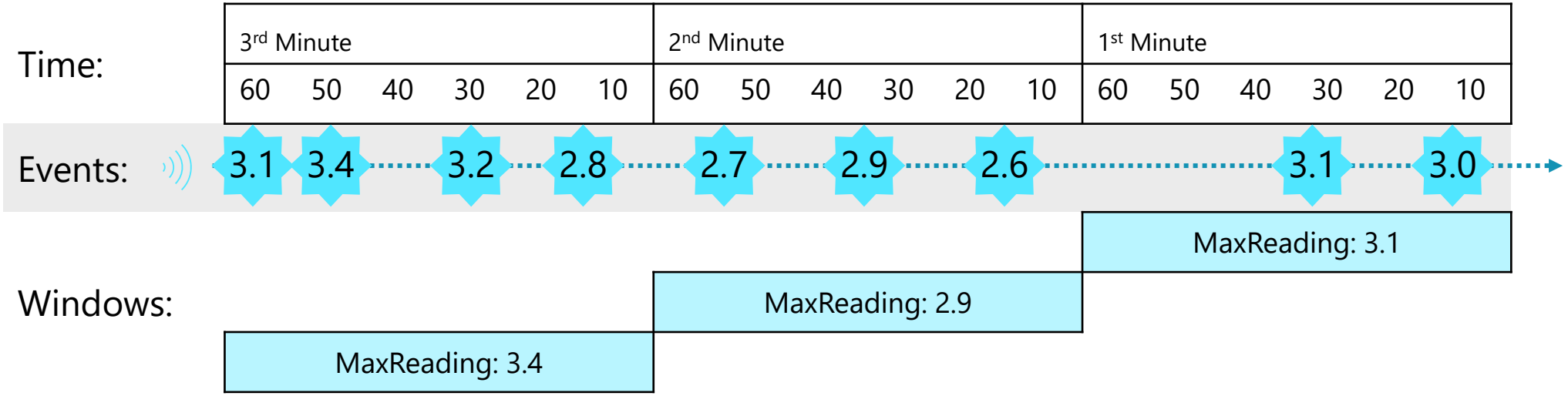
   - Azure Data Lake Gen 2
   - Azure SQL Database
   - Azure Synapse Analytics
   - Azure Functions
   - Azure Event Hubs
   - Microsoft Power BI
   - Others



Window functions – Tumbling Window

Contiguous series of fixed-size, non-overlapping temporal windows

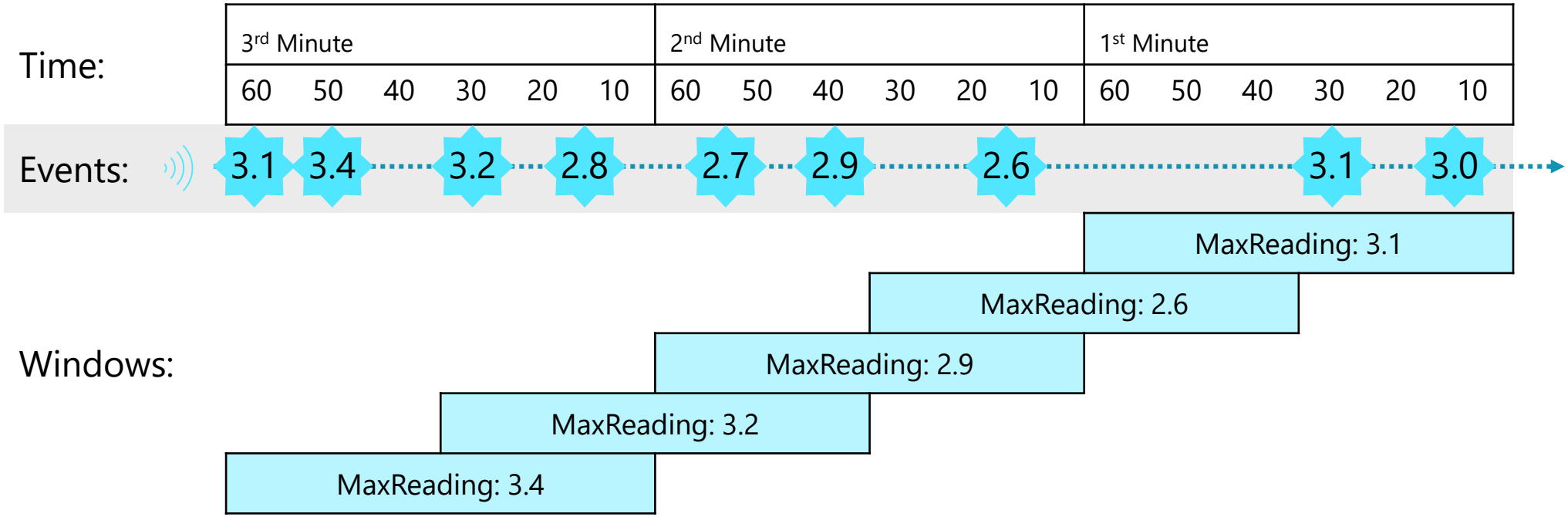
```
SELECT DateAdd(minute,-1,System.TimeStamp) AS WindowStart,
       System.TimeStamp() AS WindowEnd,
       MAX(Reading) AS MaxReading
INTO
    [output]
FROM
    [input] TIMESTAMP BY EventProcessedUtcTime
GROUP BY TumblingWindow(minute, 1)
```



Window functions – Hopping Window

Overlapping windows at fixed intervals

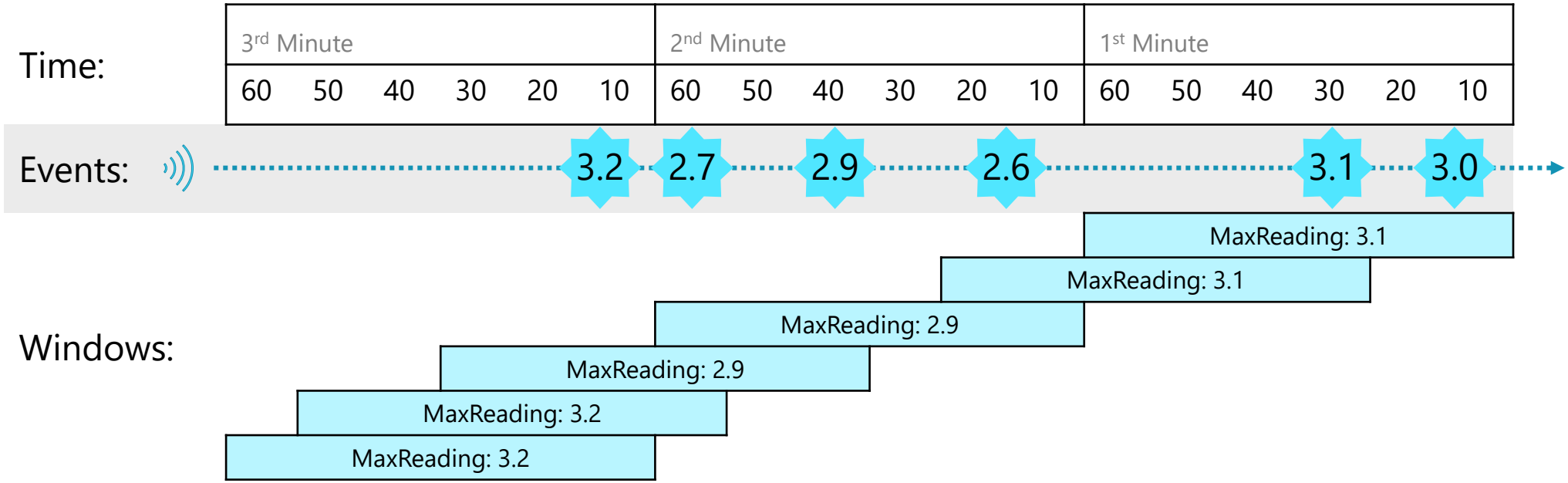
```
SELECT DateAdd(second,-60,System.TimeStamp) AS WindowStart,
       System.TimeStamp() AS WindowEnd,
       MAX(Reading) AS MaxReading
INTO
    [output]
FROM
    [input] TIMESTAMP BY EventProcessedUtcTime
GROUP BY HoppingWindow(second, 60, 30)
```



Window functions – Sliding Window

All windows of a specified duration in which events occur

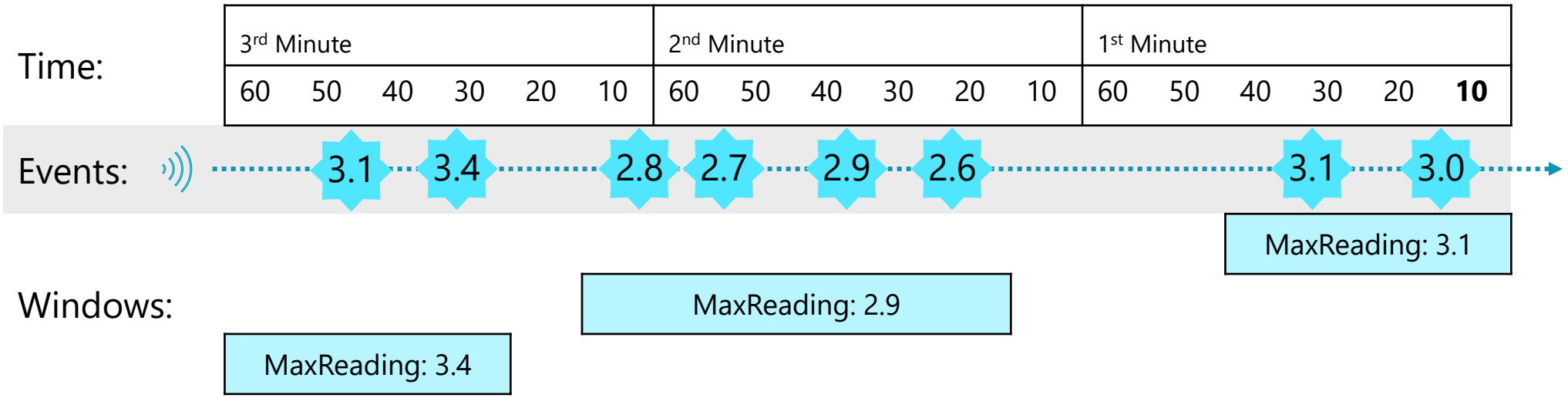
```
SELECT DateAdd(minute,-1,System.TimeStamp) AS WindowStart,
       System.TimeStamp() AS WindowEnd,
       MAX(Reading) AS MaxReading
INTO
    [output]
FROM
    [input] TIMESTAMP BY EventProcessedUtcTime
GROUP BY SlidingWindow(minute, 1)
```



Window functions – Session Window

Variable length windows in which events occur within a specific timeout

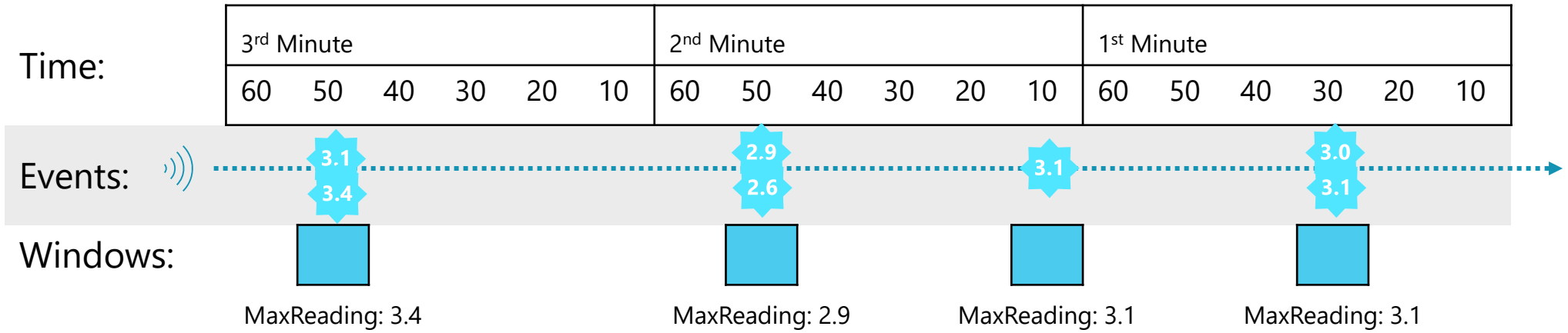
```
SELECT DateAdd(second,-60,System.TimeStamp) AS WindowStart,
       System.TimeStamp() AS WindowEnd,
       MAX(Reading) AS MaxReading
INTO
    [output]
FROM
    [input] TIMESTAMP BY EventProcessedUtcTime
GROUP BY SessionWindow(second, 20, 60)
```



Window functions – Snapshot Window

Windows containing concurrent events with the same timestamp

```
SELECT System.TimeStamp() AS WindowTime,
       MAX(Reading) AS MaxReading
INTO
    [output]
FROM
    [input] TIMESTAMP BY EventProcessedUtcTime
GROUP BY System.Timestamp()
```



Demo: Get started with Azure Stream Analytics

You can try this for yourself later by following the instructions at the link below:

https://aka.ms/mslearn-stream-lab

Knowledge check

You need to process a stream of sensor data, aggregating values over one-minute windows and storing the results in a data lake. Which service should you use?

- ❑ Azure SQL Database
- ❑ Azure Cosmos DB
- ☑ Azure Stream Analytics

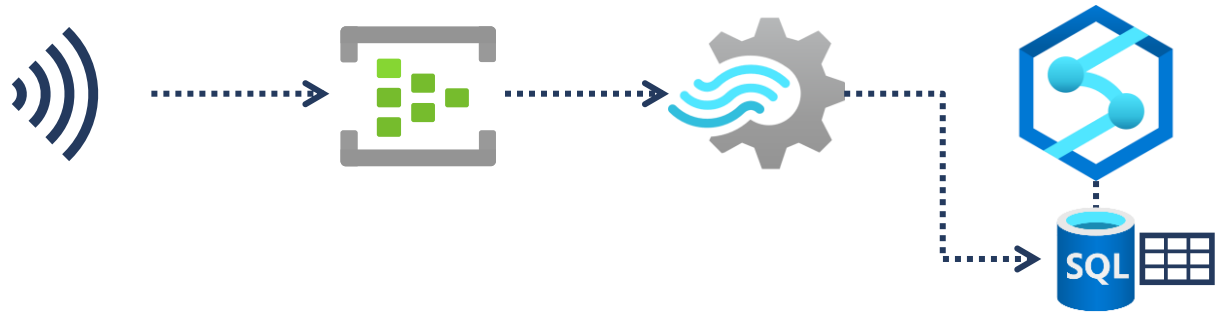
You want to aggregate event data by contiguous, fixed-length, non-overlapping temporal intervals. What kind of window should you use?

- ❑ Sliding
- ❑ Session
- ☑ Tumbling

Ingest streaming data using Azure Stream Analytics and Azure Synapse Analytics

Stream ingestion scenarios

Relational data warehouse



Use an **Azure Synapse Analytics** output

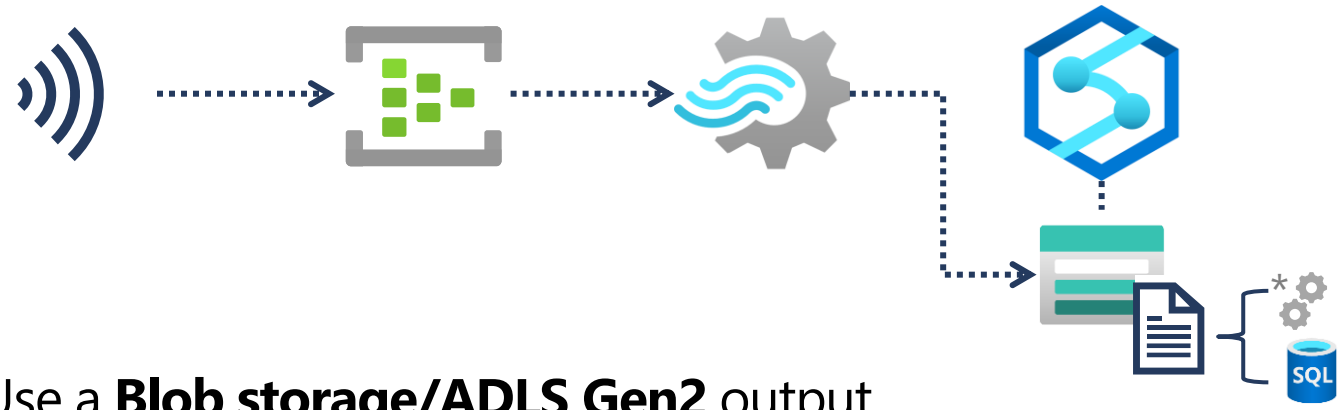
Specify:

- An alias for the output
- The target Synapse Analytics dedicated SQL pool
- The existing table into which the data is to be loaded
- Authentication credentials

Query ingested data using SQL in the dedicated SQL pool

Stream ingestion scenarios

Data Lake



Use a **Blob storage/ADLS Gen2** output

Specify:

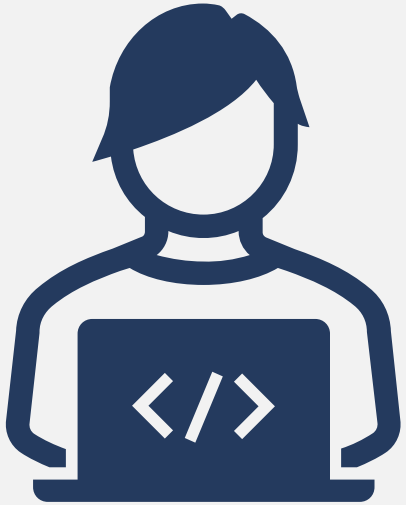
- An alias for the output
- The target Azure Storage account and container
- Authentication credentials
- File format
- Folder pattern (for example, *YYYY/MM/DD*)

Query ingested data using the serverless SQL pool or an Apache Spark pool

Exercise: Ingest streaming data into Azure Synapse Analytics

Use the hosted lab environment provided, or view the lab instructions at the link below:

https://aka.ms/mslearn-stream-analytics-synapse



Knowledge check

Which type of output should you use to ingest the results of an Azure Stream Analytics job into a dedicated SQL pool table in Azure Synapse Analytics?

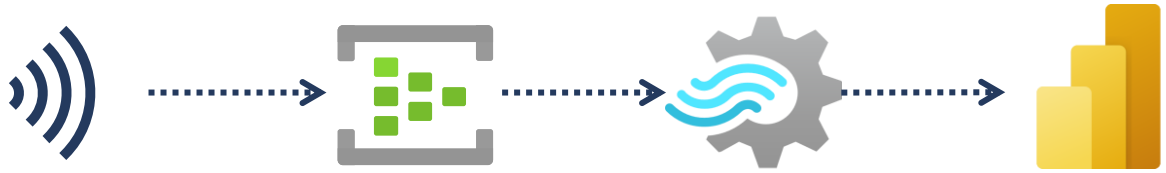
- ☑ Azure Synapse Analytics
- ☐ Blob storage/ADLS Gen2
- ☐ Azure Event Hubs

Which type of output should be used to ingest the results of an Azure Stream Analytics job into files in a data lake for analysis in Azure Synapse Analytics?

- ☐ Azure Synapse Analytics
- ☑ Blob storage/ADLS Gen2
- ☐ Azure Event Hubs

Visualize real-time data with Azure Stream Analytics and Power BI

Azure Stream Analytics and Power BI



Use a **Power BI** output to send streaming query results to Power BI

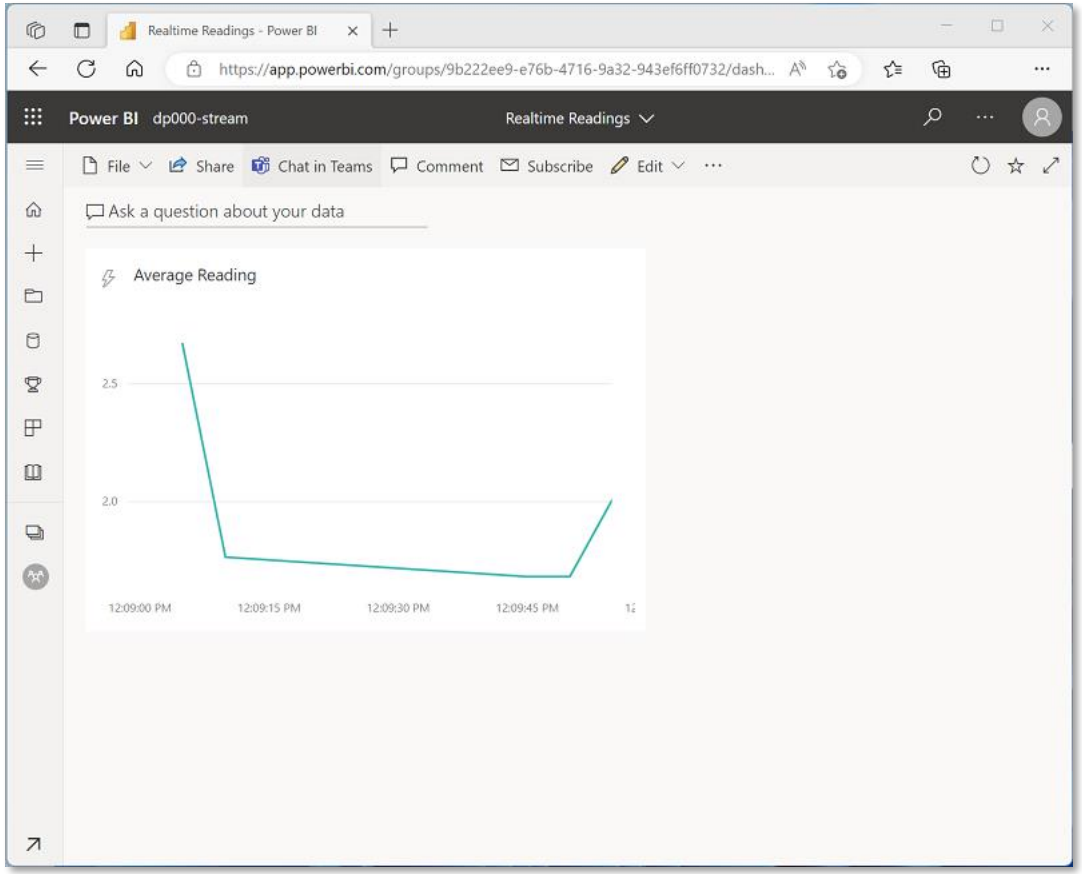
Specify:

- An alias for the output
- The target Power BI workspace
- Names of the dataset and table to be created
- Authentication information

The query creates a streaming dataset with a single table

Real-time data visualizations in Power BI

1. Create a dashboard
2. Add a tile:
   - Select the streaming dataset
   - Configure the desired visualization
3. View the dashboard
   - The visualization is updated in real-time as new data arrives in the table



Demo: Create a real-time data visualization

You can try this for yourself later by following the instructions at the link below:

https://aka.ms/mslearn-stream-powerbi

Knowledge check

Which type of Azure Stream Analytics output should you use to support real-time visualizations in Microsoft Power BI?

- ❑ Azure Synapse Analytics
- ❑ Azure Event Hubs
- ☑ Power BI

You want to use an output to write the results of a Stream Analytics query to a table in a dataset in a Power BI workspace. What should you do?

- ☑ Create only the workspace before creating the output
- ❑ Create the workspace and dataset before creating the output
- ❑ Create the workspace, dataset, and table before creating the output

You want to create a visualization that updates dynamically based on a table in a streaming dataset in Power BI. What should you do?

- ❑ Create a report from the dataset
- ☑ Create a dashboard with a tile based on the streaming dataset
- ❑ Export the streaming dataset to Excel and create a report from the Excel workbook

Further reading

Implement a Data Streaming Solution with Azure Stream Analytics
https://aka.ms/mslearn-azure-stream-analytics