



TRIFACTA

Install Guide for Azure

Version: 6.0.2
Doc Build Date: 05/24/2019

Copyright © Trifacta Inc. 2019 - All Rights Reserved. CONFIDENTIAL

These materials (the “Documentation”) are the confidential and proprietary information of Trifacta Inc. and may not be reproduced, modified, or distributed without the prior written permission of Trifacta Inc.

EXCEPT AS OTHERWISE PROVIDED IN AN EXPRESS WRITTEN AGREEMENT, TRIFACTA INC. PROVIDES THIS DOCUMENTATION AS-IS AND WITHOUT WARRANTY AND TRIFACTA INC. DISCLAIMS ALL EXPRESS AND IMPLIED WARRANTIES TO THE EXTENT PERMITTED, INCLUDING WITHOUT LIMITATION THE IMPLIED WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT AND FITNESS FOR A PARTICULAR PURPOSE AND UNDER NO CIRCUMSTANCES WILL TRIFACTA INC. BE LIABLE FOR ANY AMOUNT GREATER THAN ONE HUNDRED DOLLARS (\$100) BASED ON ANY USE OF THE DOCUMENTATION.

For third-party license information, please select **About Trifacta** from the User menu.

- 1. *Install Overview* . . 4
 - 1.1 *Install for Azure* . . 4
- 2. *Install Software* . . 9
 - 2.1 *Install Dependencies without Internet Access* . . 9
 - 2.2 *Install Enterprise on CentOS and RHEL* . 10
 - 2.3 *Install Enterprise on Ubuntu* . 13
 - 2.4 *License Key* . 17
 - 2.5 *Install for Wrangler Enterprise Application* . 19
 - 2.6 *Start and Stop the Platform* . 22
 - 2.7 *Login* . 24
- 3. *Install Reference* . 25
 - 3.1 *Install SSL Certificate* . 25
 - 3.2 *Change Listening Port* . 28
 - 3.3 *Supported Deployment Scenarios for Azure* . 29
 - 3.4 *Uninstall* . 31

Install Overview

Contents:

- *Basic Install Workflow*
 - *Installation Scenarios*
 - *Notation*
-

Basic Install Workflow

1. Review the pre-installation checklist and other system requirements. See *Install Preparation*.
2. Review the requirements for your specific installation scenario in the following sections.
3. Install the software. See *Install Software*.
4. Install the databases. See *Install Databases*.
5. Configure your installation.
6. Verify operations.

Notation

In this guide, JSON settings are provided in dot notation. For example, `webapp.selfRegistration` refers to a JSON block `selfRegistration` under `webapp`:

```
{
  ...
  "webapp": {
    "selfRegistration": true,
    ...
  }
  ...
}
```

Install for Azure

Contents:

- *Scenario Description*
 - *Product Limitations*
 - *Pre-requisites*
 - *Desktop Requirements*
 - *Azure Pre-requisites*
 - *Preparation*
 - *Deploy the Cluster*
 - *Deploy the Trifacta node*
 - *Prepare the cluster*
 - *Install Workflow*
 - *Configuration Workflow*
 - *Documentation*
-

This install process applies to installing Trifacta® Wrangler Enterprise on an Azure infrastructure that you manage.

Azure Marketplace deployments:

NOTE: Content in this section does not apply to deployments from the Azure Marketplace, which provide fewer deployment and configuration options. For more information, see the Azure Marketplace.

Scenario Description

NOTE: All hardware in use for supporting the platform is maintained within the enterprise infrastructure on Azure.

- Installation of Trifacta Wrangler Enterprise on a node in Microsoft Azure
- Installation of Trifacta databases on the same node
- Integration with a supported cluster for running jobs.
- Base storage layer and backend datastore of ADLS or WASB
- High availability or failover of the Trifacta node is not supported in Azure.
- High availability of Hadoop components is automatically managed by the HDI cluster.
 - Auto-management does not apply to non-Hadoop clusters, such as Azure Databricks.

For more information on deployment scenarios, see *Supported Deployment Scenarios for Azure*.

Product Limitations

- The application user credentials are used to access to the HDI cluster. Details are provided below.
- ADLS/Storage Blob access is only for the HDInsight cluster's primary storage. Additional storage accounts are not supported.
- HDFS must be set as the base storage layer of the Trifacta platform. Details are provided later.
 - S3 integration and AWS-based integrations such as Redshift are not supported.
- Use of HttpFS is not supported.
- Security features such as Kerberos and secure impersonation are not supported.

For more information on the limitations of this deployment scenario, see *Product Limitations*.

Pre-requisites

Desktop Requirements

- All desktop users of the platform should have a supported version of Google Chrome installed on their desktops.
 - For more information. see *Desktop Requirements*.
 - If a supported browser is not available within your enterprise, desktop users can install the Trifacta enterprise application as a separate application. For more information, see *Install for Wrangler Enterprise Application*.
- All desktop users must be able to connect to the EC2 instance through the enterprise infrastructure.

Azure Pre-requisites

Depending on which of the following Azure components you are deploying, additional pre-requisites and limitations may apply. Please review these sections as well.

- Cluster:
 - *Configure for HDInsight*

- *Configure for Azure Databricks*
- Storage:
 - *Enable ADLS Access*
 - *Enable WASB Access*
- *Configure SSO for Azure AD*

Preparation

Before you begin, please verify that you have completed the following:

1. **Review Planning Guide:** Please review and verify *Install Preparation* and sub-topics.
2. **Read:** Please read this entire document before you create the EMR cluster or install the Trifacta platform.
3. **Acquire Assets:** Acquire the installation package for your operating system and your license key. For more information, contact *Trifacta Support*.
 - a. If you are completing the installation without Internet access, you must also acquire the offline versions of the system dependencies. See *Install Dependencies without Internet Access*.
4. **Cluster sizing:** Before you begin, you should allocate sufficient resources for the cluster. For guidance, please contact your Trifacta representative.
5. **Node:** Review the system requirements for the node hosting the Trifacta platform. See *System Requirements*.
 - a. The required set of ports must be enabled for listening. See *System Ports*.
 - b. This node should be dedicated for Trifacta use.
6. **Databases:**
 - a. The platform utilizes a set of databases that must be accessed from the Trifacta node. Databases are installed as part of the workflow described later.
 - b. For more information on the supported databases and versions, see *System Requirements*.
 - c. For more information on database installation requirements, see *Install Databases*.

Limitations: For more information on limitations of this scenario, see *Product Limitations* in the *Install Preparation* area.

Deploy the Cluster

Deploy and provision a cluster of one of the supported types. The Trifacta platform supports integrations with multiple cluster types.

NOTE: Before you deploy, you should review cluster sizing options. For guidance, please contact your Trifacta representative.

Primary storage of the cluster may be set to an existing Azure Data Lake Store or Blob Storage.

- Any additional storage associated with the cluster is not available through the Trifacta application.

For more information, see *Supported Deployment Scenarios for Azure*.

Deploy the Trifacta node

In your Azure infrastructure, you must deploy a suitable VM for the installation of the Trifacta platform.

The operating system requirements for the VM for installing the platform vary depending on the type of job execution cluster with which you are running.

Cluster Type	Supported O/S for VM	Notes
HDInsight	Ubuntu only	Trifacta platform must be installed on an edge node of the HDInsight cluster.
Azure Databricks	CentOS and Ubuntu	

- When you configure the platform to integrate with the cluster, you must acquire some information about the cluster resources. For more information on the set of information to collect, see *Pre-Install Checklist* in the *Install Preparation* area.
- For more information, see *System Requirements* in the *Install Preparation* area.
- A set of ports must be opened on the VM for the platform. For more information, see *System Ports* in the *Install Preparation* area.

For more information on the supported EMR distributions, see *Supported Deployment Scenarios for Azure*.

Prepare the cluster

1. Create the following directories, which are specified by parameter in the platform.

Default HDFS path	Platform configuration property
/user/trifacta	
/trifacta	
/trifacta/dictionaries	hdfs.pathsConfig.dictionaries
/trifacta/libraries	hdfs.pathsConfig.libraries
/trifacta/queryResults	hdfs.pathsConfig.batchResults
/trifacta/tempfiles	hdfs.pathsConfig.tempFiles
/trifacta/uploads	hdfs.pathsConfig.fileUpload
/trifacta/.datasourceCache	hdfs.pathsConfig.globalDatasourceCache

2. Change the ownership of the above directories to `trifacta:trifacta` or the corresponding values for the S3 user in your environment.

Additional users may be required. For more information, see *Required Users and Groups* in the *Install Preparation* area.

Install Workflow

Please complete these steps listed in order:

1. **Install Software:** Install the Trifacta platform software on the node you created.

NOTE: You must follow the instructions provided for Ubuntu installation.

See *Install Software*.

2. **Install Databases:** The platform requires several databases for storing metadata.

NOTE: The software assumes that you are installing the databases on a PostgreSQL server on the same node as the software. If you are not or are changing database names or ports, additional configuration is required as part of this installation process.

For more information, see *Install Databases*.

3. **Login to the application:** After software and databases are installed, you can login to the application to complete configuration:
 - a. See *Login*.

- b. As soon as you login, you should change the password on the admin account. In the left menu bar, select **Settings > Admin Settings**. Scroll down to Manage Users. For more information, see *Change Admin Password*.

Tip: At this point, you can access the online documentation through the application. In the left menu bar, select **Help menu > Product Docs**. All of the following content, plus updates, is available online. See Documentation below.

Configuration Workflow

After you have completed the above topics, you can complete the configuration for your deployment below.

NOTE: The following configuration topics are not part of this installation guide. See links below.

1. **Configure for Azure:** Configure the platform to work with Azure.
2. **Integrate with cluster:** If the application is up and running, you can configure to the backend cluster for running jobs. Choose one of the following:
 - a. HDInsight
 - b. Azure Databricks
3. **Integrate with backend storage:**
 - a. **Set base storage layer:** The base storage layer must be set at the time of install and cannot be changed. See *Set Base Storage Layer*.
 - b. ADLS
 - c. WASB
4. **Verify operations:** At this point, you should be able to run a job. See *Verify Operations*.
5. **Create additional connections:** Through connections, you can access other sources of data and, optionally, publish job results.

Documentation

You can access complete product documentation online and in PDF format. From within the product, select **Help menu > Product Docs**.

After you have accessed the documentation, the following topics are relevant to Azure deployments. Please review them in order.

Topic	Description
<i>Supported Deployment Scenarios for Azure</i>	Matrix of supported Azure components.
<i>Configure for Azure</i>	Top-level configuration topic on integrating the platform with Azure. <div style="border: 1px solid #c8e6c9; padding: 5px; margin-top: 10px;">Tip: You should review this page.</div>
<i>Configure for HDInsight</i>	Review this section if you are integrating the Trifacta platform with a pre-existing HDI cluster.
<i>Configure for Azure Databricks</i>	Review this section if you are integrating with a pre-existing Azure Databricks cluster.
<i>Enable ADLS Access</i>	Configuration to enable access to ADLS.
<i>Enable WASB Access</i>	Configuration to enable access to WASB.
<i>Verify Operations</i>	You should be able to verify platform operations by running a simple job at this time.

Relational Connections	To enable, see <i>Enable Relational Connections</i> . Azure-specific relational connections: <ul style="list-style-type: none"> • <i>Create SQL DW Connections</i> • <i>Create SQL DB Connections</i>
Configure SSO for Azure AD	How to integrate the Trifacta platform with Azure Active Directory for Single Sign-On.

Install Software

To install Trifacta® Wrangler Enterprise, please review and complete the following sections in the order listed below.

Topics:

- *Install Dependencies without Internet Access*
- *Install Enterprise on CentOS and RHEL*
- *Install Enterprise on Ubuntu*
- *License Key*
- *Install for Wrangler Enterprise Application*
- *Start and Stop the Platform*
- *Login*

Install Dependencies without Internet Access

Offline dependencies should be included in the URL location that Trifacta® provided to you. Please use the `*dependencies*` file.

NOTE: If your installation server is connected to the Internet, the required dependencies are automatically downloaded and installed for you. You may skip this section.

Use the steps below to acquire and install dependencies required by the Trifacta platform. If you need further assistance, please contact *Trifacta Support*.

Install dependencies without Internet access for CentOS or RHEL:

1. In a CentOS or RHEL environment, the dependencies repository must be installed into the following directory:

```
/var/local/trifacta
```

2. The following commands configure Yum to point to the repository in `/var/local/trifacta`, which yum knows as `local`. Repo permissions are set appropriately. Commands:

```
tar xvzf <DEPENDENCIES_ARCHIVE>.tar.gz
mv local.repo /etc/yum.repos.d
mv trifacta /var/local
chown -R root:root /var/local/trifacta
chmod -R o-w+r /var/local/trifacta
```

3. The following command installs the RPM while disable all repos other than local, which prevents the installer from reaching out to the Internet for package updates:

NOTE: The disabling of repositories only applies to this command.

```
sudo yum --disablerepo=* --enablerepo=local install <INSTALLER>.rpm
```

4. If the above command fails and complains about a missing repo, you can add the missing repo to the `enablerepo` list. For example, if the `centos-base` repo is reported as missing, then the command would be the following:

```
sudo yum --disablerepo=* --enablerepo=local,centos-base install  
<INSTALLER>.rpm
```

5. If you do not have a supported version of a Java Developer Kit installed on the Trifacta node, you can use the following command to install OpenJDK, which is included in the offline dependencies:

```
sudo yum --disablerepo=* --enablerepo=local,centos-base install  
java-1.8.0-openjdk-1.8.0 java-1.8.0-openjdk-devel
```

Install dependencies without Internet access in Ubuntu:

If you are trying to perform a manual installation of dependencies in Ubuntu, please contact *Trifacta Support*.

Install Enterprise on CentOS and RHEL

Contents:

- *Preparation*
- *Installation*
 - *1. Install Dependencies*
 - *2. Install JDK*
 - *3. Install Trifacta package*
 - *4. Verify Install*
 - *5. Install License Key*
 - *6. Store install packages*
 - *7. Install and configure Trifacta databases*
- *Configuration*

This guide takes you through the steps for installing Trifacta® Wrangler Enterprise software on CentOS or Red Hat.

For more information on supported operating system versions, see *System Requirements*.

Preparation

Before you begin, please complete the following.

NOTE: Except for database installation and configuration, all install commands should be run as the root user or a user with similar privileges. For database installation, you will be asked to switch the database user account.

Steps:

1. Set the node where Trifacta Wrangler Enterprise is to be installed.
 - a. Review the *System Requirements* and verify that all required components have been installed.
 - b. Verify that all required system ports are opened on the node. See *System Ports*.
2. Review the *Desktop Requirements*.

NOTE: Trifacta Wrangler Enterprise requires the installation of Google Chrome on each desktop. Additionally, two plugins must be enabled and of sufficient versions to properly use the Photon in-browser engine. For more information, see *Desktop Requirements*.

3. Review the *System Dependencies*.

NOTE: If you are installing on node without access to the Internet, you must download the offline dependencies before you begin. See *Install Dependencies without Internet Access*.

4. Acquire your *License Key*.
5. Install and verify operations of the datastore, if used.

NOTE: In some cases, access to the Hadoop cluster is required.

6. Verify access to the server where the Trifacta platform is to be installed.
7. **Hadoop:** Additional steps are required to integrate the Trifacta platform with Hadoop. See *Prepare Hadoop for Integration with the Platform*.

Installation

1. Install Dependencies

Without Internet access

If you have not done so already, you may download the dependency bundle with your release directly from Trifacta

- a. For more information, see *Install Dependencies without Internet Access*.

With Internet access

Use the following to add the hosted package repository for CentOS/RHEL, which will automatically install the proper packages for your environment.

```
# If the client has curl installed ...
curl
https://packagecloud.io/install/repositories/trifacta/dependencies/script
.rpm.sh | sudo bash

# Otherwise, you can also use wget ...
wget -qO-
https://packagecloud.io/install/repositories/trifacta/dependencies/script
.rpm.sh | sudo bash
```

2. Install JDK

By default, the Trifacta node uses OpenJDK for accessing Java libraries and components. In some environments, basic setup of the node may include installation of a JDK. Please review your environment to verify that an appropriate JDK version has been installed on the node.

NOTE: Use of Java Development Kits other than OpenJDK is not currently supported. However, the platform may work with the Java Development Kit of your choice, as long as it is compatible with the supported version(s) of Java. See *System Requirements*.

NOTE: OpenJDK is included in the offline dependencies, which can be used to install the platform without Internet access. For more information, see *Install Dependencies without Internet Access*.

The following commands can be used to install OpenJDK. These commands can be modified to install a separate compatible version of the JDK.

```
sudo yum install java-1.8.0-openjdk-1.8.0 java-1.8.0-openjdk-devel
```

NOTE: If `java-1.8.0-openjdk-devel` is not included, the batch job runner service, which is required, fails to start.

JAVA_HOME:

By default, the `JAVA_HOME` environment variable is configured to point to a default install location for the OpenJDK package.

NOTE: If you have installed a JDK other than the OpenJDK version provided with the software, you must set the `JAVA_HOME` environment variable on the Trifacta node to point to the correct install location.

The property value must be updated in the following locations:

1. Edit the following file: `/opt/trifacta/conf/env.sh`
2. Save changes.

3. Install Trifacta package

NOTE: If you are installing without Internet access, you must reference the local repository. The command to execute the installer is slightly different. See *Install Dependencies without Internet Access*.

NOTE: Installing the Trifacta platform in a directory other than the default one is not supported or recommended.

Install the package with yum, using root:

```
sudo yum install <rpm file>
```

4. Verify Install

The product is installed in the following directory:

```
/opt/trifacta
```

JAVA_HOME:

The platform must be made aware of the location of Java.

Steps:

1. Edit the following file: `/opt/trifacta/conf/trifacta-conf.json`
2. Update the following parameter value:

```
"env": {  
  "JAVA_HOME": "/usr/lib/jvm/java-1.8.0-openjdk.x86_64"  
},
```

3. Save changes.

5. Install License Key

Please install the license key provided to you by Trifacta. See *License Key*.

6. Store install packages

For safekeeping, you should retain all install packages that have been installed with this Trifacta deployment.

7. Install and configure Trifacta databases

The Trifacta platform requires installation of several databases. If you have not done so already, you must install and configure the databases used to store Trifacta metadata. See *Install Databases*.

Configuration

After installation is complete, additional configuration is required.

The Trifacta platform requires additional configuration for a successful integration with the datastore. Please review and complete the necessary configuration steps. For more information, see *Configure*.

Install Enterprise on Ubuntu

Contents:

- *Preparation*
- *Installation*
 - *1. Install Dependencies*

- 2. *Install JDK*
 - 3. *Install Trifacta package*
 - 4. *Verify Install*
 - 5. *Install License Key*
 - 6. *Store install packages*
 - 7. *Install and configure Trifacta databases*
 - *Configuration*
-

This guide takes you through the steps for installing Trifacta® Wrangler Enterprise software on Ubuntu. For more information on supported operating system versions, see *System Requirements*.

Preparation

Before you begin, please complete the following.

NOTE: Except for database installation and configuration, all install commands should be run as the root user or a user with similar privileges. For database installation, you will be asked to switch the database user account.

Steps:

1. Set the node where Trifacta Wrangler Enterprise is to be installed.
 - a. Review the *System Requirements* and verify that all required components have been installed.
 - b. Verify that all required system ports are opened on the node. See *System Ports*.
2. Review the *Desktop Requirements*.

NOTE: Trifacta Wrangler Enterprise requires the installation of Google Chrome on each desktop. Additionally, two plugins must be enabled and of sufficient versions to properly use the Photon running environment.

3. Review the *System Dependencies*.

NOTE: If you are installing on node without access to the Internet, you must download the offline dependencies before you begin. See *Install Dependencies without Internet Access*.

4. Acquire your *License Key*.
5. Install and verify operations of the datastore, if used.

NOTE: In some cases, access to the Hadoop cluster is required.

6. Verify access to the server where the Trifacta platform is to be installed.
7. **Hadoop:** Additional steps are required to integrate the Trifacta platform with Hadoop. See *Prepare Hadoop for Integration with the Platform*.

Installation

1. Install Dependencies

Without Internet access

If you have not done so already, you may download the dependency bundle with your release directly from Trifacta a. For more information, see *Install Dependencies without Internet Access*.

With Internet access

Use the following to add the hosted package repository for Ubuntu, which will automatically install the proper packages for your environment.

NOTE: Install curl if not present on your system.

Then, execute the following command:

NOTE: Run the following command as the root user. In proxied environments, the script may encounter issues with detecting proxy settings.

```
curl
https://packagecloud.io/install/repositories/trifacta/dependencies/script
.deb.sh | sudo bash
```

Special instructions for Ubuntu installs

These steps manually install the correct and supported version of the following:

- nodeJS
- nginx

Due to a known issue resolving package dependencies on Ubuntu, please complete the following steps prior to installation of other dependencies or software.

1. Login to the Trifacta node as an administrator.
2. Execute the following command to install the appropriate versions of nodeJS and nginx.
 - a. Ubuntu 14.04:

```
sudo apt-get install nginx=1.12.2-1~trusty
nodejs=10.13.0-1nodesource1
```

- b. Ubuntu 16.04

```
sudo apt-get install nginx=1.12.2-1~xenial
nodejs=10.13.0-1nodesource1
```

3. Continue with the installation process.

2. Install JDK

By default, the Trifacta node uses OpenJDK for accessing Java libraries and components. In some environments, basic setup of the node may include installation of a JDK. Please review your environment to verify that an appropriate JDK version has been installed on the node.

NOTE: Use of Java Development Kits other than OpenJDK is not currently supported. However, the platform may work with the Java Development Kit of your choice, as long as it is compatible with the supported version(s) of Java. See *System Requirements*.

NOTE: OpenJDK is included in the offline dependencies, which can be used to install the platform without Internet access. For more information, see *Install Dependencies without Internet Access*.

The following commands can be used to install OpenJDK. These commands can be modified to install a separate compatible version of the JDK.

```
sudo apt-get install openjdk-8-jre-headless
```

JAVA_HOME:

By default, the `JAVA_HOME` environment variable is configured to point to a default install location for the OpenJDK package.

NOTE: If you have installed a JDK other than the OpenJDK version provided with the software, you must set the `JAVA_HOME` environment variable on the Trifacta node to point to the correct install location.

The property value must be updated in the following locations:

1. Edit the following file: `/opt/trifacta/conf/env.sh`
2. Save changes.

3. Install Trifacta package

NOTE: If you are installing without Internet access, you must reference the local repository. The command to execute the installer is slightly different. See *Install Dependencies without Internet Access*.

NOTE: Installing the Trifacta platform in a directory other than the default one is not supported or recommended.

Install the package with apt, using root:

```
sudo dpkg -i <deb file>
```

The previous line may return an error message, which you may ignore. Continue with the following command:


```
sudo apt-get -f -y install
```

4. Verify Install

The product is installed in the following directory:

```
/opt/trifacta
```

JAVA_HOME:

The platform must be made aware of the location of Java.

Steps:

1. Edit the following file: `/opt/trifacta/conf/trifacta-conf.json`
2. Update the following parameter value:

```
"env": {  
  "JAVA_HOME": "/usr/lib/jvm/java-1.8.0-openjdk.x86_64"  
},
```

3. Save changes.

5. Install License Key

Please install the license key provided to you by Trifacta. See *License Key*.

6. Store install packages

For safekeeping, you should retain all install packages that have been installed with this Trifacta deployment.

7. Install and configure Trifacta databases

The Trifacta platform requires installation of several databases. If you have not done so already, you must install and configure the databases used to store Trifacta metadata. See *Install Databases*.

Configuration

After installation is complete, additional configuration is required.

The Trifacta platform requires additional configuration for a successful integration with the datastore. Please review and complete the necessary configuration steps. For more information, see *Configure*.

License Key

Contents:

- *Acquire license key*
 - *Install your license key*
 - *Update your license key*
 - *Changing the license key location*
 - *Expired license*
 - *Invalid license key file*
-

Acquire license key

A valid license key (`license.json`) is provided to each customer prior to installation. Your license key file is a JSON file that contains important information on your license such as the expiration date.

NOTE: If your license key has expired, please contact *Trifacta Support*.

Install your license key

If you are updating your license, you may want to save your previous license key to a new location before overwriting.

NOTE: Do not maintain multiple license key files in this directory.

To apply your new or updated license key, copy the key file to the following location in the Trifacta® deployment:

```
/opt/trifacta/license
```

Update your license key

After you have installed your license key, you can update your license with a new one through the Admin Settings page. See *Admin Settings Page*.

Changing the license key location

By default, the license key file in use must be named: `license.json`.

If needed, you can change the path and filename of the license key. The property is the following:

```
"license.location"
```

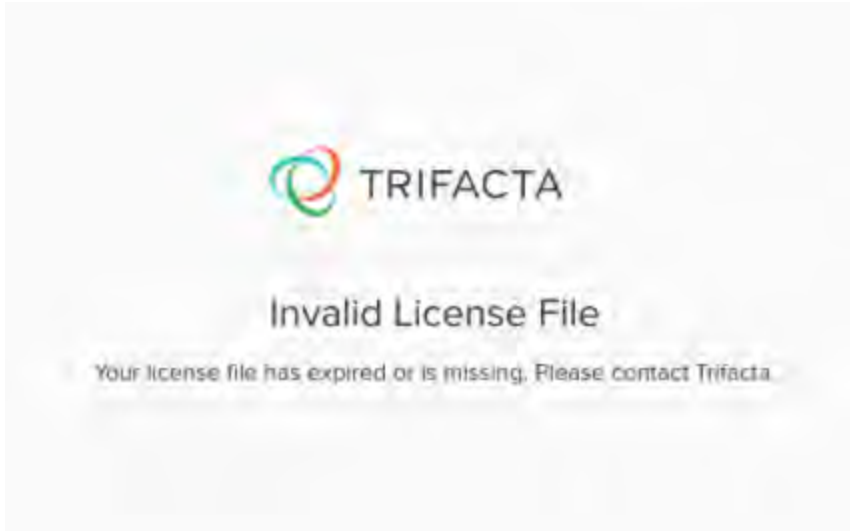
See *Admin Settings Page*.

Expired license

NOTE: If your license expires, you cannot use the product until a new and valid license key file has been applied. When administrators attempt to login to the application, they are automatically redirected to a location from which they can upload a new license key file.

Invalid license key file

When you start the Trifacta platform, you may see the following:



Your license key is either missing or has expired. Please contact *Trifacta Support*.

Install for Wrangler Enterprise Application

Contents:

- *Download*
- *Setup*
- *Install for Windows*
- *Windows Command Line Installation and Configuration*
- *Launch the Application*
- *Documentation Note*
- *Uninstall*
- *Troubleshooting*
 - *Cannot connect to server*
 - *"Does Not Support Your Browser" error*

If your environment does not support the use of Chrome, you can install the Wrangler Enterprise desktop application to provide the same access and functionality as the Trifacta® application. This desktop application connects to the enterprise Trifacta instance and provides the same capabilities without requiring a locally installed version of Chrome browser.

Trifacta application is a hybrid desktop application. Your local application instance accesses registered data files located in the datastore to which the Trifacta server is connected.

NOTE: The Wrangler Enterprise desktop application is a 64-bit Microsoft Windows application. It requires a 64-bit version of Windows to execute. The application also supports Single Sign On (SSO), if it is enabled.

Download

To begin, you must download the following Windows MSI file (`TrifactaEnterpriseSetup.msi`) from the location where your software was provided.

If you are planning to automate installation to desktops in your environment, please also download `setTrifactaServer.ps1`.

Setup

Before you begin, you should perform any necessary configuration of the Trifacta Server before deploying the instances of the application. See *Configure for Trifacta Enterprise Application*.

Install for Windows

Steps:

1. On your Windows desktop, double-click the MSI file.
2. Follow the on-screen instructions to install the software.

Windows Command Line Installation and Configuration

As an alternative, you can perform installation and initial configuration from the command line. Download the MSI and the PS1 files to a local directory that is accessible.

NOTE: For command line install, you must download from the `setTrifactaServer.ps1` from the download location.

Install software:

```
msiexec /i <path_to_TrifactaEnterpriseSetup.msi> /passive
```

Configure URL of Trifacta Server:

```
setTrifactaServer.ps1 -trifactaServer <server_url> -installDir  
<local_dir>
```

Parameter	Description
trifactaServer	(Required) URL of the server hosting the Trifacta platform. Format: <pre><http https>://<host>:<port></pre>
installDir	(Optional) Specifies the installation directory in the local environment. If not specified, installation directory defaults to use the same path as the installer.
common installer parameters	This command supports the following Windows installer parameters: Verbose, Debug, ErrorAction, ErrorVariable, WarningAction, WarningVariable, OutBuffer, PipelineVariable, and OutVariable. For more information, see <i>about_CommonParameters</i> here: http://go.microsoft.com/fwlink/?LinkID=113216 .

After this install is completed, desktop users should be able to use the application normally.

Launch the Application

Steps:

1. When installation is complete, double-click the application icon.
2. For the Trifacta Server, please enter the full URL including port number of the Trifacta instance to which you are connecting.
 - a. By default, the server is available over port 3005. For more information, please contact your IT administrator.
 - b. If you connect to the Internet through a proxy, additional configuration is required. See *Configure Server Access through Proxy*.

NOTE: If you make a mistake in specifying the URL to the Trifacta server, please uninstall and reinstall the MSI. This step clears the local application cache, and you can enter the appropriate path through the application. See *Uninstall* below.

3. When the proper URL and port number are provided, you may launch the application.
4. If your environment contains multiple Trifacta Server deployments, you can select the one to which to connect:

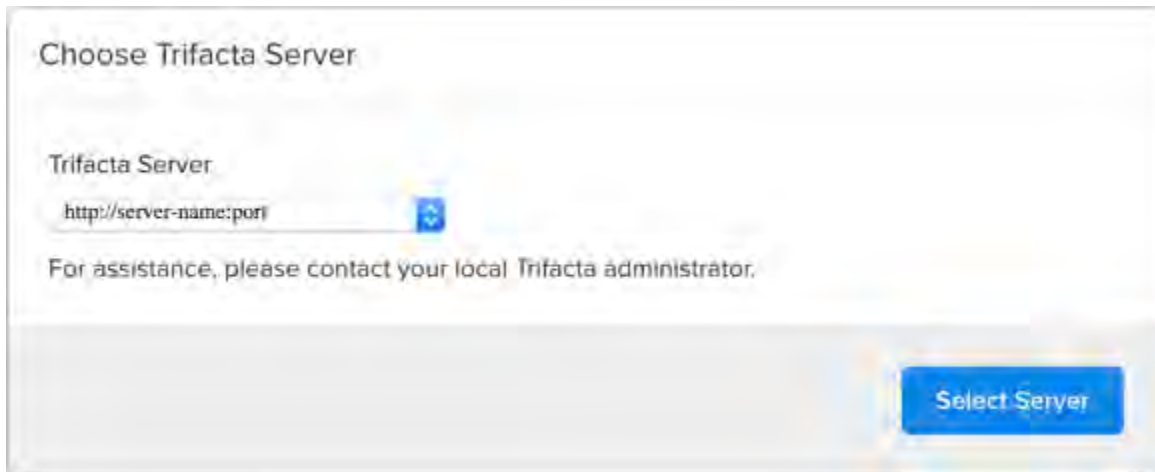


Figure: Choose Server

5. Login with your Trifacta account. See *Login*.

Documentation Note

Unless specifically noted, all features described for Trifacta Wrangler Enterprise or the Trifacta application apply to the Wrangler Enterprise desktop application.

Uninstall

To uninstall from your Windows machine, use the Add or Remove Programs control panel.

Troubleshooting

Cannot connect to server

If you are unable to connect to the Trifacta server, please do the following:

1. Verify that you are connecting to the appropriate URL.
 - a. If you are connecting to the incorrect URL, please uninstall the application and re-install using the

MSI file. See *Uninstall* above.

2. Verify if you need to connect to the server through a proxy server. If so, additional configuration is required. See *Configure Server Access through Proxy*.
3. Check your firewall settings.

"Does Not Support Your Browser" error

This error message indicates that you are trying to connect to an instance of the Trifacta server that does not support the Wrangler Enterprise desktop application. Please verify that your connection URL is pointed to a supported instance of the Trifacta server.

Start and Stop the Platform

Contents:

- *Start*
 - *Verify operations*
- *Restart*
- *Stop*
- *Debugging*
- *Troubleshooting*
 - *Error - SequelizeConnectionRefusedError: connect ECONNREFUSED*

Tip: The Restart Trifacta button in the Admin Settings page is the preferred method for restarting the platform.

NOTE: The restart button is not available when high availability is enabled for the Trifacta® node.

See *Admin Settings Page*.

Start

NOTE: These operations must be executed under the root user.

Command:

```
service trifacta start
```

Verify operations

Steps:

1. Check logs for errors:

```
/opt/trifacta/logs/*.log
```

- a. You can also access logs through the Trifacta® application for each service. See *System Services and Logs*.
2. Login to the Trifacta application. If available, perform a simple transformation operation. See *Login*.
 3. Run a simple job. See *Verify Operations*.

Restart

Command:

```
service trifacta restart
```

When the login page is available, the system has been restarted. See *Login*.

Tip: If you have made any configuration changes, you should verify operations. See *Verify Operations*.

Stop

Command:

```
service trifacta stop
```

Debugging

You can verify operations of WebHDFS. Command:

```
curl -i  
"http://<hadoop_node>:<port_number>/webhdfs/v1/?op=LISTSTATUS&user.name=t  
rifacta"
```

Troubleshooting

Error - SequelizeConnectionRefusedError: connect ECONNREFUSED

If you have attempted to start the platform after an operating system reboot, you may receive the following error message, and the platform start fails to complete:

```
2016-10-04T14:03:17.883Z - error: [ENVIRONMENT] Environment Sanity Test Failed
2016-10-04T14:03:17.883Z - error: [ENVIRONMENT] Exception Type: Error
2016-10-04T14:03:17.883Z - error: [ENVIRONMENT] Exception Message:
SequelizeConnectionRefusedError: connect ECONNREFUSED
```

Solution:

NOTE: This solution applies to PostgreSQL 9.6 only. Please modify for your installed database version.

This error can occur when the operating system is restarted. Please execute the following commands to check the PostgreSQL configuration and restart the databases.

```
chkconfig postgresql-9.6 on
```

Then, restart the platform as normal.

```
service trifacta restart
```

Login

NOTE: Administrators of the platform should change the default password for the admin account. See *Change Admin Password*.

To login to the Trifacta® application, navigate to the following in your browser:

```
http://<host_name>:<port_number>
```

where:

- <host_name> is the host of the Trifacta application.
- <port_number> is the port number to use. Default is 3005.

If you do not have an account, click **Register**.

- If self-registration is enabled, you may be able to immediately login after registering.
- If Kerberos or secure impersonation is enabled, an administrator must apply a Hadoop principal value to the account before you can login. Please contact your Trifacta administrator.
- System administrators can enable self-registration. See *Configure User Self-Registration*.

After you login, you are placed in the Flows page, where you can create and manage your datasets and flows. See *Flows Page*.

- If you are using S3 as your base storage layer, you or your Trifacta administrator must provide the AWS access key, secret, and storage bucket identifiers to connect to your storage. To do it yourself, click **Configure Storage Settings**. See *User Profile Page*.
- For a basic walkthrough of the Trifacta application, see *Workflow Basics*.

To logout:

From the Settings menu, select **Logout**.

Install Reference

These appendices provide additional information during installation of Trifacta® Wrangler Enterprise.

Topics:

- *Install SSL Certificate*
- *Change Listening Port*
- *Supported Deployment Scenarios for Cloudera*
- *Supported Deployment Scenarios for Hortonworks*
- *Supported Deployment Scenarios for AWS*
- *Supported Deployment Scenarios for Azure*
- *Uninstall*

Install SSL Certificate

Contents:

- *Pre-requisites*
 - *Configure nginx*
 - *Modify listening port for Trifacta platform*
 - *Add secure HTTP headers*
 - *Enable secure cookies*
-

You may optionally configure an SSL certificate to secure connections to the web application of the Trifacta® platform.

Pre-requisites

1. A valid SSL certificate for the FQDN where the Trifacta application is hosted
2. Root access to the Trifacta server
3. Trifacta platform is up and running

Configure nginx

There are two separate Nginx services on the server: one service for internal application use, and one service that functions as a proxy between users and the Trifacta application. To install the SSL certificate, all configuration are applied to the proxy process only.

Steps:

1. Log into the Trifacta server as the **centos** user. Switch to the **root** user:

```
sudo su
```

2. Enable the proxy nginx service so that it starts on boot:

```
systemctl enable nginx
```

3. Create a folder for the private key and limit access to it:

```
sudo mkdir /etc/ssl/private/ && sudo chmod 700 /etc/ssl/private
```

4. Copy the following files to the server. If you copy and paste the content, please ensure that you do not miss characters or insert unwanted characters.
- The `.key` file should go into the `/etc/ssl/private/` directory.
 - The `.crt` file and the CA bundle/intermediate certificate bundle should go into the `/etc/ssl/certs/` directory.

NOTE: The delivery name and format of these files varies by provider. Please verify with your provider's documentation if this is unclear.

- Your certificate and the intermediate/authority certificate must be combined into one file for nginx. Here is an example of how to combine them together:

```
cat example_com.crt bundle.crt >> ssl-bundle.crt
```

5. Update the permissions on these files. Modify the following filenames as necessary:

```
sudo chmod 600 /etc/ssl/certs/ssl-bundle.crt
sudo chmod 600 /etc/ssl/private/your-private-cert.key
```

6. Use the following commands to deploy the example SSL configuration file provided on the server:

NOTE: Below, some values are too long for a single line. Single lines that overflow to additional lines are marked with a `\`. The backslash should not be included if the line is used as input.

```
cp /opt/trifacta/conf/ssl-nginx.conf.sample
/etc/nginx/conf.d/trifacta.conf && \
rm /etc/nginx/conf.d/default.conf
```

7. Edit the following file:

```
/etc/nginx/conf.d/trifacta.conf
```

8. Please modify the following key directives at least:

Directive	Description
<code>server_name</code>	FQDN of the host, which must match the SSL certificate's Common Name
<code>ssl_certificate</code>	Path to the file of the certificate bundle that you created on the server. This value may not require modification.
<code>ssl_certificate_key</code>	Path to the <code>.key</code> file on the server.

Example file:

```
server {
    listen          443;
    ssl             on;
    server_name     EXAMPLE.CUSTOMER.COM;
    # Don't limit the size of client uploads.
    client_max_body_size 0;
    access_log      /var/log/nginx/ssl-access.log;
    error_log       /var/log/nginx/ssl-error.log;
    ssl_certificate  /etc/ssl/certs/ssl-bundle.crt;
    ssl_certificate_key /etc/ssl/certs/EXAMPLE-NAME.key;
    ssl_protocols   SSLv3 TLSv1 TLSv1.1 TLSv1.2;
    ssl_ciphers     RC4:HIGH:!aNULL:!MD5;
    ssl_prefer_server_ciphers on;
    keepalive_timeout 60;
    ssl_session_cache shared:SSL:10m;
    ssl_session_timeout 10m;
    location / {
        proxy_pass http://localhost:3005;
        proxy_next_upstream error timeout invalid_header http_500
http_502 http_503 http_504;
        proxy_set_header    Accept-Encoding    "";
        proxy_set_header    Host               $host;
        proxy_set_header    X-Real-IP         $remote_addr;
        proxy_set_header    X-Forwarded-For
$proxy_add_x_forwarded_for;
        proxy_set_header    X-Forwarded-Proto $scheme;
        add_header          Front-End-Https   on;
        proxy_http_version 1.1;
        proxy_set_header    Upgrade          $http_upgrade;
        proxy_set_header    Connection      "upgrade";
        proxy_set_header    Host            $host;
        proxy_redirect      off;
    }
    proxy_connect_timeout   6000;
    proxy_send_timeout      6000;
    proxy_read_timeout      6000;
    send_timeout            6000;
}
server {
    listen          80;
    return 301 https://$host$request_uri;
}
```

9. Save the file.
10. To apply the new configuration, start or restart the nginx service:

```
service nginx restart
```

Modify listening port for Trifacta platform

If you have changed the listening port as part of the above configuration change, then the `proxy.port` setting in Trifacta platform configuration must be updated. See *Change Listening Port*.

Add secure HTTP headers

If you have enabled SSL on the platform, you can optionally insert the following additional headers to all requests to the Trifacta node:

Header	Protocol	Required Parameters
X-XSS-Protection	HTTP and HTTPS	<code>proxy.securityHeaders.enabled=true</code>
X-Frame-Options	HTTP and HTTPS	<code>proxy.securityHeaders.enabled=true</code>
Strict-Transport-Security	HTTPS	<code>proxy.securityHeaders.enabled=true</code> and <code>proxy.securityHeaders.httpsHeaders=true</code>

NOTE: SSL must be enabled to apply these security headers.

Steps:

To add these headers to all requests, please apply the following change:

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 setting and change its value to `true`:

```
"proxy.securityHeaders.httpsHeaders": false,
```

3. Save your changes and restart the platform.

Enable secure cookies

If you have enabled SSL on the platform, you can optionally enable the use of secure cookies.

NOTE: SSL must be enabled.

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 setting and change its value to `true`:

```
"webapp.session.cookieSecureFlag": false,
```

3. Save your changes and restart the platform.

Change Listening Port

If you need to change the listening port for the Trifacta® platform, please complete the following instructions.

Tip: This change most typically applies if you are enabling use of SSL. For more information, see *Install SSL Certificate*.

NOTE: By default, the platform listens on port 3005. All client browsing devices must be configured to enable use of this port or any port number that you choose to use.

Steps:

1. Login to the Trifacta node as an admin.
2. Edit the following file:

```
/opt/trifacta/conf/nginx.conf
```

3. Edit the following setting:

```
server {  
    listen 3005;  
    ...  
}
```

4. Save the file.
5. You can apply this change through the *Admin Settings Page* (recommended) or `trifacta-conf.json`. For more information, see *Platform Configuration Methods*.
6. Locate the following setting:

```
"proxy.port": 3005,
```

7. Set this value to the same value you applied in `nginx.conf`.
8. Save your changes and restart the platform.

Supported Deployment Scenarios for Azure

Contents:

- *Azure Deployment Scenarios*
- *Azure Installations*
- *Azure Integrations*

Azure Deployment Scenarios

The following are the Azure deployment scenarios.

Deployment Scenario	Trifacta node i nstallation	Base Storage Layer	Storage - WASB	Storage - ADLS	Storage - SQL DB	Storage - SQL DW	Cluster
---------------------	--------------------------------	--------------------------	-------------------	-------------------	---------------------	---------------------	---------

Trifacta® Wrangler Enterprise install for WASB	Azure	WASB	read/write	read only	read	read/write	<ul style="list-style-type: none"> • HDI • Azure Databricks
Trifacta Wrangler Enterprise in stall for ADLS	Azure	HDFS	read only	read/write	read	read/write	<ul style="list-style-type: none"> • HDI • Azure Databricks

Legend and Notes:

Column	Notes
Deployment Scenario	Description of the Azure-connected deployment
Trifacta node installation	Location where the Trifacta node is installed in this scenario.
Base Storage Layer	<p>When the Trifacta platform is first installed, the base storage layer must be set.</p> <div style="border: 1px solid black; padding: 5px; margin: 10px 0;"> <p>NOTE: After you have begun using the product, you cannot change the base storage layer.</p> </div> <p>For more information, see <i>Set Base Storage Layer</i>.</p>
Storage - WASB	For read/write access to WASB, the base storage layer must be set to WASB. For more information, see <i>Enable WASB Access</i> .
Storage - ADLS	For read/write access to ADLS, the base storage layer must be set to HDFS. For more information, see <i>Enable ADLS Access</i> .
Storage - SQL DB	For Azure installs, you can optionally create a connection to an Azure-hosted instance of SQL DB. For more information, see <i>Create SQL DB Connections</i> .
Storage - SQL DW	For Azure installs, you can optionally create a connection to an Azure-hosted instance of SQL DW. For more information, see <i>Create SQL DW Connections</i> .
Cluster	<p>List of Hadoop cluster types that are supported for integration and job execution at scale.</p> <ul style="list-style-type: none"> • The Trifacta platform can integrate with at most one cluster. It cannot integrate with two different clusters at the same time. • Smaller jobs can be executed on the Trifacta Server running environment, which is hosted on the Trifacta node itself. • For more information, see <i>Running Environment Options</i>.

Azure Installations

All Azure installations are executed through the Azure Marketplace. Then, the platform can be integrated with a pre-existing HDI cluster or a cluster that you create as part of the installation.

For more information, see *Install from Azure Marketplace*.

Azure Integrations

The following table describes the different Azure components that can host or integrate with the Trifacta platform. Combinations of one or more of these items constitute one of the deployment scenarios listed in the following section.

Azure Service	Description	Base Storage Layer	Other Required Azure Services
HDI	Microsoft Azure deployments can integrate with an HDI cluster, which can be pre-existing or created at the time of deployment. See <i>Configure for HDInsight</i> .	Base storage layer can be HDFS (for ADLS) or WASB.	
Azure Databricks	Optionally, you can integrate the Trifacta platform with an Azure Databricks cluster. See <i>Configure for Azure Databricks</i> .	Base storage layer can be HDFS (for ADLS) or WASB.	
WASB	Windows Azure Storage Blobs (WASB) extends HDFS to enable access to storage blobs that have not been deployed into the HDI cluster. See <i>Enable WASB Access</i> .	Base Storage Layer = WASB	HDI cluster WASB Key Vault
ADLS	Active Data Lake Store (ADLS) provides a highly scalable file-based storage system within HDI cluster. See <i>Enable ADLS Access</i> .	Base Storage Layer = HDFS	HDI cluster ADLS Key Vault

The following database connections are optional.

Database Name	Description
Hive	You can read from and write to Hive, a data warehouse built on top of HDI. <ul style="list-style-type: none"> For usage information, see <i>Using Hive</i>. To create the connection, see <i>Create Hive Connections</i>.
SQL DW	You can read from and write to SQL Data Warehouse, a scalable data warehouse solution for Azure. <ul style="list-style-type: none"> For usage information, see <i>Using SQL DW</i>. To create the connection, see <i>Create SQL DW Connections</i>.
SQL DB	You can read from SQL DB, a SQL Server variant for Azure. <ul style="list-style-type: none"> For usage information, see <i>Using Databases</i>. To create the connection, see <i>Create SQL DB Connections</i>.

Uninstall

To remove Trifacta® Wrangler Enterprise, execute as root user one of the following commands on the Trifacta node.

NOTE: All platform and Hadoop configuration files are preserved. User metadata is preserved in the Trifacta database.

CentOS/RHEL:

```
sudo rpm -e trifacta
```

Ubuntu:

```
sudo apt-get remove trifacta
```



Copyright © 2019 - Trifacta, Inc.
All rights reserved.