Data Storage using Azure Data Lake Store gen1 & Blob Storage

Lab 2 – Data Ingestion & storage using Azure Data Lake Store gen1 & Blob.

Overview

In this lab, you will understand the installation steps of AdlCopy , AzCopy & data ingestion techniques from Blob storage to ADLS using AzCopy, Powershell/Azure Portal.

What You’ll Need

To complete the labs, you will need the following:

* A web browser
* A Microsoft account
* A Microsoft Azure subscription
* A Microsoft Windows, Linux, or Apple Mac OS X computer on which the Azure CLI has been installed.
* Azure PowerShell

The lab files for this course

## Prerequisites:

* For AzCopyV10, no installation is required & can start directly from command line with azcopy.exe executable is located. You can add the AzCopy folder location to your system path.
* For AzCopy v0.8 you need to install the tool from this link: <https://aka.ms/downloadazcopy>

## Commands / Script for installation of AdlCopy

* From your browser, go to this link to download AdlCopy- <https://www.microsoft.com/en-us/download/details.aspx?id=50358>
* Install AdlCopy by accepting License Agreement & selecting installation folder & click next to complete the installation.
* Move to Windows Environment Variables section, Under ‘Edit environment variable’ , click ‘New’.
* In the text box, type %HOMEPATH%\Documents\AdlCopy, and then click OK.
* In the Environment Variables dialog box, Click OK.
* In the System Properties dialog box, click OK.

## Upload a set of files to Azure Data Lake Store from Azure Blob Storage

* From Azure Portal, Click +Create a resource, click Storage & then select Storage account-blob,file,table,queue.
* On the Create storage account blade, Create a storage account with the Resource Group -AzDataLakeRG.
* In the \*\*Name box, type blobstore<yourblobname><date>
* In the Location field, select the same location as you’ll use for Azure Data Lake storage account.
* In the Account kind list, select Blob storage
* Finally review all blob storage account provisioning details (optionally with Azure Resource manager JSON templates) & click Create.
* Next, create Azure Data Lake Store account from portal, from +Create a Resource, click Storage & select Data Lake Storage.
* Select New Data Lake Storage Gen1 blade, in the Name field, provide dlstore<youradlsname><date>
* Select the same Azure resource group ‘AzDataLakeRG’ & provide your nearest & same location of previously created Blob storage account.
* Finally review the ADLS resource provisioning template & click on Create.

### Upload data to Azure Blob Storage & Transfer to Azure Data Lake Store

* From ‘All resources’ option under Azure Portal, select your just created Blobstore<yourname><date>.
* Next, in the settings , select New Container dialog & provide name of the Blob container (for e.g. ‘datafeed’) & click OK.
* Next, copy the ‘Access Keys’ of the storage account under Settings , Copy the Key1 & paste into your notepad.
* Download the Lab2 ‘Sample data’ from the OneDrive link & upload to Blob storage by using the following command in command prompt by using AzCopy.

azcopy /Source:"<yourdrive> \TestdataTree" /Dest:https://<Your Azure storage account name>.blob.core.windows.net/datafeed /DestKey:<access key> /S

* Verify the uploaded blob storage contents by using the tools like Azure Storage explorer/cloud Xplorer etc. or from Visual Studio Cloud Explorer tool.

### Use AdlCopy to transfer files from Azure Blob Storage to ADLS

* Open your command prompt (preferred with elevated rights)
* Type the following script to transfer data from blob(source) to ADLS(destination)

adlcopy /source https://<Your Azure storage account name>.blob.core.windows.net/datafeed/ /dest adl://&lt;name of your Data Lake Store&gt;.azuredatalakestore.net/Copytest/ /sourcekey<access ke>

Make sure you create a folder under your ADLS named as Copytest before execution of the above script.

* Verify with Azure portal , all data files from Azure Blob Storage transferred to ADLS.