How To Access ADLS APIs and Integrate in CDI/CAI

**Draft 1.0**

**Author: Amar Kumar Srivastava**

[**amsrivastava@informatica.com**](mailto:amsrivastava@informatica.com)

How-To Guide

# TABLE OF CONTENTS

[1 TABLE OF CONTENTS 2](#_Toc36475639)

[2 Document purpose 3](#_Toc36475640)

[3 Solution 3](#_Toc36475641)

# Document purpose

Describe how to access Azure Data Lake Services (ADLS) Gen2 APIs to perform some basic operations. This includes setting up the configuration within Azure Portal.

File related operations stored in ADLS

**Azure**

**IICS CDI**

**Secure Agent**



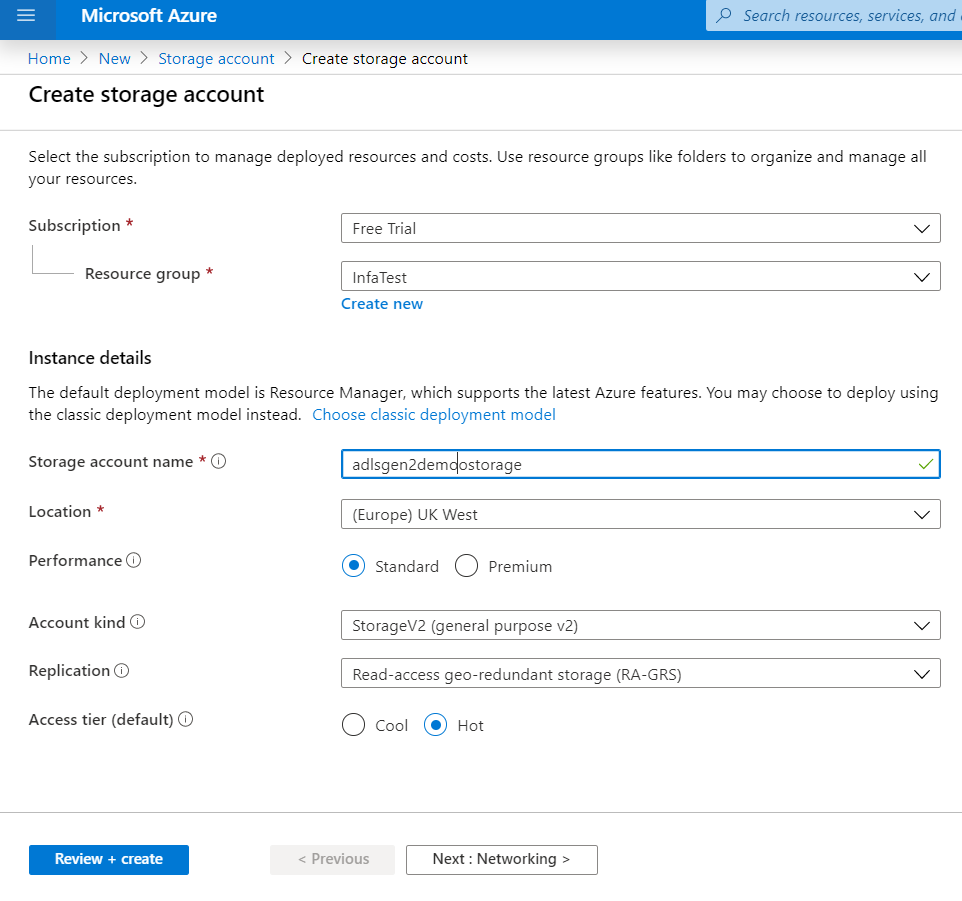
# Solution

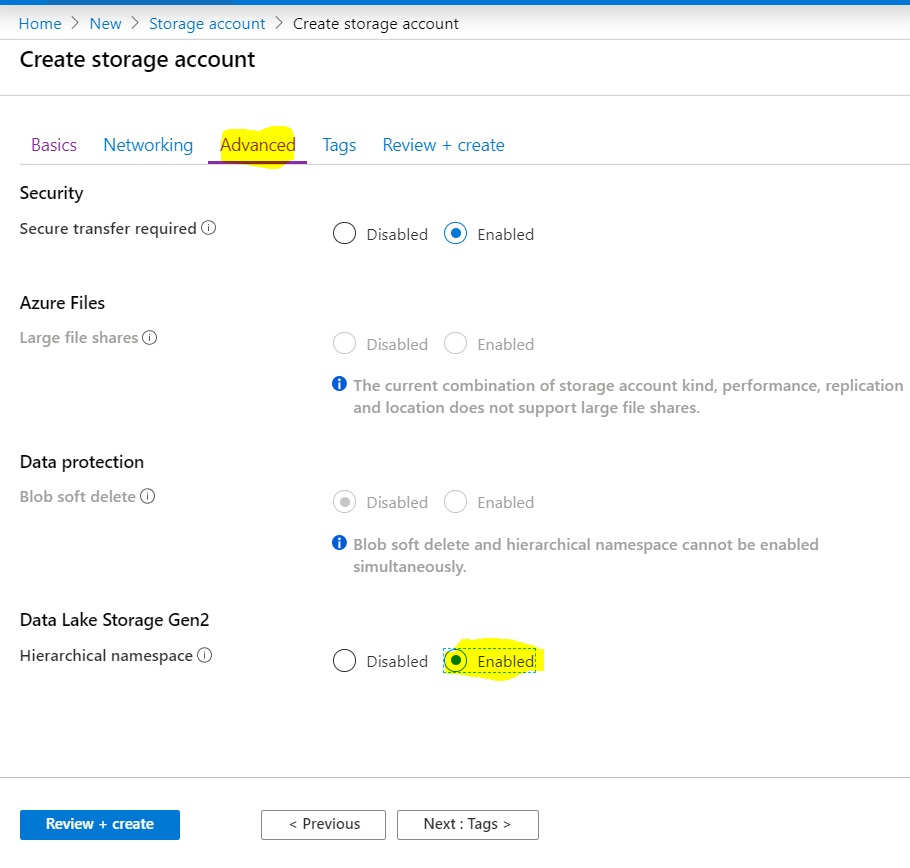
## Analyst

Please find below the steps required to achieve the ADLS file rename (Step 1,2 and 3 is for first time setup). Also included the CURL script for the same, the response token for the first API(Auth) call is the input for the second API(Rename) call and can be included as post script to the CDI task or Service Connector in CAI.

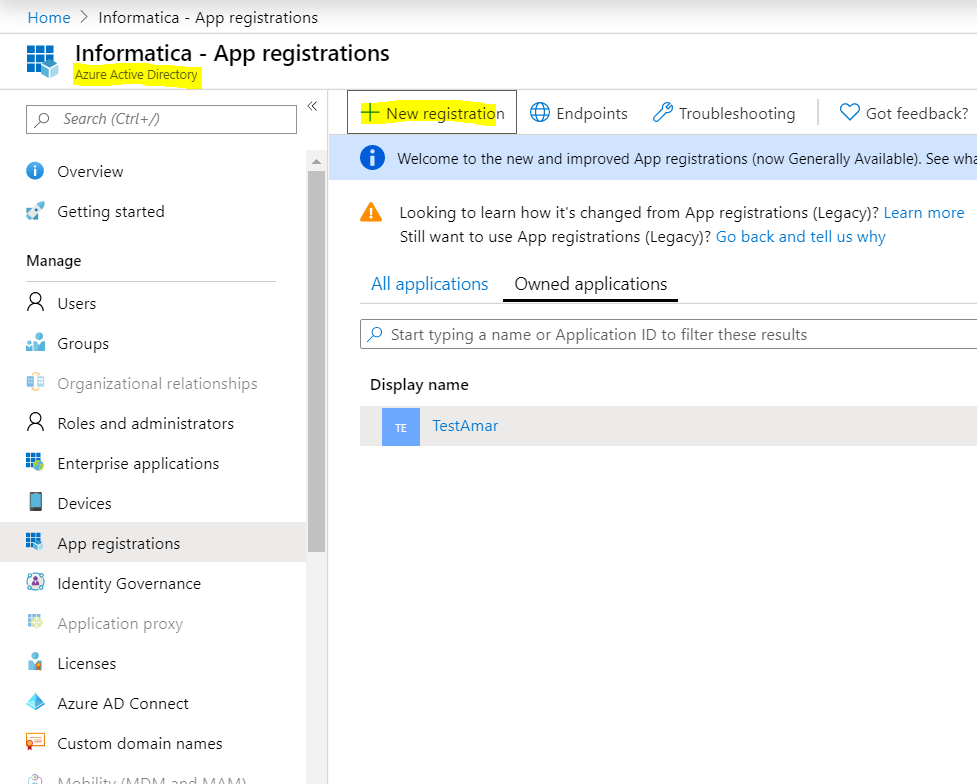
**Steps:**

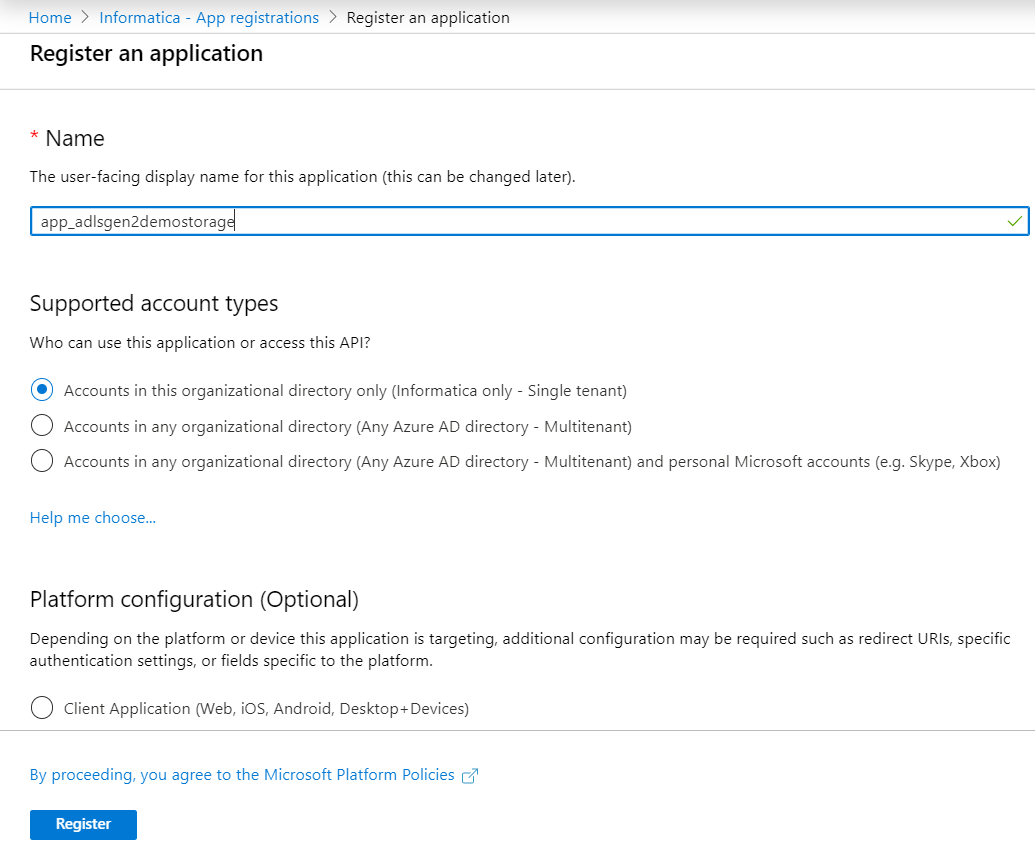
1. Create the ADLS storage and filesystem for example <adlsgen2demostorage> as storage account and <test> as filesystem.



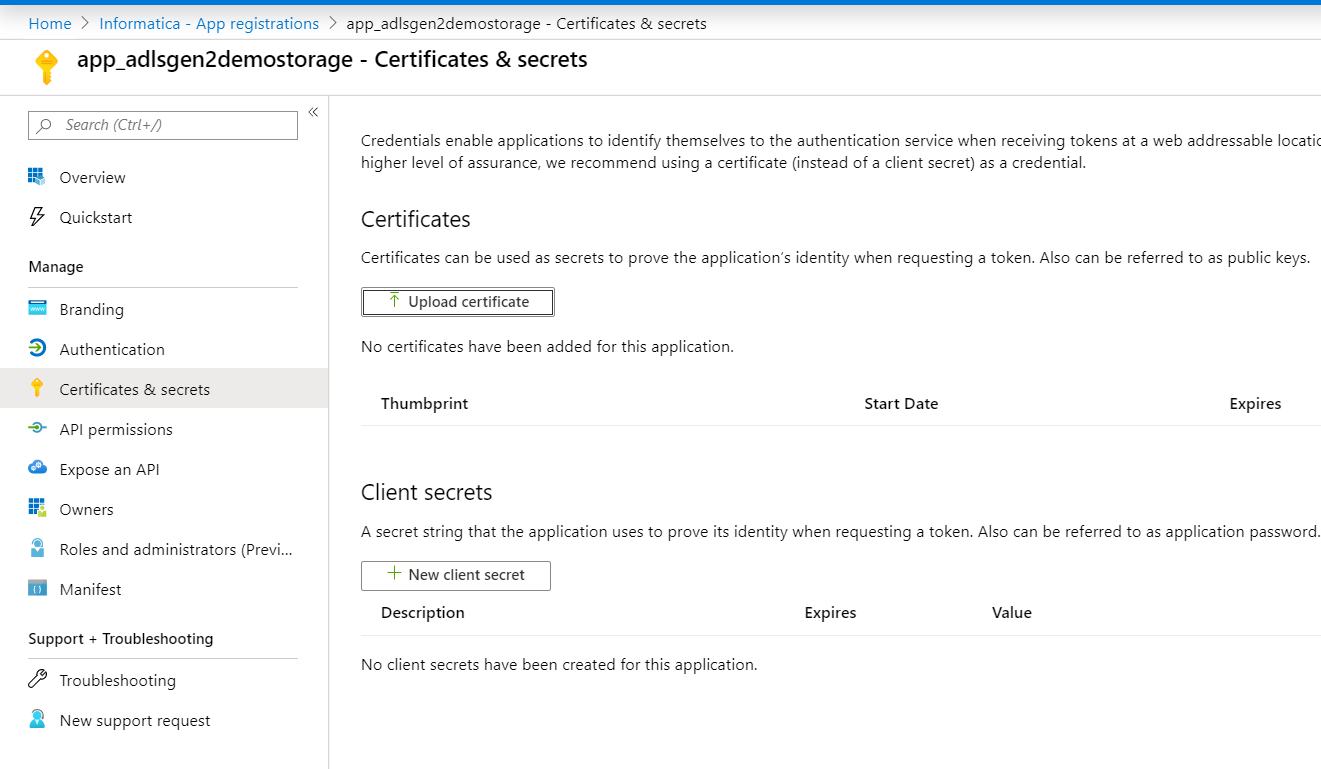


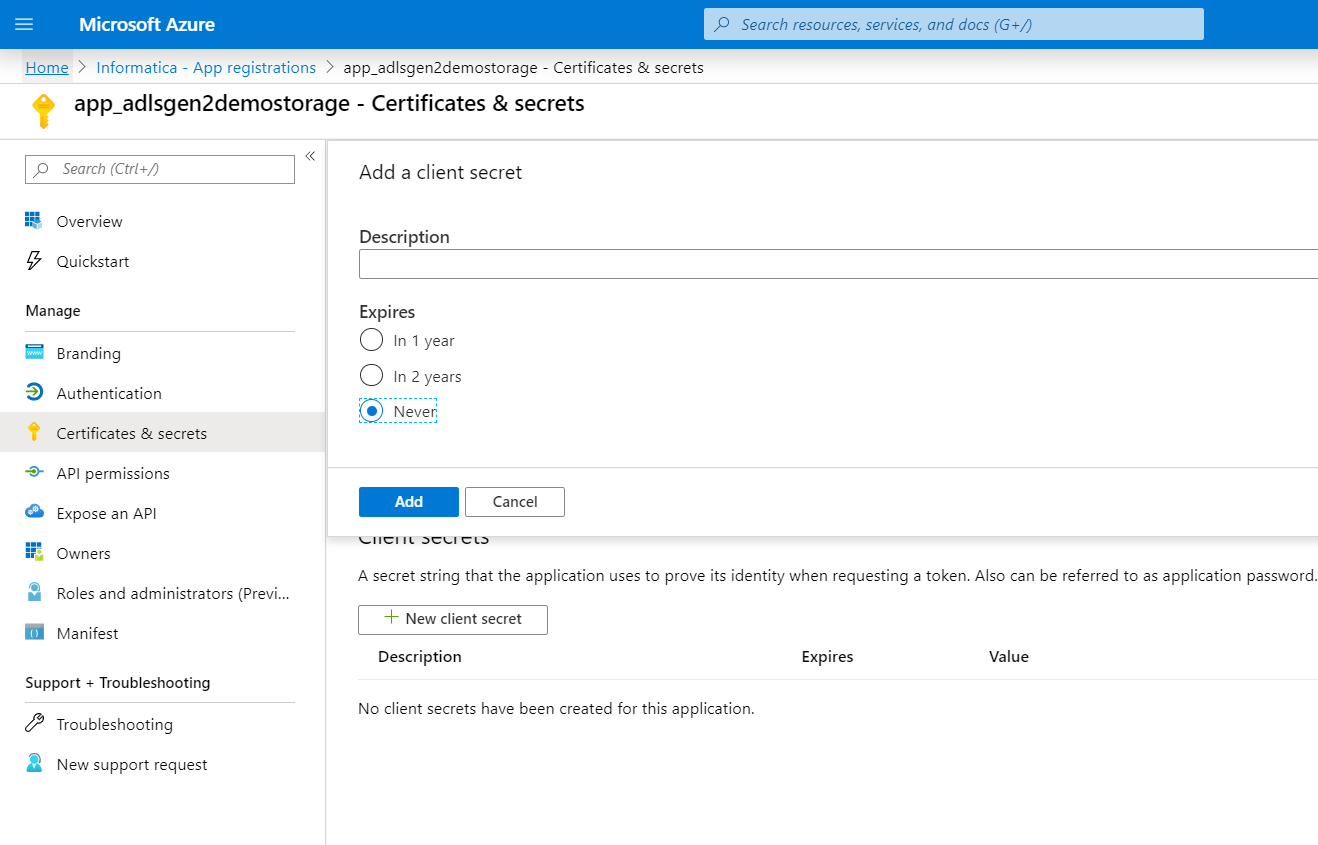
1. Goto App Registrations in Azure Active Directory (Azure Portal).
   * New registration for App, give it any name like <adlsgen2demostorage>, click register.

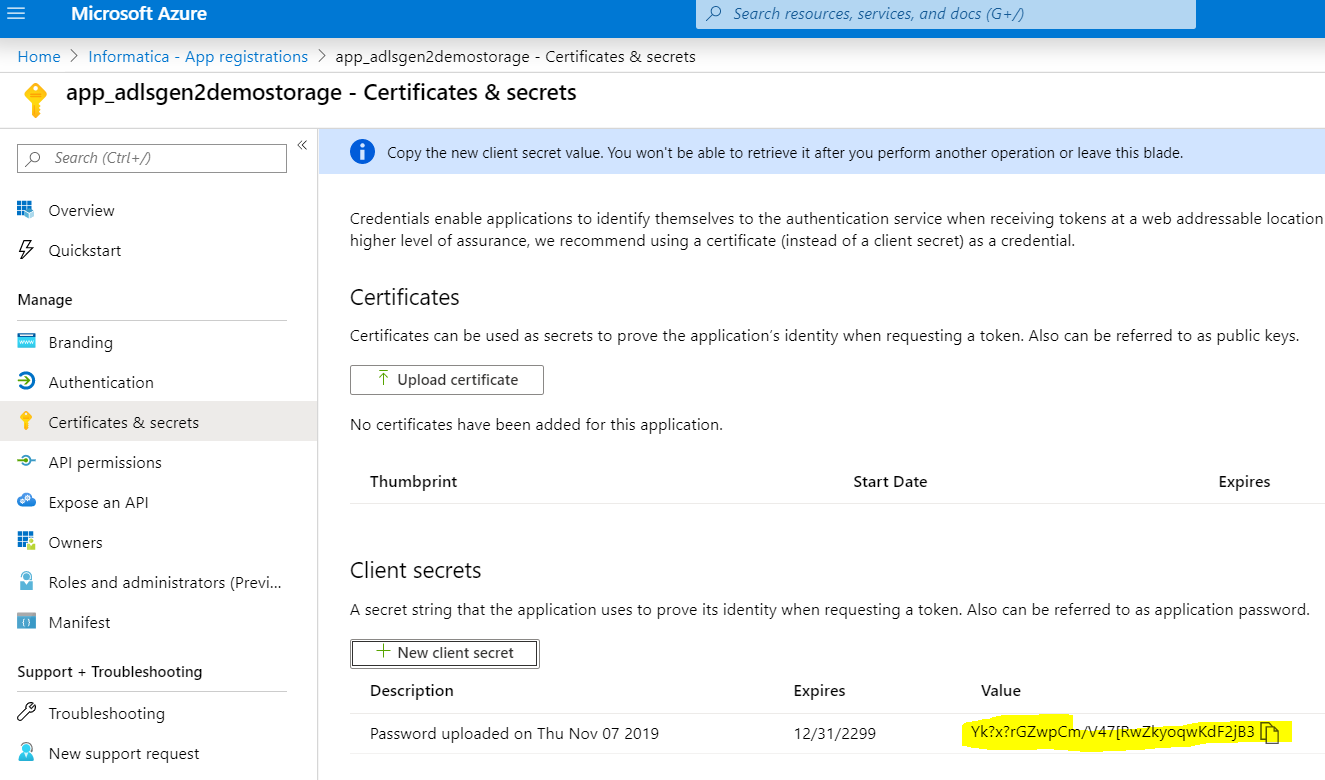




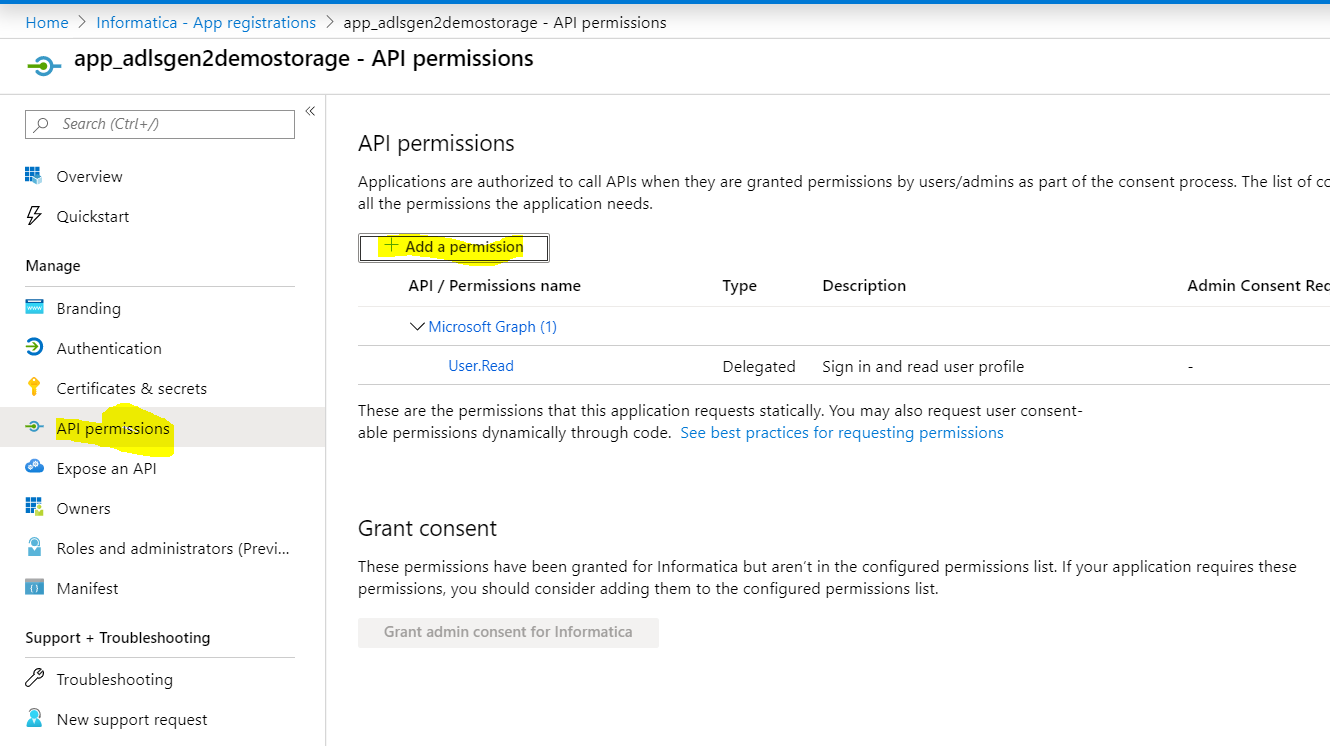
* + Goto newly created App (adlsgen2demostorage)->Certificates and secrets
    - Add a new client secret with never expire. This will generate a key (**client secret**) copy and save it somewhere as this cannot be retrieved anytime later after creation.

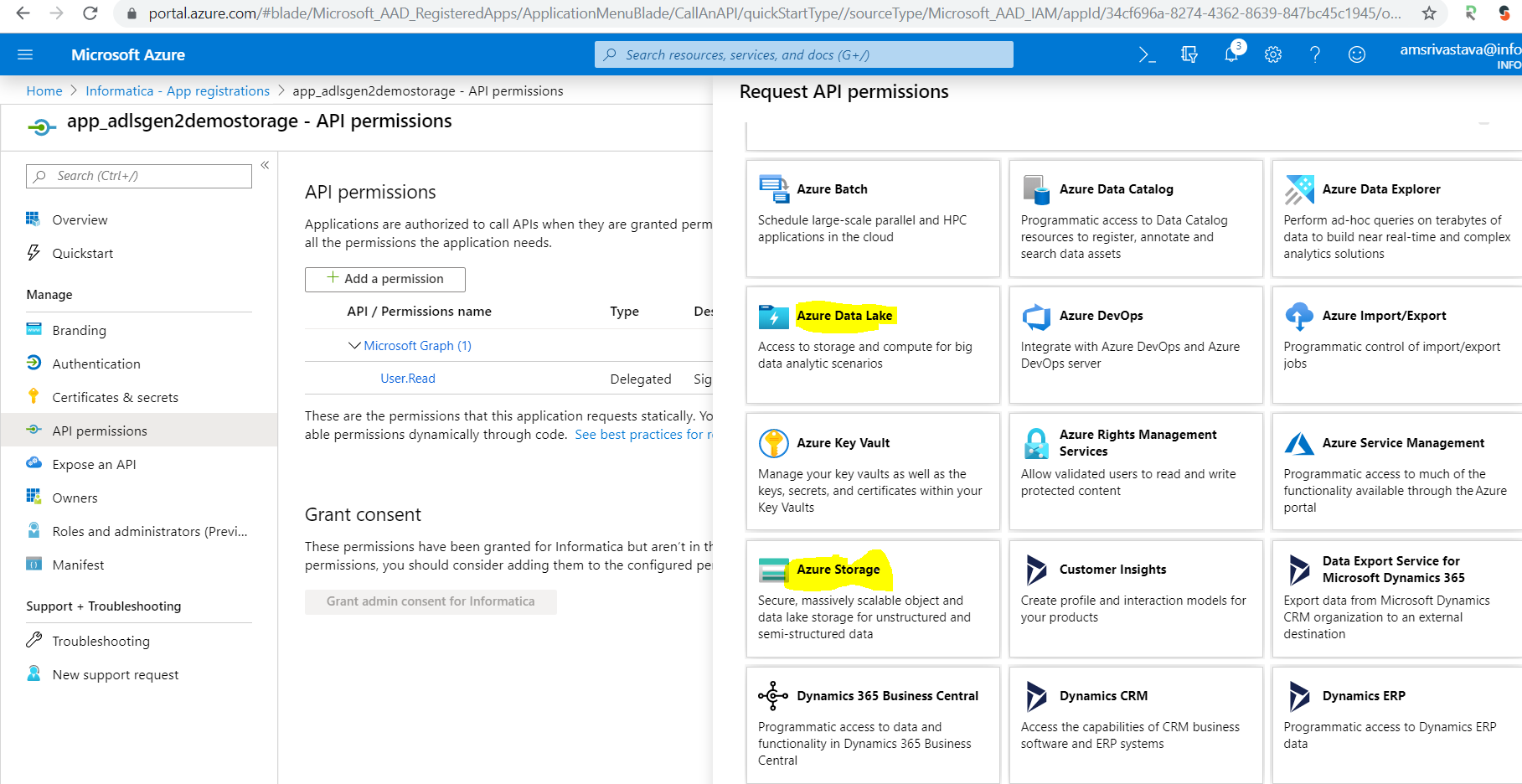




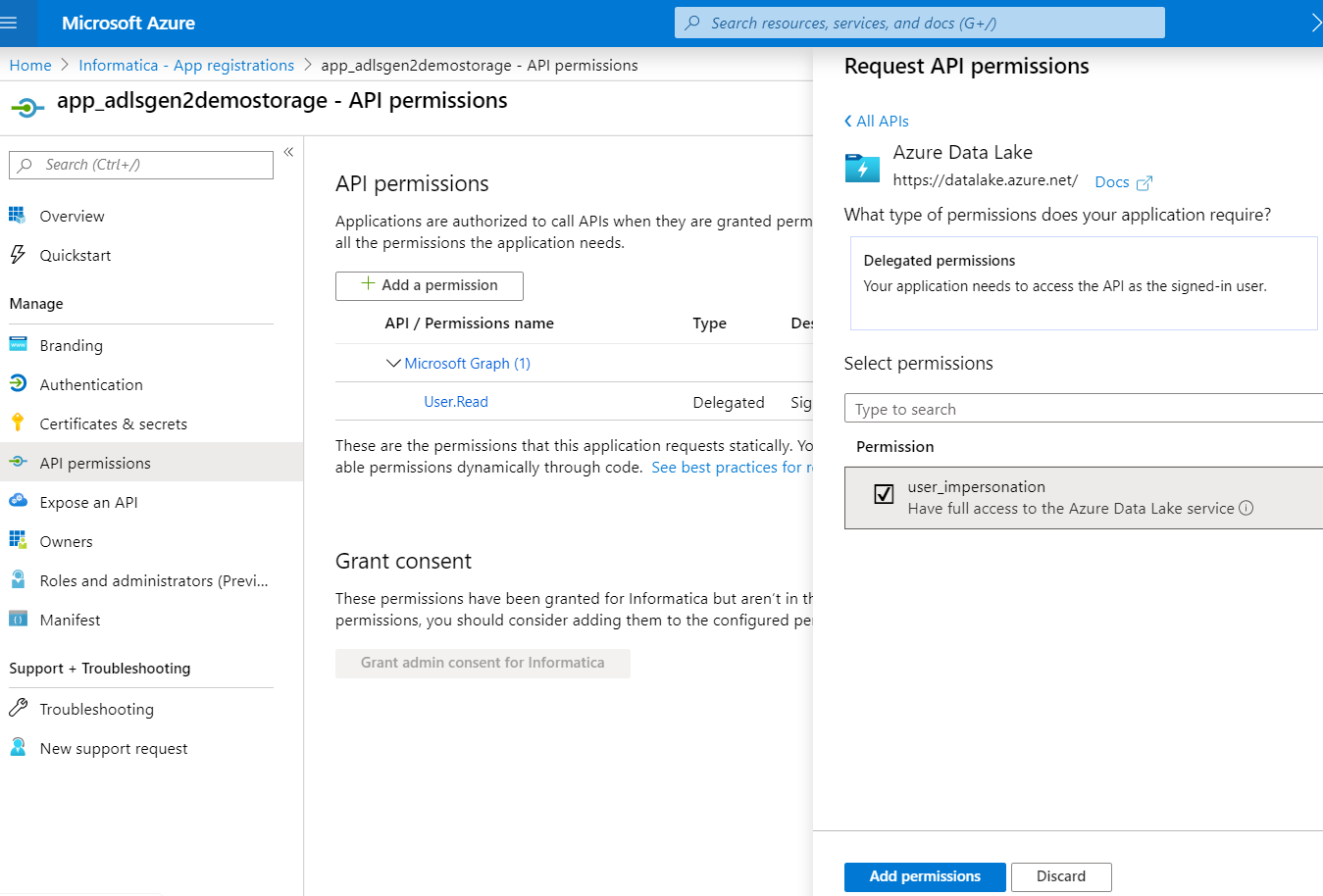


* + - Go to newly created App (adlsgen2demostorage)-> API Permissions -> Click “Add a permission”

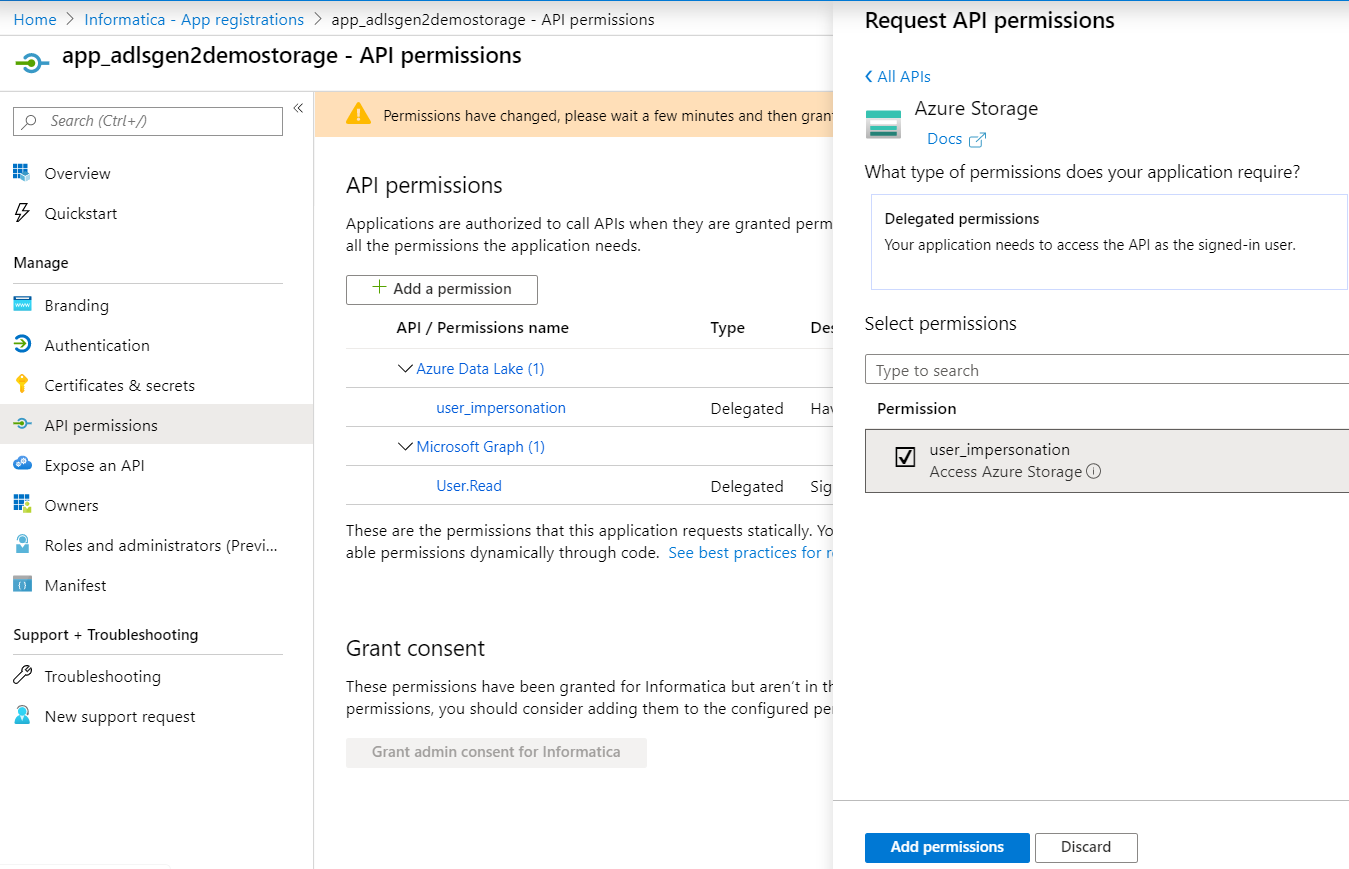




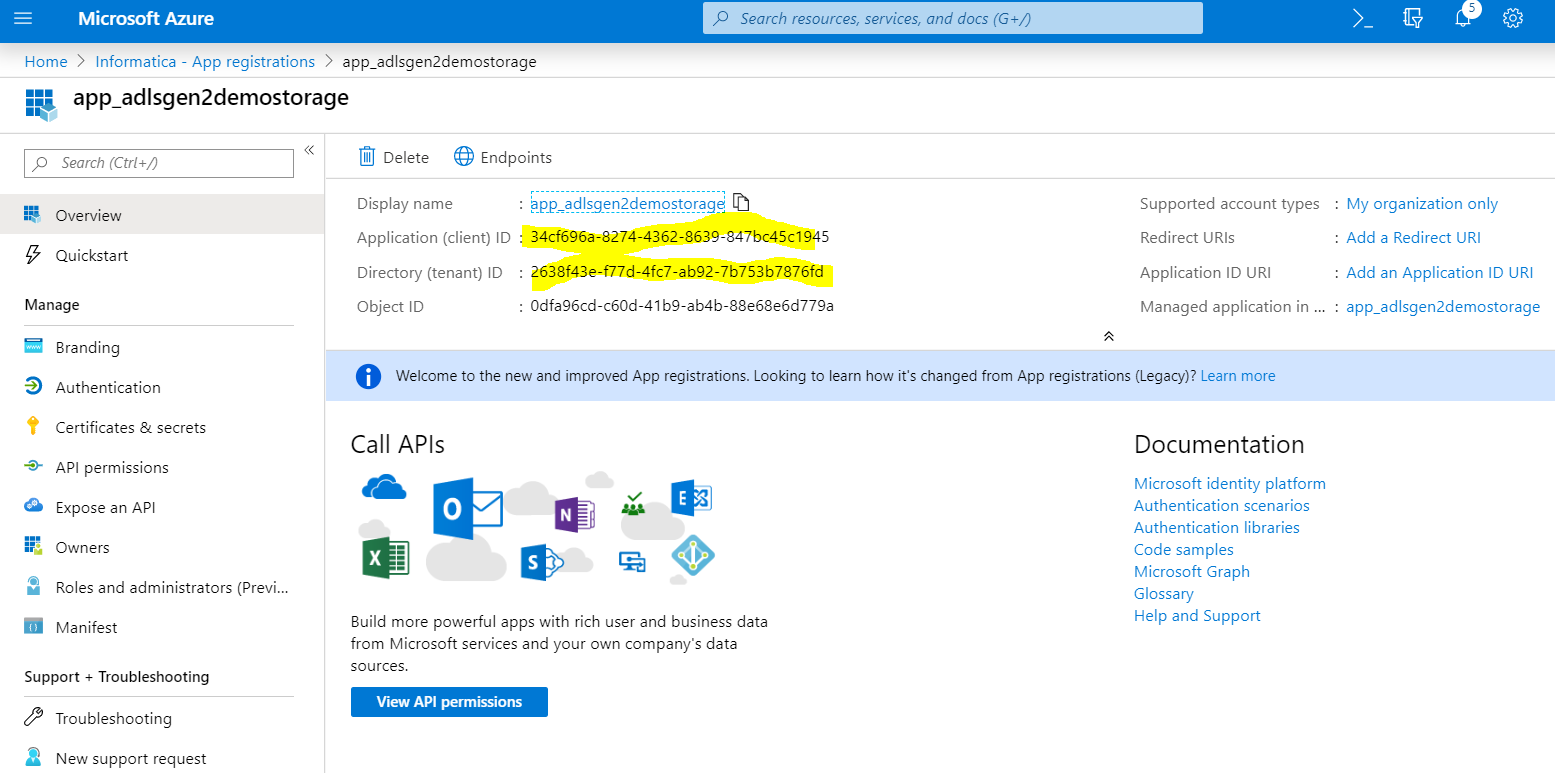
* + - * Add “Azure Data Lake” with user\_impersonation (select check box)



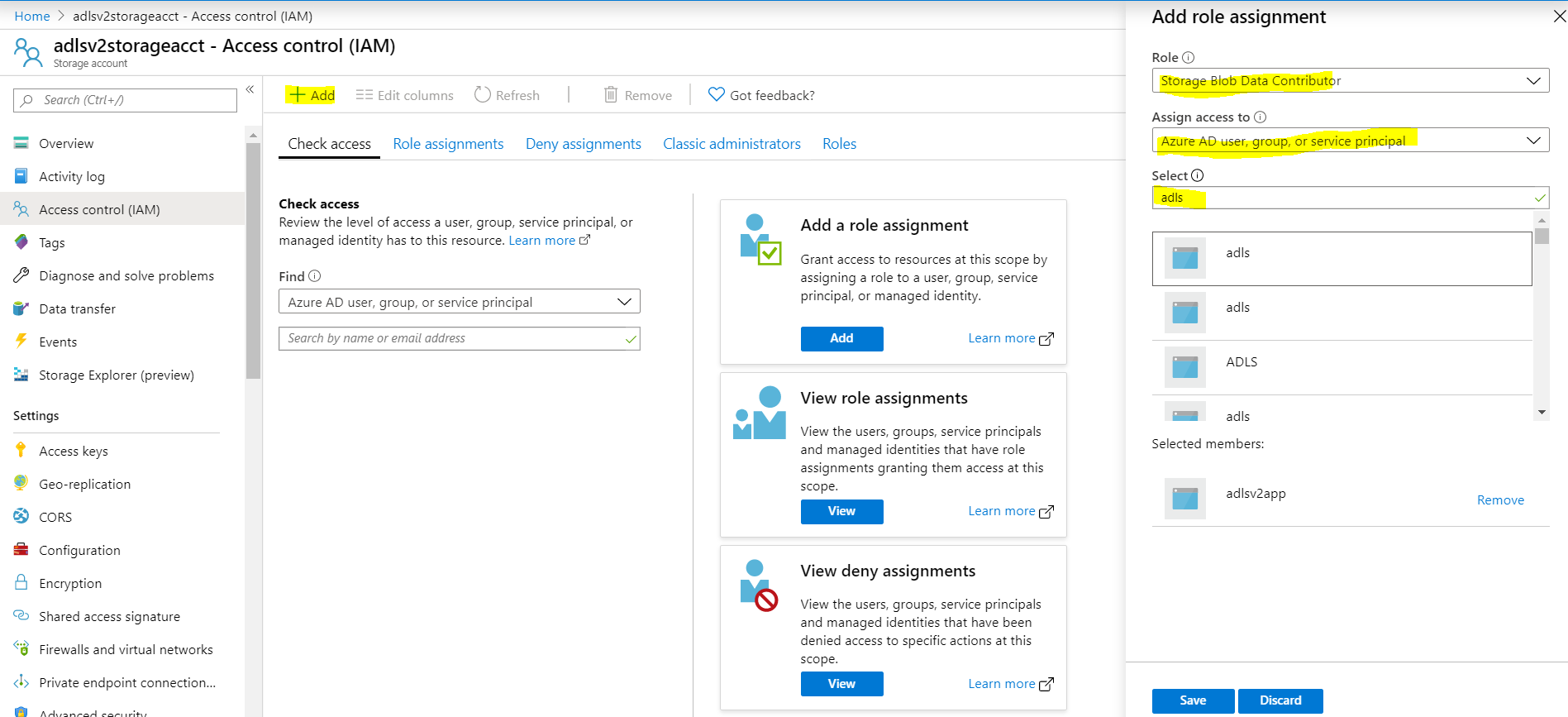
* + - * Add “Azure Storage” with user\_impersonation (select check box)



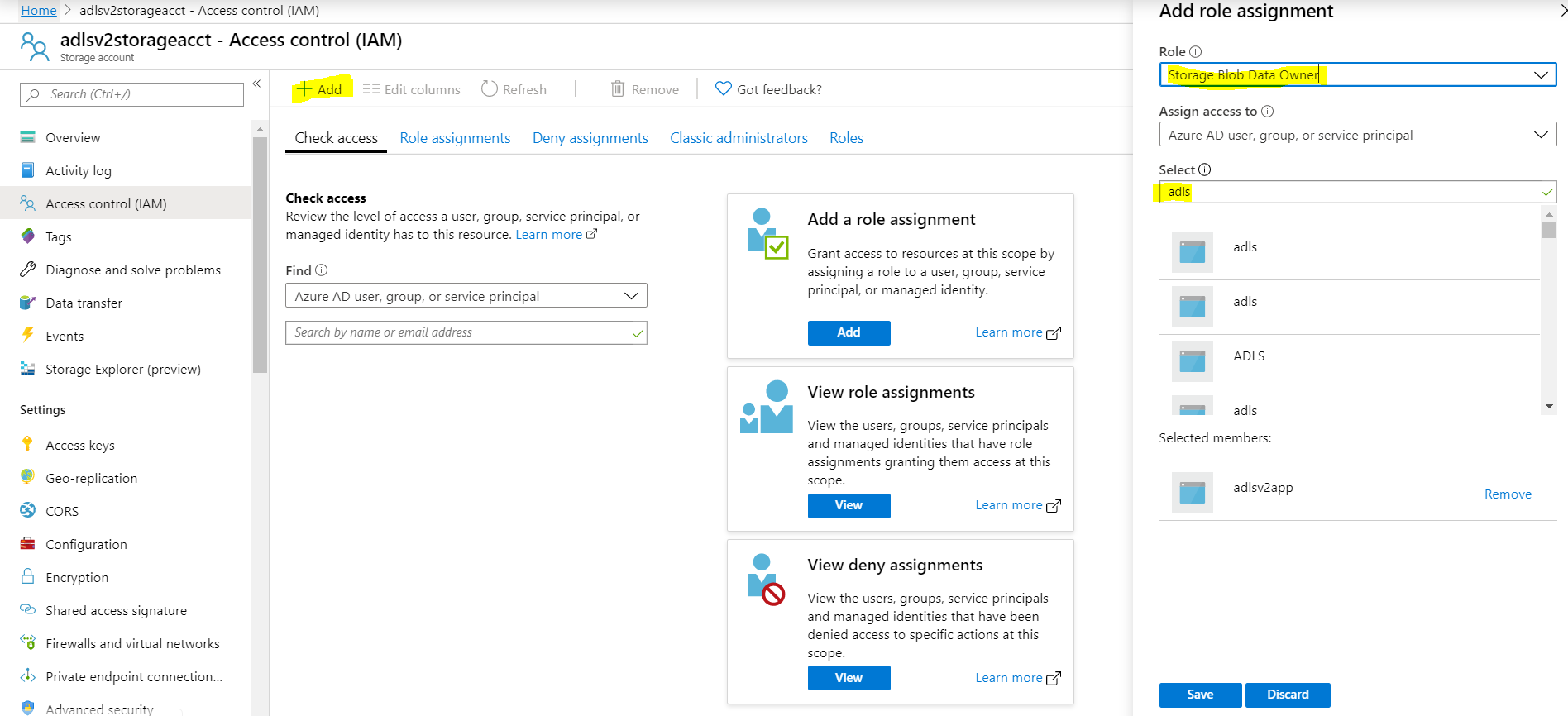
* + - Go back to newly created App (adlsgen2demostorage)-> Overview and copy Application **(client) ID** and Directory **(tenant) ID**.



1. Goto your storage account (adlsgen2demostorage)-> Access Control
   * Add “role assignment” and select
     + Role: Storage Blob Data Contributor
     + Assign access to: Azure AD user, group, or service principal
     + Select: write your storage app name (it will show up in dropdown)
     + Save



* + Add “role assignment” and select
    - Role: Storage Blob Data Owner
    - Assign access to: Azure AD user, group, or service principal
    - Select: write your storage app name (it will show up in dropdown)
    - Save

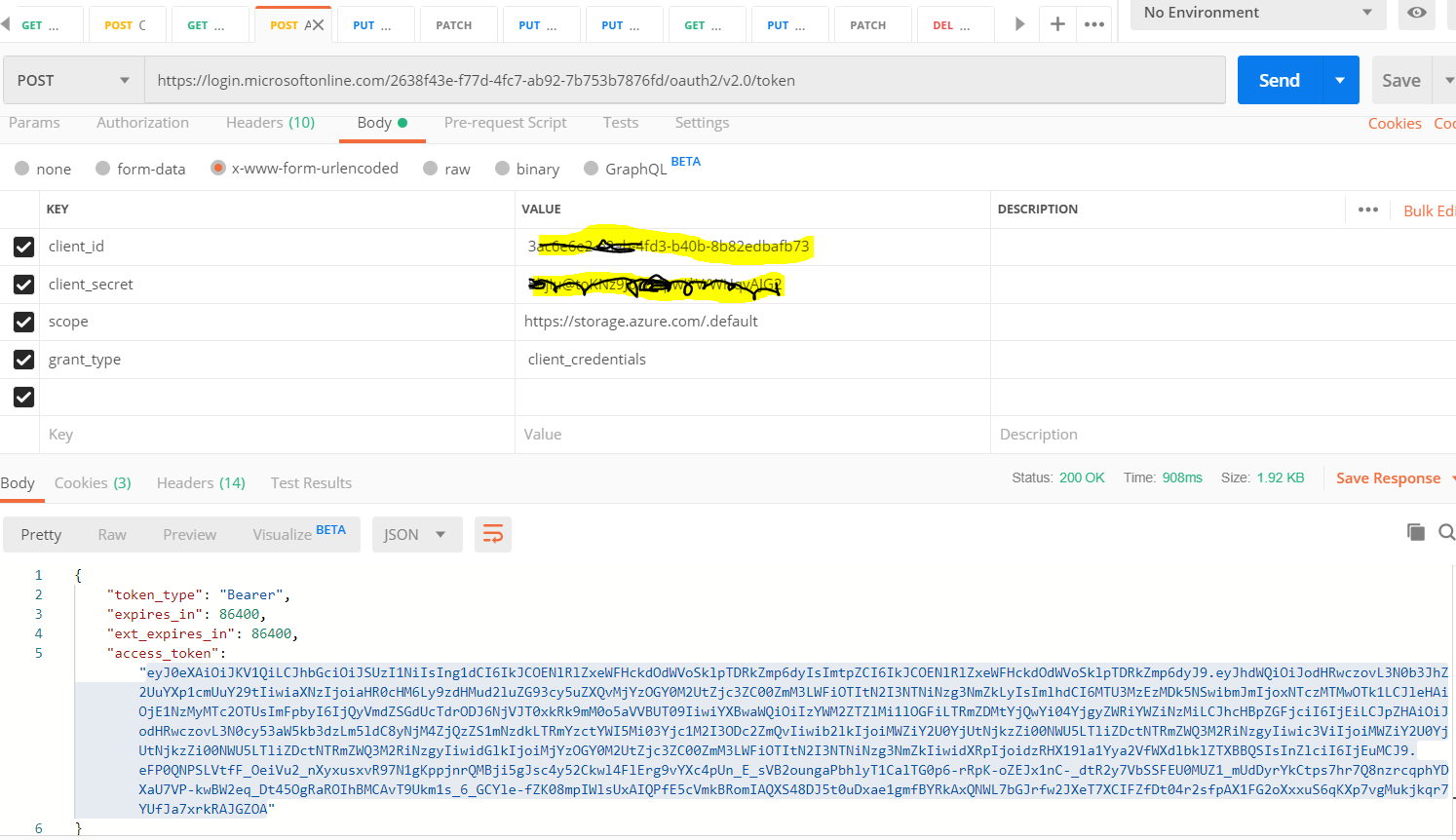


1. Generate a access token using the information collected earlier with a **POST** request with body type x-www-from-urlencoded as

|  |  |
| --- | --- |
| **key** | **Value** |
| client\_id | <Copied above from app registration> |
| client\_secret | <Copied above from app registration as client secret> |
| scope | [https://storage.azure.com/.default](https://nam02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fstorage.azure.com%2F.default&data=02%7C01%7Camsrivastava%40informatica.com%7Caf91fd46f6604d07a3bd08d7639f6db1%7C2638f43ef77d4fc7ab927b753b7876fd%7C0%7C0%7C637087408714643976&sdata=3JlsfdFNIQ%2F%2FtmcdDxRjegw%2BAqAkoFNNplNMXlpQ0ck%3D&reserved=0) |
| grant\_type | client\_credentials |

|  |  |
| --- | --- |
| **URL** | [https://login.microsoftonline.com/**<tenantID>**/oauth2/v2.0/token](https://nam02.safelinks.protection.outlook.com/?url=https%3A%2F%2Flogin.microsoftonline.com%2F%253ctenantID%253e%2Foauth2%2Fv2.0%2Ftoken&data=02%7C01%7Camsrivastava%40informatica.com%7Caf91fd46f6604d07a3bd08d7639f6db1%7C2638f43ef77d4fc7ab927b753b7876fd%7C0%7C0%7C637087408714643976&sdata=Px%2FpEGfHu5F%2FyORTj31HQkdXIbBOfD5V789aMVN9UkQ%3D&reserved=0) (replace tenantID copied earlier) |

|  |
| --- |
| **Response Sample** |
| {      "token\_type": "Bearer",      "expires\_in": 86399,      "ext\_expires\_in": 86399,      "**access\_token**": "eyJ0eXAiOiJKV1QiLCJhbGciOiJSUzI1NiIsIng1dCI6ImFQY3R3X29kdlJPb0VOZzNWb09sSWgydGlFcyIsImtpZCI6ImFQY3R3X29kdlJPb0VOZzNWb09sSWgydGlFcyJ9..M9lUXchAViL2fyFz8LNpH9AKO88ECQGVyhg2DsUMOy-KqwFZEr8huv0abj0\_4N02BPuzQSvluN85I0W-VCF0UF8mpGKKuBjktrjT20nEYLUDhYjT-KN-jVZ6WiUNmO-jGfbOtc-OoZKHR7Or4NFECdXa6QVe1O81s1eI4eqKoS7fx7CKEOKgLATZZ10LIitLW3QxxfWrAqsI1c3S5\_6uawswX\_ZWOdRAFIN3W6oOrDkdcrhHwILpYLlcRJkfmS67d\_7uLwITU97F\_fa-eEthqYCvd6uRALuIRRsA0Uq\_pKvZvNSxBzgmooP7Mimx8lopJe2ZoF2dix6FjXm5J9ujbA"  } |



Copy the access token received (without quotes) as this will be used to authenticate the call going forward.

1. Now run a PUT request to rename the file as below

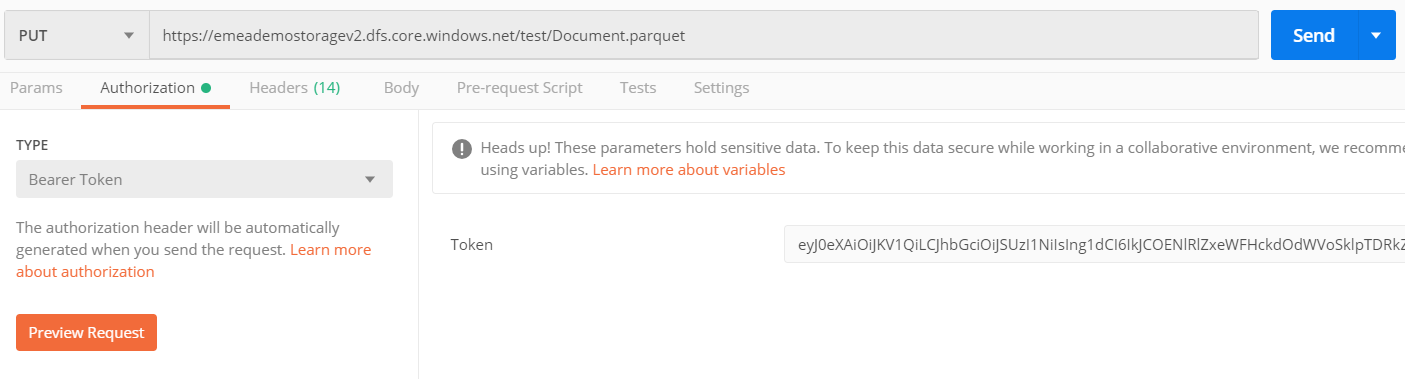
|  |  |
| --- | --- |
| **Command URL** | PUT https://{accountName}.{dnsSuffix}/{filesystem}/{path} |
| **Example URL** | PUT [https://adlsgen2demostorage.dfs.core.windows.net/test/Document.parquet](https://nam02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fadlsgen2demostorage.dfs.core.windows.net%2Ftest%2FDocument.parquet&data=02%7C01%7Camsrivastava%40informatica.com%7Caf91fd46f6604d07a3bd08d7639f6db1%7C2638f43ef77d4fc7ab927b753b7876fd%7C0%7C0%7C637087408714643976&sdata=RF39i1ylyLmFM3si8dWWlNZ4wzR55FCdGcFjcGiD4fQ%3D&reserved=0)  Document.parquet is the **target** filename |

Use Bearer authorization

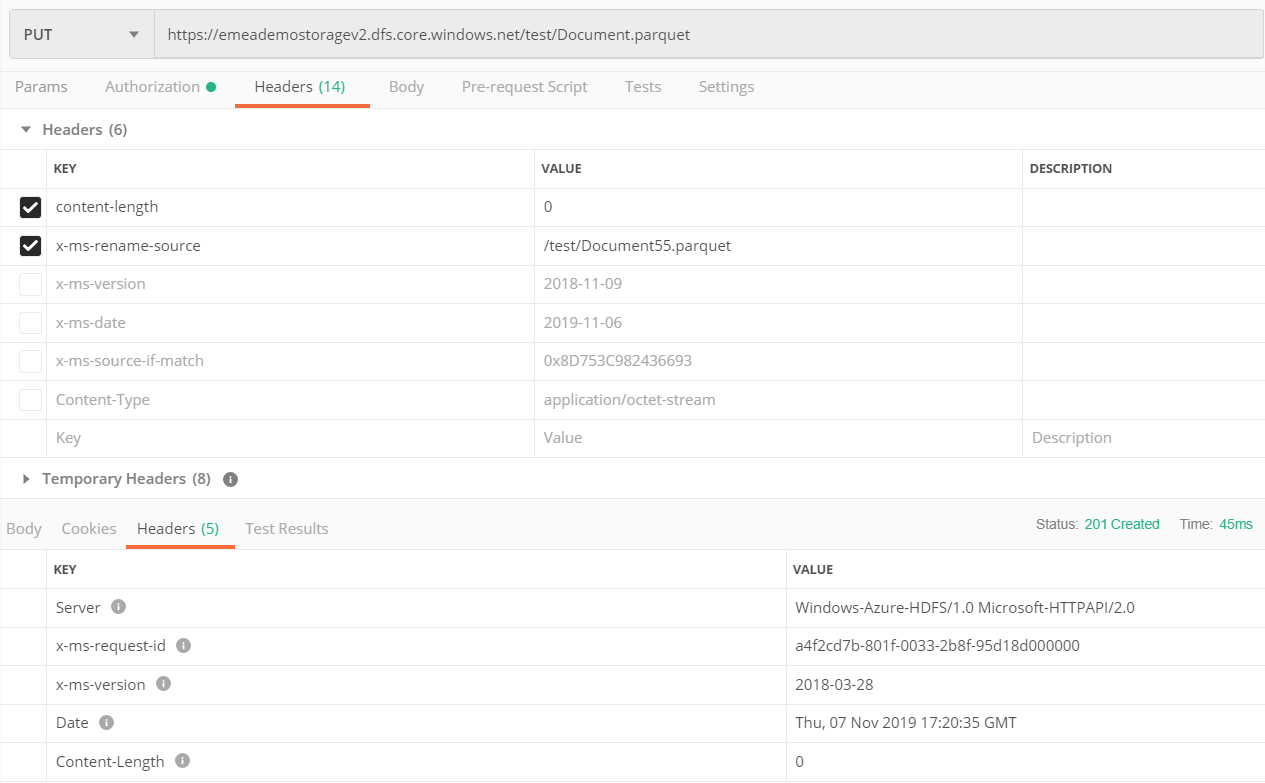
|  |  |
| --- | --- |
| **Authorization** | Bearer <copied token from step 4> |

               Provide the below in the requester header.

|  |  |
| --- | --- |
| **Header Name** | **Header Value** |
| content-length | 0 |
| x-ms-rename-source | /{filesystem}/{path}  For example: /test/Document55.parquet  This is the file expected to be renamed i.e. **source** |

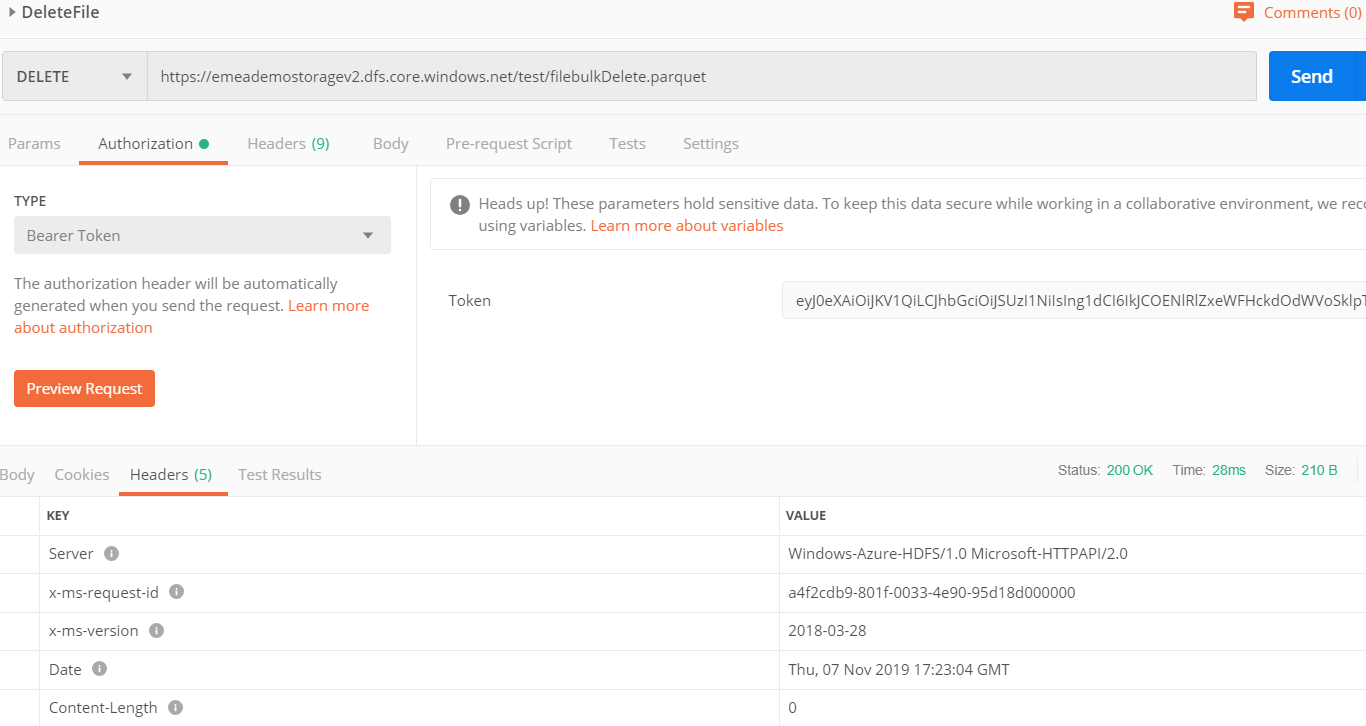


The response will be HTTP code with headers.



Also curl command for the same:

|  |
| --- |
| curl -X POST \  [https://login.microsoftonline.com/2638f43e-f77d-4fc7-ab92-7b753b7876fd/oauth2/v2.0/token](https://nam02.safelinks.protection.outlook.com/?url=https%3A%2F%2Flogin.microsoftonline.com%2F2638f43e-f77d-4fc7-ab92-7b753b7876fd%2Foauth2%2Fv2.0%2Ftoken&data=02%7C01%7Camsrivastava%40informatica.com%7Caf91fd46f6604d07a3bd08d7639f6db1%7C2638f43ef77d4fc7ab927b753b7876fd%7C0%7C0%7C637087408714653932&sdata=D06rUHQef%2F0b2TchhaE6hk0fWlsGm5X1%2BzjEckbuu8A%3D&reserved=0) \    -H 'Accept: \*/\*' \    -H 'Accept-Encoding: gzip, deflate' \    -H 'Cache-Control: no-cache' \    -H 'Connection: keep-alive' \    -H 'Content-Length: 193' \    -H 'Content-Type: application/x-www-form-urlencoded' \    -H 'Cookie: x-ms-gateway-slice=prod; stsservicecookie=ests; fpc=Ap0Y1A-XGNtLvj5Yw\_STvaQ' \    -H 'Host: login.microsoftonline.com' \    -H 'cache-control: no-cache' \    -d 'client\_id=%203ac6e6e2-e8ab-4fd3-b40b-8b82edbafb75&client\_secret=%205hjIu%40toKNz9J0%2FT%5B4pwj\*V%2FWHqvAJG5&scope=https%3A%2F%2Fstorage.azure.com%2F.default&grant\_type=%20client\_credentials&=' |
| curl -X PUT \  [https://emeademostoragev2.dfs.core.windows.net/test/Document55.parquet](https://nam02.safelinks.protection.outlook.com/?url=https%3A%2F%2Femeademostoragev2.dfs.core.windows.net%2Ftest%2FDocument55.parquet&data=02%7C01%7Camsrivastava%40informatica.com%7Caf91fd46f6604d07a3bd08d7639f6db1%7C2638f43ef77d4fc7ab927b753b7876fd%7C0%7C0%7C637087408714653932&sdata=jDe6VWaeaqVwv7D7Q5zBywVVd5gOjM4IFqHUPJguLe0%3D&reserved=0) \    -H 'Accept: \*/\*' \    -H 'Accept-Encoding: gzip, deflate' \    -H 'Authorization: Bearer eyJ0eXAiOiJKV1QiLCJhbGciOiJSUzI1NiIsIng1dCI6ImFQY3R3X29kdlJPb0VOZzNWb09sSWgydGlFcyIsImtpZCI6ImFQY3R3X29kdlJPb0VOZzNWb09sSWgydGlFcyJ9..M9lUXchAViL2fyFz8LNpH9AKO88ECQGVyhg2DsUMOy-KqwFZEr8huv0abj0\_4N02BPuzQSvluN85I0W-VCF0UF8mpGKKuBjktrjT20nEYLUDhYjT-KN-jVZ6WiUNmO-jGfbOtc-OoZKHR7Or4NFECdXa6QVe1O81s1eI4eqKoS7fx7CKEOKgLATZZ10LIitLW3QxxfWrAqsI1c3S5\_6uawswX\_ZWOdRAFIN3W6oOrDkdcrhHwILpYLlcRJkfmS67d\_7uLwITU97F\_fa-eEthqYCvd6uRALuIRRsA0Uq\_pKvZvNSxBzgmooP7Mimx8lopJe2ZoF2dix6FjXm5J9ujbA' \    -H 'Cache-Control: no-cache' \    -H 'Connection: keep-alive' \    -H 'Host: emeademostoragev2.dfs.core.windows.net' \    -H 'cache-control: no-cache' \    -H 'content-length: 0' \    -H 'x-ms-rename-source: /test/Document.parquet' |



## Analyst

**Requirement:**

1. ADLS will be having 2 Files Document\_Merge.parquet and Document\_Total.parquet.
2. Delete existing Document\_Total.parquet.
3. Rename Document\_Merge.parquet to Document\_Total.parquet.

**Solution:**

* Once everything will be setup in customer’s account as per the steps mentioned yesterday,  the script can be run as post processing command.
* On the machine this script will be executed (SA) we need to install **jq** as the script uses it to extract token from a JSON response. On Ubuntu installation command is as below
  + sudo apt-get install jq
  + For other platform please refer: [https://stedolan.github.io/jq/download/](https://nam02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fstedolan.github.io%2Fjq%2Fdownload%2F&data=02%7C01%7Camsrivastava%40informatica.com%7Caf91fd46f6604d07a3bd08d7639f6db1%7C2638f43ef77d4fc7ab927b753b7876fd%7C0%7C0%7C637087408714624064&sdata=IdTvIyRfnQ8fv7HS2WG6tvs1o%2FQiPNiTKNUty27x0hc%3D&reserved=0)
* Attached ADLSFileUtility.sh: Get a token, Extract the token and pass it to delete a file and then rename a file. To test it in the shell do below.
  + vi ADLSFileUtility.sh
  + chmod +x ADLSFileUtility.sh
  + ./ ADLSFileUtility.sh

**Note: The script need to be updated for tenantID, client\_id, client\_secret, {accountName}.{dnsSuffix}/{filesystem}/{path} to work in a specific environment.**

|  |
| --- |
| accessToken=$(curl -X POST \  [https://login.microsoftonline.com/2638f43e-f77d-4fc7-ab92-7b753b7876fd/oauth2/v2.0/token](https://nam02.safelinks.protection.outlook.com/?url=https%3A%2F%2Flogin.microsoftonline.com%2F2638f43e-f77d-4fc7-ab92-7b753b7876fd%2Foauth2%2Fv2.0%2Ftoken&data=02%7C01%7Camsrivastava%40informatica.com%7Caf91fd46f6604d07a3bd08d7639f6db1%7C2638f43ef77d4fc7ab927b753b7876fd%7C0%7C0%7C637087408714624064&sdata=bmzARZy7%2FF6%2BACVs8sJYyFXN9YM4NbLCTLYjjS4j0I4%3D&reserved=0) \    -H 'Accept: \*/\*' \    -H 'Content-Type: application/x-www-form-urlencoded' \    -d 'client\_id=%203ac6e6e2-e8ab-4fd3-b40b-8b82edbafb73&client\_secret=%205hjIu%40toKNz9J0%2FT%5B4pwj\*V%2FWHqvAJG2&scope=https%3A%2F%2Fstorage.azure.com%2F.default&grant\_type=%20client\_credentials&=' | jq -r '.access\_token')    curl -X DELETE \  [https://emeademostoragev2.dfs.core.windows.net/test/Document\_Total.parquet](https://nam02.safelinks.protection.outlook.com/?url=https%3A%2F%2Femeademostoragev2.dfs.core.windows.net%2Ftest%2FDocument_Total.parquet&data=02%7C01%7Camsrivastava%40informatica.com%7Caf91fd46f6604d07a3bd08d7639f6db1%7C2638f43ef77d4fc7ab927b753b7876fd%7C0%7C0%7C637087408714624064&sdata=y7ggvceNTBF9dz%2B4tvhbbWNBrhzib%2FjBiOWTy1Thp2w%3D&reserved=0) \    -H 'Accept: \*/\*' \    -H 'Accept-Encoding: gzip, deflate' \    -H 'Authorization: Bearer '$accessToken'' \    -H 'Cache-Control: no-cache' \    -H 'Connection: keep-alive' \    -H 'Content-Length: 0' \    -H 'cache-control: no-cache'    curl -X PUT \  [https://emeademostoragev2.dfs.core.windows.net/test/Document\_Total.parquet](https://nam02.safelinks.protection.outlook.com/?url=https%3A%2F%2Femeademostoragev2.dfs.core.windows.net%2Ftest%2FDocument_Total.parquet&data=02%7C01%7Camsrivastava%40informatica.com%7Caf91fd46f6604d07a3bd08d7639f6db1%7C2638f43ef77d4fc7ab927b753b7876fd%7C0%7C0%7C637087408714634021&sdata=iXfwI3HcJuku5NhiQLmDDiDppBj2ndvI087KtwPWiJ4%3D&reserved=0) \    -H 'Accept: \*/\*' \    -H 'Accept-Encoding: gzip, deflate' \    -H 'Authorization: Bearer '$accessToken'' \    -H 'Cache-Control: no-cache' \    -H 'Connection: keep-alive' \    -H 'cache-control: no-cache' \    -H 'content-length: 0' \    -H 'x-ms-rename-source: /test/Document\_Merge.parquet' |