TRIFACTA

Install Guide for Amazon Marketplace

Version: 5.0
Doc Build Date: 09/24/2018

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

Install from Amazon Marketplace

Contents:

> **NOTE:** If you are installing or upgrading a Marketplace deployment, you must use the install and configuration materials available through the Marketplace listing. Online materials may be referenced afterward.

This guide steps through the requirements and process for installing Trifacta® Wrangler Pro through the Amazon Marketplace.

Product Limitations

- Jobs must be executed on the Trifacta Server. No other running environment integrations are supported.
- Anomaly and stratified sampling are not supported in this deployment.
- When publishing single files to S3, you cannot apply an `append` publishing action.
- The EC2 instance, S3 buckets, and any connected Redshift databases must be located in the same Amazon region. Cross-region integrations are not supported at this time.

Assumptions

This document assumes that you are setting up Trifacta Wrangler Enterprise to use Amazon's preferred EC2 role-based authentication for access to AWS resources.

> **Tip:** Using EC2 role-based authentication is recommended by AWS. For more information, see *https://docs.aws.amazon.com/IAM/latest/UserGuide/best-practices.html#use-roles-with-ec2*.

Trifacta Wrangler Enterprise also supports the following authentication methods:

- System mode - All users of Trifacta Wrangler Enterprise use the same key and secret combination to access resources.
- User mode - Each user has a separately specified key and secret combination to access resources.

If you are using one of these two access methods, please do the following:

1. Specify an EC2 role without any permissions. Specifically, it should have no data access permissions, as this

role cannot be changed at a later time.
2. Complete the following sequence through the Install Steps. Specify the above EC2 role as part of the configuration.
3. When you launch the product, you can specify the appropriate access mode through the platform. For more information, see *Configure for AWS* in the Install Guide.
    a. This content is also available through the online Documentation referenced at the end of this document.
4. Complete any related configuration through AWS as needed.

> **Tip:** If you want to use EC2 roles at a later time, you can just apply AWS policies to the empty role you created here. Additional configuration is required in the platform to use this role.

Install

Desktop Requirements

- All desktop users of the platform must have the latest version of Google Chrome installed on their desktops.
    - Google Chrome must have the PNaCl client installed and enabled.
    - PNaCl Version: `0.50.x.y` or later
- All desktop users must be able to connect to the EC2 instance through the enterprise infrastructure.

Sizing Guide

> **NOTE:** The following guidelines apply only to Trifacta Wrangler Pro.

Use the following guidelines to select your instance size:

> **NOTE:** Trifacta Wrangler Pro enforces a maximum limit of 30 users.

Size	max recommended users	Avg. size of jobs on Trifacta Server (GB)
Small	5	5 GB
Medium	15	20 GB
Large	30	40 GB

Pre-requisites

Before you install the platform, please verify that the following steps have been completed.

1. **EULA.** Before you begin, please review the End-User License Agreement. See *End-User License Agreement*.
2. **S3 bucket.** Please create an S3 bucket to storeTrifacta assets. In the bucket, the platform stores metadata in the following location:

    ```
    <S3_bucket_name>/trifacta
    ```

    See *https://s3.console.aws.amazon.com/s3/home*.
3. **IAM policies.** Create IAM policies for access to the S3 bucket. Required permissions are the following:
    - The system account or individual user accounts must have full permissions for the S3 bucket:

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

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

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

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

4. **EC2 instance role.** Create an EC2 instance role for this policy. See *https://console.aws.amazon.com/iam/home#/roles*.

Install Steps

1. Launch Trifacta Wrangler Pro from AMI.
2. In the EC2 Console:
    a. **Instance size:** Select the instance size. See*Sizing Guide*.
    b. **Network:** Configure the VPC, subnet, firewall and other configuration settings necessary to communicate with the instance.
    c. **Auto-assigned Public IP:** You must create a public IP to access the Trifacta platform.
    d. **EC2 role:** Select the EC2 role that you created.
    e. **Local storage:** Select a local EBS volume. The default volume includes 100GB storage.

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

    f. **Security group:** Use a security group that exposes access to port 3005, which is the default port for the platform.
    g. **Create an AWS key-pair for access:**This key is used to provide SSH access to the platform, which may be required for some admin tasks. Save key file to your local computer for later use.
    h. Save your changes.
3. Launch the configured version of Trifacta Wrangler Pro.

    > **NOTE:** From the EC2 Console, please acquire the `instanceId`, which is needed in a later step.

4. When the instance is spinning up for the first time, performance may be slow. When the instance is up, please navigate to the following:

    ```
    http://<public_hostname>:3005
    ```

5. When the login screen appears, enter the following:
    a. Username: *admin@trifacta.local*
    b. Password: (the `instanceId` value)

        > **NOTE:** As soon as you login as an admin for the first time, you should immediately change the password. Select the User Profile menu item in the upper-right corner. Change the password and click **Save** to restart the platform.

6. From the application menu, select **Settings menu > Admin Settings**.
7. In the Admin Settings page, you can configure many aspects of the platform, including user management tasks, and perform restarts to apply the changes.
    a. In the Search bar, enter the following:

        ```
        aws.s3.bucket.name
        ```

    b. Set the value of this setting to be the bucket that you created for Trifacta Wrangler Pro.
8. The following setting must be specified.

    ```
    "aws.mode":"system",
    ```

You can set the above value to either of the following:

aws.mode value	Description
`system`	Set the mode to `system` to enable use of EC2 instance-based authentication for access.
`user`	Set the mode to `user` to utilize user-based credentials. This mode requires additional configuration.

Details on the above configuration are described later.

9. Click **Save**.
10. When the platform restarts, you can begin using the product.

SSH Access

If you need to SSH to the Trifacta node, you can use the following command:

```
ssh -i <path_to_key_file> <userId>@<tri_node_DNS_or_IP>
```

Parameter	Description
`<path_to_key_file>`	Path to the key file stored on your local computer.
`<userId>`	The user ID is always `centos`.
`<tri_node_DNS_or_IP>`	DNS or IP address of the Trifacta node

Upgrade

Please complete the instructions in this section if you are upgrading from a previous version of Trifacta® Wrangler Pro.

NOTE: These instructions apply only to Trifacta® Wrangler Pro available the Amazon Marketplace. If you are upgrading Trifacta Wrangler Enterprise, please follow the upgrade instructions provided to you by your Trifacta representative.

Backup data from platform instance

Before you begin, you should backup your current instance.

1. SSH to your current Marketplace AMI instance. See instructions in the Install section.
2. Stop the Trifacta platform on your current Marketplace AMI instance:

```
sudo service trifacta stop
```

3. Update the backup script with a more current version.
   a. Download 5.0.0 backup script from the following location:

```
https://raw.githubusercontent.com/trifacta/trifacta-utils/releas
e/5.0/trifacta-backup-config-and-db.sh
```

b. Example command to download the script:

```
curl --output trifacta-backup-config-and-db.sh
https://raw.githubusercontent.com/trifacta/trifacta-utils/releas
e/5.0/trifacta-backup-config-and-db.sh
```

c. Overwrite the downloaded script to the following location:

```
/opt/trifacta/bin/setup-utils/trifacta-backup-config-and-db.sh
```

d. Verify that this script is executable:

```
sudo chmod 775
/opt/trifacta/bin/setup-utils/trifacta-backup-config-and-db.sh
```

4. Run the backup script:

```
sudo /opt/trifacta/bin/setup-utils/trifacta-backup-config-and-db.sh
```

a. When the script is complete, the output identifies the location of the backup. Example:

```
/opt/trifacta-backups/trifacta-backup-4.2.1+126.20171217124021.a
8ed455-20180514213601.tgz
```

5. Store the backup in a safe location. If needed, you can store this backup in the S3 bucket used by the platform. Example:

```
aws s3 cp
/opt/trifacta-backups/trifacta-backup-4.2.1+126.20171217124021.a8ed45
5-20180514213601.tgz s3://<my-trifacta-s3-bucket>/trifacta-backups/
```

Spin up upgraded platform instance and restore data

In this section, you spin up the upgraded instance and then restore the data that you have backed up into the instance.

Steps:

1. Spin up the new instance from the Amazon Marketplace AMI. For more information on installation, please see the instructions earlier in this document.
2. Verify that the instance has successfully started. If you can connect to the login page, the platform has started.
3. SSH to the new Amazon Marketplace AMI instance. See previous instructions in the Install section.
4. Stop the Trifacta platform on the instance:

```
sudo service trifacta stop
```

5. Download the restore script from the following location:

```
https://raw.githubusercontent.com/trifacta/trifacta-utils/release/5.0
/trifacta-restore-from-backup.sh
```

   a. Example command to download the script:

```
curl --output trifacta-restore-from-backup.sh
https://raw.githubusercontent.com/trifacta/trifacta-utils/releas
e/5.0/trifacta-restore-from-backup.sh
```

   b. Place the restore script in the following location in the instance:

```
/opt/trifacta/bin/setup-utils/trifacta-restore-from-backup.sh
```

   c. Verify that this script is executable:

```
sudo chmod 775
/opt/trifacta/bin/setup-utils/trifacta-restore-from-backup.sh
```

6. Download the backup from your storage location and extract its contents. Example:

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

```
sudo mkdir -p /root/trifacta-restore-files
sudo cd /root/trifacta-restore-files
sudo aws s3 cp \

s3://<my-trifacta-s3-bucket>/trifacta-backups/trifacta-backup-4.2.1+1
26.20171217124021.a8ed455-20180514213601.tgz .
sudo tar xzf
trifacta-backup-4.2.1+126.20171217124021.a8ed455-20180514213601.tgz
```

7. The backup contents should be located in a directory with a path similar to the following:

```
/root/trifacta-restore-files/trifacta-backup-4.2.1+126.20171217124021
.a8ed455-20180514213601
```

8. Execute the restore script. Pass in the path to your unzipped backup as a parameter, as in the following example:

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

```
sudo /opt/trifacta/bin/setup-utils/trifacta-restore-from-backup.sh \
  -r
/root/trifacta-restore-files/trifacta-backup-4.2.1+126.20171217124021
.a8ed455-20180514213601
```

Verify

The upgrade is complete. To verify:

Steps:

1. Restart the platform:

```
sudo service trifacta start
```

2. Run a simple job with profiling.
3. Verify that the job has successfully completed.
    a. In the Jobs page, locate the job results. See *Jobs Page*.
    b. Click **View Details** next to the job to review the profile.

Documentation

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