Article

Kickstart Hadoop In HDInsight On Windows

By **Abdul Rasheed Feroz Khan**   on   **Sep 10 2016**

Introduction

This article will help you learn Hadoop using HDInsight on Windows. Hadoop helps you draw some meaningful information from a set of unstructured data, using Hive query.

What is Hadoop?

Hadoop is a framework that helps for distributed storage and distributed processing of very large data sets on computer clusters, built from commodity hardware.

What is HDInsight?

HDInsight provides data storage concepts like HDFS (Hadoop Distributed File System) and a simple MapReduce programming model to process and analyse the data.

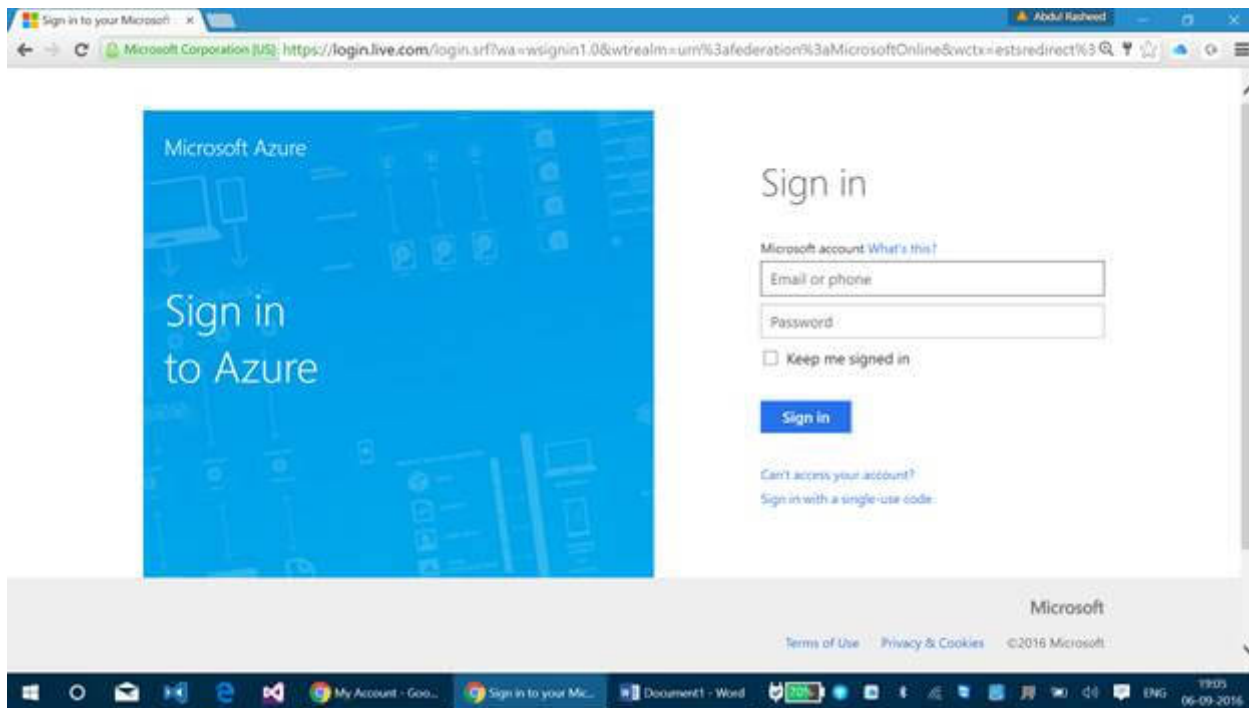Here, as a first module, we will be working with creating Hadoop Clusters:

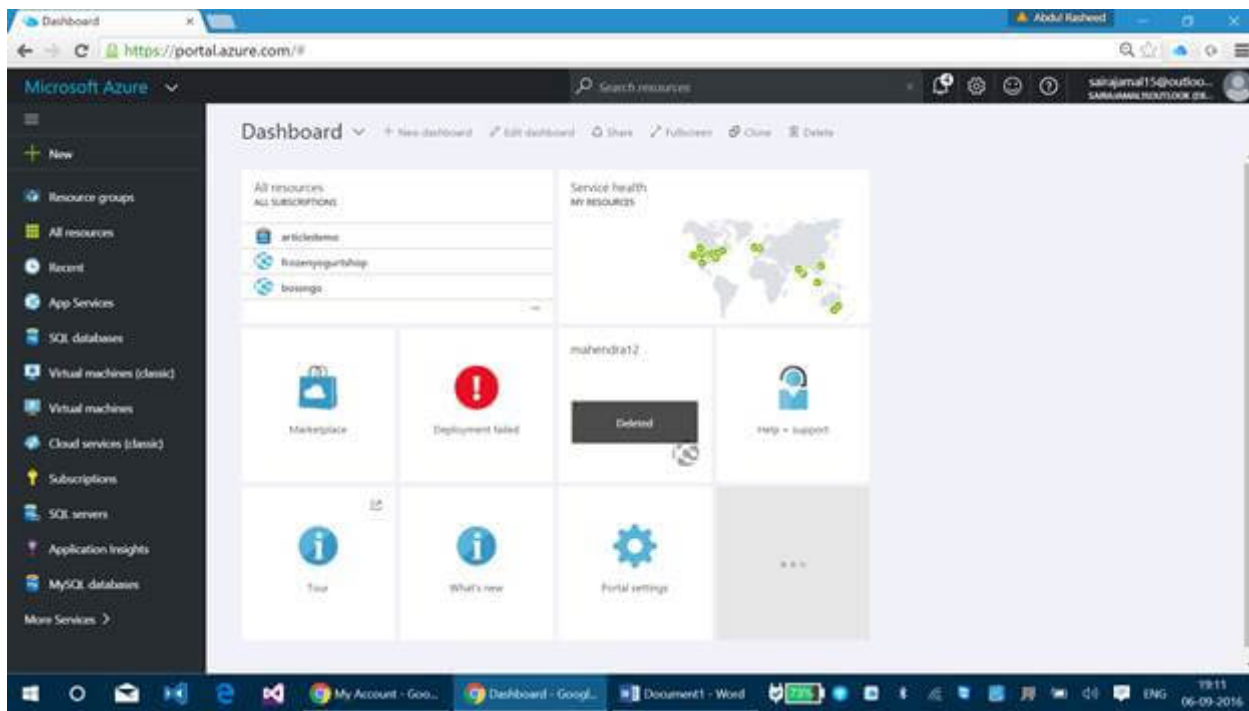### Pre-requisites

- An Azure Subscription.

Note: Click here to get a free trial account of Azure.
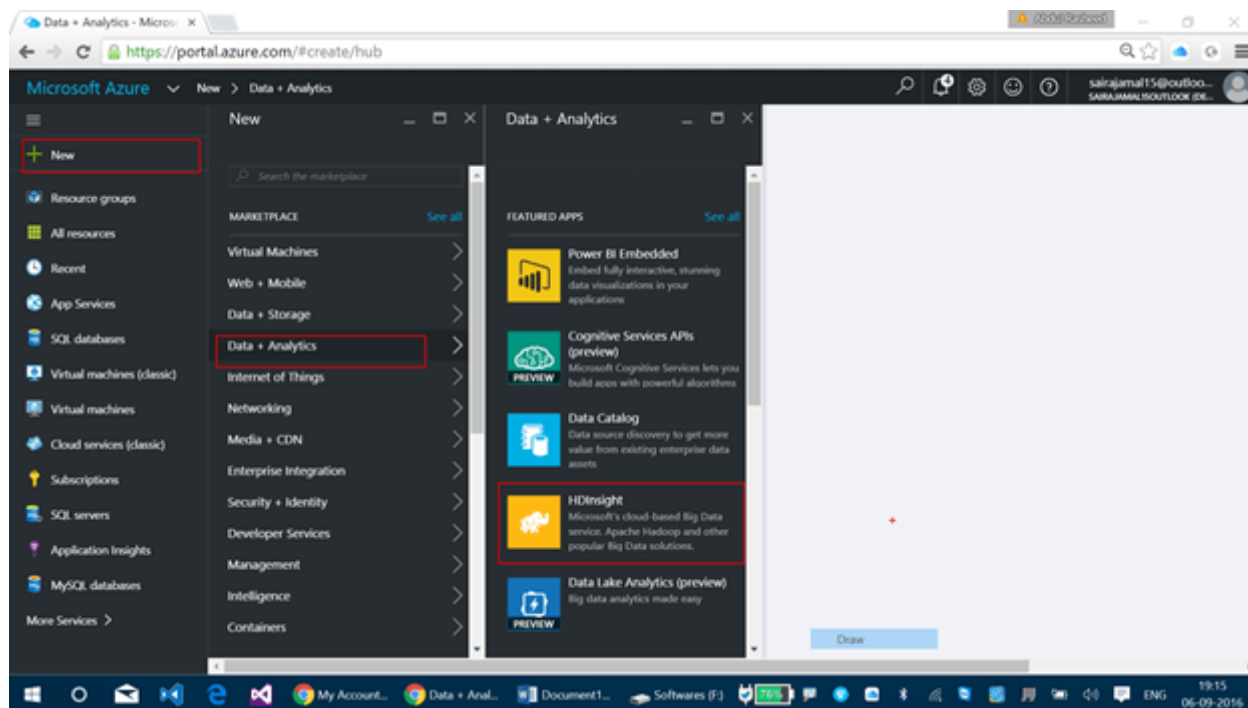
Follow the below steps, now, to create Hadoop Clusters.

Step 1 - Login to the Azure Portal, using www.portal.azure.com, and sign in with your credentials.
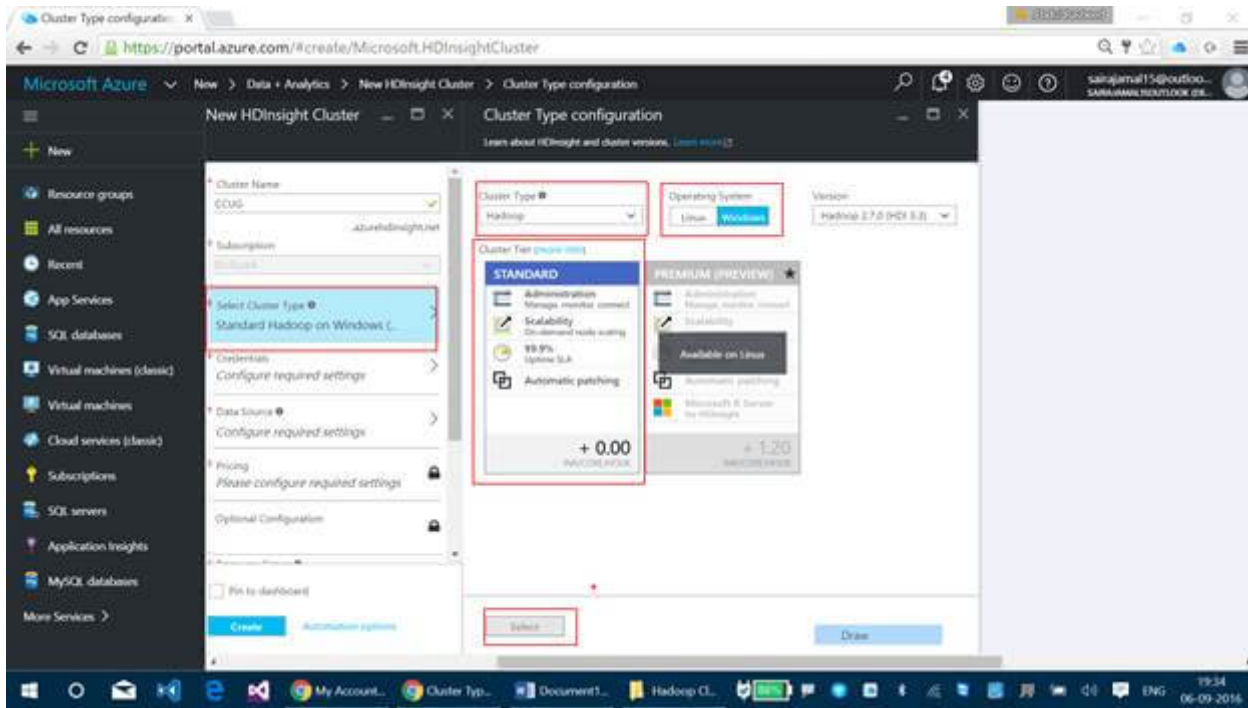
Step 2 - Here comes the Azure Portal.
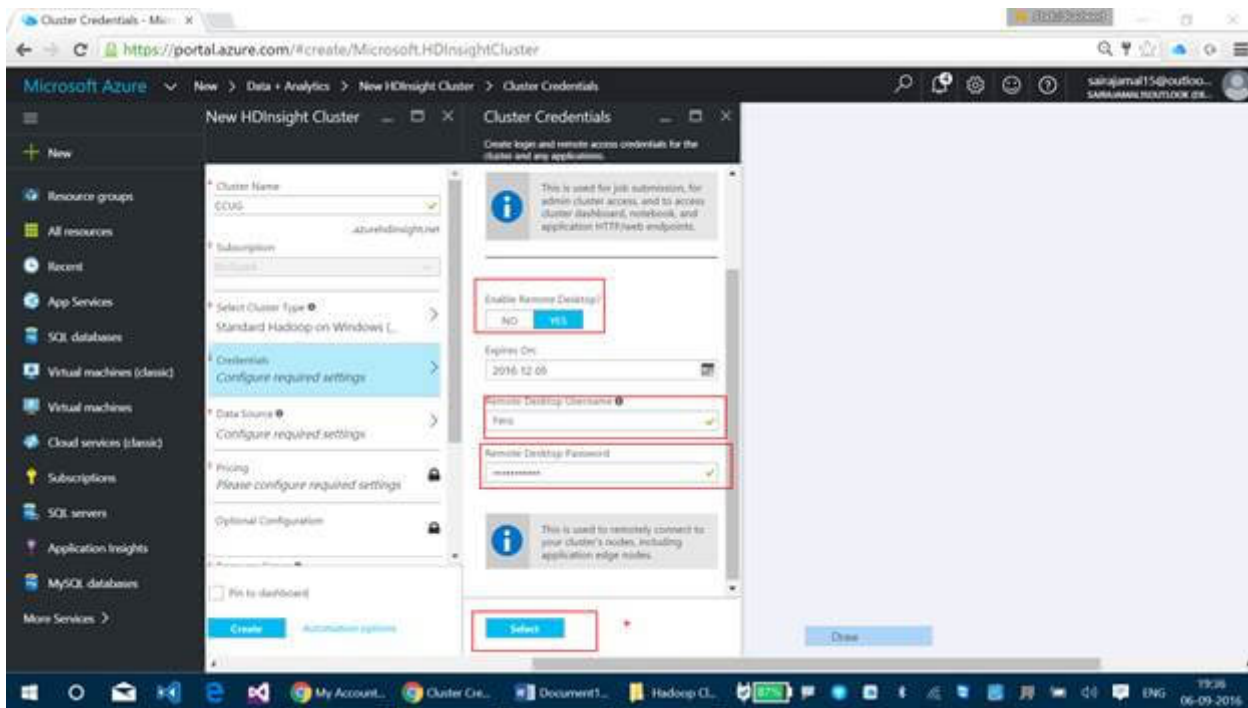

Click on New - Data + Analytics.

Step 3 - Enter the following details needed.

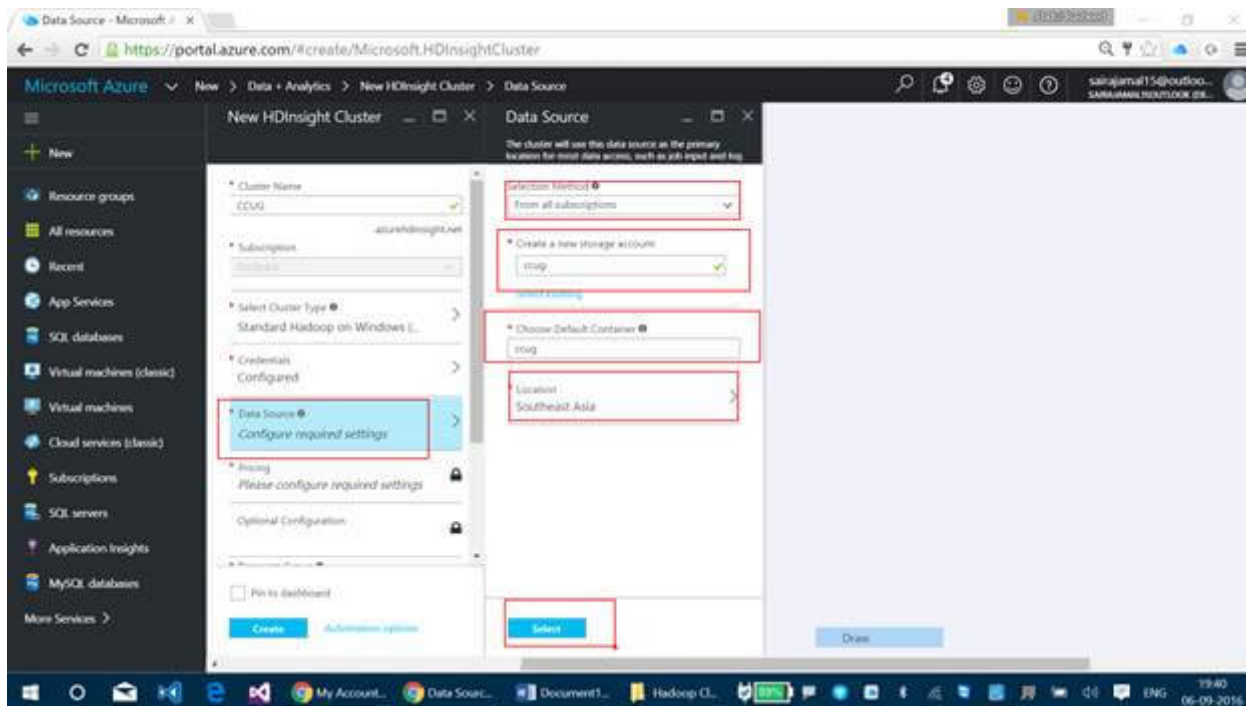| Fields	Details	Defined
Cluster Name	It should be a unique name for identifying the cluster	Here I have named my Cluster as CCUG.
Subscription	Select the Subscription that you have	
Cluster Type	Hadoop	Select Cluster Type → "Hadoop" at Cluster type → OS as "Windows" and followed by the Standard one for pricing tier. Click on Select button once after its done.
Credentials	Enter the Cluster login name and password, Windows based Clusters can have two user names. You can also create a remote desktop by enabling it and configuring for a Remote Desktop User Account	Select "Yes" for Remote Desktop Connectivity followed by a User Name and password with a select button.
Data Source	Every HDinsight cluster has a default Blob container on an Azure storage account. The location of the default Azure storage account determines the location of the HDInsight cluster.	Click Create New to create a new default Azure storage account. Use the cluster name as the default container name.
Node Pricing	You can use 1 or 2 worker nodes with the default worker node and head note pricing tier.	Mention the node worker for one or two over here.
Resource Group	A basic HDInsight cluster contains a cluster and its default storage account. You can group the two into a resource group for easy management.	Select an existing Azure resource group or create a new resource group. Here I have created a new user group.

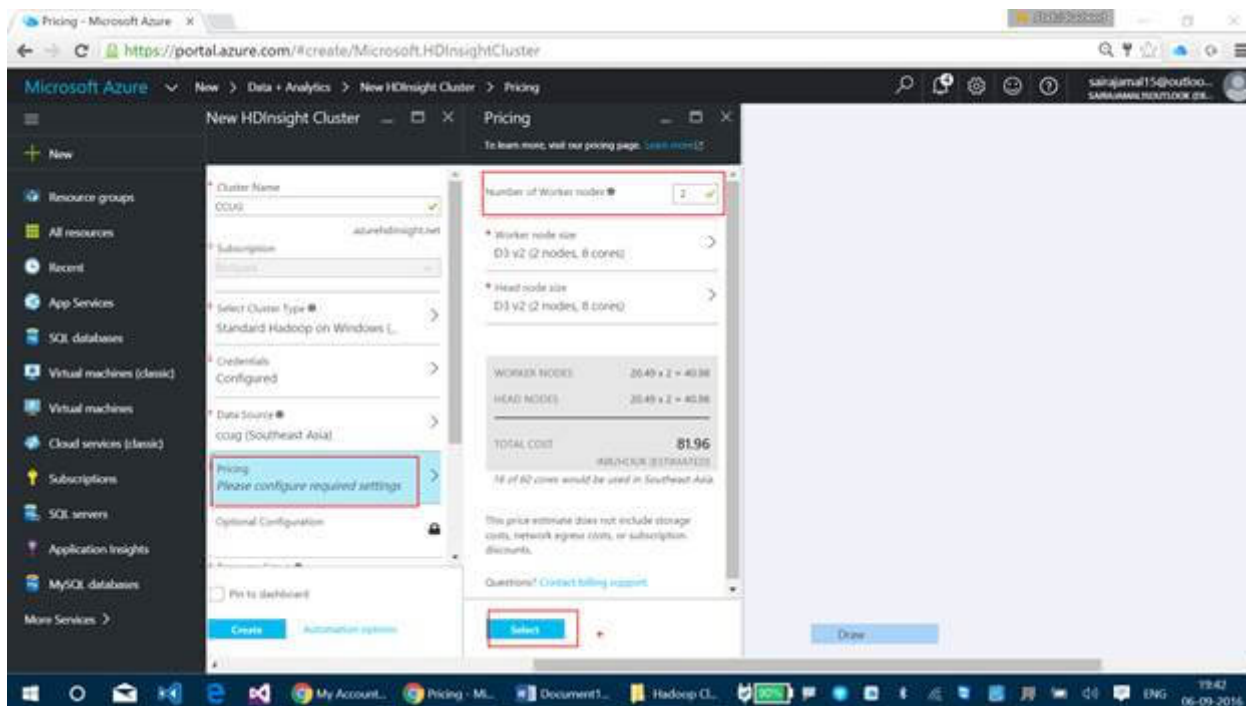Surf the below images for reference and the below image for Cluster Type.
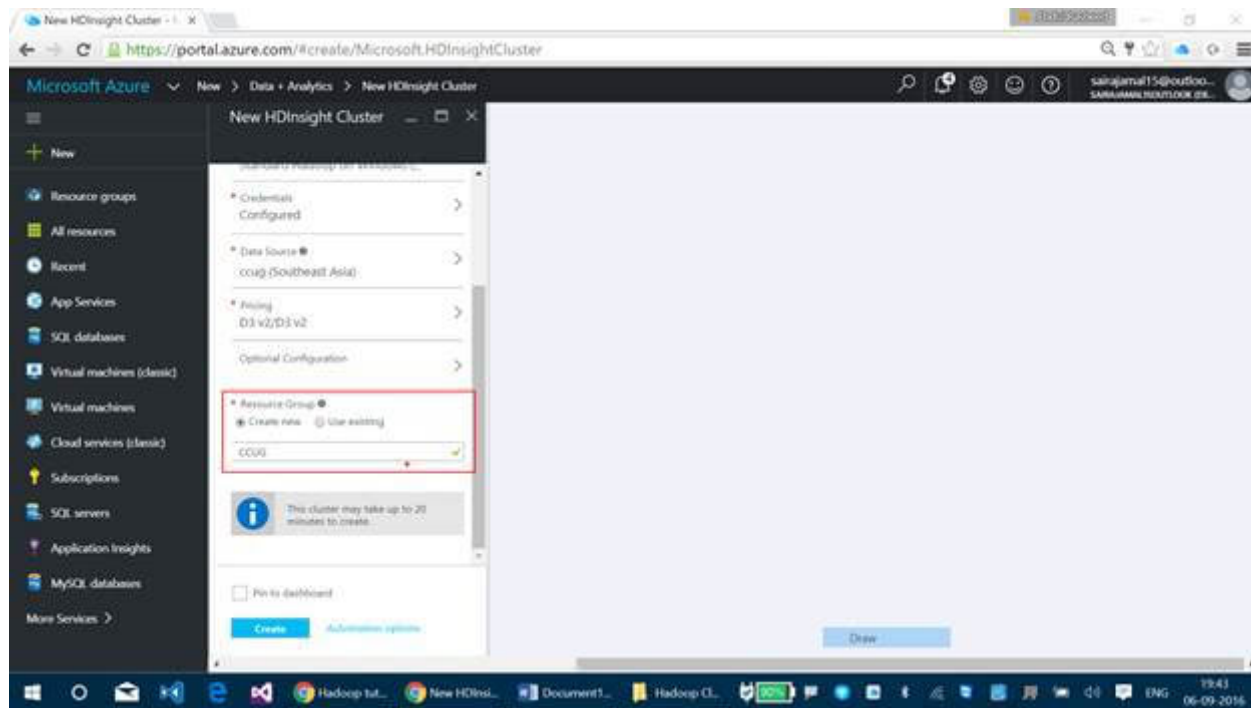
Below image is for credentials.

For Data Source

For Pricing

For Resource Group

Click on Create button now. It will take somewhere between 15 to 20 minutes to create a Cluster.

Thank you for using C# Corner