



TRIFACTA

Install Guide for AWS

Version: 6.8.2
Doc Build Date: 04/14/2020

Copyright © Trifacta Inc. 2020 - 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 Help menu.

- 1. *Install* . 4
 - 1.1 *Install Overview* . 4
 - 1.1.1 *Install for AWS* . . 6
 - 1.1.2 *Configure Server Access through Proxy* . 10
 - 1.2 *Install Software* 11
 - 1.2.1 *Install Dependencies without Internet Access* . 11
 - 1.2.2 *Install for Docker* 13
 - 1.2.3 *Install on CentOS and RHEL* . 22
 - 1.2.4 *Install on Ubuntu* . 28
 - 1.2.5 *License Key* . 34
 - 1.3 *Start and Stop the Platform* . 36
 - 1.4 *Login* 39
 - 1.5 *Install Reference* 40
 - 1.5.1 *Install SSL Certificate* 40
 - 1.5.2 *Change Listening Port* . 44
 - 1.5.3 *Install Desktop Application* 45
 - 1.5.4 *Supported Deployment Scenarios for AWS* . 48
 - 1.5.5 *Uninstall* 52

Install

This section contains content related to the installation and configuration of Trifacta® products.

Install Overview

Contents:

- *Required Documents*
- *Basic Install Workflow*
- *Installation Scenarios*
 - *Install On-Premises*
 - *Install for AWS*
 - *Install for Azure*
 - *Install for Docker*
- *Install Errata*
 - *Desktop Application Installation*
 - *Notation*

Required Documents

If you do not have access to online documentation, please verify that you have the following PDF documents, which are part of or are referenced during the installation process.

Tip: You should be able to install and configure the Trifacta platform using only the Install Guide. However, if you have additional requirements or require further explanation than what is provided in the Install Guide, these documents are important references.

NOTE: For AWS Marketplace or Azure Marketplace installs, the content available through the Marketplace should contain all documentation required to complete the installation.

Document	Starting Online Link	Description
Planning Guide PDF	<i>Install Planning</i>	Requirements and pre-install preparation for the Trifacta node in your install environment Tip: If you have not done so already, please review this Guide and verify requirements against your environment. In particular, please verify the <i>Product Support Matrix</i> and <i>System Requirements</i> sections.
Databases Guide PDF	<i>Install Databases</i>	Install or upgrade the Trifacta databases for one of the supported database versions.

Install Guide PDF	<i>Install Overview</i>	Instructions for installing the software and configuring it for basic operations NOTE: There are separate Install Guides for each supported infrastructure. Please verify that you are using the appropriate one.
Configuration Guide PDF	<i>Configure</i>	Configure integrations with running environments, backend storage, and other data sources, as well as instructions for configuring the Trifacta platform
Admin Guide PDF	<i>Admin</i>	For the installation process, this Guide contains useful topics on backup and recovery and verifying operations of the Trifacta platform.
User Guide PDF	<i>Workflow Basics</i>	After installation and configuration is complete, this Guide can be helpful for references on how to use the product.

Basic Install Workflow

1. Prepare the environment for your installation scenario.
2. Install the software.
3. Install the databases.
4. Start the platform and login.
5. Configure your installation.
6. Verify operations.

The install workflow is described in detailed in the page for your installation scenario.

Install Errata

Desktop Application Installation

Optionally, end users who do not have access to a supported browser version can install the browser-free Wrangler Enterprise desktop application, which enables access to the product through the Trifacta application.

This installation can be completed at any time and via remote methods. See *Install Desktop Application*.

Notation

In this guide, JSON settings may be provided in dot notation in either of the following forms.

For example, `webapp.selfRegistration` refers to a JSON block `selfRegistration` under `webapp`:

Form 1:

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

Form 2:

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

Install for AWS

Contents:

- *Scenario Description*
 - *Limitations*
 - *Deployment Limitations*
 - *Product Limitations*
 - *Pre-requisites*
 - *AWS desktop requirements*
 - *AWS pre-requisites*
 - *Preparation*
 - *AWS Information*
 - *Internet access*
 - *Deploy the Cluster*
 - *Deploy the EC2 Node*
 - *Install Workflow*
 - *Next Steps*
-

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

AWS Marketplace deployments:

NOTE: Content in this section does not apply to deployments from the AWS Marketplace. For more information on installing from the Marketplace, see the AWS Marketplace listing.

Scenario Description

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

- Installation of Trifacta Wrangler Enterprise on an EC2 server in AWS

- Installation of Trifacta databases on AWS
- Integration with a supported EMR cluster.
- Base storage layer and backend data store of S3

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

Limitations

Deployment Limitations

The following limitations apply to installations of Trifacta Wrangler Enterprise on AWS:

- No support for high availability and failover
- Job cancellation is not supported on EMR.
- When publishing single files to S3, you cannot apply an `append` publishing action.
- The following limitations apply to EMR integration only:
 - No support for Hive integration
 - No support for secure impersonation or Kerberos

Product Limitations

For general limitations of Trifacta Wrangler Enterprise, see *Product Limitations* in the Planning Guide.

Pre-requisites

Please acquire the following assets:

- **Install Package:** Acquire the installation package for your operating system.
 - **License Key:** As part of the installation package, you should receive a license key file. See *License Key* for details.
 - For more information, contact *Trifacta Support*.
- **Offline system dependencies:** 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*.

AWS desktop requirements

- All desktop users must be able to connect to the EC2 instance through the enterprise infrastructure.

AWS pre-requisites

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

- *Configure for EMR* in the Configuration Guide
- *Enable S3 Access* in the Configuration Guide
- *Create Redshift Connections* in the Configuration Guide

Preparation

Before you install Trifacta Wrangler Enterprise on AWS, please verify that you have completed the following:

1. **Read:** Please read this entire document before you create the EMR cluster or install the Trifacta platform.
2. **VPC:** Enable and deploy a working AWS VPC.
 1. In your VPC, you must define a subnet where you plan to deploy the Trifacta node.
3. **S3:** Enable and deploy an AWS S3 bucket to use as the base storage layer for the platform. In the bucket, the platform stores metadata in the following location:

```
<S3_bucket_name>/trifacta
```

See <https://s3.console.aws.amazon.com/s3/home>.

4. **IAM Policies:** Create IAM policies for access to the S3 bucket. Required permissions are the following:
 - The system account or individual user accounts must have full permissions for the S3 bucket:

```
Delete*, Get*, List*, Put*, Replicate*, Restore*
```

- These policies must apply to the bucket and its contents. Example:

```
"arn:aws:s3:::my-trifacta-bucket-name"  
"arn:aws:s3:::my-trifacta-bucket-name/*"
```

- See <https://console.aws.amazon.com/iam/home#/policies>

5. **EC2 instance role:** Create an EC2 instance role for your S3 bucket policy. See <https://console.aws.amazon.com/iam/home#/roles>.
6. **EC2 instance:** Deploy an AWS EC2 with SELinux where the Trifacta software can be installed.
 1. The required set of ports must be enabled for listening. See *System Ports* in the Planning Guide.
 2. This node should be dedicated for Trifacta use.

NOTE: The EC2 node must meet the system requirements for installing the platform. For more information, see *System Requirements* in the Planning Guide.

7. **EMR cluster:** An existing EMR cluster is required.
 1. **Cluster sizing:** Before you begin, you should allocate sufficient resources for sizing the cluster. For guidance, please contact your Trifacta representative.
 2. See Deploy the Cluster below.
8. **Databases:**
 1. 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.
 2. If installing databases on Amazon RDS, an admin account to RDS is required. For more information, see *Install Databases on Amazon RDS*.

AWS Information

Before you begin installation, please acquire the following information from AWS:

- **EMR:**
 - AWS region for the EMR cluster, if it exists.
 - ID for EMR cluster, if it exists
 - If you are creating an EMR cluster as part of this process, please retain the ID.
 - The EMR cluster must allow access from the Trifacta node. This configuration is described later.
- **Subnet:** Subnet within your virtual private cloud (VPC) where you want to launch the Trifacta platform.
 - This subnet should be in the same VPC as the EMR cluster.
 - Subnet can be private or public.
 - If it is private and it cannot access the Internet, additional configuration is required. See below.
- **S3:**
 - Name of the S3 bucket that the platform can use
 - Path to resources on the S3 bucket
- **EC2:**

- Instance type for the Trifacta node

Internet access

From AWS, the Trifacta platform requires Internet access for the following services:

NOTE: Depending on your AWS deployment, some of these services may not be required.

- AWS S3
- Key Management System [KMS] (if sse-kms server side encryption is enabled)
- Secure Token Service [STS] (if temporary credential provider is used)
- EMR (if integration with EMR cluster is enabled)

NOTE: If the Trifacta platform is hosted in a VPC where Internet access is restricted, access to S3, KMS and STS services must be provided by creating a VPC endpoint. If the platform is accessing an EMR cluster, a proxy server can be configured to provide access to the AWS ElasticMapReduce regional endpoint.

Deploy the Cluster

In your AWS infrastructure, you must deploy a supported version of EMR across a recommended number of nodes to support the expected data volumes of your Trifacta jobs.

- For more information on the supported EMR distributions, see *Supported Deployment Scenarios for AWS*.
- For more information on suggested sizing, see *Sizing Guidelines* in the Planning Guide.

NOTE: Cluster information including cluster configuration files must be accessible to the Trifacta node. These requirements are described below.

Deploy the EC2 Node

An EC2 node of the cluster must be deployed to host the Trifacta platform software. Here are some guidelines for deploying the EC2 cluster from the EC2 cluster:

1. **Instance size:** Select the instance size.
2. **Network:** Configure the VPC, subnet, firewall and other configuration settings necessary to communicate with the instance.
3. **Auto-assigned Public IP:** You must create a public IP to access the Trifacta platform.
4. **EC2 role:** Select the EC2 role that you created.
5. **Local storage:** Select a local EBS volume. The default volume includes 100GB storage.

NOTE: The local storage environment contains the Trifacta databases, the product installation, and its log files. No source data is ever stored within the product.

6. **Security group:** Use a security group that exposes access to port 3005, which is the default port for the platform.
7. **Create an AWS key-pair for access:** This key is used to provide SSH access to the platform, which may be required for some admin tasks.
8. Save your changes.

Install Workflow

NOTE: These steps are covered in greater detail later in this section.

The installation and configuration process requires the following steps. To continue, see Next Steps below.

1. **Install software:** Install the Trifacta platform software on the Trifacta node. See *Install Software*.
2. **Install databases:** The platform requires several databases for storage.

NOTE: The default configuration assumes that you are installing the databases on a PostgreSQL server on the same edge node as the software using the default ports. If you are changing the default configuration, additional configuration is required as part of this installation process.

For more information, see *Install Databases* in the Databases Guide.

3. **Start the platform:** For more information, see *Start and Stop the Platform*.
4. **Login to the application:** After software and databases are installed, you can login to the application to complete configuration:
 1. See *Login*.
 2. As soon as you login, you should change the password on the admin account. In the left menu bar, select **Settings > Settings > Admin Settings**. Scroll down to Manage Users. For more information, see *Change Admin Password* in the Configuration Guide.

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

5. **Install configuration:** After you are able to successfully login to the Trifacta application, you must configure the product to work with your backend storage layer and the running environment on the cluster. See *Install Configuration*.

Next Steps

To continue, please install the Trifacta software on the Trifacta node.

NOTE: Please complete the installation steps for the operating system version that is installed on the Trifacta node.

See *Install Software*.

Configure Server Access through Proxy

When you attempt to launch the application, you may receive an error message similar to the following:

```
No internet connection
Remote server timed out.
```

In some environments, your desktop machine may need to connect to the Internet through a proxy server. If you are using Wrangler Enterprise desktop application, it needs to know the proxy server to which to connect in order to access the Trifacta® node.

Please complete the following configuration steps to access the Trifacta servers.

Steps:

1. In the No internet connection dialog, click **Configure Proxy Settings**.
2. Please provide the following configuration information for your proxy server:
 1. **Proxy Host:** The URL of the proxy server. Please include the protocol identifier (e.g. `http://` or `https://`).
 2. **Proxy Port:** The port number to use to connect to the proxy server. In a URL, this value appears after a colon (e.g. `http://myproxy.example.com:8080`).
 3. **Username:** (optional) If your proxy requires a username to access, please enter it here.
 4. **Password:** (optional) Password associated with the user name.
3. Click **Save Proxy Settings and Restart**.

When the application restarts, you should be able to connect to the login screen.

NOTE: If you continue to have difficulties connecting to the Internet, please contact your network administrator or Internet provider.

Install Software

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

Install Dependencies without Internet Access

Contents:

- *Install CentOS or RHEL dependencies without Internet access*
 - *Install CentOS or RHEL software dependencies*
 - *Install CentOS or RHEL database dependencies*
- *Install Ubuntu dependencies without Internet access*
 - *Install Ubuntu software dependencies*

Offline dependencies should be included in the URL location that Trifacta® provided to you. Please use the `*deps*` 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 CentOS or RHEL dependencies without Internet access

Install CentOS or RHEL software dependencies

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 CentOS or RHEL database dependencies

If you are installing the databases on a CentOS node without Internet access, you can install the dependencies using either of the following commands:

NOTE: This step is only required if you are installing the databases on the same node where the software is installed.

For PostgreSQL:

```
sudo yum --disablerepo=* --enablerepo=local install postgresql96-server
```

For MySQL:

```
sudo yum --disablerepo=* --enablerepo=local install mysql-community-server
```

NOTE: You must also install the MySQL JARs on the Trifacta node. These instructions are provided later.

Databases are installed after the software is installed. For more information, see *Install Databases* in the Databases Guide.

Install Ubuntu dependencies without Internet access

Install Ubuntu software dependencies

In an Ubuntu environment, you can use the following sequence of commands to install the dependencies without Internet access.

1. Unzip the tar ball and change to the `trifacta-repo` directory:

```
tar xvzf trifacta-server-deps-6.8.0-ubuntu-16.04.tar.gz
cd trifacta-repo
```

2. Execute the following commands to install the dependencies:

```
sudo dpkg -i $(ls | grep minimal | sort)
sudo dpkg -i $(ls | grep -v ^python | sort)
sudo dpkg -i $(ls | grep python | sort)
sudo apt-get -f -y install
sudo dpkg -i <TRIFACTA_DEB_INSTALLER>
```

Install for Docker

Contents:

- *Deployment Scenario*
- *Limitations*
- *Requirements*
 - *Docker Daemon*
- *Preparation*
- *Acquire Image*
 - *Acquire from FTP site*

- *Build your own Docker image*
 - *Configure Docker Image*
 - *Start Server Container*
 - *Import Additional Configuration Files*
 - *Import license key file*
 - *Import Hadoop distribution libraries*
 - *Import Hadoop cluster configuration files*
 - *Install Kerberos client*
 - *Perform configuration changes as necessary*
 - *Start and Stop the Container*
 - *Stop container*
 - *Restart container*
 - *Recreate container*
 - *Stop and destroy the container*
 - *Verify Deployment*
 - *Configuration*
-

This guide steps through the process of acquiring and deploying a Docker image of the Trifacta® platform in your Docker environment. Optionally, you can build the Docker image locally, which enables further configuration options.

Deployment Scenario

- Trifacta Wrangler Enterprise deployed into a customer-managed environment: On-premises, AWS, or Azure.
- PostgreSQL 9.6 or MySQL 5.7 installed either:
 - Locally
 - Remote server
- Connected to a supported Hadoop cluster.
- Kerberos integration is supported.

Limitations

- You cannot upgrade to a Docker image from a non-Docker deployment.
- You cannot switch an existing installation to a Docker image.
- Supported distributions of Cloudera or Hortonworks:
 - *Supported Deployment Scenarios for Cloudera*
 - *Supported Deployment Scenarios for Hortonworks*
- The base storage layer of the platform must be HDFS. Base storage of S3 is not supported.
- High availability for the Trifacta platform in Docker is not supported.
- SSO integration is not supported.

Requirements

Support for orchestration through Docker Compose only

- Docker version 17.12 or later. Docker version must be compatible with the following version(s) of Docker Compose.
- Docker Compose 1.24.1. It should be compatible with above version(s) of Docker.

Docker Daemon

	Minimum	Recommended
CPU Cores	2 CPU	4 CPU
Available RAM	8 GB RAM	10+ GB RAM

Preparation

1. Review the *Desktop Requirements* in the Planning Guide.

NOTE: Trifacta Wrangler Enterprise requires the installation of a supported browser on each desktop.

2. Acquire your *License Key*.

Acquire Image

You can acquire the latest Docker image using one of the following methods:

1. Acquire from FTP site.
2. Build your own Docker image.

Acquire from FTP site

Steps:

1. Download the following files from the FTP site:
 1. `trifacta-docker-setup-bundle-x.y.z.tar`
 2. `trifacta-docker-image-x.y.z.tar`

NOTE: `x.y.z` refers to the version number (e.g. 6.4.0).

2. Untar the `setup-bundle` file:

```
tar xvf trifacta-docker-setup-bundle-x.y.z.tar
```

3. Files are extracted into a `docker` folder. Key files:

File	Description
<code>docker-compose-local-postgres.yaml</code>	Runtime configuration file for the Docker image when PostgreSQL is to be running on the same machine. More information is provided below.
<code>docker-compose-local-mysql.yaml</code>	Runtime configuration file for the Docker image when MySQL is to be running on the same machine. More information is provided below.
<code>docker-compose-remote-db.yaml</code>	Runtime configuration file for the Docker image when the database is deployed on a remote server. NOTE: You must manage this instance of the database. More information is provided below.

README-running-trifacta-container.md	Instructions for running the Trifacta container <div style="border: 1px solid #ccc; padding: 5px; text-align: center;"> NOTE: These instructions are referenced later in this workflow. </div>
README-building-trifacta-container.md	Instructions for building the Trifacta container <div style="border: 1px solid #ccc; padding: 5px; text-align: center;"> NOTE: This file does not apply if you are using the provided Docker image. </div>

4. Load the Docker image into your local Docker environment:

```
docker load < trifacta-docker-image-x.y.z.tar
```

5. Confirm that the image has been loaded. Execute the following command, which should list the Docker image:

```
docker images
```

6. You can now configure the Docker image. Please skip that section.

Build your own Docker image

As needed, you can build your own Docker image.

Requirements

- Docker version 17.12 or later. Docker version must be compatible with the following version(s) of Docker Compose.
- Docker Compose 1.24.1. It should be compatible with above version(s) of Docker.

Build steps

1. Acquire the RPM file from the FTP site:

NOTE: You must acquire the e17 RPM file for this release.

2. In your Docker environment, copy the `trifacta-server*.rpm` file to the same level as the `Dockerfile`.
3. Verify that the `docker-files` folder and its contents are present.
4. Use the following command to build the image:

```
docker build -t trifacta/server-enterprise:latest .
```


- This process could take about 10 minutes. When it is completed, you should see the build image in the Docker list of local images.

NOTE: To reduce the size of the Docker image, the Dockerfile installs the trifacta-server RPM file in one stage and then copies over the results to the final stage. The RPM is not actually installed in the final stage. All of the files are properly located.

- You can now configure the Docker image.

Configure Docker Image

Before you start the Docker container, you should review the properties for the Docker image. In the provided image, please open the appropriate `docker-compose` file:

File	Description
<code>docker-compose-local-postgres.yaml</code>	Database properties in this file are pre-configured to work with the installed instance of PostgreSQL, although you may wish to change some of the properties for security reasons.
<code>docker-compose-local-mysql.yaml</code>	Database properties in this file are pre-configured to work with the installed instance of MySQL, although you may wish to change some of the properties for security reasons.
<code>docker-compose-remote-db.yaml</code>	The Trifacta databases are to be installed on a remote server that you manage. <div style="border: 1px solid #ccc; border-radius: 10px; padding: 5px; margin: 10px auto; width: fit-content;"> <p>NOTE: Additional configuration is required.</p> </div>

NOTE: You may want to create a backup of this file first.

Key general properties:

NOTE: Avoid modifying properties that are not listed below.

Property	Description
<code>image</code>	This reference must match the name of the image that you have acquired.
<code>container_name</code>	Name of container in your Docker environment.
<code>ports</code>	Defines the listening port for the Trifacta application. Default is 3005. <div style="border: 1px solid #ccc; border-radius: 10px; padding: 5px; margin: 10px auto; width: fit-content;"> <p>NOTE: If you must change the listening port, additional configuration is required after the image is deployed. See <i>Change Listening Port</i>.</p> </div>

Database properties:

These properties pertain to the database installation to which the Trifacta application connects.

Property	Description
DB_INIT	<p>If set to <code>true</code>, database initialization steps are performed at startup.</p> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> <p>NOTE: This step applies only if you are starting the container for the first time, and the databases will be installed locally.</p> </div>
DB_TYPE	Set this value to <code>postgresql</code> or <code>mysql</code> .
DB_HOST_NAME	Hostname of the machine hosting the databases. Leave value as <code>localhost</code> for local installation.
DB_HOST_PORT	<p>(Remote only) Port number to use to connect to the databases. Default is <code>5432</code>.</p> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> <p>NOTE: If you are modifying, additional configuration is required after installation is complete. See <i>Change Database Port</i> in the Databases Guide.</p> </div>
DB_ADMIN_USERNAME	<p>Admin username to be used to create DB roles/databases. Modify this value for remote installation.</p> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> <p>NOTE: If you are modifying this value, additional configuration is required. Please see the documentation for your database version.</p> </div>
DB_ADMIN_PASSWORD	Admin password to be used to create DB roles/databases. Modify this value for remote installation.

Kerberos properties:

If your Hadoop cluster is protected by Kerberos, please review the following properties.

Property	Description
KERBEROS_KEYTAB_FILE	<p>Full path inside of the container where the Kerberos keytab file is located. Default value:</p> <div style="border: 1px dashed #ccc; padding: 10px; margin-top: 10px; text-align: center;"> <pre>/opt/trifacta/conf/trifacta. keytab</pre> </div> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> <p>NOTE: The keytab file must be imported and mounted to this location. Configuration details are provided later.</p> </div>

KERBEROS_KRB5_CONF	Full path inside of the container where the Kerberos krb5.conf file is located. Default: <div style="border: 1px dashed blue; padding: 10px; margin: 10px auto; width: fit-content;"> <code>/opt/krb-config/krb5.conf</code> </div>
--------------------	---

Hadoop distribution client JARs:

Please enable the appropriate path to the client JAR files for your Hadoop distribution. In the following example, the Cloudera path has been enabled, and the Hortonworks path has been disabled:

```

# Mount folder from outside for necessary hadoop client jars
# For CDH
- /opt/cloudera:/opt/cloudera
# For HDP
#- /usr/hdp:/usr/hdp

```

Please modify these lines if you are using Hortonworks.

Volume properties:

These properties govern where volumes are mounted in the container.

NOTE: These values should not be modified unless necessary.

Property	Description
volumes.conf	Full path in container to the Trifacta configuration directory. Default: <div style="border: 1px dashed blue; padding: 10px; margin: 10px auto; width: fit-content;"> <code>/opt/trifacta/conf</code> </div>
volumes.logs	Full path in container to the Trifacta logs directory. Default: <div style="border: 1px dashed blue; padding: 10px; margin: 10px auto; width: fit-content;"> <code>/opt/trifacta/logs</code> </div>
volumes.license	Full path in container to the Trifacta license directory. Default: <div style="border: 1px dashed blue; padding: 10px; margin: 10px auto; width: fit-content;"> <code>/trifacta-license</code> </div>

Start Server Container

After you have performed the above configuration, execute the following to initialize the Docker container:

```
docker-compose -f <docker-compose-filename>.yaml run trifacta initfiles
```

When the above is started for the first time, the following directories are created on the localhost:

Directory	Description
./trifacta-data	Used by the Trifacta container to expose the <code>conf</code> and <code>logs</code> directories.

Import Additional Configuration Files

After you have started the new container, additional configuration files must be imported.

Import license key file

The Trifacta license file must be staged for use by the platform. Stage the file in the following location in the container:

NOTE: If you are using a non-default path or filename, you must update the `<docker-compose-filename>.yaml` file.

```
trifacta-license/license.json
```

Import Hadoop distribution libraries

If the container you are creating is on the edge node of your Hadoop cluster, you must provide the Hadoop libraries.

1. You must mount the Hadoop distribution libraries into the container. For more information on the libraries, see the documentation for your Hadoop distribution.
2. The Docker Compose file must be made aware of these libraries. Details are below.

Import Hadoop cluster configuration files

Some core cluster configuration files from your Hadoop distribution must be provided to the container. These files must be copied into the following directory within the container:

```
./trifacta-data/conf/hadoop-site
```

For more information, see *Configure for Hadoop* in the Configuration Guide.

Install Kerberos client

If Kerberos is enabled, you must install the Kerberos client and keytab on the node container. Copy the keytab file to the following stage location:

```
/trifacta-data/conf/trifacta.keytab
```

See *Configure for Kerberos Integration* in the Configuration Guide.

Perform configuration changes as necessary

The primary configuration file for the platform is in the following location in the launched container:

```
/opt/trifacta/conf/trifacta-conf.json
```

NOTE: Unless you are comfortable working with this file, you should avoid direct edits to it. All subsequent configuration can be applied from within the application, which supports some forms of data validation. It is possible to corrupt the file using direct edits.

Configuration topics are covered later.

Start and Stop the Container

Stop container

Stops the container but does not destroy it.

NOTE: Application and local database data is not destroyed. As long as the `<docker-compose-filename> .yaml` properties point to the correct location of the `*-data` files, data should be preserved. You can start new containers to use this data, too. Do not change ownership on these directories.

```
docker-compose -f <docker-compose-filename>.yaml stop
```

Restart container

Restarts an existing container.

```
docker-compose -f <docker-compose-filename>.yaml start
```

Recreate container

Recreates a container using existing local data.

```
docker-compose -f <docker-compose-filename>.yaml up --force-recreate -d
```

Stop and destroy the container

Stops the container and destroys it.

The following also destroys all application configuration, logs, and database data. You may want to back up these directories first.

```
docker-compose -f <docker-compose-filename>.yaml down
```

Local PostgreSQL:

```
sudo rm -rf trifacta-data/ postgres-data/
```

Local MySQL or remote database:

```
sudo rm -rf trifacta-data/
```

Verify Deployment

1. Verify access to the server where the Trifacta platform is to be installed.
2. **Cluster Configuration:** Additional steps are required to integrate the Trifacta platform with the cluster. See *Prepare Hadoop for Integration with the Platform* in the Planning Guide.
3. Start the platform within the container. See *Start and Stop the Platform*.

Configuration

After installation is complete, additional configuration is required. You can complete this configuration from within the application.

Steps:

1. Login to the application. See *Login*.
2. The primary configuration interface is the Admin Settings page. From the left menu, select **Settings menu > Settings > Admin Settings**. For more information, see *Admin Settings Page* in the Admin Guide.
3. In the Admin Settings page, you should do the following:
 1. Configure password criteria. See *Configure Password Criteria*.
 2. Change the Admin password. See *Change Admin Password*.
4. Workspace-level configuration can also be applied. From the left menu, select **Settings menu > Settings > Workspace Admin**. For more information, see *Workspace Admin Page* in the Admin Guide.

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* in the Configuration Guide.

Install on CentOS and RHEL

Contents:

- *Preparation*
 - *Required version of RPM for CentOS*
 - *Installation*
 - *1. Install Dependencies*
 - *2. Install JDK*
 - *3. Install Trifacta package*
 - *4. Verify Install*
 - *5. Install License Key*
 - *6. Install Hadoop dependencies*
 - *7. Store install packages*
 - *Install Hadoop Dependencies*
 - *Included Dependencies*
 - *Acquire Other Dependencies*
 - *Install Dependencies*
 - *Next Steps*
 - *Install and configure Trifacta databases*
 - *Install configuration*
 - *Install Desktop Application*
-

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 *Product Support Matrix* in the Planning Guide.

Preparation

Before you install software, please review and verify 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. Review key sections of the Planning Guide:
 1. Review the *System Requirements* and verify that all required components have been installed.
 2. Verify that all required *System Ports* are opened on the node.
 3. Review the *System Dependencies* in the Planning Guide.
 4. **Cluster Configuration:** Additional steps are required to integrate the Trifacta platform with the cluster. See *Prepare Hadoop for Integration with the Platform* in the Planning Guide.
2. Acquire your *License Key*.
3. Install and verify operations of the datastore, if used.

NOTE: Access to the Spark cluster is required.

4. Verify access to the server where the Trifacta platform is to be installed.

Required version of RPM for CentOS

The installer for the Trifacta platform on CentOS/RHEL requires RPM version 4.11.3-40. Please upgrade if necessary.

NOTE: On CentOS/RHEL 7.4 or earlier, the installer may fail to launch on earlier versions of RPM.

Installation

Tip: The Python setup tools can be useful for debugging startup issues. To install:

```
yum install python-setuptools
```

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. 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. For more information, see *System Requirements* in the Planning Guide.

Tip: 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. Install Hadoop dependencies

If you are integrating with a supported Hadoop cluster, you must install the dependencies for the Hadoop cluster on the Trifacta node. See below.

7. Store install packages

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

Install Hadoop Dependencies

If you are integrating Hadoop cluster, the associated Hadoop dependencies must be installed on the Trifacta® node.

Included Dependencies

The Hadoop dependencies for the latest supported version of each Hadoop distribution are included in the Trifacta software distribution.

Supported Versions:

- *Supported Deployment Scenarios for Cloudera*
- *Supported Deployment Scenarios for Hortonworks*
- *Configure for EMR* in the Configuration Guide

Not required for:

NOTE: If you are integrating with one of the following running environments, please skip installing Hadoop dependencies.

Azure running environments:

- HDI
- Azure Databricks

Acquire Other Dependencies

Hadoop dependencies for other versions of the Hadoop distribution can be acquired from the Trifacta FTP site using one of the following methods.

Via a web browser

1. Log in: <https://ftp.trifacta.com/login>
2. Browse to the following directory:

```
Releases/Trifacta_x.y/hadoop/
```

where: `x.y` corresponds to the release number that you are installing (e.g. Release 6.8).

3. Download the following file: `hadoop_deps.tar.gz`

Via WGET

Example is for Release 6.8:

```
wget --user CustomerUsername --ask-password ftps://ftp.trifacta.com  
/Releases/Trifacta_6.8/hadoop/hadoop-deps.tar.gz
```

Via SFTP

Example is for Release 6.8:

```
sftp CustomerUsername@ftp.trifacta.com:Releases/Trifacta_6.8/hadoop  
/hadoop-deps.tar.gz .
```

Via CURL

Example is for Release 6.8:

```
curl -O -C - -u CustomerUsername:CustomerPassword ftps://ftp.trifacta.com  
/Releases/Trifacta_6.8/hadoop/hadoop-deps.tar.gz
```

Via FTP/FTPS

1. Access the FTP server via your preferred FTP client.
2. Browse to the following directory:

```
Releases/Trifacta_x.y/hadoop/
```

where: `x.y` corresponds to the release number that you are installing (e.g. Release 6.8).

3. Download the following file: `hadoop_deps.tar.gz`

Install Dependencies

If needed, transfer the download to the Trifacta node.

Extract it to the following directory:

```
sudo tar -vxf hadoop-deps.tar --directory /opt/trifacta/
```

NOTE:

After you extract the files to the target directory, verify that the ownership of the new directory (`/opt/trifacta/hadoop-deps/`) and its subfolders match the ownership settings for the rest of the Trifacta installation in `/opt/trifacta`.

Next Steps

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* in the Databases Guide.

Install configuration

After installation is complete, additional configuration is required to make the platform operational. See *Install Configuration*.

Install Desktop Application

You can deploy the Wrangler Enterprise desktop application as a desktop client to enable end-users to connect to the Trifacta application without using one of the supported web browsers. See *Install Desktop Application*.

Install on Ubuntu

Contents:

- *Preparation*
- *Installation*
 - *1. Install Dependencies*
 - *2. Install JDK*
 - *3. Install Trifacta package*
 - *4. Verify Install*
 - *5. Install License Key*
 - *6. Install Hadoop dependencies*
 - *7. Store install packages*
- *Install Hadoop Dependencies*
 - *Included Dependencies*
 - *Acquire Other Dependencies*
 - *Install Dependencies*
- *Next Steps*
 - *Install and configure Trifacta databases*
 - *Install configuration*
 - *Install Desktop Application*

This guide takes you through the steps for installing Trifacta® Wrangler Enterprise software on Ubuntu.

For more information on supported operating system versions, see *Product Support Matrix* in the Planning Guide.

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. Review key sections of the Planning Guide:
 1. Review the *System Requirements* and verify that all required components have been installed.
 2. Verify that all required *System Ports* are opened on the node.
 3. Review the *System Dependencies* in the Planning Guide.
 4. **Cluster configuration:** Additional steps are required to integrate the Trifacta platform with the cluster. See *Prepare Hadoop for Integration with the Platform* in the Planning Guide.
2. Acquire your *License Key*.
3. Install and verify operations of the datastore, if used.

NOTE: Access to the cluster may be required.

4. Verify access to the server where the Trifacta platform is to be installed.

Installation

Tip: The Python setup tools can be useful for debugging startup issues. To install:

```
wget https://bootstrap.pypa.io/ez_setup.py -O - | python
```

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 . 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.

1. Ubuntu 14.04:

```
sudo apt-get install nginx=1.12.2-1~trusty nodejs=10.13.0-
lnodesource1
```

2. Ubuntu 16.04

```
sudo apt-get install nginx=1.12.2-1~xenial nodejs=10.13.0-
lnodesource1
```

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. For more information, see *System Requirements* in the Planning Guide.

Tip: 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:

NOTE: If you encounter errors running the following command, execute the next command anyway. If that command completes without error, the installation is ok.

```
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. Install Hadoop dependencies

If you are integrating with a supported Hadoop cluster, you must install the dependencies for the Hadoop cluster on the Trifacta node. See below.

7. Store install packages

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

Install Hadoop Dependencies

If you are integrating Hadoop cluster, the associated Hadoop dependencies must be installed on the Trifacta® node.

Included Dependencies

The Hadoop dependencies for the latest supported version of each Hadoop distribution are included in the Trifacta software distribution.

Supported Versions:

- *Supported Deployment Scenarios for Cloudera*
- *Supported Deployment Scenarios for Hortonworks*
- *Configure for EMR* in the Configuration Guide

Not required for:

NOTE: If you are integrating with one of the following running environments, please skip installing Hadoop dependencies.

Azure running environments:

- HDI
- Azure Databricks

Acquire Other Dependencies

Hadoop dependencies for other versions of the Hadoop distribution can be acquired from the Trifacta FTP site using one of the following methods.

Via a web browser

1. Log in: `https://ftp.trifacta.com/login`
2. Browse to the following directory:

```
Releases/Trifacta_x.y/hadoop/
```

where: `x.y` corresponds to the release number that you are installing (e.g. Release 6.8).

3. Download the following file: `hadoop_deps.tar.gz`

Via WGET

Example is for Release 6.8:

```
wget --user CustomerUsername --ask-password ftps://ftp.trifacta.com  
/Releases/Trifacta_6.8/hadoop/hadoop_deps.tar.gz
```

Via SFTP

Example is for Release 6.8:

```
sftp CustomerUsername@ftp.trifacta.com:Releases/Trifacta_6.8/hadoop  
/hadoop_deps.tar.gz .
```

Via CURL

Example is for Release 6.8:

```
curl -O -C - -u CustomerUsername:CustomerPassword ftps://ftp.trifacta.com  
/Releases/Trifacta_6.8/hadoop/hadoop_deps.tar.gz
```

Via FTP/FTPS

1. Access the FTP server via your preferred FTP client.
2. Browse to the following directory:

```
Releases/Trifacta_x.y/hadoop/
```

where: `x.y` corresponds to the release number that you are installing (e.g. Release 6.8).

3. Download the following file: `hadoop_deps.tar.gz`

Install Dependencies

If needed, transfer the download to the Trifacta node.

Extract it to the following directory:

```
sudo tar -vxf hadoop-deps.tar --directory /opt/trifacta/
```

NOTE:

After you extract the files to the target directory, verify that the ownership of the new directory (`/opt/trifacta/hadoop-deps/`) and its subfolders match the ownership settings for the rest of the Trifacta installation in `/opt/trifacta`.

Next Steps

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* in the Databases Guide.

Install configuration

After installation is complete, additional configuration is required to make the platform operational. See *Install Configuration*.

Install Desktop Application

You can deploy the Wrangler Enterprise desktop application as a desktop client to enable end-users to connect to the Trifacta application without using one of the supported web browsers. See *Install Desktop Application*.

License Key

Contents:

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

Download license key file

If you have not done so already, the license key file is available where you have acquired the installation package. Please download `license.json`.

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.

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 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* in the Admin Guide.

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* in the Admin Guide.

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 missing or has expired. Please contact *Trifacta Support*.

Start and Stop the Platform

Contents:

- *Command Line*
 - *Start*
 - *Restart*
 - *Stop*
- *Debugging*
- *Troubleshooting*
 - *Error - "ImportError: No module named pkg_resources" error in supervisord*
 - *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* in the Admin Guide.

Command Line

NOTE: In an Azure HDI environment, you must perform platform start and stop operations from `/opt/trifacta`. Running these commands from other directories, such as `/root`, can cause service issues.

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
```

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

Restart

Command:

```
service trifacta restart
```

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

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=trifacta"
```

Troubleshooting

Error - "ImportError: No module named pkg_resources" error in supervisord

When you start the platform for the first time, you may receive the following error:

```
Traceback (most recent call last):
  File "/usr/local/bin/supervisord", line 5, in <module>
    from pkg_resources import load_entry_point
ImportError: No module named pkg_resources
```

This error occurs when the supervisord process is starting. The Trifacta platform fails to complete startup.

Solution:

This issue is caused by a missing package for supervisord. The simplest solution is to install the Python setup tools on the Trifacta node. Commands are listed below.

NOTE: These commands must be executed as root user.

CentOS/RHEL:

```
yum install python-setuptools
```

Ubuntu:

```
wget https://bootstrap.pypa.io/ez_setup.py -O - | python
```

After installation is complete, restart the platform.

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* in the Admin Guide.

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**.

NOTE: If you enter a mismatched password and password confirmation, registration fails as expected. If you correct the mismatch and try to register again, registration may fail again. The workaround is to clear your browser cache and register again. This is a known issue.

- 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* in the Configuration Guide.

After you login, you are placed in the Home page. See *Home Page* in the User Guide.

- If you are using S3 as your base storage layer and per-user authentication has been enabled, you must provide the AWS credentials to connect to your storage. From the left navigation bar, select **Settings > Storage** and then select the AWS option. See *Configure Your Access to S3* in the Configuration Guide.
- For a basic walkthrough of the Trifacta application, see *Workflow Basics* in the User Guide.

Product Documentation:

NOTE: After you log in the Trifacta application, you can access online documentation for your product. Select **Help menu > Documentation**.

To logout:

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

Install Reference

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

Install SSL Certificate

Contents:

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

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.

1. The `.key` file should go into the `/etc/ssl/private/` directory.
2. 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.

3. 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

ssl_certificate	Path to the file of the certificate bundle that you created on the server. This value may not require modification.
ssl_certificate_key	Path to the .key 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.

Disable default port

If you wish to access through the default port (3005), you must do so external to the platform and through the node itself.

NOTE: The Trifacta platform requires access to the default port internally. You cannot disable external access to this port through the platform. You must disable through the operating system.

For more information, please see the documentation provided with your operating system distribution.

Troubleshooting

Problem - SELinux blocks proxy service from communicating with internal app service

If the Trifacta platform is installed on SELinux, the operating system blocks communications between the service that manages the proxy between users and the application and the service that manages internal application communications.

To determine if this problem is present, execute the following command:

```
sudo cat /var/log/audit/audit.log | grep nginx | grep denied
```

The problem is present if an error similar to the following is returned:

```
type=AVC msg=audit(1555533990.045:1826142): avc: denied { name_connect } for pid=25516 comm="nginx" dest=3005 scontext=system_u:system_r:httpd_t:s0
```

For more information on this issue, see <https://www.nginx.com/blog/using-nginx-plus-with-selinux>.

Solution:

The solution is to enable the following network connection through the operating system:

```
sudo setsebool -P httpd_can_network_connect 1
```

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.

Install Desktop Application

Contents:

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

NOTE: In a future release, the Wrangler Enterprise desktop application will be deprecated. Please switch to a supported version of Google Chrome or Mozilla Firefox. Support for Edge Chromium is expected in a future release. See *Desktop Requirements*.

If your environment does not enable use of one of the supported browser versions, 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 supported browser version on the local desktop.

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

Install Process

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 node before deploying the instances of the application. See *Configure for Desktop Application* in the Configuration Guide.

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 node:

```
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 server, please enter the full URL including port number of the Trifacta instance to which you are connecting.
 1. By default, the server is available over port 3005. For more information, please contact your IT administrator.
 2. 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 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 *320735828* below.

3. When the proper URL and port number are provided, you may launch the application.

4. If your environment contains multiple server deployments, you can select the one to which to connect:

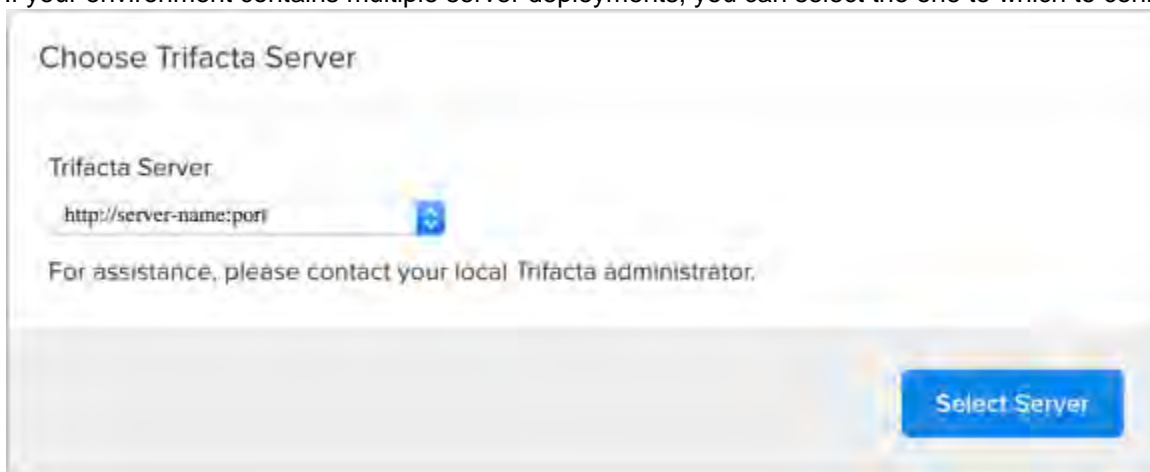


Figure: Choose Server

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 server, please do the following:

1. Verify that you are connecting to the appropriate URL.
 1. If you are connecting to the incorrect URL, please uninstall the application and re-install using the MSI file. See *320735828* 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 server that does not support the Wrangler Enterprise desktop application. Please verify that your connection URL is pointed to a supported instance of the server.

Supported Deployment Scenarios for AWS

Contents:

- *AWS Deployment Scenarios*
- *AWS Installations*
 - *Trifacta Data Preparation for Amazon Redshift and S3 on AWS Marketplace (AMI)*
 - *Trifacta Wrangler Enterprise on AWS Marketplace with EMR*
 - *Trifacta Wrangler Enterprise on EC2 Instance*

- *AWS Integrations*

AWS Deployment Scenarios

The following are the basic AWS deployment scenarios.

Trifacta platform deployed through AWS Marketplace:

Deployment Scenario	Trifacta node installation	Base Storage Layer	Storage - S3	Storage - Redshift	Cluster	Notes
Trifacta Data Preparation for Amazon Redshift and S3 AWS install through AWS Marketplace CloudFormation template	EC2	S3	read/write	read/write	None	Trifacta Data Preparation for Amazon Redshift and S3 does not support integration with any running environment clusters. All job execution occurs on the Trifacta node in the Trifacta Photon running environment. This scenario is suitable for smaller user groups and data volumes.
Trifacta Wrangler Enterprise AWS install through AWS Marketplace CloudFormation template - with integration to EMR cluster	EC2	S3	read/write	read/write	EMR	This deployment scenario integrates by default with an EMR cluster, which is created as part of the process. It does not support integration with a Hadoop cluster.

Trifacta platform installed on AWS:

Deployment Scenario	Trifacta node installation	Base Storage Layer	Storage - S3	Storage - Redshift	Cluster	Notes
Trifacta Wrangler Enterprise AWS install with S3 read access	EC2	HDFS	read only	Not supported	EMR	When HDFS is the base storage layer, the only accessible AWS resources is read-only access to S3.
Trifacta Wrangler Enterprise AWS install with S3 read/write access	EC2	S3	read/write	read/write	EMR	

Trifacta platform installed on-premises and integrated with AWS resources:

Deployment Scenario	Trifacta node installation	Base Storage Layer	Storage - S3	Storage - Redshift	Cluster	Notes
Trifacta Wrangler Enterprise on-premises install with S3 read access	On-premises	HDFS	read only	Not supported	Hadoop	When HDFS is the base storage layer, the only accessible AWS resources is read-only access to S3.
Microsoft Azure						Integration with AWS-based resources is not supported.

Legend and Notes:

Column	Notes
Deployment Scenario	Description of the AWS-connected deployment
Trifacta node installation	Location where the Trifacta node is installed in this scenario. All AWS installations are installed on EC2 instances.
Base Storage Layer	When the Trifacta platform is first installed, the base storage layer must be set. <div style="border: 1px solid gray; padding: 5px; margin: 5px 0;"> <p>NOTE: After you have begun using the product, you cannot change the base storage layer.</p> </div> <div style="border: 1px solid gray; padding: 5px; margin: 5px 0;"> <p>NOTE: Read/write access to AWS-based resources requires that S3 be set as the base storage layer.</p> </div>
Storage - S3	Trifacta Wrangler Enterprise supports read access to S3 when the base storage layer is set to HDFS. For read/write access to S3, the base storage layer must be set to S3.
Storage - Redshift	For access to Redshift, the base storage layer must be set to S3.
Cluster	List of cluster types that are supported for integration and job execution at scale. <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. • Access to an EMR cluster requires S3 to be the base storage layer. • Smaller jobs can be executed on the Trifacta Photon running environment, which is hosted on the Trifacta node itself. • For more information, see <i>Running Environment Options</i> in the Configuration Guide.
Notes	Any additional notes

AWS Installations

Trifacta Data Preparation for Amazon Redshift and S3 on AWS Marketplace (AMI)

Through the Amazon Marketplace, you can license and deploy an AMI of Trifacta Data Preparation for Amazon Redshift and S3, which does not require integration with a clustered running environment. All job execution happens within the AMI on the EC2 instance that you deploy. For more information, see the Trifacta Data Preparation for Amazon Redshift and S3 listing on AWS Marketplace.

Trifacta Wrangler Enterprise on AWS Marketplace with EMR

You can deploy an AMI of the Trifacta platform onto an EC2 instance. For more information, see the Trifacta Wrangler Enterprise listing for AWS Marketplace.

You can deploy it in either of the following ways:

1. Auto-create a 3-node EMR cluster.
2. Integrate it later with your pre-existing EMR cluster.

For more information, see the Trifacta Wrangler Enterprise listing on AWS Marketplace.

Trifacta Wrangler Enterprise on EC2 Instance

When the Trifacta platform is installed on AWS, it is deployed on an EC2 instance. Through the EC2 console, there are a few key parameters that must be specified.

NOTE: After you have created the instance, you should retain the `instancetype` from the console, which must be applied to the configuration in the Trifacta platform.

AWS Integrations

The following table describes the different AWS 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.

AWS Service	Description	Base Storage Layer	Other Required AWS Services
EC2	Amazon Elastic Compute Cloud (EC2) can be used to host the Trifacta node in a scalable cloud-based environment. The following deployments are supported: <ul style="list-style-type: none">• Trifacta Wrangler Enterprise with or without access to an EMR cluster• Trifacta Data Preparation for Amazon Redshift and S3 on an AMI	Base storage layer can be S3 or HDFS. If set to HDFS, only read access to S3 is permitted.	
S3	Amazon Simple Storage Service (S3) can be used for reading data sources, writing job results, and hosting the Trifacta databases.	Base storage layer can be S3 or HDFS. If set to HDFS, only read access to S3 is permitted.	

Redshift	Amazon Redshift provides a scalable data warehouse platform, designed for big data analytics applications. The Trifacta platform can be configured to read and write from Amazon Redshift database tables.	Base Storage Layer = S3	S3
EMR	For more information on supported versions of EMR, see <i>Configure for EMR</i> in the Configuration Guide.	Base Storage Layer = S3	EC2 instance
Amazon RDS	Optionally, the Trifacta databases can be installed on Amazon RDS. For more information, see <i>Install Databases on Amazon RDS</i> in the Databases Guide.	Base Storage Layer = S3	

AWS Marketplace integrations:

AWS Service	Description	Base Storage Layer	Other Required AWS Services
AMI	Through the AWS Marketplace, you can license and install an Amazon Machine Image (AMI) instance of Trifacta Data Preparation for Amazon Redshift and S3. This product is intended for smaller user groups that do not need large-scale processing of Hadoop-based clusters.	Base Storage Layer = S3 NOTE: HDFS is not supported.	EC2 instance
EMR	Through the AWS Marketplace, you can license and install an AMI specifically configured to work with Amazon Elastic Map Reduce (EMR), a Hadoop-based data processing platform.	Base Storage Layer = S3	AMI

Uninstall

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

NOTE: All platform and cluster 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 © 2020 - Trifacta, Inc.
All rights reserved.