Enable ADLS Gen1 Access

Contents:

By default, Microsoft Azure deployments integrate with Azure Data Lake Store (ADLS). Optionally, you can configure your deployment to integrate with WASB.

- **Microsof Azure Data Lake Store (ADLS Gen1)** is a scalable repository for big data analytics.
- ADLS Gen1 is accessible from Microsoft HDI and Azure Databricks.
- For more information, see *https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-overview*.
- For more information on the newer version of ADLS, see *Enable ADLS Gen2 Access*.

Limitations of ADLS Gen1 Integration

- In this release, the Trifacta platform supports integration with the default store only. Extra stores are not supported.

Read-only access

If the base storage layer has been set to WASB, you can follow these instructions to set up read-only access to ADLS Gen1.

> **NOTE:** To enable read-only access to ADLS Gen1, do not set the base storage layer to `adl`. The base storage layer for ADLS read-write access must remain `wasbs`.

Pre-requisites

General

- The Trifacta platform has already been installed and integrated with an Azure Databricks cluster. See *Configure for Azure Databricks*.
- ADL must be set as the base storage layer for the Trifacta platform instance. See *Set Base Storage Layer*.

Create a registered application

Before you integrate with Azure ADLS Gen1, you must create the Trifacta platform as a registered application. See *Configure for Azure*.

Azure properties

The following properties should already be specified in the Admin Settings page. Please verify that the following have been set:

- `azure.applicationId`
- `azure.secret`
- `azure.directoryId`

The above properties are needed for this configuration. For more information, see *Configure for Azure*.

Key Vault Setup

An Azure Key Vault has already been set up and configured for use by the Trifacta platform. For more information, see *Configure for Azure*.

Configure ADLS Authentication

Authentication to ADLS storage is supported for the following modes, which are described in the following section.

Mode	Description
System	All users authenticate to ADLS using a single system key/secret combination. This combination is specified in the following parameters, which you should have already defined: • `azure.applicationId` • `azure.secret` • `azure.directoryId` These properties define the registered application in Azure Active Directory. System authentication mode uses the registered application identifier as the service principal for authentication to ADLS Gen1. All users have the same permissions in ADLS Gen1. For more information on these settings, see *Configure for Azure*.
User	Per-user mode allows individual users to authenticate to ADLS Gen1 through their Azure Active Directory login. **NOTE:** Additional configuration for AD SSO is required. Details are below.

Steps:

Please complete the following steps to specify the ADLS Gen1 access mode.

1. You can apply this change through the *Admin Settings Page* (recommended) or `trifacta-conf.json`

   . For more information, see *Platform Configuration Methods*.
2. Set the following parameter to the preferred mode (`system` or `user`):

   ```
   "azure.adl.mode": "<your_preferred_mode>",
   ```

3. Save your changes.

System mode access

When access to ADLS is requested, the platform uses the combination of Azure directory ID, Azure application ID, and Azure secret to complete access.

After defining the properties in the Trifacta platform, system mode access requires no additional configuration.

User mode access

In user mode, a user ID hash is generated from the Key Vault key/secret and the user's AD login. This hash is used to generate the access token, which is stored in the Key Vault.

Set up for Azure AD SSO

> **NOTE:** User mode access to ADLS requires Single Sign On (SSO) to be enabled for integration with Azure Active Directory. For more information, see *Configure SSO for Azure AD*.

Configure the Trifacta platform

Configure storage protocol

You must configure the platform to use the ADL storage protocol when accessing.

> **NOTE:** Per earlier configuration, base storage layer must be set to `adl` for read/write access to ADLS. See *Set Base Storage Layer*.

Steps:

1. You can apply this change through the *Admin Settings Page* (recommended) or `trifacta-conf.json`

   . For more information, see *Platform Configuration Methods*.
2. Locate the following parameter and change its value to `adl`:

   ```
   "webapp.storageProtocol": "adl",
   ```

3. Set the following parameter to `false`:

   ```
   "hdfs.enabled": false,
   ```

4. Save your changes and restart the platform.

Define storage locations

You must define the base storage location and supported protocol for storing data on ADLS.

> **NOTE:** You can specify only one storage location for ADLS.

Steps:

1. You can apply this change through the *Admin Settings Page* (recommended) or `trifacta-conf.json`

   . For more information, see *Platform Configuration Methods*.
2. Locate the following configuration block. Specify the listed changes:

```
"fileStorage": {
    "defaultBaseUris": [
      "<baseURIOfYourLocation>"
    ],
    "whitelist": ["adl"]
  }
```

Parameter	Description
`defaultBaseUris`	A comma-separated list of protocols that are permitted to read and write with ADLS storage. **NOTE:** The `adl://` protocol identifier must be included. Example value: `adl://<YOUR_STORE_NAME>.azuredatalakestore.net`
`whitelist`	For each supported protocol, this array must contain a top-level path to the location where Trifacta platform files can be stored. These files include uploads, samples, and temporary storage used during job execution. **NOTE:** This array of values must include `adl`.

3. Save your changes and restart the platform.

Testing

Restart services. See *Start and Stop the Platform*.

After the configuration has been specified, an ADLS connection appears in the Import Data page. Select it to begin navigating for data sources.

Try running a simple job from the Trifacta application. For more information, see *Verify Operations*.

- See *ADLS Gen1 Browser*.
- See *Using ADLS*.