Data Movement: Azure Storage Blobs to Data Lake Store with adlcopy

Updated on: 12/29/2016

# Introduction

# Information to remember

**Use this WASB account for all demos**

sachinstr22wasb

ojT2oWLG8lOFbhNsR9DsyEZTZYRbsh5CqoDuyi5LLvHIOw69ewX0zq1DyffigotvSybnP9X6noGUmCYJVc3vBw==

# Copying Data to the ADLS

* Copy the file: BigDataTweets-PowerShell.csv located in the sharepoint folder to your local disk.
* Start PowerShell
* Execute the following commands

Login-AzureRmAccount -SubscriptionId d66b1168-d835-4066-8c45-7d2ed713c082

Test-AzureRmDataLakeStoreAccount –Name sachinstr22adls

Get-AzureRmDataLakeStoreChildItem -AccountName sachinstr22adls -Path /

Import-AzureRmDataLakeStoreItem -AccountName sachinstr22adls -Path “c:\users\sachins\documents\misc\demo\BigDataTweets-PowerShell.csv” -Destination \demo\BigDataTweets-PowerShell.csv

* Using the Azure Portal, show that the file has been copied.

# Installing ADLCopy

Download and install Adlcopy from [here](https://www.microsoft.com/en-us/download/details.aspx?id=50358).

Open command prompt and navigate to the directory where Adlcopy is installed.

Default location is %USERPROFILE%\documents\adlcopy

Adlcopy /source <https://sachinstr22wasb.blob.core.windows.net/copytoadls/BigDataTweets-AdlCopy.csv> /dest swebhdfs://sachinstr22adls.azuredatalakestore.net/starbucks/ /sourcekey ojT2oWLG8lOFbhNsR9DsyEZTZYRbsh5CqoDuyi5LLvHIOw69ewX0zq1DyffigotvSybnP9X6noGUmCYJVc3vBw==

Adlcopy /source <https://sachinstr22wasb.blob.core.windows.net/copytoadls/BigDataTweets-AdlCopyAdla.csv> /dest swebhdfs://sachinstr22adls.azuredatalakestore.net/demo/ /sourcekey ojT2oWLG8lOFbhNsR9DsyEZTZYRbsh5CqoDuyi5LLvHIOw69ewX0zq1DyffigotvSybnP9X6noGUmCYJVc3vBw== /account sachinstr22adla /units 5

Using the Azure Portal, show that the file has been copied.

**Distcp via rdp**

For this you will need to rdp into an hdi cluster and run the following command.

The cluster to connect to is sachinstr23hdi

Account name is sachinstr23hdirdp

The rdp account password is: !P@ssword1

Run the following commands on the hadoop commandline.

hdfs dfs –ls [wasb://copytoadls@sachinstr22wasb.blob.core.windows.net/](mailto:wasb://copytoadls@sachinstr22wasb.blob.core.windows.net/)

hdfs dfs –ls adl://sachinstr22adls.azuredatalakestore.net:443/

Hadoop distcp [wasb://copytoadls@sachinstr22wasb.blob.core.windows.net/BigDataTweets-DistCp-rdp.csv](mailto:wasb://copytoadls@sachinstr22wasb.blob.core.windows.net/BigDataTweets-DistCp.csv) adl://sachinstr22adls.azuredatalakestore.net:443/

Using the Azure Portal, show that the file has been copied.

**Distcp via ssh**

For this you will need to ssh into an hdi cluster and run the following command.

The cluster to connect to is sachinstr222hdi using puTTy and the following string: sachinstr222hdi-ssh.azurehdinsight.net

Account name is sachinstr222hdissh

The rdp account password is: Password!123

Run the following commands on the hadoop commandline.

hdfs dfs –ls [wasb://copytoadls@sachinstr22wasb.blob.core.windows.net/](mailto:wasb://copytoadls@sachinstr22wasb.blob.core.windows.net/)

hdfs dfs –ls adl://sachinstr22adls.azuredatalakestore.net:443/

Hadoop distcp [wasb://copytoadls@sachinstr22wasb.blob.core.windows.net/BigDataTweets-DistCp-ssh.csv](mailto:wasb://copytoadls@sachinstr22wasb.blob.core.windows.net/BigDataTweets-DistCp.csv) adl://sachinstr22adls.azuredatalakestore.net:443/

Using the Azure Portal, show that the file has been copied.