Enable WASB Access

Contents:

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

- **Windows Azure Storage Blob (WASB)** is an abstraction layer on top of HDFS, which enables persistence of storage, access without a Hadoop cluster presence, and access from multiple Hadoop clusters.

Limitations of WASB Integration

- In this release, the Trifacta platform supports integration with the default store only. Extra stores are not supported.
- If a directory is created on the HDI cluster through WASB, the directory includes a Size=0 blob. The Trifact a platform does not list them and does not support interaction with Size=0 blobs.

Read-only access

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

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

Pre-requisites

General

- The Trifacta platform has already been installed and integrated with an Azure HDI cluster. See *Configure for HDInsight*.
- WASB must be set as the base storage layer for the Trifacta platform instance. See *Set Base Storage Layer*.
- For each combination of blob host and container, a separate Azure Key Vault Store entry must be created. For more information, please contact your Azure admin.

Create a registered application

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

Other 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

For new installs, an Azure Key Vault has already been set up and configured for use by the Trifacta platform.

> **NOTE:** An Azure Key Vault is required. Upgrading customers who do not have a Key Vault in their environment must create one.

For more information, see *Configure for Azure*.

Configure WASB Authentication

Authentication to WASB storage is managed by specifying the appropriate host, container, and token ID in the Trifacta platform configuration. When access to WASB is requested, the platform passes the information through the Secure Token Service to query the specified Azure Key Vault Store using the provided values. The keystore returns the value for the secret. The combination of the key (token ID) and secret is used to access WASB.

For more information on creating the Key Vault Store and accessing it through the Secure Token Service, see *Configure for Azure*.

Configure the Trifacta platform

Define default storage location and access key

In platform configuration, you must define the following properties:

Item	Configuration	Description
storage account	`"azure.wasb.defaultStore.blobHost": "<your_value_here>",`	Azure path to the location where your data is to be stored.
container	`"azure.wasb.defaultStore.container": "<your_value_here>",`	Within your storage location, this value defines the default container for storing data.

token ID	`"azure.wasb.defaultStore.sasTokenId": "<your_value_here>",`	The default SAS token identifier to query the Azure Key Value Store. **NOTE:** Do not modify `azure.wasb.defaultStore.sasToken`, which should be set to null. It is for debugging purposes only.

When these properties are specified, the platform queries the keystore to return the secret for the specified token ID, which is used to gain access to WASB.

Configure storage protocol

You can configure the platform to use the WASB or WASBS (secure) storage protocol when accessing.

NOTE: Depending on how you created the secure token in the Azure portal, you can choose to support secure and non-secure protocols. Please use the setting below to determine the appropriate setting for your environment.

Azure token setting	storageProtocol setting
Support HTTP and HTTPS	`wasb`
Support HTTPS only	`wasbs` **Tip:** `wasbs` is recommended.

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 `wasb` for non-secure access or `wasbs` for secure access:

```
"webapp.storageProtocol": "wasbs",
```

3. Save your changes and restart the platform.

Enable

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 `true`:

```
"azure.wasb.enabled": true,
```

3. Save your changes and restart the platform.

Testing

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

After the configuration has been specified, a WASB connection appears in the Import Data page. Select it to begin navigating through the WASB Browser for data sources.

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

- See *WASB Browser*.
- See *Using WASB*.