Azure HDInsight Operations by Using Windows PowerShell ISE

This tutorial demonstrates the following operations by Windows PowerShell ISE

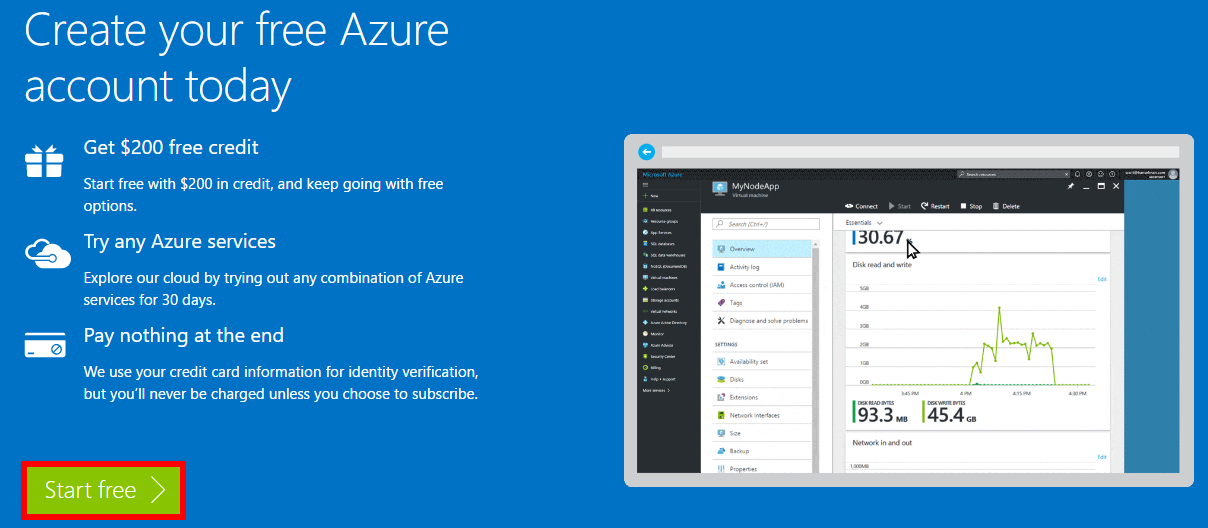
* Creating an HDInsight cluster
* Submitting a MapReduce job the cluster and executing it
* Transferring results of the job to a blob storage or ADLS storage
* Deleting the HDInsight cluster

Prerequisites:

* Azure subscription or free trial account ([30 day free trial](https://azure.microsoft.com/en-us/pricing/free-trial/))
* Azure PowerShell ([link](https://github.com/Azure/azure-powershell/releases/download/v3.6.0-February2017/azure-powershell.3.6.0.msi))
* Windows PowerShell ISE ([link](https://www.microsoft.com/en-us/search/DownloadResults.aspx?q=%22windows%20management%20framework%22%20PowerShell&sortby=Relevancy~Descending))
* Create an HDInsight cluster (Including the followings:)
  + Create an Azure resource group
  + Create an Azure Storage account
  + Create an Azure Blob container

Subscribe of free trial account to Azure Storage:

1. Follow the link provides in ‘Prerequisite’ and click on ‘Start Free’
2. Follow the instruction after the navigated page and complete subscriptions



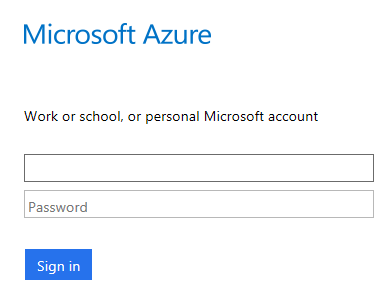
Setting up Azure PowerShell and Windows PowerShell ISE

1. Follow the Azure PowerShell link provides in ‘Prerequisite’ and install
2. Follow the Windows PowerShell ISE link provides in ‘Prerequisite’ and check for installation or upgrade

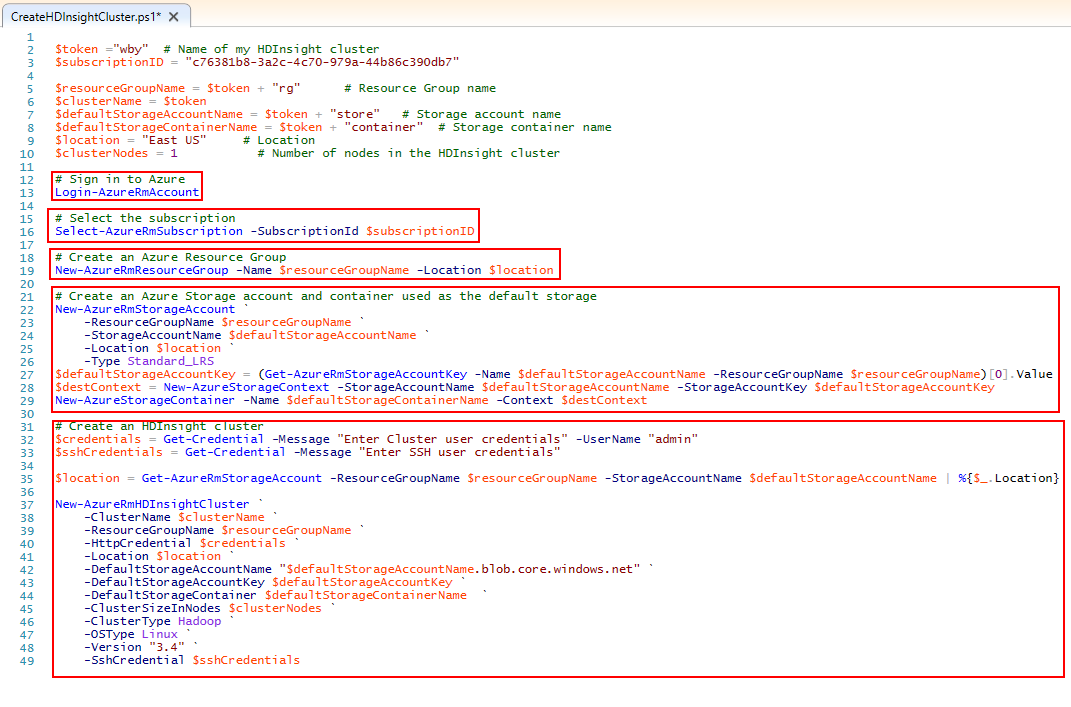
# 1 Creating an HDInsight cluster

1. Open Windows PowerShell ISE
2. Open a **new file** on Windows PowerShell ISE
3. Connect to Azure Subscription by using ‘Add-Azureaccount’ command and log in

Command: 



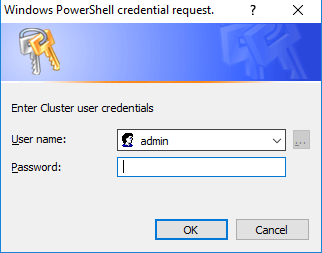
1. Specify the name of HDInsight Cluster (Here I used String ‘token’)
2. Specify the names of resource group, new storage account and new storage container
3. Use the following commands to create an Azure resource group, an Azure storage account and container
4. Create a new HDInsight Cluster
5. The codes are provided below:



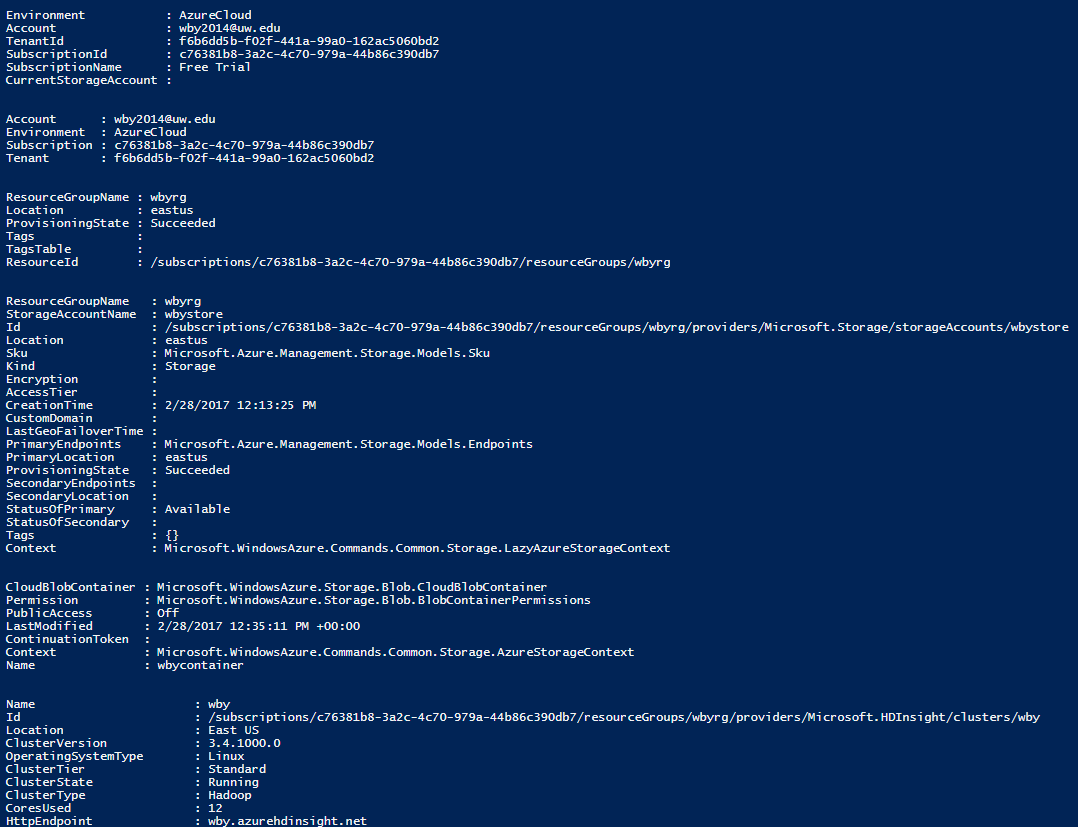
It takes some time (about 20 minutes) to create the cluster

When creating the HDInsight cluster, it will let you type the credentials for Hadoop User Account (**$cluseterCredentials**) and SSH user (**$sshCredentials**)

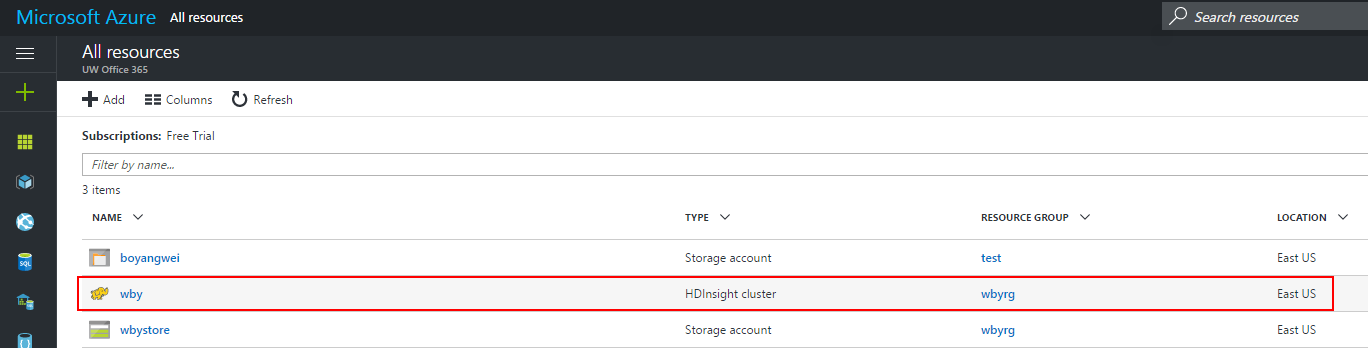
Default name for Hadoop User Account is ‘**admin**’



The results are shown below:

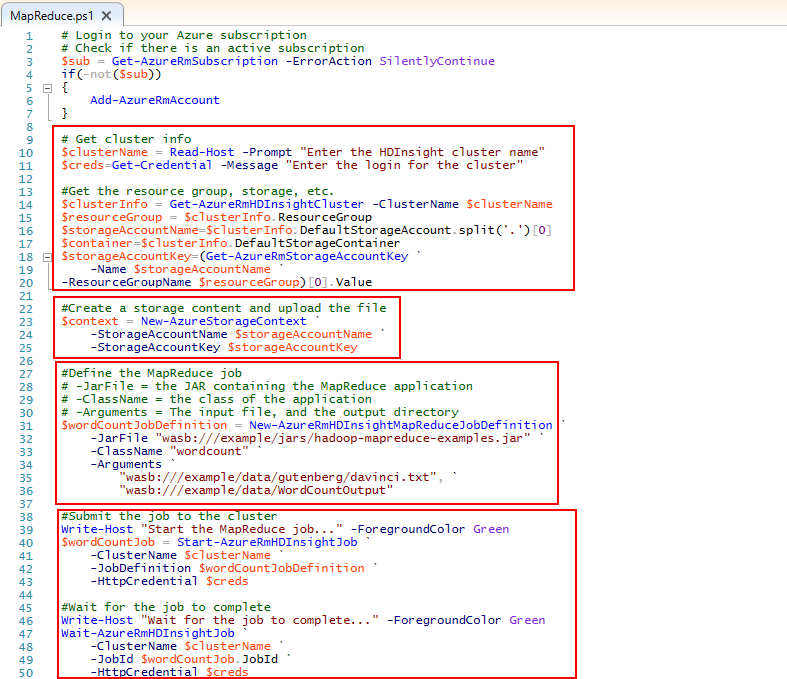


Which can also be seen on the Azure Portal visually (In **All resources**):

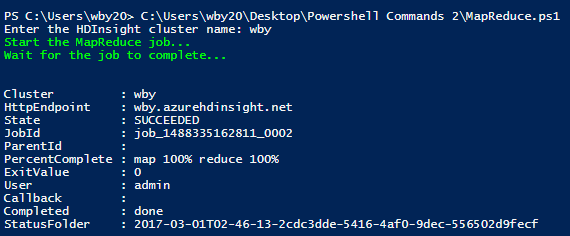


# 2 Submitting a MapReduce job the cluster and executing it

1. Open Windows PowerShell ISE
2. Open a **new file** on Windows PowerShell ISE
3. Login to Azure subscription (For instructions please refer contexts above)
4. Get the information of HDInsight cluster and Resource Group, Storage, etc.
5. Create a storage content and upload the file
6. Create a MapReduce job by selecting the jar file that contains MapReduce Application
7. Specify the class of the application and the input file and output directory (In my case, I chose to save the output file under my user’s address on my PC)
8. Submit the job to cluster



The output is shown below:

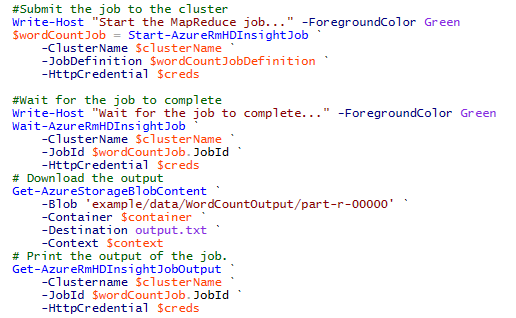


Now Let’s see how we can transfer the output into our blob and download it

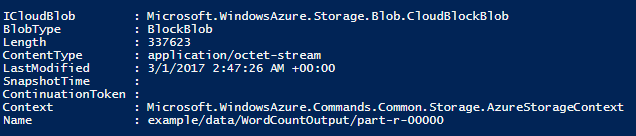
# 3 Transferring results of the job to a blob storage or ADLS storage

1. Open Windows PowerShell ISE
2. Open a **new file** on Windows PowerShell ISE
3. Login to Azure subscription (For instructions please refer contexts above)
4. Continue from completion of MapReduce job

Download the output by using Get-AzureStorageBlobContent and specify the destination of the output file

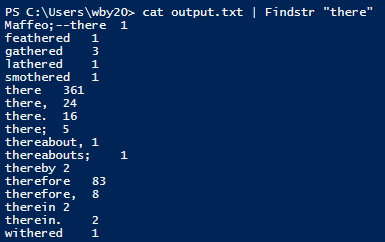


The output is shown below:



Now I have the output file on my local PC address, let us test one of the function with it

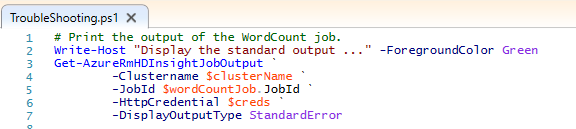
For example, see how many String ‘there’ is in our test file: (By using **Findstr**) and here is the result:



Trouble-shooting Part:

If no information is returned when the job completes, an error may have occurred during processing. To view error information for this job:

* + 1. Open Windows PowerShell ISE
    2. Open a new file in Windows PowerShell ISE
    3. Use Get-AzureRmHDInsightJobOutput command and type in the cluster name, etc.



# 4 Delete the HDInsight cluster

Because HDInsight cluster billing starts once a cluster is created and stops when the cluster is deleted. Billing is pro-rated per minute, so we should always delete our cluster when it is no longer in use. To do so:

* + 1. Open a new file in Windows PowerShell ISE
    2. Type in following command to remove the current HDInsight cluster (Name of cluster)



It also takes some time to delete HDInsight Cluster

The output is shown below:

