TRIFACTA

Admin Guide

Version: 6.4.2
Doc Build Date: 02/25/2020

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

Admin

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

Admin Tasks

The topics below provide information on how to manage aspects of your account in Trifacta® Wrangler Enterprise.

Verify Operations

Contents:

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. Your dataset should be created in the format that you wish to test.

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.

Store Your Dataset

If you are testing an integration, you should store your dataset in the datastore with which the product is integrated.

> **Tip:** Uploading datasets is always available as a means of importing datasets.

- You may need to create a connection between the platform and the datastore.
- Read and write permissions must be enabled for the connecting user to the datastore.
- For more information, see *Connections Page*.

Verification Steps

Steps:

1. Login to the application. See *Login*.
2. In the application menu bar, click **Library**.
3. Click **Import Data**. See *Import Data Page*.
    1. Select the connection where the dataset is stored. For datasets stored on your local desktop, click **U pload**.
    2. Select the dataset.
    3. In the right panel, click the Add Dataset to a Flow checkbox. Enter a name for the new flow.
    4. Click **Import and Add to Flow**.
    5. **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.
4. In the left menu bar, click the Flows icon. Flows page, open the flow you just created. See *Flows Page*.
5. In the Flows page, click the dataset you just imported. Click **Add new Recipe**.
6. Select the recipe. Click **Edit Recipe**.
7. The initial sample of the dataset is opened in the Transformer page, where you can edit your recipe to transform the dataset.
    1. In the Transformer page, some steps are automatically added to the recipe for you. So, you can run the job immediately.
    2. You can add additional steps if desired. See *Transformer Page*.
8. Click **Run Job**.
    1.
    2. If options are presented, select the defaults.
    3. To generate results in other formats or output locations, click **Add Publishing Destination.** Configur e the output formats and locations.
    4. To test dataset profiling, click the Profile Results checkbox. Note that profiling runs as a separate job and may take considerably longer.
    5. See *Run Job Page*.
    6. **Troubleshooting:** Later, you can re-run this job on a different running environment. Some formats are not available across all running environments.
9. When the job completes, you should see a success message under the Jobs tab in the Flow View page.
    1. **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 a different running environment (if available). You can also download the log files to try to identify the problem. See *Job Details Page*.
10. Click **View Results** from the context menu for the job listing. In the Job Details page, you can see a visual profile of the generated results. See *Job Details Page*.
11. In the Output Destinations tab, click a link to download the results to your local desktop.
12. Load these results into a local application to verify that the content looks ok.

> **Checkpoint:** You have verified importing from the selected datastore and transforming a dataset. If your job was successfully executed, you have verified that the product is connected to the job running environment and can write results to the defined output location. Optionally, you may have tested profiling of job results. If all of the above tasks completed, the product is operational end-to-end.

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:

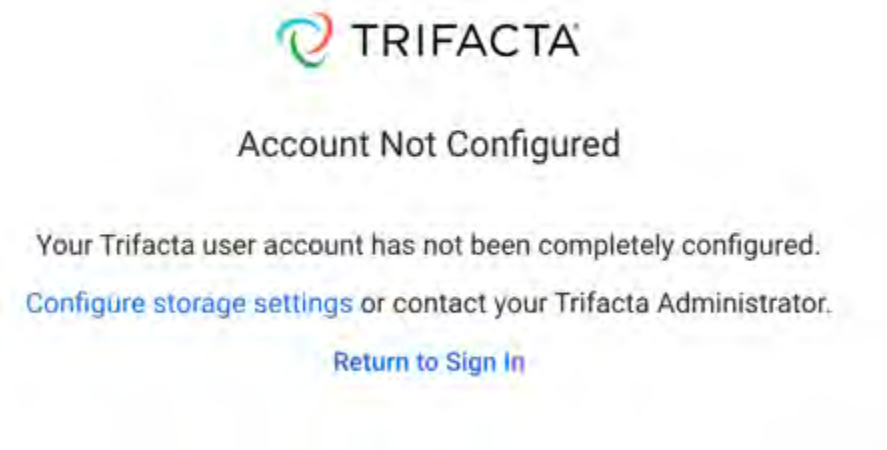

Figure: Account Not Configured

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:

You can create additional administrator accounts to the Trifacta® platform 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 `webapp/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 enviroment 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.

Manage Users

Contents:

Through the Admin Settings page, administrators can manage aspects of user accounts, as well as other aspects of the instance. 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 instance, 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.

- **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 Password**. 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.

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.

> **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 the Deployment Manager.
    - In a Production environment where the Deployment Manager applies to the entire instance, this role does not apply.
    - For more information, see *Configure Deployment Manager*.
    - For more information on Deployment Manager, see *Overview of Deployment Manager*.
- **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.

AWS Config

When per-user authentication is enabled for AWS access, administrators can review and modify each user's settings for AWS authentication, click **Configure**.

> **NOTE:** When you return from configuring S3 access, your changes there have already been saved.

For more information, see *Configure Your Access to S3*.

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**.

Manage Users under SSO

Contents:

This section covers additional requirements for managing users in SSO environments.

Enable SSO

The Trifacta® platform requires additional configuration to integrate with your SSO provider. Available methods:

Method	Description
SAML IDP	Integrate the platform with enterprise SAML identity provider. See *Configure SSO for SAML*.
Native LDAP-AD	Using native functionality in the platform, it can integrate with enterprise LDAP/AD. For more information, see *Configure SSO for AD-LDAP*.
LDAP-AD via reverse proxy	A reverse proxy server outside of the platform can be set up for integration with enterprise LDAP/AD. **NOTE:** This method is likely to be deprecated in a future release. For more information, see *Configure SSO for AD-LDAP*.

Configure Auto-Registration

Tip: By default, user auto-registration is enabled. It is recommended.

How users are managed depends on whether auto-registration is enabled:

- If auto-registration is enabled, after users provide their credentials, the account is automatically created for them.
- If auto-registration is disabled, a Trifacta administrator must still provision a user account before it is available. See below.

Manage Users with Auto-Registration

After SSO with auto-registration has been enabled, you can still manage users through the Admin Settings page, with the following provisions:

- The Trifacta platform does not recheck for attribute values on each login. If attribute values change with your identity provider, they must be updated in the configuration.
  - For more information, see *Configure SSO for AD-LDAP*
  - For more information, see *Configure SSO for SAML*.
- If the user has been removed from AD, the user cannot sign in to the platform.
- If you need to remove a user from the platform, you should just disable the user through the User Management area.
  - If the user is deleted, then if the user returns to the platform in the future, a new account is created for the user.

For more information, see *Manage Users*.

Disable Auto-Registration

To disable auto-provisioning in the platform, please verify the following property:

1. You can apply this change through the *Admin Settings Page* (recommended) or `trifacta-conf.json`. For more information, see *Platform Configuration Methods*.
2. Set the following property:

```
"webapp.sso.enableAutoRegistration" : false,
```

3. Save your changes and restart the platform.
4. New users of the Trifacta platform must be provisioned by a Trifacta administrator. See below.

Provision new users under SSO without auto-registration

If SSO auto-registration is disabled, admin users can provision new users of the platform through the following URL:

```
https://<hostname>:<sso_port_number>/register
```

where:

- `<hostname>` is the host of the Trifacta platform
- `<sso_port_number>` is the port number.

The user's password is unnecessary in an SSO environment. You must provide the SSO principal value, which is typically the Active Directory login for the user.

- If you are connected to a Hadoop cluster, you must provision the Hadoop principal value.
- See *Create User Account*.

User access for reverse proxy method

Users access the application through the Trifacta node using the standard hostname and the port that you specified:

> **NOTE:** All users must be use this URL to access the Trifacta application. If they use the non-SSO URL, they may receive an `Unprovisioned User` error.

```
https://<hostname>:<sso_port_number>
```

Manage Schedules

Contents:

Through the Schedules page, administrators of Trifacta® Wrangler Enterprise can manage all of the schedules in the deployment.

> **NOTE:** The Schedules page is available to administrators only.

Enable or Disable Schedule

Steps:

1. Login as an administrator.
2. In the left nav bar, click the Schedules icon.
3. In the Schedules page, locate the schedule to change.
4. For the schedule's entry, open the context menu on the right side of the page.
5. Select **Enable Schedule** or **Disable Schedule**.

Delete Schedule

> **Deleting a schedule cannot be undone.**

Steps:

1. Login as an administrator.
2. In the left nav bar, click the Schedules icon.
3. In the Schedules page, locate the schedule to delete.
4. For the schedule's entry, open the context menu on the right side of the page.
5. Select **Enable Schedule** or **Disable Schedule**.

Create Schedule

A schedule is composed of:

- A schedule frequency
- A set of one or more scheduled outputs

These objects are created from within a flow. For more information, see *Flow View Page*.

- See *Add Schedule Dialog*.

Configure Password Criteria

By default, the Trifacta® application enforces very few requirements on password length, capitalization, or special characters. Users who are setting or resetting their passwords are permitted to create a password of one character in length with no additional requirements.

> **NOTE:** When passwords are set or reset, the platform does perform an assessment of the quality of the password and reports it to the user before saving. For more information, see *User Profile Page*.

> **Before you permit users to create accounts, you should review the password requirements for your enterprise and, where needed, apply them to the Trifacta application.**

Enable

To enable enforcement of password criteria, please enable the following parameter.

Steps:

You can apply this change through the *Admin Settings Page* (recommended) or `trifacta-conf.json`. For more information, see *Platform Configuration Methods*.

Locate the following parameter and set it to `true`:

```
"feature.enablePasswordCriteria": true,
```

When enabled, submitted changes to user passwords are evaluated based on the configuration settings defined below.

Configure

The following parameters govern the password criteria enforced by the Trifacta application when the feature is enabled.

Parameter	Description	Default
webapp.passwordCriteria.length.min	Minimum length of a password to Trifacta application	0
webapp.passwordCriteria.length.max	Maximum length of a password to Trifacta application	100
webapp.passwordCriteria.description	Text describing the criteria that a password must meet. Specify this value last.	
webapp.passwordCriteria.contains.uppercase	Defines whether the password must contain uppercase characters	undefined
webapp.passwordCriteria.contains.symbols	Defines whether the password must contain symbols	undefined
webapp.passwordCriteria.contains.spaces	Defines whether the password must contain space characters	undefined
webapp.passwordCriteria.contains.lowercase	Defines whether the password must contain lowercase characters	undefined
webapp.passwordCriteria.contains.letters	Defines whether the password must contain letters (a-z)	undefined
webapp.passwordCriteria.contains.digits	Defines whether the password must contain digits (0-9)	undefined

Criteria settings:

Some of the criteria settings support the following options:

Setting	Description
enforce	Each password must pass this requirement.
forbid	Passwords cannot have this requirement.
undefined	(default) This requirement is disabled. Users may choose to include or not include this requirement in their passwords.

System Services and Logs

Contents:

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`

Trifacta® administrators can access the logs through the Trifacta application. Use the following URL:

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

Logs available for download

Filename	Description
`jobgroups/`	Directory of logs for transformation jobs by Id. **Tip:** If you are troubleshooting a failed job, please acquire the job logs from the Job Details page when you contact *Trifacta Support*. See *Job Details Page*.
`jobs/`	Directory of logs for other kinds of jobs, such as sampling or ingest, by Id.
`nginx/`	Temporary storage for nginx server. No log files are stored here.
`artifact-storage-service.access.log`	Access logs for the artifact storage service.
`artifact-storage-service.log`	Application logs for the artifact storage service.

batch-job-runner.access.log	Access logs for the batch job runner service. Batch job runner service manages transformation jobs and scheduling. More information is below.
batch-job-runner.job-status.log	Status information on batch job runner jobs.
batch-job-runner.log	Application logs for the batch job runner service.
configuration-service.access.log	Access logs for the configuration service service. Configuration service is used for managing configuration that can be changed at runtime for different workspaces and users. Some of these settings are available through the Trifacta application. See *Workspace Admin Page*.
configuration-service.log	Application logs for the configuration service.
conversion-service.access.log	Access logs for the conversion service. Conversion service is used for converting from various inputs formats and to various output formats.
conversion-service.log	Application logs for the conversion service.
data-service.access.log	Access logs for the data service. Data service is used for interacting with relational sources. More information is below.
data-service.log	Application logs for the data service.
job-metadata-service.access.log	Access logs for the job metadata service.
job-metadata-service.log	Application logs for the job metadata service.
migration-report.log	Log containing information about specific migrations between releases.
migration.log	Application logs for database migrations performed for the webapp service.
ml_service.access.log	Access logs for the ml (machine-learning) service. Machine learning service is used for predictive interaction, suggestion ranking, pattern profiling, pattern suggestions, and collecting user action logs. More information is below.
ml_service.log	Application logs for the ml (machine-learning) service.
nginx_service.log	Application logs for the nginx service. More information is below.
protobuf-events.log	Client events around column values, user selections, and script editing.
proxy_access.log	Access logs for the nginx server.
proxy_error.log	Error logs for the nginx server.
scheduling-service.access.log	Access logs for the scheduling service. Scheduling service is used for scheduling jobs at a specific time. More information is below.
scheduling-service.log	Application logs for the scheduling service.
secure-token-service.access.log	Access logs for the secure token service. secure-token-service is used for securely storing tokens for some external services, such as Azure AD and Databricks.
secure-token-service.log	Application logs for the secure token service

`spark-job-service.log`	Application logs for the Spark job service. Spark job service is used for interfacing with cluster-based Spark service to plan and execute Spark jobs. More information is below.
`supervisord.log`	Logs for the supervisord system service. supervisord manages the starting, stopping, and restarting of services for the Trifacta platform. More information is below.
`time-based-trigger-service.access.log`	Access logs for the time-based trigger service. Time-based trigger service is used for managing the triggers for scheduled jobs. More information is below.
`time-based-trigger-service.log`	Application logs for the time-based trigger service.
`vfs-service.access.log`	Access logs for the VFS service. VFS service is used for managing loading of files from various supported datastores. More information is below.
`vfs-service.log`	Application logs for the VFS service.
`webapp.access.log`	Access logs for the Webapp service. Webapp service serves the Trifacta application to users. More information is below.
`webapp.log`	Application logs for the Webapp service. More information is below.
`webapp.sql-error.log`	Error log for SQL issued from the Webapp service.
`webworker.log`	Event logs for webworkers running in browser clients. Webworkers run in the background of a browser client's Trifacta session and are used for predictive interaction, suggestion ranking, pattern profiling, pattern suggestions, and collecting user action logs.

Batch Job Runner

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

Log File	Can Help With

`batch-job-runner.log`	• Service errors and crashes
	• Determine execution environment of the job. Search for:
	  • `LocalJobRunner` = local execution in Photon
	  • `YARNRunner` = execution in Spark
	• 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
`data-service.log`	• 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`	• Status of the service **Tip:** You can pass through messages on errors through Logger to this log, which can assist in diagnosing issues. See *Java UDFs*.

Machine Learning Service

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

Log File	Can Help With
`ml_service.log`	• This log is likely to contain information that is only useful if the ML service has crashed.

Nginx Service

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

Log File	Can Help With
`nginx_service.log`	• This log may be useful in identifying any warnings that occurred when the nginx services starts. • The `nginx` 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`	• Shows any requests made through the nginx to the port used by the Trifacta platform.
`proxy_error.log`	• 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`	• Schedule-related issues
`scheduling-service.access.log`	• 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`	• Status of the service • Debugging issues with Spark jobs • See *Configure for Spark*.

Supervisord Server

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

Log File	Can Help With
`supervisord.log`	• 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`	• schedules not triggering correctly
`time-based-trigger-service.access.log`	• 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 the Trifacta Photon running environment is enabled. For more information, see *Running Environment Options*.

Log File	Can Help With
`vfs-service.log`	• 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`	• 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:

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. For example, you cannot upgrade from Release 5.1 to Release 6.0 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.

> **NOTE:** These are example upgrade paths. Some of these releases may not exist yet.

Example Upgrade	Supported?
Release 6.0.1 Release 6.0.2	Supported
Release 6.1.1 Release 6.1.2	Supported
Release 6.1.1 Release 7.0.0	Not Supported
Release 6.0.0 Release 6.2	Not Supported
Release 6.0.1 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>
```

Do not restart the platform at this time, in case there are necessary migrations for this update. See below.

Migrate Configurations

Even between maintenance releases, it's possible for some configuration settings to be updated or migrated into the Workspace Admin page.

NOTE: For every upgrade or update, you should perform the steps to migrate any configurations that have changed. If there are no configurations to migrate, the process is harmless.

For more information, see *Migrate Configurations*.

Verify Update

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:

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

NOTE: This section covers how to perform a basic cold backup of the product. Hot backups are not supported.

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

Stop All Services

Before you begin, the Trifacta platform and databases should be stopped. See *Start and Stop the Platform*.

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	DatabaseId	Description
Main 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.
Configuration Service DB	`trifactaconfigurationservice`	Stores configuration settings for the workspace.
Artifact Storage Service DB	`trifactaartifactstorageservice`	Stores feature usage data such value mappings for the standardization feature.
Job Metadata Service DB	`trifactajobmetadataservice`	Stores metadata on job execution.

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

Location of backup and recovery tools

PostgreSQL

Depending on your operating system, you can find the backup tools 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
```

Configuration Service DB:

```
pg_dump trifactaconfigurationservice > trif_confservDB_bkup_<date>.sql
```

Artifact Storage DB:

```
pg_dump trifactaartifactstorageservice >
trif_artifactstorageservDB_bkup_<date>.sql
```

Job Metadata Service DB:

```
pg_dump trifactajobmetadataservice > trif_jobmetadataservDB_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
```

Configuration Service DB:

```
mysqldump trifactaconfigurationservice > trif_confservDB_bkup_<date>.sql
```

Artifact Storage DB:

```
mysqldump trifactaartifactstorageservice >
trif_artifactstorageservDB_bkup_<date>.sql
```

Job Metadata Service DB:

```
mysqldump trifactajobmetadataservice >
trif_jobmetadataservDB_bkup_<date>.sql
```

Scheduling

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

Restart

You can restart the Trifacta platform now. See *Start and Stop the Platform*.

Recovery

See *Platform Rollback*.

Platform Rollback

Contents:

In the event that an upgrade or hotfix to your instance of the Trifacta® platform has run into issues that cannot be repaired in the upgraded instance, you can follow the steps in this section to rollback to your previous version.

> NOTE: Before you perform a rollback, you should review the set of issues with Trifacta first. For more information, please contact *Trifacta Support*.

Pre-requisites

In order to complete the rollback in a timely manner, please verify that you have access to the following:

Access

You must:

- Acquire root user access to the Trifacta node.
- Acquire database access to uninstall and reinstall the Trifacta databases.

> **Tip:** You should communicate to any affected users the required maintenance and expected outage window.

Backups

> **If you do not have the following, you cannot perform a rollback. These items cannot be acquired from Trifacta.**

- Backups of your pre-upgrade Trifacta configuration files
- Backups of your pre-upgrade Trifacta databases

The following can be acquired from Trifacta if you do not have them:

- RPM installers for the previous version. If any Hotfixes have been applied to the previous version, you should acquire and use the latest Hotfix RPM for your re-install.
- PDF documentation for the previous version.

Rollback Steps

To recover the Trifacta platform based on backups:

> **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 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.
    1. **PostgreSQL:**

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

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

        3. (Release 3.2 and later) Jobs database:

> **NOTE:** Please note the escaped quotes in the `CREATE DATABASE` command for this database.

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

        4. (Release 4.1 and later) Scheduling database:

```
psql -c "DROP DATABASE trifactaschedulingservice;"
psql -c "CREATE DATABASE trifactaschedulingservice WITH
OWNER trifactascheduling;"
psql --dbname=trifactaschedulingservice <
trif_schDB_bkup_<date>.sql
```

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

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

6. (Release 6.0 and later) Configuration Service database:

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

7. (Release 6.0 and later) Artifact Storage Service database:

> **NOTE:** This database is unused in this product. It may be used in the future for managing feature-specific data.

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

8. (Release 6.4 and later) Job Metadata Service database:

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

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

```
su - mysql
```

   2. Trifacta database:

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

   3. Jobs database:

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

   4. (Release 4.1 and later) Scheduling database:

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

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

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

6. (Release 6.0 and later) Configuration Service database:

```
mysql trifactaconfigurationservice <
trif_confservDB_bkup_<date>.sql
```

7. (Release 6.0 and later) Artifact Storage Service database:

> **NOTE:** This database is unused in this product. It may be used in the future for managing feature-specific data.

```
mysql trifactaartifactstorageservice <
trif_artifactstorageservDB_bkup_<date>.sql
```

8. (Release 6.4 and later) Job Metadata Service database:

```
mysql trifactajobmetadataservice <
trif_jobmetadataservDB_bkup_<date>.sql
```

4. Uninstall the current version of the Trifacta software:

> **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
```

5. Perform a clean install of the Trifacta software provided in your distribution. See *Install*.

6. Restore your configuration files. The following commands assume that they were backed up to the `/tmp` dir ectory 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
```

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

Admin Reference

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

Admin Settings Page

Contents:

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. So me 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 distributions. • See *Configure for Cloudera*. • See *Configure for Hortonworks*.
`[CLUSTER]`	Settings that affect how the platform interacts with the integrated backend cluster. See *Prepare Hadoop for Integration with the Platform*.
`[HIVE]`	Settings pertaining to integration with the connected instance of Hive. See *Configure for Hive*.
`[HA]`	Settings pertaining to integration with cluster high availability for the Trifacta platform. See *Enable Integration with Cluster High Availability*.
`[SECURITY]`	General settings pertaining to security. See *Configure Security*.
`[SSL]`	Settings pertaining to SSL access to the platform. See the SSL section in *Configure Security*.
`[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.

AWS Region

Enter the AWS region code where the EMR cluster is hosted. For example:

```
us-east-1
```

For a list of available regions, see *https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-regions-availability-zones.html#concepts-available-regions* .

Resource Bucket

The name of the default S3 bucket where platform resources are stored

Resource Path

The path in the default S3 bucket to where resources are stored

After you have made any changes to the AWS properties, click **Save**.

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.

View Logs

Click the View Logs link to review and download the logs maintained on the Trifacta node.

For more information on these logs, see *System Services and Logs*.

For more information, see *Configure Logging for Services*.

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*.

Deployment Manager Page

Contents:

Through the Deployment Manager 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 Deployment Manager 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 Manager*.

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 Manager*.

Deployment Hierarchy

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

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

2	release	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. You can activate previous releases as needed through the context menus in Deployment Manager.

3	flow or flows	Within a release, you can explore the flows that were included in the import package for the release:
		
		• The **primary flow** is the one that is executed when:
		  • 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.
		  • 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 Deployment Manager, you can explore all of the deployments in the Production instance.

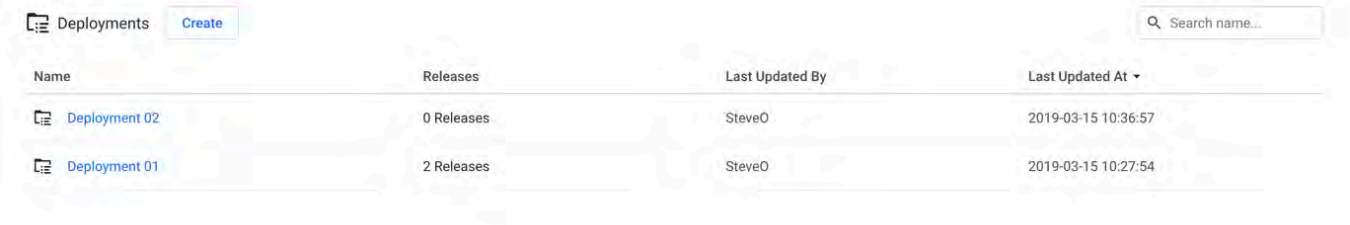

Figure: Deployment Manager

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.

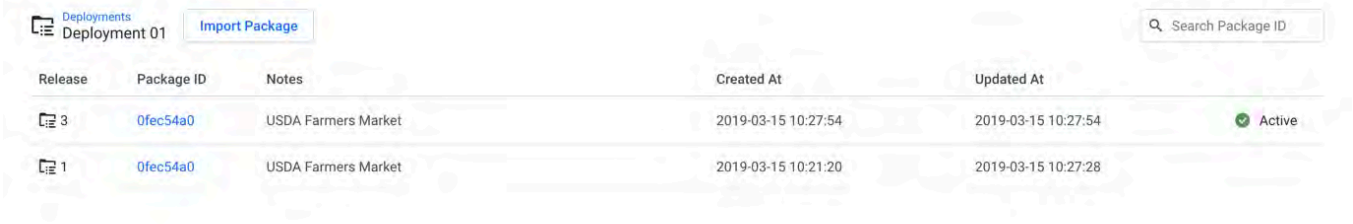

Figure: Releases View

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 *Export Flow*.

Delete	Delete the release from the Production instance. **NOTE:** If the release was imported from a Dev instance on the same platform, the Dev instance of the release is not removed. **NOTE:** Deletion of a release does not remove any results generated from it. Those results are still accessible through the Jobs page.

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.

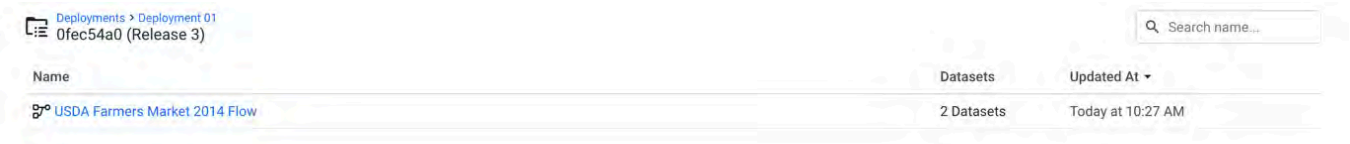

Figure: Flows View

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*.

Workspace Admin Reference

These pages assist workspace administrators in managing their workspaces, users, and user roles.

Workspace Admin Page

Contents:

- *Product walkthroughs*
- *Session duration*
- *Sample downloads*
- *Allow the user to modify their paths*
- *Schematized output*
- *Parquet output format*
- *JSON output format*
- *CSV output format*

- *Avro output format*
- *TDE output format*
- *API Access Token*
- *Parameterization*
- *Output parameterization*
- *Allow users to export their flows*
- *Allow users to import flows into Trifacta*
- *Enable Flow Sharing feature*
- *Forbid users to add non-default publishing actions*
- *Hide underlying file system to users*
- *Enable publishing*
- *Enable UI for range join*
- *Enable Scheduling feature*
- *Allow users to send copies of flows to other users*
- *Show datasource tab in the application*
- *Show file location*
- *Show output directory in profile view*
- *Show upload directory in profile view*
- *Enable Connectivity feature*
- *Enable custom SQL query*
- *Show users a modal to upgrade to a plan with Connectivity*
- *Show users a modal to upgrade to a plan with uploadLargeFiles*
- *Enable self-service password reset*
- *Schedule List*

The following settings can be customized for the user experience in your workspace. When you modify a setting, the change is immediately applied to the workspace.

> **NOTE:** Users may not experience the changed environment until each user refreshes the application page or logs out and in again.

Options:

> **NOTE:** Any values specified in the Workspace Admin page applies exclusively to the specific workspace and override any system-level defaults.

Option	Description
`Default`	The default value is applied. This value may be inherited from higher level configuration. **Tip:** You can review the default value as part of the help text.
`Enabled`	The setting is enabled. **NOTE:** If the setting applies to a feature, the feature is enabled. Additional configuration may be required. See below.
`Disabled`	The setting is disabled.
`Edit`	Click **Edit** to enter a specific value for the setting.

Product walkthroughs

When enabled, new members are led through a tour of the product. After the tour is dismissed, it does not re-appear again to the member.

> **NOTE:** This feature may need to be enabled in your environment by an administrator. For more information, see *Enable Onboarding Tour*.

Session duration

Maximum length in minutes of a member's session, after which the session is terminated, and the member must login again.

Sample downloads

When enabled, members can download the contents of the Transformer page at any time. For an individual step, a member can download the current sample, as modified by the current recipe up to the point of the current step. For more information, see *Recipe Panel*.

Allow the user to modify their paths

Allow members of the workspace to change paths to their upload and output results locations through their user profile. See *Storage Config Page*.

Schematized output

When enabled, all output columns are typecast to their annotated types. This feature is enabled by default.

Parquet output format

When enabled, members can generate outputs in Parquet format.

JSON output format

When enabled, members can generate outputs in JSON format.

CSV output format

When enabled, members can generate outputs in CSV format.

Avro output format

When enabled, members can generate outputs in Avro format.

TDE output format

When enabled, members can generate outputs in TDE format.

API Access Token

When accessing the REST APIs, you can optionally use a token for simpler use and enhanced security.

NOTE: This feature may not be available in all environments.

For more information, see *Access Tokens Page*.

Parameterization

When enabled, users can create parameters, which can be applied to import, creating sample, and outputs. For more information, see *Overview of Parameterization*.

Output parameterization

When enabled, parameters, can be applied to outputs.

NOTE: Parameterization must also be enabled.

See *Overview of Parameterization*.

Allow users to export their flows

When enabled, workspace users are permitted to export their flow definitions in a ZIP file. See *Export Flow*.

Allow users to import flows into Trifacta

When enabled, workspace users are permitted to import exported flows from a ZIP file. See *Import Flow*.

Enable Flow Sharing feature

When enabled, workspace users are permitted to share flows with other users in the workspace. See *Share a Flow*.

Forbid users to add non-default publishing actions

When enabled, workspace users are not permitted to specify publishing actions, which can be used to control export of results to unexpected locations or systems.

Hide underlying file system to users

When enabled, workspace users cannot see locations in the default storage layer.

Enable publishing

When enabled, workspace users are permitted to publish results through the Output Destinations tab in the Job Details page to external datastores.

NOTE: These external datastores must be enabled and configured. See *Connection Types*.

For more information, see *Job Details Page*.

Enable UI for range join

When enabled, workspace users can specify join key matching across a range of values.

Enable Scheduling feature

When enabled, workspace users can schedule the execution of flows. See *Add Schedule Dialog*.

Allow users to send copies of flows to other users

When enabled, workspace users can send an independent copy of a flow to other workspace users. For more information, see *Send a Copy of a Flow*.

Show datasource tab in the application

When enabled, workspace users can review the sources of data for a job through the Data sources tab in the Job Details page. See *Job Details Page*.

Show file location

When enabled, workspace users can see the locations of source and output files within the application.

Show output directory in profile view

When enabled, workspace users can review the directory where outputs were generated in the Profile tab of the Job Details page. See *Job Details Page*.

Show upload directory in profile view

When enabled, workspace users can review the directory where file uploads are posted in the Profile tab of the Job Details page. See *Job Details Page*.

Enable Connectivity feature

When enabled, workspace users can create connections to relational datasources.

> **NOTE:** Disabling this feature hides existing relational connections.

See *Enable Relational Connections*.

Enable custom SQL query

When enabled, users can create custom SQL queries to import datasets from relational tables. For more information, see *Enable Custom SQL Query*.

Show users a modal to upgrade to a plan with Connectivity

When enabled, workspace users are presented with the option to upgrade to a plan that supports connection to external data sources, if the feature is current disabled.

Show users a modal to upgrade to a plan with uploadLargeFiles

When enabled, workspace users are presented with the option to upgrade to a plan that supports uploading large files, if the feature is current disabled.

Enable self-service password reset

When enabled, workspace users can reset their own passwords via link on the login page.

Schedule List

When enabled, administrators and workspace administrators can see a list of all schedules in the workspace.

Create column from examples feature

When enabled, users can access a tool through the column menus that enables creation of new columns based on example mappings from the selected column. For more information, see *Overview of TBE*.

Admin for AWS

General AWS Configuration:

Topic	Description
Configure for AWS	General configuration for AWS and its components.
Configure for EMR	How to integrate the Trifacta® platform with an Elastic Map Reduce (EMR) cluster.
Enable S3 Access	How to enable the Trifacta platform to integrate with S3. **NOTE:** S3 can be used as the primary backend datastore for the Trifacta platform.
Create Redshift Connections	How to create a connection to your Redshift instance hosted in S3.

The following special topics apply to configuring your AWS environment to enable access for the platform.

Insert Trust Relationship in AWS IAM Role

If you are using per-user authentication through an AWS IAM role, you must insert a trust relationship into the role so that the Trifacta® platform can leverage it.

Pre-requisites:

Please acquire the following information:

- **IAM role:** The AWS IAM role that the Trifacta platform should use.
- **EC2 instance role:** If the EC2 instance role is to be used to assume the AWS role, then please acquire the following:
    - AWS account ID
    - EC2 instance role
    - Details on the above are listed below.

For more information on the AWS Principal options described below, please review *https://docs.aws.amazon.com/IAM/latest/UserGuide/reference_policies_elements_principal.html*.

Steps:

1. You can apply this change through the *Admin Settings Page* (recommended) or `trifacta-conf.json`. For more information, see *Platform Configuration Methods*.
2. Locate the following parameter and retrieve its value (`true` or `false`):

```
"aws.ec2InstanceRoleForAssumeRole"
```

3. Login to the AWS console.
4. Open the IAM role for use with the Trifacta platform.
5. If `aws.ec2InstanceRoleForAssumeRole=true`, then the EC2 instance role is used for assuming the provided AWS role. Paste the following into the IAM role for the trust relationship:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "AWS": [
           "arn:aws:iam::<awsAccountId>:role/<ec2InstanceRole>"
        ]
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

Property	Description
`<awsAccountId>`	AWS account identifier for which the EC2 instance role is assumed
`<ec2InstanceRole>`	EC2 instance role to use

6. If `aws.ec2InstanceRoleForAssumeRole=false`, then the AWS user associated with the provided AWS key and secret is assumed. Paste the following into the IAM role for the trust relationship:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "AWS": [
          "arn:aws:iam::862753480162:user/sample-user"
        ]
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

7. Save the IAM role definition.