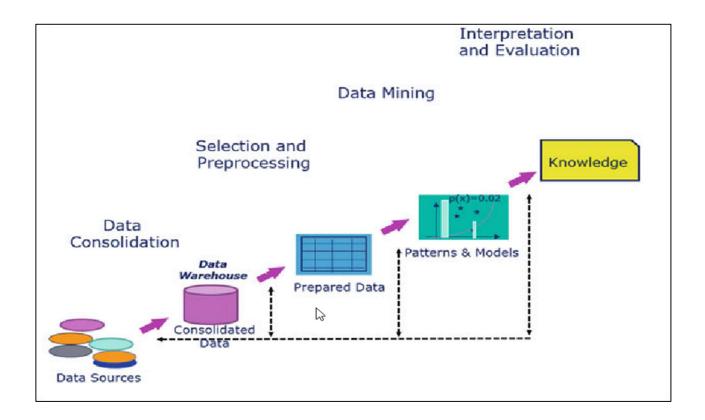
## Develop a Fully automated data analytics Platform with Azure

In this blog, I have created a Function App on Azure using Java, Data Bricks workspace and devised a way to "Automate the data analytics platform" with the following Azure Services: IoTHub, EventHub, Function App, Azure Storage, Azure Data Bricks Azure Cosmos DB, and IAM. We tried multiple use cases across domains Industry 4.0, Manufacturing, Factory of Future and Renewable Energy. By following this blog, you can seamlessly automate your data analytics platform use cases within azure.

## Problem statement

All businesses (manufacturing etc.), whether they are growing or established need a data analytics strategy. We tried to address the below problem statements through azure PAAS services.

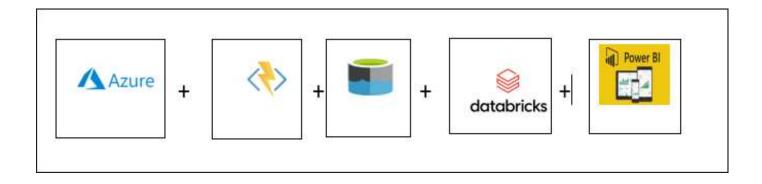
- 1. The IoT devices were getting disconnected to due to component failures.
- 2. The Data Aggregation was to complex using Azure functions.
- 3. There is no near real time visualization dashboard for monitoring the device metrics.
- 4. Unable to derive meaningful insights from the telemetry data.

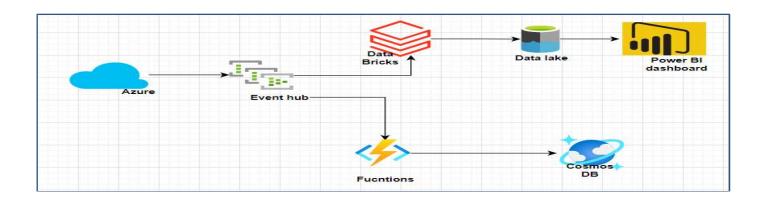


However, data analytics is complex due to multiple reasons:

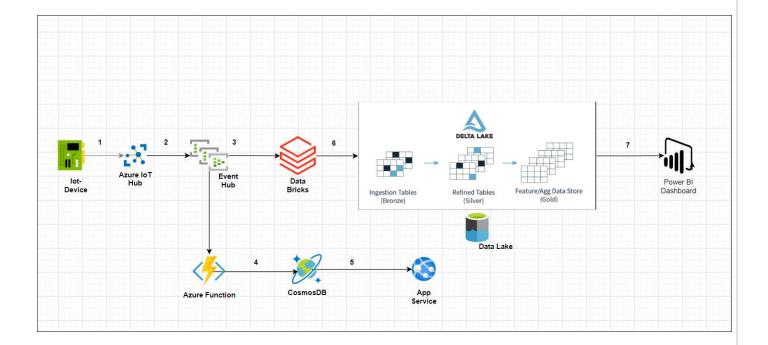
- · The size (millions of messages) of the data received form data sources.
- The frequency at which the data is being collected and how it's sent to cloud.
- · Streaming and processing the raw data before storing it in data lake.
- To get meaning full insights and visualizing the same.

# **Solution Block:**





# Architecture Diagram:



- 1. **IoT device** Agent has been developed using Azure IoT SDK, the agent collects the telemetry data and sends it to IoTHub (blogathon-iothub).
- 2. **Messaging Route** is in place for routing the message from IoTHub(blogathon-iothub) to EventHub(azureblogathon-telemetry).
- 3. Data Bricks service consumes the messages from the EventHub through respective consumer group.
- 4. Azure function is configured with EventHub Trigger, the function will write the telemetry data to the cosmos DB.
- 5. End customers can view the devices status details through webApp.
- 6. Using **Delta Lake** and python notebook the telemetry data is processed and stored in respective tables (Bronze, Silver, Gold).
- 7. Near real-time PowerBI dashboard has been integrated with Databricks for visualization.

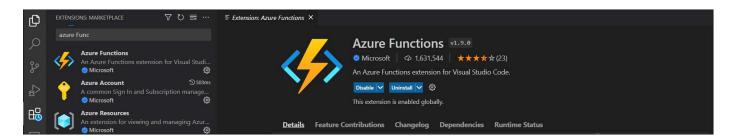
**Note:** As described in the introduction, the below steps can be used to replicate the whole data analytics process based on your requirement.

### The implementation of data analytics has been divided in 5 steps:

- · Deploying the required resources.
- Deploying the code to Azure Functions.
- Establishing Function App connection with the Cosmos DB.
- · Configuring Data Bricks cluster.
- Testing the workflow by simulating the telemetry data and visualizing the data in Power BI.

### **Pre-requisites:**

- An Azure account with an active subscription. Create an account for free.
- The Azure Functions Core Tools version 3.x.
- Java versions that are supported by Azure Functions.
- Visual Studio Code on one of the supported platforms.
- The Java extension for Visual Studio Code.
- The Azure Functions extension for Visual Studio Code.



Azure Functions VS Code extension

## Step 1: Deploying the required resources

1.1 IoThuB: - Creating IotHub from Azure portal, below are the configurations.

On the Azure homepage, select the + Create a resource button.

On the Basics tab, complete the fields as follows:

Subscription: Select the subscription to use for your hub (Ex: - Visual studio enterprise subscription).

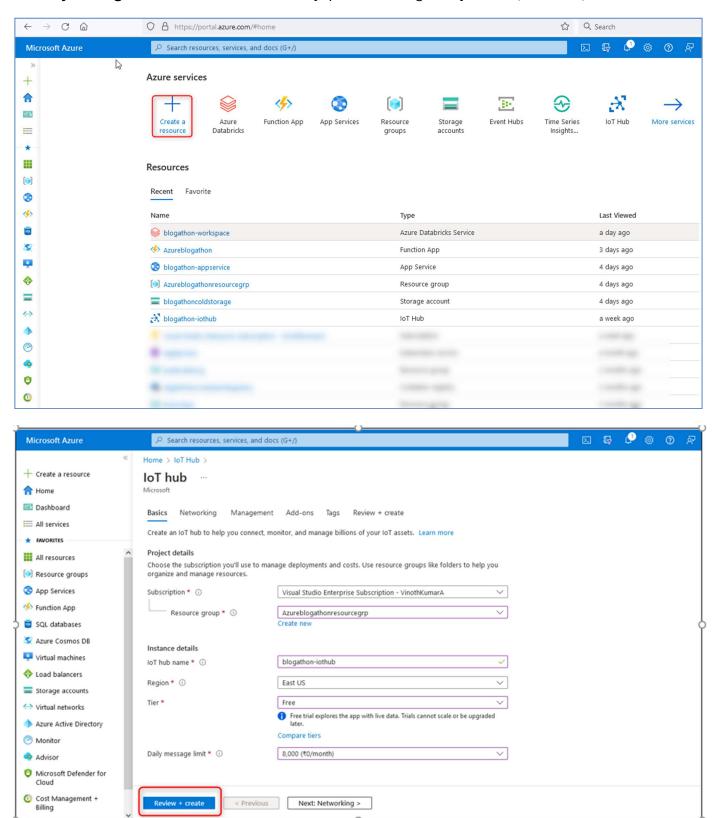
**Resource group:** Select a resource group or create a new one (Ex: - Azureblogathonresourcegrp).

**IoT hub name:** Enter a name for your hub (Ex: - blogathon-iothub).

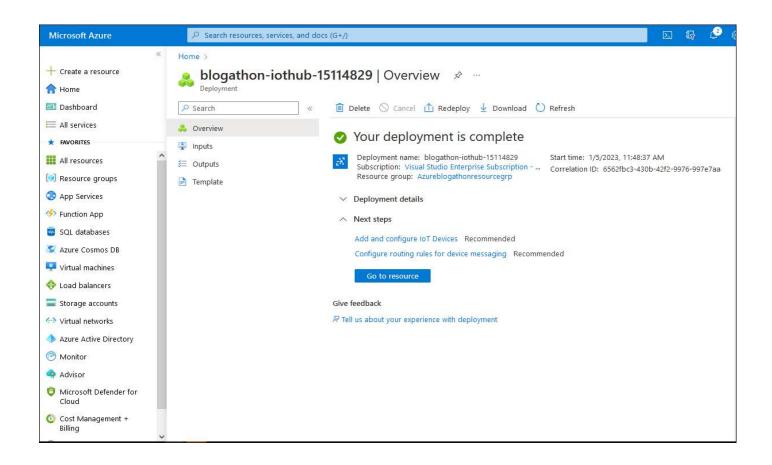
Region: Select the region, closest to you (Ex: - East US).

**Tier:** Select the tier that you want to use for your hub (**EX: -Free**).

Daily message limit: Select the maximum daily quota of messages for your hub (Ex: - 8000).



Once the deployment is successful, you will see this screen.



**1.2 EventHub:** - Creating event hub from Azure portal, below are the configurations.

On the Azure homepage, select the + Create a resource button.

On the Basics tab, complete the fields as follows:

Subscription: Select the subscription to use for your hub (Ex: - Visual studio enterprise subscription).

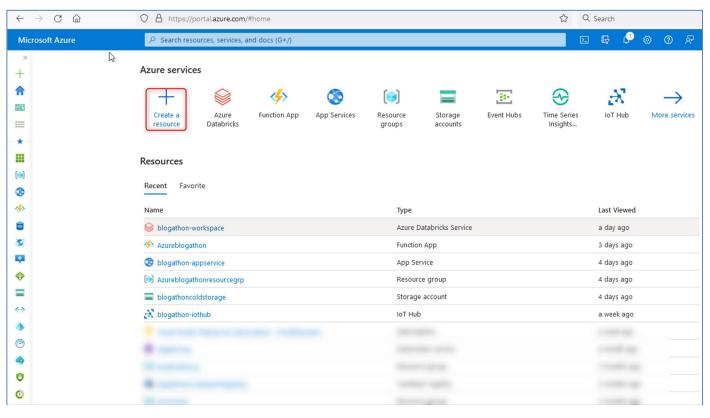
**Resource group:** Select a resource group or create a new one (Ex: - Azureblogathonresourcegrp).

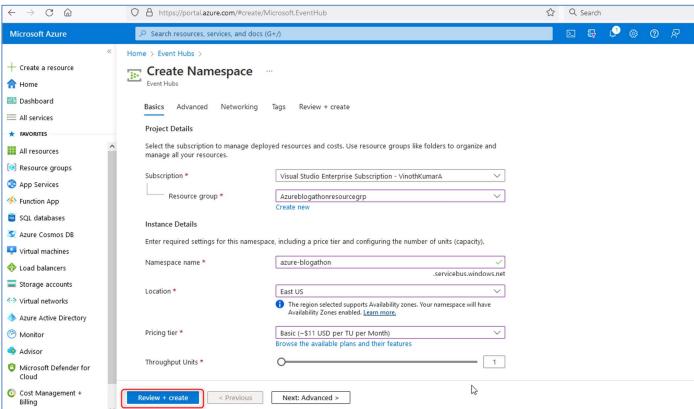
Namespace name: Enter a name for your Namespace (Ex: - azure-obligation)

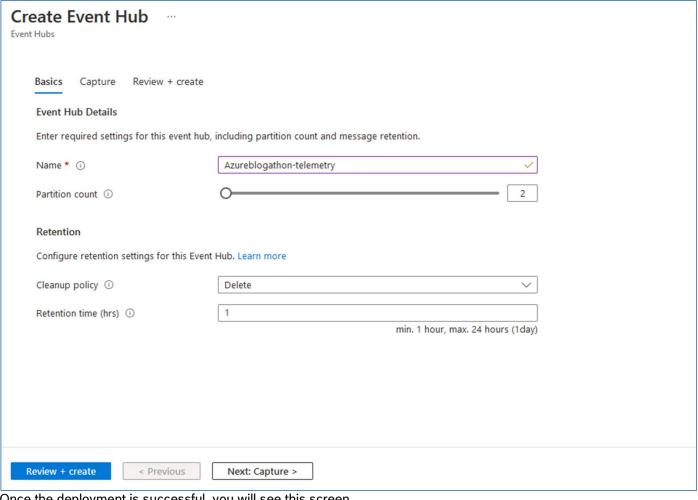
**Location:** Select the region, closest to you (Ex: - East US).

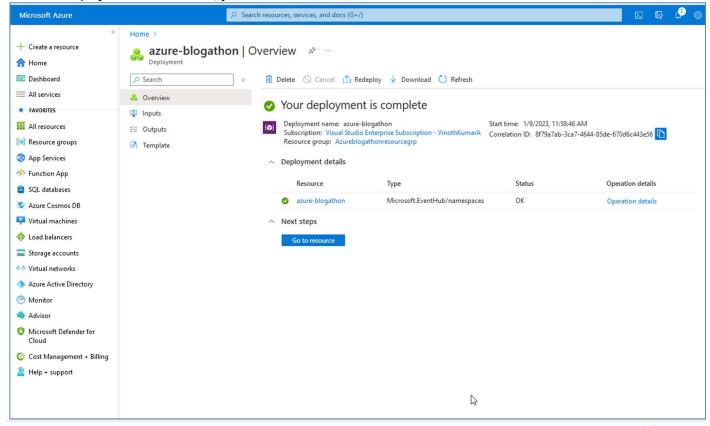
**Pricing tier:** Select a pricing tier (Ex: - Basic)

Throughput units: Select number of throughput units (Ex: - 1)









1.3 Function App: - Creating Function App from Azure portal, below are the configurations.

On the Azure homepage, select the + Create a resource button.

On the Basics tab, complete the fields as follows:

Subscription: Select the subscription to use for your hub (Ex: - Visual studio enterprise subscription).

**Resource group:** Select a resource group or create a new one (Ex: - Azureblogathonresourcegrp).

FunctionApp name: Enter a name for your FcuntionApp (Ex: - Azureblogathon)

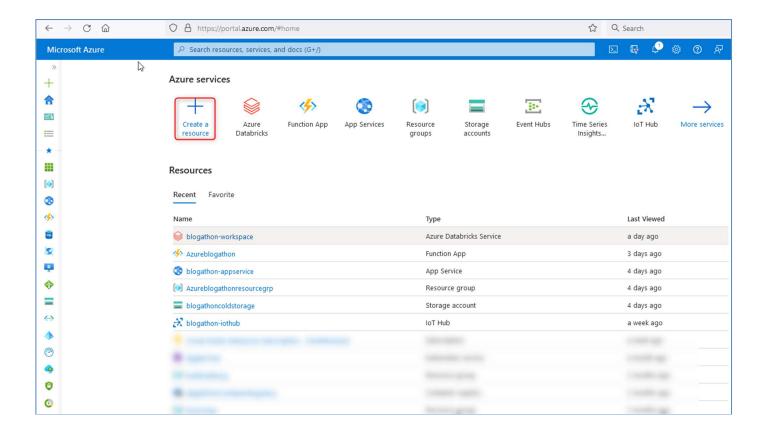
**Publish:** Option to publish code files or a Docker container (**Ex: - code**).

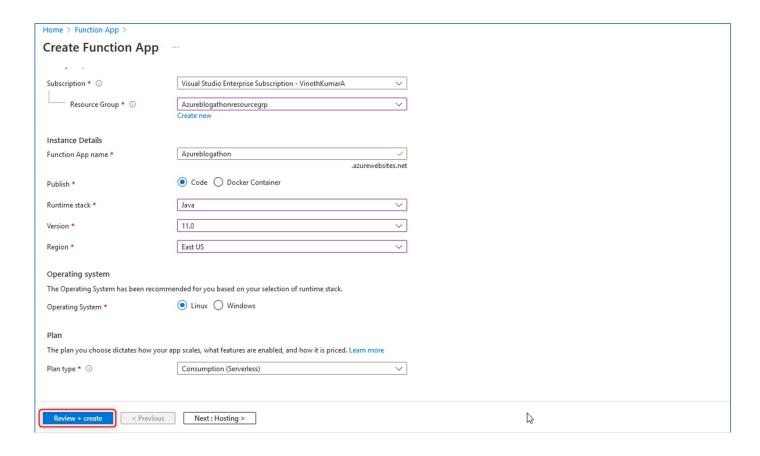
Runtime stack: Choose a runtime that supports your favorite function programming language (Ex: - Java).

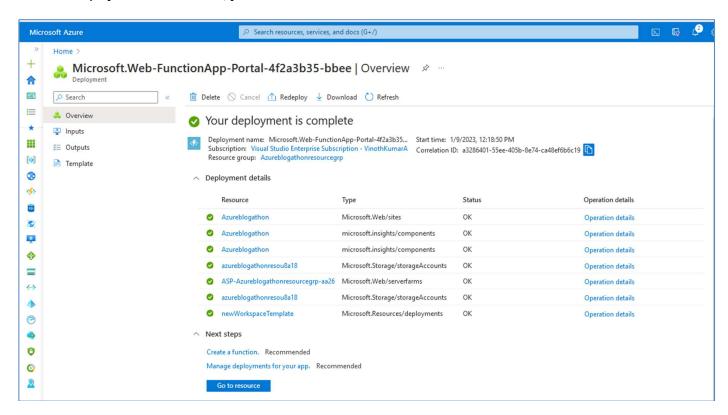
**Version:** Choose the version of your installed runtime Ex: - 11).

**Location:** Select the region, closest to you (Ex: - East US).

Plan: Choose a plan type (Ex: - Serverless).







**1.4 Cosmos DB: -** Creating Cosmos DB from Azure portal, below are the configurations.

On the Azure homepage, select the + Create a resource button.

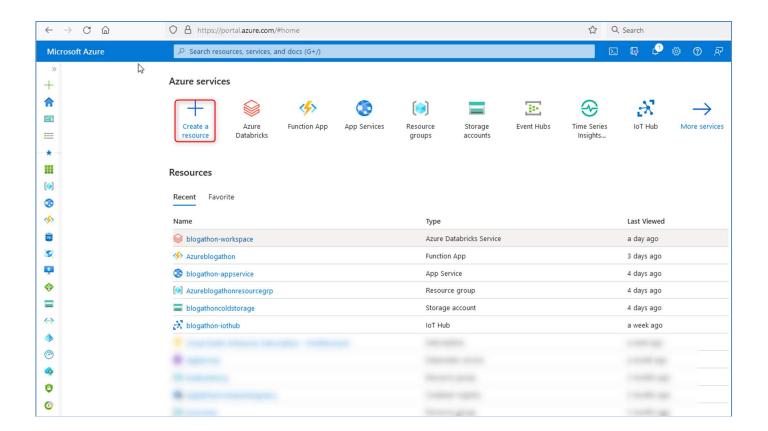
On the Basics tab, complete the fields as follows:

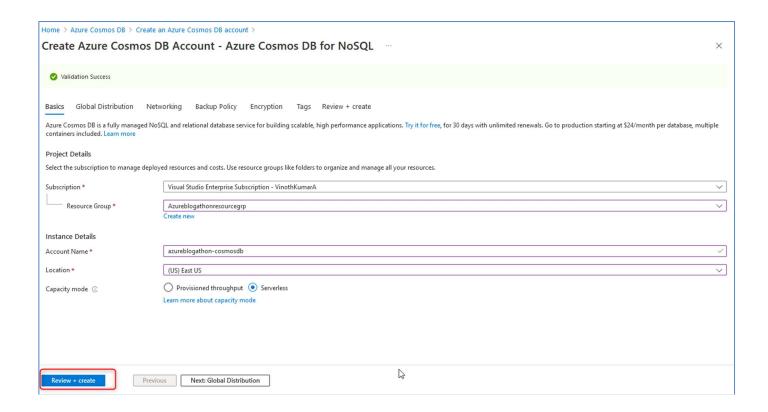
Subscription: Select the subscription to use for your hub (Ex: - Visual studio enterprise subscription).

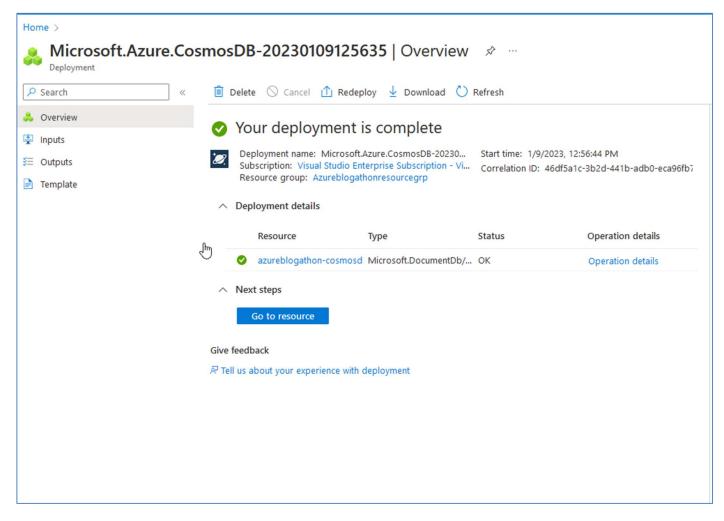
Account Name: Enter a name for your Cosmos DB(Ex:- azureblogathon-cosmosdb).

Region: Select the region, closest to you (Ex: - East US).

Capacity Mode: Choose a capacity mode (Ex: - Serverless).







**1.5 Storage account: -** Creating Storage account from Azure portal, below are the configurations.

On the Azure homepage, select the + Create a resource button.

On the Basics tab, complete the fields as follows:

Subscription: Select the subscription to use for your hub (Ex: - Visual studio enterprise subscription).

Resource group: Select a resource group or create a new one (Ex: - Azureblogathonresourcegrp).

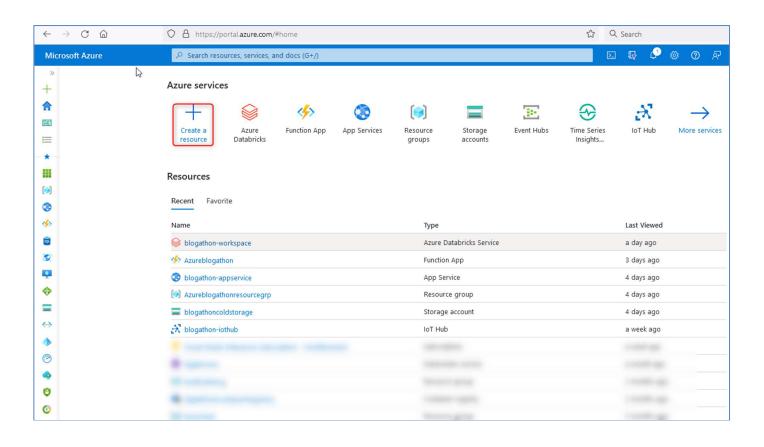
Storage account name: Enter a name for your Storage account (Ex:- blogathoncoldstorage)

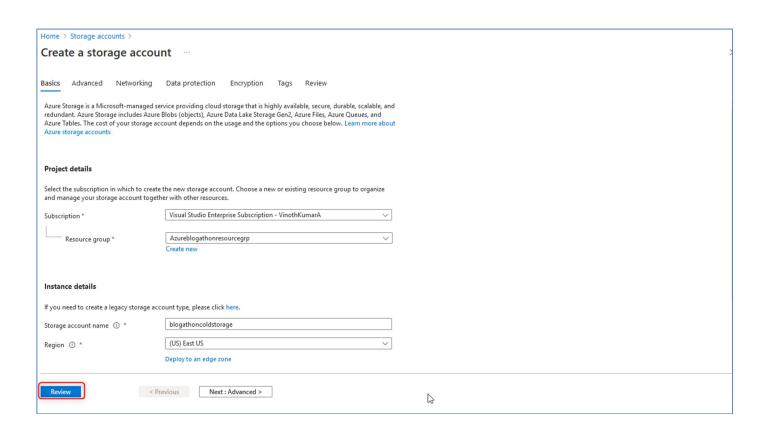
**Region:** Select the region, closest to you (Ex: - East US).

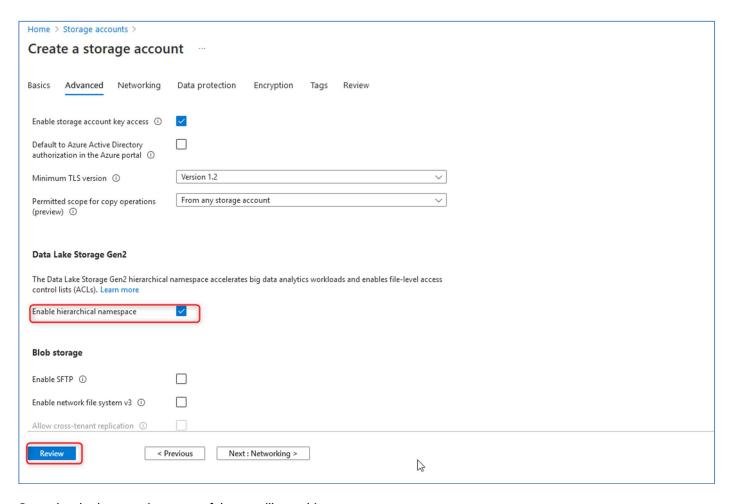
On the Advanced tab, complete the fields as follows:

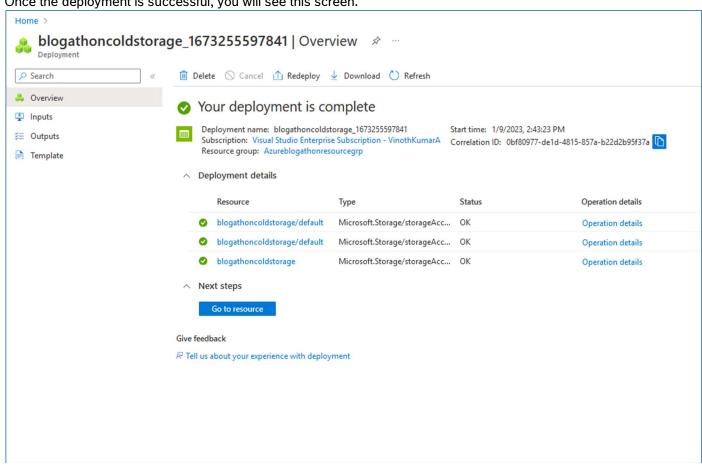
#### Data Lake Gen2:

Choose the option to **Enable Hierarchical namespace.** 









**1.6 Data Bricks workspace: -** Creating workspace from Azure portal, below are the configurations.

On the Azure homepage, select the + Create a resource button.

On the Basics tab, complete the fields as follows:

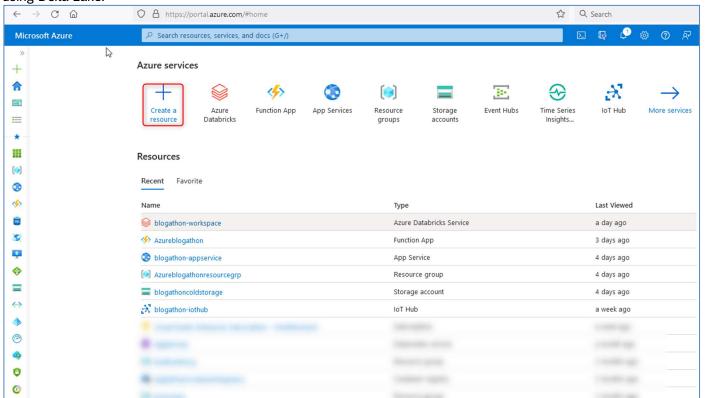
Subscription: Select the subscription to use for your hub (Ex: - Visual studio enterprise subscription).

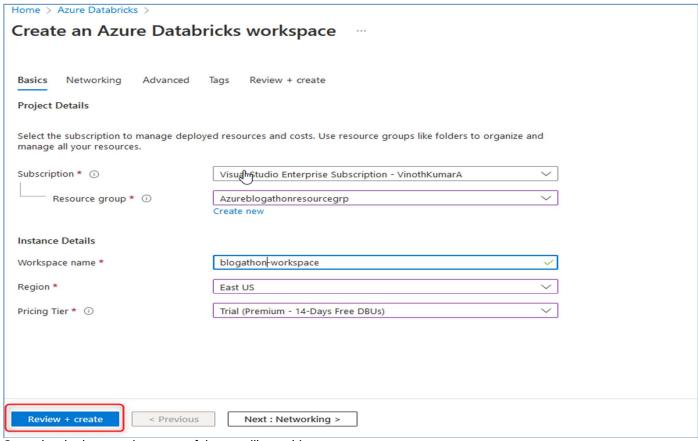
**Resource group:** Select a resource group or create a new one (Ex: - Azureblogathonresourcegrp).

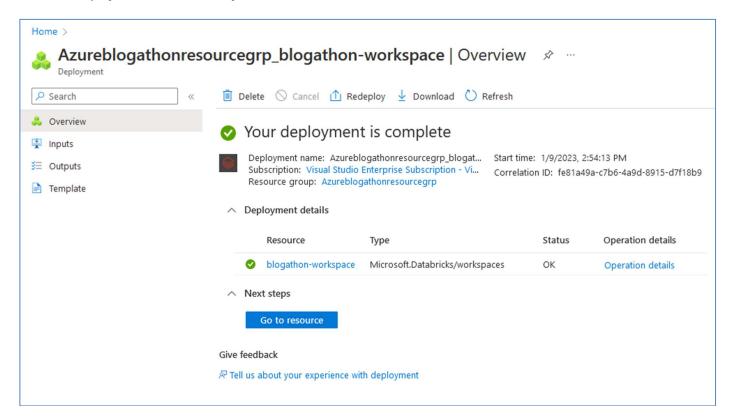
Workspace name: Enter a workspace name(Ex:- blogathon-workspace).

Region: Select the region, closest to you (Ex: - East US).

You need to have access to an Azure Databricks workspace to perform Structured Streaming with batch jobs by using Delta Lake.







1.7 Web App: - Creating Webapp service from Azure portal, below are the configurations.

On the Azure homepage, select the + Create a resource button.

On the Basics tab, complete the fields as follows:

Subscription: Select the subscription to use for your hub (Ex: - Visual studio enterprise subscription).

**Resource group:** Select a resource group or create a new one (Ex: - Azureblogathonresourcegrp).

WebApp name: Enter a name for your WebApp name (Ex: - blogathon-Appservice)

Publish: Option to publish code files or a Docker container (Ex: - code).

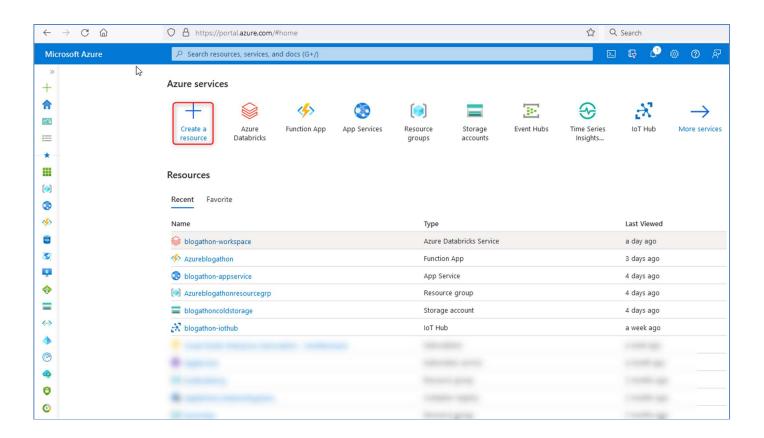
Runtime stack: Choose a runtime that supports your favorite function programming language (Ex: - Java 11).

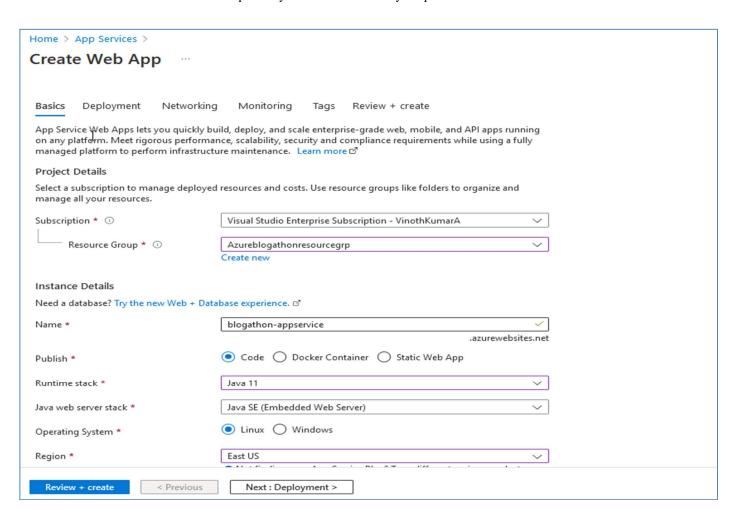
Java web server stack: Choose a java web server stack (Ex: - Java SE (Embedded web server)).

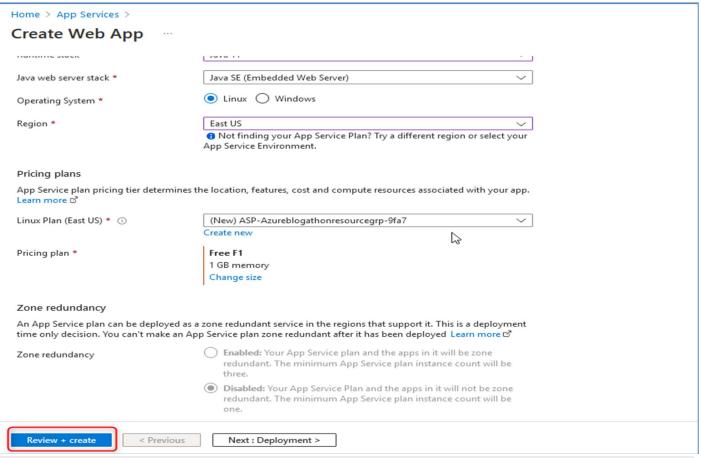
Operating system: Choose an operating system (Ex: - Linux).

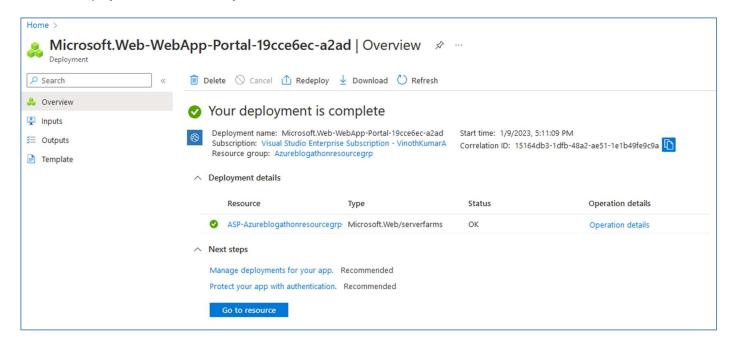
Region: Select the region, closest to you (Ex: - East US).

Pricing Plan: Choose a plan type (Ex: - Free F1).









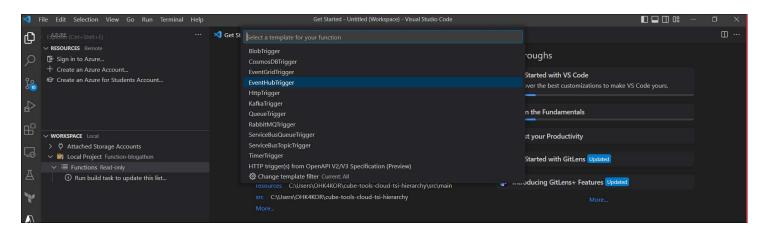
## Step 2: Deploying the code to Azure Functions

```
@CosmosDBOutput(
    name = "databaseOutput",
    databaseName = "Cosmos DB name",
    collectionName = "Cosmos DB collection name",
    connectionStringSetting = "CosmosDBConnectionString")
    OutputBinding<String> document,
    final ExecutionContext context
) {
    context.getLogger().info("Java Event Hub trigger function executed.");
    context.getLogger().info(message);
}
```

Note:- Please go through reference (1. **Deploying Azure Functions**) for more details.

#### Creating the function locally

Click on you workspace's "+" icon to create a ducntion with Azure Event hub trigger



Connect your VS code to Azure by signing in.

```
Get Started
                                                                          J EventHubTriggerJava1.java 1, U X
     ∨ RESOURCES Remote
                                         + [≈ ひ ∄
                                                        Function-blogathon > src > main > java > com > function > J EventHubTriggerJava1.java > {} com.function
     F Sign in to Azure...
      + Create an Azure Account...
                                        Sign in to Azure...

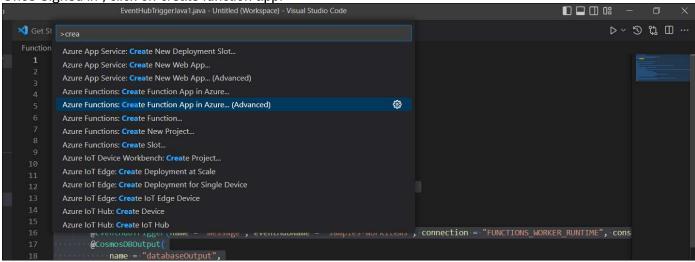
∨ WORKSPACE | local
                                         4000
      > 🛱 Attached Storage Accounts
                                                                     * This function will be invoked when an event is received from Event Hub.
     V 📑 Local Project Function-blogathon

✓ 

Functions Read-only

                                                                    @FunctionName("EventHubTriggerJava1")
           (i) Run build task to update this list...
                                                                         @EventHubTrigger(name = "message", eventHubName = "samples-workitems",
Y
                                                                             name = "databaseOutput",
                                                                             databaseName = "TelemetryDb",
collectionName = "TelemetryInfo",
                                                                             connectionStringSetting = "CosmosDBConnectionString")
                                                                         final ExecutionContext context
```

Once Signed in , click on create function app.



You will need to fill in the following details:

Subscription: Select the subscription to use for your hub(Ex:- Visual studio enterprise subscription).

**Resource group:** Select a resource group or create a new one( Ex:- Azureblogathonresourcegrp).

FuctionApp name: Enter a name for your FcuntionApp (Ex:- Azureblogathon)

Publish: : Option to publish code files or a Docker container(Ex:-code).

**Runtime stack:** Choose a runtime that supports your favorite function programming language(**Ex:- Java**).

**Version:** Choose the version of your installed runtime **Ex:-11**).

**Location:** Select the region, closest to you(**Ex:- East US**).

**Plan:** Choose a plan type(Ex:- Serverless).

Finally, deploy the local function to azure.

```
နှ

∨ WORKSPACE Local

                                        4000
      Deploy...
G
                                                                  * This function will be invoked when an event is received from Event Hub.
      V 📑 Local Project Function-blogathon
           (i) Run build task to update this list...
                                                                    @EventHubTrigger(name = "message", eventHubName = "samples-workitems", connection =
Y
                                                                        name = "databaseOutput",
databaseName = "TelemetryDb",
collectionName = "TelemetryInfo",
                                                                         connectionStringSetting = "CosmosDBConnectionString")
OutputBinding<String> document,
context.getLogger().info(msg: "Java Event Hub trigger function executed.");
```

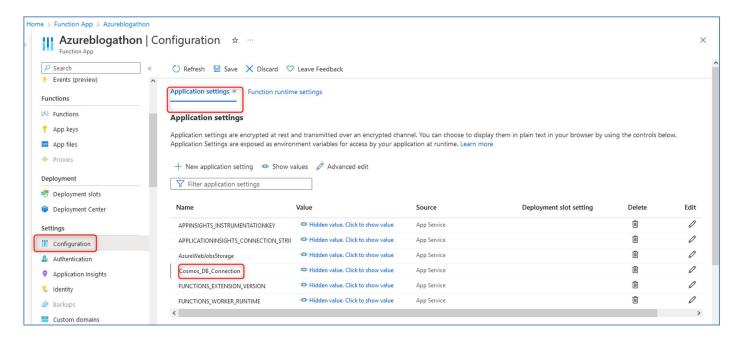
Step 3: Establishing Function App connection with the Cosmos DB

To enable the function app to connect with the Azure Cosmos DB

Once the database is created, create a container for saving the telemetry data.

All we need to do is update the database connection string in the Application setting of the function.

Since our function can connect with the Azure Cosmos DB, we can query it directly.



Step 4: Configuring Data Bricks cluster

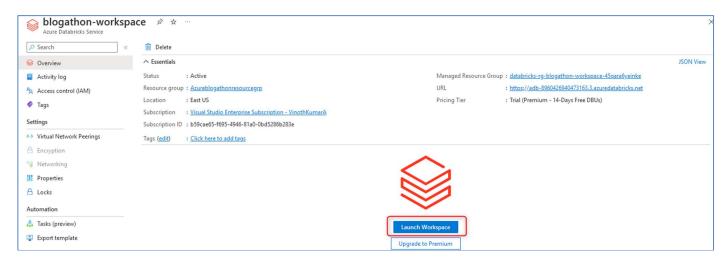
Perform Batch and Stream Processing with Delta Lake

You need to have access to an Azure Databricks workspace to perform Structured Streaming with batch jobs by using Delta Lake

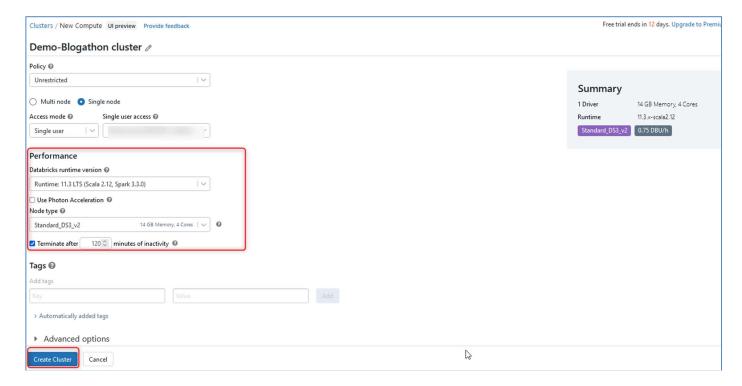
#### Create A Cluster

We have already deployed Databricks workspace in step1.

1) Click on the button Launch Workspace to open your Databricks workspace in a new tab.



- 2) In the left-hand menu of your Databricks workspace, select Clusters.
- 3) Select Create Cluster to add a new cluster.



4) Select Create Cluster.

### Perform Batch and Stream Processing

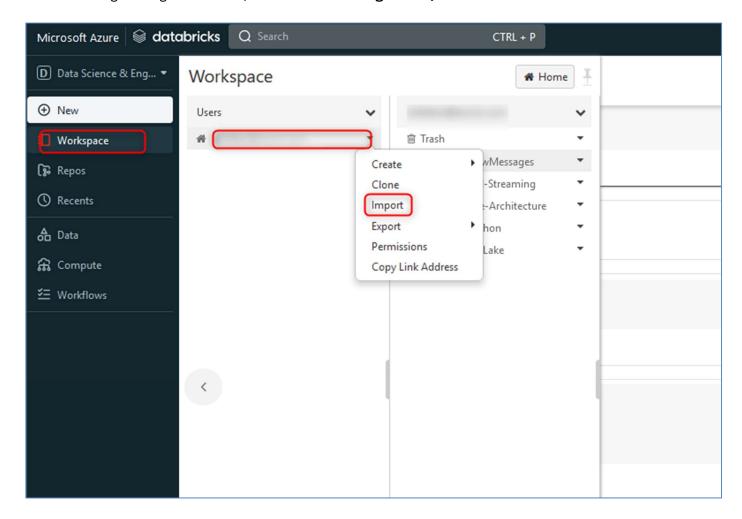
- 1) After the cluster is created, in the left pane, select **Workspace > Users**, and select your username (the entry with the house icon).
- 2) In the pane that appears, select the arrow next to your name, and select Import.

https://github.com/MicrosoftLearning/DP-203-Data-Engineer/raw/master/Allfiles/microsoft-learning-paths-databricks-notebooks/data-engineering/DBC/10-Structured-Streaming.dbc

https://github.com/solliancenet/microsoft-learning-paths-databricks-notebooks/blob/master/data-engineering/DBC/11-Delta-Lake-Architecture.dbc

The above generic notebooks can help you get started with data analytics; you can customize the notebook as per you need.

Note:- Please go through reference (2. Data Bricks configuration) for more details.



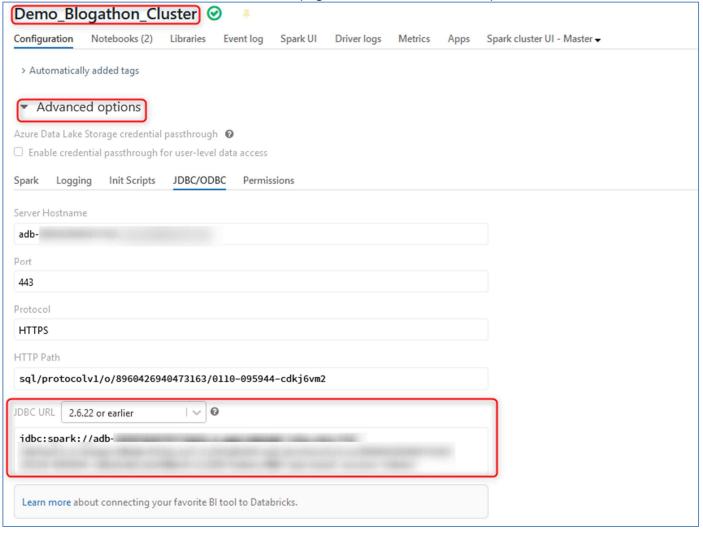
Step 5: Testing the workflow by simulating the telemetry data and visualizing the data in Power BI

Set up Power BI Databricks Integration

Step 1: Get Databricks Connection Information

The first step involved in Power BI Databricks Integration requires you to extract some details about the Databricks connection such as Server Name and HTTP Path. Follow the steps given below to do so:

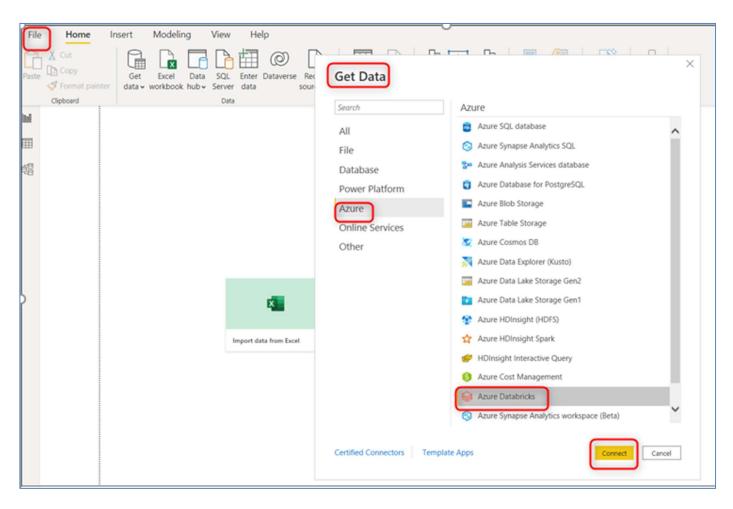
- Log in to the Microsoft Azure portal using the appropriate credentials.
- Navigate to the left menu bar and click on the Clusters option.
- Scroll down the Clusters page and click on Advanced Options.



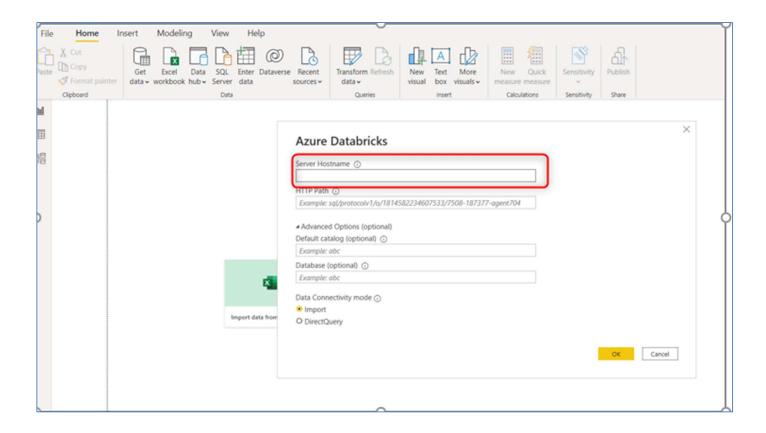
Step 2: Configure Azure Databricks Connection in Power BI

Once you have retrieved the connection information, you need to configure the Databricks connection in Power BI to set up Power BI Databricks Integration:

- Open Microsoft Power BI Desktop.
- Once you open Power BI Desktop, you will see a bunch of options on top of the home page. Click on the Get Data option.
- From the Get Data dialog box, select the Azure option and then click on Azure Databricks Connector to set up Power BI Databricks Integration.



Now, click on the **Connect** button and paste the **Server Name** and **HTTP Path** that you retrieved in the first step.



In terms of Data Connectivity Mode, you will have the following two options:

**Import**: If you select the Import option, Power BI Desktop will make use of the imported data whenever you create, design, or interact with a visualization. To see underlying data changes that have occurred since the initial import or the most recent refresh, you must refresh the data, which will re-import the entire dataset.

**Direct Query**: Selecting the Direct Query option will allow you to view the current data at any instant.

- At the Authentication Prompt, enter your Azure account credentials and click on the Connect button. If you want to authenticate using the token, you can click here to learn how to retrieve it and build the Power BI Databricks Integration.
- Finally, select the Databricks data that you wish to query from the Power BI Connector.

Once you follow the above steps in the correct sequence, you will be able to establish Power BI Databricks Integration and create visualizations.

#### Knowledge Sharing and Best Practices:

- ✓ Using system-assigned managed identity prevents the usage of keys within code which is a bad practice for production environments.
- ✓ The authorization level "anonymous" for a function app is not recommended for the production environment.

#### Challenges Faced:

- The automation was tricky to implement due to the multiple components involved.
   Integrating it with Azure Cosmos DB, creating an Event hub-triggered function using Java.
- Secondly, the data bricks notebook for Delta Lake architecture. was the most difficult part to implement.

#### **Business Benefits:**

- ✓ We have utilized 100% of Azure PaaS components in implementing solution to reduce the administrative (Scalability and Availability) overhead.
- ✓ Develop Once Deploy Multiple times on different IOT based data analytics use cases across domains
- ✓ Out of the box Azure Data Bricks services helped in implementing Data Analytics solution with ease and simplified the Data Aggregation process.
- ✓ The insights from the solution will help us in identifying real time faults in devices.
- ✓ The visualization dashboard will help in performing predictive maintenance and detailed service level reporting for the end customers.

#### References:

#### 1. Deploying Azure Functions:

https://learn.microsoft.com/en-us/azure/azure-functions/create-first-function-vs-code-java

2. Data Bricks configuration: -

https://microsoftlearning.github.io/DP-203-Data-Engineer/Instructions/Labs/LAB 11 stream with azu re\_databricks.html

- 2.1 <a href="https://learn.microsoft.com/en-us/azure/architecture/reference-architectures/data/stream-processing-databricks">https://learn.microsoft.com/en-us/azure/architecture/reference-architectures/data/stream-processing-databricks</a>
- 3. **Connecting to Power BI**: <a href="https://learn.microsoft.com/en-us/azure/databricks/partners/bi/power-bi">https://learn.microsoft.com/en-us/azure/databricks/partners/bi/power-bi</a>