

Architecting IoT Data Ingestion with Azure Data Services

MIKE BENKOVICH

mike@benko.com | www.Benkotips.com

TITANIUM SPONSORS



Platinum Sponsors



Gold Sponsors



Today

Hello IoT

Event Hubs

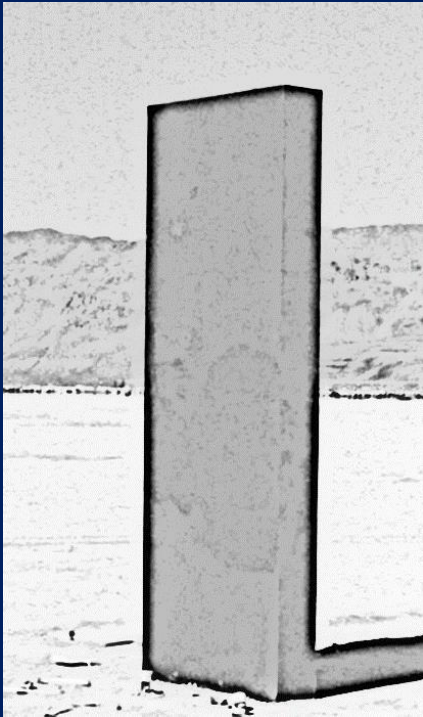
Streaming Data

Functions

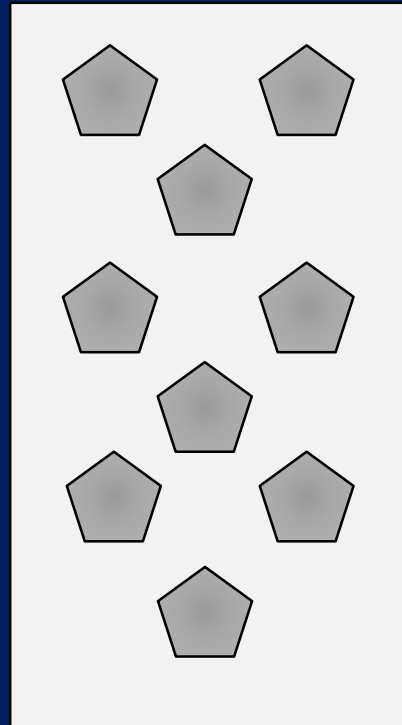
Machine Learning

Patterns

Monolith



MicroService



Serverless



Hello IoT

Internet of...

IoC	Internet of Coffee
IoG	Internet of Gadgets
IoT	Internet of Toast
IoB	Internet of Biology
IoH	Internet of Home
IoM	Internet of Media
IoS	Internet of Stuff

The Internet of Things (IoT)

Currently 20 billion devices connected to the Internet

By 2020, expect 50 billion or more

- Health-monitoring devices
- Thermostats, wind turbines, and solar farms
- Cars, trucks, traffic lights, and drones
- **EVERYTHING** will be connected

How do you process all that data?

How do you process it in real time?



IoC = Internet of Cloud

Canonical scenarios



ARCHIVING



DASHBOARDING



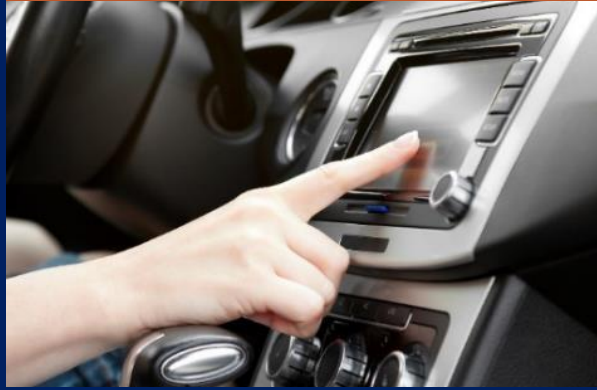
TRIGGERING WORKFLOWS

What are customers wanting to do?

Real-time fraud detection



Connected car scenario



Click-stream analysis



Real-time financial portfolio alerts



Smart grid



CRM alerting sales with customer scenario



Data and identity protection services



Real-time financial sales tracking



Cognition

NEC FieldAnalyst (People count / Age / Gender) data to Azure in real-time

The screenshot displays the NEC FieldAnalyst software interface. On the left, a summary panel shows statistics: Entry (1145), Exit (1772), and a table of detection times, genders, and ages. The main window shows a live video feed of a crowded exhibition hall with people's faces tracked by blue and purple bounding boxes and labeled with age ranges like 'age:40-59' and 'age:25-39'.

Summary Panel Data:

Category	Count	Unit
Entry	1145	
Exit	1772	

Detection time	Gender	Age
2015/04/13 17:11:18	♂	40-59
2015/04/13 17:11:19	♂	40-59
2015/04/13 17:11:21	♂	60+
2015/04/13 17:11:25	♂	40-59
2015/04/13 17:11:27	♂	60+
2015/04/13 17:11:49	♂	40-59
2015/04/13 17:11:49	♂	40-59
2015/04/13 17:11:51	♂	40-59
2015/04/13 17:12:06	♂	25-39
2015/04/13 17:12:10	♂	40-59

How old do I look?

www.how-old.net

(it lies...)

How-Old.net
How old do I look? #HowOldRobot



Sorry if we didn't quite get it right - we are still improving this feature.

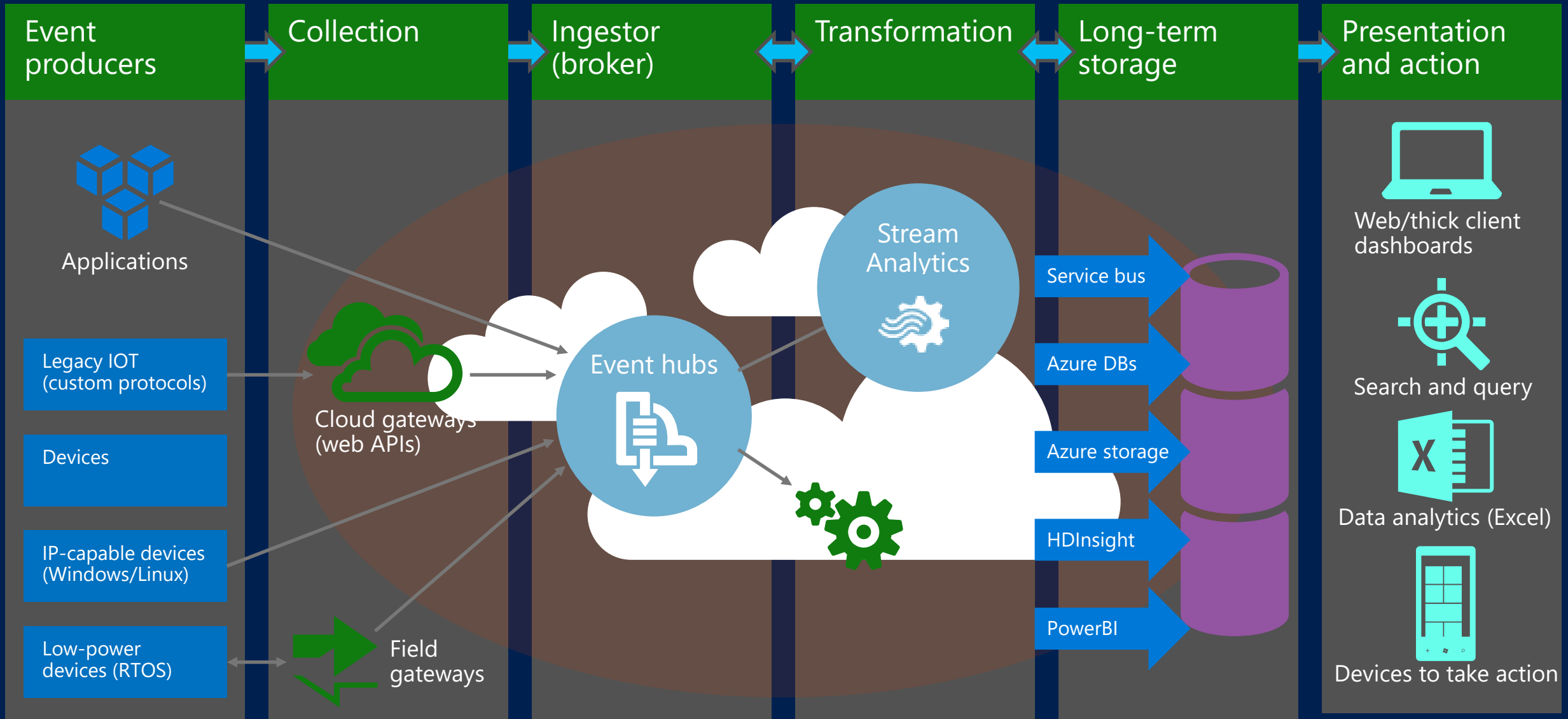
Try Another Photo!

 **Microsoft**

P.S. We don't keep the photo

 Share 2.3M  Tweet

Canonical Eventing scenario



Some data ingestion components



IoT Hub



Event Hubs



Stream Analytics



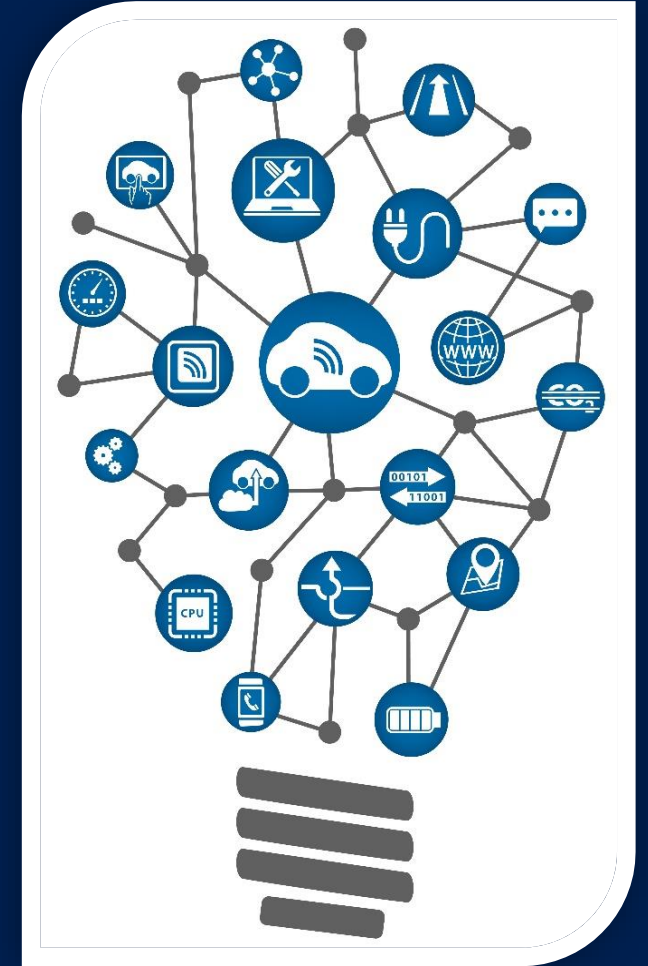
Azure Functions



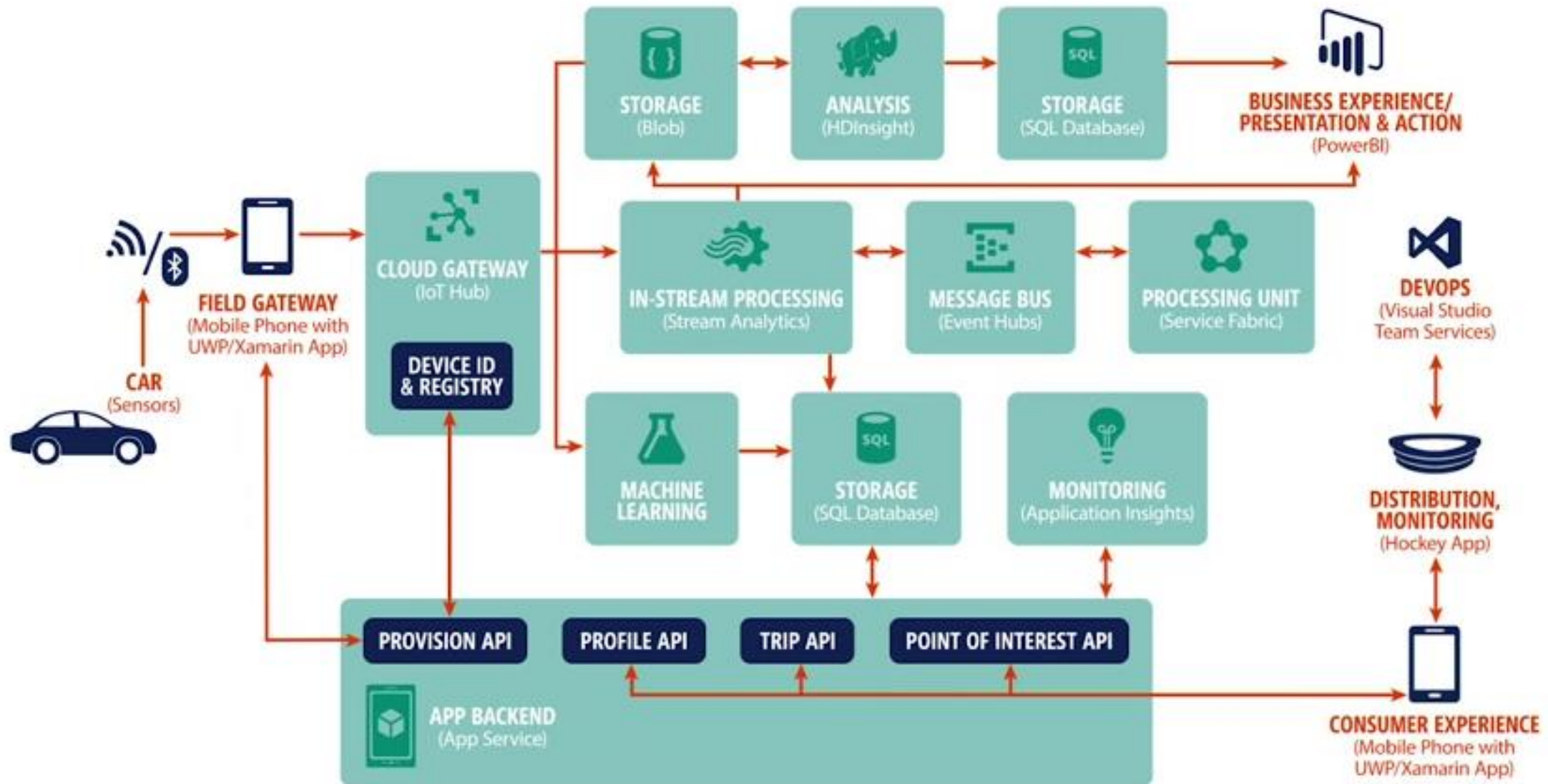
Data Factories



Machine Learning



Check out www.azure.com/MyDriving



Check out <http://azureiotsuite.com>

Microsoft Azure IoT Suite

msdn BENKOVICH

Provisioned solutions

Create a new solution
Create your own fully integrated provisioning solution

Provisioning...

mdt17iotdemo3
Connect, monitor and control industrial devices using OPC UA.

Details

Provisioning...

Ready

mdt17demo1

Ready

See your pre-configured solution running here:

[Solution dashboard](#)

[ML Workspace](#)

[Cortana Analytics Gallery: Predictive Maintenance Template](#)

[Remaining Useful Life API Help](#)

Region

West US

Subscription ID

08e3589c-64d0-49f3-a429-bfff16fb024

Modify your solution

See the provisioned Azure resources that make-up your pre-configured solution in the [Azure Management Portal](#).

View the Source Code for this Pre-Configured Solution on [GitHub](#).

Resources

[Developer documentation](#)

Predictive maintenance

live.com#msdn@benko.com

map

Simulation in progress

Engine #1Engine #2

Sensor history

Sensor 9

Engine 1Engine 2

Sensor 11

Engine 1Engine 2

Sensor 14

Engine 1Engine 2

Sensor 15

Engine 1Engine 2

Remaining Useful Life (RUL)
IN CYCLES

207
ENGINE #1

239
ENGINE #2

Cycles
#

3
ENGINE #1

3
ENGINE #2

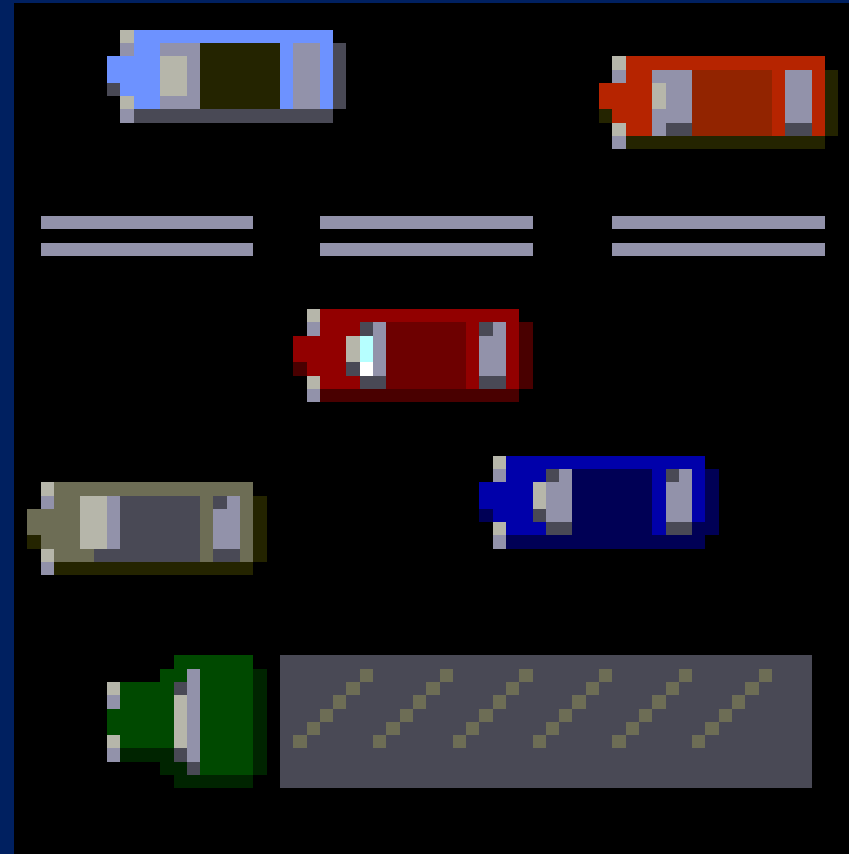
IoT Hubs & Event Hubs

What is Streaming Data?

Data at Rest



Data in Motion





Event Hubs



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. Once Event Hubs has collected the data, transform and store it by using any real-time analytics provider or with batching/storage adapters.

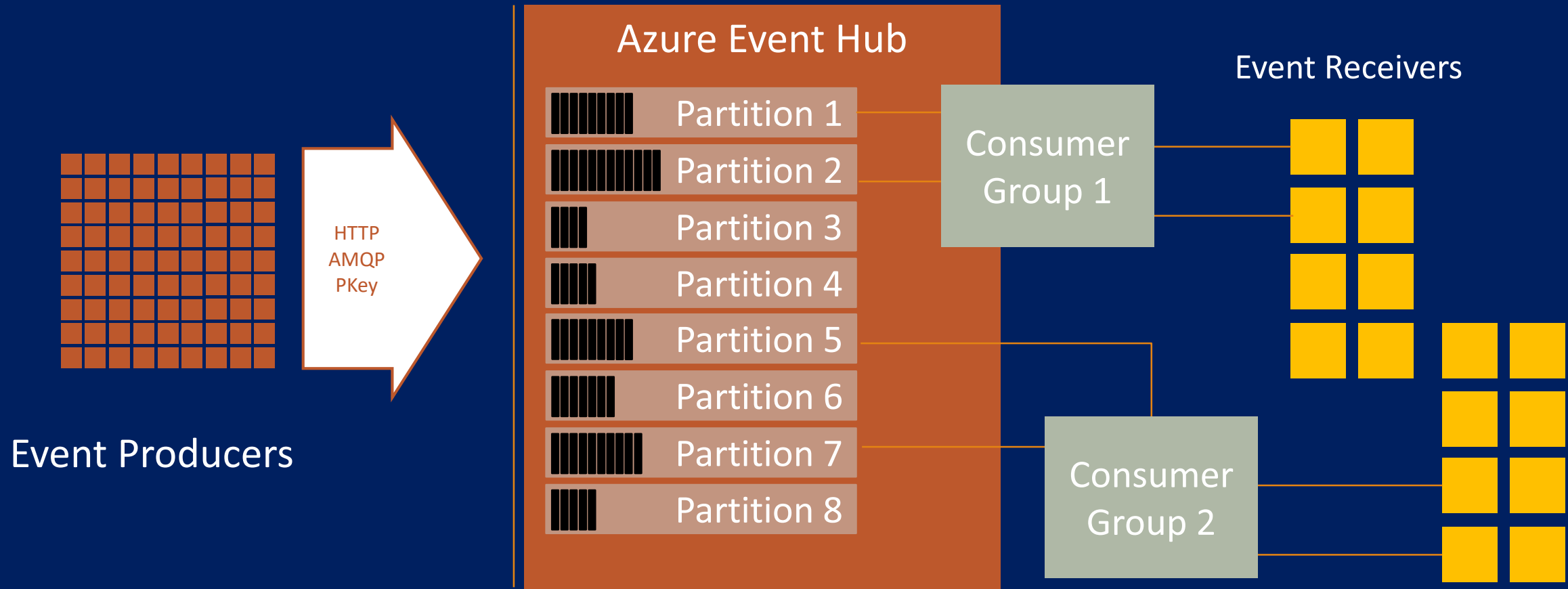
Process events with variable load profiles

Big data defines today's connected world. Big data originates from many sources with variable load profiles, such as connected cars and thermostats that produce telemetry data every few minutes, application performance counters that generate events every second, and mobile apps that capture telemetry for every user's individual action. Event Hubs is a managed service that ingests events with elastic scale to accommodate these variable load profiles and the spikes caused by intermittent connectivity.

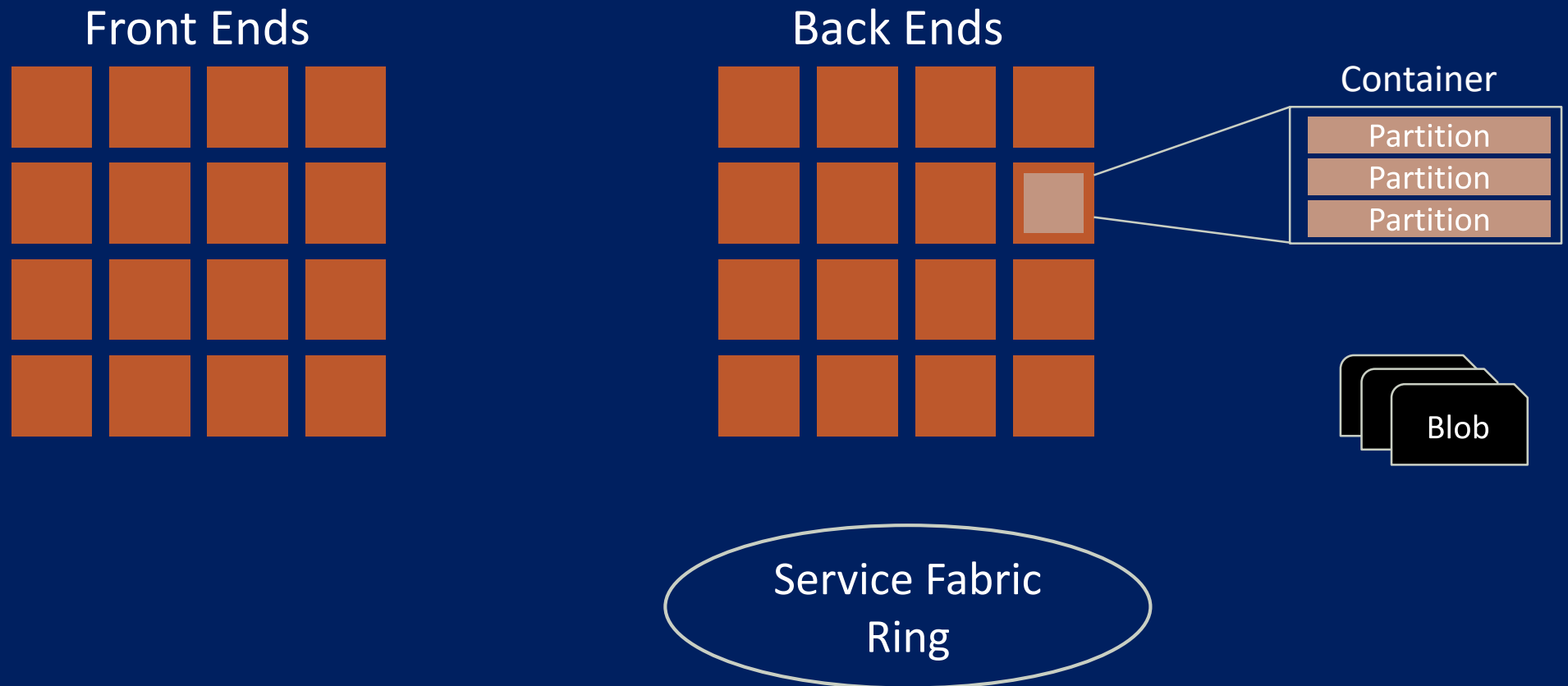
Connect millions of devices across platforms

Due to the rapid proliferation of connected devices and the variety of platforms and protocols involved, ingesting data from devices is challenging. Event Hubs meets the challenge of connecting disparate data sources while handling the scale of the aggregate stream. It lets you easily provision capacity to process events from millions of devices while preserving event order on a per-device basis. Support for Advanced Message Queuing Protocol (AMQP) and HTTP allow many platforms to work with Event Hubs. Native client libraries also exist for popular platforms.

Event Hubs Overview



Event Hubs High Level Architecture



Stream Analytics



Stream Analytics



Gain real-time insights

Azure Stream Analytics lets you rapidly develop and deploy low-cost solutions to gain real-time insights from devices, sensors, infrastructure, and applications. Use it for Internet of Things (IoT) scenarios, such as real-time remote management and monitoring or gaining insights from devices like mobile phones and connected cars.

Enable rapid development

Reduce friction and complexity and use fewer lines of code when developing analytic functions for scale-out distributed systems. Describe the desired transformation with SQL-based syntax, and Stream Analytics automatically distributes it for scale, performance, and resiliency. No need to manage complex infrastructure and software.

Perform real-time analytics

Stream Analytics is integrated out-of-the-box with [Azure Event Hubs](#) to ingest millions of events per second. Stream Analytics processes ingested events in real-time, comparing multiple streams or comparing streams with historical values and models. It detects anomalies, transforms incoming data, triggers an alert when a specific error or condition appears in the stream, and displays this real-time data in your dashboard.

Azure Stream Analytics

Highly scalable service for analyzing data in motion

Supports SQL-like query language for data analysis

Scales using Streaming Units (1 SU \approx 1 MB/sec)



Stream Analytics Query Language

SQL-like language for querying live data streams

- Subset of T-SQL
- Supports bigint, float, nvarchar(max), datetime, record, and array
- Supports SELECT, FROM, WHERE, GROUP BY, and other common Data Manipulation Language (DML) statements
- Supports COUNT, AVG, DATEDIFF, and other common functions

Adds extensions such as `TIMESTAMP BY` and `System.Timestamp`

Supports temporal grouping of events via "windowing"

Querying a Data Stream

List all Connecticut cars that enter a toll booth, and include the entry time, toll booth ID, and license-plate number

```
SELECT EntryTime, TollId, LicensePlate
FROM EntryData
WHERE State = 'CT'
```

ENTRYTIME	TOLLID	LICENSEPLATE
2014-09-10T12:02:00+00:00	3	ABC 1004
2014-09-10T12:03:00+00:00	2	XYZ 1003
2014-09-10T12:11:00+00:00	1	NJB 1006

Designating a Field as the Event Time

Designate the EntryTime field as the event time for calculations that involve event time

```
SELECT System.Timestamp AS [Entry Time],  
       TollId, LicensePlate  
FROM EntryData TIMESTAMP BY EntryTime  
WHERE State = 'CT'
```

ENTRYTIME	TOLLID	LICENSEPLATE
2014-09-10T12:02:00+00:00	3	ABC 1004
2014-09-10T12:03:00+00:00	2	XYZ 1003
2014-09-10T12:11:00+00:00	1	NJB 1006

JOINing Two Data Streams

How long does it take each car that enters a toll booth to pay the toll and exit the booth?

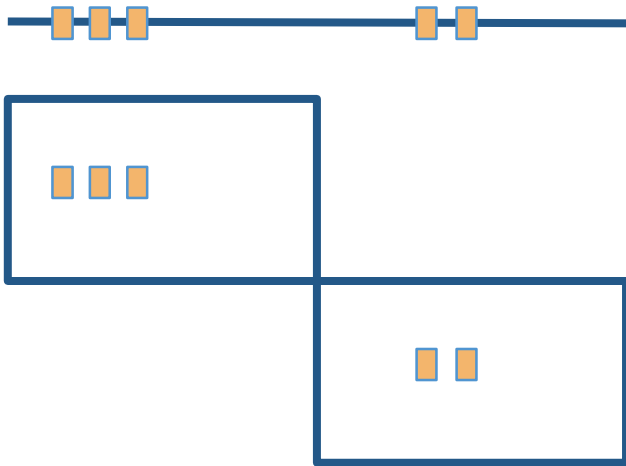
```
SELECT EN.TollId, EN.EntryTime, EN.LicensePlate,  
       DATEDIFF(minute, EN.EntryTime, EX.ExitTime) AS Minutes  
FROM EntryData EN TIMESTAMP BY EntryTime  
JOIN ExitData EX TIMESTAMP BY ExitTime  
  ON EN.TollId = EX.TollId  
  AND EN.LicensePlate = EX.LicensePlate  
  AND DATEDIFF(minute, EN, EX) BETWEEN 0 AND 60
```

TOLLID	ENTRYTIME	LICENSEPLATE	MINUTES
1	2014-09-10T12:01:00.000Z	JNB 7001	2
1	2014-09-10T12:02:00.000Z	YXZ 1001	1
3	2014-09-10T12:02:00.000Z	ABC 1004	2

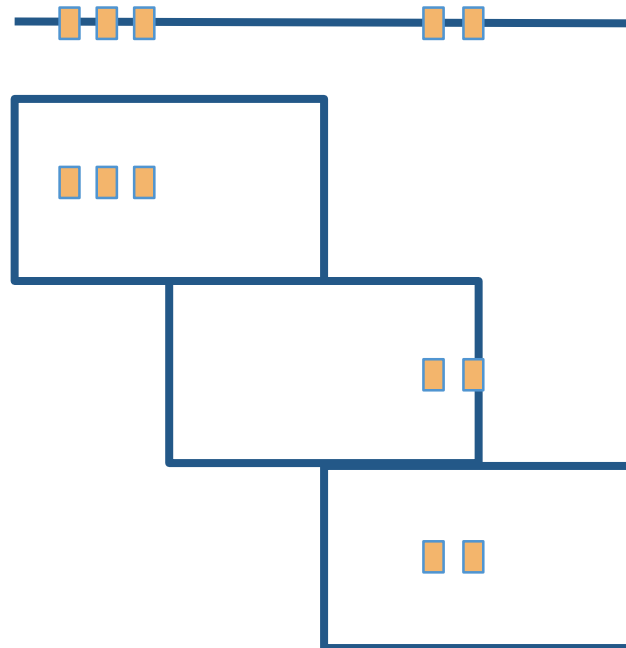
Windowing

Count or aggregate events over a specified time period

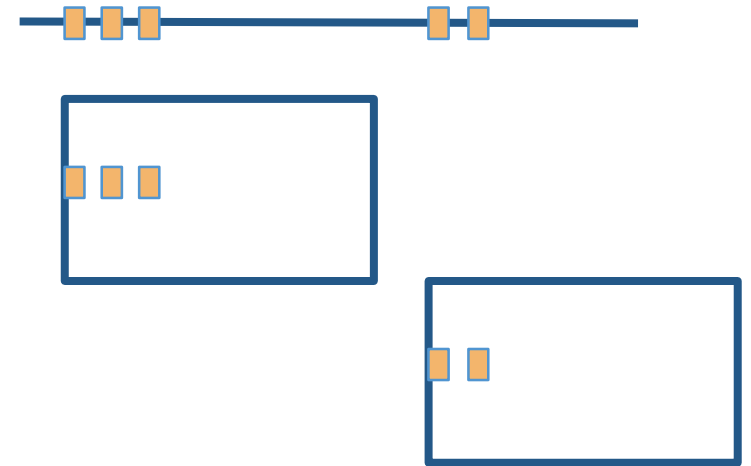
TumblingWindow



HoppingWindow



SlidingWindow



Using TumblingWindow

How many New York cars enter a toll booth every 5 minutes?

```
SELECT DateAdd(minute, -5, System.TimeStamp)
       AS [Start Time], System.TimeStamp AS [End Time],
       COUNT(*)
FROM EntryData TIMESTAMP BY EntryTime
WHERE State = 'NY'
GROUP BY TumblingWindow(minute, 5)
```

START TIME	END TIME	COUNT
2014-09-10T12:00:00.000Z	2014-09-10T12:05:00.000Z	3
2014-09-10T12:05:00.000Z	2014-09-10T12:10:00.000Z	6
2014-09-10T12:15:00.000Z	2014-09-10T12:20:00.000Z	2

Using HoppingWindow

What is the average wait time at all toll booths for the last 5 minutes, updated every 1 minute?

```
SELECT DateAdd(minute, -5, System.TimeStamp)
      AS [Start Time], System.TimeStamp AS [End Time],
      AVG(DATEDIFF(minute, EN.EntryTime, EX.ExitTime))
      AS [Average Wait Time]
FROM EntryData EN TIMESTAMP BY EntryTime
JOIN ExitData EX TIMESTAMP BY ExitTime
  ON EN.TollId = EX.TollId
  AND EN.LicensePlate = EX.LicensePlate
  AND DATEDIFF(minute, EN, EX) BETWEEN 0 AND 60
GROUP BY HoppingWindow(minute, 5, 1)
```

START TIME	END TIME	AVERAGE WAIT TIME
2014-09-10T11:58:00.000Z	2014-09-10T12:03:00.000Z	1.5
2014-09-10T11:59:00.000Z	2014-09-10T12:04:00.000Z	1.6666666666666667
2014-09-10T12:00:00.000Z	2014-09-10T12:05:00.000Z	1.6666666666666667

Using SlidingWindow

In which 5-minute windows does at least one Connecticut car enter a toll booth?

```
SELECT DateAdd(minute, -5, System.TimeStamp)
       AS [Start Time], System.TimeStamp AS [End Time],
       TollId, COUNT(*)
FROM EntryData TIMESTAMP BY EntryTime
WHERE State = 'CT'
GROUP BY TollId, slidingwindow(minute, 5)
HAVING COUNT(*) > 0
```

START TIME	END TIME	TOLLID	COUNT
2014-09-10T11:57:00.000Z	2014-09-10T12:02:00.000Z	3	1
2014-09-10T11:58:00.000Z	2014-09-10T12:03:00.000Z	2	1
2014-09-10T12:06:00.000Z	2014-09-10T12:11:00.000Z	1	1

Building Real-Time Dashboards

Direct Stream Analytics output to an Azure event hub

Write code that subscribes to events from the event hub

Application name			Home	About	Contact	Register	Log in
<h2>ATM Dashboard</h2> <p>The table below lists potentially fraudulent ATM transactions and is updated every 5 seconds.</p>							
Card Number		ATM 1	ATM 2				
995172721		14957	69434				
592830441		79817	23843				
676470808		15223	74522				
906825608		84048	62218				
282119302		43681	47663				
801485511		26945	66819				

Azure Functions

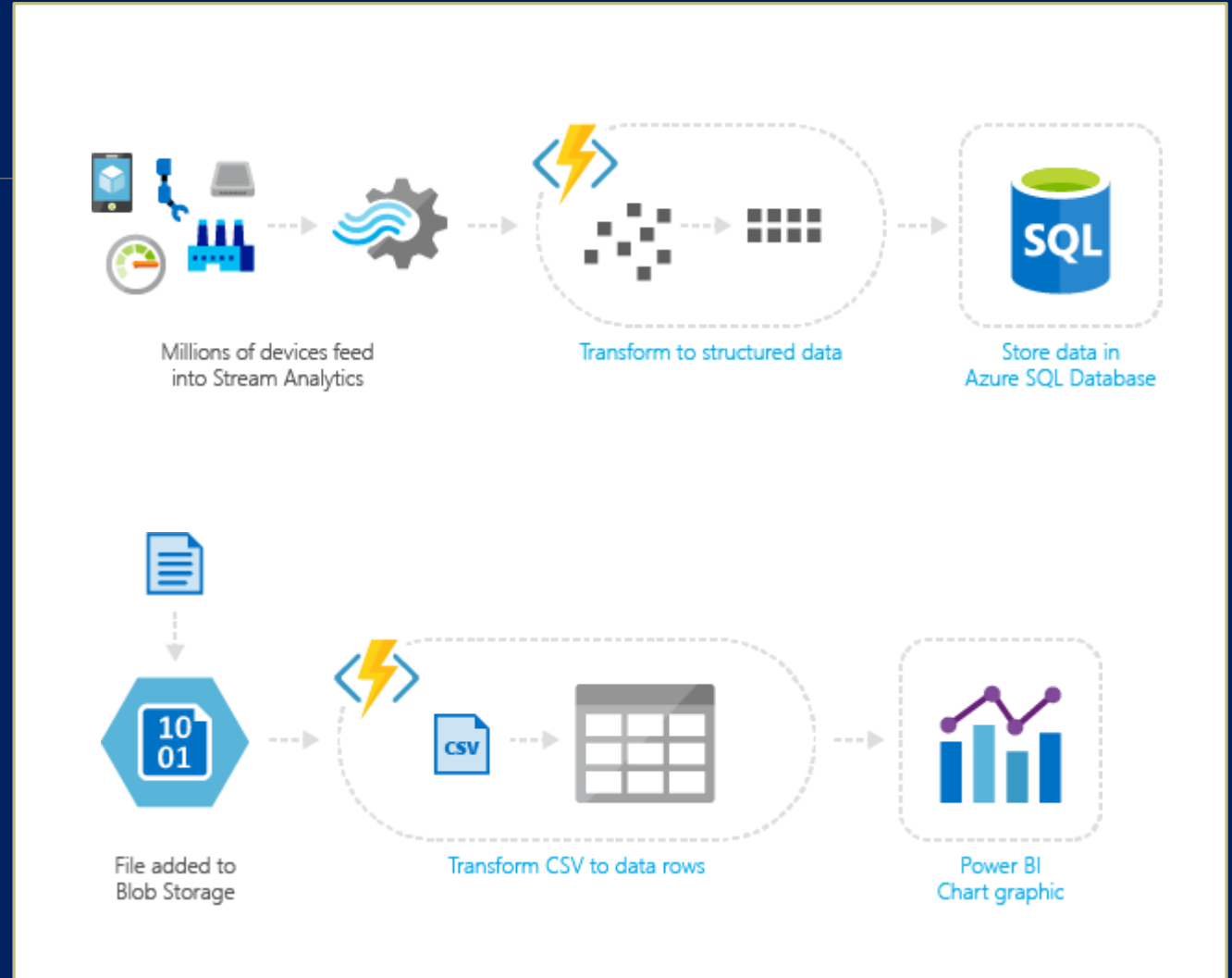


Azure Functions

Real-time stream processing

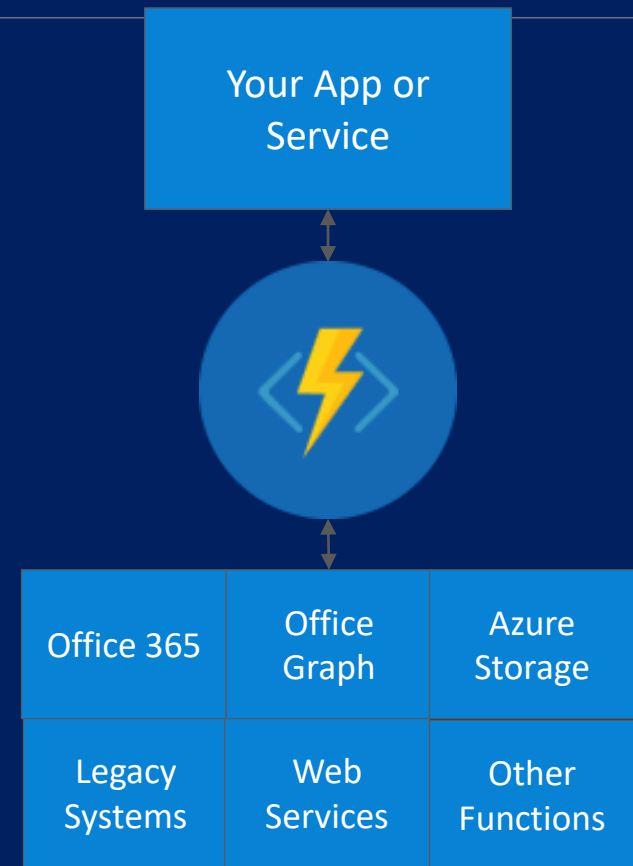
For example, IoT devices send messages to Azure Stream Analytics, which then calls an Azure Function to transform the message. This function processes the data and creates a new record in an Azure SQL Database.

Azure Functions supports triggering an event based on an activity in an Azure service. For example, you could execute code that reads newly discovered test log files in an Azure Blob Storage container and transforms this into a row in an Azure SQL Database table



Common Scenarios

Timer-based processing
Azure service event processing
SaaS event processing
Serverless web application architectures
Serverless mobile backends
Real-time stream processing
Real-time bot messaging



Function App Templates

Function App templates are categorized into general areas of Timer, Data Processing, and Webhook & API

Choose a template

Language: Scenario:

BlobTrigger - C# A C# function that will be run whenever a blob is added to a specified container	BlobTrigger - Node A Node.js function that will be run whenever a blob is added to a specified container	Empty - C# An empty C# function without triggers, inputs, or outputs	Empty - Node An empty Node.js function without triggers, inputs, or outputs
EventHubTrigger - Node A Node.js function that will be run whenever an event hub receives a new event	Generic Webhook - C# A C# function that will be run whenever it receives a webhook request	Generic Webhook - Node A Node.js function that will be run whenever it receives a webhook request	GitHub WebHook - C# A C# function that will be run whenever it receives a GitHub webhook request

- BlobTrigger
- EventHubTrigger
- Generic webhook
- GitHub webhook
- HTTPTrigger
- QueueTrigger
- ServiceBusQueueTrigger
- ServiceBusTopicTrigger
- TimerTrigger
- Blank & Experimental

Anatomy of a Function

A “Run” file that containing the function code

A “Function” file containing all service and trigger bindings and parameters

A “Project” file containing project assembly and NuGet package references

App Service settings, such as connection strings and API keys



Executable code

Function configuration

.NET Core and Project
references

Function Bindings

Bindings serve as the basis for all connections to and from a function. Many bindings can be “bi-directional” as well.

Type	Service	Trigger	Input	Output
Schedule	Azure Functions	✓		
HTTP (REST or webhook)	Azure Functions	✓		✓*
Blob Storage	Azure Storage	✓	✓	✓
Events	Azure Event Hubs	✓		✓
Queues	Azure Storage	✓		✓
Tables	Azure Storage		✓	✓
Tables	Azure Mobile Apps		✓	✓
No-SQL DB	Azure DocumentDB		✓	✓
Push Notifications	Azure Notification Hubs			✓

Machine Learning

What is Machine Learning?

Branch of computer science in which a computer "learns" from data in order to perform predictive analytics

- Credit-card fraud detection
- Online shopping recommendations
- Self-driving cars and more

Supervised learning

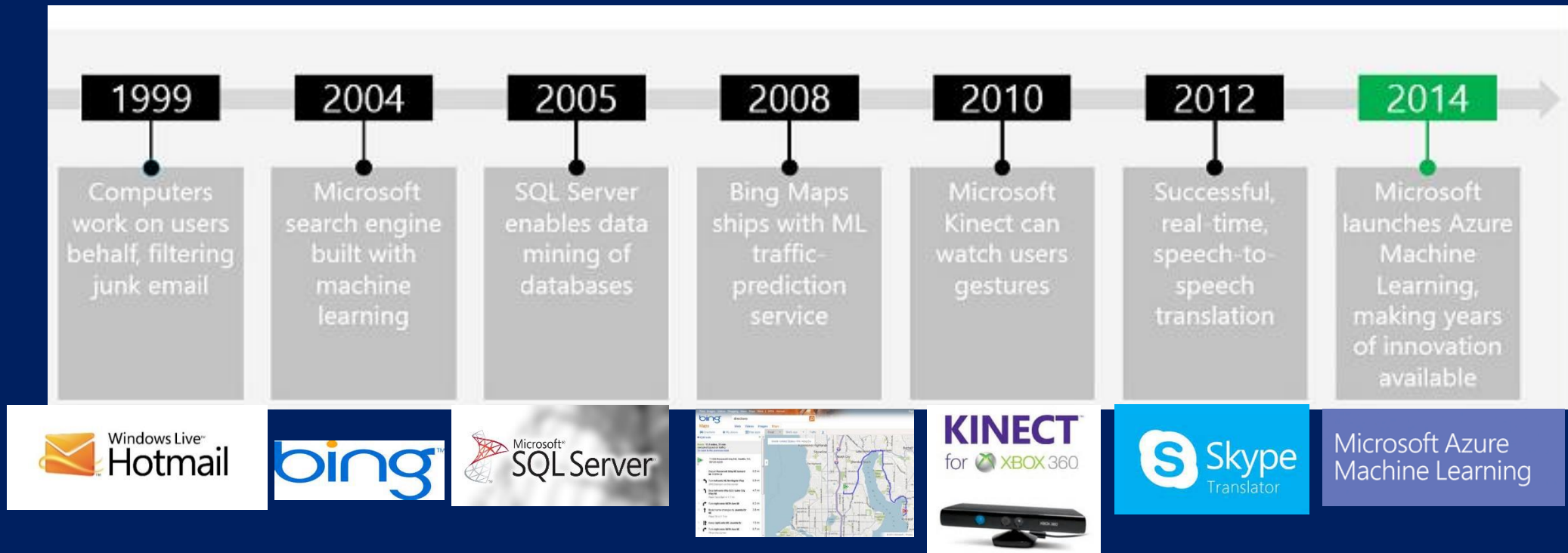
- Regression and classification

Unsupervised learning

- Clustering



Microsoft and Machine Learning



Modified from <http://pulsweb.fr/predict-wine-quality-azureml>

Azure Machine Learning

Fully managed cloud service for building and operationalizing ML models



Fully managed

No software to install, no hardware to manage, and one portal to view and update.

Integrated

Simple drag, drop and connect interface for Data Science. No need for programming for common tasks.

Best in Class Algorithms + R

Built-in collection of best of breed algorithms. Support for R and popular CRAN packages.

Deploy in minutes

Operationalize models with a single click. Monetize in Machine Learning Marketplace.

“

*I spent last semester
building a regression model
in Python, and I just did the
same thing in 10 minutes
with Azure ML*

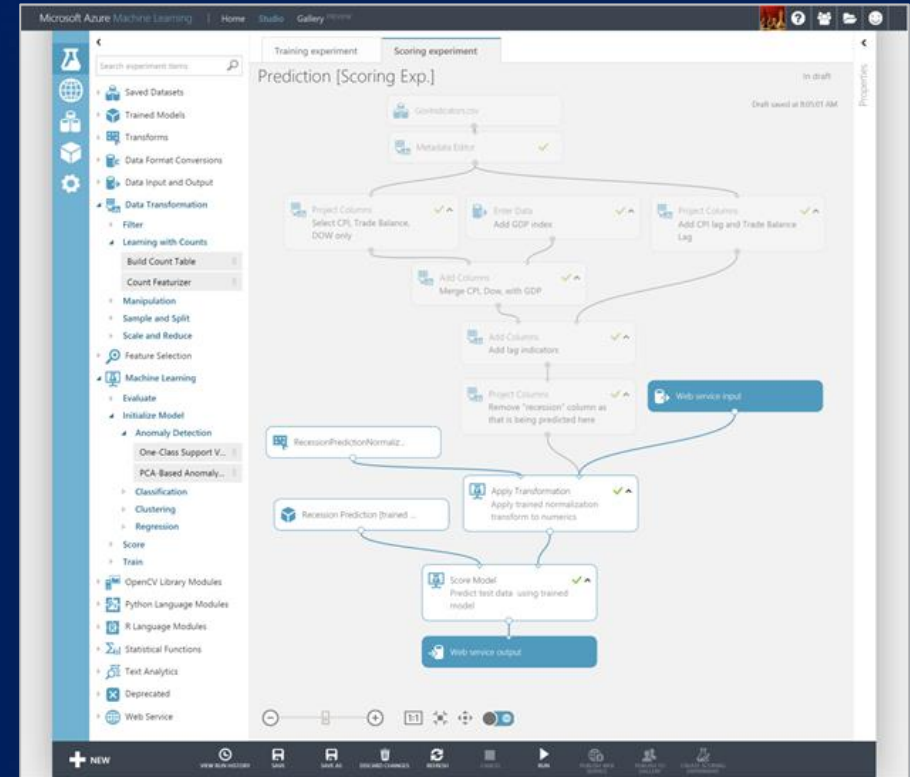
”

Azure Machine Learning Studio

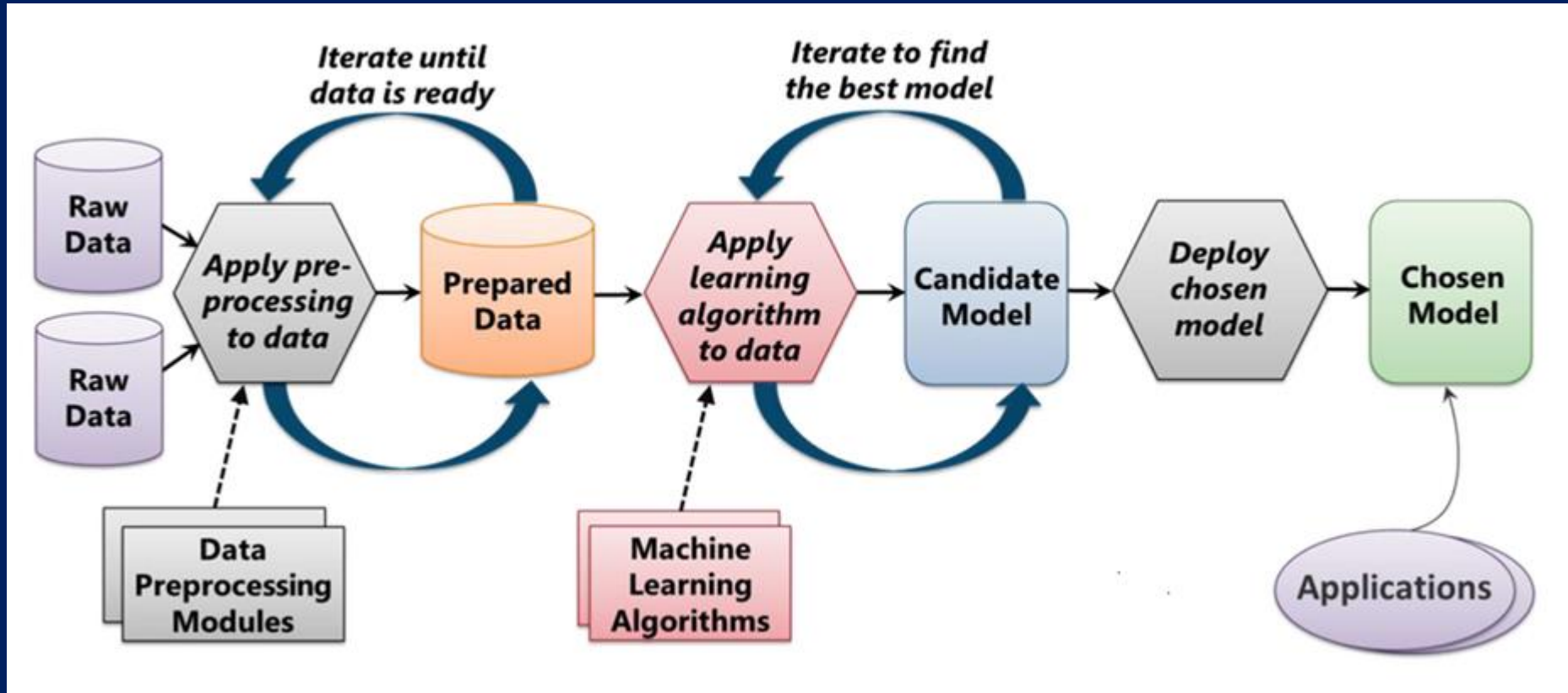
Visual editor for composing, testing, refining, and deploying machine-learning models

- Includes hundreds of modules
- Includes common algorithms for classification, regression, and more
- Supports numerous input formats
- Supports R and Python

Machine learning for the masses



The Machine Learning Process



From "Introduction to Microsoft Azure" by David Chappell

Summary

Where can I get more info?

Visit my site <http://www.benkotips.com>

- Resources from today's talk
- Webcasts
- Downloads
- More!

Get a FREE Cloud Readiness Evaluation
email mike@benko.com for more info

Follow me on Twitter!
[@mbenko](https://twitter.com/mbenko)