**IT Operations Insight**

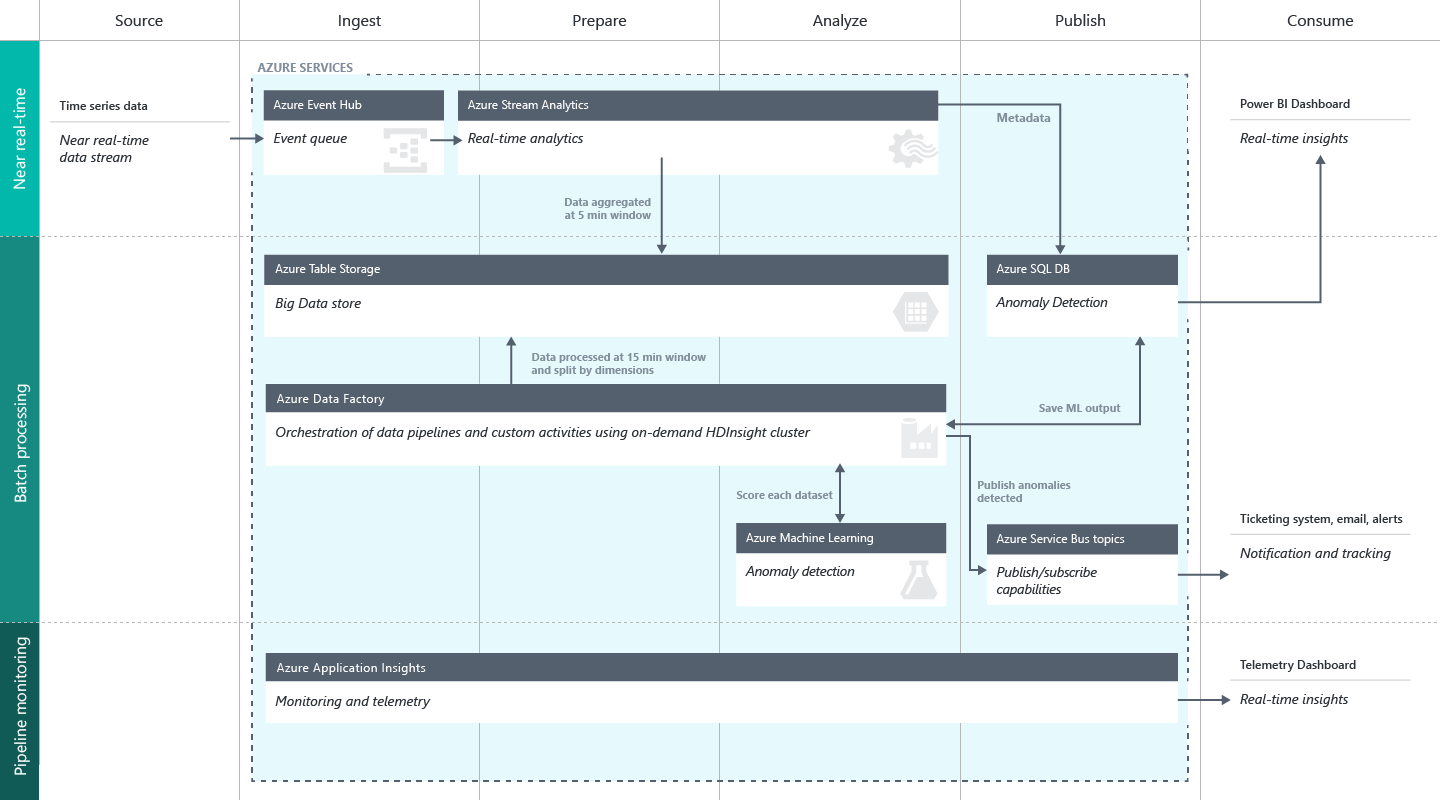
**Pre-Configured Solution**

**CIQS Deployment**

**Overview**

The IT Operations Insight Pre-Configured Solution will deploy an end to end solution into your subscription. The goal of this document is to explain the reference architecture and different components provisioned in your subscription as part of the deployment. The document also talks about how to start sending sample data or real data to be able to see insights. Instructions on how to build the Power BI dashboard for this solution are provided at the end.

### **Architecture**



When the pipeline is deployed, various Azure services within Cortana Analytics Suite are activated (i.e. Event Hub, Stream Analytics, Storage, HDInsight, Data Factory, Machine Learning, etc.). The architecture diagram above shows, at a high level, how the IT Operations Insight Pre-Configured Pipeline is constructed from end-to-end.

## ****Data Source and Ingestion****

### **Azure Event Hub**

The [Azure Event Hub](https://azure.microsoft.com/services/event-hubs/) service is the recipient of the input data provided by the data source. The data source can be the data generator provided or any service that can send data to Azure Event Hub in the required format. Read more about it in the following section.

## ****Data Preparation and Analysis****

### **Azure Stream Analytics**

The [Azure Stream Analytics](https://azure.microsoft.com/services/stream-analytics/) service is used to provide near real-time analytics on the input stream from the [Azure Event Hub](https://azure.microsoft.com/en-us/documentation/articles/cortana-analytics-technical-guide-demand-forecast/?tduid=%28ff7611aab34207ef35998cad0ae7b15b%29%28256380%29%282459594%29%28je6NUbpObpQ-vVSJiDiAo4JYZVCC3m1roQ%29%28%29#azure-event-hub) service and archiving all raw incoming events to the [Azure Storage](https://azure.microsoft.com/services/storage/) service for later processing by the [Azure Data Factory](https://azure.microsoft.com/documentation/services/data-factory/) service.

### **Azure Data Factory**

The [Azure Data Factory](https://azure.microsoft.com/documentation/services/data-factory/) service orchestrates the movement and processing of data. The data factory is made up of [pipelines](https://azure.microsoft.com/en-us/documentation/articles/data-factorydata-factory-create-pipelines/) and activities for preparing, analyzing and publishing results. It uses custom activities to read raw data from the input storage tables, prepare individual time series datasets that will be sent to Azure Machine Learning - Anomaly Detection API for scoring and then publishes the results as follows.

## ****Data Publishing****

### **Azure SQL Database Service**

The [Azure SQL Database](https://azure.microsoft.com/services/sql-database/) service is used to store (managed by Azure Data Factory) the anomaly detection scores received by the Azure Machine Learning service that will be consumed in the [Power BI](https://powerbi.microsoft.com/) dashboard.

## ****Data Consumption****

### **Power BI**

The [Power BI](https://powerbi.microsoft.com/) service is used to show a dashboard that contains aggregations provided by the [Azure Stream Analytics](https://azure.microsoft.com/services/stream-analytics/) service as well as anomaly detection score results stored in [Azure SQL Database](https://azure.microsoft.com/services/sql-database/) that were produced using the [Azure Machine Learning](https://azure.microsoft.com/services/machine-learning/) service. For Instructions on how to use the Power BI dashboard for this solution, refer to the section below.

**Post Deployment Steps**

Below are the steps to get started with the deployed solution.

Step 1: Once the solution is deployed successfully; you can start sending data into it. The data source can be a simulated data source like sample data generator provided or you can bring your own data as described below.

Step 2: Monitor if data is flowing into your pipeline.

Step 3: Lastly, visualize the results using Power BI dashboard.

### **Synthetic Data Source**

You can use [data generator](https://microsoft.sharepoint.com/teams/CECloudML/_layouts/15/guestaccess.aspx?guestaccesstoken=%2b%2bO6JPCI0sq461mNJvKwCiMkKuHnOyOp1PhP6uID8P8%3d&docid=2_134e1e20f90bc4d428052b8c87285071b&rev=1) as the data source to provide synthetic data to the deployed pipeline. The data generator is a desktop application that you can download and run locally after successful deployment. You will find the instructions to download and install this application from [**Data Generator Instructions**](https://microsoft.sharepoint.com/teams/CECloudML/_layouts/15/guestaccess.aspx?guestaccesstoken=A8QYBWwvymHFjuFDmp7jMKbkeNEOlf3t1JRetXBOudo%3d&docid=2_02a108544cf0a405cb3868aeb62992a1a&rev=1). This application feeds the [Azure Event Hub](https://azure.microsoft.com/en-us/documentation/articles/cortana-analytics-technical-guide-demand-forecast/?tduid=%28ff7611aab34207ef35998cad0ae7b15b%29%28256380%29%282459594%29%28je6NUbpObpQ-tDK.bHRFEXOQuN7wy1uyeg%29%28%29) service with data points, or events, that will be used in the rest of the pipeline flow.

The event generation application will populate the Azure Event Hub only while it's executing on your computer.

**How to bring your own data**

This section describes how to bring your own data to Azure. As long as the events being sent to the Azure Event Hub follow the required schema, encoding and format, there are no additional changes to be made to the pipeline components.

Encoding: UTF-8

Formatting: JSON

Schema:

|  |
| --- |
| {  "Host": "Hostname (required”",  "Metric": "Counter/metric name (required)",  "Value": "Numeric value (required)",  "Application": "Originating application of the event (optional)",  "Brand": "Originating brand name of the event (optional)"  } |

The [Azure Event Hub](https://azure.microsoft.com/services/event-hubs/) service is very generic, such that data can be posted to the hub in either CSV or JSON format. No special processing occurs in the Azure Event Hub, but it is important you understand the data that is fed into it.

This document does not describe how to ingest your data, but one can easily send events or data to an Azure Event Hub, using the [Event Hub API](https://azure.microsoft.com/en-us/documentation/articles/event-hubsevent-hubs-programming-guide/).

## ****Monitor Progress****

Once the data generator/ data source starts sending events, the pipeline begins to get hydrated and the different components of your solution start kicking into action following the commands issued by the Data Factory. There are multiple ways you can monitor the pipeline.

1. Check the input data populated in Azure Table Storage and Azure SQL Database

The Stream Analytics job writes the raw incoming data to table storage and SQL database. If you click on Resource Group name in the CIQSportal, it will take you to your deployed solution in the [Azure management portal](https://portal.azure.com/). Once there, click on **Tables**. In the next panel, check if you see the two tables **"asaEgressPartitions" and “asaEgress”**. You can also check the if data is being populated in the tables using tools like [Azure Storage Explore](http://storageexplorer.com/)r. If you see these tables and data, it indicates that the raw data is successfully being generated on your computer and stored in table storage.

You can also check the data being populated in SQL database by going to your Resource Group, and locating your database (ex: demo123456db ) and connecting to it using the SQL server username and password provided in the CIQS portal. Once connected, you can check “**AdditionalInfo**” table in the database and make sure it has records being populated by the Stream Analytics Job (e.g. “select count(\*) from AdditionalInfo”).

1. Check the results output data from Azure SQL Database:

The last step of the pipeline is to write data (e.g. anomalies detected using Machine Learning API) into SQL Database table named “**AdScoreResults**”. You might have to wait for a minimum of 15 minutes for the output data to appear in table. Here, you can query for the number of rows (e.g. "select count(\*) from AdScoreResults"). As your database grows, the number of rows in the table should increase.

1. Check Azure Data Factory dashboard

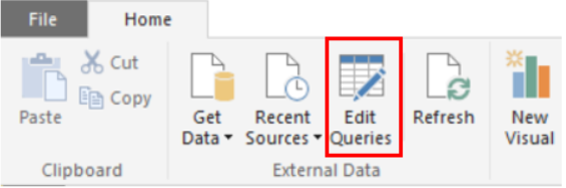
The [Azure Data Factory](https://azure.microsoft.com/documentation/services/data-factory/) service orchestrates the movement and processing of data. You can access your data factory from the Azure management portal resource group (e.x. demo12345adf) . If you see errors under your datasets, you can ignore those as they are due to data factory being deployed before the data generator was started. Those errors do not prevent your data factory from functioning. Read more about monitoring and managing the ADF pipeline [here](https://azure.microsoft.com/en-us/documentation/articles/data-factory-monitor-manage-pipelines/).

## ****Power BI Dashboard****

### **Overview**

This section describes how to set up Power BI dashboard to visualize the output results of the pipeline. Power BI connects to an Azure SQL database as its data source, where the Machine Learning score results are stored. Below are the steps to setup the Power BI dashboard.

1. Get the database server name, database name, user name and password from the CIQS portal.
2. Update the data source of the Power BI file
   * Make sure you have installed the latest version of [Power BI desktop](https://powerbi.microsoft.com/desktop).
   * Download the Power BI desktop file for the solution from [here](https://microsoft.sharepoint.com/teams/CECloudML/_layouts/15/guestaccess.aspx?guestaccesstoken=529Gl6mqa816nsFuDB6AwD3Asek%2bIj%2bntkRiGuFnOLU%3d&docid=2_0692940cd577543ab860eba90c19425d7&rev=1). The initial visualizations are based on dummy data. **Note:** If you see an error massage, please make sure you have installed the latest version of Power BI Desktop.

Once you open it, on the top of the file, click **‘Edit Queries’**. In the pop out window, double click **‘Source’** on the right panel. 

* + In the pop out window, replace **"Server"** and **"Database"** with your own server and database names, and then click **"OK"**. For server name, make sure you specify the port 1433 (**YourSolutionName.database.windows.net, 1433**). Ignore the warning messages that appear on the screen.
  + In the next pop out window, you'll see two options on the left pane (**Windows** and **Database**). Click **"Database"**, fill in your**"Username"** and **"Password"** (this is the username and password you entered when you first deployed the solution and created an Azure SQL database). In ***Select which level to apply these settings to***, check database level option. Then click **"Connect"**.
  + Once you are guided back to the previous page, close the window. A message will pop out - click **Apply**. Lastly, click the **Save** button to save the changes. Your Power BI file has now established connection to the server. If your visualizations are empty, make sure you clear the selections on the visualizations to visualize all the data by clicking the eraser icon on the upper right corner of the legends. Use the refresh button to reflect new data on the visualizations. Initially, you will only see the seed data on your visualizations as the data factory is scheduled to refresh every 3 hours. After 3 hours, you will see new scores reflected in your visualizations when you refresh the data.

1. (Optional) Publish the dashboard to [Power BI online](http://www.powerbi.com/). Note that this step needs a Power BI account (or Office 365 account).
   * Click **"Publish"** and few seconds later a window appears displaying "Publishing to Power BI Success!" with a green check mark. To find detailed instructions, see [Publish from Power BI Desktop](https://support.powerbi.com/knowledgebase/articles/461278-publish-from-power-bi-desktop).
   * To create a new dashboard: click the **+** sign next to the **Dashboards** section on the left pane. Enter the name "Demand Forecasting Demo" for this new dashboard.
2. (Optional) Schedule refresh of the data source.
   * To schedule refresh of the data, hover your mouse over the dataset, click https://acom.azurecomcdn.net/80C57D/cdn/mediahandler/docarticles/dpsmedia-prod/azure.microsoft.com/en-us/documentation/articles/cortana-analytics-technical-guide-demand-forecast/20160816064407/powerbipic3.png and then choose **Schedule Refresh**. **Note:** If you see a warning massage, click **Edit Credentials** and make sure your database credentials are the same as those described in step 1.

 - Expand the **Schedule Refresh** section. Turn on "keep your data up-to-date". - Schedule the refresh based on your needs. To find more information, see [Data refresh in Power BI](https://powerbi.microsoft.com/documentation/powerbi-refresh-data/).