ADLS Integration with Databricks

| **Method** | **Use Case** | **Secure** | **Recommended?** |
| --- | --- | --- | --- |
| **OAuth with SP** | **Prod use with service principal** | ✅ Yes | ✅ Yes |
| Managed Identity | No-secret access via VNet | ✅ Yes | ✅ Best if supported |
| **DBFS Mount** | **Easier path management** | ✅ Yes | Optional |
| SAS/Account Key | Temporary access/testing | ❌ No | Not for prod |

Service principal

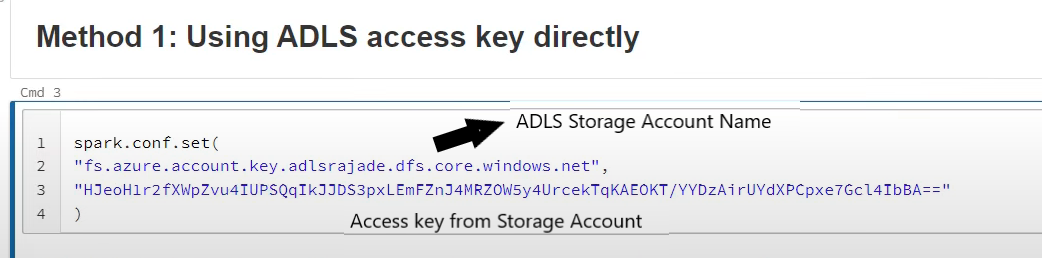
 **Register an Azure AD App** (service principal)

 **Assign it Storage Blob Data Contributor** access to your ADLS container or account

 Note the following:

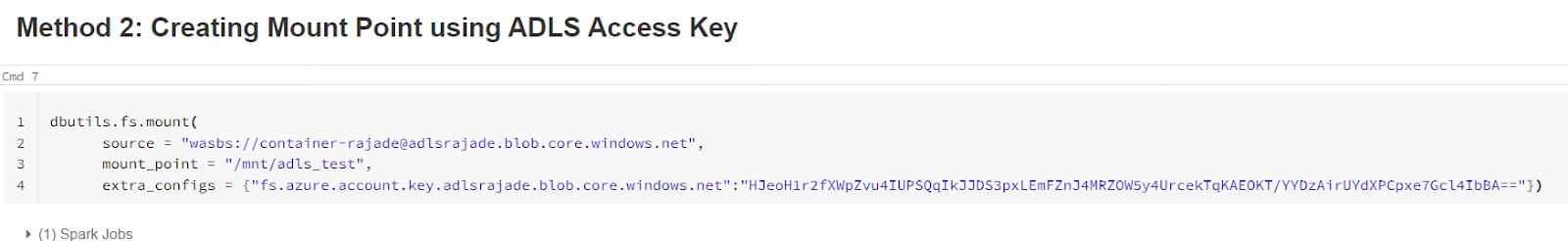
* client\_id (Application ID)
* client\_secret
* tenant\_id
* storage\_account\_name

Mount

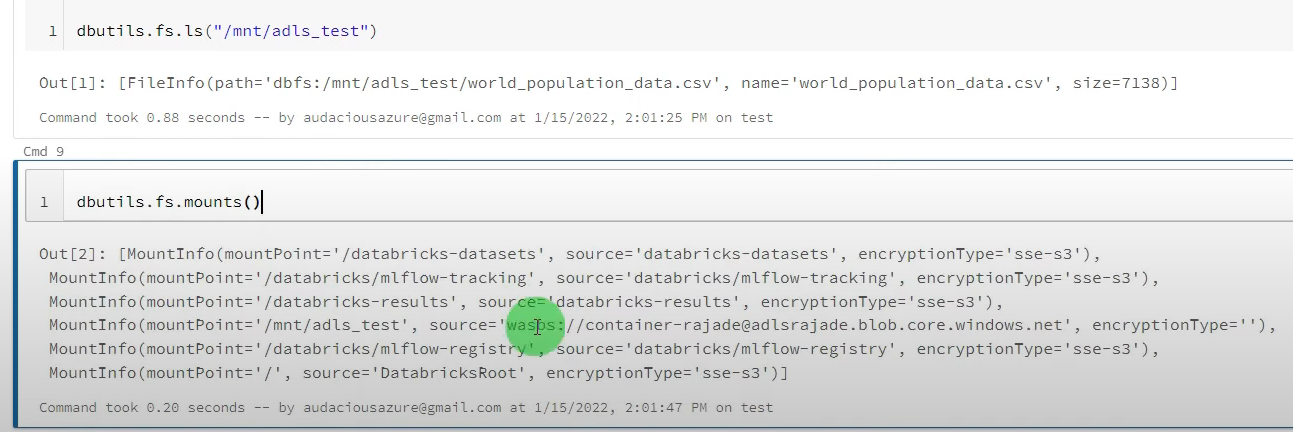


https://lh7-rt.googleusercontent.com/docsz/AD_4nXfeGCa4dzjY4pWnYdYNyXrBQTGmmOQps1pSQYYQ2OywAIpnr56I6metbzaT1Ukdf1TomG1JfWsNSNLiLH2Om-NjI0Y8l-6tk57vNq2y5YsatQXwp6zjAw4ENqOz5G4G2nJqcw0jS5mAumF2cQb4I8lmkezw?key=5-UzcStZXxBX00DSXnl4pA

Access using containername@adlsstorageacctname

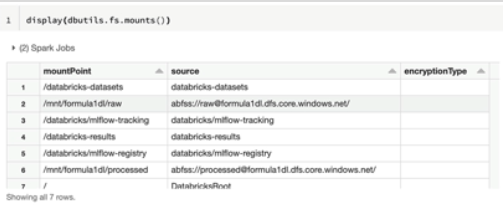


Mount\_point path we give as a free text which points to adls container.

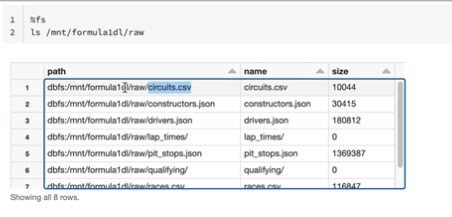


In Databricks notebook

display(dbutils.fs.mounts())

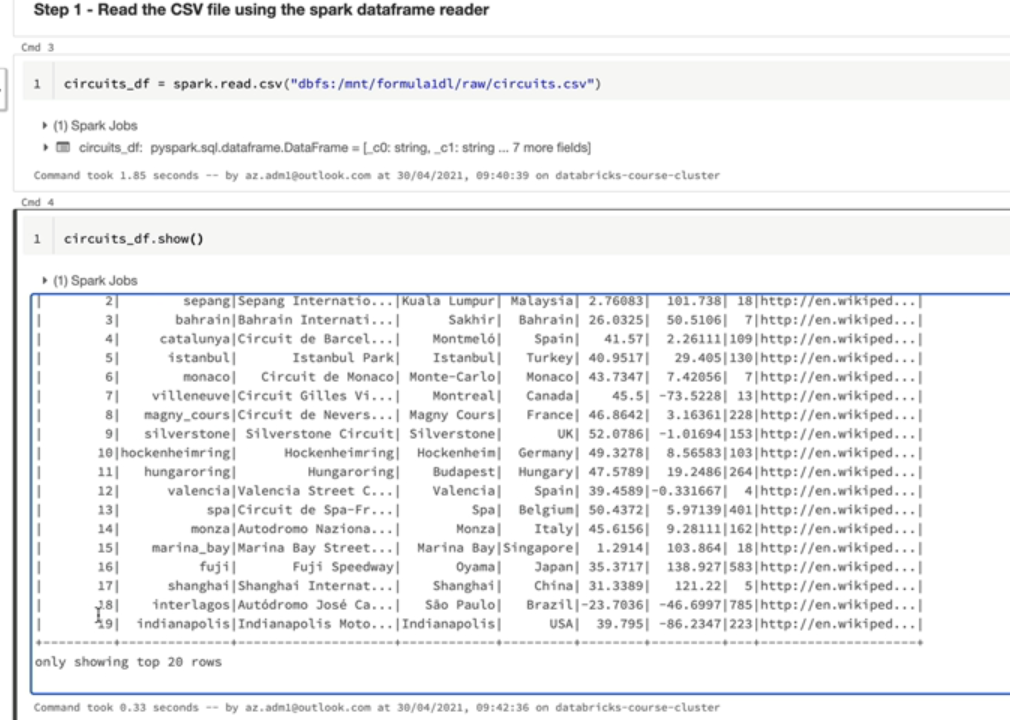


Get the file name under your raw directory



df = spark.read.csv("/mnt/<mount-name>/file.csv")

as I am using Databricks community edition using dbfs in below



show()  (pyspark method) method shows 20 records,  if not specified in its parameters. Also column data length truncated to 20 chars size.