



TRIFACTA

Admin Guide

Version: 5.1

Doc Build Date: 12/07/2018

Copyright © Trifacta Inc. 2018 - 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. *Admin* . . . 4
 - 1.1 *Verify Operations* . . . 4
 - 1.2 *Create User Account* . . . 5
 - 1.3 *Create Admin Account* . . . 6
 - 1.4 *Manage Users* . . . 7
 - 1.5 *System Services and Logs* . . . 11
 - 1.6 *Maintenance Release Updater* . . . 15
 - 1.7 *Backup and Recovery* . . . 17
 - 1.8 *Admin Reference* . . . 22
 - 1.8.1 *Admin Settings Page* . . . 23
 - 1.8.2 *Deployments Page* . . . 26

Admin

These topics pertain to the administration of Trifacta® Wrangler Enterprise.

Topics:

- *Verify Operations*
- *Create User Account*
- *Create Admin Account*
- *Manage Users*
- *System Services and Logs*
- *Maintenance Release Updater*
- *Backup and Recovery*
- *Admin Reference*
 - *Admin Settings Page*
 - *Deployments Page*

Verify Operations

After you have applied a configuration change to the platform and restarted, you can use the following steps to verify that the platform is working correctly.

- If your configuration change was applied to `trifacta-conf.json`, you should restart the platform before continuing. See *Start and Stop the Platform*.

Prepare Your Sample Dataset

To complete this test, you should locate or create a simple dataset. Characteristics:

- Two or more columns.
- If there are specific data types that you would like to test, please be sure to include them in the dataset.
- A minimum of 25 rows is required for best results of type inference.
- Ideally, your dataset is a single file or sheet.

Your dataset should be created in the format that you wish to test. If you are testing an integration, you should create it in the format that is supported by the connected datastore.

Tip: It might be easiest to create your dataset in Microsoft Excel in your local desktop.

Step-by-step Guide

Steps:

1. Store your example dataset in a location to which the application is connected.
2. Login to the application. See *Login*.
3. In the application menu bar, click **Library**.
4. Click **Import Data**. See *Import Data Page*.
 - a. Select the connection where the dataset is stored.
 - b. Navigate the tree to select the dataset.
 - c. Click the + icon next to the dataset to upload.
 - d. In the right panel, click the Add Dataset to a Flow checkbox. Enter a name for the new flow.
 - e. Click **Import and Add to Flow**.
 - f. **Troubleshooting:** At this point, you have read access to your datastore from the platform. If not, please check the logs, permissions, and your Trifacta® configuration .
5. In the Library page, click the name of the dataset you just imported.
 - a. In the Transformer page, some steps are automatically added to the recipe for you. So, you can run the job immediately.

- b. You can add additional steps if desired. See *Transformer Page*.
6. Click **Run Job**.
 - a. Select the running environment to generate the results.
 - b. To add other output formats, click **Add Publishing Destination**. Configure the output formats and locations.
 - c. To test dataset profiling, click the Profile Results checkbox. Note that profiling runs as a separate job and may take considerably longer.
 - d. **Troubleshooting**: Later you can re-run this job on a different running environment. Some formats are not available across all running environments.
7. When the job completes, you should see a success message under the Jobs tab in the Flow View page.
 - a. **Troubleshooting**: Either the Transform job or the Profiling job may break. To localize the problem, try re-running a job by deselecting the broken job type or running the job on the Trifacta Server. You can also download the log files to try to identify the problem.
8. Click **View Results** from the context menu for the job listing. In the Job Results page, you can see a visual profile of the generated results.
 - a. See *Job Results Page*.
9. Click **Export Results**. In the Export Results window, click the CSV and JSON links to download the results to your local desktop.
10. Load these results into a local application to verify that the content looks ok.

Create User Account

By default, users can create their own accounts.

- As needed, self-registration can be disabled, so that all users must be created by an administrator. See *Configure User Self-Registration*.

Creating your own user account

Steps:

1. Users may self-register at the following address:
`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.
2. Click the Register link.
3. Enter your credentials in the spaces provided. A valid email address is required.
4. As soon as the account is created, you may login at the first address. See *Login*.

Creating users when self-registration is disabled

When self-registration is disabled, an administrator must manually create the accounts for users. Administrators can create accounts at the following address:

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

NOTE: If SSO or secure impersonation is enabled in your environment, administrators must apply a principal value to each newly created user. See *Manage Users*.

When a new account is created, an email is sent to the address for the created user.

Troubleshooting

Account Not Configured login error

If you have created a user account, you may see the following error message when you try to login:

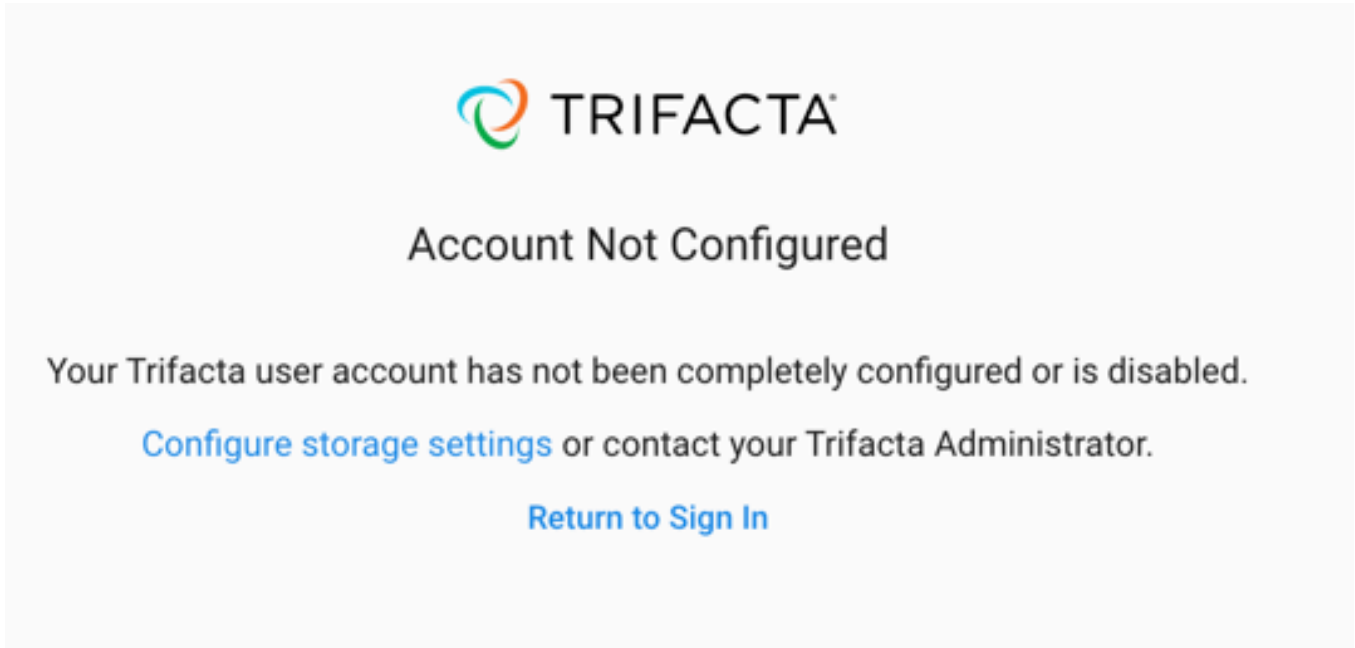


Figure: Account Not Configured

In this case, the account may require additional configuration. In SSO or Kerberos environments, an administrator may need to provision a SSO or Hadoop principal value to be applied to the user account. See *Admin Settings Page*.

Create Admin Account

Contents:

- *Create admin accounts*
- *Create admin account outside the UI*
 - *Without SSO*
 - *With SSO*
- *Create admin accounts under SSO via Trifacta command line interface*

You can create additional administrator accounts using one of the following methods.

The password for the default admin account should be changed as soon as you have access to the application. See *Change Admin Password*.

Create admin accounts

Steps:

1. Login using another admin account.

2. Create the account normally. See *Create User Account*.
3. Select **Settings menu > Admin Settings**.
4. In the Admin Settings page, click **Manage Users**.
5. For the newly created user, click the checkbox in the Trifacta Administrator column.
6. Save changes.
7. Login to the account and verify that the Admin Settings page is available.

Create admin account outside the UI

If you do not have access to an admin account through the application, you can create admin accounts for users from the Trifacta node using the `thewebapp/bin/ensure-user` command.

Without SSO

If Single Sign-On (SSO) is not enabled, use the following command:

```
<install_dir>/webapp/bin/ensure-user --admin "<FirstName LastName>"  
<e-mail> <password>
```

With SSO

If the environment uses SSO, the following command can create the admin user based on an Active Directory login:

```
<install_dir>/webapp/bin/ensure-user --admin "<FirstName LastName>"  
<e-mail> <password> <AD_LOGIN>
```

where:

<AD_LOGIN> is the active directory login for the user.

Create admin accounts under SSO via Trifacta command line interface

You can also use the Trifacta command line interface:

```
./trifacta_admin_cli.py --admin_username <admin_username>  
--admin_password <admin_pwd> edit_user --user_name <username>  
--enableAdmin --ssoPrincipal <sso_Principal>
```

See *CLI for User Admin*.

Manage Users

Contents:

- *Important Note on Permissions*
- *User Account Fields*
- *Edit Users*
 - *Password Reset*
 - *Disable User*
 - *Platform Roles*

- *Manage Users from the Command Line*
-

Through the Admin Settings page, administrators can manage aspects of user accounts, as well as other aspects of the deployment. See *Admin Settings Page*.

- To make changes to individual user accounts, click **Edit Users**.

NOTE: You must be an administrator to access this feature.

Important Note on Permissions

Depending on your deployment, access to stored assets can be governed by multiple sets of permissions. Access can be governed by:

- Trifacta® permissions
- Domain authentication (e.g. SSO) permissions
- Storage environment (e.g. Hadoop) permissions

When a Trifacta user shares a resource with another user, that second user may not have access to the underlying resource if one of the other permission sets does not provide it. In the Trifacta application, the issue may be surfaced as a generic read or access error, which may be difficult for end users to debug.

Tip: Where possible, you should use a single principal user for Trifacta users. If that is not possible, you should verify consistency in access permissions between Trifacta platform and the underlying storage environment.

User Account Fields

- **Name:** Display name for the user.
- **Email:** The value is the user ID. It must resolve to a valid, accessible email address. Some features of the platform fail to work correctly with invalid email addresses.
- **Trifacta Administrator:** Set this value to `true` to allow the user administrator privileges.

NOTE: You should limit the number of administrator accounts, which have extensive privileges in the application.

- **Roles:** Trifacta platform roles assigned to the user. See *Platform Roles* below.
- **AWS:** If you are connecting to AWS and using S3 for storage, you must populate the following values for each user to enable access to storage:
 - **AWS Key:** The AWS key value required for access for this user.
 - **AWS Secret:** The AWS secret value associated with the key.
 - **S3 Default Bucket:** The name of the S3 bucket that is the default storage for the user.

Tip: Additional buckets can be specified in individual user accounts. See *User Profile Page*.

- **SSO Principal:** If SSO is enabled, set this value to be the SSO principal value associated with this user.

NOTE: Required value for each user if SSO is enabled. See *Configure SSO for AD-LDAP*.

- **Hadoop Principal:** If secure impersonation is enabled, set this value to be the Hadoop principal value associated with this user.

NOTE: The user principal value should not include the realm.

NOTE: Required value if secure impersonation is enabled. See *Configure for secure impersonation*.

NOTE: If Kerberos is enabled, verify that all user principals that use the platform are also members of the group of the keytab user.

- **Created:** Timestamp when the account was created.
- **Updated:** Timestamp when the account was last modified.
- **Disabled:** If `true` the account is currently disabled. Else, the account is active. Edit the user to change access.
- **Last Login Time:** Timestamp for when the account was last used to access the application.
 - A value of `Never` indicates that the account has never been used.

Edit Users

Password Reset

NOTE: Wrangler Enterprise desktop application users cannot complete this method for password reset. Users of this version must use the self-service method of password reset, which must be enabled in the Trifacta platform. For more information, see *Enable Self-Service Password Reset*.

To reset a user's account password, click **Reset**. Copy the URL and paste it into an email to send the user.

Tip: If you are using Chrome for Windows, press `CTRL+C` in the popup to select the password reset URL.

Disable User

Non-admin users can be enabled or disabled as needed.

- To disable a user, click the checkbox in the Disabled column. Then, click **Submit**.

Platform Roles

The following platform roles are supported in the Trifacta platform.

- **Trifacta Administrator:** Provides administrator roles, which include administering users, changing configuration, and deletion of objects created by other users.

Avoid granting Trifacta Administrator role to many users.

- **Data Admin:** Enables user to use file browsers to browse external file systems such as HDFS, S3, and WASB.

NOTE: The Data Admin role is required to browse HDFS or other non-relational datastores. If an account lacks this role, dataset upload and download and access to JDBC data sources, including Hive, are still supported.

- **Deployment:** In a Development environment, this role can be added to a user's account to enable access to deployment management.
 - In a Production environment where deployment management is for the entire platform instance, this role does not apply.
 - For more information, see *Configure Deployment Management*.
 - For more information on deployment management, see *Overview of Deployment Management*.
- **user:** Please ignore this role. It does nothing at present.
- **admin:** Please ignore this role. It does nothing at present.
- **wrangler:** Enables access to the Trifacta application. All users accounts must have this role.

NOTE: All users accounts must have this role, which cannot be modified.

Default roles

Default roles provisioned to new users can be defined in platform configuration. You can apply this change through the *Admin Settings Page* (recommended) or `trifacta-conf.json`. For more information, see *Platform Configuration Methods*.

Modify the following property:

```
"webapp.defaultUserRoles": "wrangler,dataAdmin",
```

Manage Users from the Command Line

The Trifacta platform provides a command line interface that enables administrators to create, edit, and delete users. The CLI can also be used to generate password reset URLs.

For more information, see *CLI for User Admin*.

System Services and Logs

Contents:

- *Download Logs*
- *Batch Job Runner*
- *Data Service*
- *Java UDF Service*
- *Machine Learning Service*
- *Nginx Service*
- *Scheduling Service*
- *Spark Job Service*
- *Supervisord Server*
- *Time-Based Trigger Service*
- *VFS Service*
- *Webapp Service*
- *Additional logs*
 - *Job logs*

The Trifacta® platform provides the following major services. For each of the listed service, any relevant logs are listed.

The logging levels for many of these services can be modified through the Admin Settings page. See *Configure Logging for Services*.

Download Logs

System logs are maintained in the following directory: `/opt/trifacta/logs`

Administrators of the platform can now access all log files through the following URL:

```
http://<hostname>:<port_number>/logs
```

This endpoint displays for download all log files stored in the `/opt/trifacta/logs` directory.

Batch Job Runner

Description: This service manages the tracking of jobs submitted to the backend running environment.

Log File	Can Help With
<code>batch-job-runner.log</code>	<ul style="list-style-type: none">• Service errors and crashes• Determine execution environment of the job. Search for:<ul style="list-style-type: none">• <code>LocalJobRunner</code> = local execution• <code>YARNRunner</code> = execution in Hadoop• Communication errors back from environment• Status information on jobs• Status information on counts of job retries

Data Service

Description: Service prepares queries against JDBC interfaces, using internal REST API calls.

Log File	Can Help With
<code>data-service.log</code>	<ul style="list-style-type: none">• Initialization of communications through JDBC interface• Query failures

Java UDF Service

Description: Service enables the execution of Java-based user-defined functions within a transform recipe.

Log File	Can Help With
java-udf-service.log	<ul style="list-style-type: none">• Status of the service <div style="border: 1px solid green; padding: 5px; margin-top: 10px;">Tip: You can pass through messages on errors through Logger to this log, which can assist in diagnosing issues. See <i>Java UDFs</i>.</div>

Machine Learning Service

Description: ML service provides machine learning capabilities for the platform.

Log File	Can Help With
ml_service.log	<ul style="list-style-type: none">• This log is likely to contain information that is only useful if the ML service has crashed.

Nginx Service

Description: Nginx is a proxy server embedded in the platform that serves the web application and other resources.

Log File	Can Help With
nginx_service.log	<ul style="list-style-type: none">• This log may be useful in identifying any warnings that occurred when the nginx services starts.• The <code>nginx</code> server may contain log information for the server that provides HTTP access.

Proxy

Description: The proxy (nginx) service manages requests from the user interface to the other components of the platform.

Log File	Can Help With
proxy_access.log	<ul style="list-style-type: none">• Shows any requests made through the nginx to the port used by the Trifacta platform.
proxy_error.log	<ul style="list-style-type: none">• Contains any errors thrown by the nginx service when a request is made to the port used by the Trifacta platform.

Scheduling Service

Description: Handles all metadata related to scheduling

Log File	Can Help With
scheduling-service.log	<ul style="list-style-type: none">• Schedule-related issues
scheduling-service.access.log	<ul style="list-style-type: none">• Gives information about accessed routes

Spark Job Service

Description: Service that manages jobs processed on Spark.

Log File	Can Help With
----------	---------------

spark-job-service.log	<ul style="list-style-type: none"> • Status of the service • Debugging issues with Spark jobs • See <i>Configure for Spark</i>.
-----------------------	--

Supervisord Server

Description: Process that starts, stops, and restarts services in the platform.

Log File	Can Help With
supervisord.log	<ul style="list-style-type: none"> • Status information from platform services

Time-Based Trigger Service

Description: Handles all metadata related to the trigger service

Log File	Can Help With
time-based-trigger-service.log	<ul style="list-style-type: none"> • schedules not triggering correctly
time-based-trigger-service.access.log	<ul style="list-style-type: none"> • Gives information about accessed routes

VFS Service

Description: Loads data from the various filesystems supported by the platform, both in the front-end user interface and in batch mode when Photon running environment is enabled. For more information, see *Running Environment Options*.

Log File	Can Help With
vfs-service.log	<ul style="list-style-type: none"> • Client connection issues • Status issues with backend components • Information on batch jobs that cannot be started

Webapp Service

Description: Loads data from the various filesystems supported by the platform in the front-end user interface.

Log File	Can Help With
webapp.log	<ul style="list-style-type: none"> • Client connection issues • Status issues with backend components • Information on batch jobs that cannot be started

Additional logs

Job logs

The following sources of information may provide information related to job status and performance:

- job log
- spark log
- cdf script
- yarn application logs
(if log aggregation is enabled)
- platform configuration file
(trifacta-conf.json)
- batch job runner log

- spark service log
- hadoop conf directory (if accessible)

Maintenance Release Updater

Contents:

- *Support Matrix*
- *Warnings*
- *Pre-requisites*
- *Commands*

Periodically, Trifacta may publish updates to the software release that you have currently installed. The Trifacta® Maintenance Release Updater can be run from the Trifacta node to check for updates for the release of the platform that is currently installed on the node. If a maintenance release (e.g. Release x.y.1) is available, the script downloads the package and launches the update locally.

Support Matrix

This script can only be used for maintenance patches for the current release. You cannot upgrade from Release 3.2 to Release 4.0 or any later release using this script.

NOTE: This script is not available in the GA release of any major release, since no maintenance release is yet available.

Below are some example upgrade paths.

Example Upgrade	Supported?
Release 3.2.0 Release 3.2.1	Supported
Release 3.2.1 Release 3.2.2	Supported (target release does not exist)
Release 3.2.1 Release 4.0.0	Not Supported
Release 4.0.0 Release 4.1	Not Supported
Release 4.0.0 Hot Fixes	Not Supported

Warnings

- The Trifacta platform must not be running when the Updater is executed.
- Do not run this script using a scheduler.

NOTE: When the script is executed and new software is downloaded, the script automatically begins the update process, which can disrupt user operations and user data and may not work if the platform is running.

Pre-requisites

NOTE: You must acquire the script from the Trifacta download site. Please use the credentials provided to you to connect to the download site.

NOTE: Before you begin, shut down the Trifacta platform and perform backups of the deployment directory and Trifacta databases.

- This script is supported by Centos/RHEL 6 or later. Ubuntu is not supported.
- This script must be run as the `root` user on the node.
- The Trifacta platform must be installed on the node and configured to run using the `[os.user (default=trifacta)]` user.
- To enable downloads of updates, CURL must be installed on the node.

Commands

Script Name:

```
trifacta-maintenance-release-upgrade.sh
```

Location:

Prior to execution, the script should be placed somewhere other than the Trifacta install directory.

- `/root` or `/tmp` are acceptable locations.
- The script must have execute permissions, if it doesn't already have them.

For each run of the script:

- It creates a unique working directory under `/opt/trifacta-upgrade-workdir`.
- The working directory is where downloaded artifacts, backup artifacts, and log files are stored during execution.

Download and Install:

The following command checks for an update. If one is detected, it downloads the update and launches the updating process:

```
./trifacta-maintenance-release-upgrade.sh
```

NOTE: The credentials asked by the script are only used to connect to the FTP site to check for and download installer artifacts. These credentials are not stored anywhere.

The package is downloaded by default to the following location:

```
/opt/trifacta-upgrade-workdir/<UNIQUE-FOLDER-PER-RUN>/<trifacta-server-installer-rpm-file>
```

Install Downloaded Package:

If you have separately downloaded the update from Trifacta, the following performs the update from the local package, which is specified by parameter:

```
./trifacta-maintenance-release-upgrade.sh -i <path_to_downloaded_RPM>
```


After you have completed the update:

1. Review any update documentation provided to you.
2. Restart the platform. See *Start and Stop the Platform*.
3. Verify operations. See *Verify Operations*.

More Help:

```
./trifacta-maintenance-release-upgrade.sh -h
```

Backup and Recovery

Contents:

- *Backup Platform Files*
 - *Configuration*
 - *License*
- *Backup Databases*
 - *Location of backup and recovery tools*
 - *Backup commands*
 - *Scheduling*
- *Recovery*
 - *Verify*
 - *Rollback steps*

This section provides overview information on the key data and metadata that should be managed by your enterprise backup and recovery policies.

All backups should be performed in accordance with your enterprise's backup and recovery policies.

Backup Platform Files

The following directories on the Trifacta node should be backed up on a regular basis:

Configuration

You can back up all key configuration files into the `/tmp` directory using the following commands:

```
cp /opt/trifacta/conf/trifacta-conf.json /tmp/trifacta-conf.json
cp /opt/trifacta/conf/env.sh /tmp/env.sh
cp /etc/init.d/trifacta /tmp/trifacta.service
```

License

You should backup your license key:

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

See *License Key*.

Backup Databases

The Trifacta platform utilizes the following databases as part of normal operations. These databases should be backed up on a regular basis:

Database Name	Databaseld	Description
Trifacta DB	trifacta	Stores users and metadata for flows, including datasets, and recipes.
Jobs DB	trifacta-activiti	Stores and maintains job execution status and details.
Scheduling DB	trifactaschedulingservice	Stores metadata for scheduled jobs.
Time-based Trigger DB	trifactatimebasedtriggerservice	Additional database required for scheduled jobs.

For more information on setting up these databases, see *Set up the Databases*.

Location of backup and recovery tools

PostgreSQL

Depending on your operating system, you can find the backup tool `pg_dump` in the following location.

NOTE: These locations apply to PostgreSQL 9.6.

CentOS/RHEL:

```
/usr/pgsql-9.6/bin/pg_dump  
/usr/pgsql-9.6/bin/psql
```

Ubuntu:

```
/usr/lib/postgresql/9.6/bin/pg_dump  
/usr/lib/postgresql/9.6/bin/psql
```

MySQL

Please locate the following programs in your MySQL distribution:

```
mysqldump  
mysql
```

Backup commands

The following commands can be used to back up the databases.

PostgreSQL

For more information on command options, see <https://www.postgresql.org/docs/9.6/static/backup.html>.

NOTE: These commands must be executed as the `trifacta` user.

NOTE: The following commands are for PostgreSQL 9.6 for all supported operating systems. For specific commands for other versions, please see the database documentation.

Trifacta DB:

```
pg_dump trifacta > trif_triDB_bkp_<date>.sql
```

Jobs DB:

```
pg_dump trifacta-activiti > trif_actDB_bkp_<date>.sql
```

Scheduling DB:

```
pg_dump trifactaschedulingservice > trif_schDB_bkup_<date>.sql
```

Time-Based Trigger DB:

```
pg_dump trifactatimebasedtriggerservice > trif_tbtsDB_bkup_<date>.sql
```

MySQL

For more information on command options, see <https://dev.mysql.com/doc/refman/5.7/en/mysqldump-sql-format.html>.

```
su - mysql
```

NOTE: The following commands are for MySQL 5.7 for all supported operating systems. For specific commands for other versions, please see the database documentation.

Trifacta DB:

```
mysqldump trifacta > trif_triDB_bkp_<date>.sql
```

Jobs DB:

```
mysqldump trifacta-activiti > trif_actDB_bkp_<date>.sql
```

Scheduling DB:

```
mysqldump trifactaschedulingservice > trif_schDB_bkup_<date>.sql
```

Time-Based Trigger DB:

```
mysqldump trifactatimebasedtriggerservice > trif_tbtsDB_bkup_<date>.sql
```

Scheduling

You can schedule nightly execution of these backups using a third-party scheduler such as cron.

Recovery

Verify

Before you begin, please verify that you have valid backups for the following data from the version to which you are rolling back:

NOTE: When the databases are restored, internal identifiers such as job IDs, are reset in an order that may not correspond to the expected order. Consequently, references to specific identifiers may be corrupted. After restoring the databases, you should clear the job logs.

NOTE: If you do not have any of these items, you may not be able to recover your instance of the Trifacta platform to its previous state.

Item	Description
Configuration files	Backup of your configuration files
Databases	Backup of your databases
RPM	RPM installer

Rollback steps

To recover the Trifacta platform based on backups:

NOTE: If any of the hosts, pathnames, or credentials have changed since the backups were performed, these updates must be applied through `trifacta-conf.json` or through the Admin Settings page after the restoration is complete.

Steps:

1. Login to the Trifacta node as root user.
2. Stop the Trifacta service:

```
service trifacta stop
```

3. Clear each current database and restore the backup of the version from the preceding release. In some cases, the database may not exist in the previous version.

a. PostgreSQL:

- i. Login as a user that can run admin commands for PostgreSQL. This user may vary between deployments.
- ii. Trifacta database:

```
psql -c "DROP DATABASE trifacta;"
psql -c "CREATE DATABASE trifacta WITH OWNER trifacta;"
psql --dbname=trifacta < trifacta_backup_<date>.sql
```

- iii. (Release 3.2 and later) Jobs database:

NOTE: Please note the escaped quotes around the database name in the CREATE DATABASE command.

```
psql -c "DROP DATABASE \"trifacta-activiti\";"
psql -c "CREATE DATABASE \"trifacta-activiti\" WITH OWNER
trifacta;"
psql --dbname=trifacta-activiti <
trif_actDB_backup_<date>.sql
```

- iv. (Release 4.1 and later) Scheduling database:

```
psql -c "DROP DATABASE trifactaschedulingsservice;"
psql -c "CREATE DATABASE trifactaschedulingsservice WITH
OWNER trifacta;"
psql --dbname=trifactaschedulingsservice <
trif_schedDB_backup_<date>.sql
```

- v. (Release 4.1 and later) Time-based Trigger Service database:

```
psql -c "DROP DATABASE trifactatimebasedtriggerservice;"
psql -c "CREATE DATABASE trifactatimebasedtriggerservice
WITH OWNER trifacta;"
psql --dbname=trifactatimebasedtriggerservice <
tri_tbttsDB_backup_<date>.sql
```

- b. **MySQL:** For details, see <https://dev.mysql.com/doc/refman/5.7/en/reloading-sql-format-dumps.html>.
- i. Login:

```
su - mysql
```

- ii. Trifacta database:

```
mysql trifacta < trifacta_backup_<date>.sql
```

- iii. Jobs database:

```
mysql trifacta-activiti < trif_actDB_backup_<date>.sql
```

- iv. (Release 4.1 and later) Scheduling database:

```
mysql trifactaschedulingservice <
trif_schedDB_backup_<date>.sql
```

- v. (Release 4.1 and later) Time-based Trigger Service database:

```
mysql trifactatimebasedtriggerservice <
tri_tbttsDB_backup_<date>.sql
```

4. Perform a clean install of the Trifacta software provided in your distribution. See *Install*.
5. Restore your configuration files. The following commands assume that they were backed up to the /tmp directory on the node:

```
cp /tmp/trifacta-conf.json /opt/trifacta/conf/trifacta-conf.json
cp /tmp/env.sh /opt/trifacta/conf/env.sh
cp /tmp/trifacta.service /etc/init.d/trifacta
```

6. Apply any patches or maintenance updates that may have been provided to you. See *Maintenance Release Updater*.
7. Restart the platform. See *Start and Stop the Platform*.
8. Login and verify operations. See *Verify Operations*.

Admin Reference

The following pages in Trifacta® Wrangler Enterprise apply to admin users.

Topics:

- *Admin Settings Page*
- *Deployments Page*

Admin Settings Page

Contents:

- *Platform Settings*
 - *External Service Settings*
 - *AWS EMR Cluster ID*
 - *Users*
 - *Services*
 - *Tricheck*
 - *Upload License*
 - *Restart*
-

Admin users of the Trifacta® platform can change settings through the Trifacta application. Login as an admin user, and click the Gear icon. Select **Admin Settings**.

NOTE: You must be an administrator to access this feature.

Platform Settings

Do not modify settings through the Admin Settings page and through `trifacta-conf.json` at the same time. Saving changes in one interface wipes out any unsaved changes in the other interface. Each requires a platform restart to apply the changes.

Platform administrators can change any parameter value that is available through the web application. Enter some or all of parameter name into the search box to see a set of possible matches.

Do not modify parameters with which you are unfamiliar or have not been instructed to change. Some changes can be harmful to the system. In particular, changing the database connection parameters can break access to the application and the Admin Settings page.

Search:

Tip: You can copy setting names from the documentation to search the available set. Search retrieves matches from the setting name and the help text for the parameter. Do not paste in double quotes from documentation samples.

If your search for a parameter comes up empty and you know that the parameter exists, you must make changes on the Trifacta node in `trifacta-conf.json`. See *Required Platform Configuration*.

Search groupings:

If you search for the following strings, which may appear in property descriptions, you can review groups of settings pertaining to the configuration areas listed below.

NOTE: Do not perform configuration of these areas by simply reviewing and modifying the settings in these parameter groups. Additional configuration may be required. Some required settings may not be grouped, and some of these settings may not be documented. Please review the related documentation sections.

Search string	Setting group
[CORE]	Core platform settings.
[DISTRO]	Settings pertaining to specific Hadoop distributions. <ul style="list-style-type: none">• See <i>Configure for Cloudera</i>.• See <i>Configure for Hortonworks</i>.
[CLUSTER]	Settings that affect how the platform interacts with the integrated backend cluster. See <i>Prepare Hadoop for Integration with the Platform</i> .
[HIVE]	Settings pertaining to integration with the Hadoop cluster's instance of Hive. See <i>Configure for Hive</i> .
[HA]	Settings pertaining to integration with cluster high availability for the Trifacta platform. See <i>Enable Integration with Cluster High Availability</i> .
[SECURITY]	General settings pertaining to security. See <i>Configure Security</i> .
[SSL]	Settings pertaining to SSL access to the platform. See the SSL section in <i>Configure Security</i> .
[ADVANCED]	Advanced settings.

When you modify a setting, your change is validated against the data type or set of accepted values. String-based entries cannot be validated.

Notes:

- Sensitive information is obfuscated in the display values in the Admin Settings page.
- To save changes, click **Save**.

NOTE: Saving changes forces an automatic type validation of the configuration and a restart of the platform, which terminates any active user sessions.

NOTE: Environmental validation is not performed as part of changes in this user interface. For example, you can change the port number for the Trifacta application to an invalid value and save the configuration change. However, when the platform is restarted, the application fails to start, and you cannot continue. In this case, you must fix the problem in `trifacta-conf.json`.

External Service Settings

AWS EMR Cluster ID

If you have deployed your instance of the Trifacta platform on to Amazon Web Services (AWS) and are connected to an Elastic Map Reduce (EMR) cluster, you can review or modify the cluster identifier in this location. For example, in the event of prolonged outage or failure of the original cluster, you can insert the cluster ID of a secondary cluster to effectively failover to the new cluster.

NOTE: When you first install and integrate with an EMR cluster, this identifier is stored in the Trifacta database for you. It should be modified only if you need to switch to a different EMR cluster. Only one EMR cluster can be active at any time.

NOTE: If this cluster ID is modified, you must modify any other EMR-related settings to match the corresponding values for the new cluster. Please search for `emr` among the admin settings.

When you have entered a new cluster ID, click **Save**.

NOTE: For this setting, you do not have to click the Save button at the bottom of the screen, which restarts the Trifacta platform.

Users

You can manage aspects of user accounts through the Admin Settings page. See *Manage Users*.

Services

You can review overall status of the Trifacta platform.

Tricheck

Tricheck performs a variety of tests of your environment to determine its suitability for use with the Trifacta platform.

Tip: Tricheck should be run immediately after the Trifacta software has been installed or upgraded or whenever there are significant changes to the node or its connected cluster.

Checks include but are not limited to:

- Sufficient hardware resources on the Trifacta node
- Supported versions of software installed on the Trifacta node
- Access to required ports and all nodes of the cluster
- Trifacta node system profiling

NOTE: Tricheck performs no data-dependent checking. It cannot assess suitability of the environment for specific data volumes, connections, or data types.

Click **Run Tricheck** to run checks and download the output log.

Upload License

NOTE: For more information on acquiring an updated license file, please contact *Trifacta Support*.

You can update the license file stored on the Trifacta node. Click **Upload License** to browse for and select the license file.

For more information on your license, see *License Key*.

Restart

Click **Restart Trifacta** to immediately restart the platform.

Tip: The Restart Trifacta button is the preferred method for restarting the platform. A restart is automatically executed when you save changes to the platform settings.

NOTE: This button may not be available in high availability environments. In those deployments, please restart individual services or use the command line command. For more information, see *Start and Stop the Platform*.

Deployments Page

Contents:

- *Access*
- *Deployment Hierarchy*
- *Deployments View*
- *Releases View*
- *Flows View*

Through the Deployments page, you interact with flows that you have imported into your Production instance. Through this interface, you can activate Production versions of your flows or rollback to previous versions as needed.

NOTE: The Deployments page is available only in a Production environment, which is a special instance of the Trifacta platform designed to support production use of your flows. For more information, see *Overview of Deployment Management*.

Access

A Production environment can be accessed in either of the following ways:

- You are given access to a separate instance of the Trifacta platform configured for Production use only.
- On any instance, the Deployment role is added to your user account by a Trifacta administrator.

For more information, see *Configure Deployment Management*.

Deployment Hierarchy

In a Production environment, a **deployment** is version-managed flow and all of its dependencies, including other dependent flows. Through the Deployments page, an individual deployment is structured in the following hierarchy:

Hierarchy Level	Object	Description
1	deployment	When you open the Deployments page, you can review all of the deployments that have been created in the environment. A deployment is container for releases.

2	release	<p>When you select a deployment, you can explore its releases. A release is an individual instance of an imported flow and its dependencies (an import package). Each time that the import package is re-imported into the Production instance, a new release is created and made the active release for the deployment.</p> <p>You can activate previous releases as needed through the context menus in Deployments page.</p>
3	flow or flows	<p>Within a release, you can explore the flows that were included in the import package for the release:</p> <ul style="list-style-type: none"> • The primary flow is the one that is executed when: <ul style="list-style-type: none"> • Its release is the active one for the deployment • The job for the deployment is executed • Any secondary flows are the flows on which the primary flow depends for data. <ul style="list-style-type: none"> • During export from the source instance, all objects in secondary flows are included in the package. There may be objects in a secondary flow that are unused in the Production instance.

Deployments View

When you open the Deployments page, you can explore all of the deployments in the Production instance.

NAME	RELEASES	LAST UPDATED BY	LAST UPDATED AT
2013 POS	1 Releases	SteveO	2018-03-28 13:16:53

Figure: Deployments View

To create a new deployment, click **Create**.

1. Enter a name for the deployment, and click **Create**.
2. To create a new release, click the created deployment. See Releases View below.

Actions:

- **Search:** Enter values in the search textbox to search deployment names. Matching occurs in real-time.
- **Edit name:** You can change the name of your deployment as needed.
- **Delete:** Select this option to remove the deployment, all of its releases, and all of the flows within each release.

You cannot undo deleting a deployment. Any results generated from jobs run for the deployment are not removed from the output location and are still accessible through the Jobs page of the Production instance.

Releases View

Through Releases view, you can import new packages to create new releases and activate them, roll back to previous releases, and remove releases that are no longer in use.

RELEASE	PACKAGE ID	NOTES	CREATED AT	UPDATED AT	
1	dc5aff10		2018-03-28 13:16:53	2018-03-28 13:16:53	Active

Figure: Releases View

To create a new release, click **Import Package**.

NOTE: Before you import a package, you must apply any import mapping rules to the deployment. These rules map values and objects in the package to corresponding values in the new instance. For more information, see *Define Import Mapping Rules*.

1. Navigate your local environment to select the ZIP file containing the flow and its dependencies from the source instance.
2. Click **Import**.
3. The release is added to Releases view.

For more information on importing, see *Import Flow*.

Actions:

Action	Description
Search	Enter values in the search textbox to search package identifiers. Matching occurs in real-time.
Activate	Make the selected release the active one for the deployment. When jobs are executed at the deployment level, the primary flow for the active release is executed.
Export	Export the release for use in another instance of the platform. See <i>Export Flow</i> .
Delete	<p>Delete the release from the Production instance.</p> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> <p>NOTE: If the release was imported from a Dev instance on the same platform, the Dev instance of the release is not removed.</p> </div> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> <p>NOTE: Deletion of a release does not remove any results generated from it. Those results are still accessible through the Jobs page.</p> </div>

Flows View

When you open a release, you can review the flows contained in it. You can explore the flow or flows that were included in the import package for the selected release.

NAME	DATASETS	UPDATED AT
 2013 POS	6 Datasets	Today at 1:16 PM

Figure: Flows View

Actions:

- **Search:** Enter values in the search textbox to search flow names. Matching occurs in real-time.
- Click a flow name to explore the flow in Flow View.

NOTE: Avoid making changes to a flow in a Production instance. You can run ad-hoc jobs, but you should avoid making changes to the objects or their structure through Flow View in Production instances. Scheduling should be done through the command line.

For more information, see *Flow View Page*.



Copyright © 2018 - Trifacta, Inc.
All rights reserved.