Aigars Mačiņš | @Owandir

Program Manager,
Localization

Skype, Microsoft

Introduction into building IoT Backend Achitecture

TTU TechnoHack 22.10.2016

Agenda

Patterns and anti-patterns

Demo

Many IoT solutions control critical operations at the core of industrial and civil infrastructure.

Digital security will be increasingly interwoven with physical safety of life and equipment.

Many IoT solutions will provide very deep and near-real time insight into industrial and business processes, as well as into homes and the immediate personal environment.

Privacy matters.

Pattern: Think big. Start small

Think big, but start small.

Experiment, learn and refine.

- Build to an architecture that will scale, but start prototyping with a small number of devices.
- It's hard to predict what data provides value -- which impacts which sensors and devices are necessary -- until you build something.
- It's much easier to work through through device identity, management/update and security at small scale.

IoT architecture requirements

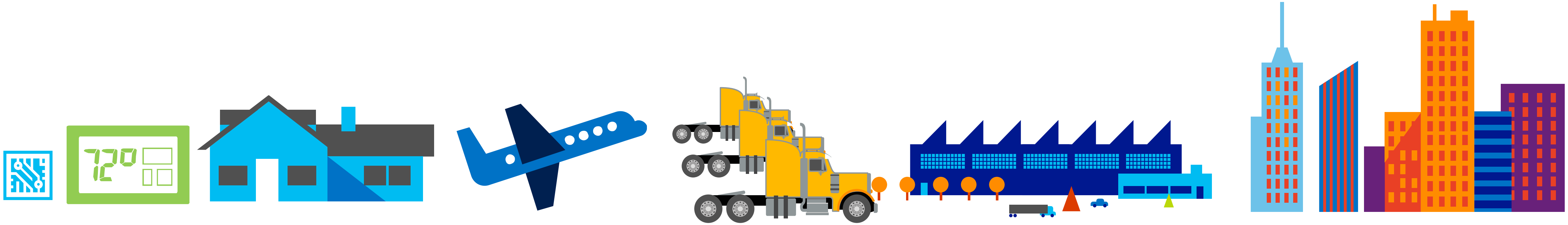
Handle extreme hardware and software heterogeneity.

Lower barriers to entry: evaluate -> prototype -> deploy.

Provide hot-path and cold-path analysis and response.

Build for hyper-scale and enable low latency.

Be secure by design; support defense in depth.



Pattern: Telemetry first

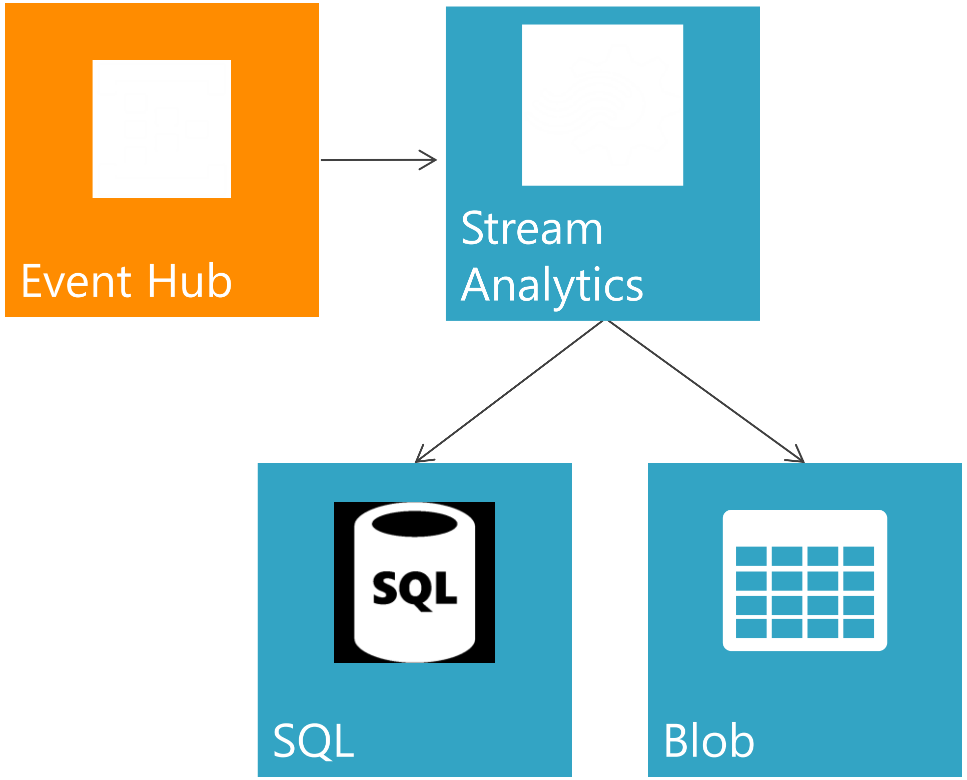
Start with telemetry.

The important data may not be what you expected.

Address privacy, security and manageability before moving to command and control.

- It is very hard to predict in advance what data will be useful.
- It is tempting, but likely inefficient to try for business transformation in the first step.
- Think about not only device telemetry but also diagnostic telemetry.
- Privacy and security implications of telemetry are generally lesser than for command and control.

Telemetry



- High scale data ingestion via Event Hub.
- High scale stream processing via Stream Analytics (or HDInsight /Storm)
- Storage for cold-path analytics
- Processing for hot-path analytics

Event Hubs and Stream Analytics

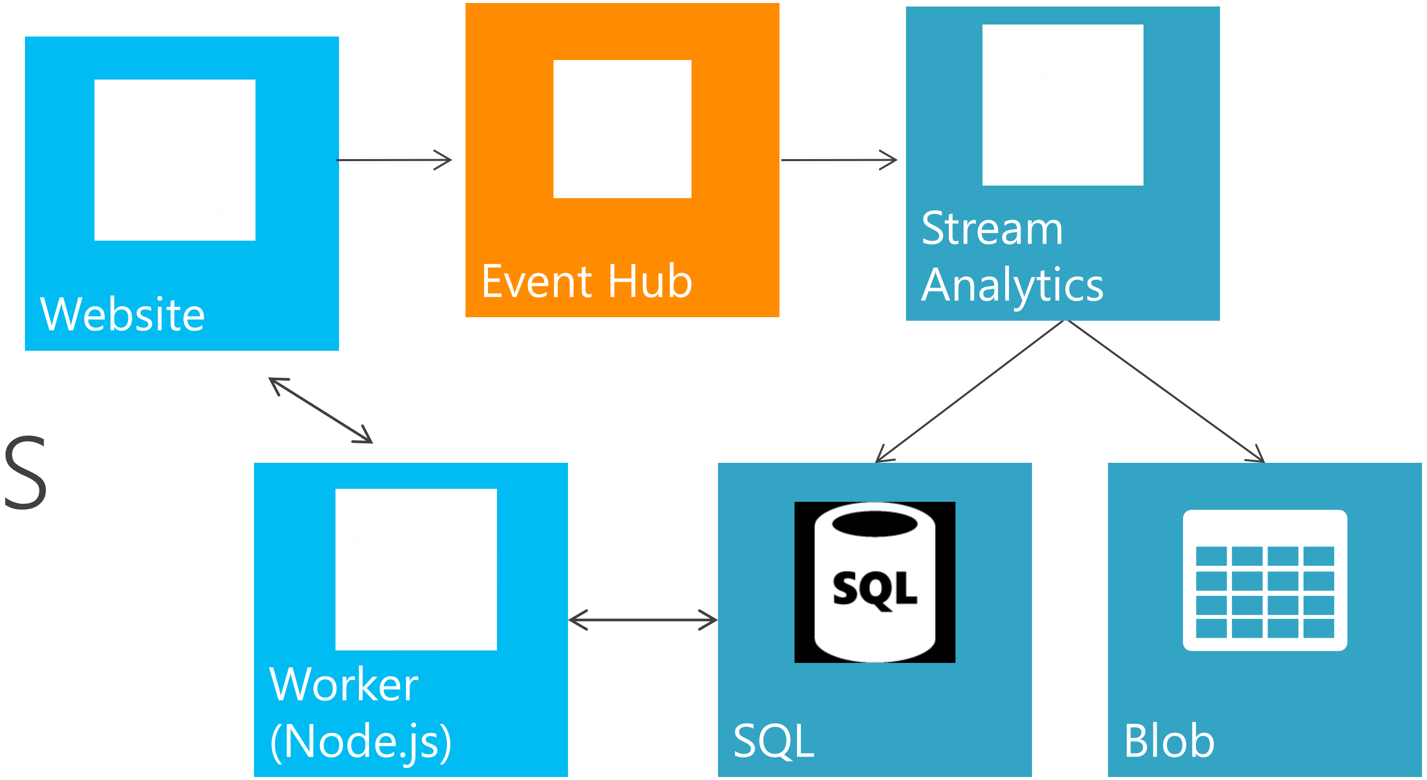
Event Hub

- Cloud-scale telemetry ingestion from websites, apps, and devices
- Compatible with more than a million publishers supporting HTTP, AMQP and MQTT
- Ingress millions of events per second
- SAS based security, with unique token per publisher
- Configurable data retention (1-30 days)
- Low latency (<10 ms for volatile data)
- Pluggable with other cloud services like Stream Analytics

Stream Analytics

- Real-time analytics for Internet of Things solutions
- Stream millions of events per second
- Mission critical reliability, performance and predictable results
- Rapid development with familiar SQL-based language

Event Hub and Stream Analytics



JavaScript (to Event Hub)

```
var eventBody = { "reading": x, "device_id": id };

ehClient = new EventHubClient({
        'name': «ttu-demo», 'namespace': «ttu-demo-ns",
        'sasKey': <snipped>, 'sasKeyName': "sendTelemetry",
        'timeOut': 10,
    });

var msg = new EventData(eventBody);

ehClient.sendMessage(msg, function (messagingResult) {
      // <body snipped>
});
```

Stream Analytics (to blob)

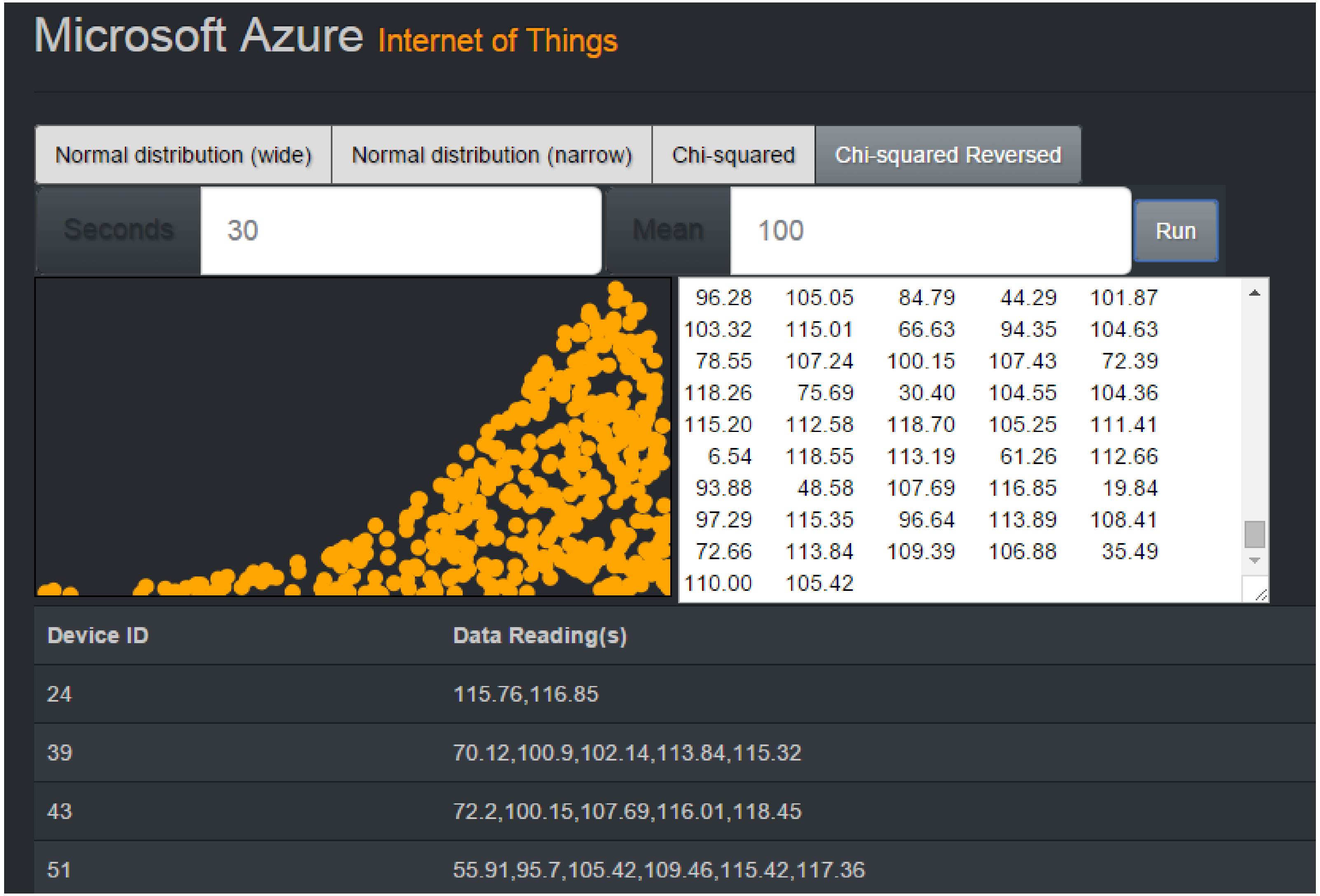
```
SELECT
    device_id as Device_Id,
    reading as Reading,
    EventProcessedUtcTime as UTCDateTime
FROM [eventhub]
INTO [out2blob]
```

Stream Analytics (to SQL)

```
SELECT
    System.TimeStamp as UTCDateTime, device_id as Device_Id,
    COUNT (*) as Count
FROM [ttu-demo] TIMESTAMP BY EventProcessedUtcTime
INTO [alertCounts]
WHERE ( CAST(reading AS float) > 115.0 )
GROUP BY device_id, SlidingWindow(second, 15)
HAVING COUNT(*) > 1

SELECT
    device_id as Device_Id, reading as Reading,
    EventProcessedUtcTime as UTCDateTime
FROM [iotdemoeventhub] TIMESTAMP BY EventProcessedUtcTime
INTO [stream2sql]
```

Data


Microsoft Azure Internet of Things
Normal distribution (wide)
Normal distribution (narrow)
Chi-squared
Chi-squared Reversed
Run

Seconds 30

Mean 100

```
 96.28   105.05    84.79    44.29   101.87
103.32   115.01    66.63    94.35   104.63
 78.55   107.24   100.15   107.43    72.39
118.26    75.69    30.40   104.55   104.36
115.20   112.58   118.70   105.25   111.41
  6.54   118.55   113.19    61.26   112.66
 93.88    48.58   107.69   116.85    19.84
 97.29   115.35    96.64   113.89   108.41
 72.66   113.84   109.39   106.88    35.49
110.00   105.42
```

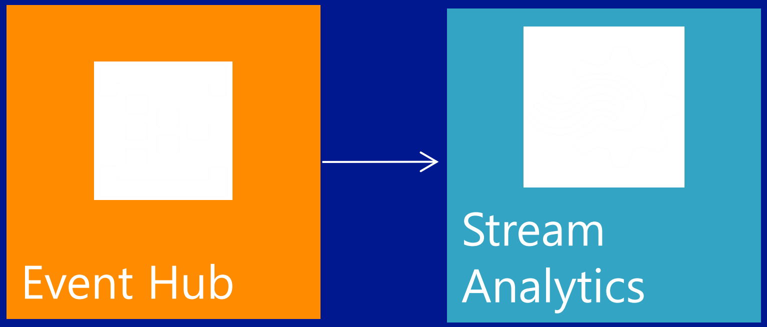
Device ID	Data Reading(s)
24	115.76,116.85
39	70.12,100.9,102.14,113.84,115.32
43	72.2,100.15,107.69,116.01,118.45
51	55.91,95.7,105.42,109.46,115.42,117.36

Pattern: Don’t interrupt the fast path

Don’t accidentally create processing bottlenecks. (Think carefully before interrupting data flow between high-scale components.)

“Don’t stick your head in the fire hose unless you know what you’re doing.”

- In the telemetry example, Event Hub data flows directly into Stream Analytics.

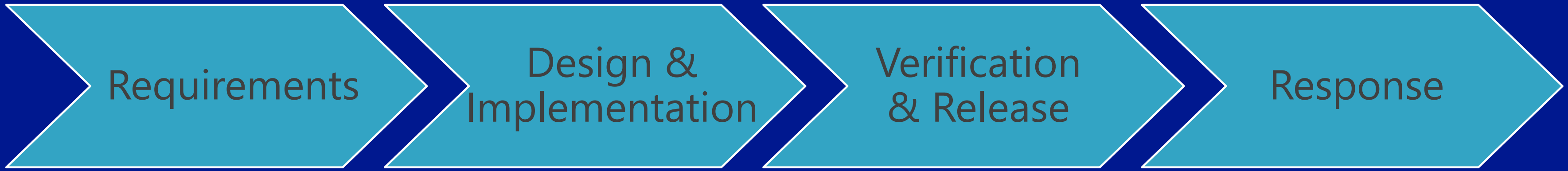


- Both components are designed for high scale.
- Don’t process between high-scale components unless you can handle that scale.

Pattern: Defense in depth

Think about security, identity and management from the very beginning.

The entire organization needs to be focused on security, and that focus must inform the entire product lifecycle.



Requirements

Design & Implementation

Verification & Release

Response

Security is a shared responsibility between Azure and the customer.

Think about security on the device, at the field gateway (if one exists) and in the cloud.

- Physical Security, Tamper Detection
- Hardware & firmware security, secure boot
- Network, protocol & application security
- Identity management for devices and users
- Data Privacy Protection and Controls

Azure IoT Suite

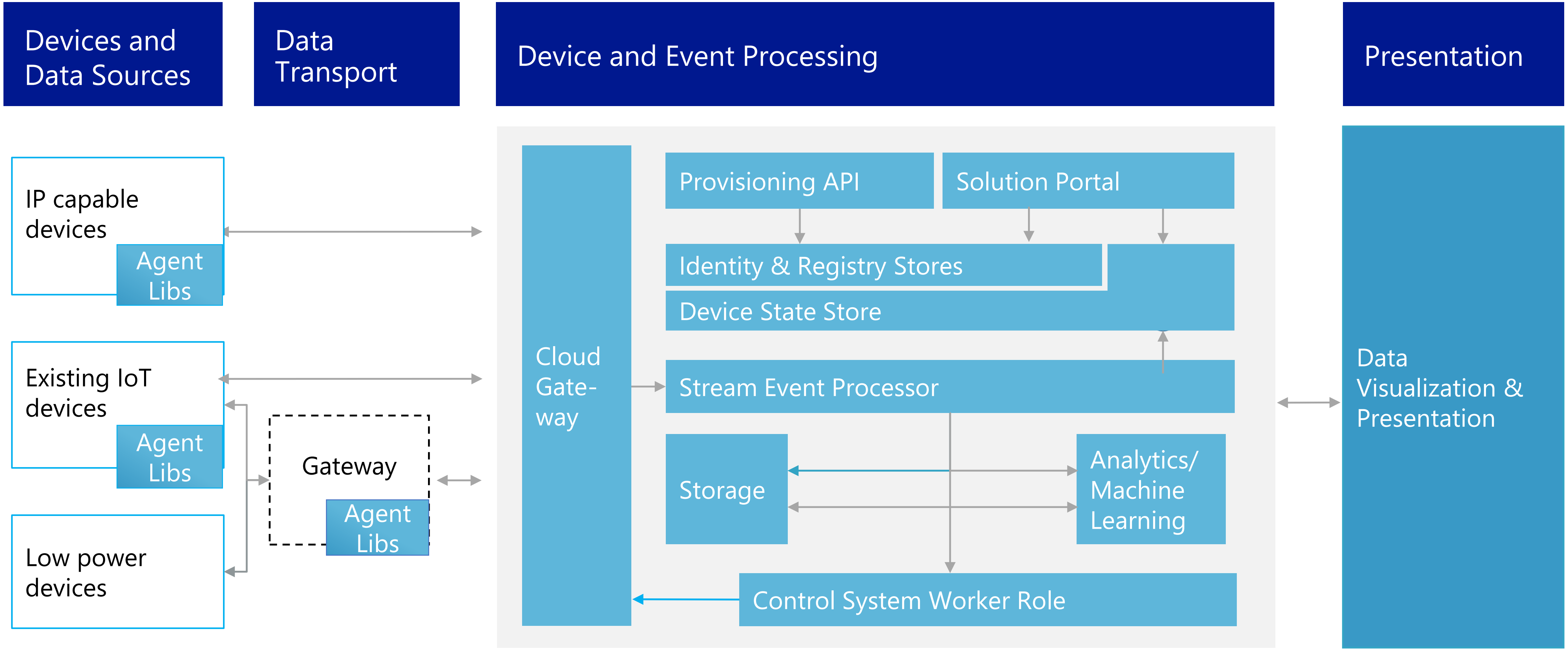
- **Accelerate time-to-value** by easily deploying IoT applications for the most common use cases, such as remote monitoring, asset management, and predictive maintenance

- **Plan and budget appropriately** through a simple predictable business model

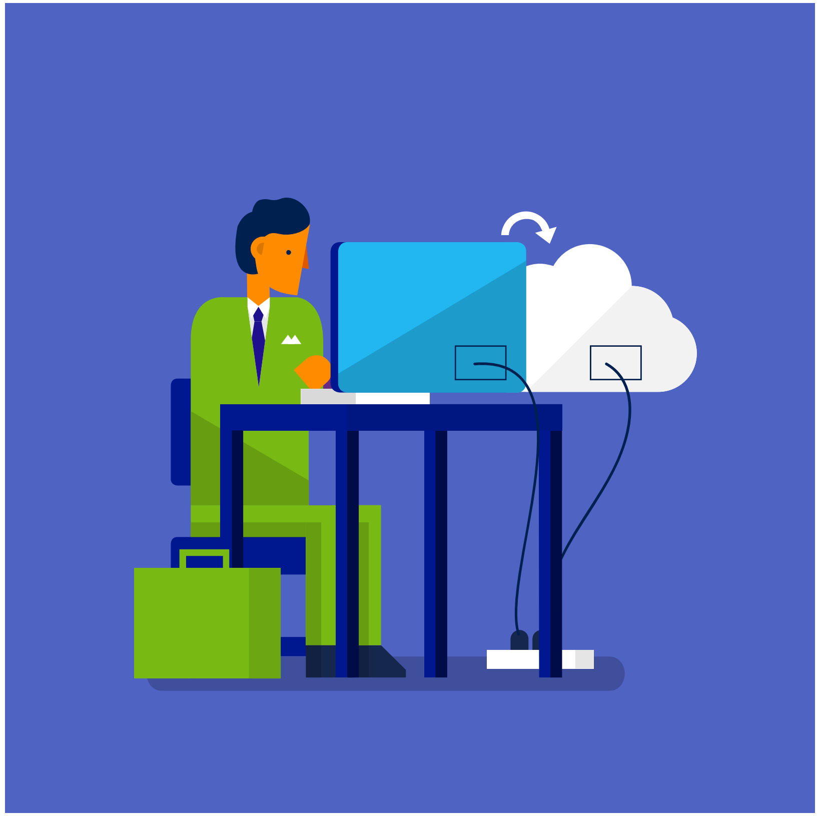
- **Grow and extend** solutions to support millions of assets

Connect assets
Mine data
Take action

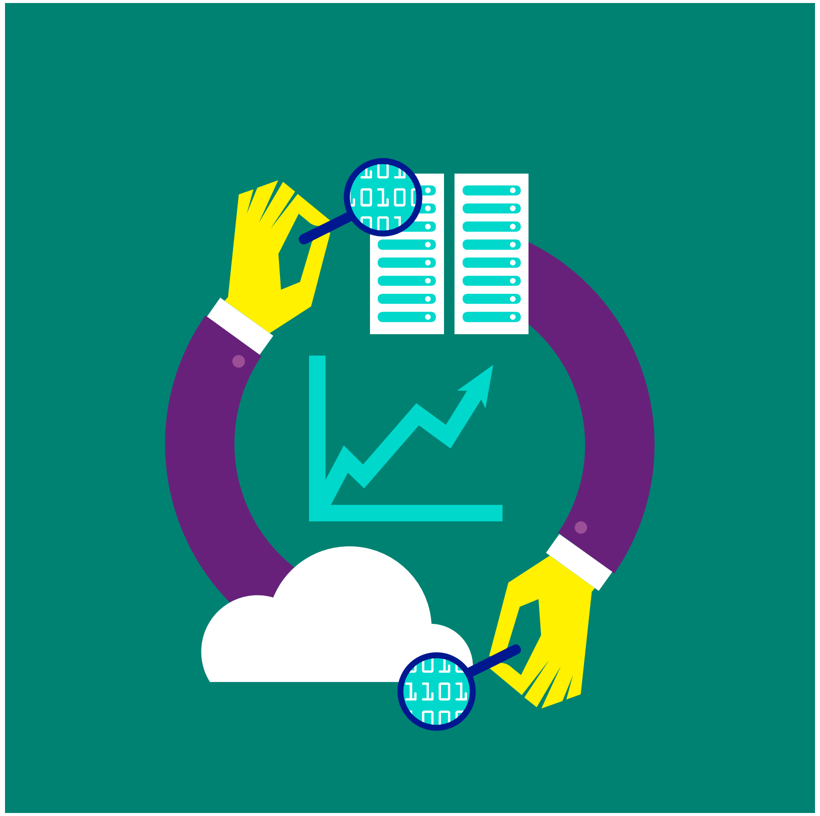
Azure IoT Reference Architecture



Canonical scenarios



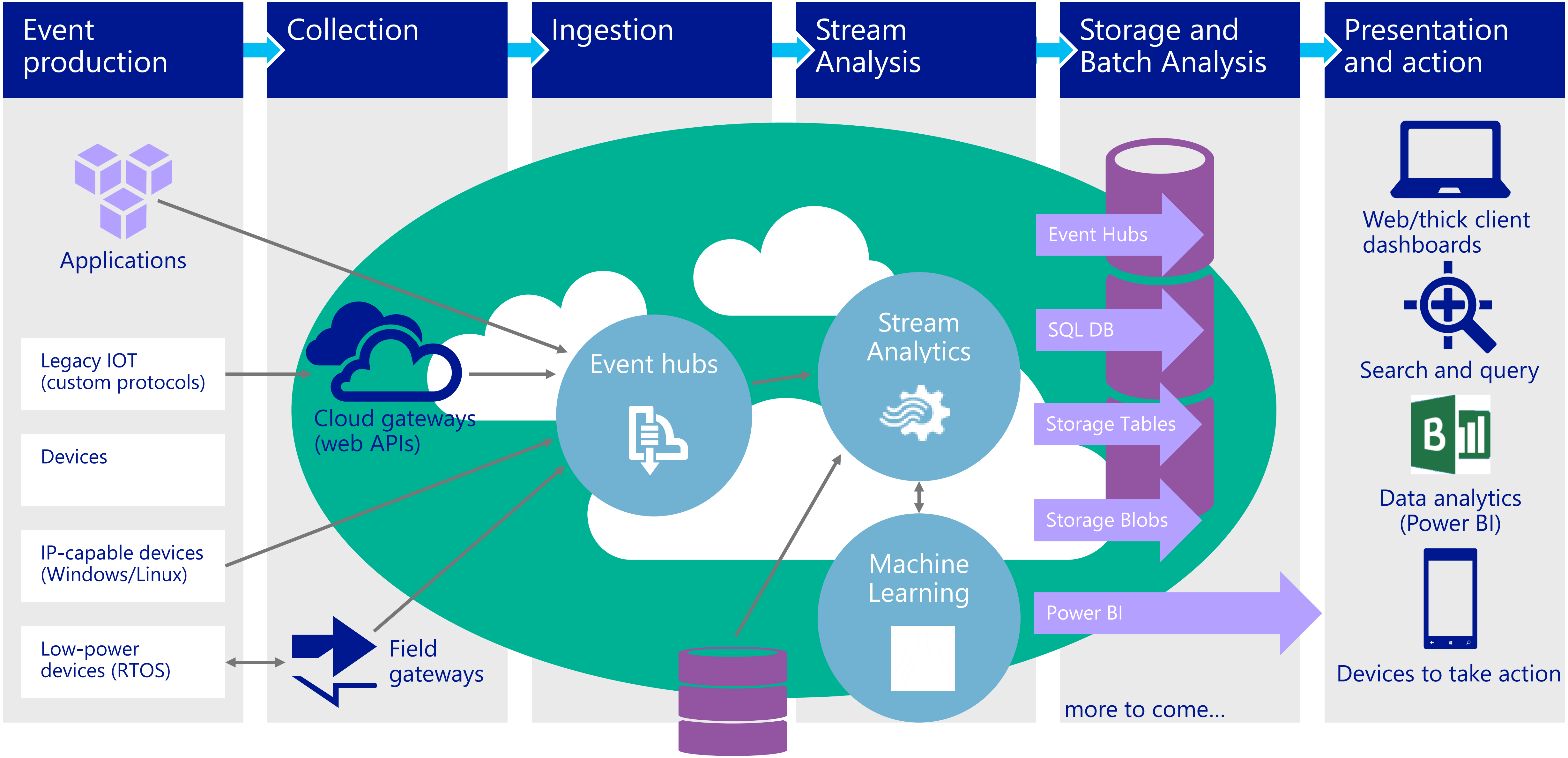
ARCHIVING



DASHBOARDING

TRIGGERING WORKFLOWS

Canonical Stream Analytics Pattern



Event production

Collection

Ingestion

Stream Analysis

Storage and Batch Analysis

Presentation and action

Applications

Legacy IOT (custom protocols)

Devices

IP-capable devices (Windows/Linux)

Low-power devices (RTOS)

Cloud gateways (web APIs)

Field gateways

Event hubs

Stream Analytics

Machine Learning

Event Hubs

SQL DB

Storage Tables

Storage Blobs

Power BI

more to come…

Web/thick client dashboards

Search and query

Data analytics (Power BI)

Devices to take action

Azure Stream Analytics

Fully managed real-time analytics

Mission critical reliability and scale

Enables rapid development

Scenario – Twitter Analytics

"A news media website wants to increase site traffic by covering trending topics on social media."

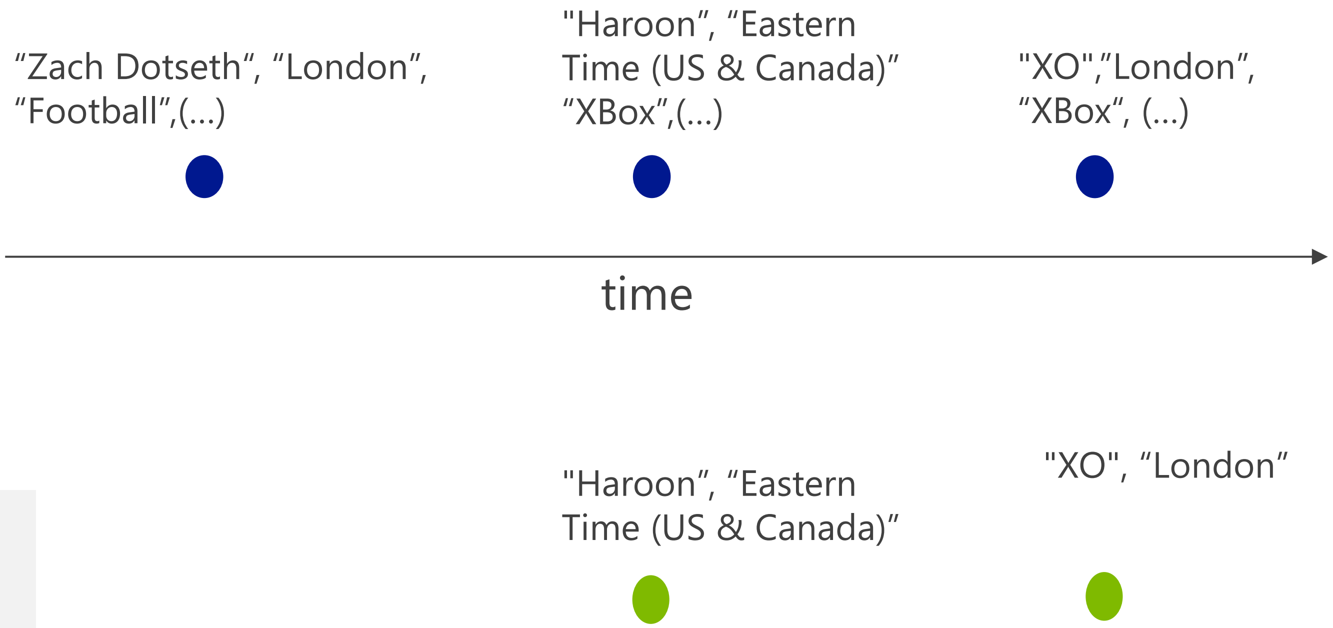
To determine which topics are immediately relevant to customers, they need real-time analytics about the tweet volume and sentiment for each topic.

TwitterStream

ID	CreatedAt	UserName	TimeZone	Text	Language	Topic
1	2015-04-30T20:45:30	Joshua X	Eastern Time (US & Canada)	Oh, joy! More forced @Xbox Live updates	en	XBox
2	2015-04-30T20:45:31	Cristabel Y	London	RT @verge: Streaming Xbox One games ..	en	XBox
	…					

Filters

Show me the user name and time zone of tweets on the topic XBox



“Zach Dotseth“, “London”, “Football”,(…)

"Haroon", “Eastern Time (US & Canada)” “XBox”,(…)

"XO",”London”, “XBox“, (…)

time

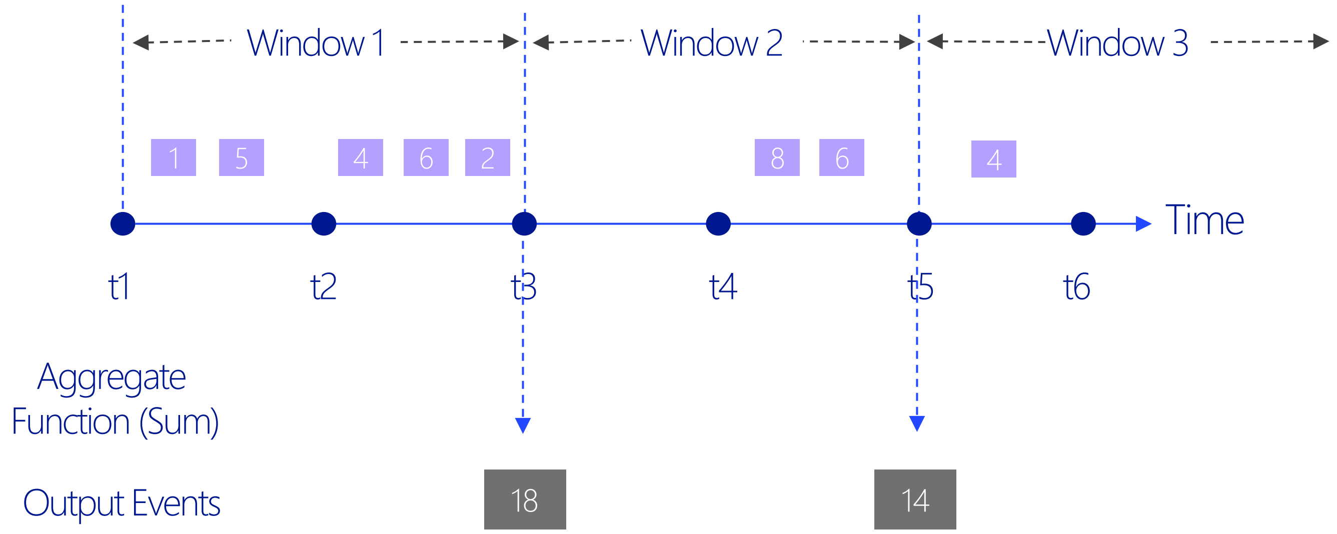
"Haroon", “Eastern Time (US & Canada)”

"XO", “London”

```
SELECT UserName, TimeZone
FROM InputStream
WHERE Topic = 'XBox'
```

Windowing Concepts

- Windows can be **tumbling**, **hopping**, or **sliding**
- Windows are fixed length
- Must be used in a **GROUP BY** clause

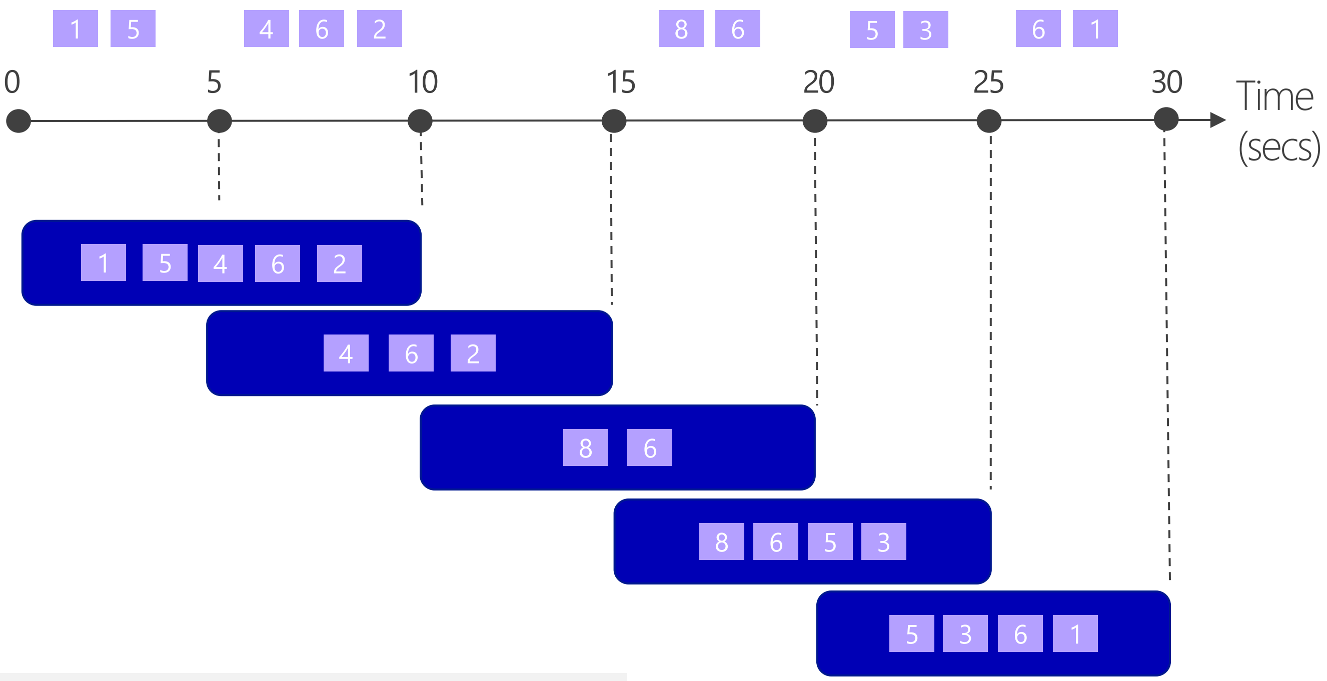


- Output event will have the timestamp of the end of the window

Hopping Windows

“Every 5 seconds give me the count of tweets over the last 10 seconds”

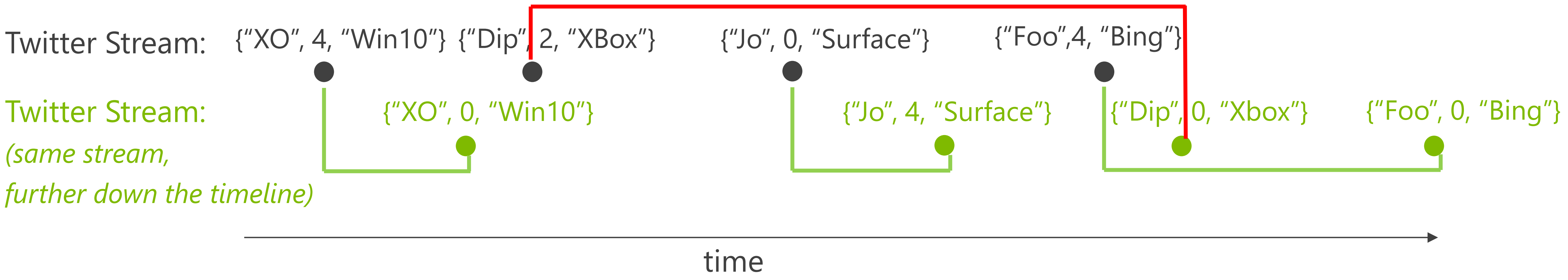
A 10-second Hopping Window with a 5-second “Hop”



```
SELECT Topic, Count(*) AS TotalTweets
FROM TwitterStream TIMESTAMP BY CreatedAt
GROUP BY Topic, HoppingWindow(second, 10, 5)
```

Joining multiple streams

“List all users and the topics on which they switched their sentiment within a minute”



Twitter Stream: {“XO”, 4, “Win10”} {“Dip”, 2, “XBox”} {“Jo”, 0, “Surface”} {“Foo”,4, “Bing”}

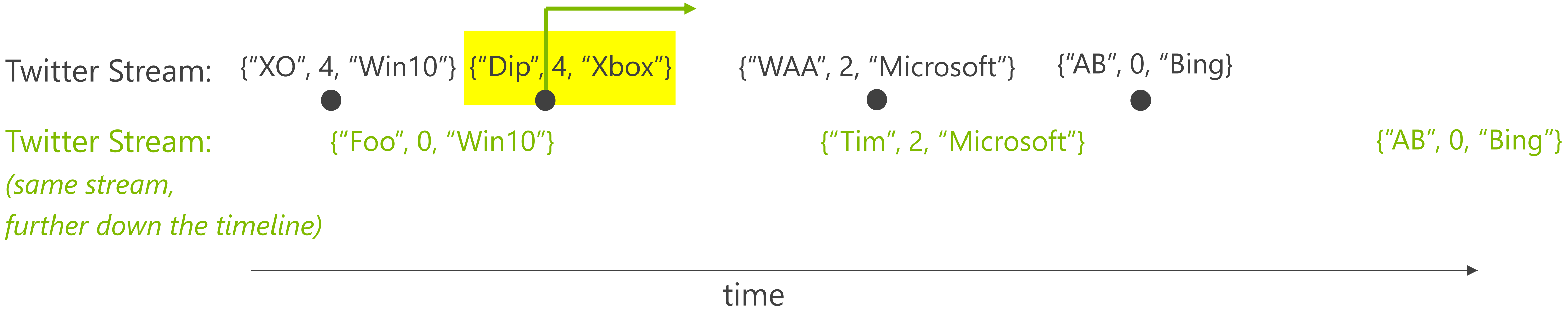
Twitter Stream: *(same stream, further down the timeline)* {“XO”, 0, “Win10”} {“Jo”, 4, “Surface”} {“Dip”, 0, “Xbox”} {“Foo”, 0, “Bing”}

time

```
SELECT TS1.UserName, TS1.Topic
FROM TwitterStream TS1 TIMESTAMP BY CreatedAt
JOIN TwitterStream TS2 TIMESTAMP BY CreatedAt
      ON TS1.UserName = TS2.UserName AND TS1.Topic = TS2.Topic
      AND DateDiff(second, TS1, TS2) BETWEEN 1 AND 60
WHERE TS1.SentimentScore != TS2.SentimentScore
```

Detecting absence of events

“Show me if a topic is **not** tweeted for 10 seconds since it was last tweeted”



Twitter Stream: {“XO”, 4, “Win10”} {“Dip”, 4, “Xbox”} {“WAA”, 2, “Microsoft”} {“AB”, 0, “Bing}

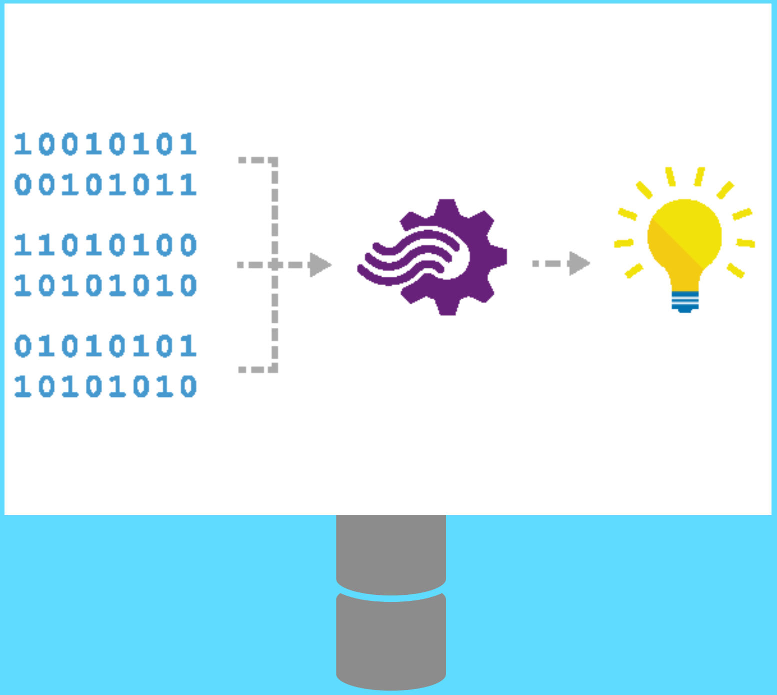
Twitter Stream: {“Foo”, 0, “Win10”} {“Tim”, 2, “Microsoft”} {“AB”, 0, “Bing”}

(same stream,
further down the timeline)

time

```
SELECT TS1.CreatedAt, TS1.Topic, TS1.UserName
FROM TwitterStream TS1 TIMESTAMP BY CreatedAt
LEFT OUTER JOIN TwitterStream TS2 TIMESTAMP BY CreatedAt
      ON TS1.Topic = TS2.Topic
      AND DateDiff(second, TS1, TS2) BETWEEN 1 AND 10
WHERE TS2.Topic IS NULL
```

Reference Data



Seamless correlation of event streams with reference data

Static or slowly-changing data stored in blobs

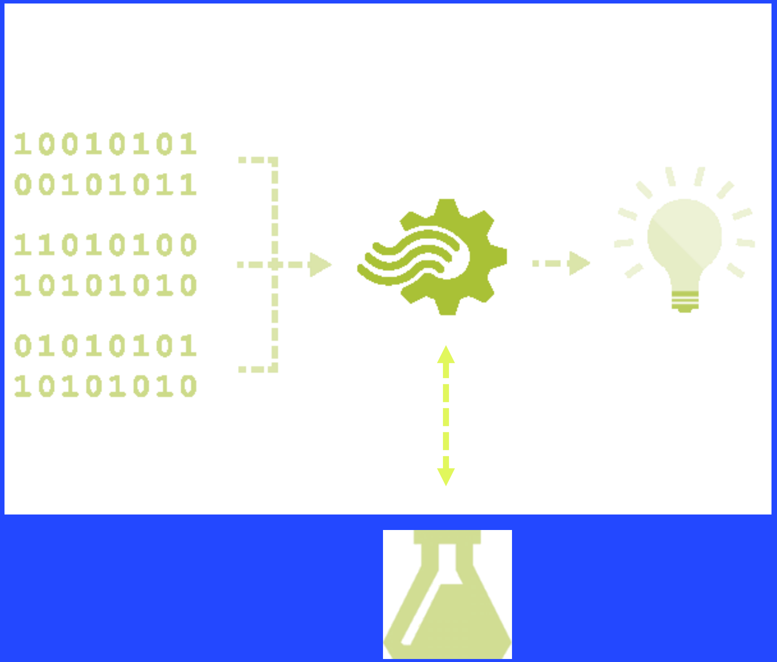
CSV and JSON files in Azure Blobs;
scanned for new snapshots on a settable cadence

JOIN (INNER or LEFT OUTER) between streams and reference data sources

Reference data appears like another input:

```
SELECT myRefData.Name, myStream.Value
FROM myStream
JOIN myRefData
       ON myStream.myKey = myRefData.myKey
```

Machine Learning



Combine that with Azure ML

Azure ML can publish web endpoints for operationalized models

Azure Stream Analytics can bind custom function names to such web endpoints

Example: apply bound function event-by-event

```
SELECT text,                AS score
FROM myStream
```

Summary

Think big (architecture), but start small (experiment, learn and refine).

Start with telemetry. Address privacy, security and manageability before moving to command and control.

Don’t interrupt the fast path and create processing bottlenecks.

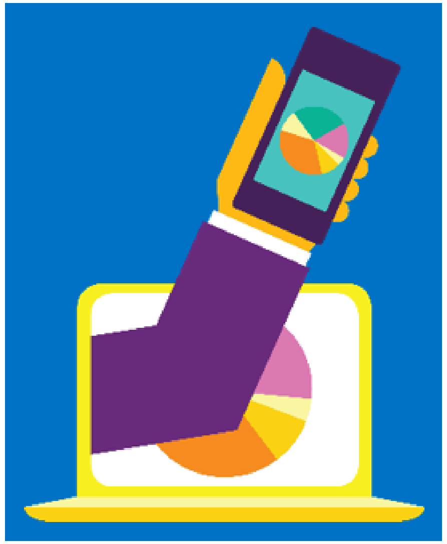
Think about security, identity and management from the very beginning, and through the life of the product.

Build to the reference architecture to ease the move to IoT Suite.

Microsoft

skype™

Resource Library



Business Overview	http://azure.microsoft.com/en-us/services/stream-analytics/
Documentation	http://azure.microsoft.com/en-us/documentation/services/stream-analytics/
Samples	https://github.com/streamanalytics/samples
ASA Blog	http://blogs.msdn.com/b/streamanalytics/rss.aspx
Follow us on Twitter	https://twitter.com/AzureStreaming (follow @AzureStreaming)
ASA Forum	https://social.msdn.microsoft.com/Forums/en-US/home?forum=AzureStreamAnalytics
Vote for ideas	http://feedback.azure.com/forums/270577-azure-stream-analytics
Email ASA Team	azstream@microsoft.com