

Source: C# Corner (www.c-sharpcorner.com)

PRINT

Article



Azure Data Factory - Creating Input And Output Datasets

By [Abdul Rasheed Feroz Khan](#) on Dec 25 2016

Introduction

This article will help you to create an input dataset and an output dataset in Azure Data Factory.

Note

Work on my previous article before you step into this, as we need a Data Factory account, storage account, and Azure HDInsight Cluster to create a dataset, and also for the future purpose of creating a pipeline and to monitor it. Surf the link given below to create it.

- [Creating Linked Services In Azure Data Factory](#)

Follow the below steps now.

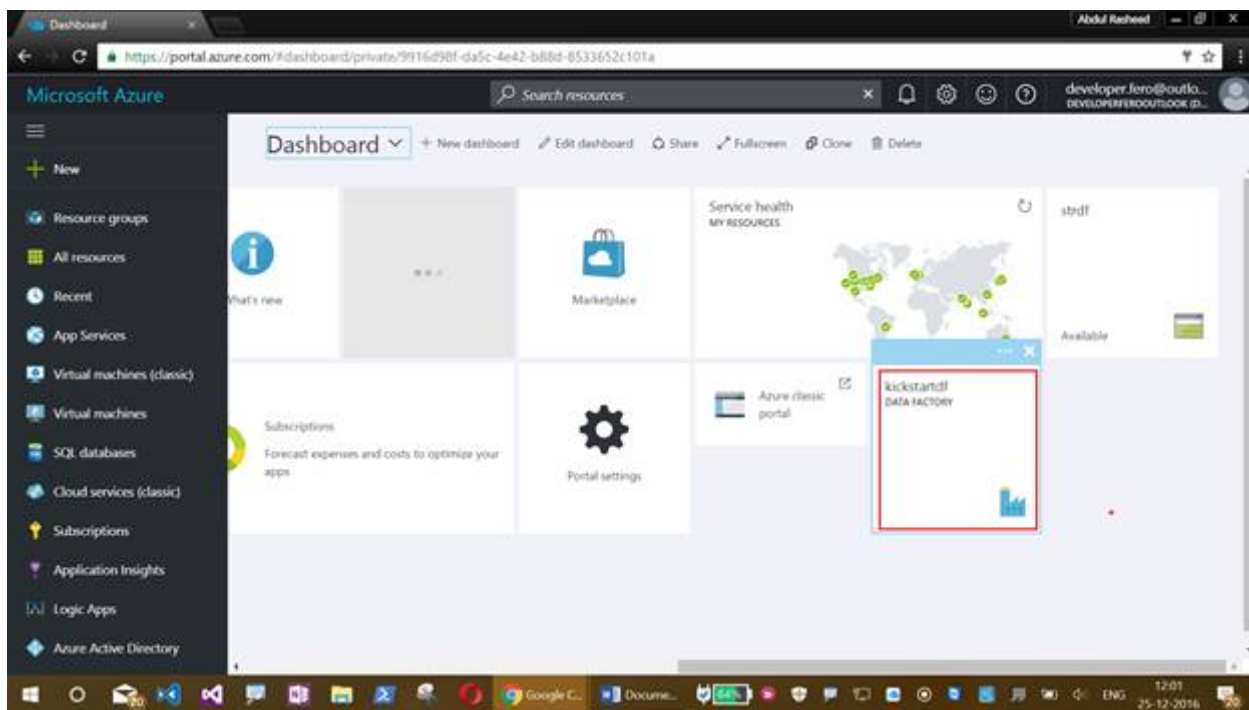
Step 1

Here, we will be creating a dataset to give the input and output dataset for Hive processing.

Login to Azure portal. Move to Azure Data Factory account. Click on the Data Factory editor.

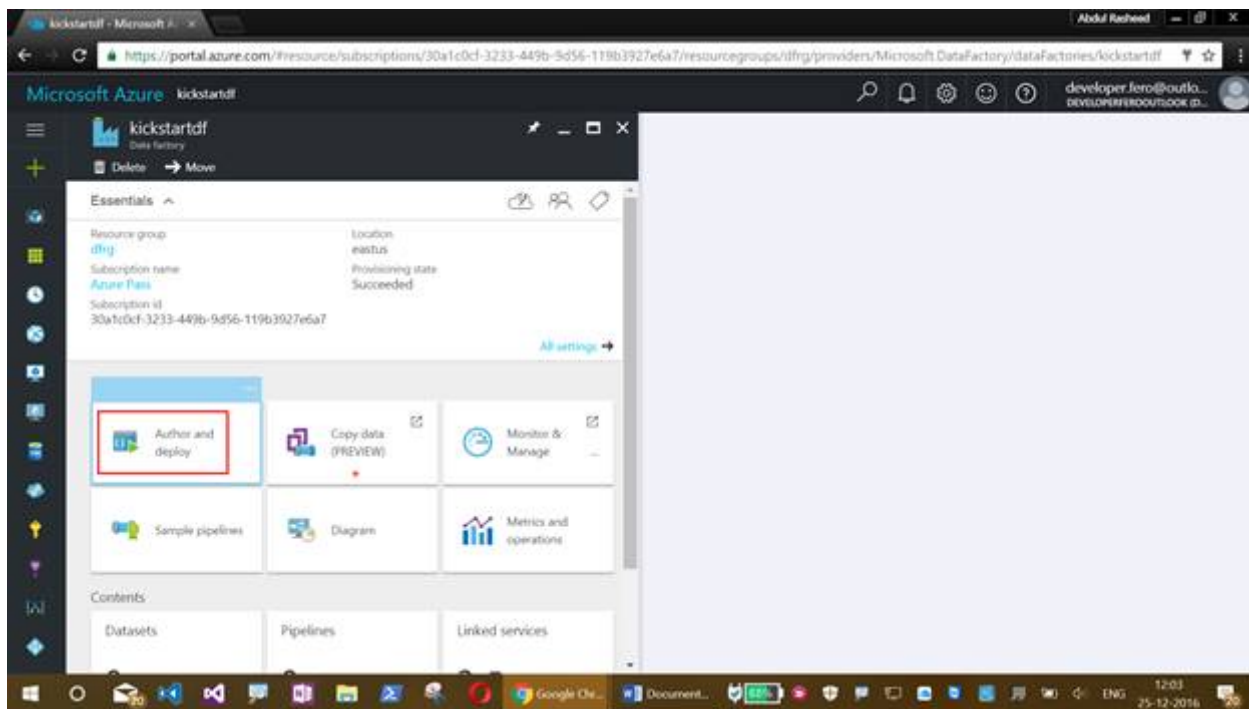
Note

This data account is where we have the Storage account configured, Azure Linked Services and an Azure HDInsight Cluster.



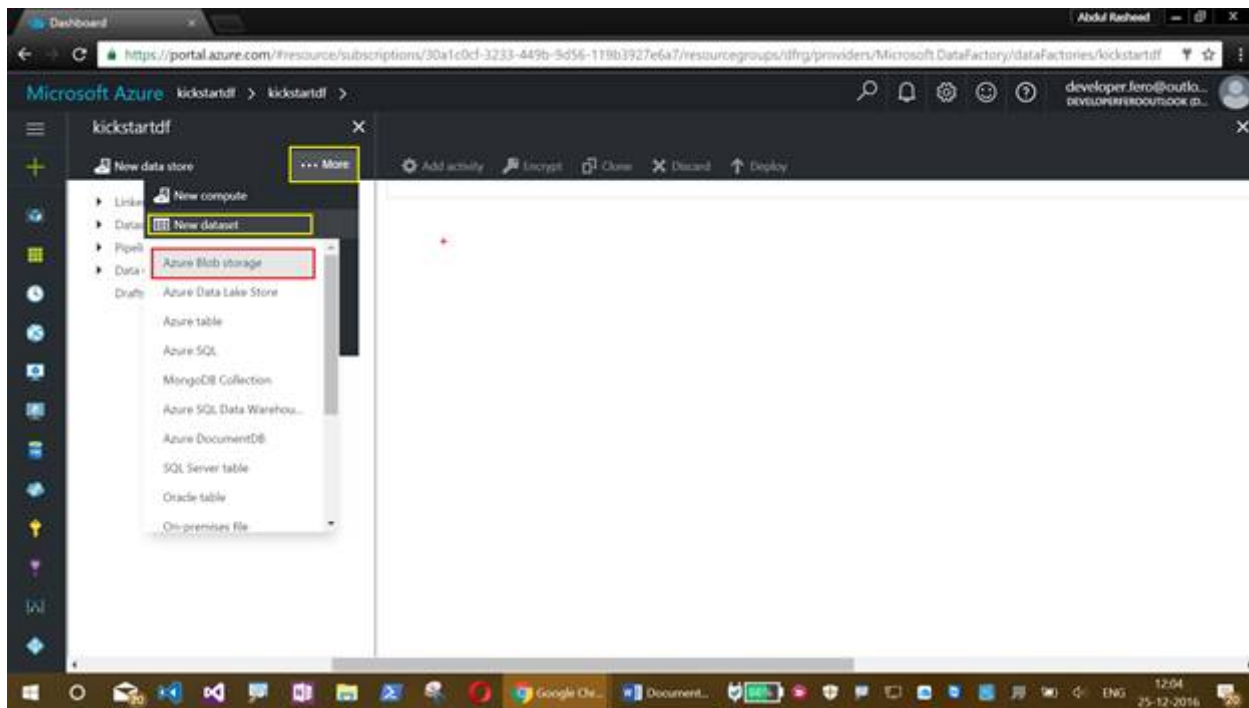
Step 2

Click on "Author and deploy".

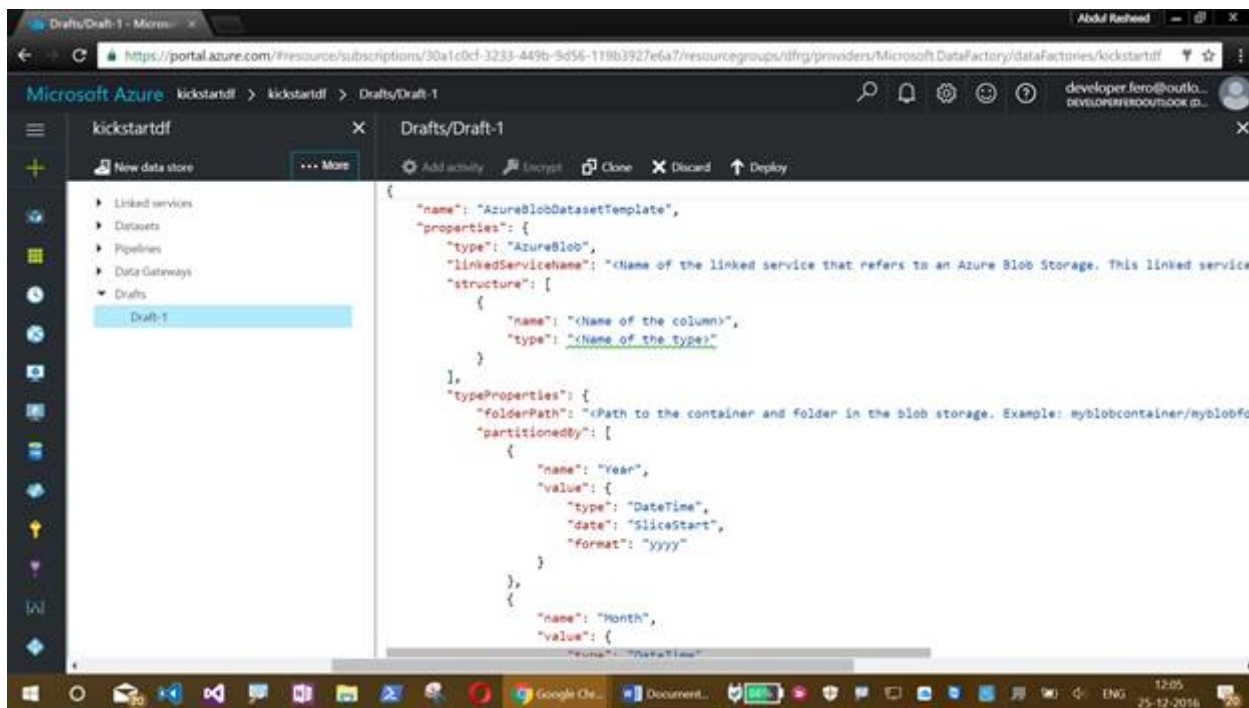


Step 3

In the New Data Store blade, click on More - New Dataset - Azure Blob Storage.



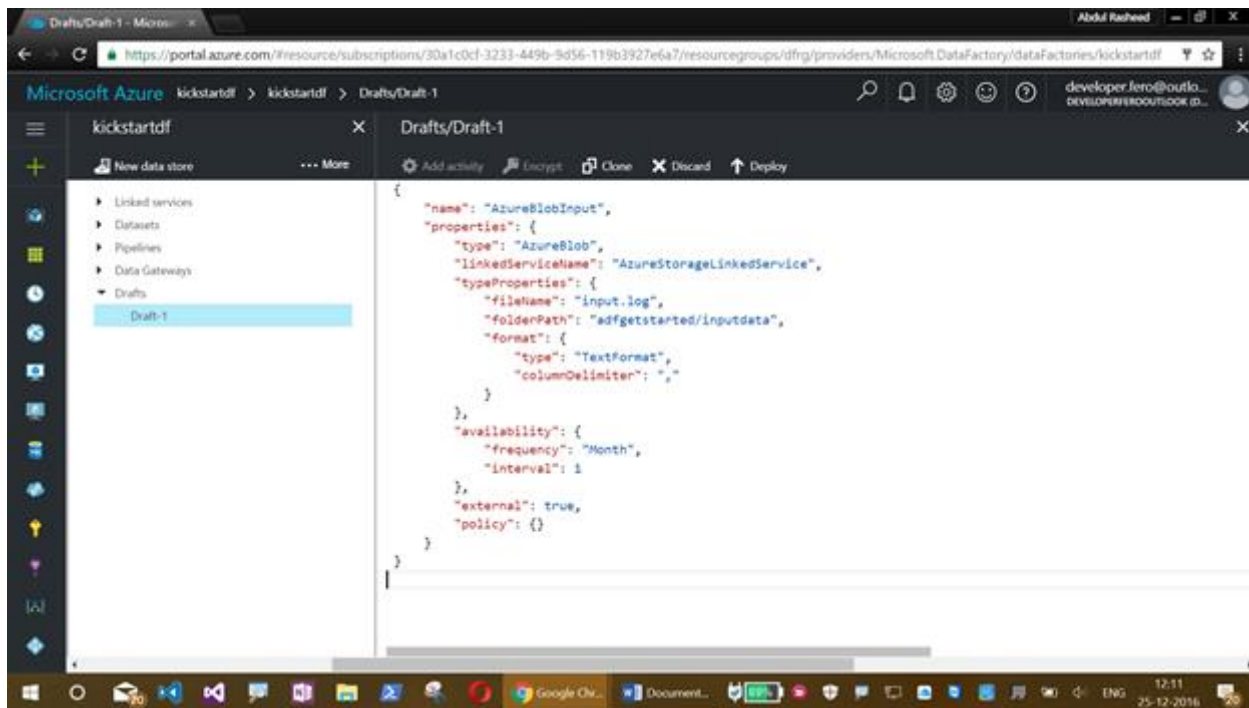
We will be getting a blade with the code snippets, as shown below.



Step 4

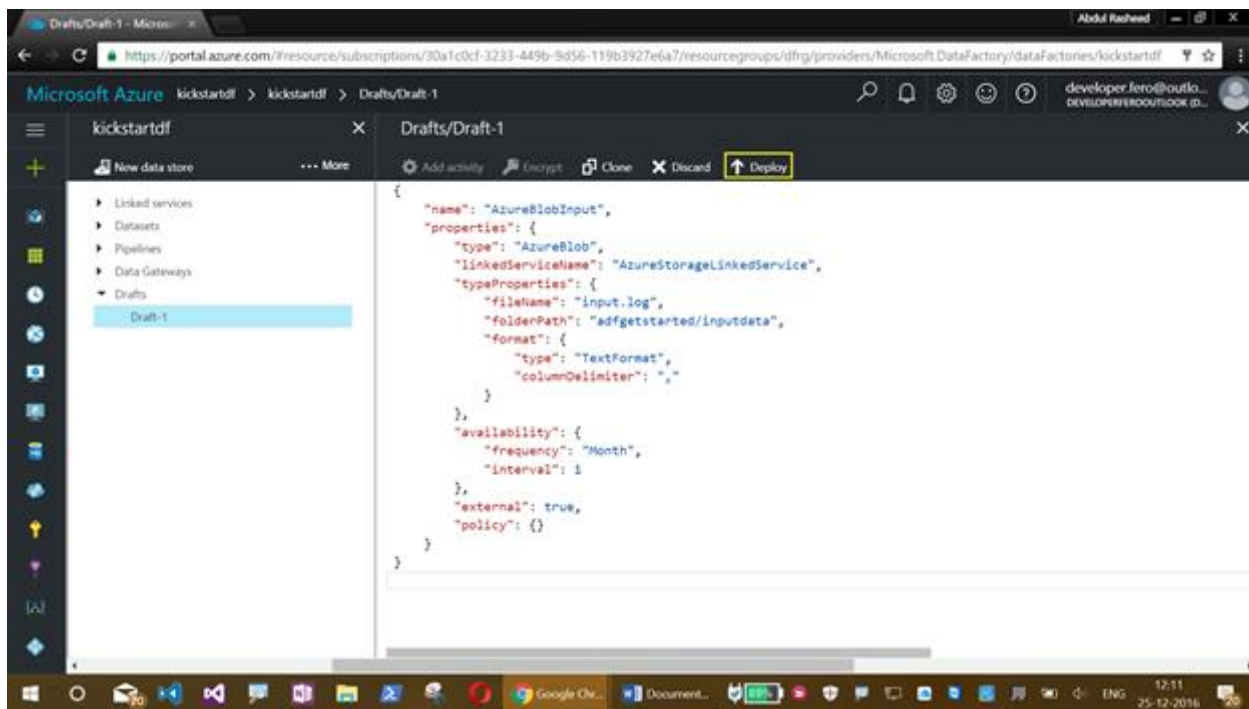
Copy the coding given below and paste it in the coding editor pane.

```
{  
  "name":  
  "AzureBlobInput",  
  
  "properties":  
  : {  
    "type":  
    "AzureBlob",  
  
    "linkedServiceName":  
    "AzureStorageLinkedService",  
  
    "typeProperties": {  
  
      "fileName":  
      "input.log",  
  
      "folderPath":  
      :  
      "adfgetstarted/inputdata",  
  
      "format": {  
  
        "type":  
        "TextFormat"
```

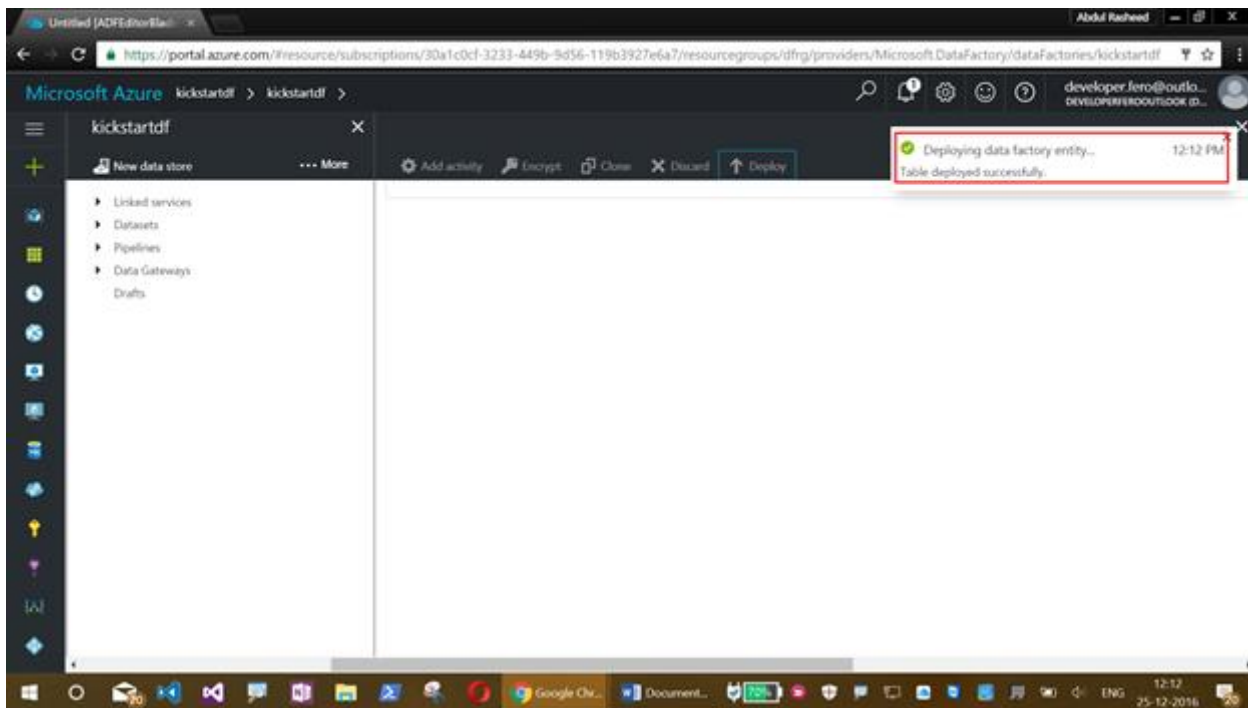


Step 5

Click on “deploy” once, as the code is copied.

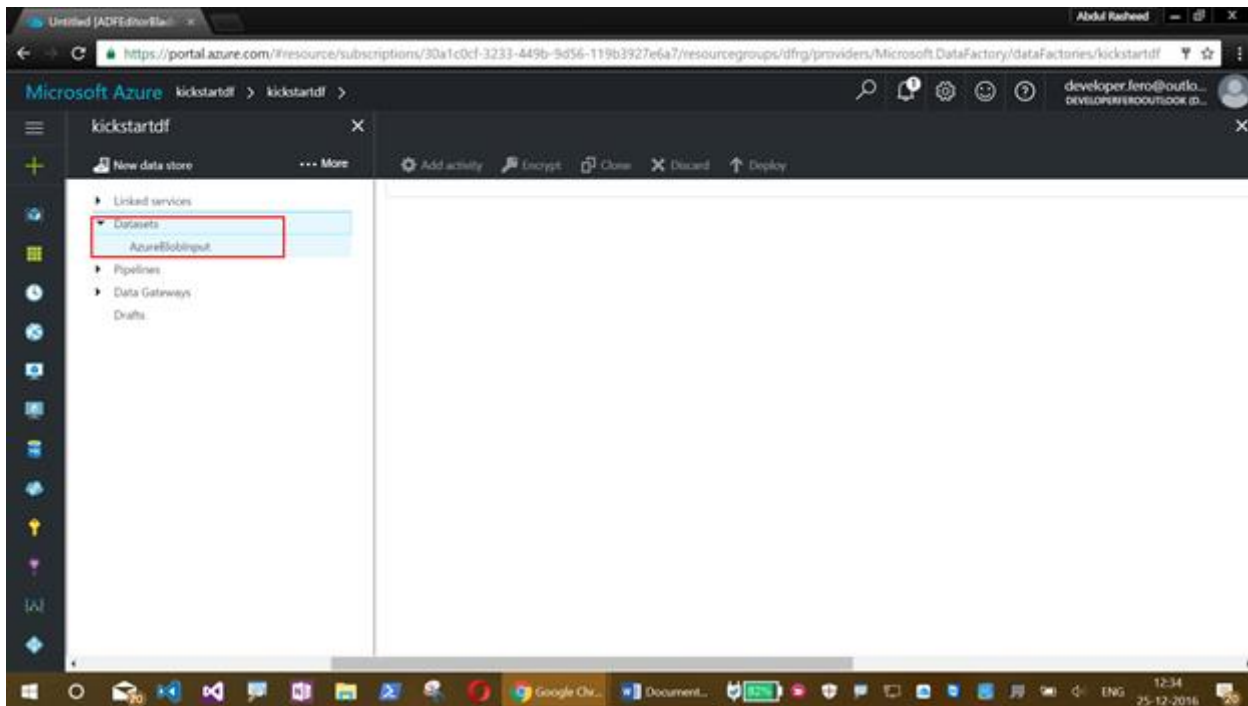


The table will be deployed and we will be getting a notification, as shown below.



The JSON properties defined above will be dealing with the properties of type, linkedServiceName, filename, type, columndelimiter, frequency/interval and external.

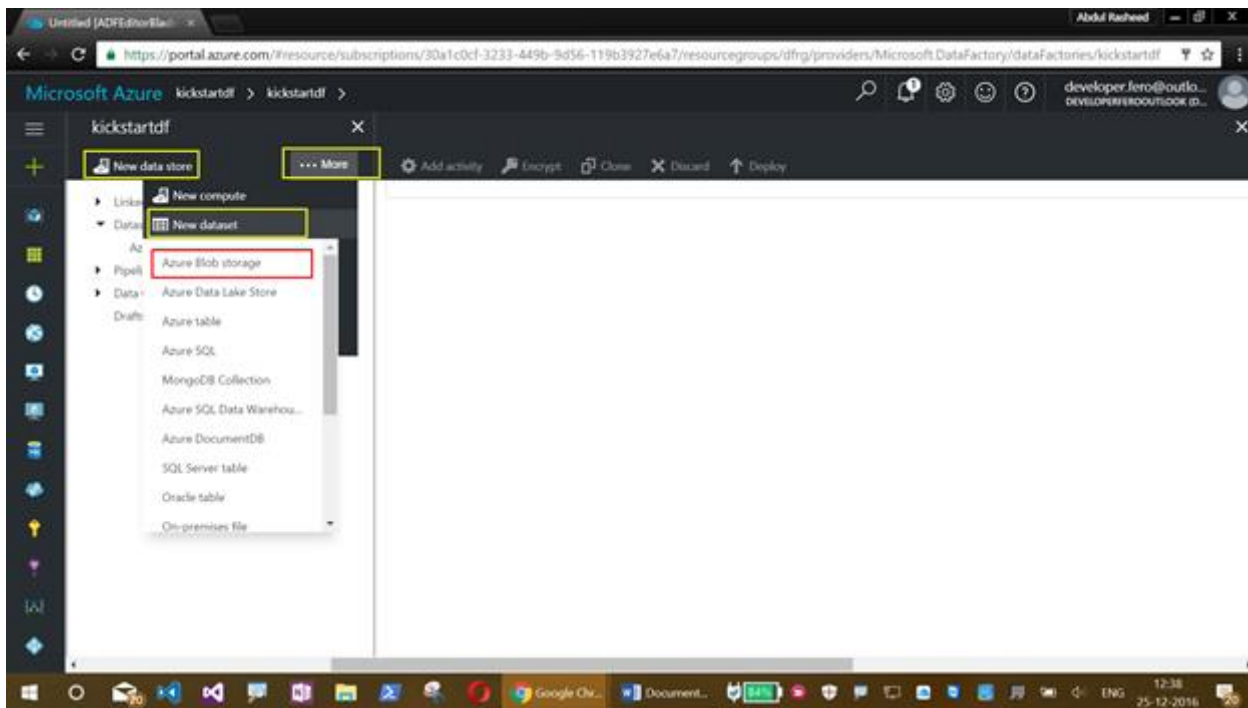
Now, we can find the Dataset on the left pane of the Data Factory blade.



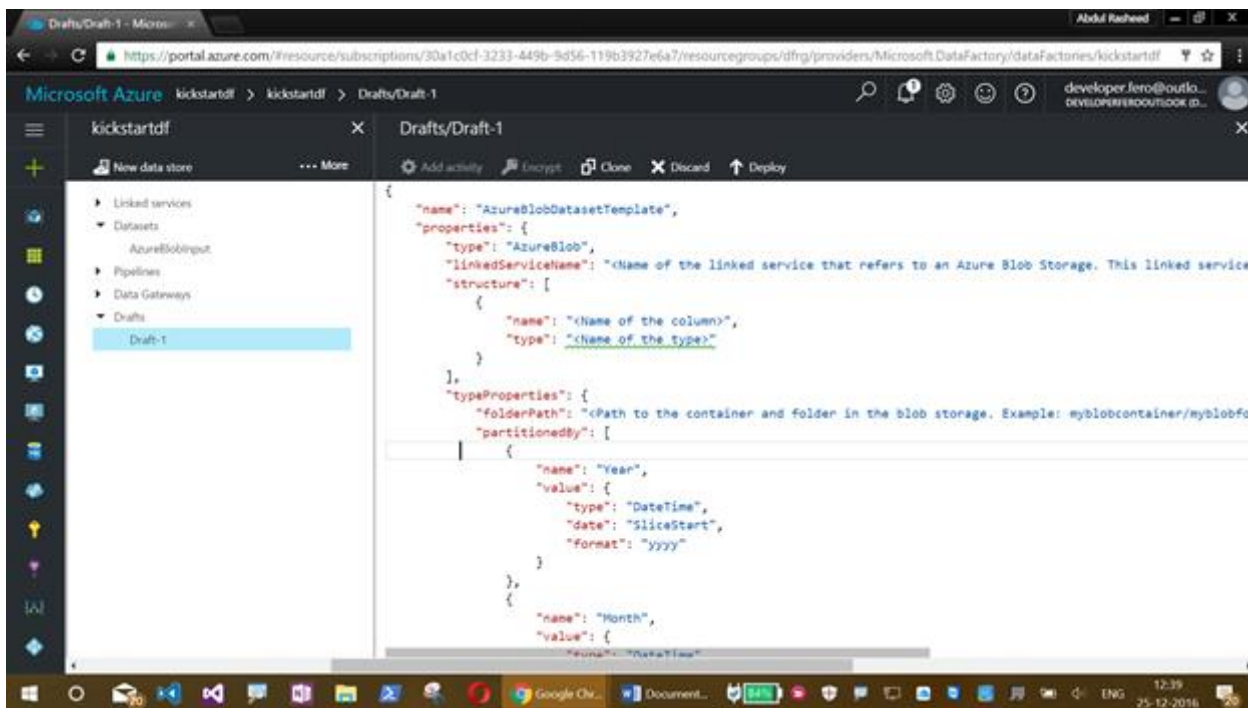
Step 6

Let's create an Output dataset now.

Go to the New data store blade - More - New DataSet - AzureBlobStorage.



We will be getting a window with the coding blade of JSON response, as shown below.



Replace the code given above with JSON response code snippets, given below.

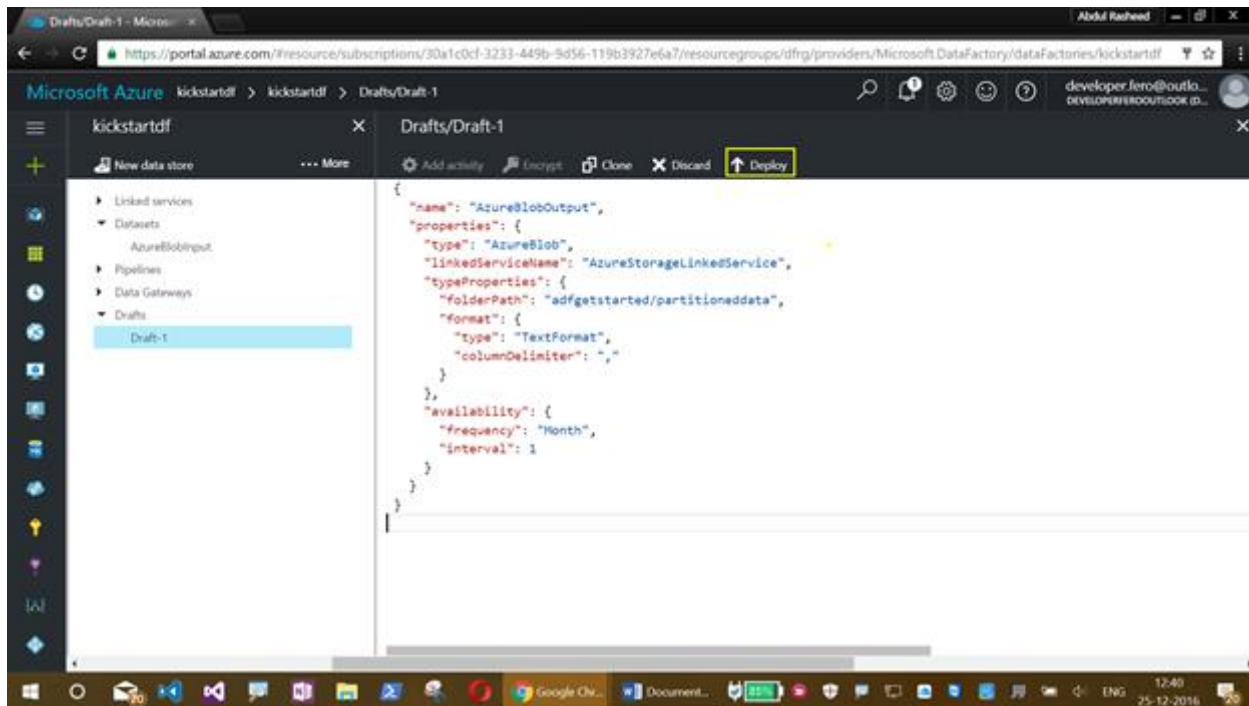
```
1. {
2.   "name": "AzureBlobOutput",
3.   "properties": {
4.     "type": "AzureBlob",
5.     "linkedServiceName": "AzureStorageLinkedService",
6.     "typeProperties": {
7.       "folderPath": "adfgetstarted/partitioneddata",
8.       "format": {
9.         "type": "TextFormat",
10.        "columnDelimiter": ",",

```

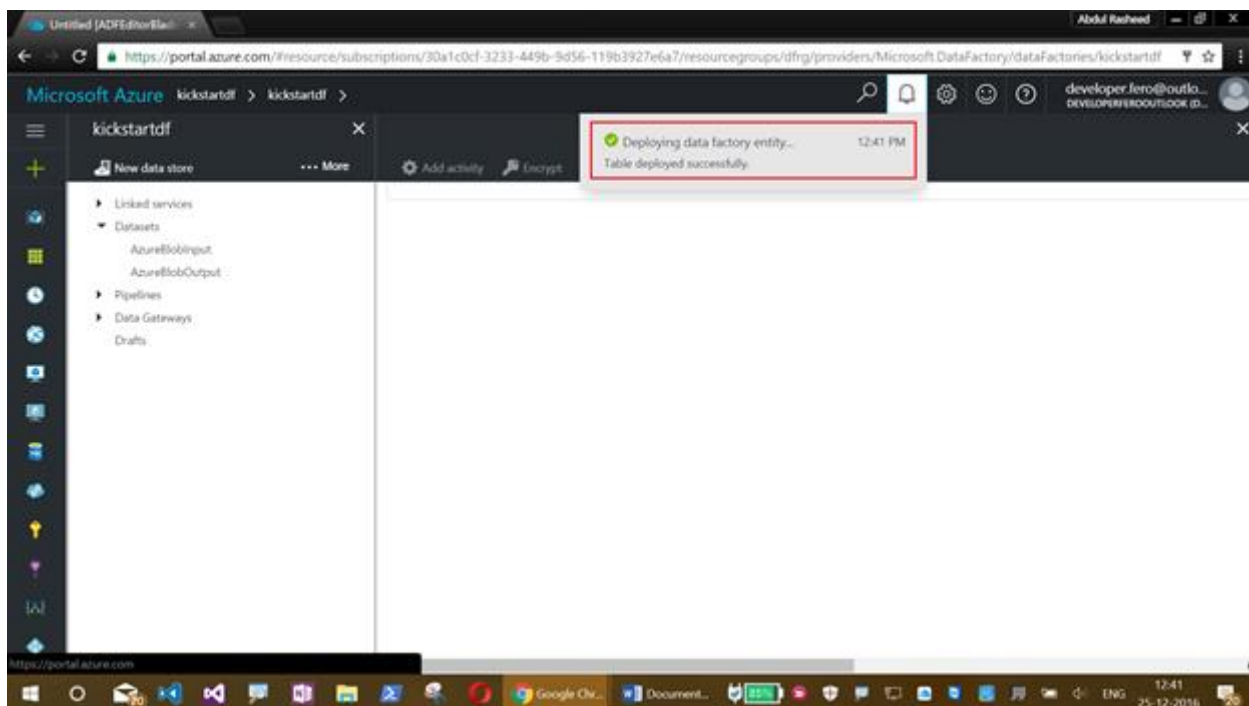
```
11.     }  
12.   },  
13.   "availability": {  
14.     "frequency": "Month",  
15.     "interval": 1  
16.   }  
17. }  
18. }
```

Step 7

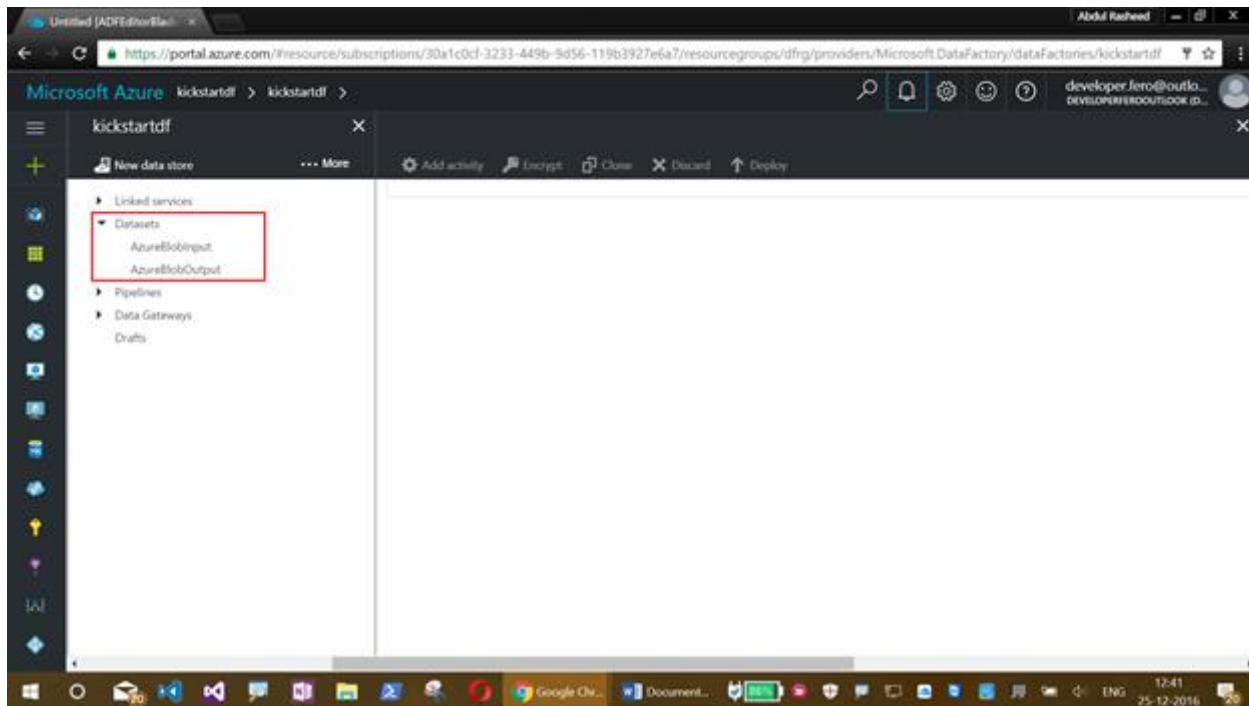
Click on "Deploy" once the JSON response is copied.



Here goes our Data Factory deployed with the new entity.



In Datasets, we can find two entries as AzureBlobInput and AzureBlobOutput, as shown below.



Follow my upcoming articles to create a pipeline and to monitor it.

Thank you for using C# Corner