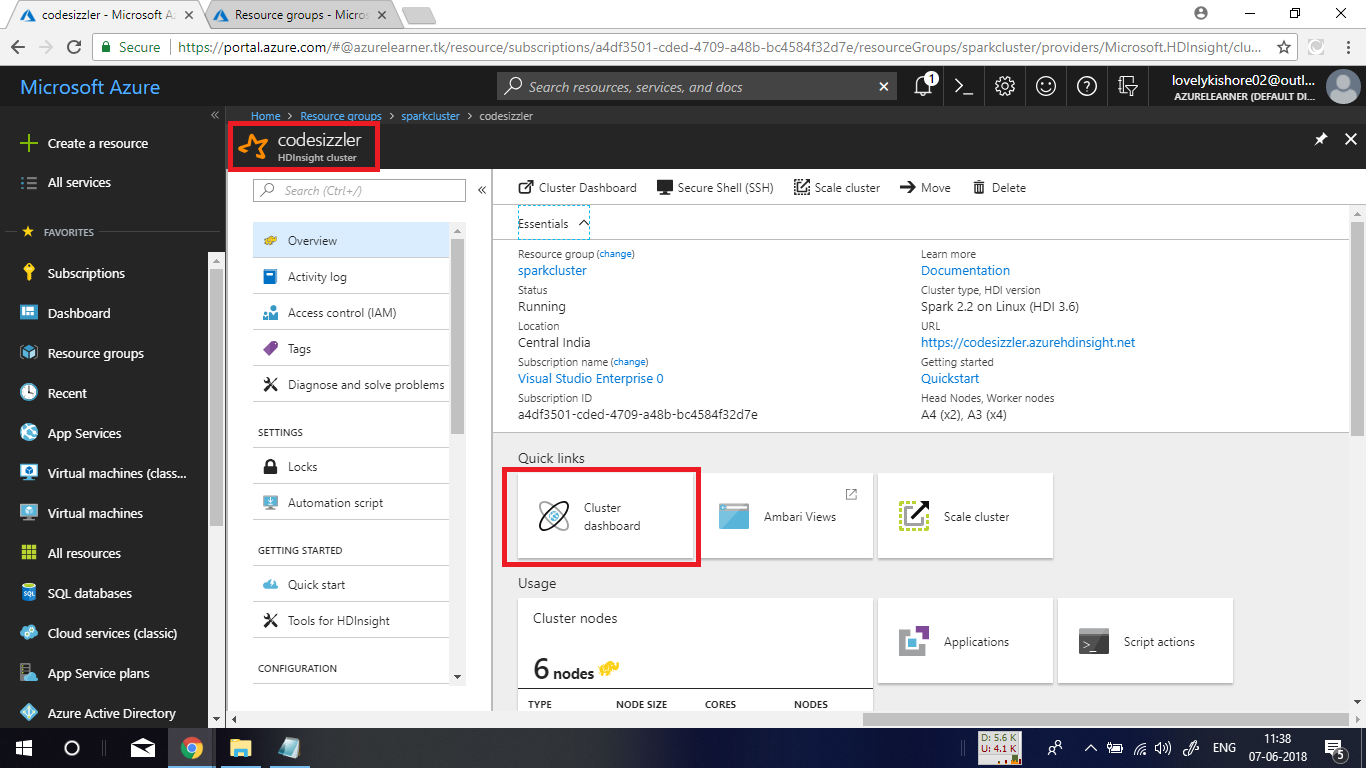
# **Querying with Spark Cluster**

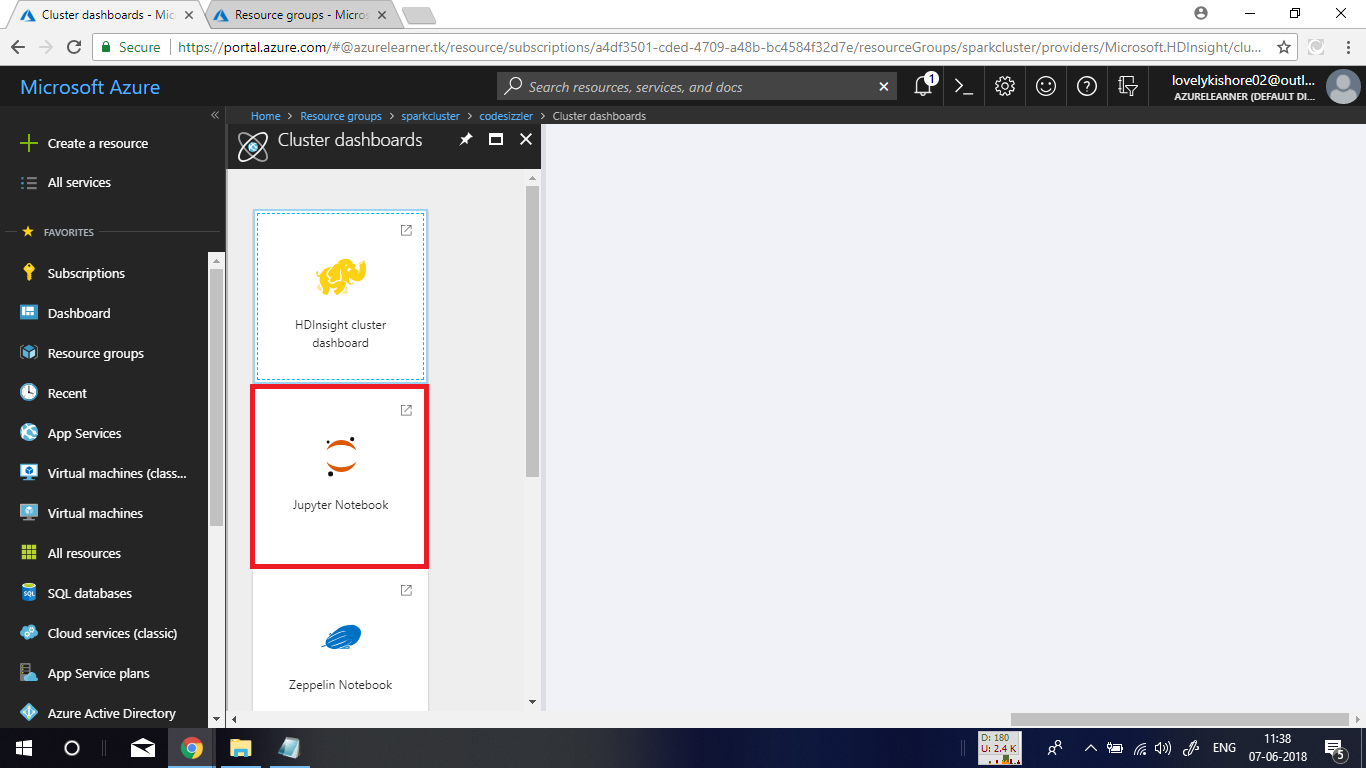
This demo is for using queries and working with Spark Cluster in HDInsight. This hands-on follows by,

1. Creating a dataframe from CSV file.
2. Running queries in dataframe.

**Creating Jupyter Notebook:**

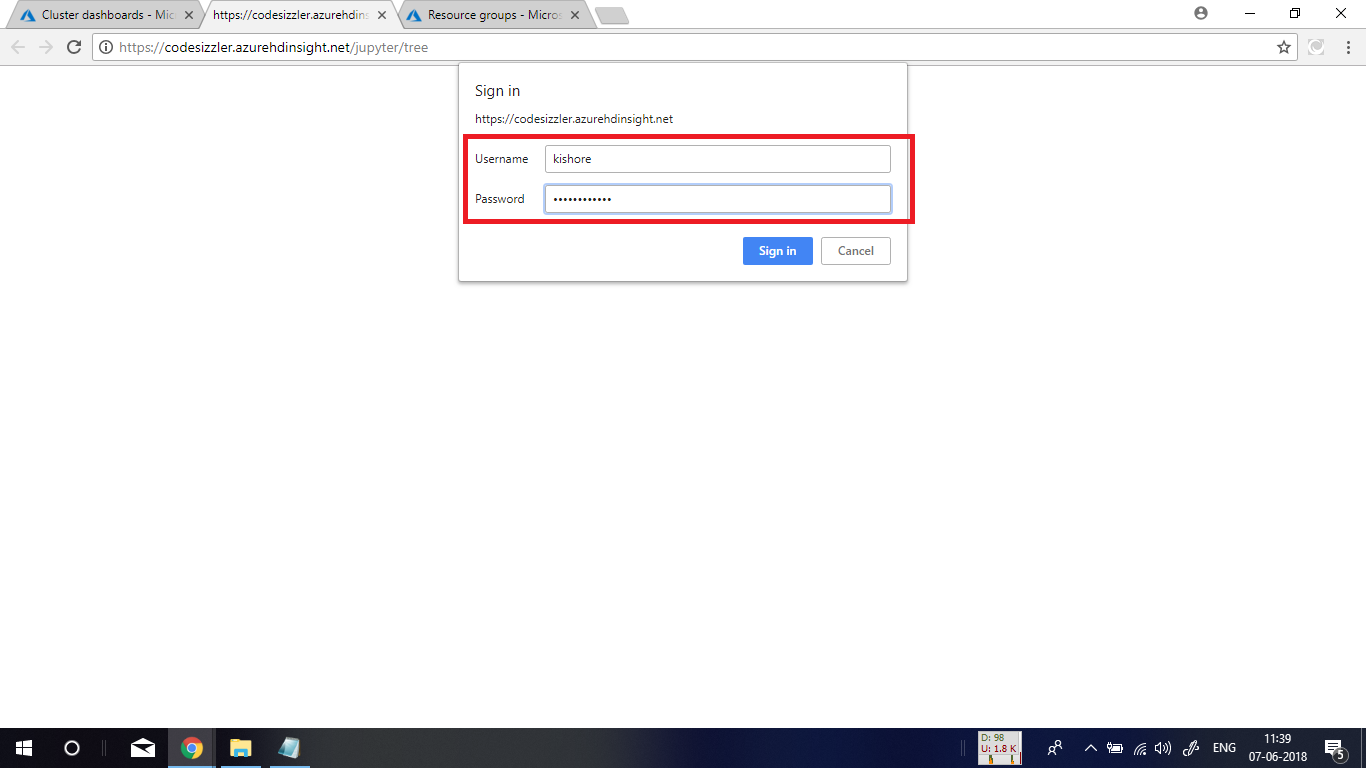
In the HDInsight cluster that you have created, click on **Cluster Dashboard.** This will show you another blade. In there, click on **Jupyter Notebook.** This will take you to new window.



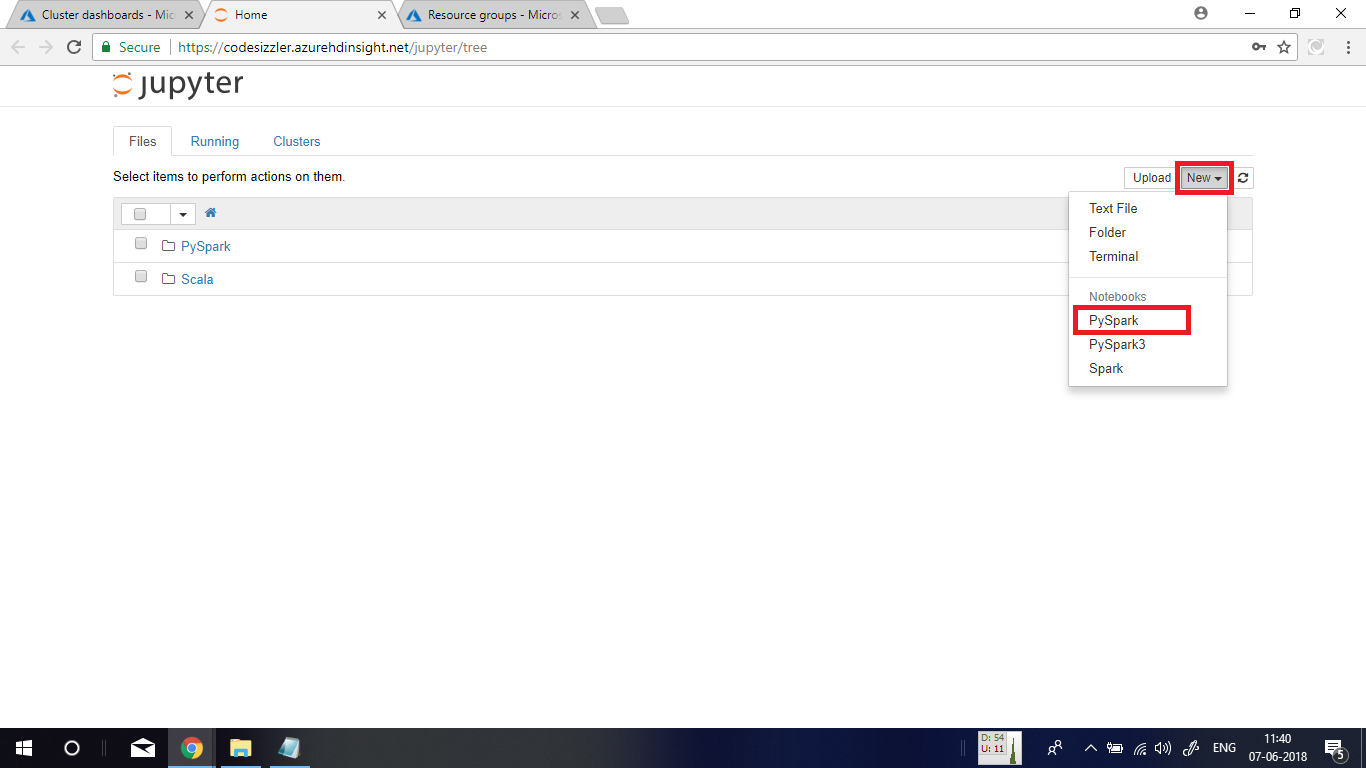


**Signing into Jupyter Notebook:**

In the new blade enter **Username** and **Password** the you gave while creating clusterand hit on **Sign In.**



After successful sign in, in the homepage of **Jupyter Notebook** that you get, click on **New** option in the top right corner and choose **PySpark** option to create new **Jupyter Notebook**.



**Executing Spark SQL Statements:**

To import some required types for this hands-on for this scenario, paste the below given command in the scripting pane and press **Shift+Enter** to execute it. The command is,

**from pyspark.sql import \***

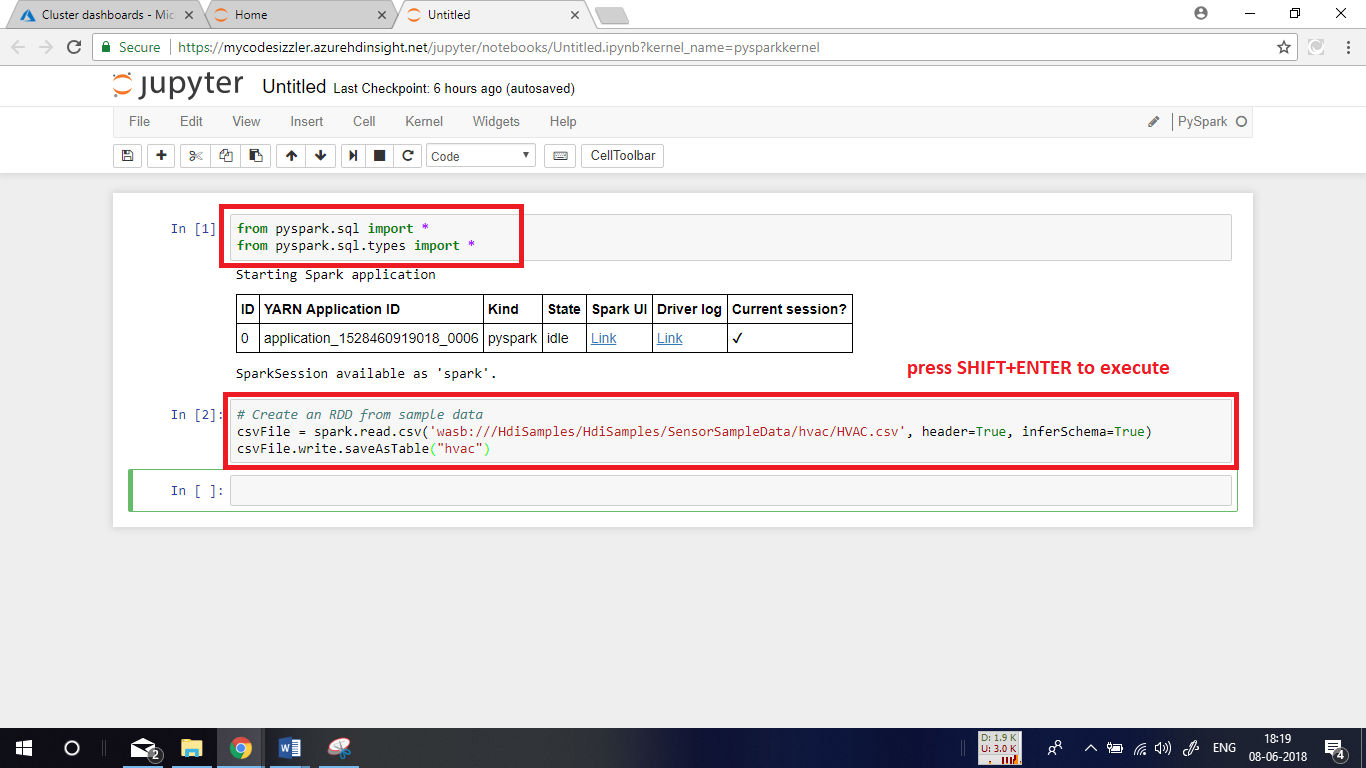
**from pyspark.sql.types import \***

Now, run these commands to create dataframe and a temporary table.

**# Create an RDD from sample data**

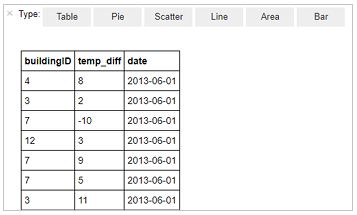
**csvFile = spark.read.csv('wasb:///HdiSamples/HdiSamples/SensorSampleData/hvac/HVAC.csv', header=True, inferSchema=True)**

**csvFile.write.saveAsTable("hvac")**



**Querying with Dataframe:**

Now, run the below given command to get data from the **hvac** table.



Click on the **Area** option in the output to visualize the table.

