

# Set Base Storage Layer

## Contents:

- *Base Storage Layer Options*
    - *HDFS*
    - *S3*
    - *WASBS*
    - *ADL*
    - *ABFSS*
    - *Base storage layer port options*
  - *Set Storage Layer*
  - *Disable Hadoop Access*
- 

In your platform configuration, you must specify the storage platform that is your base storage layer. The **base storage layer** defines the primary storage integration for the Trifacta® platform. In some cases, integration with other storage layers is supported.

**After you define the base storage layer and restart the platform, you cannot change the base storage layer to another option. Please consider your options carefully before you define the base storage layer.**

If S3 is the base storage layer, you must also define the default storage bucket to use during initial installation, which cannot be changed at a later time. For additional requirements, see *Enable S3 Access*.

**NOTE:** If HDFS is specified as your base storage layer, you cannot publish to Redshift.

## Base Storage Layer Options

### HDFS

If you are integrating with a Hadoop cluster, you can use HDFS for base storage.

**Tip:** For Trifacta Self-Managed Enterprise Edition, HDFS is the default base storage layer.

### S3

If you have installed the product on-premises or on an EC2 instance in AWS, you can set the base storage layer to S3.

Read access to S3 is supported if HDFS is the base storage layer.

For more information, see *Enable S3 Access*.

### Required for:

- Enable write access to S3
- Publish to Redshift

## WASBS

If you have installed the product from the Azure Marketplace and are integrating with WASB, you must set to the base storage layer to WASBS.

For more information, see *Enable WASB Access*.

### Required for:

- Access to WASB (Azure deployments only)

## ADL

Set the base storage layer to `adl` if you are integrating with ADLS Gen1 for read/write access.

**NOTE:** ADLS Gen1 storage requires an Azure Databricks cluster for execution.

For more information, see *Enable ADLS Gen1 Access*.

### Required for:

- Access to ADLS Gen1 (Azure deployments only)

## ABFSS

Set the base storage layer to `abfss` if you are integrating with ADLS Gen2.

**NOTE:** ADLS Gen2 storage requires an Azure Databricks cluster for execution.

For more information, see *Enable ADLS Gen2 Access*.

### Required for:

- Access to ADLS Gen2 (Azure deployments only)

For more information on options, see *Storage Deployment Options*.

## Base storage layer port options

When you configure your base storage layer, you must also define the port number to use for access.

**NOTE:** If you change the port number of the base storage layer in the future, all results from previous jobs are lost. Please choose the port number with care.

## Set Storage Layer

When you have decided on the final base storage layer, set the following property to one of the above values in platform configuration.

The platform requires that one backend datastore be configured as the base storage layer. This base storage layer is used for storing uploaded data and writing results and profiles. Please complete the following steps to set the base storage layer for the Trifacta platform.

**You cannot change the base storage layer after it has been set. You must uninstall and reinstall the platform to change it.**

#### 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 set it to the value for your base storage layer:

```
"webapp.storageProtocol": "hdfs",
```

3. Save your changes and restart the platform.

**NOTE:** To complete the integration with the base storage layer, additional configuration is required.

## Disable Hadoop Access

If you are not using Hadoop at all, please complete the following configuration change.

#### Steps:

1. Login to the Trifacta node.
2. Edit the following files:

```
site-config-*-s3.json  
site-config.installer-*-edge.json
```

3. In these files, set the following property value to `hostname`:

```
"hdfs.namenode.host": "hostname",
```

4. Save the files and restart the platform.