



# Trillium™ Quality for Big Data

## Installation Guide

Version 15.8.2

This manual contains proprietary and confidential material, and is only for use by licensees of Trillium Software, Inc. This manual, as well as the software described in it, are furnished under license and may be used only in accordance with the terms of such license. The content of this manual is furnished for informational purposes only, is subject to change without notice, and should not be construed as a commitment by Trillium Software, Inc.

THE INFORMATION IN THIS MANUAL IS PROVIDED "AS IS" WITH ALL FAULTS. ALL STATEMENTS, INFORMATION, AND RECOMMENDATIONS IN THIS PUBLICATION ARE BELIEVED TO BE ACCURATE BUT ARE PRESENTED WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. CUSTOMERS MUST TAKE FULL RESPONSIBILITY FOR THEIR APPLICATION OF ANY OF THE INFORMATION CONTAINED HEREIN. TRILLIUM SOFTWARE, INC. SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OF NONINFRINGEMENT, OR ARISING FROM A COURSE OF DEALING, USAGE, OR TRADE PRACTICE. IN NO EVENT SHALL TRILLIUM SOFTWARE, INC. OR ITS SUPPLIERS BE LIABLE FOR ANY INDIRECT, SPECIAL, CONSEQUENTIAL, OR INCIDENTAL DAMAGES, INCLUDING, WITHOUT LIMITATION, LOST PROFITS OR LOSS OR DAMAGE TO DATA ARISING OUT OF THE USE OR INABILITY TO USE THIS PUBLICATION, EVEN IF TRILLIUM SOFTWARE, INC. OR ITS SUPPLIERS HAVE BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

The customer shall not disclose, copy, reproduce, distribute, or display any portion of the Trillium Software System or this manual in any form to any third person without the prior written consent of Trillium Software, Inc., nor allow third parties to do the same. The customer shall keep the Trillium Software System and all confidential information, including this manual, in the strictest confidence.

***Trillium™ Quality for Big Data Installation Guide  
May 2019, 5719***

© 2018, 2019 Trillium Software, Inc. All rights reserved. Trillium Software and Trillium Software System are trademarks of Trillium Software, Inc. All other company and product names used herein may be the trademarks of their respective companies.

Trillium Software, Inc  
1700 District Avenue, Suite 300, Burlington, MA 01803, USA  
978-901-0000 (Worldwide Headquarters)  
+44 (0) 118-940-7676 (European Headquarters)

# Contents

<b>PREFACE</b> .....	5
<b>CHAPTER 1</b>	<b>Overview and Requirements</b> ..... 6
	About the Installation ..... 6
	Prerequisites..... 7
	Supported Hadoop Distributions ..... 7
	Certification ..... 8
<b>CHAPTER 2</b>	<b>Installing Trillium Quality for Big Data - Ambari Packages</b> ..... 9
	Deploying the Ambari Packages..... 9
	Installing Trillium Quality for Big Data ..... 11
	Installing GAV Tables..... 12
	Uninstalling Trillium Quality for Big Data ..... 13
<b>CHAPTER 3</b>	<b>Installing Trillium Quality for Big Data - Cloudera Manager Parcels</b> ..... 15
	Downloading Parcels..... 15
	Installing Trillium Quality for Big Data ..... 17
	Uninstalling Trillium Quality for Big Data ..... 17
<b>CHAPTER 4</b>	<b>Installing Trillium Quality for Big Data - Amazon EMR</b> ..... 19
	Installing Trillium Quality for Big Data on New Clusters ..... 20
	Installing Trillium Quality for Big Data on Existing Clusters..... 23
	Prerequisites ..... 23
	Installing Trillium Quality for Big Data..... 23
	Uninstalling Trillium Quality for Big Data ..... 28
<b>CHAPTER 5</b>	<b>Installing Trillium Quality for Big Data - RPM</b> ..... 29
	Installing Trillium Quality for Big Data ..... 29
	Uninstalling Trillium Quality for Big Data ..... 31
<b>CHAPTER 6</b>	<b>Using Distributed Cache for Trillium Global Address Verification (GAV) Tables</b> ..... 33
	Installing GAV Tables for Distributed Cache ..... 33
	Guidelines ..... 33

	Notes for Global GAV Tables/Geocoding Tables.....	35
<b>CHAPTER 7</b>	<b>Setting Up SSH</b> .....	36
	Setting up Security .....	36
	Brief Overview of the Process .....	40
	Error Messages.....	40
<b>CHAPTER 8</b>	<b>Troubleshooting</b> .....	41
	Ambari Server Fails to Start.....	41
	Ambari Installation Fails to Modify Config Files for GAV Tables.....	42
<b>INDEX</b>	.....	44

---

# Preface

This document describes how to install and uninstall Trillium™ Quality for Big Data. Trillium Quality for Big Data integrates Trillium's industry-leading data quality application with Syncsort's DMX-h™ to provide a highly scalable, high-performance implementation of data cleansing and matching services within the Hadoop environment.

## Intended Audience

This document is intended for users who will be installing Trillium Quality for Big Data.

## Customer Support

Customer Support is available Monday through Friday, 24 hours a day, to help you with any questions you may have about Trillium Quality for Big Data. This service is available to all customers who have active maintenance plans.

- **E-mail:** [TrilliumSupport@syncsort.com](mailto:TrilliumSupport@syncsort.com)
- **Tel:** **978-901-0000 (Worldwide Headquarters)**  
**+44 (0) 118 946 7676 (European Headquarters)**

## CHAPTER 1

# Overview and Requirements

This chapter provides an overview and requirements for installing Trillium Quality for Big Data.

## About the Installation

Trillium Quality for Big Data is a Hadoop extension of Trillium Software® System (TSS) and provided as Apache Ambari service packages, Cloudera parcels, and RPM packages. Trillium Quality for Big Data Version 15.8.2 includes:

- Trillium™ Quality - TSS Version 15.8.2
- Trillium™ Global Address Verification (GAV) Tables
- Syncsort DMX-h™ Release 9.7.32

① *Trillium Global Address Verification (GAV) tables were formerly named TSS Postal Directories.*

If you need to create a new Trillium project or export an existing Trillium project to Hadoop, installation of the following TSS components is also required. These components allow you to create, test, and export Trillium projects to the Hadoop environment.

- Repository Server
  - Trillium™ Control Center
- ① *The Trillium Control Center installation includes the TSQ Batch to Big Data Deployment Manager.*
- Country Templates
  - Trillium™ Global Address Verification (GAV) Tables
  - Batch Deployment Tool

See the *Trillium Software® System Version 15.8 Installation Guide* for the system requirements, installation procedures, and license information for

these components.

## Prerequisites

Trillium Quality for Big Data requires the following:

- Java/JRE (or JDK) 1.7 64-bit or higher
- For Ambari Service installation, Ambari repository and interface for Apache Ambari 1.7 and higher
- Root/sudo privileges. For Hortonworks Ambari, Cluster Administrators role is required.
- A minimum of 40 GB of free local disk space is required for each node for all installations including software, Global Address Verification (GAV) tables, and intermediary files. A minimum of 40 GB of free local disk space is also required for the machine where the Ambari server or Cloudera Manager Server resides.

## Supported Hadoop Distributions

### MapReduce

- Cloudera CDH
  - 6.0
  - 5.x (5.2 and higher)– YARN (MRv2)
- Hortonworks Data Platform (HDP) 2.x (2.1 and higher) – YARN
- Amazon Elastic MapReduce (EMR) 5.x
- Apache Hadoop 2.x (2.2 and higher) – YARN
- MapR, Community Edition and Enterprise Edition only (previously termed M5 and M7, respectively)
  - 5.x – YARN
  - 4.0.1, 4.0.2, 4.1 – YARN
- Pivotal HD 3.0 – YARN
- BigInsights 4 – YARN

Trillium Quality for Big Data is certified as ODPI (1.0 and higher) interoperable.

### Spark

- Spark on YARN on the following Hadoop distributions:
  - Cloudera CDH 6.0 and 5.x (5.5 and higher)
  - Hortonworks Data Platform (HDP) 2.3.4, 2.x (2.4 and higher)
  - MapR 5.x (5.1 and higher), Community Edition and Enterprise Edition only (previously named M5 and M7, respectively)
- Spark on Mesos 0.21.0
- Spark Standalone 1.5.2 and higher

## Certification

Trillium Quality for Big Data is certified in the following:

- Hortonworks Data Platform (HDP) 2.6. Trillium Quality for Big Data has been approved as a YARN/SEC/OPS ready solution.
- Cloudera 6 as a Cloudera Certified Technology.

## CHAPTER 2

# Installing Trillium Quality for Big Data - Ambari Packages

This chapter describes how to install and uninstall Trillium Quality for Big Data using Ambari packages.

Trillium Quality for Big Data is provided as two Apache Ambari service packages, one for the Big Data solution which includes Trillium Quality and DMX-h, and one for Trillium Global Address Verification (GAV) tables. For each package, you will need to deploy the service definition package to your Ambari repository, and then install the package on the nodes using the Ambari web interface.

The Big Data solution must be installed on *all nodes* in the Hadoop cluster, as well as on an *edge node* that has access to the cluster. GAV tables can be installed on all nodes in the Hadoop cluster or only on one node if distributed cache is utilized. See [Chapter 6, “Using Distributed Cache for Trillium Global Address Verification \(GAV\) Tables”](#) for more information about distributed cache.

- ① *Trillium Global Address Verification (GAV) tables were formerly named TSS Postal Directories. GAV tables include the country-specific postal tables and system resource files such as code\_pages, general\_resources, and parser\_rules.*

## Deploying the Ambari Packages

Execute the following procedure on the machine where the Ambari server resides:

### ► To deploy the Ambari packages

1. Download the Trillium Quality for Big Data installation package from TSS Download Center (File Portal) FTP website. The package includes the following self-extracting shrink-wrap executables:

- Trillium Quality for Big Data (Trillium Quality and DMX-h)

`tsqbd-version-linuxOS.ambari-service.bin`

- GAV tables

`tss_directories-client_name.date-linuxOS.ambari-service.bin`

**Where:**

- `version` = the Trillium Quality for Big Data version number.
- `linuxOS` = Linux OS version. For example, `e17` for RHEL7.x.
- `client_name.date` = Client name and the date (YYMM) when the tables were built.

2. Run one of the shrink-wrap executables.

**Examples**

```
./tsqbd-15.8.2.18200.bd-e17.ambari-service.bin
```

```
./tss_directories-CRM_WW.1711-e17.ambari-service.bin
```

3. Read and accept the Software License Agreement.
4. Enter a target directory in which to extract the installation folder, or press **Enter** to accept the default, which is `/opt/bin`.
5. If the Trillium Quality for Big Data or TSS Postal Directories service definition already exists in the repository, you will be prompted to remove the existing version. Enter `y` to remove, or `n` to exit the process.

❗ *This is an update step. If you are installing the same version in a different location, you will first need to uninstall the existing version and then install again. See [Uninstalling Trillium Quality for Big Data on page 13](#).*

6. Enter the Ambari server's address, username, and password, and the cluster name, as prompted. If the credentials entered fail, you can re-run this step manually by executing the following script, where `<Ambari service extracted package path>` is the directory you specified in Step 4:

```
<Ambari service extracted package path>/services/TSQBD/  
package/scripts/prepare_tsqbd_upgrade.sh
```

```
<Ambari service extracted package path>/services/TSS_  
DIRECTORIES/package/scripts/prepare_tss_directories_  
upgrade.sh
```

7. Enter a target directory in which to copy the service package where it can be found by the Ambari server, or press **Enter** to accept the default, which is the root path of the latest stack.
8. Enter `y` to restart the Ambari server in order for the new package to be picked up, or `n` to restart later.

The following completion message is displayed:

```
The TSQBD (TSS_DIRECTORIES) service files has been
successfully added to the repository, and the TSQBD (TSS_
DIRECTORIES) service can now be managed from the Ambari web
interface.
```

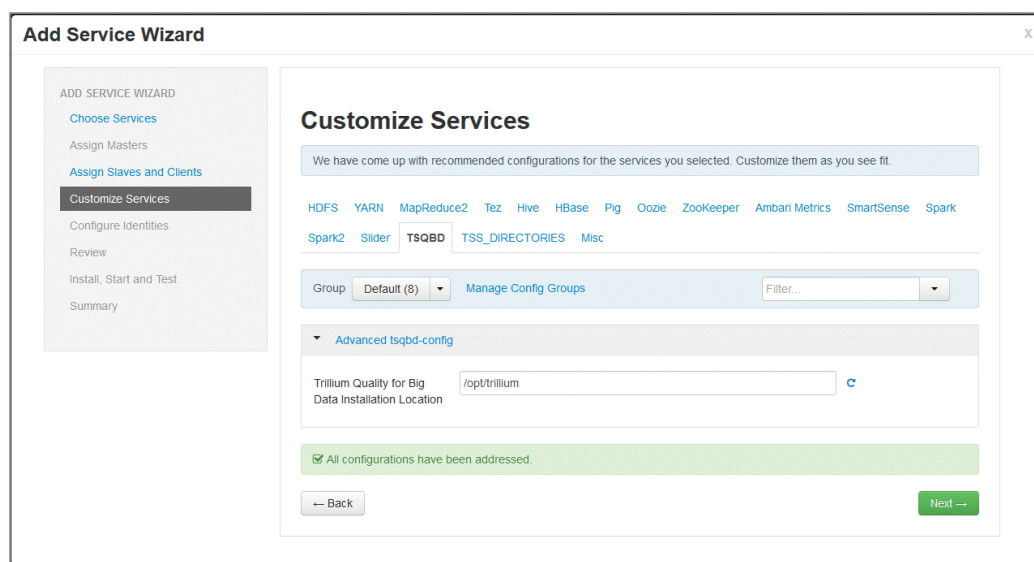
9. If another shrink-wrap executable needs to be run, repeat Step 2 through 8.

## Installing Trillium Quality for Big Data

Install the Trillium Quality for Big Data service on all nodes in the cluster, as well as on an *edge node* that has access to the Hadoop cluster.

### ► To install Trillium Quality for Big Data

1. Log in to the Ambari web interface and select **Actions > +Add Service**.
2. Select **TSQBD** and click **Next**. The Add Service Wizard page opens.
3. Click **Next**. The **Assign Slaves and Clients** page opens.
4. Select all nodes you need, and click **Next**. The Customize Services page opens with the TSQBD tab displayed.



**Figure 2-1 Customize Services Page - TSQBD**

5. Expand the **Advanced tsbd-config** section.
6. Specify the installation location in the text box. The default location is `/opt/trillium`.
7. Click **Next**. The Configure Identities page opens.

8. Click **Next**. The Review page opens.
9. On the **Review** page, verify the information and click **Deploy**.
10. Enter your Admin principal name and Admin password and click **Save**. The Install, Start, and Test page opens.
  - ① *You must append @WINKERB2.ORG to your user name for the Admin principal name.*
11. On the **Install, Start and Test** page, wait for the Trillium Quality for Big Data services to be successfully installed and started on each node.
12. When services are complete, click **Next**. The Summary page opens.
13. Click **Complete**.

## Installing GAV Tables

GAV tables include the postal tables and system resource files, such as code\_pages, general\_resources, and parser\_rules. Install the TSS Postal Directories service for the GAV tables on all nodes in the cluster or only on one node if distributed cache is utilized. See [Chapter 6, "Using Distributed Cache for Trillium Global Address Verification \(GAV\) Tables"](#) for more information about distributed cache.

### ► To install GAV tables

1. Log in to the Ambari web interface and select **Actions > +Add Service**.
2. Select **TSS\_DIRECTORIES** and click **Next**. The Add Service Wizard page opens.
3. Click **Next**. The **Assign Slaves and Clients** page opens.
4. Select all nodes you need, and click **Next**. The Customize Services page opens with the TSS\_DIRECTORIES tab displayed.

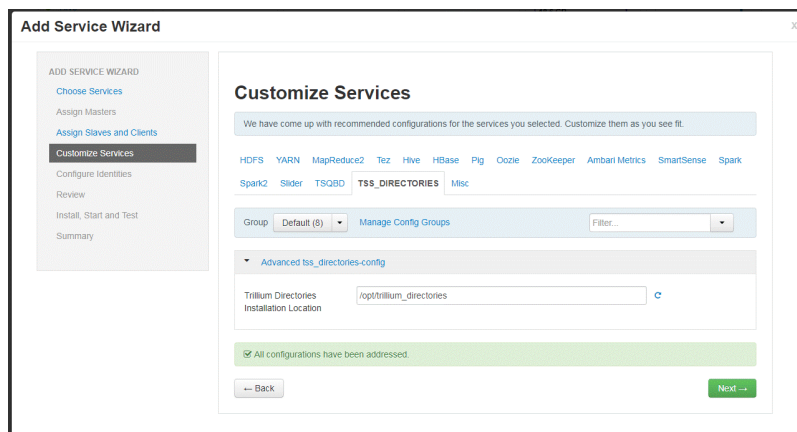


Figure 2-2 Customize Services Page - TSS\_DIRECTORIES

5. Expand the **Advanced tss\_directories-config** section.
6. Specify the installation location in the text box. The default location is `/opt/trillium_directories`.
7. Click **Next**. The Configure Identities page opens.
8. Click **Next**. The Review page opens.
9. On the **Review** page, verify the information and click **Deploy**.
10. Enter your Admin principal name and Admin password and click **Save**. The Install, Start, and Test page opens.

*ⓘ You must append @WINKERB2.ORG to your user name for the Admin principal name.*
11. On the **Install, Start and Test** page, wait for the TSS Postal Directories services to be successfully installed and started on each node.
12. When services are complete, click **Next**. The Summary page opens.
13. Click **Complete**.

## Uninstalling Trillium Quality for Big Data

Follow the procedure below to uninstall Trillium Quality for Big Data from where it is installed. You must uninstall the Big Data solution and GAV tables separately.

### ► To uninstall Trillium Quality for Big Data from Ambari Client Manager

1. Ensure that no Trillium Quality for Big Data jobs are running.
2. Log in to the Ambari web interface and click **TSQBD** in the left navigation pane.
3. Select **Dashboard > TSQBD**. The Host page opens with a list of active nodes displayed.
4. Click one of the nodes where you want to uninstall Trillium Quality for Big Data from. The Summary page opens.
5. Click the **Installed** drop-down list in the **Components** section and select **TSQBD > UNINSTALL**.

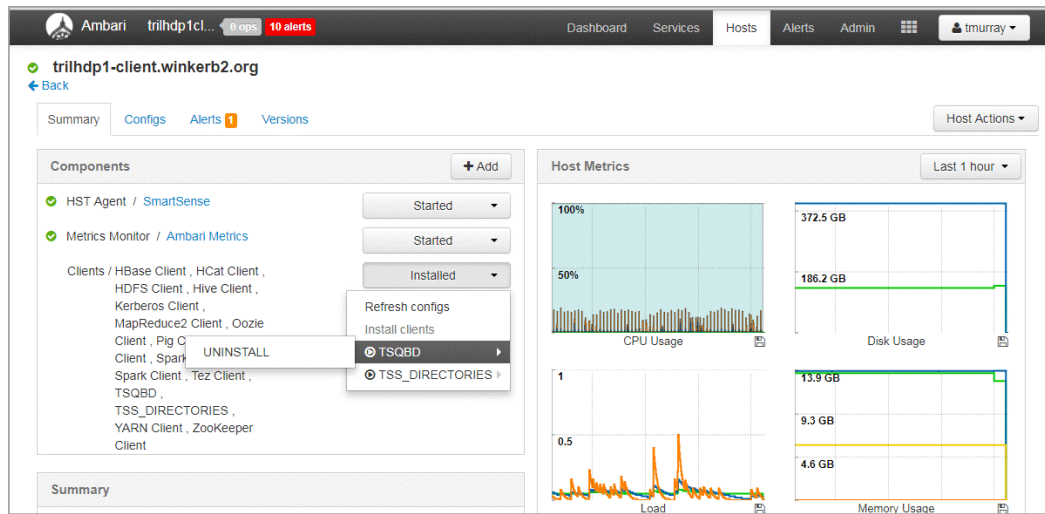


Figure 2-3 Uninstall TSQBD

6. Click **OK** in the confirmation box.
7. When uninstall is complete, click **OK** to return to the Summary page.
8. Repeat Step 4 through 7 for each node to uninstall Trillium Quality for Big Data.
9. Repeat Step 2 through 8 for **TSS\_DIRECTORIES** to uninstall the GAV tables.

► **To manually uninstall Trillium Quality for Big Data**

You can also uninstall Trillium Quality for Big Data manually from the command line.

1. Ensure that no Trillium Quality for Big Data jobs are running.
2. Verify the name of the RPM packages to uninstall:

```
rpm -qa | grep tsqbd
rpm -qa | grep tss_directories
```

3. Run the following command with sudo or root privileges using the erase option (-e):

```
rpm -e tsqbd
rpm -e tss_directories
```

**Examples**

```
sudo rpm -e tsqbd
sudo rpm -e tss_directories
```

## CHAPTER 3

# *Installing Trillium Quality for Big Data - Cloudera Manager Parcels*

This chapter describes how to install and uninstall Trillium Quality for Big Data using Cloudera Manager parcels.

Trillium Quality for Big Data is provided as two Cloudera parcels, one for the Big Data solution which includes Trillium Quality and DMX-h, and one for Trillium Global Address Verification (GAV) tables. For each parcel, you will need to download the parcel to your Cloudera parcel repository, and then distribute and activate the parcel to the Hadoop cluster using Cloudera Manager.

The Big Data solution must be installed on *all nodes* in the cluster, as well as on an *edge node* that has access to the cluster. GAV tables can be installed on all nodes in the cluster or only on one node if distributed cache is utilized. See [Chapter 6, “Using Distributed Cache for Trillium Global Address Verification \(GAV\) Tables”](#) for more information about distributed cache.

- ① *Trillium Global Address Verification (GAV) tables were formerly named TSS Postal Directories. GAV tables include the country-specific postal tables and system resource files such as code\_pages, general\_resources, and parser\_rules.*

## Downloading Parcels

The parcel should be downloaded to a local parcel repository on the Cloudera Manager Server, where it is available for distribution to the nodes in the clusters managed by this Cloudera Manager Server.

Execute the following procedure on the machine where Cloudera Manager resides:

► **To download parcels**

1. Download the Trillium Quality for Big Data installation parcels from TSS Download Center (File Portal) FTP website. The parcels include the following self-extracting shrink-wrap executables:

- Trillium Quality for Big Data (Trillium Quality and DMX-h)

`tsqbd-version-linuxOS.parcel.bin`

- GAV tables

`tss_directories-client_name.date-linuxOS.parcel.bin`

**Where:**

- *version* = the Trillium Quality for Big Data version number.
- *linuxOS* = Linux OS version. For example, el7 for RHEL7.x.
- *client\_name.date* = Client name and the date (YYMM) when the tables were built.

2. Run one of the shrink-wrap executables.

**Examples**

`./tsqbd-15.8.2.17326.bd-el7.parcel.bin`

`./tss_directories-trillium.1711-el7.parcel.bin`


3. Read and accept the Software License Agreement.
4. When prompted, specify the location where you want the `.parcel` and `.parcel.sha` files to be extracted, or press **Enter** to select the default location.
5. Using sudo privileges, copy the `.parcel`, `.parcel.sha`, and `manifest.json` files to the default Cloudera parcel repository:  
`/opt/cloudera/parcel-repo`
6. Repeat Step 2 through 5 for the other shrink-wrap executable. Two parcels, `tsqbd` and `tss_directories`, are now downloaded to your Cloudera parcel repository.

## Installing Trillium Quality for Big Data

Install the Trillium Quality for Big Data on all nodes in the cluster, as well as on an *edge node* that has access to the cluster.

Install the GAV tables on all nodes in the cluster or only on one node if distributed cache is utilized. See [Chapter 6, "Using Distributed Cache for Trillium Global Address Verification \(GAV\) Tables"](#) for more information about distributed cache.

### ► To install Trillium Quality for Big Data and GAV tables


1. Log into Cloudera Manager and click the parcels icon (  ) on the navigation bar. The **tsqbd** and **tss\_directories** parcels display on the list. If you do not see those parcels, click **Check for New Parcels** on the navigation bar.
2. Click **Distribute** for the tsqbd parcel. The status changes to "Distributing" and the parcel is distributed to the cluster managed by the Server.
3. After distribution, click **Activate** and click **OK** in the confirmation message. The status changes to "Activating" and the parcel becomes active and ready to be used. After activation, the status shows "Distributed, Activated".
4. Repeat Step 2 and 3 for the tss\_directories parcel.

After you have installed Trillium Quality for Big Data, the default `/opt/trillium` directory is created on each node to store the Trillium Quality software, and the default `/opt/trillium_directories` directory is created to store the GAV tables and system resource files, such as `code_pages`, `general_resources`, and `parser_rules`. The DMX-h modules are stored in the default `/opt/trillium/dmexpress` directory.

## Uninstalling Trillium Quality for Big Data

Follow the procedure below to uninstall Trillium Quality for Big Data from where it is installed. You must uninstall the Big Data solution and GAV tables separately.

### ► To uninstall Trillium Quality for Big Data

1. Ensure that no Trillium Quality for Big Data jobs are running.
2. Log into Cloudera Manager and click the parcels icon (  ) on the navigation bar.
3. Click **Deactivate** for the tsqbd parcel. At the confirmation message, select **Restart** or **Deactivate Only** and click **OK**. The parcel component is deactivated and the status changes to "Distributed".

4. After deactivation, click **Remove from Hosts** and click **OK** in the confirmation message. The parcel component is removed from the cluster.
5. After removal, select **Delete** and click **OK** in the confirmation message. The parcel component is removed from the parcel repository.
6. Repeat Step 3 through 5 for the tss\_directories parcel.

## CHAPTER 4

# Installing Trillium Quality for Big Data - Amazon EMR

This chapter describes how to install and uninstall Trillium Quality for Big Data for Amazon EMR (Elastic MapReduce) clusters.

This chapter contains the following sections:

- [Installing Trillium Quality for Big Data on New Clusters on page 20](#)
- [Installing Trillium Quality for Big Data on Existing Clusters on page 23](#)
- [Uninstalling Trillium Quality for Big Data on page 28](#)

Trillium Quality for Big Data is provided as two RPM packages, one for the Big Data solution which includes Trillium Quality and DMX-h, and one for Trillium Global Address Verification (GAV) tables. You will need to download the packages to your system in order to install both packages on the EMR cluster.

The Big Data solution must be installed on *all nodes* in the EMR cluster, as well as on an *edge node* that has access to the cluster. GAV tables can be installed on all nodes in the EMR cluster or only on one node if distributed cache is utilized. See [Chapter 6, "Using Distributed Cache for Trillium Global Address Verification \(GAV\) Tables"](#) for more information about distributed cache.

① *Trillium Global Address Verification (GAV) tables were formerly named TSS Postal Directories. GAV tables include the country-specific postal tables and system resource files such as code\_pages, general\_resources, and parser\_rules.*

For more information about RPM packages for Trillium Quality for Big Data and GAV tables, see [Chapter 5, "Installing Trillium Quality for Big Data - RPM"](#).

## Installing Trillium Quality for Big Data on New Clusters

This section contains instructions for installing Trillium Quality for Big Data on a newly created EMR cluster. You can use a bootstrap action at the creation of a cluster to install Trillium Quality for Big Data.

### ► To install the Trillium Quality for Big Data on a new cluster

1. Download the Trillium Quality for Big Data RPM packages from TSS Download Center (File Portal) FTP website.

- Trillium Quality for Big Data (Trillium Quality and DMX-h)

```
tsqbd-version-1.x86_64.rpm
```

#### Example

```
tsqbd-15.8.2.18144.bd-1.x86_64.rpm
```

- GAV tables

```
tss_directories-client_name.date-1.x86_64.rpm
```

#### Example

```
tss_directories-US.1805-1.x86_64.rpm
```

#### Where:

- *version* = Trillium Quality for Big Data version number.
- *client\_name.date* = Client name and the date (YYMM) when the tables were built.

2. Upload the RPM packages into your S3 bucket.
3. Create a bootstrap script using the samples below to install the RPM packages.

### To install to the default location:

This script will install Trillium Quality for Big Data to the default `/opt/trillium` directory and the GAV tables to the default `/opt/trillium_directories` directory.

Sample Script: `tsqbd_15.8_emr_bootstrap.sh`

```
#!/bin/sh
install_rpm()
{
    echo Installing $1 from ${S3_LOCATION_RPM}/$1
    if [ ! -f $1 ]; then
        aws s3 cp ${S3_LOCATION_RPM}/$1 $1
    fi
    if [ -z "$2" ] ; then
```

```
        sudo rpm -i $1
    else
        sudo rpm -i $1 --prefix $2
    fi
}
echo s3 location of RPM
if [ -z "$1" ]; then
    S3_LOCATION_RPM=s3://testbucket/tsqinstall
else
    S3_LOCATION_RPM=s3://"${1}"
fi
version="$2"
if [ -z "$version" ]; then
    version=15.8.1.18144.bd-1
fi
tables_version="$3"
if [ -z "$tables_version" ]; then
    tables_version=US.1805-1
fi
cd $HOME
TEMP_DIR=install
echo S3_LOCATION_RPM = $S3_LOCATION_RPM
mkdir -p $TEMP_DIR
cd $TEMP_DIR
install_rpm tsqbd-${version}.x86_64.rpm
#/opt/trillium
install_rpm tss_directories-${tables_version}.x86_64.rpm
#/opt/trillium_directories
```

### **To install to a user defined location:**

This script will install Trillium Quality for Big Data to the custom `/opt/trillium_custom` directory and the GAV tables to the custom `/opt/custom_tables` directory, and replace the table path in the config files, `EDQ.ini` and `gaserver.ini` located in `/opt/trillium_custom/conf`.

Sample Script: `tsqbd_15.8_emr_bootstrap_custom.sh`

```
#!/bin/sh
install_rpm()
{
    echo Installing $1 from ${S3_LOCATION_RPM}/${1}
    if [ ! -f $1 ]; then
        aws s3 cp ${S3_LOCATION_RPM}/${1} $1
    fi
    if [ -z "$2" ]; then
        sudo rpm -i $1
    else
        sudo rpm -i $1 --prefix $2
    fi
}
```

```
uninstall_rpm()
{
    echo UnInstalling $1
    if [ -n "$1" ]; then
        sudo rpm -e $1
    fi
}
echo s3 location of RPM
if [ -z "$1" ]; then
    S3_LOCATION_RPM=s3://testbucket/tsqinstall
else
    S3_LOCATION_RPM=s3://" $1"
fi
version="$2"
if [ -z "$version" ]; then
    version=15.8.1.18280.bd-1
fi
tables_version="$3"
if [ -z "$tables_version" ]; then
    tables_version=Custom.1809-1
fi
cd $HOME
TEMP_DIR=install
echo S3_LOCATION_RPM = $S3_LOCATION_RPM
mkdir -p $TEMP_DIR
cd $TEMP_DIR
uninstall_rpm tsqbd
uninstall_rpm tss_directories
install_rpm tsqbd-${version}.x86_64.rpm /opt/trillium_custom
install_rpm tss_directories-${tables_version}.x86_64.rpm /opt/
custom_tables

/opt/trillium_custom/bin/set_table_path.sh /opt/custom_tables
#script delivered with Trillium Quality that replaces default table
path with the user-defined table path, /opt/custom_tables, in the
Trillium config files, EDQ.ini and gaserver.ini
```

4. Upload the bootstrap script into the S3 bucket.
5. Configure a bootstrap action in the creation of an EMR cluster. This bootstrap action will call the bootstrap script that installs the Big Data solution and GAV tables.
6. Prepare the root volume size on each node of the EMR cluster to have enough space on the / partition for the Big Data solution, and the GAV tables if installed on each node.  
  
*ⓘ The Big Data solution requires approximately 3 GB. The size of GAV tables varies depending on the country.*
7. Create an EMR cluster that meet the requirements described above.

## Installing Trillium Quality for Big Data on Existing Clusters

This section contains instructions for installing Trillium Quality for Big Data on an existing EMR cluster. You can install Trillium Quality for Big Data on your existing or running cluster using AWS Systems Manager Run Command.

### Prerequisites

Note the following prerequisites:

- Knowledge of AWS Systems Manager is required. Verify that your system meets AWS Systems Manager's prerequisites.
- Ensure that all EC2 instances of the EMR cluster on which you are installing Trillium Quality for Big Data are *managed instances*. A *managed instance* is any Amazon EC2 instance that has been configured for AWS Systems Manager.
- Verify that your instances meet the minimum requirements for Run Command.
- Verify that the EMR Cluster administrator meets the requirements to execute Run Command.

① [See the AWS Systems Manager documentation for details of these guidelines.](#)

## Installing Trillium Quality for Big Data

### ► To install the Trillium Quality for Big Data on a running cluster

1. Download the Trillium Quality for Big Data RPM packages from TSS Download Center (File Portal) FTP website.

- Trillium Quality for Big Data (Trillium Quality and DMX-h)

`tsqbd-version-1.x86_64.rpm`

#### **Example**

`tsqbd-15.8.2.18144.bd-1.x86_64.rpm`

- GAV tables

`tss_directories-client_name.date-1.x86_64.rpm`

#### **Example**

`tss_directories-US.1805-1.x86_64.rpm`

#### **Where:**

- `version` = Trillium Quality for Big Data version number.

- `client_name.date` = Client name and the date (YYMM) when the tables were built.
2. Upload the RPM packages into your S3 bucket.
  3. From AWS Systems Manager, select **Shared Resources > Managed Instances**. Ensure that all EC2 instances of the EMR cluster on which you are installing Trillium Quality for Big Data are listed.

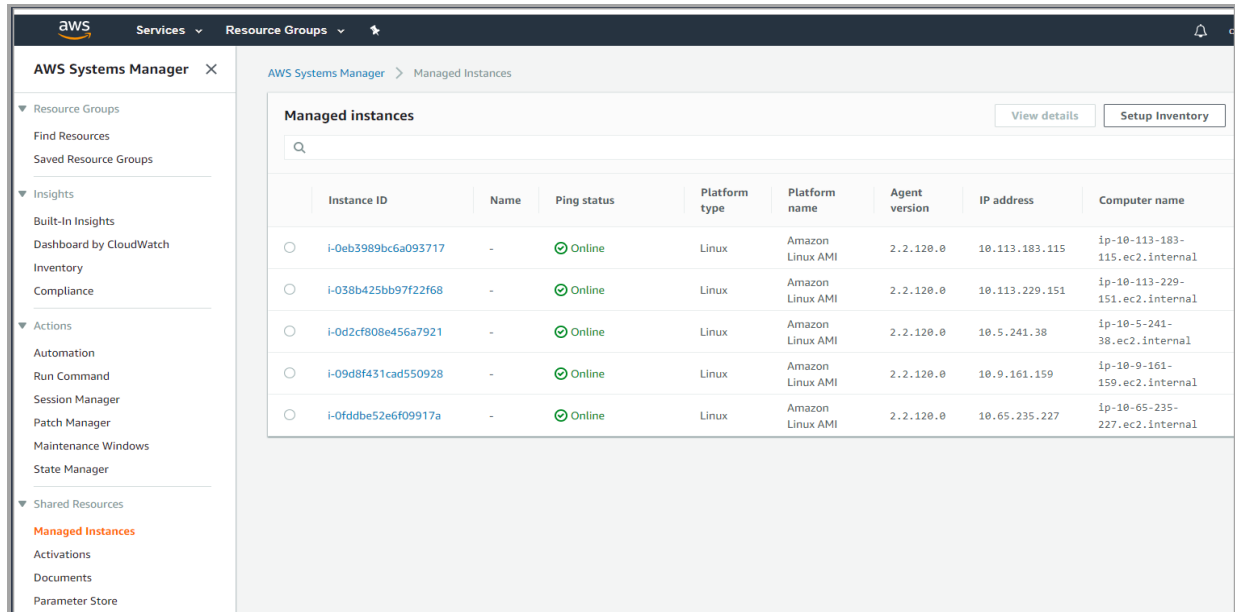


Figure 4-1 Managed Instances

4. Click **Run Command** at the top right corner to start the Run Command process.

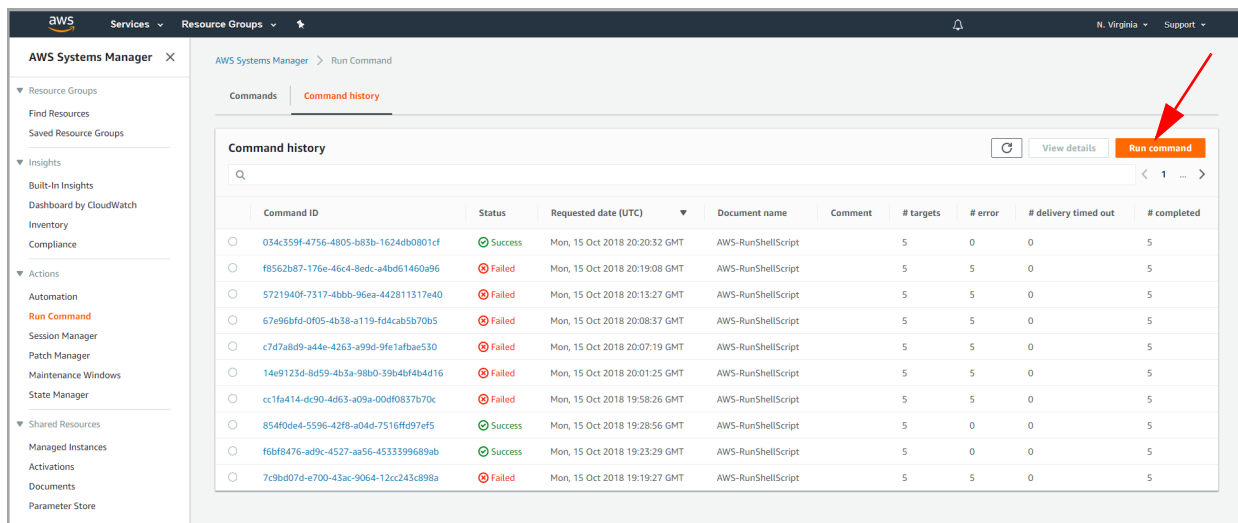


Figure 4-2 Run Command

5. Select **AWS-RunShellScript** (Linux) in the Command document section.

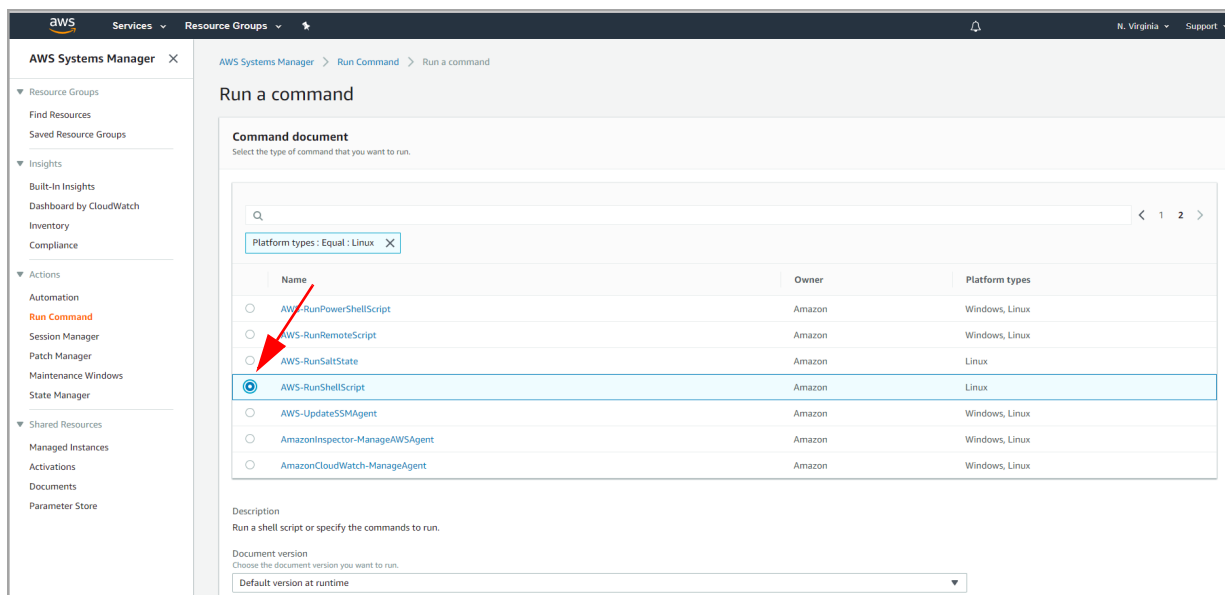


Figure 4-3 Command Document

6. Enter your command in the Command parameters section.

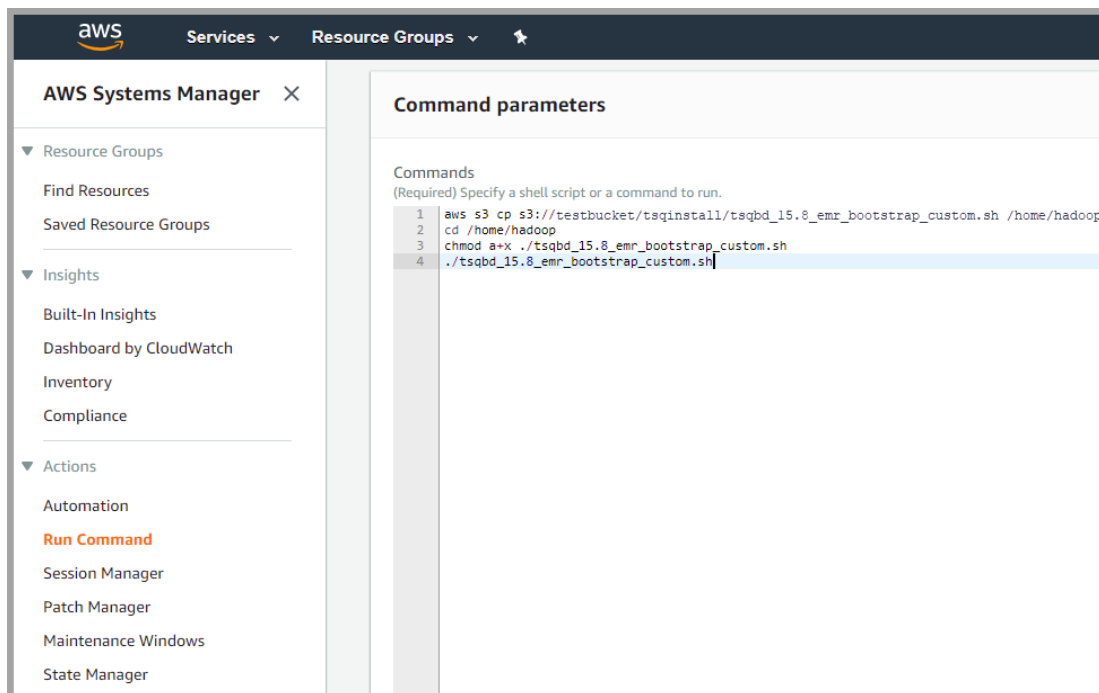


Figure 4-4 Command Parameters

Sample Command:

```
aws s3 cp s3://testbucket/tsqinstall/tsqbd_15.8_emr_bootstrap_
custom.sh /home/hadoop
cd /home/hadoop
chmod a+x ./tsqbd_15.8_emr_bootstrap_custom.sh
./tsqbd_15.8_emr_bootstrap_custom.sh
```

The script ./tsqbd\_15.8\_emr\_bootstrap\_custom.sh is the sample bootstrap in the previous section (see [page 21](#)). You can use shell commands instead. For example, you can use the rpm commands directly to install the rpm packages.

You can also perform the same action by entering the AWS Command Line Interface (CLI) command listed in the CLI command box at the bottom of the page from the AWS CLI client that is configured to access AWS Systems Manager Run Command.

- 7. Select your target instances in the Targets section.

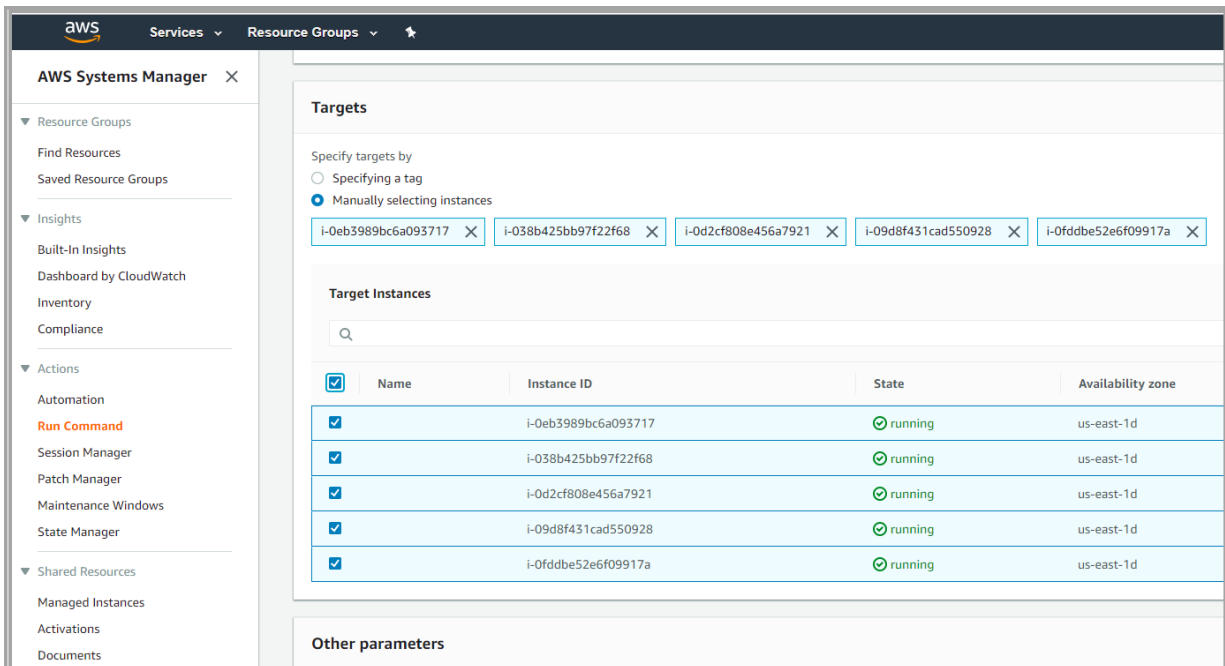


Figure 4-5 Target Instances

8. Click **Run** at the bottom right corner to run the command.

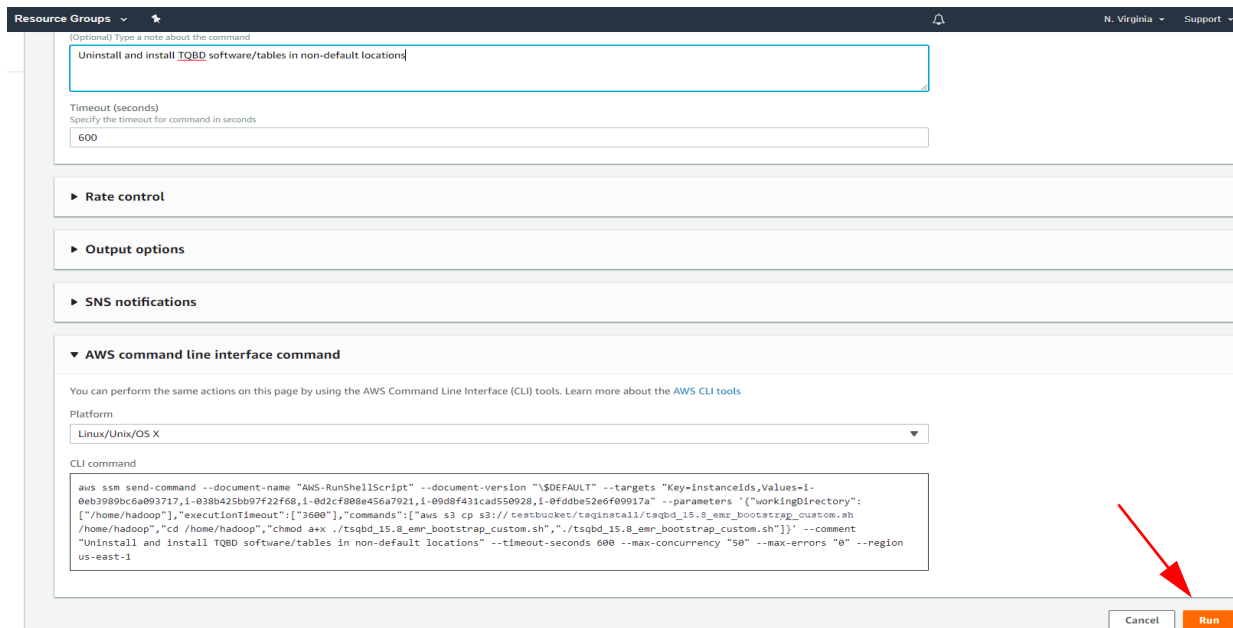


Figure 4-6 Execute Command

9. Check the Targets and outputs section for the status. The Status column should display "In Progress" for each instance. When the installation completes, the status will change to "Success".

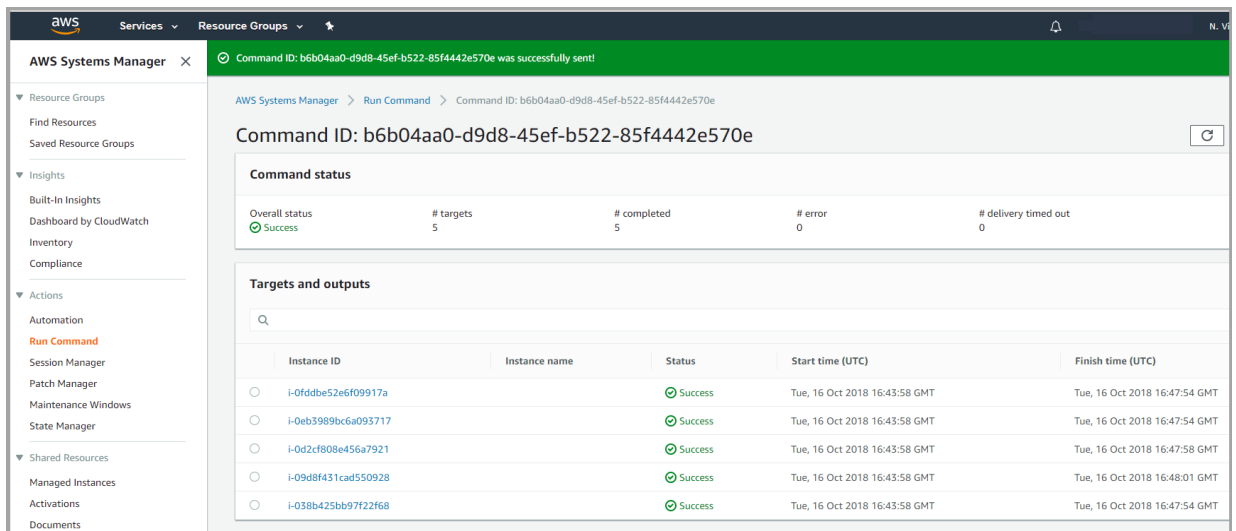


Figure 4-7 Run Command Success

## Uninstalling Trillium Quality for Big Data

When the EMR cluster is terminated, Trillium Quality for Big Data and GAV tables are automatically removed.

If the cluster is not terminated, Trillium Quality for Big Data and GAV tables can be uninstalled with the `rpm -e` command from AWS Systems Manager Run Command.

For more information, see [Installing Trillium Quality for Big Data on Existing Clusters on page 23](#) and [Chapter 5, "Installing Trillium Quality for Big Data - RPM"](#).

## CHAPTER 5

# Installing Trillium Quality for Big Data - RPM

This chapter describes how to install and uninstall Trillium Quality for Big Data RPM packages.

Trillium Quality for Big Data is provided as two RPM packages, one for the Big Data solution which includes Trillium Quality and DMX-h, and one for Trillium Global Address Verification (GAV) tables. You will need to download the packages to your system and then install each package separately.

The Big Data solution must be installed on *all nodes* in the cluster, as well as on an *edge node* that has access to the cluster. GAV tables can be installed on all nodes in the cluster or only on one node if distributed cache is utilized. See [Chapter 6, “Using Distributed Cache for Trillium Global Address Verification \(GAV\) Tables”](#) for more information about distributed cache.

- ① *Trillium Global Address Verification (GAV) tables were formerly named TSS Postal Directories. GAV tables include the country-specific postal tables and system resource files such as code\_pages, general\_resources, and parser\_rules.*
- ① *Supported platforms are Red Hat Enterprise Linux 6 (RHEL6.x) and Red Hat Enterprise Linux 7 (RHEL7.x).*

## Installing Trillium Quality for Big Data

This section contains instructions for installing Trillium Quality for Big Data RPM packages.

### ► To install the Trillium Quality for Big Data RPMs

1. Download the Trillium Quality for Big Data RPM packages from TSS Download Center (File Portal) FTP website.
  - Trillium Quality for Big Data (Trillium Quality and DMX-h)

```
tsqbd-version-1.x86_64.rpm
```

**Example**

```
tsqbd-15.8.2.18144.bd-1.x86_64.rpm
```

■ GAV tables

```
tss_directories-client_name.date-1.x86_64.rpm
```

**Example**

```
tss_directories-US.1805-1.x86_64.rpm
```

**Where:**

- *version* = Trillium Quality for Big Data version number.
- *client\_name.date* = Client name and the date (YYMM) when the tables were built.

2. Copy the RPM packages locally into your system.
3. Run the following commands with sudo privileges:

■ Trillium Quality for Big Data

**To install to the default location (/opt/trillium):**

```
sudo rpm -i tsqbd-version-1.x86_64.rpm
```

**Example**

```
sudo rpm -i tsqbd-15.8.2.18137.bd-1.x86_64.rpm
```

**To install to a user defined location:**

```
sudo rpm -i --prefix <user defined location> tsqbd-version-1.x86_64.rpm
```

**Example**

```
sudo rpm -i --prefix /home/test/trillium tsqbd-15.8.2.18137.bd-1.x86_64.rpm
```

■ GAV tables

**To install to the default location (/opt/trillium\_directories):**

```
sudo rpm -i tss_directories-client_name.date-1.x86_64.rpm
```

**Example**

```
sudo rpm -i tss_directories-trillium.1803-1.x86_64.rpm
```

**To install to a user defined location:**

```
sudo rpm -i --prefix <user defined location> tss_
```

```
directories-client_name.date-1.x86_64.rpm
```

### Example

```
sudo rpm -i --prefix /home/test/trillium_directories tss_
directories-trillium.1803-1.x86_64.rpm
```

After you have installed Trillium Quality for Big Data, the default `/opt/trillium` or `<user defined>` directory is created to store the Trillium Quality software, and the default `/opt/trillium_directories` or `<user defined>` directory is created to store the GAV tables. The DMX-h modules are stored in the `<installed trillium path>/dmexpress` directory.

If you have installed GAV tables to a user defined location that is non-default, you will need to update the following configuration files in the `conf` directory of Trillium Quality for Big Data. Replace the default `/opt/trillium_directories` in the files with your user defined installation path for the GAV tables.

- `EDQ.ini`
- `gaserver.ini`

① *You will need to update `gaserver.ini` if you have installed Global GAV tables and/or Geocoding (Lat/Long) tables.*

## Uninstalling Trillium Quality for Big Data

This section contains instructions for uninstalling Trillium Quality for Big Data RPM packages.

### ► To uninstall the Trillium Quality for Big Data RPMs

Follow the procedure below to uninstall Trillium Quality for Big Data from where it is installed. You must uninstall the Big Data solution and GAV tables separately.

① *The uninstallation procedure is the same regardless of your installation location.*

1. Ensure that no Trillium Quality for Big Data jobs are running.
2. Verify the name of the RPM packages to uninstall:

```
rpm -qa | grep tsqbd
rpm -qa | grep tss_directories
```

3. Run the following command with `sudo` or root privileges using the erase option (`-e`):

```
rpm -e tsqbd
rpm -e tss_directories
```

### Examples

```
sudo rpm -e tsqbd
```

```
sudo rpm -e tss_directories
```

## CHAPTER 6

# Using Distributed Cache for Trillium Global Address Verification (GAV) Tables

This chapter describes how to use distributed cache for accessing Trillium Global Address Verification (GAV) tables in a Hadoop cluster. Instead of installing GAV tables on *all nodes* in the cluster, you have the option to install the tables on only an *edge node* and access them at runtime through distributed cache. As the size of the tables could be significantly large depending on the country/package, this option gives you the benefit of space optimization and allows you to efficiently manage the Hadoop resources.

- ① *Trillium Global Address Verification (GAV) tables were formerly named TSS Postal Directories. GAV tables include the country-specific postal tables and system resource files such as code\_pages, general\_resources, and parser\_rules.*

## Installing GAV Tables for Distributed Cache

To utilize distributed cache, you must install GAV tables on the edge node of the cluster, and then copy the tables to HDFS by the `copy2HDFS.sh` script. The script will update the project's `EDQ.ini` file with the HDFS path information. When you run the projects, Trillium Quality for Big Data will use this information and access the necessary tables by distributed cache.

## Guidelines

Note the following guidelines:

- The installation procedure is the same, whether you are installing for the first time or installing an update. Whenever you receive updated GAV tables, you must repeat this process.

- You must have write permissions to the destination in HDFS and sudo privileges on the local node. By default the `EDQ.ini` file is writable only by root. It is recommended that you create a Linux Group to adjust your security/user management.
- At runtime, the `tsqbd.sh` script will redefine `TS_CONFIG` to point to a converted project specific `Config64.tbl`, called `config_dtl_mapreduce.tbl` if the DMX framework is MAPREDUCE, or `config_dtl_spark.tbl` if the DMX framework is SPARK.

► **To install/update GAV tables in HDFS**

1. Using Ambari, Cloudera, or RPM, download GAV tables or updates to the edge node of the cluster. GAV tables are provided by Apache Ambari service packages, Cloudera Manager parcels, or RPMs through Trillium Software's FTP website. See Chapter 2, 3, 4 and 5 for how to download and install GAV tables.
2. Run the `copy2HDFS.sh` script to copy the installed GAV tables to HDFS. This script is located in `/opt/trillium/bin`.

**Syntax**

```
./copy2HDFS.sh [--help] [--debug] [--force] -s|--src <local source directory> -d|--dest <hdfs destination directory>
```

**where:**

`--help`: Optional. Displays usage information.

`--debug`: Optional. Shows verbose debug information.

`--force`: Optional. Overwrites files in HDFS destination.

`-s|--src`: Required. Source directory (for example, `/opt/trillium_directories`)

`-d|--dest`: Required. Destination directory in HDFS (for example, `/user/qa/trillium_directories`)

**Example**

```
./copy2HDFS.sh -s /opt/trillium_directories -d /opt/trillium/trillium_directories --force
```

The script will update the `EDQ.ini` file (`/opt/trillium/conf`) on the edge node with the following distributed cache parameter specifying the HDFS path: `HDFS_TABLES=""<HDFS path for tables>""`.

**Example**

```
cat /opt/trillium/conf/EDQ.ini:
```

```
TABLES = "/opt/trillium_directories"  
HDFS_TABLES="/opt/trillium/trillium_directories"
```

① *No change will be made to the EDQ.ini file installed with the software.*

3. Verify that `EDQ.ini` on the edge node has the specified location for the `HDFS_TABLES` parameter. When you run the projects, Trillium Quality for Big Data will use this path and access the necessary GAV tables in HDFS.

① *See the Trillium Quality for Big Data Developer's Guide for information on running projects with distributed cache.*

## Notes for Global GAV Tables/Geocoding Tables

For Global GAV tables and Geocoding (Lat/Long) tables, you can specify the `GA_TABLES=${TABLES}` setting in the project's `Config64.tbl` file, identifying the path of the distributed cache being used. This allows the Global Postal Matcher and Geocoder to redefine values in the `gaserver.ini` file and store in a project specific location.

Note the following:

- Substitutions are only performed when `GA_TABLES` is defined in `Config64.tbl`, otherwise the default `conf/gaserver.ini` is used.
- When `GA_TABLES` is defined in `Config64.tbl`, the paths will be substituted in `conf/gaserver.ini` and written to the project's `SETTINGS` directory:

### Example

```
CountryDataDirectory=/opt/trillium_directories/gatables/  
datastore/rdata  
KnowledgeBaseDirectory=/opt/trillium_directories/gatables/  
datastore/kbase  
LicenseDirectory=/opt/trillium_directories/gatables/license
```

- The project's `SETTINGS` directory will be passed into the `libgaserver StartEngine()` function call to identify where the `libgaserver` will find the `gaserver.ini` file.
- When distributed cache is enabled and the project is run, the logging information will be written in the Hadoop job history log under the **ts\_02\_gaserver.log** section. The "ts\_02\_" prefix is a naming convention for similar log information such as parsing exceptions ("ts\_01\_") and debug ("ts\_99\_"). When the project is run using the local framework on the edge node, the logging information will be written to the `settings/gaserver.log` file.
- The `conf/gaserver.ini` file remains unchanged.

## CHAPTER 7

# Setting Up SSH

This chapter describes how to set up SSH Secure Shell for use with Trillium Quality for Big Data. Trillium Quality for Big Data employs security with *SSH Secure Shell*. Public key and private key are required to deploy a Trillium Quality project to an edge node. A passphrase is optional.

When security errors occur, Trillium Quality for Big Data displays an error message. See [Error Messages on page 40](#) for more information.

Public and private keys are included in the TSS Windows Server or Client installation and are placed in the following directories:

- Windows Server  
`C:\Program Files\Trillium Software\MBSW\15\tsq\default_keys`
- Client  
`C:\Program Files\Common Files\Trillium Software\15\default_keys`

They do not contain a passphrase and do not require you to create one. You can use these keys for testing.

## Setting up Security

Setting up security involves the following steps:

1. System administrator sets up SSH on the edge node.
2. Generate public and private keys on the edge node where the project is deployed.

There are three tools available to set up security:

- Putty (a free UI tool available on the Internet)
  - CGWIN (a free SSH toolkit SDK available on the Internet)
  - SSH command line
3. (Optional) Create a passphrase.

4. Append the public key to the `authorized_keys` file.
5. Insert the command script directive before the public key entry inside the `authorized_keys` file.

### ► To set up SSH on edge node

Follow this step if a system administrator or authorized user needs to set up SSH to work with a user account on the edge node.

Create a `.ssh` directory for a Trillium user or authorized user and set necessary permissions.

### Example

```
cd ~ (home directory or Trillium user directory)
mkdir .ssh
chmod 700 .ssh
```

### ► To generate public and private keys using the Putty tool

You can use Putty or CGWIN to generate public and private keys. The Putty tool method is described as an example.

1. Start **puttygen.exe**. The PuTTYKey Generator window opens.
2. In Parameters, select **SSH-2 RSA**.
3. In Actions, click **Generate**. The key will be generated.

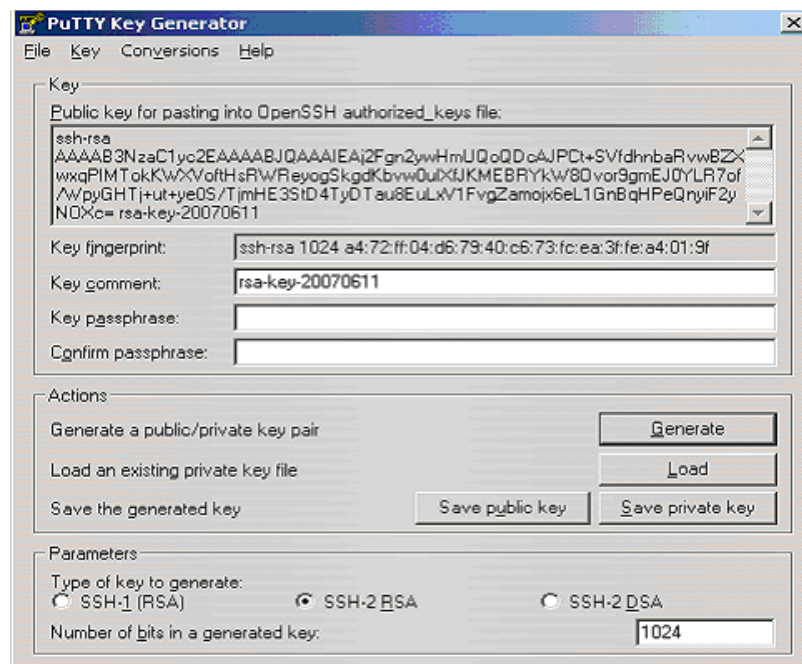


Figure 7-1 Putty Key Generator

4. In **Key passphrase**, enter your key passphrase and confirm it.

5. Select **Conversions > Export OpenSSH key**. Save the file.
6. Highlight the entire text of the public key and copy it to the clipboard.
7. Open a text editor such as Notepad and paste the key.
8. Save the public key file using the same name as the private key file with a file extension of **.pub**. This helps you tell which files are paired.

If you are using a special TSS user created by the administrator, give this public key to the administrators and have them perform the following steps.

If you are using your own user id, you can do the following steps by yourself to your own user account.

9. Append the public key to the `authorized_keys` file and set necessary permissions.

The `~user` directory on Linux is configured by the administrator to have a subdirectory of `.ssh`. This directory holds the `authorized_keys` file used by OpenSSH.

### Example

```
cd ~/.ssh(the .ssh directory in home directory)
cat tqbdkey.pub >> authorized_keys
chmod 644 authorized_keys
```

10. Open the `authorized_keys` file and add the following command before the word "ssh-rsa" which is in front of the key. The command points to the `trilsecure.sh` script in the `bin` directory of Trillium Quality.

```
command="/<trillium_bin_location>/trilsecure.sh"
```

### Example

Before:

```
ssh-rsa AAB3NzaC1yc2EAAAmx5ZFqxa....=mlamand@tril01
```

After:

```
command="/opt/trillium/bin/trilsecure.sh" ssh-rsa
AAB3NzaC1yc2EAAAmx5ZFqxa....=mlamand@tril01
```

❗ *If you are using Tectia, see Tectia's documentation.*

Rather than using Putty or CGWIN, you can use the SSH command line interface invoking the `ssh-keygen` tool. You could do this step in your user account, or a system administrator could do this step in a designated TSS user account.

► **To generate public and private keys using the SSH command line interface**

1. Change directory to **.ssh**
2. Using the following interactive script command, generate public and private keys with passphrase.

```
ssh-keygen -t rsa
```

**Example**

```
Enter file to save key: tqbdkey
```

```
Enter Passphrase: ****
```

```
Re-Enter Passphrase: ****
```

The following key files are generated in this example:

- `tqbdkey.pub`. The public key.
- `tqbdkey`. The private key.

Record the passphrase you created for this key. In the TSQ Batch to Big Data Deployment Manager, you will need to specify both the private key file and its passphrase.

3. Copy the *private key* to the remote client machine. The private key is useless without a passphrase. Only you know that passphrase.
4. Append the public key file to the `authorized_keys` file and set necessary permissions.

**Example**

```
cd ~/.ssh(the .ssh directory in your home directory)
```

```
cat tqbdkey.pub >> authorized_keys
```

```
chmod 644 authorized_keys (necessary permissions)
```

5. Open the `authorized_keys` file and add the following command before the word "ssh-rsa" which is in front of the key. The command points to the `trilsecure.sh` script in the **bin** directory of TSS on your server.

```
command="/<trillium_bin_location>/trilsecure.sh"
```

**Example**

Before:

```
ssh-rsa AAB3NzaC1yc2EAAAQx5ZFqxa....=mlamand@tril01
```

After:

```
command="/opt/trillium/bin/trilsecure.sh" ssh-rsa
AAB3NzaC1yc2EAAAQx5ZFqxa....=mlamand@tril01
```

## Brief Overview of the Process

The following is a brief description of setting up -SSH for those who are familiar with the process. Refer to the previous sections of this chapter for more detailed information on setting up SSH.

After installing Trillium Quality for Big Data, enter the following commands to set up SSH on your user account:

```
cd ~
mkdir .ssh
chmod 700 .ssh
cd .ssh
touch authorized_keys
chmod 644 authorized_keys
cat public_key >> authorized_keys.

vi authorized_keys
[ insert in front of ssh-rsa the following command:
command="/install_path/trillium/bin/trilsecure.sh" ssh-rsa
]
```

Public\_key is the one delivered with the installation, located in the following directories, or the one you created:

- Windows Server  
C:\Program Files\Trillium Software\MBSW\15\tsq\default\_keys
- Client  
C:\Program Files\Common Files\Trillium Software\15\default\_keys

## Error Messages

When security errors occur, Trillium Quality for Big Data displays an error message. Error messages are described in Table 7-1.

**Table 7-1 Error Messages**

Message	Description
Invalid private key file	The wrong private key file was used.
Permission denied (publickey)	The private key file used did not match the public key.

## CHAPTER 8

# Troubleshooting

This chapter discusses potential problems you may encounter during Trillium Quality for Big Data installation and the steps to resolve them.

## Ambari Server Fails to Start

**Problem:**

Startup of Ambari server fails when you try to install a large package such as GAV tables for World Wide on your cluster. The following or similar error is generated:

```
Organizing resource files at /var/lib/ambari-server/resources...  
ERROR: Exiting with exit code -1.  
REASON: Can not organize resource files at /var/lib/ambari-server/  
resources: Can not create zip archive of directory /var/lib/  
ambari-server/resources/stacks/HDP/2.4/services/tss_directories_  
WW/package: Zipfile size would require ZIP64 extensions
```

**Cause:**

By default, Ambari server uses a 32-bit zip utility to create a maximum 4 GB of zip files for the resource files. If the zip file for the resource files would exceed 4 GB in size, the 32-bit zip utility would not be able to create it.

This is a known issue with Ambari server.

<https://issues.apache.org/jira/browse/AMBARI-18509>

<https://issues.apache.org/jira/browse/AMBARI-17593>

**Solution:**

To resolve the issue, Ambari server needs to use a 64-bit zip utility to create zip files of more than 4 GB. Perform the following steps as a workaround:

► **To enable Ambari server to use 64-bit zip utility**

1. Login to the Ambari server machine.
2. Go to the following location:  

```
/usr/lib/python2.6/site-packages/ambari_server
```
3. Make a backup of the existing file `resourceFilesKeeper.py`.
4. After making backup, open the file `resourceFilesKeeper.py` for editing.
5. Comment out the following line:  

```
(zf = zipfile.ZipFile(zip_file_path, "w"))
```
6. Add the following line below the commented line:  

```
(zf = zipfile.ZipFile(zip_file_path, "w", allowZip64 = True))
```
7. Save the changes to the file.
8. Start the Ambari server again.

## Ambari Installation Fails to Modify Config Files for GAV Tables

**Problem:**

If GAV tables are installed to a non-default location and the Ambari installation process fails to modify the config files (`EDQ.ini` and `gaserver.ini`) to replace the default path with the specified non-default path, you will need to modify the config files.

**Solution:**

Identify the nodes where the config files (`EDQ.ini` and `gaserver.ini`) do not contain the correct location of the GAV tables, then run the standalone script, `set_table_path.sh` on each such node, to change the install location of the GAV tables. When the script is run, the `EDQ.ini` and `gaserver.ini` files are updated automatically with the new install path.

- ① *The `gaserver.ini` file exists only when you have installed Global GAV tables and/or Geocoding (Lat/Long) tables.*

► **To change the install path of GAV tables**

1. Locate the `set_table_path.sh` script. The script is in the `bin` directory of the installed Trillium Quality for Big Data software, for example, `/opt/trillium/bin`.
2. Run the `set_table_path.sh` script.

### **Syntax**

```
set_table_path.sh <new path>
```

### **Example**

This is a sample command to move the install location to `/user/test/postal`.

```
/opt/trillium/bin/set_table_path.sh /user/test/postal
```

3. Verify that the GAV tables install path is set into the `EDQ.ini` and `gaserver.ini` files on each node where the software is installed.

# Index

## Numerics

64-bit zip utility, 41

## A

Amazon EMR, 19  
Ambari dashboard, 11, 12, 13  
Apache Ambari, 6, 7, 9  
AWS Systems Manager, 23  
AWS-RunShellScript, 25

## B

Batch Deployment Tool, 6  
bootstrap script, 20

## C

Cloudera Manager, 17  
Cloudera parcels, 6, 15  
config\_dtl\_mapreduce.tbl, 34  
config\_dtl\_spark.tbl, 34  
Config64.tbl, 34, 35  
copy2HDFS.sh, 34  
Country Templates, 6  
customer support, 5

## D

definition package, 9  
distributed cache, 33  
DMX-h, 6

## E

EC2 instances, 23  
EDQ.ini, 34

## G

GA\_TABLES, 35  
gaserver.ini, 35  
Geocoding, 35  
Global Postal Tables, 35

## H

HDFS\_TABLES, 34

## J

Java/JRE, 7

## M

managed instances, 23

## R

Repository Server, 6  
RPM packages, 6  
Run Command, 23

## S

support, 5

## T

Trillium™ Control Center, 6  
Trillium™ Global Address Verification (GAV)  
Tables, 6

## U

uninstalling, 13, 17, 28, 31