Trafodion Install Guide

Version 1.3.0, January 2016

Table of Contents

License Statement

Licensed to the Apache Software Foundation (ASF) under one or more contributor license agreements. See the NOTICE file distributed with this work for additional information regarding copyright ownership. The ASF licenses this file to you under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

Version	Date
1.3.0	January, 2016

Chapter 1. About This Document

This guide describes how to provision Trafodion for end-user environments. This install allows you to store and query data using Trafodion, either via Trafodion clients (see Trafodion Client Installation Guide or via application code you write.

If you want to install a Trafodion developer-build environment, then please refer to the Trafodion Contributor Guide for instructions.

1.1. Intended Audience

This guide is intended for:

- **Single-Node Environments**: Typically used when you want to evaluate Trafodion.

- **Cluster (Multi-Node) Environments**: Typically used when you deploy Trafodion for application usage.

> Trafodion can be provisioned on a single-node or multi-node enviroment. Unless specifically noted, the term **cluster** is used to mean both single- and multi-node environments.

The provisioning instructions applies to a diverse set of platforms:

- **Virtual Machines**: Often used for evaluations and Trafodion development.

- **Cloud**: Used for Product Environments as well as for Developer Environments.

- **Bare Metal**: Used for Product Environments as well as for Developer Environments.

> The term **node** is used to represent a computing platform on which operating system, Hadoop, and Trafodion software is running. Unless specifically qualified (bare-metal node, virtual-machin node, or cloud-node), **node** represents a computing platform in your cluster regardless of platform type.

1.2. New and Changed Information

This is a new guide.

1.3. Notation Conventions

This list summarizes the notation conventions for syntax presentation in this manual.

- UPPERCASE LETTERS

  Uppercase letters indicate keywords and reserved words. Type these items exactly as shown. Items not enclosed in brackets are required.

  ```
  SELECT
  ```

- lowercase letters

  Lowercase letters, regardless of font, indicate variable items that you supply. Items not enclosed in brackets are required.

  ```
  file-name
  ```

- [ ] Brackets

  Brackets enclose optional syntax items.

  ```
  DATETIME [start-field TO] end-field
  ```

  A group of items enclosed in brackets is a list from which you can choose one item or none.

  The items in the list can be arranged either vertically, with aligned brackets on each side of the list, or horizontally, enclosed in a pair of brackets and separated by vertical lines.

  For example:

  ```
  DROP SCHEMA schema [CASCADE]
  DROP SCHEMA schema [ CASCADE | RESTRICT ]
  ```

- { } Braces

  Braces enclose required syntax items.

  ```
  FROM { grantee [, grantee ] ... }
  ```

  A group of items enclosed in braces is a list from which you are required to choose one item.

  The items in the list can be arranged either vertically, with aligned braces on each side of the list, or horizontally, enclosed in a pair of braces and separated by vertical lines.

  For example:

  ```
  INTERVAL { start-field TO end-field }
  { single-field }
  INTERVAL { start-field TO end-field | single-field }
  ```

- | Vertical Line

  A vertical line separates alternatives in a horizontal list that is enclosed in brackets or braces.

  ```
  {expression | NULL}
  ```

- … Ellipsis

  An ellipsis immediately following a pair of brackets or braces indicates that you can repeat the enclosed sequence of syntax items any number of times.

  ```
  ATTRIBUTE[S] attribute [, attribute] ...
  {, sql-expression } ...
  ```

  An ellipsis immediately following a single syntax item indicates that you can repeat that syntax item any number of times.

  For example:

```
expression-n ...
```

- Punctuation

  Parentheses, commas, semicolons, and other symbols not previously described must be typed as shown.

  ```
  DAY (datetime-expression)
  @script-file
  ```

  Quotation marks around a symbol such as a bracket or brace indicate the symbol is a required character that you must type as shown.

  For example:

  ```
  "{" module-name [, module-name] ... "}"
  ```

- Item Spacing

  Spaces shown between items are required unless one of the items is a punctuation symbol such as a parenthesis or a comma.

  ```
  DAY (datetime-expression) DAY(datetime-expression)
  ```

  If there is no space between two items, spaces are not permitted. In this example, no spaces are permitted between the period and any other items:

```
myfile.sh
```

- Line Spacing

  If the syntax of a command is too long to fit on a single line, each continuation line is indented three spaces and is separated from the preceding line by a blank line.

  This spacing distinguishes items in a continuation line from items in a vertical list of selections.

```
match-value [NOT] LIKE _pattern
   [ESCAPE esc-char-expression]
```

1.4. Publishing History

Product Version	Publication Date
Trafodion Release 1.3.0	To be announced.

1.5. Comments Encouraged

The Trafodion community encourages your comments concerning this document. We are committed to providing documentation that meets your needs. Send any errors found, suggestions for improvement, or compliments to:

issues@trafodion.incubator.apache.org

Include the document title and any comment, error found, or suggestion for improvement you have concerning this document. Or, even better, join our community and help us improve our documentation. Please refer to Trafodion Contributor Guide for details.

Chapter 2. Requirements

Trafodion requires an x86 version of Linux.

The current release of Trafodion has been tested with:

- 64-bit Red Hat Enterprise Linux (RHEL) or CentOS 6.5, 6.6, and 6.7

- SUSE SLES 11.3

- Cloudera CDH 5.2

- Cloudera CDH 5.3

- Hortonworks HDP 2.2

Other OS releases may work, too. The Trafodion project is currently working on better support for non-distribution version of Hadoop.

2.1. General Cluster and OS Requirements and Recommendations

64-bit x86 instruction set running a Linux distribution is required. Further, Trafodion assumes an environment based on the requirements of the supported Hadoop distributions.

2.1.1. Hardware Requirements and Recommendations

Single-Node Cluster

It is possible to run Trafodion on a single-node sandbox environment. Typically, any sandbox running a Hadoop distribution can be used. A typical single-node configuration uses 4-8 cores with 16 GB of memory, and 20 GB free disk space beyond what's used for the Hadoop environment.

Multi-Node Cluster

For multi-node end-user clusters, your typical HBase environment should suffice for Trafodion. Typically, memory configuration range between 64-128 GB per node with minimum requirement of 16 GB. The cluster size can span from 1 to *n* nodes; a minimum of two nodes is recommended. A minumum of two cores is required regardless of whether you're deploying Trafodion on a bare-metal or virtual environment.

Recommended configurations:

Processors per Node	* Small: 2 cores * Medium: 4 cores * Large: 8+ cores
Memory per Node	* Small: 16 GB * Medium: 64 GB * Large: 128 GB
Concurrency:Nodes	* Two Small Nodes: Four concurrent queries * Two Medium Nodes: 64 concurrent queries * Two Large Nodes: 256 concurrent queries

2.1.2. OS Requirements and Recommendations

Please verify these requirements on each node you will install Trafodion on:

Function	Requirement	Verification Guidance
Linux	64-bit version of Red Hat 6.5 or later, or SUSE SLES 11.3 or later.	
sshd	The `ssh` daemon is running on each node in the cluster.	* `ps aux	grep sshd` * `sudo netstat -plant	grep :22`
ntpd	The `ntp` daemon is running and synchronizign time on each node in the cluster.	* `ps aux	grep ntp` * `ntpq -p`
FQDN	* `/etc/hosts` is set up for fully-qualified node names (FQDN). * `/etc/resolv.conf` is configured to use a name server.	* `host -T <FQDN>` (responds if using a DNS server, times out otherwise) * Simply ssh among nodes using `ssh <FQDN>`.
Port Availability	The Linux Kernel Firewall (`iptables`) has either been disabled or ports required by Trafodion have been opened.	* `lsmod	grep ip_tables` checks whether `iptables` is loaded. If not, no further checking is needed. * `sudo iptables -nL	grep <port>` checks the configuration of a port. An empty response indicates no rule for the port, which often means the port is **not** open.
passwordless ssh	The user name used to provision Trafodion must have passwordless ssh access to all nodes.	ssh to the nodes, ensure that no password prompt appears.
sudo priviliges	The user name used to provision Trafodion must sudo access to a number of root functions .	`sudo echo "test"` on each node.
bash	Available for shell-script execution.	`bash --version`
java	Available to run the Trafodion software. Sama version as HBase is using.	`java --version`
perl	Available for script execution.	`perl --version`
python	Available for script execution.	`python --version`
yum	Available for installs, updates, and removal of software packages.	`yum --version`
rpm	Available for installs, updates, and removal of software packages.	`rpm --version`
scp	Available to copy files among nodes in the cluster.	`scp --help`
curl	Available to transfer data with URL syntax.	`curl --version`
wget	Available to download files from the Web.	`wget --version`
pdsh	Available to run shell commands in parallel.	`pdsh -V`
pdcp	Available to copy files among nodes in parallel. part of the `pdsh` package.	`pdcp -V`

2.1.3. IP Ports

The following table lists the default ports used by the different Trafodion components plus the configuration file and configuration attribute associated with each port setting.

Default Port	Configuration File	Configuration Entry	Required	Range	Protocol	Comment
4200	rest-site.xml	trafodion.rest.port	Yes	1	REST	Trafodion REST Server.
4201	rest-site.xml	trafodion.rest.https.port	Yes	1	HTTPS	Trafodion REST Server (HTTPS).
23400	dcs-site.xml	dcs.master.port	Yes	n	binary	Start of Trafodion DCS port range. (37800 for Trafodion 1.1)
24400	dcs-site.xml	dcs.master.info.port	Yes	1	HTTP	DCS master web GUI. (40010 for Trafodion 1.1)
24410	dcs-site.xml	dcs.server.info.port	Yes	n	HTTP	Start of range for DCS server web GUIs. (40020 for Trafodion 1.1)
50030	mapred-site.xml	mapred.job.tracker.http.address	No	1	HTTP	MapReduce Job Tracker web GUI.
50070	hdfs-site.xml	dfs.http.address	No	1	HTTP	HDFS Name Node web GUI.
50075	hdfs-site.xml	dfs.datanode.http.address	No	1	HTTP	HDFS Data Node web GUI.
50090	hdfs-site.xml	dfs.secondary.http.address	No	1	HTTP	HDFS Secondary Name Node web GUI.
60010	hbase-site.xml	hbase.master.info.port	No	1	HTTP	HBase Master web GUI.
60030	hbase-site.xml	hbase.regionserver.info.port	No	1	HTTP	HBase Region Server web GUI.

There are two port ranges used by Trafodion.

- 23400 is a range, to allow multiple mxosrvr processes on each node. Allow a range of a few ports, enough to cover all the servers per node that are listed in the "servers" file in the DCS configuration directory.

- 24410 is a range as well, enough to cover the DCS servers per node, usually 1 or 2.

On top of the ports identified above, you also need the ports required by your Hadoop distribution. For example:

- *Cloudera Ports*

- *Hortonworks Ports*

Although not all the ports will be used on every node of the cluster, you need to open most of them for all the nodes in the cluster that have Trafodion, HBase, or HDFS servers on them.

2.2. Prerequisite Software

Trafodion runs as an add-on service on Hadoop distributions. The following Hadoop services and their dependiencies must be installed and running on the cluster where you intend to install Trafodion:

- Hadoop Distributed File System (HDFS)

- YARN with MapReduce version 2

- ZooKeeper

- HBase

- Hive

- Apache Ambari (Hortonworks) or Cloudera Manager (Cloudera) with associated embedded databases.

The following distributions are supported for Trafodion:

Distribution	Version	HBase Version	Installation Documentation
Cloudera Distribution Including Apache Hadoop (CDH)	5.2 or 5.3	0.98	CHD 5.2 Installation http://www.cloudera.com/downloads/manager/5-3-0.html[1]
Hortonworks Data Platform (HDP)	2.2	0.98	HDP 2.2 Installation

[1] When possible, install using **parcels** to simply the installation process.

> Trafodion does not yet support installation on a non-distribution version of Hadoop; that is, Hadoop downloaded from the Apache web site. This restriction will be lifted in a later release of Trafodion.

2.2.1. Software Packages

In addition to the software packages required to run different Hadoop services listed above (for example, `Java`), the following software packages and their dependencies must be installed on all nodes where Trafodion will be installed.

> For RedHat/CentOS, the Trafodion Installer can install a subset of the packages below.

Package	Usage	Installation
EPEL	Add-on packages to completed the Linux distribution.	Fedora RPM
pdsh	Parellelize shell commands during install and Trafodion runtime utilities.	yum install pdsh
log4cxx	Message logging.	Manual process[1]

Package	Usage	Installation
sqlite	Internal configuration information managed by the Trafodion Foundation component.	yum install sqlite
expect	Not used?	yum install expect
perl-DBD-SQLite	Allows Perl scripts to connect to SQLite.	yum install perl-DBD-SQLite
perl-Params-Validate	Validates method/function parameters in Perl sripts.	yum install perl-Params-Validate
perl-Time-HiRes	High resolution alarm, sleep, gettimeofday, interval timers in Perl scripts.	yum install perl-Time-HiRes
protobuf	Data serialization.	yum install protobuf
xerces-c	C++ XML parsing.	yum install xerces-c
gzip	Data compress/decompress.	yum install gzip
rpm-build[2]	Build binary and source software packages.	yum install rpm-build
apr-devel[2]	Support files used to build applications using the APR library.	yum install apr-devel
apr-util-devel[2]	Support files used to build applications using the APR utility library.	yum install apr-util-devel
doxygen[2]	Generate documentation from annotated C++ sources.	yum install doxygen
gcc[2]	GNU Compiler Collection	yum install gcc
gcc_c++[2]	GNU C++ compiler.	yum install gcc_c++

1. `log4c++` was recently withdrawn from public repositories. Therefore, you will need to build the `log4c++` RPM on your system and then install the RPM using the procedure described in log4c++ Installation.

2. Software package required to build `log4c++`. Not required otherwise. These packages are **not** installed by the Trafodon Installer in this release.

The Trafodion Installer requires Internet access to install the required software packages.

2.3. Trafodion User IDs and Their Priviliges

2.3.1. Runtime

The `trafodion:trafodion` user ID is created as part of the installation process. The default password is: `traf123`.

Trafodion runs under this ID, which must be registered as a user in the Hadoop Distributed File System (HDFS) to store and access objects in HDFS, HBase, and Hive. In addition, the `trafodion` user ID required passwordless access among the nodes where Trafodion is installed.

Trafodion requires that either HDFS ACLs or Kerberos is enabled.

Trafodion users are managed by the Trafodion security features (grant, revoke, etc.), which can be integrated with LDAP if so desired. These users are referred to as **database users** and do not have direct access to the operating system.

ℹ️ Do **not** create the `trafodion:trafodion` user ID in advance. The Trafodion Installer uses the presence of this user ID to determine whether you're doing an installation or upgrade.

2.3.2. Installation/Upgrades

Typically, the Trafodion Installer is used for Trafodion installations. It requires access to the following user IDs:

User ID	Usage	Requirements
Linux	Run the Trafodion Installer, which sets up the enviroment and installs the Trafodion software.	* User name or group cannot be `trafodion`. * Passwordless ssh access to all nodes in the cluster. * Internet access to download software packages. * `requiretty` must be disabled in `/etc/sudoers`. * `sudo`[1] access to: **Download and install software packages.** Mofify `/etc/sudoers.d` (allow the `trafodion` user to modify floating IP: `ip` and `arping`). **Create the `trafodion` user ID and group.** Install Trafodion software into the HBase environment. **Run Java version command on each node in the cluster.** Run Hadoop version command on each node in the cluster. **Run HBase version command on each node in the cluster.** Create directories and files in: * **`/etc`** * `/usr/lib` * **`/var/log`** Invoke `su` to execute commands as other users; for example, `trafodion`. **Edit `sysctl.conf` and activate changes using `sysctl -p`:** * Modify kernel limits. * Reserve IP ports. [1] `sudo` is **required** in the current release of Trafodion. This restriction may be relaxed in later releases. Alternative mechanisms for priviliged access (such as running as `root` or `sudo` alternative commands) are not supported.
Distribution Manager	Modify configuration and stop/start services.	* Administrator user name and password. * URL to Distribution Manager's REST API.
HDFS Administator	Create directories and change security setting, if needed.	* HDFS Administrator user name and password. * Write access to home directory on the node where the Distribution Manager is running.
HBase Administator		* HBase Administrator user name and password. * Read access to `hbase-site.xml`.

2.4. Required Configuration Changes

Trafodion requires changes to a number of different areas of your system configuration: operating system, HDFS, and HBase.

ℹ️ These changes are performed by the Trafodion Installer, if used.

2.4.1. Operating System Changes

`/etc/security/limits.d/trafodion.conf` on each node in the cluster must contain the following settings:

```
# Trafodion settings
trafodion   soft core     unlimited
trafodion   hard core     unlimited
trafodion   soft memlock unlimited
trafodion   hard memlock unlimited
trafodion   soft nofile  32768
trafodion   hard nofile  65536
trafodion   soft nproc   100000
trafodion   hard nproc   100000
hbase       soft nofile  8192
trafodion   soft nofile  8192
trafodion   hard nofile  65535
```

2.4.2. ZooKeeper Changes

These changes require a restart of ZooKeeper on all nodes in the cluster.

Trafodion requires the following changes to `zoo.cfg`:

Setting	New Value	Purpose
maxClientCnxns	0	Tell ZooKeeper to impose no limit to the number of connections.

2.4.3. HDFS Changes

These changes require a restart of HDFS on all nodes in the cluster.

Trafodion requires the following changes to the HDFS environment:

Action	Purpose
* Create `/hase-staging` directory. * Change owner to HBase Administrator.	
* Create `/bulkload` directory. * Change owner to `trafodion`.	
* Create `/lobs` directory. * Change owner to `trafodion`.	
* Create `/apps/hbase/data/archive`[1]. * Change owner to: **hbase:hbase (Cloudera)** `hbase:hdfs` (Hortonworks) * Give the `trafodion` user RWX access to `/apps/hbase/data/archive` * Set default user of `/apps/hbase/data/archive` to `trafodion` * Recursively change `setafcl` of `/apps/hbase/data/archive` to RWX	

1. These steps are performed **after** HDFS ACLs have been enabled.

The following changes are required in `hdfs-site.xml`:

Setting	New Value	Purpose
dfs.namenode.acls.enabled	true	Enable HDFS POSIX Access Control Lists (ACLs).

2.4.4. HBase Changes

These changes require a restart of ZooKeeper on all nodes in the cluster.

Trafodion requires that the following changes to the HBase environment:

Action	Affected Directories	Purpose
Install/replace Trafodion's version of `hbase-trx`	* `/usr/lib/hbase/lib/` * `/usr/share/cmf/lib/plugins/` (Cloudera) * `/usr/hdp/current/hbase-regionserver/lib/` (Hortonworks)	Trafodion transaction management relies on an enhanced version of `hbase-trx`.
Install/Replace Trafodion utility jar file.	* `/usr/lib/hbase/lib/` * `/usr/share/cmf/lib/plugins/` (Cloudera) * `/usr/hdp/current/hbase-regionserver/lib` (Hortonworks)	TODO: Add purpose here.

The following changes are required in `hbase-site.xml`. Please refer to the Apache HBase Reference Guide for additional descriptions of these settings.

Setting	New Value	Purpose
hbase.master. distributed.log.splitting	false	Do not use the HBase Split Log Manager. Instead, the HMaster controls all log-splitting activities.
hbase.coprocessor. region.classes	org.apache.hadoop. hbase.coprocessor. transactional.TrxRegionObserver, org.apache.hadoop. hbase.coprocessor. transactional.TrxRegionEndpoint, org.apache.hadoop. hbase.coprocessor. AggregateImplementation	Install Trafodion coprocessor classes.
hbase.hregion.impl	org.apache.hadoop. hbase.regionserver. transactional.TransactionalRegion	Install the Trafodion class for Write Ahead Log reading.
hbase.regionserver. region.split.policy	org.apache.hadoop. hbase.regionserver. ConstantSizeRegionSplitPolicy	Tell HBase to use the ConstantSizeRegionSplitPolicy for region splitting. This setting causes region splitting to occur only when the maximum file size is reached.
hbase.snapshot. enabled	true	Enable the HBase Snapshot feature. Used for Trafodion backup and restore.
hbase.bulkload. staging.dir	hbase-staging	Use `/hbase-staging` as the bulk load staging directory.

Setting	New Value	Purpose
hbase.regionserver.region. transactional.tlog	true	The HBase Regions requests that the Transaction Manager redrives in-doubt transactions.
hbase.snapshot. master.timeoutMillis	600000	HMaster timeout when waiting for RegionServers involved in the snapshot operation.
hbase.snapshot. region.timeout	600000	RegionServer timeout when waiting for snapshot to be created.
hbase.client. scanner.timeout.period	600000	Time limit to perform a scan request.
hbase.regionserver. lease.period	600000	Clients must report within this time limit or they are considered dead by HBase.
hbase.namenode. java.heapsize[1]	1073741824 (1GB)	Java Heap Size for the HDFS NameNode.
hbase.secondary.namenode. java.heapsize[1]	1073741824 (1GB)	Java Heap Size for the HDFS Secondary NameNode.

1. Applies to Cloudera distributions only.

2.5. Recommended Configuration Changes

The following configuration changes are recommended but not required.

🛈 | The Trafodion Installer does **not** make these changes.

2.5.1. Recommended HBase Configuration Changes

Configuration Property	Recommended Setting	Guidance
hbase.rpc.timeout	10 minutes	This setting depends on the tables' size. Sixty (60) seconds is the default. Increase this value for big tables. Make it the same value as `hbase.client.scanner.timeout.period`. We have found that increasing the setting to six-hundred (600) seconds will prevent many of the timeout-related errors we encountered, such as `OutOfOrderNextException` errors.
hbase.client.scanner.timeout.period	10 minutes	Similar to the `hbase.rpc.timeout` setting. Sixty (60) seconds is the default. Depending on the size of a user table, we have experienced timeout failures on count(*) and update statistics commands from this setting. The underlying issue is the length of the execution of the coprocessor within HBase. NOTE: HBase uses the smaller of `hbase.rpc.timeout` and `hbase.client.scanner.timeout.period` to calculate the scanner timeout.

Configuration Property	Recommended Setting	Guidance
hbase.snapshot.master.timeoutMillis and hbase.snapshot.region.timeout	10 minutes	HBase's default setting is 60000 milliseconds. If you experience timeout issues with HBase snapshots when you use the Trafodion Bulk Loader or other statements, you can set the value for these two HBase properties to 10 minutes (600,000 milliseconds).
hbase.hregion.max.filesize	107374182400 bytes	HBase's default setting is 10737418240 (10 GB). We have increased the setting to 107374182400 (100 GB), which reduces the number of HStoreFiles per table and appears to reduce disruptions to active transactions from region splitting.
hbase.hstore.blockingStoreFiles	10	http://gbif.blogspot.com/2012/07/optimizing-writes-in-hbase.html
hbase.regionserver.handler.count	<num>	This setting should match the number of concurrent sessions (mxosrvr). The default is 10.