

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

PRINT

Article



Creating Pipeline In Azure Data Factory And Monitoring

By [Abdul Rasheed Feroz Khan](#) on **Dec 27 2016**

Introduction

This article is in continuation of my Azure Data Factory article in which we were working with creating an Azure Data Factory account, security roles on Azure Data Lake store, and creating datasets in an Azure data factory account with HDInsight Cluster. And here, in this article, we will be creating Pipeline on the Azure Data Factory account to copy the data from one data store to another.

Note

Read my previous articles to create a data factory account with the help of the below links and you can work on this demo only if you have worked on the datasets on data factory.

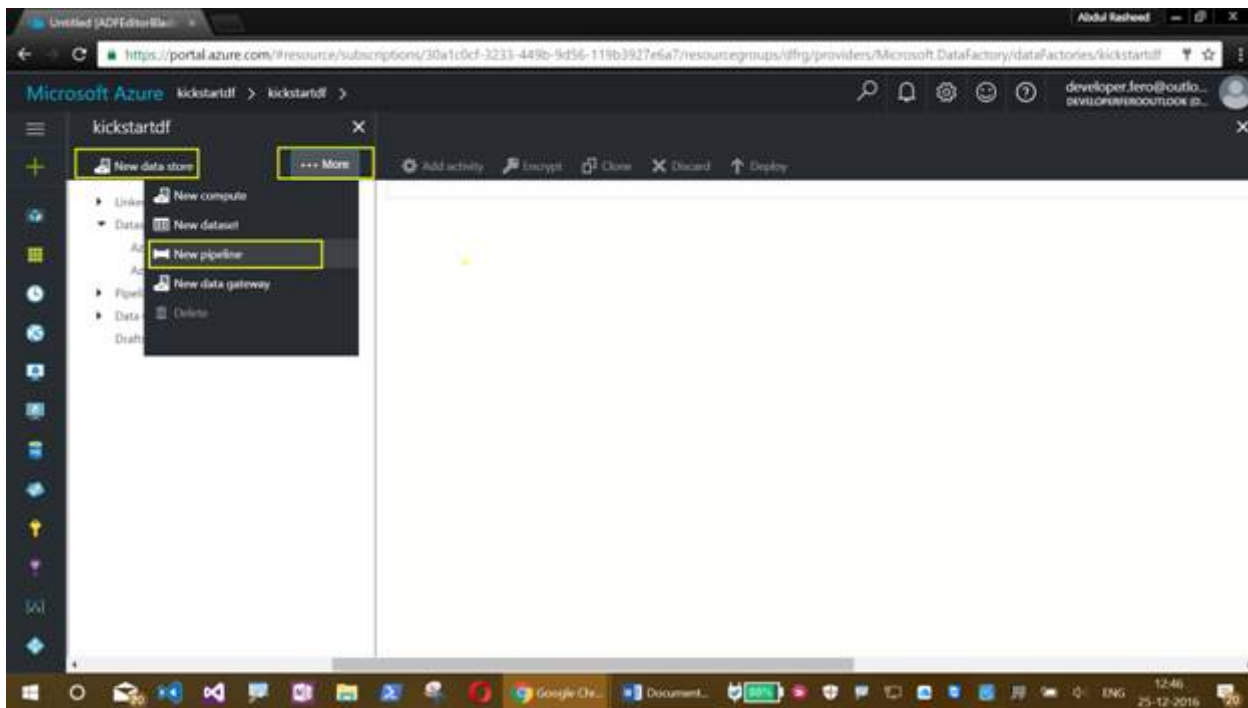
Links

1. [Security Roles on Files for Azure Data Lake Store - Part One](#)
2. [Security Roles on Files for Azure Data Lake Store - Part Two](#)
3. [Creating Linked Services in Azure Data Factory.](#)
4. [Creating Input and Output Datasets in Azure Data Factory.](#)

Follow the below steps now

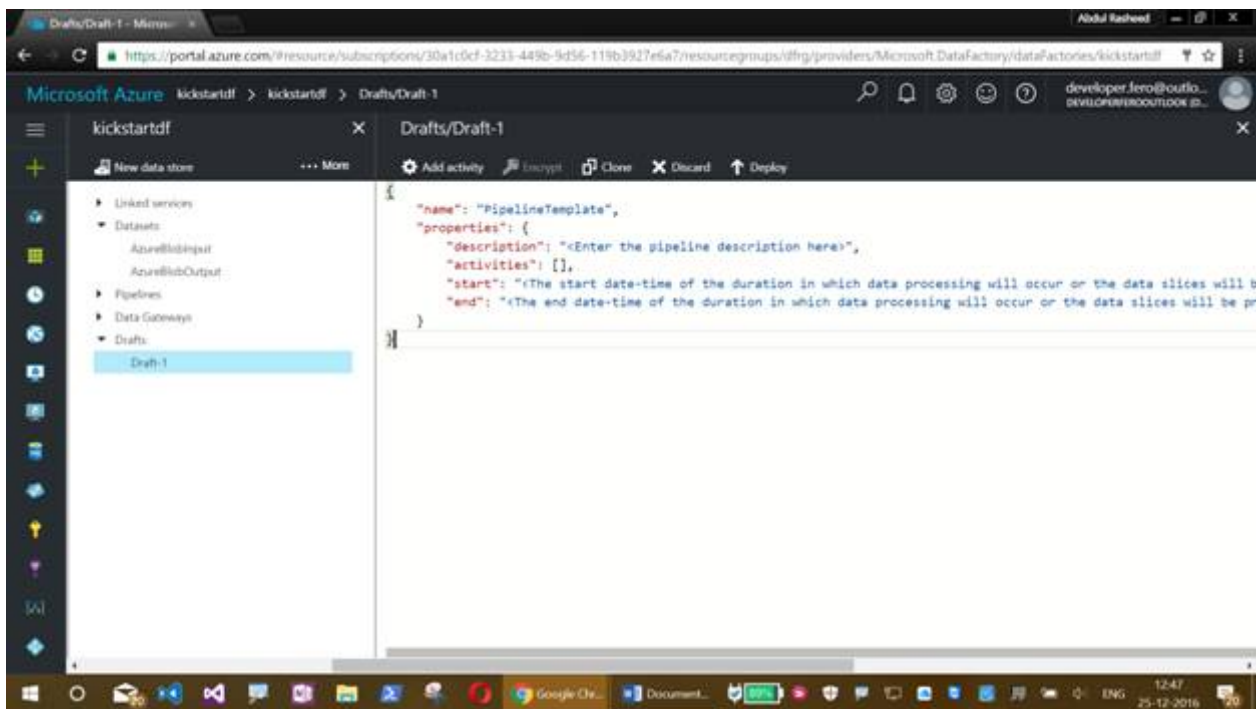
Step 1

Go to New Data Store Blade - More - New Pipeline.



Step 2

We will get a coding snippet, as shown below



Step 3

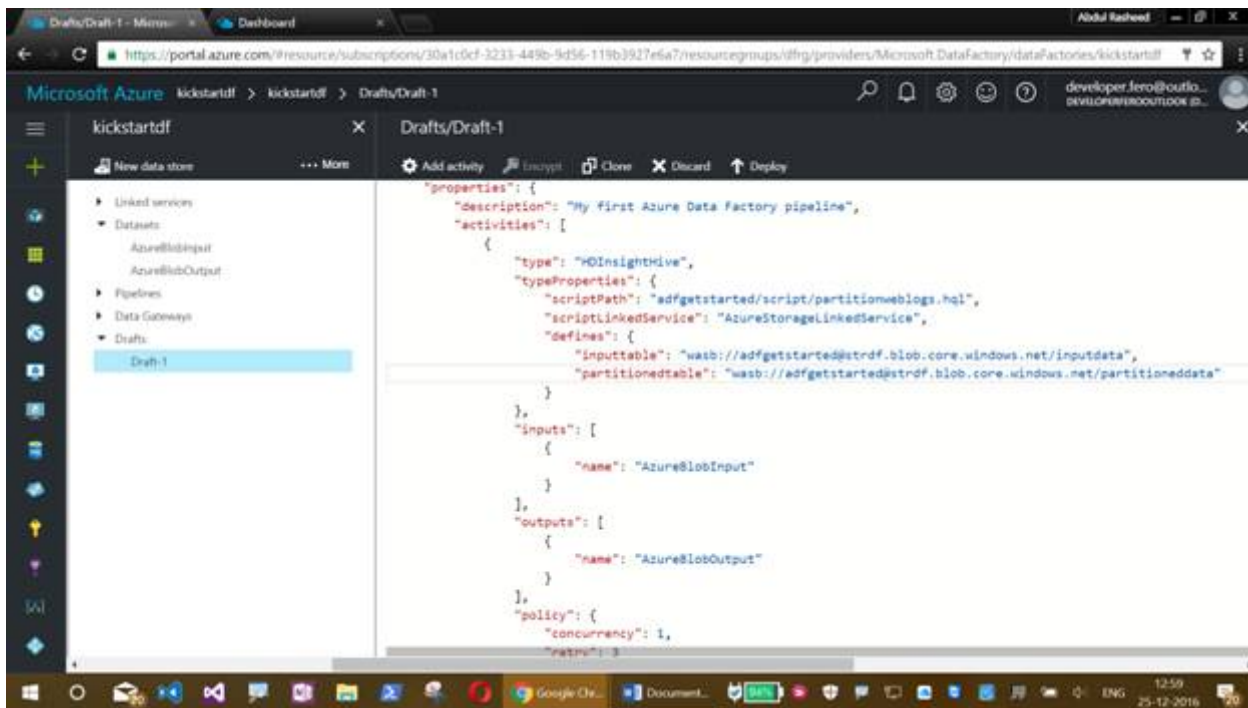
Now, copy and paste the following code in the JSON response format code editor.

Note

Replace the storage account name with your own storage account name in the below code.

1. {
2. "name"

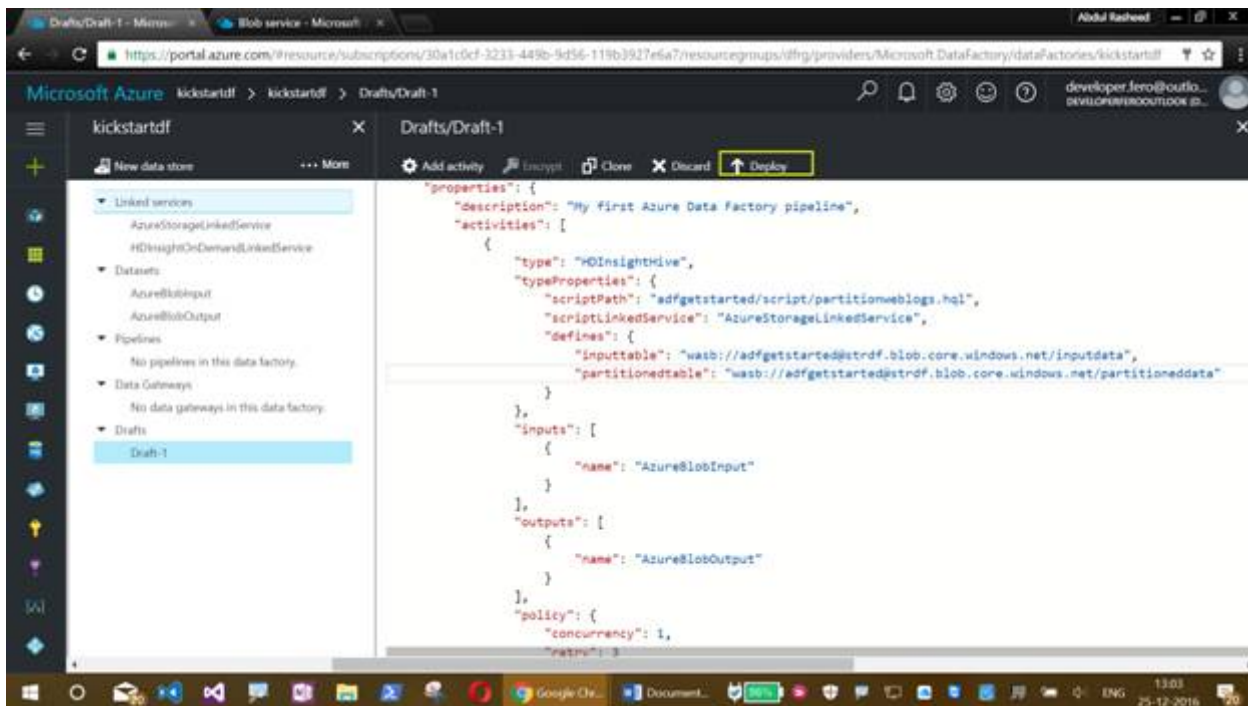
```
3. "MyFirstPipeline", "properties" {
4.   "description"
5.   "My first Azure Data Factory pipeline", "activities" [{
6.     "type"
7.     "HDInsightHive",
8.     "typeProperties" {
9.       "scriptPath"
10.      "adfgetstarted/script/partitionweblogs.hql", "scriptLinkedService"
11.      "AzureStorageLinkedService", "defines" {
12.        "inputtable"
13.        "wasb://adfgetstarted@strdf.blob.core.windows.net/inputdata", "partitionedtable"
14.        "wasb://adfgetstarted@strdf.blob.core.windows.net/partitioneddata"
15.      }
16.    },
17.    "inputs" [{
18.      "name"
19.      "AzureBlobInput"
20.    }],
21.    "outputs" [{
22.      "name"
23.      "AzureBlobOutput"
24.    }],
25.    "policy" {
26.      "concurrency"
27.      1, "retry"
28.      3
29.    },
30.    "scheduler" {
31.      "frequency"
32.      "Month", "interval"
33.      1
34.    },
35.    "name"
36.    "RunSampleHiveActivity",
37.    "linkedServiceName"
38.    "HDInsightOnDemandLinkedService"
39.  }], "start"
40.  "2016-04-01T000000Z", "end"
41.  "2016-04-02T000000Z", "isPaused"
42.  false
43. }
44. }
```



Here, by this JSON code, we are creating a pipeline consisting of a single activity that uses Hive to process data on an HDInsight Cluster.

Step 4

Click on "Deploy" now.

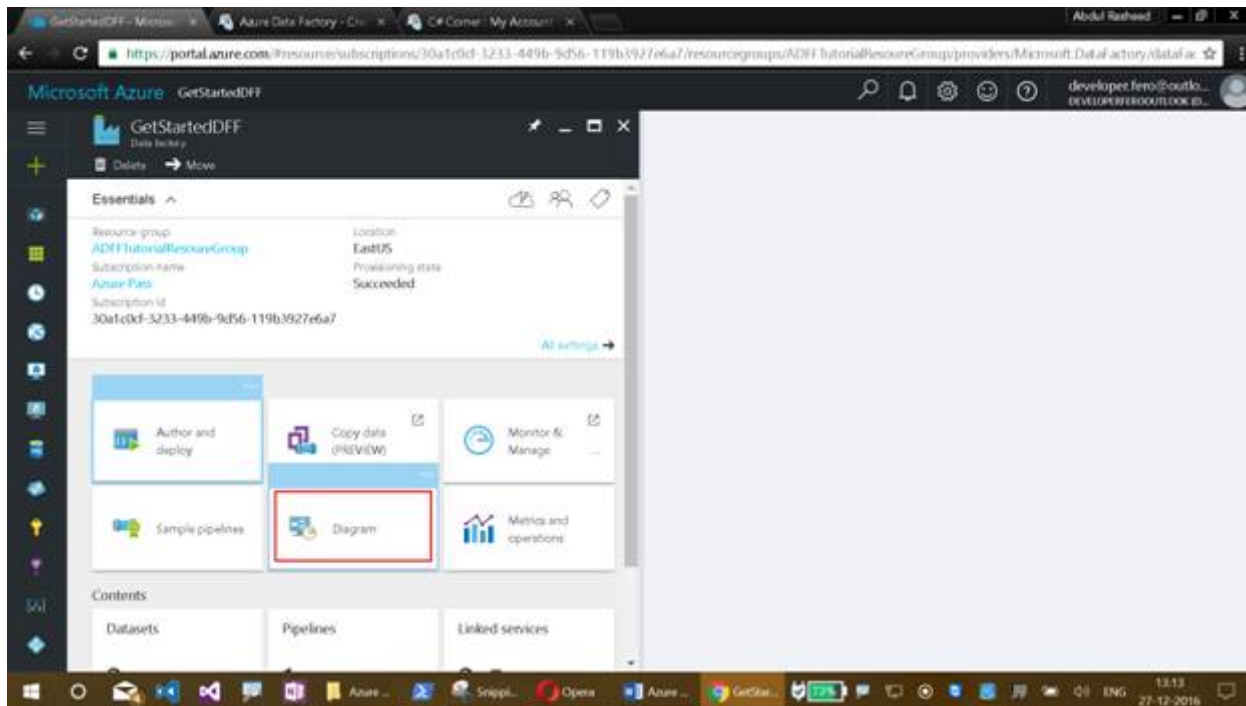


Once it is deployed, you can find "MyFirstPipeline" in Azure Data Factory under Pipelines.



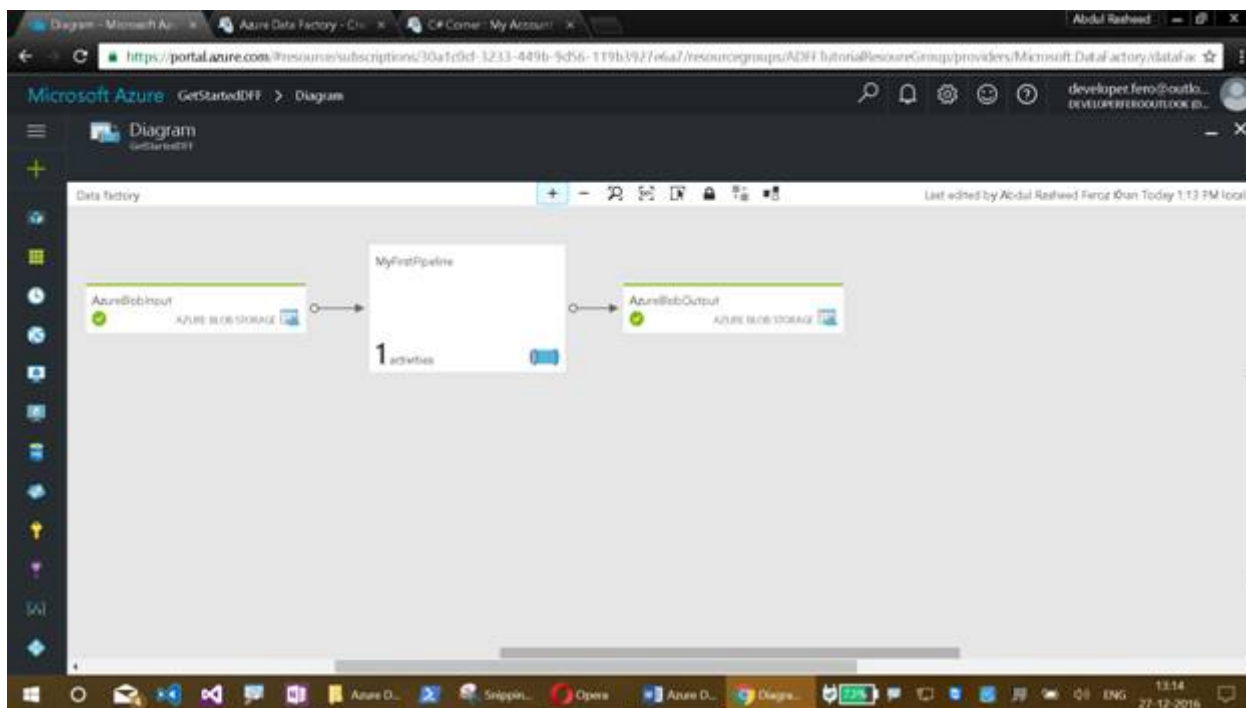
Step 5

Now, let's work on monitoring the Pipeline in Azure Data Factory. Go to the homepage of Azure Data Lake Store account and click on the diagram as shown below



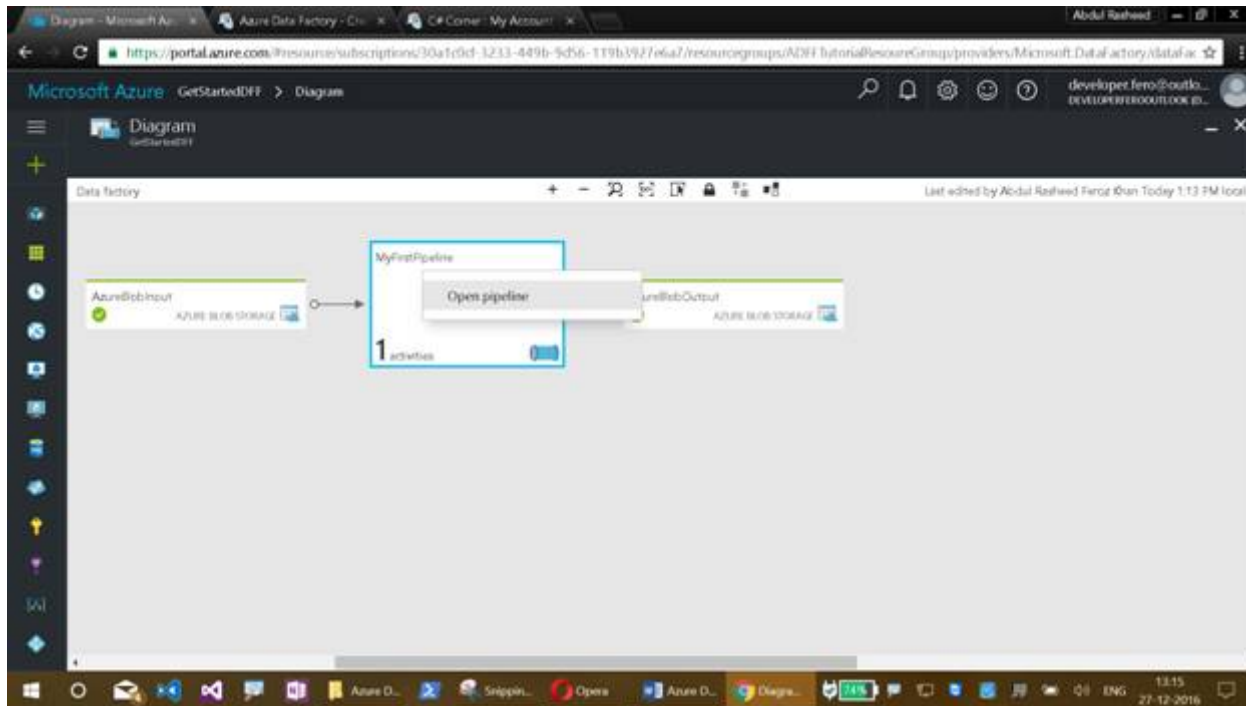
Step 6

Here, we can find an overall diagram of datasets and pipelines that we have used in the Azure Data Lake Store account.



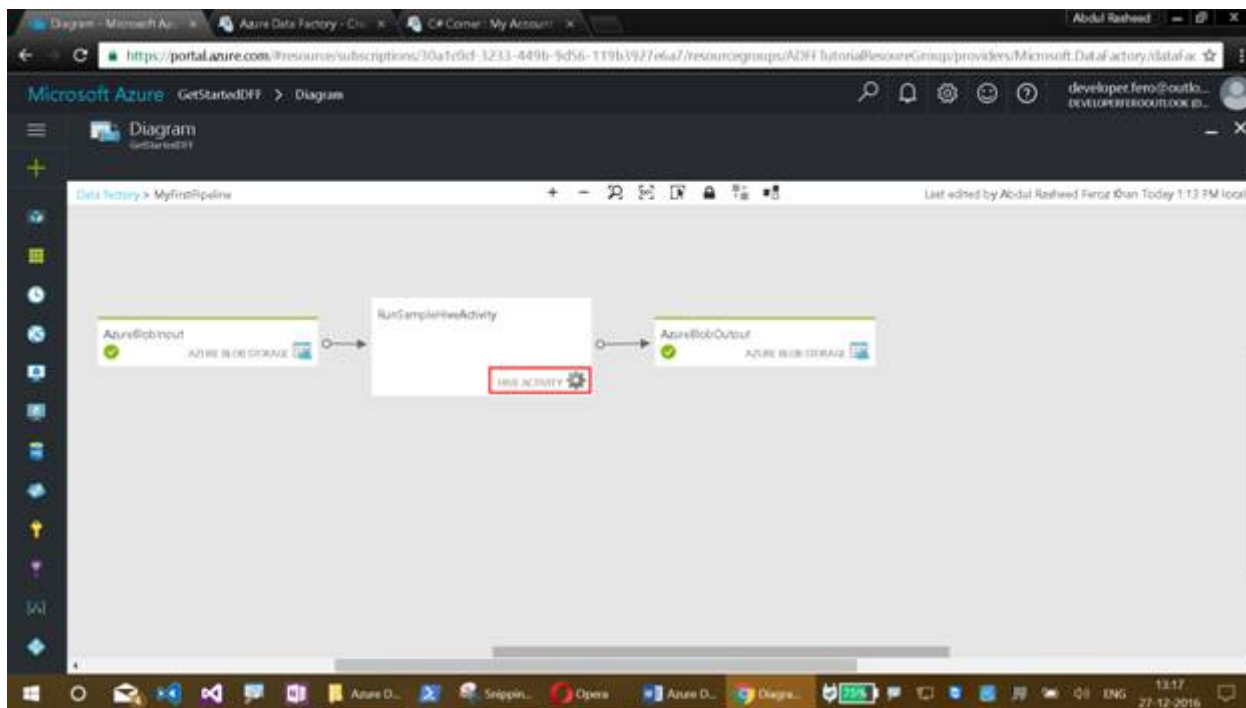
Step 7

We can also right click on an action like "MyFirstPipeline" and open the pipeline in the diagram to find the activities.



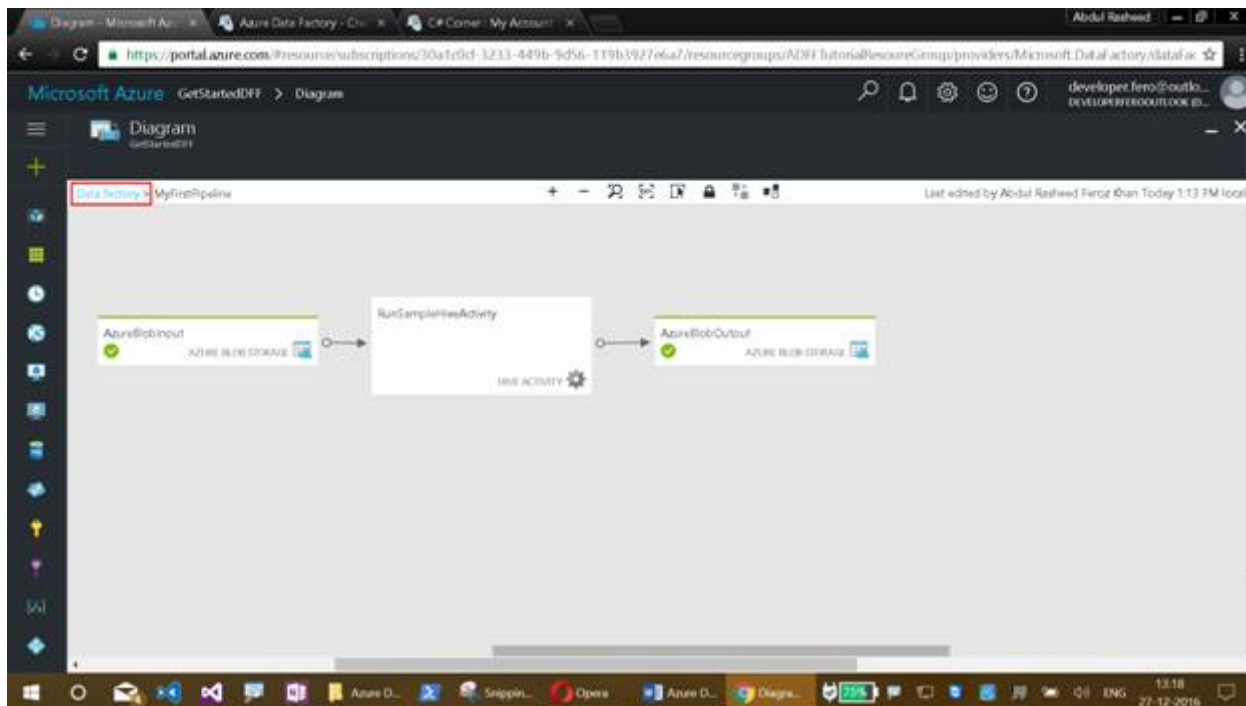
Step 8

We can find the Hive Activity in the Pipeline here.



Step 9

Click on Data Factory to move back.



Similarly, we can monitor the Azure Blob Input activity and Azure Blob Output Activity too.

Note

In my next article, we will be working on the same with Visual Studio for Azure Data Factory.

Thank you for using C# Corner