

# Install On-Premises

## Contents:

- *Scenario Description*
  - *Limitations*
    - *Deployment Limitations*
    - *Product Limitations*
  - *Pre-requisites*
  - *Preparation*
  - *Deploy the Cluster*
    - *Prepare the cluster*
  - *Deploy the Trifacta node*
  - *Install Workflow*
  - *Next Steps*
- 

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

## Scenario Description

- Installation of Trifacta Wrangler Enterprise on a server on-premises
- Installation of Trifacta databases on a server on-premises
- Integration with a supported Hadoop cluster on premises.
- Base storage layer of HDFS

## Limitations

### Deployment Limitations

None.

### 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*.

## Preparation

Before you install Trifacta Wrangler Enterprise, please complete the following steps.

1. **Deploy Hadoop cluster:** In this scenario, the Trifacta platform does not create a Hadoop cluster.

**NOTE:** Installation and maintenance of a working Hadoop cluster is the responsibility of the customer. Guidance is provided below on the requirements for integrating the platform with the cluster.

2. **Deploy Trifacta node:** Trifacta Wrangler Enterprise must be installed on an edge node of the cluster.

Details are below.

## Deploy the Cluster

In your enterprise infrastructure, you must deploy a cluster using a supported version of Hadoop to manage the expected data volumes of your Trifacta jobs.

The Trifacta platform supports integration with the following cluster types. For more information on the supported versions, please see the listed sections below.

- See *Supported Deployment Scenarios for Cloudera*.
- See *Supported Deployment Scenarios for Hortonworks*.
- 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 in the following section.

### Job execution:

- By default, smaller jobs are executed in the Photon running environment on the Trifacta node.
- Larger jobs are executed using Spark on the integrated Hadoop cluster. A supported version of Spark must be installed on the cluster. For more information, see *System Requirements* in the Planning Guide.

## Prepare the cluster

Before installing software, please complete the following steps if you are integrating with a Hadoop cluster.

Before you begin, please verify or complete the following:

1. On the Hadoop cluster:
  - a. Create a user [`hadoop.user (default=trifacta)`] and a group for it [`hadoop.group (default=trifactausers)`].
  - b. Create the following directories:
    - i. `/trifacta`
    - ii. `/user/trifacta`
  - c. Change the ownership of `/trifacta` and `/user/trifacta` to `trifacta:trifacta` or the corresponding values for the Hadoop user in your environment.

**NOTE:** You must verify that the [`hadoop.user`] user has complete ownership and full access to Read, Write and Execute on these directories recursively.

2. Verify that WebHDFS is configured and running on the cluster.
3. Software installation is completed on a dedicated node in the cluster. The user installing the Trifacta software must have sudo access.

4. If you are installing on a server with an older instance of Postgres, you should remove the older version or change the default ports.

For more information, see *Prepare Hadoop for Integration with the Platform*.

Additional users may be required. For more information, see *Required Users and Groups* in the Planning Guide.

## Deploy the Trifacta node

An edge node of the cluster is required to host the Trifacta platform software. For more information on the requirements of this node, see *System Requirements* in the Planning Guide.

## Install Workflow

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:
  - a. See *Login*.
  - b. As soon as you login, you should change the password on the admin account. In the left menu bar, select **User menu > Admin console > 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*.