

1110
1110
1010 1010
1010 1010 1010

PREDICTIVE ANALYTICS FOR IOT

Event Hubs + Streaming Analytics + Storage + SQL + Machine
Learning + Data Factory => AWESOME

Instructor: Ryan Swanstrom

Social Media: @ryanswanstrom



Ryan Swanstrom

<http://aka.ms/iotedu>



PRE-REQUISITES

Hopefully, you got a chance to do these already, but if not please take some time now to begin this process. Please place your **RED** sticky note on your computer if you have a question (then **YELLOW** when done).

- Power BI online account
- Azure account – please check that this works

And

- Create a Data Science Virtual Machine

Or Install:

- Visual Studio 2015 Desktop (Community ok) with SQL Server Data Tools



Ryan Swanstrom

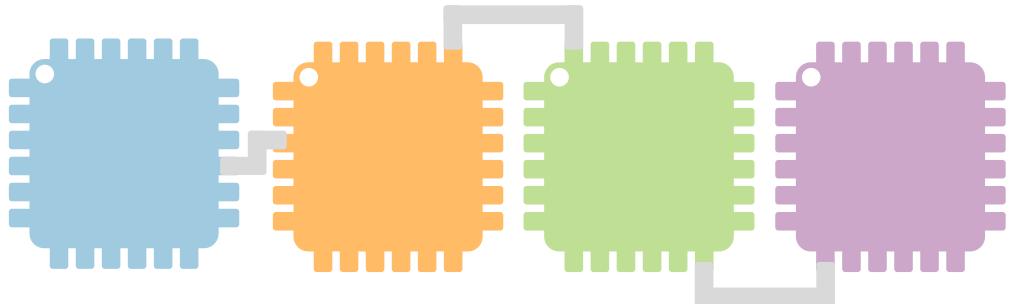
<http://aka.ms/iotedu>



Pre-requisites site with more info: <https://aka.ms/iotedu-prereqs>

Visual Studio takes a long time to download and install.

**GOOD MORNING AND WELCOME TO
ALL!**



Ryan Swanstrom

<http://aka.ms/iotedu>



QUOTE - UNLOCKING VALUE

"We knew there was inherent value in the data," said Christopher Dell, senior director of product development and management at Intelligent Mechatronic Systems (IMS). "We just didn't know how to unlock that value."



Ryan Swanstrom

<http://aka.ms/iotedu>



The story is about IoT-related struggles of a Canadian company called Intelligent Mechatronic Systems – a vehicle telemetry story (see “story.txt” under Misc folder on class site)

In here they use Pentaho analytics BI suite. They do not use Azure, but note that we have integrated it into Azure (specifically, Active Directory – if you belong to one you should have free or affordable access to this offering)

Note, this suite integration is available on the very useful Azure associated integration site called the Azure Marketplace – a software integration and dev site where companies can submit integrations with other software, a valuable “market” to know about - <https://azure.microsoft.com/en-us/marketplace/>. You can even find many APIs and different types of virtual machines here already configured. There are over 4000 products.

However, in general solutions integrating Azure products with Machine Learning are very often surfaced on the Cortana Intelligence Gallery at:
<https://gallery.cortanaintelligence.com/> or the Cortana Intelligence Quick Starts at:
<https://start.cortanaanalytics.com/> (from where we are building out a pattern in

private preview status today)

WHAT YOU WILL BE ABLE TO DO AFTER THIS TRAINING

Be able to piece apart the components of a deployed solution for predictive analytics in Azure for streaming data, including gaining familiarity with the **data science** aspects of the solution.

Be able to build back up and **deploy a simple predictive analytics solution** in Azure for streaming data including a shareable BI dashboard for **rapid synthesis** of information.



Ryan Swanstrom

<http://aka.ms/iotedu>



The BI Dashboard from a Power BI-integrated solution template will be shown now.

PREDICTIVE ANALYTICS FOR IoT

Event Hubs + Streaming Analytics + Storage + SQL + Machine Learning + Data Factory = AWESOME

PART 1 OF 5: LOGISTICS



Ryan Swanstrom

<http://aka.ms/iotedu>



USING THE CLASS REPOSITORY

- The class site and resources can be found at

<https://aka.ms/iotedu>



Ryan Swanstrom

<http://aka.ms/iotedu>



GitHub is a free code hosting web service for developers and the trainers use it because it's easy to create documents and upload files. It's also free.

Join GitHub for free: <https://github.com/join>

Benefits:

- This will allow you to be able to programmatically “clone” repositories to your local machine or create your own to share and collaborate
- You will also be able to join the discussion forum for this class (<https://aka.ms/iotedu-discussion>)
- It can be useful even for just sharing text files
- Note, github uses Markdown as a text markup language, but you don't have to know it to create text files
- And as an fyi, GitHub has GitHub Pages – a free place to create and host a website (I use for my blog)
- Many more...

More details on signing up: <https://help.github.com/articles/signing-up-for-a-new->

github-account/

PREDICTIVE ANALYTICS FOR IoT

Event Hubs + Streaming Analytics + Storage + SQL + Machine Learning + Data Factory = AWESOME

PART 2 OF 5: BUSINESS UNDERSTANDING



Ryan Swanstrom

<http://aka.ms/iotedu>



USE CASE – ENERGY DEMAND

Many utility and smart energy companies have been pioneering the [smart grid](#) by deploying a number of use cases that make use of the data generated by the grid. Many use cases revolve around the inherent characteristics of electricity production: it cannot be accumulated nor stored aside as inventory. So, what is produced must be used. Utilities that want to become more efficient need to forecast power consumption simply because that will give them greater ability to **balance supply and demand**, thus preventing energy wastage, **reduce greenhouse gas emission**, and control cost.

When talking of costs, there is another important aspect, which is price. New abilities to trade power between utilities have brought in a great need to **forecast future demand and future price of electricity**. This can help companies determine their production volumes.

A specific global energy company approaches your team and wants you to help them. Can your team help them understand the past energy usage. Also, can your team forecast future energy demand? It would even be great to know how your forecast matches up with demand. Ideally a dashboard would be great.



Ryan Swanstrom

<http://aka.ms/iotedu>



FORM GROUPS AND ANSWER

- What Questions Do You Need To Ask To Work On This Project?
- Who Do You Need To Talk With?
- What Other Industries Have Similar Problems?



Ryan Swanstrom

<http://aka.ms/iotedu>



Meet Your Neighbors and work through the questions.

WHAT ARE CUSTOMERS WANTING TO

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



Ryan Swanstrom

<http://aka.ms/iotedu>

 Microsoft

Customers can be also the businesses we enable
CRM is customer relationship management

AGL'S CUSTOMER STORY - AN ENERGY COMPANY POWERS INNOVATION WITH AZURE

"Yes, the industry was more competitive than ever before [deregulation of energy sector] but, to retain our place as the leading energy retailer, we had to find a way to differentiate ourselves beyond pricing by enhancing the value we deliver. For us – and our customers – this continues to be a rewarding and exciting journey, and one that has been enhanced with the rollout of smart meters."



Ryan Swanstrom

Microsoft Cloud - Mobility - Productivity -
Enterprise Solutions by Industry - Public Sector - Customer Stories - Events - Blogs -

Innovation Spotlight: AGL puts energy into action with the Cloud

Discover the difference the cloud is making to leading energy provider AGL Energy Limited (AGL) and some of the leading edge activities it's doing to drive innovation.

AGL puts energy into action with the Cloud



Customer Story
AGL
By Microsoft Customer Stories on March 16, 2016
Filed under Power & Utilities

Share: [f](#) [t](#) [e](#) [in](#) [p](#) [y](#) [l](#)

<http://aka.ms/iotedu>



Story here: <https://enterprise.microsoft.com/en-us/industries/power-and-utilities/agl/>

Company website: <https://www.agl.com.au>

About AGL: Australian Gas Light Company (AGL) is one of Australia's leading integrated energy companies and largest ASX listed owner, operator and developer of renewable energy generation in the country. They have 3.7 million customers across Queensland, New South Wales, Victoria and South Australia.

<https://www.agl.com.au>

"Empowering consumers to better understand their energy consumption, and subsequently lower costs and increase energy efficiency, smart meters are changing our industry, and allowing us to deliver data-rich and consider personalised experiences...."

More energy and utility customer stories:<https://enterprise.microsoft.com/en-us/industries/power-and-utilities/>

Smart grid



OUR BUSINESS GOAL

- Formulate predictive analytics in the context of the utility and energy domain
- Generalizing and modifying an end-to-end solution template to leverage for future predictive analytics solutions

We will leverage the Energy Demand Forecasting Solution as our learning tool.



Ryan Swanstrom

<http://aka.ms/iotedu>



Energy Demand Forecasting -

<https://gallery.cortanaintelligence.com/Solution/Energy-Demand-Forecasting-4>

MORE GENERAL BUSINESS GOALS

- Demonstrate a typical predictive analytics and machine learning solution
- Short deployment time or time-to-value of an ML-based solution
- Expand to other use cases or a broader scope based on business need
- Quickly gain Cortana Intelligence Suite (CIS) practical knowledge



Ryan Swanstrom

<http://aka.ms/iotedu>



WHY ENERGY FORECASTING?

It's a great example of forecasting demand and gets us these results:

Cost reduction: Utility and energy companies that want to become more efficient need to forecast power consumption simply because that will give them greater ability to **balance supply and demand**, thus preventing energy wastage, **reduce greenhouse gas emission**, and control cost.

Cost forecasts: New abilities to trade power between utilities have brought in a great need to **forecast future demand and future price of electricity**.

Overload avoidance: By remotely regulating consumption (with the help of smart meters – our IoT devices discussed here), localized **overload situations can be handled**.

IoT + live resources + big data = big opportunities



Ryan Swanstrom

<http://aka.ms/iotedu>



Solution Template for Energy Demand Forecasting -

<https://blogs.technet.microsoft.com/machinelearning/2016/03/22/solution-template-for-energy-demand-forecasting/>

Energy Demand Forecasting Template Using SQL Server R Services -

<https://blogs.technet.microsoft.com/machinelearning/2016/05/03/energy-demand-forecasting-template-using-sql-server-r-services/>

OTHER SECTORS TOUCHED BY DEMAND FORECASTING

- Health
- Finance
- Retail/Commerce
- Supply chain
- Production



Ryan Swanstrom

<http://aka.ms/iotedu>



We will be thinking about the forecast horizon in a bit, but for now know that there are different time-range based forecasts that can be done. The forecast horizon is the time-range in the future in which the forecast will fall.

PREDICTIVE ANALYTICS FOR IoT

Event Hubs + Streaming Analytics + Storage + SQL + Machine Learning + Data Factory = AWESOME

PART 3 OF 5: ARCHITECTURE AND NAMING THINGS



Ryan Swanstrom

<http://aka.ms/iotedu>



And in this session we will deploy the cloud solution to Azure so make sure you've satisfied your pre-requisites from the class site or registration information.

CORTANA INTELLIGENCE IN A SENTENCE:

Cortana Intelligence is a **Platform** and a **Process** to perform advanced analytics from start to finish



Ryan Swanstrom

<http://aka.ms/iotedu>

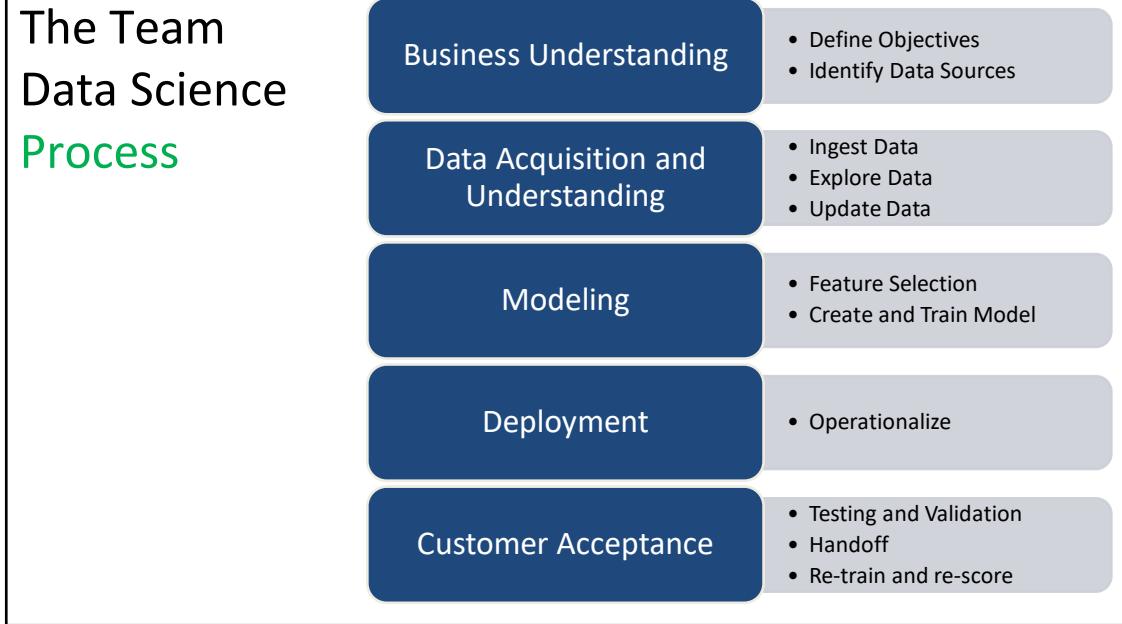


1. What you can do with CIS:

<https://www.microsoft.com/en-us/server-cloud/cortana-intelligence-suite/why-cortana-intelligence.aspx>

2. More about the process:

<https://channel9.msdn.com/Blogs/Seth-Juarez/Understanding-Data-Science-for-building-Predictive-Analytics-Solutions-by-Francesca-Lazzeri>



1. This process largely follows the CRISP-DM model:
<http://www.sv-europe.com/crisp-dm-methodology/>
2. It also references the Cortana Intelligence process:
<https://azure.microsoft.com/en-us/documentation/articles/data-science-process-overview/>
3. A complete process diagram is here:
<https://azure.microsoft.com/en-us/documentation/learning-paths/cortana-analytics-process/>
4. Some walkthrough's of the various services:
<https://azure.microsoft.com/en->

[us/documentation/articles/data-science-process-walkthroughs/](#)

5. An integrated process and toolset allows for a more close-to-intent deployment
6. Iterations are required to close in on the solution – but are harder to management and monitor

The Cortana Intelligence Platform

 Hola, Soy Cortana	Cortana, Cognitive Services, Bot Framework
 Power BI	Power BI
 Stream Analytics	Stream Analytics
 HDInsight	HDInsight
 Azure Machine Learning (MRS)	Azure Machine Learning (MRS)
 SQL Data Warehouse (SQL DB, Document DB)	SQL Data Warehouse (SQL DB, Document DB)
 Data Lake	Data Lake
 Event Hubs	Event Hubs
 Data Factory	Data Factory
 Data Catalog	Data Catalog
 Microsoft Azure	Microsoft Azure

1. Platform and Storage: Microsoft Azure – <http://microsoftazure.com> Storage: <https://azure.microsoft.com/en-us/documentation/services/storage/> (**Host It**)
2. Azure Data Catalog: <http://azure.microsoft.com/en-us/services/data-catalog> (**Doc It**)
3. Azure Data Factory: <http://azure.microsoft.com/en-us/services/data-factory/> (**Move It**)
4. Azure Event Hubs: <http://azure.microsoft.com/en-us/services/event-hubs/> (**Bring It**)
5. Azure Data Lake: <http://azure.microsoft.com/en-us/campaigns/data-lake/> (**Store It**)
6. Azure DocumentDB: <https://azure.microsoft.com/en-us/services/documentdb/>, Azure SQL Data Warehouse: <http://azure.microsoft.com/en-us/services/sql-data-warehouse/> (**Relate It**)
7. Azure Machine Learning: <http://azure.microsoft.com/en-us/services/machine-learning/> (**Learn It**)
8. Azure HDInsight: <http://azure.microsoft.com/en-us/services/hdinsight/> (**Scale It**)
9. Azure Stream Analytics: <http://azure.microsoft.com/en-us/services/stream-analytics/> (**Stream It**)
10. Power BI: <https://powerbi.microsoft.com/> (**See It**)
11. Cortana: <http://blogs.windows.com/buildingapps/2014/09/23/cortana-integration-and-speech-recognition-new-code-samples/> and <https://blogs.windows.com/buildingapps/2015/08/25/using-cortana-to-interact-with-your-customers-10-by-10/> and <https://developer.microsoft.com/en-us/Cortana> (**Say It**)
12. Cognitive Services: <https://www.microsoft.com/cognitive-services>
13. Bot Framework: <https://dev.botframework.com/>
14. All of the components within the suite: <https://www.microsoft.com/en-us/server-cloud/cortana-intelligence-suite/what-is-cortana-intelligence.aspx>
15. What can I do with it? <https://gallery.cortanaintelligence.com/>
16. Getting Started Quickly: <https://caqs.azure.net/#gallery>

COMING UP WITH A GOOD <name>

It's recommended that you use a **consistent prefix** when you name your objects in Azure

Many names need to be globally unique. Plus, a good prefix makes it easier to identify them, determine their role, and delete them later when you need to clean up.

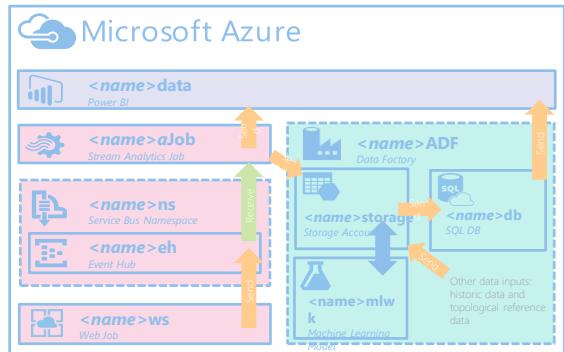
The prefix is identified by the "<name>" placeholders in the architecture diagram.

I recommend using something informative and short, containing a marker of who you are such as initials or alias. Such as:

"*ryanenergy*"



Ryan Swanstrom



<http://aka.ms/iotedu>



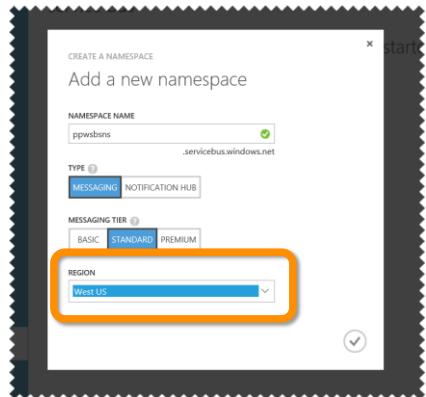
Note: we are using a solution template that when deployed provisions all of these components for us and even starts the ingress of data.

This name will prefix your resources and will also be the name of your resource group (making it easy to tear down when done)

USING THE SAME REGION

As you provision your resources in Azure you should use the same region for all of them.

This will decrease latency between them as well as reduce costs by eliminating data leaving one region as it travels to another.

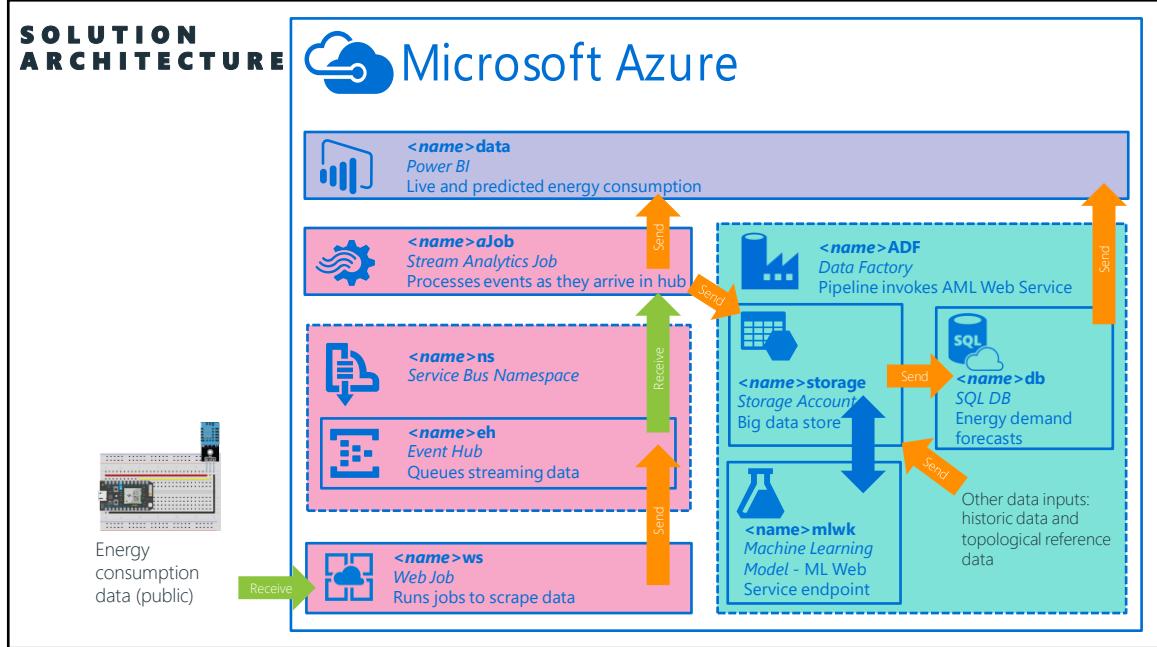


Ryan Swanstrom

<http://aka.ms/iotedu>



This is done automatically for us when we choose a region for the solution template, but when you create a custom solution this is something to consider.



We'll spend a moment and get an overview of the architecture of the solution template.

Pink background: Data in motion or fast-moving data

Green/blue background: Data at rest or slow-moving data

Data Source and Ingestion

Azure Web Jobs - scrapes data from a streaming data source and shuttles it to the Event Hub

Azure Event Hub - receives the raw data from the Web Job

Data Preparation and Analysis

Azure Stream Analytics - provides near real-time analytics and publishes results to Power BI dashboard, as well as, shuttles raw data to Azure Storage for archiving

Azure Storage - stores the archived, raw streaming data for future processing

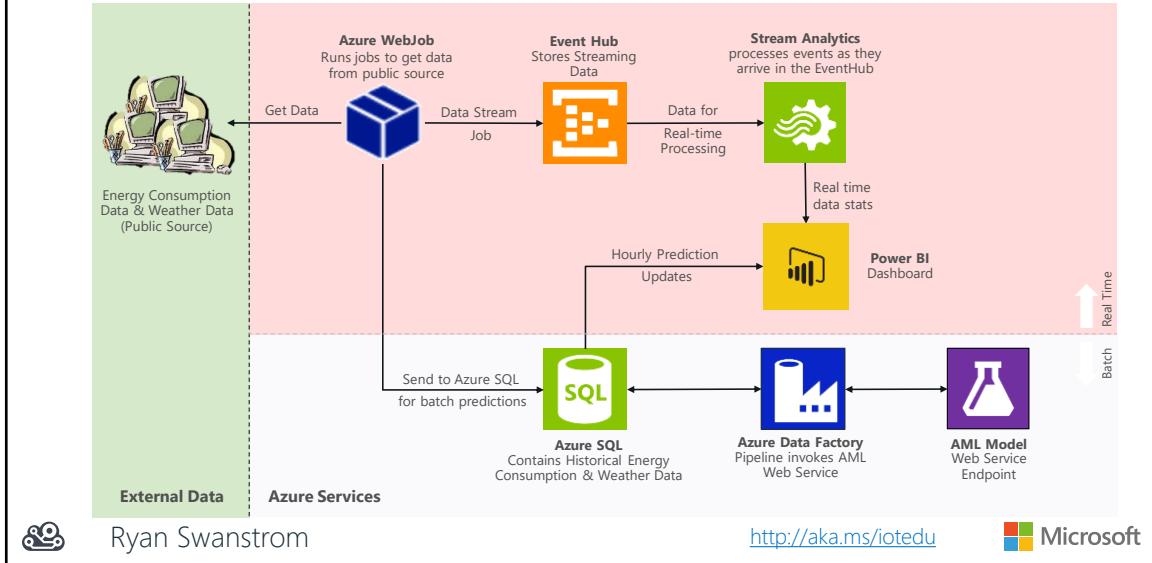
Azure Data Factory - orchestrates data flow, running scripts, calling out to the Azure Machine Learning service, and managing the Azure SQL Database service

Azure Machine Learning - returns predictions (here, future power consumption forecasts) based on inputs received using a trained model

Data Publishing and Consumption

Azure SQL Database - stores the results of the Azure Machine Learning service
Power BI - dashboarding service used to show aggregations provided by Azure Stream Analytics (data in motion) and Azure Machine Learning service results stored in Azure SQL Database (data at rest)

ARCHITECTURE DIAGRAM



Energy Demand Forecasting Solution -

<https://gallery.cortanaintelligence.com/Solution/Energy-Demand-Forecasting-4>



LAB

Deploy the Energy Demand Forecast solution template



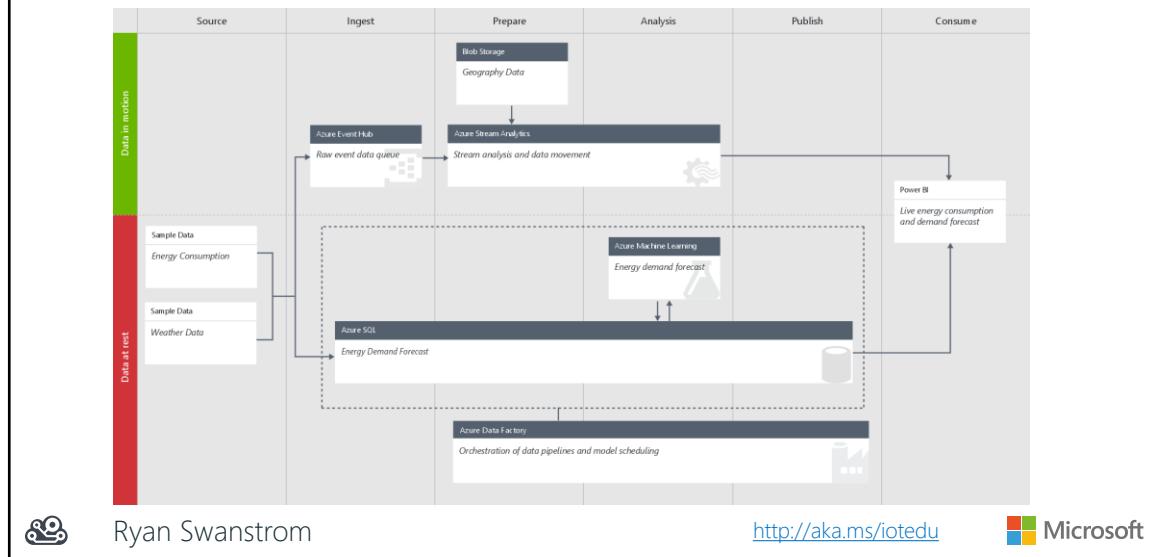
Ryan Swanstrom

<http://aka.ms/iotedu>



Follow along with the instructor and/or instructions under “Labs” on the class site and the file called “LabProvision.md”

CORTANA INTELLIGENCE ARCH



Ryan Swanstrom

<http://aka.ms/iotedu>



Energy Demand Forecasting Solution -

<https://gallery.cortanaintelligence.com/Solution/Energy-Demand-Forecasting-4>

Cortana Intelligence Energy Demand Forecasting Github Repo -

<https://github.com/Azure/cortana-intelligence-energy-demand-forecasting>

PREDICTIVE ANALYTICS FOR IoT

Event Hubs + Streaming Analytics + Storage + SQL + Machine Learning + Data Factory = AWESOME

PART 4 OF 5: DATA UNDERSTANDING



Ryan Swanstrom

<http://aka.ms/iotedu>

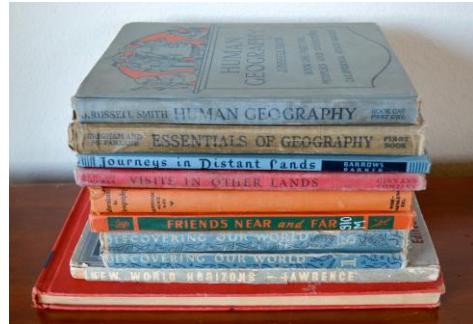


DATA TYPES

Streaming



vs. Historic (Reference)



Ryan Swanstrom

<http://aka.ms/iotedu>



What are some examples of streaming data and the use cases around them?

What about historic data?

STREAMING - TYPES AND EXAMPLES

- Sensor (humidity, temperature, altitude) – multidimensional
- Resource-constrained or limited by processing, storage and often battery life (many hand-held devices are this way)
- Text-based (email – you may detect spam as email comes in)
- Video
- Sounds
- ...almost anything with a timestamp



Ryan Swanstrom

<http://aka.ms/iotedu>



REFERENCE – TYPES AND EXAMPLES

- Regional, historical (temperature and humidity aggregated by day for a city)
- Flight data (airports, delays, air speed, departure and arrival times)
- Harvest data (crop yields from previous seasons)
- Load data (energy consumption for a suburb within a city grouped by day of the week for the past 5 years)
- Often when we use the words *past, grouped-by, aggregated...*



Ryan Swanstrom

<http://aka.ms/iotedu>



ROAD-BUMPS (FOR STREAMING)

- Sometimes we can only access data points or observations one-at-a-time (not batch)
- Time and memory usage must not grow with the amount of data
- High frequency of observations, large amount and/or high dimensionality
- Unlabeled or partially labeled data
- Privacy and permissions
- Compliance



Ryan Swanstrom

<http://aka.ms/iotedu>



How to perform stats or train/retrain machine learning models

In resource-constrained might be limited by the amount of storage on a device

Labels are needed for many machine learning algorithms (those of the supervised learning type)

Maintain privacy while being able to extract patterns from the data

Complying with laws and regulations while performing useful tasks on data

PREDICTIVE ANALYTICS FOR IoT

Event Hubs + Streaming Analytics + Storage + SQL + Machine Learning + Data Factory = AWESOME

PART 5 OF 5: DATA IN MOTION



Ryan Swanstrom

<http://aka.ms/iotedu>



STREAMING DATA CASE STUDIES

<http://aka.ms/iotedu>



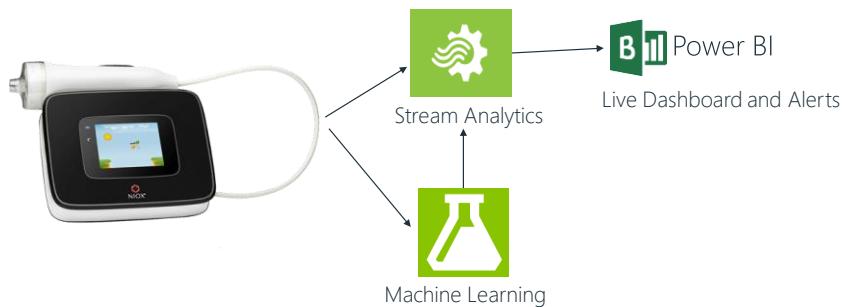


How real time analytics changes the business dynamics in the healthcare industry

Millions of asthma sufferers worldwide depend on Aerocrine monitoring devices to diagnose and treat their disease effectively. But those devices are sensitive to small changes in ambient environment. Aerocrine is using Stream Analytics to detect anomalies in Humidity which could potentially lead to device malfunction.



Ryan Swanstrom



"The ability to collect vital telemetry data from deployed devices has been a key objective of ours. A cloud-based solution allowing us to collect data on device performance in real-time helps us to be more proactive in our customer support and ensure that our NIOX devices help the physicians deliver the best possible outcome for the patient."

- Scott Myers, CEO of Aerocrine.

<http://aka.ms/iotedu>



Press Release:
Case Study:



FUJITSU

Producing High Quality Lettuce based on Real time insights

Multiple Sensor Data from the Lettuce factory being analyzed real time to product Lettuce that is both delicious and low in potassium so that dialysis patients and people with chronic disease can consume it.



Ryan Swanstrom



Stream Analytics



Machine Learning



Power BI
Live Dashboard and Alerts

↑ Timely Actions based on real time dashboard and alerts

Fujitsu and Microsoft collaborate to transform manufacturing processes through IoT innovation

Posted April 12, 2015 By Microsoft News Center

f 7 t 56 in 38 em

HANOVER, Germany and TOKYO — April 12, 2015 — At the Hannover Messe event, Fujitsu Limited and Microsoft Corp.

<http://aka.ms/iotedu>

 Microsoft

Press Release:
Case Study:



ThyssenKrupp

Differentiated Customer Service and Bottomline improvement through Preventive Maintenance and Connection Operations

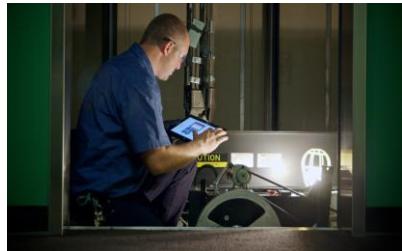


Stream Analytics



Machine Learning

Alerts & Dashboard helping timely and efficient scheduling of the technicians



"We wanted to go beyond the industry standard of preventative maintenance, to offer predictive and even *preemptive* maintenance, so we can guarantee a higher uptime percentage on our elevators." — CEO of the Company

<http://aka.ms/iotedu>



Ryan Swanstrom

Press Release:

Case Study:

CANONICAL SCENARIOS

Ingest



ARCHIVING

Analytics



TELEMETRY/LOG PROCESSING

Actions



INTERNET OF THINGS



Ryan Swanstrom

<http://aka.ms/iotedu>

Ingest Use Case: Blob Historian

The expectations for fast and agile execution in businesses continue to grow. Businesses and developers are now opting for easy-to-use, cloud-based platforms to cope with the demand for more agility, and are looking for platforms that enable them to ingest and process a continuous stream of data generated by their systems in near real-time. The canonical Blob Historian scenario can be described as follows: Data from various devices and platforms which are geo-distributed across the world are pushed to a centralized data collector. Once the data is at the central location, some stateless transformation is performed on them such as scrubbing PII information, adding geo-tagging, IP lookup etc. The transformed data is then archived into Blob storage in a fashion which can be **readily consumed by HDInsight** for offline processing. Also **Replay for RCA** etc.

Analytics Use Case: Telemetry/Log Processing

Monitoring to reduce TTD, TTM

As the volume of devices, machines and applications grow, the most common enterprise use case to run businesses is the need to monitor and respond to changing business needs by creating rich analytics near real-time.

The canonical Telemetry/Log Processing scenario can be described using the Online

Service or Application example, however the pattern is commonly seen across businesses that collect and report on application or device telemetry. The application or service regularly collects health data (data representing the current status of the application or infrastructure at a point in time) in addition to user request logs and other data representing actions or activities performed within the application. The data is historically saved to a blob or other type of data store for further processing. With the recent trend towards real-time dashboarding, in addition to saving the data to a blob or other type of store for historical analysis, customers are looking to process and transform the stream of incoming data directly such that it can be immediately provided to end users in the form of Dashboards and/or Notifications when action needs to be taken, for example if the site goes down operations personnel can be notified to begin investigation and resolve the issue quickly. As more data is gathered and processed Machine Learning can also be used to develop and learn from patterns seen in the system such that it is possible to better predict when machines may need to be serviced or when things are about to go wrong.

Operations Use Case: IoT Scenario

Command and Control, Maintenance

As devices become smarter and more devices are built with communication capabilities, the expectation of what can be done with the data generated and collected from these devices continues to evolve both in the commercial and consumer spaces. It is expected that with so much data available, we can quickly combine and process the data, gaining more insight into the environment around us, and the devices we use regularly.

The canonical IoT Scenario is often described using the **Vending Machine** example, however the pattern is commonly seen across IoT use cases. The Devices, the Vending machines, regularly send information (product stock, status, temperature, etc. data) to either a field gateway (if the device is Non-IP capable) or to a cloud gateway (IP Capable) for ingestion into the system. The incoming data stream is processed and transformed such that it can be immediately provided to end users in the form of Dashboards and Notifications when action needs to be taken, for example when product in a specific vending machine gets low, the relevant representative can be notified to restock the machine, or if the machine is in need of repair a technician can be scheduled. In some cases, the action that needs to be taken may be as simple as rebooting the machine or pushing down a firmware upgrade which can be done without the need for human interaction.

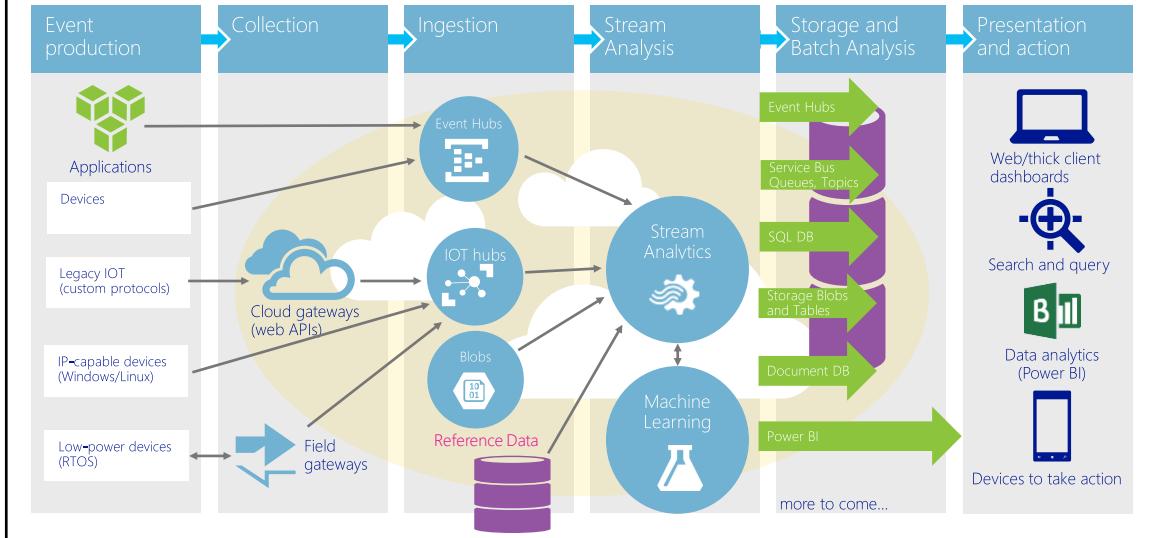
As more data is gathered and processed Machine Learning can also be used to develop and learn from patterns seen in the system such that it is possible to better predict when machines may need to be serviced or when things are about to go wrong.

PATTERN AND ARCHITECTURE

<http://aka.ms/iotedu>



CANONICAL STREAM ANALYTICS PATTERN



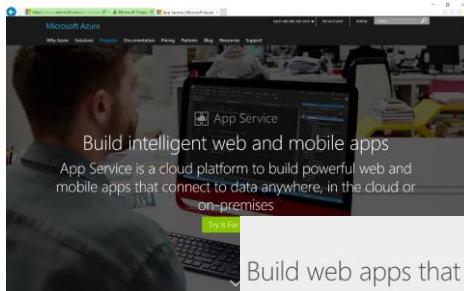
- **Stream Analytics is a building block** in a world of larger compositions of platform services
- Prominent integrations with **Event Hubs**, **Machine Learning** (new!), various output stores, and **Power BI**

WEB JOBS

<http://aka.ms/iotedu>



WEBJOBS ARE REALLY PART OF A WEB APP IN A APP SERVICE



More reasons to invest in learning about Visual Studio and GitHub, and even Python!



Ryan Swanstrom

<http://aka.ms/iotedu>



Azure Web Jobs has a new UX - <https://azure.microsoft.com/en-us/resources/videos/azure-web-jobs-and-new-ux/>

What are they? How can you create them? What types of scripting languages can you use?

Answers here: <https://azure.microsoft.com/en-us/documentation/articles/web-sites-create-web-jobs/>

How to programmatically leverage them with SDKs (note: samples on github):

<https://azure.microsoft.com/en-us/documentation/articles/websites-dotnet-webjobs-sdk/>

TYPES OF WEB JOBS

Triggered

- Web hook
- Scheduled – can set as a CRON expression (in portal even)

Continuous

- Stays up and alive
- Used for services



Ryan Swanstrom

<http://aka.ms/iotedu>





LAB

Monitor and explore data in motion and at rest



Ryan Swanstrom

<http://aka.ms/iotedu>



Follow the instructions under “Labs” -> LabWebJobs.md

MORE WEB JOBS RESOURCES

The screenshot shows a Microsoft Azure documentation page. At the top, there's a navigation bar with links for Why Azure, Solutions, Products, Documentation, Pricing, Partners, Blog, Resources, and Support. Below the navigation is a search bar and a 'FREE ACCOUNT' button. The main content area has a title 'Run Background tasks with WebJobs' and a sub-section 'Overview'. It includes a profile picture of Tom Dykstra, the update date (04/27/2016), and a 'Contributors' section. A sidebar on the left lists 'Overview', 'Get Started', 'Deploy', 'Scale', 'Manage', and 'Monitor'.

Deploy WebJobs using Visual Studio

By Tom Dykstra Updated: 04/27/2016 Contributors: Edit on GitHub

Overview

This topic explains how to use Visual Studio to deploy a Console Application project to a web app in App Service as an Azure WebJob. For information about how to deploy WebJobs by using the Azure Portal, see Run Background tasks with WebJobs.



Ryan Swanstrom

<http://aka.ms/iotedu>



Deploy from VS in various languages, but here's .NET sdk example:

Deploy WebJobs using Visual Studio - <https://azure.microsoft.com/en-us/documentation/articles/websites-dotnet-deploy-webjobs/>

EVENT HUBS

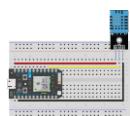
<http://aka.ms/iotedu>



**SOLUTION
ARCHITECTURE**
EVENT HUB



Microsoft Azure



Energy
consumption
data (public)

Receive



<name>ns
Service Bus Namespace



<name>eh
Event Hub
Queues streaming data



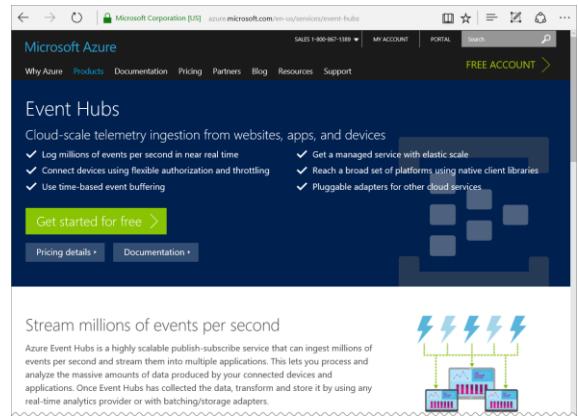
<name>ws
Web Job
Runs jobs to scrape data and more

Send

Pink background: Data in motion or fast-moving data
Green/blue background: Data at rest or slow-moving data

AZURE 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."



The screenshot shows the Microsoft Azure Event Hubs landing page. At the top, there's a navigation bar with links for Microsoft Corporation [US], azure.microsoft.com, MY ACCOUNT, PORTAL, and SIGN IN. Below the navigation is a search bar. The main heading is "Event Hubs" with a sub-section "Cloud-scale telemetry ingestion from websites, apps, and devices". To the right, there's a list of features with checkmarks: "Log millions of events per second in near real time", "Connect devices using flexible authorization and throttling", "Use time-based event buffering", "Get a managed service with elastic scale", "Reach a broad set of platforms using native client libraries", and "Pluggable adapters for other cloud services". Below this is a "Get started for free >" button and "Pricing details" and "Documentation" links. On the right side, there's a diagram showing four lightning bolts pointing down to a central processing unit icon.



Ryan Swanstrom

<http://aka.ms/iotedu>

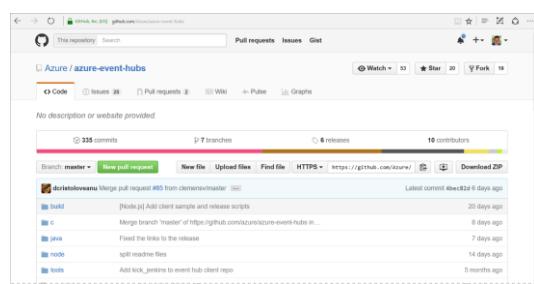
 Microsoft

Learn more here: <https://azure.microsoft.com/services/event-hubs/>

AZURE-EVENT-HUBS REPOSITORY

Azure Event Hub clients for

- C
- Java
- Node.js



<https://github.com/Azure/Azure-event-hubs>



Ryan Swanstrom

<http://aka.ms/iotedu>

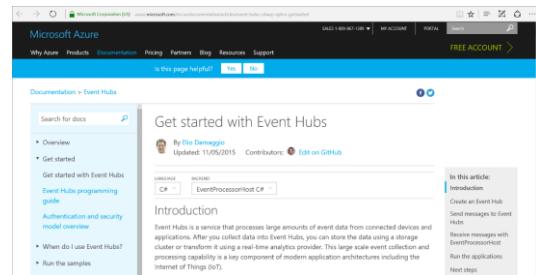


<https://github.com/Azure/Azure-event-hubs>

.NET EVENT HUB CLIENT TUTORIAL

The .NET client library is available as a NuGet package.

Get started using Event Hubs from C#. This tutorial helps you create both a sender and a receiver using the "[Microsoft Azure Service Bus](#)" and "[Microsoft Azure Service Bus Event Hub - EventProcessorHost](#)" NuGet packages.



The screenshot shows the Microsoft Azure documentation for 'Get started with Event Hubs'. It features a sidebar with links like 'Overview', 'Get started', and 'Event Hubs programming guide'. The main content area displays the 'EventProcessorHost' sample code in C#, which is a template for creating an Event Processor Host application. The code uses the Microsoft.Azure.EventHubs namespace to interact with Event Hubs.

<http://aka.ms/ehcs>



Ryan Swanstrom

<http://aka.ms/iotedu>



<http://aka.ms/ehcs>

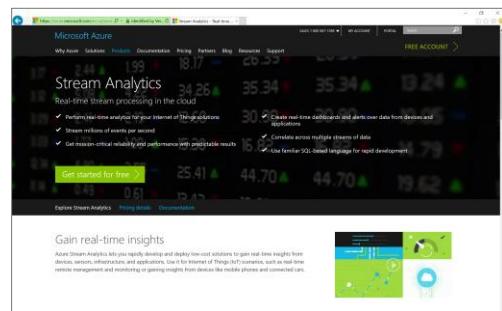
STREAM ANALYTICS

<http://aka.ms/iotedu>



STREAM ANALYTICS

Azure Stream Analytics is an **event processing engine** in the cloud that uncovers insights from devices, sensors, cloud infrastructure, and applications in real-time. With **out-of-the-box integration to Event Hubs**, the combined solution can ingest millions of events and do analytics to better understand patterns, **power a dashboard**, or kick off an action while data is being **streamed in real-time**.



<https://azure.microsoft.com/services/stream-analytics/>



Ryan Swanstrom

<http://aka.ms/iotedu>



Landing page: <https://azure.microsoft.com/en-us/services/stream-analytics/>

REAL-TIME ANALYTICS

Fully managed
real-time analytics



Real-time Analytics

- Intake **millions of events per second** (up to 1 GB/s)
- **Low processing latency**, auto adaptive (sub-second to seconds)
- **Correlate** between different streams, or with reference data
- Find **patterns** or lack of patterns in data in real-time

Fully Managed Cloud Service

- No hardware acquisition and maintenance
- No platform/infrastructure deployment and maintenance
- Easily **expand your business globally** leveraging Azure regions



Ryan Swanstrom

<http://aka.ms/iotedu>



MISSION CRITICAL

Mission critical
reliability and scale



Mission Critical Reliability

- **Guaranteed event delivery**
- **Guaranteed business continuity:** Automatic and fast recovery

Effective Audits

- **Privacy and security** properties of solutions are evident
- Azure integration for **monitoring and ops alerting**

Easy To Scale

- **Scale** from small to large on demand



Ryan Swanstrom

<http://aka.ms/iotedu>



31:00

RAPID DEVELOPMENT

Enables rapid development



Rapid Development with SQL like language

- **High-level:** focus on stream analytics solution
- **Concise:** less code to maintain
- **Fast test:** Rapid development and debugging
- First-class support for **event streams and reference data**

Built in temporal semantics

- Built-in **temporal windowing and joining**
- Simple **policy configuration** to manage out-of-order events and late arrivals



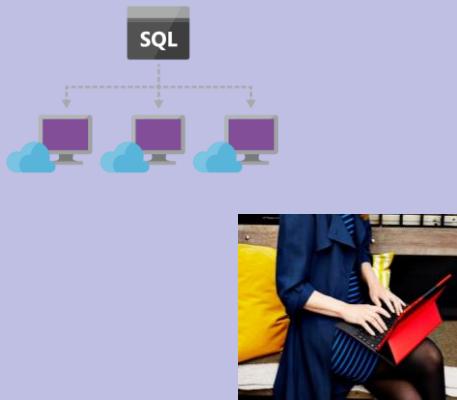
Ryan Swanstrom

<http://aka.ms/iotedu>



33:00

QUERY LANGUAGE



- You write declarative queries in SQL

No code compilation, easy to author and deploy

- Unified programming model

Brings together event streams, reference data and machine learning extensions

- Temporal Semantics

All operators respect, and some use, the temporal properties of events

- Built-in operators and functions

These should (mostly) look familiar if you know relational databases

Filters, projections, joins, windowed (temporal) aggregates, text and date manipulation

<http://aka.ms/iotedu>



Data-first, queries-first

Focus on what matters most

No need to spend most of your time on *plumbing* code

Democratize Event Processing

Language and constructs familiar to most developers

Simplify, empower, extend

Built-in operators, powerful but easy to use

Temporal semantics (no PhD required)

Machine learning ready

Unify

Live streams, historical and reference data, all together, same query language, same runtime, same results

QUERY LANGUAGE

A Subset of T-SQL

Can take data in from either streams (Event Hubs) or static data (Blobs) or relational data (SQL Database)

Can join multiple streams. For example you could join event hub message data with relational data from a database

Can output data to streams, static data, databases, or Power BI.

Has powerful windowing functions for creating aggregations over Tumbling, Hopping or Sliding time windows



Ryan Swanstrom

The screenshot shows a Microsoft Azure documentation page for Stream Analytics Query Language. The page has a navigation bar with links for Features, Pricing, Documentation, Downloads, Add-ons, Community, and Support. A 'FREE TRIAL' button is also present. The main content is titled 'Stream Analytics Query Language Reference'. It includes a table of contents with sections like Microsoft Azure, Stream Analytics Query Language, Built-in Functions, Date/Time, Query Language Elements, and Time Management. The main body of the page discusses Stream Analytics Query Language as a subset of T-SQL syntax and provides a scenario titled 'The toll booth scenario'.

<https://msdn.microsoft.com/library/azure/dn834998.aspx>

<http://aka.ms/iotedu>



Query examples for common Stream Analytics usage patterns -

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-stream-analytics-query-patterns>

Link from slide: <https://msdn.microsoft.com/library/azure/dn834998.aspx> – query language reference

It is essential to understand the progress of time. Every event that flows through the system comes with a timestamp that can be accessed via [System.Timestamp](#).

Toll Booth scenario

A tolling station is a common phenomenon – we encounter them in many expressways, bridges, and tunnels across the world. Each toll station has multiple toll booths, which may be manual – meaning that you stop to pay the toll to an attendant, or automated – where a sensor placed on top of the booth scans an RFID card affixed to the windshield of your vehicle as you pass the toll booth. It is easy to visualize the passage of vehicles through these toll stations as an event stream over which interesting operations can be performed.

What are two different ways a timestamp could be made in this scenario?

S(A)QL – LANGUAGE & LIBRARY

TSQL Subset

- SELECT
- FROM
- WHERE
- GROUP BY
- HAVING
- CASE WHEN THEN ELSE
- INNER/LEFT OUTER JOIN
- UNION
- CROSS/OUTER APPLY
- CAST
- INTO
- ORDER BY ASC, DSC

Scaling Extensions

- WITH
- PARTITION BY
OVER



Ryan Swanstrom

Date and Time Functions

- DateName, DatePart, Day
- Month, Year, DateDiff
- DateTimeFromParts, DateAdd

Mathematical Functions

- ABS, CEILING, EXP, FLOOR
- POWER, SIGN, SQUARE
- SQRT

Temporal Functions

- Lag, IsFirst, Last
- CollectTop

Windowing Extensions

- TumblingWindow
- HoppingWindow
- SlidingWindow

Aggregate Functions

- SUM
- COUNT
- AVG
- MIN
- MAX
- STDEV
- STDEVP
- VAR
- VARP
- TopOne

String Functions

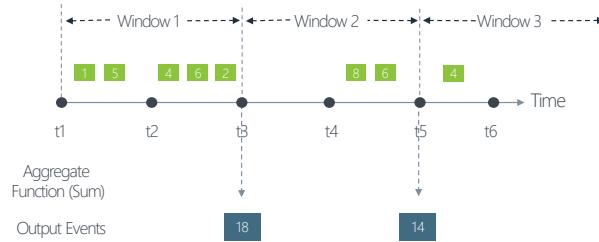
- Len, Concat, CharIndex
- Substring, Lower
- Upper, PatIndex

<http://aka.ms/iotedu>

- Let's take a quick look at the ASA language – Stream Analytics Query Language
- A subset of T-SQL, extended with temporal concept, aggregation and dynamic schema
- Built-In Functions (purple) <https://msdn.microsoft.com/en-us/library/azure/dn835030.aspx>
- T-SQL-based functions (black) <https://msdn.microsoft.com/en-us/library/azure/mt582049.aspx>

WINDOWING CONCEPTS

- Windows can be – Hopping, Sliding or Tumbling
- Windows are **fixed** length
- Output event will have the timestamp of the **end** of the window
- Must be used in a **GROUP BY** clause



Ryan Swanstrom

<http://aka.ms/iotedu>

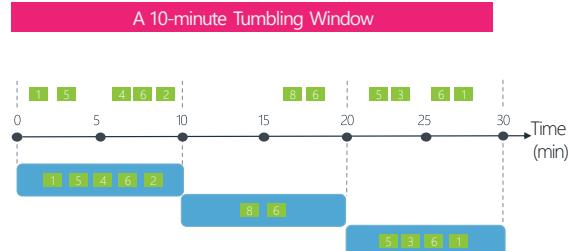


- Windowing is a core requirement for streaming analytic applications
- Common requirement to perform some set-based operation (count, aggregation etc.) over events that arrive within a specified period of time
- Azure Stream Analytics supports three types of windows: Hopping, Sliding and Tumbling
- Every window operation outputs events at the end of the window
- The output of the window will be single event based on the aggregate function used. The event will have the time stamp of the end of the window
- All windows have a fixed length
- All windows should be used in a GROUP BY clause

TUMBLING WINDOWS

Tell me the sum of the energy load over 10 minute intervals

```
SELECT
    System.Timestamp as T,
    PTID,
    SUM(Load) as Load
FROM
    InputEventHub
    TIMESTAMP BY utcTimeStamp
    GROUP BY PTID, TumblingWindow(minute, 10)
```



Ryan Swanstrom

<http://aka.ms/iotedu>



Tumbling windows:

- Repeat
- Are non-overlapping

An event can belong to only one tumbling window

Note on timestamps in stream analytics

In Azure Stream Analytics, all events have a well-defined timestamp.

You may use the `TIMESTAMP BY` keyword to specify the column in the payload

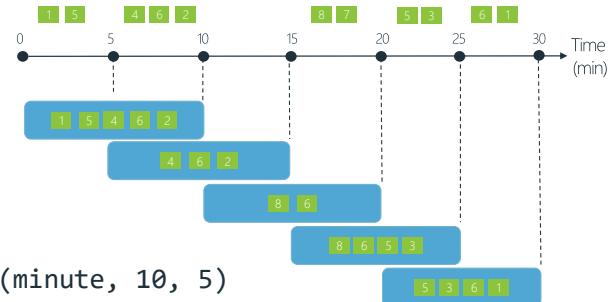
Example query with `TIMESTAMP BY`:

HOPPING WINDOWS

Every 5 minutes give me the sum of the energy load over the last 10 minutes

```
SELECT
    System.Timestamp as T,
    PTID,
    SUM(Load) as Load
FROM
    InputEventHub
    TIMESTAMP BY utcTimeStamp
    GROUP BY PTID, HoppingWindow(minute, 10, 5)
```

A 10-minute Hopping Window with a 5-minute "Hop"



Ryan Swanstrom

<http://aka.ms/iotedu>



Hopping windows:

- Hop forward in time by a fixed period
- Repeat
- Can overlap

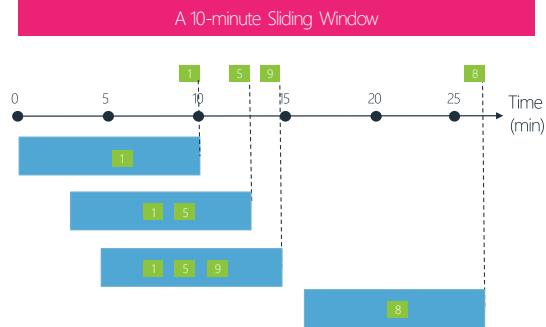
Same as tumbling window if hop size = window size

Events can belong to more than one hopping window

SLIDING WINDOWS

Give me the summed energy load when it is more than 3 energy units in the last 10 minutes

```
SELECT
    System.Timestamp as T,
    PTID,
    SUM(Load) as Load
FROM
    InputEventHub
    TIMESTAMP BY utcTimeStamp
    GROUP BY PTID, SlidingWindow(minute, 10)
    HAVING SUM(Load) > 3
```



Ryan Swanstrom

<http://aka.ms/iotedu>

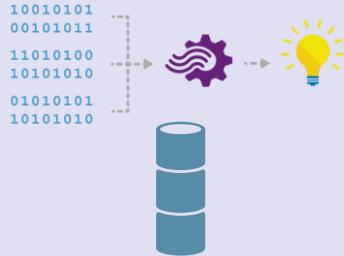


Sliding window:

- Continuously moves forward by an ϵ (epsilon)
- Produces an output *only during the occurrence of an event*
- Every window will have at least one event

Events can belong to more than one sliding window

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
```

<http://aka.ms/iotedu>

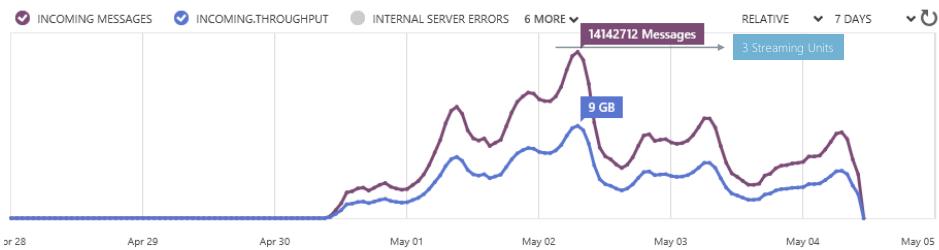


This has been done also for us in the ASA job query so you may wish to take a look.

ADVANTAGE OF CLOUD – SCALE AND COSTS

analytics_events

DASHBOARD CONFIGURE CONSUMER GROUPS



Ryan Swanstrom

<http://aka.ms/iotedu>



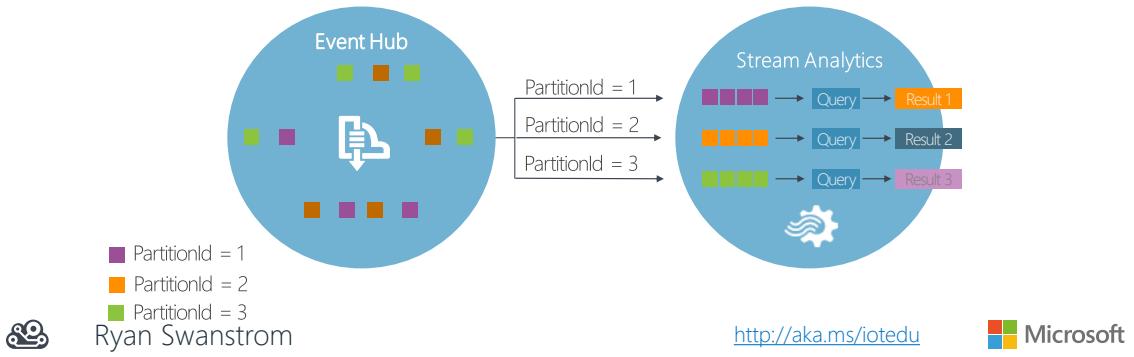
Pay only for what you use and per hour billing

Pricing details: <https://azure.microsoft.com/en-us/pricing/details/stream-analytics/>

SCALING USING PARTITIONS

Partitioning allows for parallel execution over scaled-out resources

```
SELECT SUM(*) AS Sum, PTID  
FROM InputEventHub PARTITION BY PartitionId  
GROUP BY TumblingWindow(minute, 10), PTID, PartitionId
```



More Advanced, but here for reference.

When a query is partitioned, input events will be processed and aggregated in separate partition groups

Output events are produced for each partition group

Partitioning a step enables more streaming units to be allocated to a job as there is a limit on the number of units that can be assigned to an un-partitioned step.

Partitioning requires that all three conditions listed in the slide be satisfied.

When a query is partitioned, the input events will be processed and aggregated in separate partition groups, and outputs events are generated for each of the groups. If a combined aggregate is desirable, you must create a second non-partitioned step to aggregate.

You can always partition by the PartitionId field, which is a built-in field in your query. The PartitionId field indicates from which partition of source data stream the event is from. Since Event Hubs supports partitioning, you can easily develop partitioned queries that read data from Event Hubs.

MULTIPLE STEPS, MULTIPLE OUTPUTS

```
WITH Step1 AS (
    SELECT Count(*) AS CountTweets, Topic
    FROM TwitterStream PARTITION BY PartitionId
    GROUP BY TumblingWindow(second, 3), Topic, PartitionId
),
Step2 AS (
    SELECT Avg(CountTweets)
    FROM Step1
    GROUP BY TumblingWindow(minute, 3)
)
SELECT * INTO Output1 FROM Step1
SELECT * INTO Output2 FROM Step2
SELECT * INTO Output3 FROM Step2
```



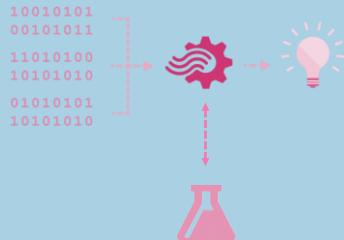
Ryan Swanson

- A query can have **multiple steps** to enable pipeline execution
- A step is a sub-query defined using **WITH** ("common table expression")
- Can be used to develop complex queries more elegantly by creating a **intermediary named result**
- Creates unit of execution for scaling out when **PARTITION BY** is used
- Each step's output can be sent to multiple output targets using **INTO**

<http://aka.ms/iotedu>

More advanced, but here for reference.

ASA and MACHINE LEARNING



Azure ML and Stream Analytics are now integrated

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 Load, predict(Load) AS stforecast  
FROM InputStream
```

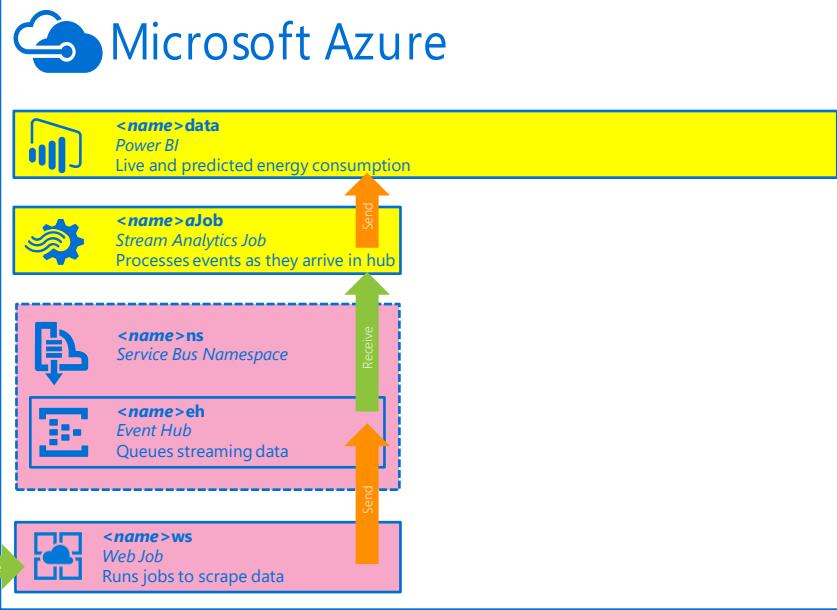
`predict` mapped to endpoint/API key

<http://aka.ms/iotedu>



- ASA-AML integration
- REST endpoint created by operationalizing ML model can be bound to a function (UDF)

**SOLUTION
ARCHITECTURE**
**STREAM
ANALYTICS
WITH POWER BI**



TIPS TO KEEP IN MIND FOR STREAM ANALYTICS

- It is critical to identify the “timestamp” in the incoming stream of data and one should ensure that the time captured also **confirms** the occurrence of an event.
- We recommend using **TIMESTAMP BY** over arrival time as a best practice.
- All field names referenced in the query are case ***insensitive*** – so be careful with the schema key names in the payload (from event hub)

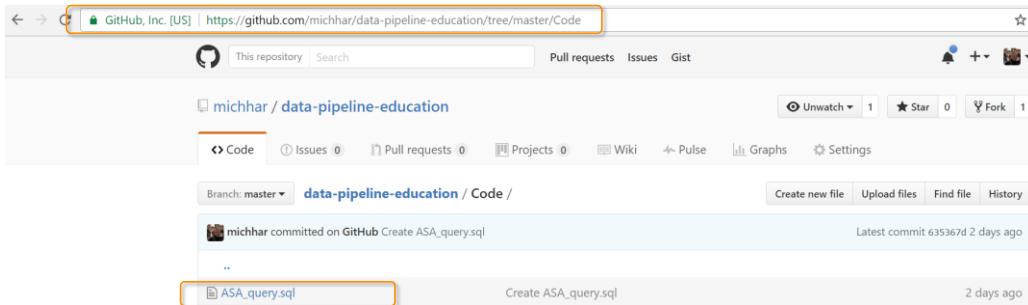


Ryan Swanstrom

<http://aka.ms/iotedu>



GROKING THE STREAM ANALYTICS QUERY



Ryan Swanstrom

<http://aka.ms/iotedu>



Get to this query file at https://github.com/michhar/data-pipeline-education/blob/master/Code/ASA_query.sql

GROKING THE STREAM ANALYTICS QUERY

```
CREATE TABLE InputEventHub (
    utcTimeStamp datetime,
    PTID bigint,
    Load float
);
```

Issuing the “[CREATE TABLE](#)” statement before the query allows us to define the schema of the data. Any incoming messages that don’t meet this schema will be rejected. This helps prevent the Stream Analytics job from failing on invalid messages.



Ryan Swanstrom

<http://aka.ms/iotedu>



Subset of this file: https://github.com/michhar/data-pipeline-education/blob/master/Code/ASA_query.sql

GROKING THE STREAM ANALYTICS QUERY

```
SELECT
    a.utcTimeStamp as Time,
    b.Name as Region,
    a.PTID,
    b.Latitude,
    b.Longitude,
    a.Load as Demand
INTO
    <name>data
FROM (
    SELECT
        utcTimeStamp,
        PTID,
        Load
    FROM
        <name>hub
)
a ...
```

The “**SELECT**” statement defines the final values that will be output to the Power BI data set. We really aren’t taking much advantage of Stream Analytics processing capabilities here, we’re just forwarding the data long.

You could however be using Stream Analytics time windowing functions and SQL aggregations to generate summary values or even generate alert data based on data anomalies.

Power BI data set the data will be sent INTO

Event Hub source messages will come FROM



Ryan Swanstrom

<http://aka.ms/iotedu>



This is a subset of the statement found in the SQL code on the class site. Subset of this file: https://github.com/michhar/data-pipeline-education/blob/master/Code/ASA_query.sql

This portion represents the “hot path” to PBI (the other portion after the join will be discussed with the SQL database session later on). The rest of the query can be seen on the site.

“b” refers to some historical data coming in from blob storage (pulled in by the Web Job – same job that sends the streaming data to the Event Hub)

ASA 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



Ryan Swanstrom

<http://aka.ms/iotedu>





LAB

Create an ASA Query with Hopping Windows



Ryan Swanstrom

<http://aka.ms/iotedu>



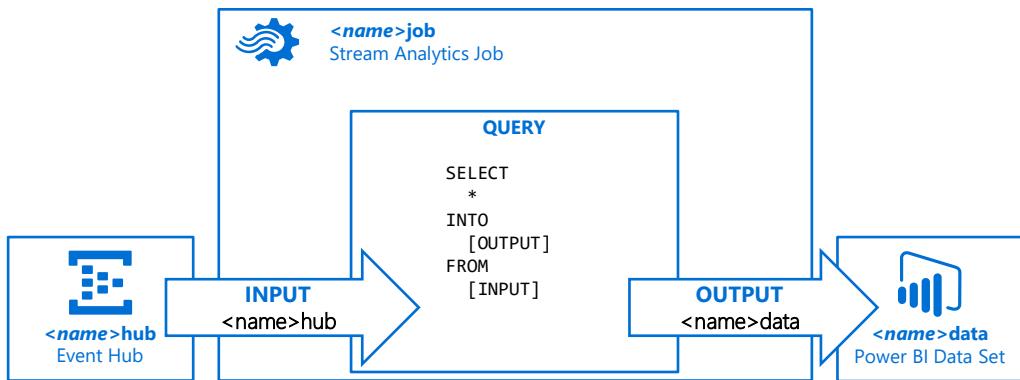
Follow the instructions under “Labs” -> LabASA.md

POWER BI (AND THE STREAM ANALYTICS INTEGRATION)

<http://aka.ms/iotedu>



STREAM ANALYTICS INPUTS & OUTPUTS



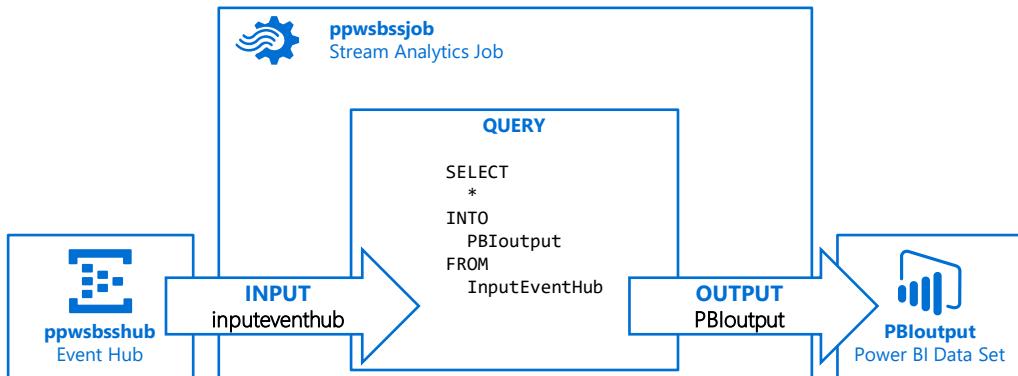
Ryan Swanstrom

<http://aka.ms/iotedu>



Use the “classic” portal to modify the query and hook up to PBI.
<https://manage.windowsazure.com>

STREAM ANALYTICS INPUTS & OUTPUTS



Ryan Swanstrom

<http://aka.ms/iotedu>



Side Note for Later: the OUTPUT, called PBIconnector is the output according to ASA. When we hook it up to PBI (the “hot path”) later you will see it gets renamed for the actual PBI online interface so the dataset, PBIconnector, will likely have a different name in PBI itself. This will become clearer later.

Query Code

Check out the more complete query on the class site at
https://github.com/michhar/data-pipeline-education/blob/master/Code/ASA_query.sql

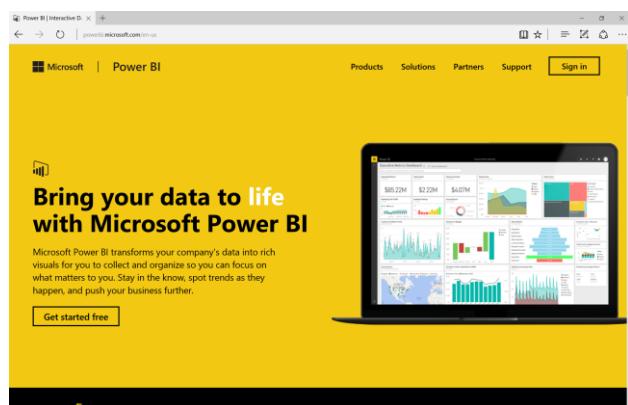
or on the “classic” portal at <https://manage.windowsazure.com> under the stream analytics job and its query

WHAT IS POWER BI ONLINE?

Create custom reports and dashboards using data from many different sources in an intuitive web interface.

Share datasets, reports and dashboards with your organization and others.

You need a **Work or School Account** (a.k.a. organizational email) to sign up for Power BI online (like Office 365).



Ryan Swanstrom

<http://aka.ms/iotedu>



There is also a Power BI desktop app. They are both free to use, but there are limitations on creating dashboards to share in desktop. (e.g. can't pin tiles from a desktop report to a dashboard online)

Power BI originated as a part of Excel.

You can't use outlook.com, gmail.com, etc. personal accounts (even a personal Microsoft Account). Check out this blog post for more info on signing up:
<http://BretStateham.com/Using-power-bi-at-events>

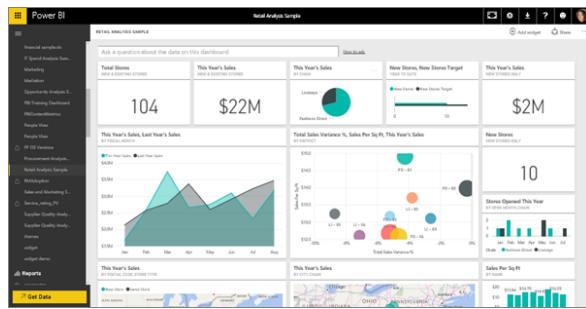
WHAT IS A POWER BI DASHBOARD

Dashboards consists of tiles which can be from PBI reports, other dashboards, PBI datasets, Excel reports, SSRS and more.

Pinning an entire report page creates *live* tiles one can interact with.

A tile is a snapshot of your data, pinned to the dashboard.

Quick Insights is a lightning fast way to create a dashboard from a dataset.



Ryan Swanstrom

<http://aka.ms/iotedu>



Dashboard help: <https://powerbi.microsoft.com/en-us/documentation/powerbi-service-dashboards/>

Sharing – read-only, personal or collaborator

Content as tiles

Tiles can be created from sources such as: Excel, an online PBI report, dataset, other PBI dashboard, the Q&A box, SQL Server Reporting Services and more (can't pin, however, directly from PBI Desktop reports**) – can even link On-Prem data with available software called the On-Premises Gateway

(<https://powerbi.microsoft.com/en-us/documentation/powerbi-gateway-onprem/>)

Standalone tiles can be created from: web content, images, videos, and text boxes

Can make hard copies to save or print

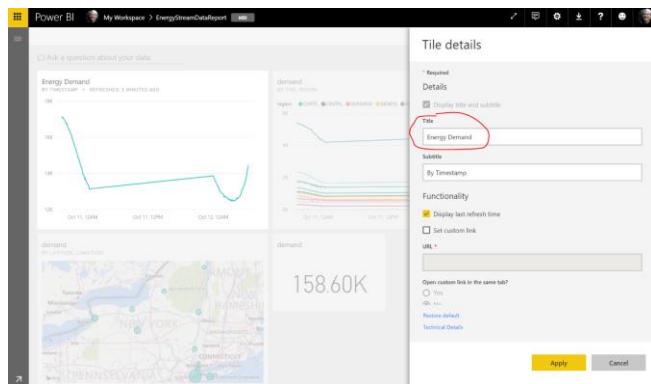
**Reports from PBI Desktop (as well as Excel and CSV) files may be used if they are store in OneDrive for Business.

Note:

If the original visualization used to create the tile changes, the tile doesn't change (unless it's a live tile). The data refreshes, but the visualization type does not. For example, if you pinned a line chart from a report and then you change the line chart to a bar chart, the dashboard tile continues to show a line chart.

WORKING IN THE DASHBOARD

From within the dashboard you can change the title and other properties.



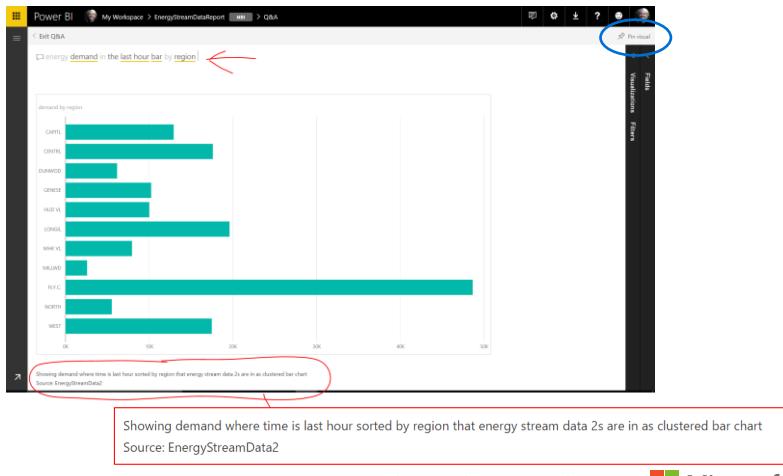
Ryan Swanstrom

<http://aka.ms/iotedu>



WORKING IN THE DASHBOARD

PBI Q&A –
“search” the
dataset in plain
text and pin a
visual
representing that
“query”



Ryan Swanstrom

<http://aka.ms/iotedu>



Red arrow is my query

Red circle is the actual description translating to the query

Blue circle is how we pin this visual to the dashboard

Remember this is all from the dashboard itself.

TIPS ON CREATING A GREAT DASHBOARD

- Know your audience - what they need will dictate what you need
- Keep the story to one screen – they are “information at-a-glance” and the underlying reports/data have deeper detail so keep it to the essentials
- Consider what devices will be used to view it
- Size of content should match importance
- Placement matters - the more general and/or more important information where reading begins (e.g. top-left in English)
- Keep visuals easy to read and interpret – don’t forget titles and labels



Ryan Swanstrom

<http://aka.ms/iotedu>



Tips site: <https://powerbi.microsoft.com/en-us/documentation/powerbi-service-tips-for-designing-a-great-dashboard/>

Questions to ask around these tips:

Audience: What are the key metrics that will help them make decisions? How will the dashboard be used? What learned or cultural assumptions may affect design choices? What information does your audience need to be successful? Where is the dashboard going to be displayed?

Story on one screen: Can you avoid scroll bars on your dashboard? Is the dashboard too cluttered?

Easy to read and interpret: Keep in mind that it is difficult for the human brain to interpret circular shapes. Be consistent with scales. If numbers get too long use words instead (3.4 million instead of 3,400,000) or scientific notation. Note on pie charts: less than 8 categories and they are good generally for conveying part-to-whole relationships.



LAB

Reporting with Power BI



Ryan Swanstrom

<http://aka.ms/iotedu>



Follow the instructions under “Labs” -> LabPBI.md

Given extra time or if you finish early, change the ASA query back to the original query saved from the last lab. Observed changes in the BI report and dashboard. You may wish to download the report to examine it more closely. Please explain this change.

WHAT YOU HAVE LEARNED SO FAR

- Gained some familiarity and practice with Web Jobs
- Gained some familiarity with what Event Hubs do for us
- Learned in some detail about Stream Analytics, scenarios and practiced monitoring and managing it
- Gained some knowledge and practice of Power BI and some tips on use



Ryan Swanstrom

<http://aka.ms/iotedu>



TOMORROW - WHAT TO EXPECT

- Learn a bit about relevant machine learning (basics to remind you) and Azure Machine Learning – get a chance to practice
- Become familiar with Azure Data Factory and how it's the glue that binds together many components of our architecture
- Gain some understanding around options for security and compliance
- End of training wrap-up



Ryan Swanstrom

<http://aka.ms/iotedu>

