Esgyn Corporation

EsgynDB Installation Guide for Release R2.4.0

Published: April 2018
Edition: EsgynDB Release 2.4.0

Contents

Legal Notice

Acknowledgements

1. About This Document

This EsgynDB Installation Guide describes how to install and configure Release R2.4.0 of the EsgynDB core product (based on Trafodion) and required components on a Hadoop cluster.

It does not describe how to install the OS or Hadoop distribution, which are prerequisites for the EsgynDB installation. Refer to the vendor documentation for help in installation and configuration of those components.

2. Intended Audience

This guide is intended for EsgynDB and Hadoop system administrators.

3. Pre-requisite

Starting with EsgynDB R2.2, installation requires a license key provided by Esgyn Corporation. Make sure you have one before starting the installation process.

4. Preparing Your PC

If you are using a Windows PC, additional PC software may be needed for the installation process. It is recommended that you preinstall the PC software before continuing with the EsgynDB installation.

Required PC software:

- putty and puttygen (download from the PuTTY website)
- VNC client (download from http://www.realvnc.com)
- Firefox or Chrome browser
- SFTP client to transfer files from your PC to the server: WinSCP or FileZilla

5. Validate Your Cluster Environment

Before installing EsgynDB, validate your cluster environment.

Cluster Requirements

The following configuration settings have been tested and are known to work with the EsgynDB installation:

Hardware platform	**x86-64**
Operating systems	CentOS 6.5 ~ 6.9, 7.2 (64 bit) Red Hat 6.5 ~ 6.9, 7.2 (64-bit)
Hadoop distributions	Cloudera CDH 5.4 ~ 5.9 Hortonworks HDP 2.3 ~ 2.5 Note: CDH 5.7 or greater is required for CentOS / Red Hat Enterprise Linux 7.2.
User IDs	A user ID with passwordless sudo access for command-line installer. See Obtaining Sudo Access and Passwordless SSH.
Cluster size	A cluster consisting of 1 to n nodes. There is currently no upper limit. Two nodes is the recommended minimum.
Disk space	Minimum of 20 GB. See Checking the Disk Space.
Memory	Minimal 1 GB * number of connectivity servers (MXOSRVR processes) configured per node in the cluster.

Check the Disk Space

Before installing the Hadoop distribution, ensure a minimum of 20 GB is available to support the database. The default installation location for Cloudera CDH is **"/var/lib/cloudera-scm-server-db"**. To check the size available to **/var**, start a putty session, or a VNC terminal window on the node in your cluster where Cloudera will be installed. To execute the following commands, you will need either **root** or **sudo** access.

Confirm there is a minimum of 20 GB available to **/var**.

```
$ cd "/var"
$ df -hP
```

If there is insufficient space available in **/var**, a possible solution is to provide a soft link to another drive for your Cloudera database. Locate a drive that does have sufficient space.

```
$ cd <new drive>  (e.g. cd /DATA)
$ mkdir cloudera-scm-server-db
$ chmod 777 cloudera-scm-server-db
$ cd /var/lib
$ ln -s <new drive>/cloudera-scm-server-db .
```

If **/var** is a subdirectory in your cluster's root filesystem, the Cloudera database should have sufficient space available.

If the Cloudera distribution has already been installed and it is showing red for log directories, it means that Cloudera was installed using **/var/lib** and may not have a large enough **/var** file system. In this case, there is an unsupported script that can be used to move the directories. Please see the **clouderaMoveDB.sh** script in the **installer/tools** directory, which is created when the installer tar.gz file is untarred. Execute **clouderaMoveDB.sh** without syntax to display help information.

6. Install a Supported Hadoop Distribution

EsgynDB R2.3 is compatible with the Cloudera and Hortonworks distributions.

Distribution	Version	HBase Version	Installation
Cloudera Distribution Including Apache Hadoop (CDH)	CDH 5.4.5 ~ 5.13	1.0, 1.2	Refer to installation instructions on the Cloudera site for the specific version you plan to install.
Hortonworks Data Platform (HDP)	HDP 2.3, 2.5	1.1	Refer to installation instructions on the Hortonworks site for the specific version you plan to install.

Mandatory Hadoop Services and Settings

IMPORTANT: Before installing a Hadoop distribution, please review this list of mandatory services and settings:

- HDFS
- Yarn/MapReduce
- ZooKeeper
- HBase
- Hive
- Embedded Databases

Please make sure that those services and settings are selected during installation.

IMPORTANT: The EsgynDB command-line installer needs to run from one of the nodes that will be a part of the EsgynDB cluster. Off-platform installation is not currently supported. All EsgynDB nodes must have HBase installed.

7. Prepare for Install

Obtain sudo access and passwordless SSH (for command-line install option)

The EsgynDB installation requires a user ID with these attributes:

- sudo access
- passwordless ssh to all nodes on the cluster

> Note: You may need to request permission from your cluster management team to obtain this type of access.

The following example shows how to set up your user ID to have "passwordless ssh" abilities.

```
$ echo -e 'y\n' | ssh-keygen -t rsa -N "" -f $HOME/.ssh/id_rsa
$ cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys
$ echo localhost $(cat /etc/ssh/ssh_host_rsa_key.pub) >> $HOME/.ssh/known_hosts
$ echo "NoHostAuthenticationForLocalhost=yes" >> $HOME/.ssh/config
$ chmod 600 $HOME/.ssh/config
```

After running these commands, you will need to copy the contents of the public key file, `$HOME/.ssh/id_rsa.pub`, and append those contents to each additional node's `$HOME/.ssh/authorized_keys` file. You will also need to copy your private `$HOME/.ssh/id_rsa` file from the current node to the other node's `$HOME/.ssh` directory, and secure it as private to yourself (chmod 700).

Configure an LDAP Identity Store

If you plan to enable authentication in EsgynDB, you will need to have a configured LDAP identity store available. The EsgynDB installer will prompt you to set up an authentication configuration file that points to an LDAP server (or servers), which will enable security (that is, authentication and authorization). Refer to the section on Enabling Security Features for more details.

Integrate with Kerberos

If Kerberos is enabled on your system, then it needs to be enabled in EsgynDB. The EsgynDB installer will prompt you to set up Kerberos attributes needed to install Kerberos principals and keytabs. Refer to the section on Enabling Security Features for more details.

User IDs and Passwords

This table lists the user IDs and passwords that you will use during installation.
NOTE: You will use two user IDs: your user ID and **trafodion**.

Logon	User ID	Password	Description
Cloudera Manager Web GUI logon	`admin`(default)	`admin`(default)	After installing Cloudera, you will be instructed to log on to the Cloudera Manager Web GUI. Use the default user ID and password. If you already had Cloudera installed, please use your previously defined user ID and password.

Apache Ambari Web GUI logon	`admin`(default)	`admin`(default)	After installing Hortonworks' HDP, you will be instructed to log on to the Apache Ambari Web GUI. Use the default user ID and password. If you already had HDP installed, please use your previously defined user ID and password.
User ID with sudo access	<sudo-username>	<password>	In the installation steps, you may be instructed to use "sudo" or "sudo userid" access. You will be using your user ID, which has been enabled with "sudo" access and passwordless ssh to all nodes of the cluster.
EsgynDB logon	`trafodion`	`traf123` (default)	This user ID is **automatically created** for you by the EsgynDB installer when EsgynDB is installed. Do not create this user ID manually.

Required Software

EsgynDB requires supplementary software to be installed on the cluster before it is installed. These are Linux tools that are not typically packaged as part of the core Linux distribution. The installer will attempt to automatically get these packages over the Internet, but if the cluster's access to the Internet is disabled, you will need to manually download the packages and make them available for installation.

OS	Packages	
CentOS Linux 6.5 ~ 6.9, 7.2 **Red Hat Linux 6.5 ~ 6.9, 7.2**	pdsh	apr
	log4cxx	apr-util
	sqlite	protobuf
	expect	lzo
	perl-DBD-SQLite	lzop
	xerces-c	unzip
	perl-Params-Validate	gcc-c++
	perl-Time-HiRes	unixODBC
	gzip	unixODBC-devel
	gnuplot	libiodbc
	lsof	libiodbc-devel
	keepalived	openldap-clients
	libcgroup	snappy

Information Gathering

The installer will prompt for additional information over the course of the installation process. Before you start installation, make sure you have collected the data

Information	Default	Notes
Location of the working directory where the installer will untar files	None	Specify a location
Location of the directory on the node where the installer package was untarred	None	You will need to specify the location

License key	None	You will need a license key from Esgyn Corporation
Number of concurrent client sessions per node	8	This number specifies the concurrent sessions per node to be supported. Each session could require up to 1GB of physical memory. The number can be changed post-installation. For more information, refer to the DCS Installation Guide
Upgrade of software on an existing EsgynDB installation or a new install	None	The installer will take additional actions if the installation is on a new cluster
Trafodion user id and password	User ID: `trafodion` Password: `traf123`	Recommend not to change the User ID. The password should be changed after installation on all nodes is complete, if command-line installer is used.
List of nodes in the cluster	None	All nodes should be accessible by passwordless ssh as the sudo or root user.
Prefix of the home directory path of the trafodion user ID	/home	If the home directory of the trafodion user ID is "/opt/home/trafodion", specify the prefix as "/opt/home"
Location of the JDK	None	Fully qualified path of the JDK. For example: "/usr/java/jdk1.8.0_112-cloudera"
Location of the EPEL rpm	None	If your machine does not have external Internet access then you must install the EPEL repo manually.
Location of the EsgynDB package tar file	None	Specify the full pathname of the EsgynDB package
URL for the Hadoop distribution	None	Specify in the form: <IP-address>:<port> or <node name>:<port> Example: "vm-1.yourcompany.local:7180"
Hadoop distribution details 1. Admin UI user ID, password 2. Cluster name 3. HDFS user ID 4. HBase user ID, group 5. HBase service name	 Distribution-dependent Cluster 1 hdfs hbase, hbase hbase	
EsgynDB installation directory	None	Specify the full path for the directory. This will allow you to maintain

		multiple versions of the software if desired.
DCS HA (High Availability)	Disabled	You will need the Floating IP address, the interface, and the backup nodes for DCS master.
Security	Disabled	If security is to be enabled, ensure LDAP is configured and the name of the LDAP configuration file is available.
Kerberos	Hadoop option	If Kerberos is enabled on your system, EsgynDB needs to create Kerberos principals and keytabs for the `trafodion` ID.

8. Install EsgynDB

Command Line Installer

The EsgynDB command line installation tool is called `db_install.py` It is distributed as an independent package.

EsgynDB must be installed on all nodes that host an HBase RegionServer (that is, where a supported Hadoop distribution is installed).

1. You will execute the installer from one of the nodes in the EsgynDB cluster, as user <sudo-username>. Create the `esgyndb_downloads` directory (if one does not exist) on a node that is part of the cluster. Typically, this will be the first node in the cluster. Place the downloaded server package (**esgynDB_server-2.4.0-RH6-x86_64.tar.gz** or **esgynDB_server-2.4.0-RH7-x86_64.tar.gz,** depending on the installed OS version) and the corresponding installer package (**esgynDB_pyinstaller-2.4.0-RH6.tar.gz** or **esgynDB_pyinstaller-2.4.0-RH7.tar.gz**) compressed tar files into it. For example

   ```
   $ mkdir $HOME/esgyndb_downloads
   $ mv <your-download-path>/esgynDB_server-2.4.0-RH6-x86_64.tar.gz
   $HOME/esgyndb_downloads
   $ mv <your-download-path>/esgynDB_pyinstaller-2.4.0-RH6.tar.gz
   $HOME/esgyndb_downloads
   $ cd $HOME/esgyndb_downloads
   ```

2. Untar the downloaded installer file. For example
   ```
   $ tar -xzf esgynDB_pyinstaller-2.4.0-RH6.tar.gz
   ```

3. Change to the installer directory:
   ```
   $ cd python-installer
   ```

4. Run the installer script. You may execute it in one of two installation modes -

Mode	Comments
Guided Setup	The installer will interactively prompt for information from the user as it works through the installation process. Recommended for the new user.
Expert Setup	Required information is provided in a pre-formatted plain text configuration file. The installer is pointed to this file at invocation. The installer will not prompt for any further information. A template of the configuration file is available within the untarred installer directory: `python-installer/configs/db_config_default.ini` Make a copy of the file in your directory and populate the needed information. Recommended for the expert user and/or for unattended installs.

Guided Setup mode: You will be prompted for all needed information.

```
./db_install.py
```

Expert mode: Ensure the installer configuration file is prepopulated. Invoke the installer
```
./db_install.py –config_file <installer-config-file>
```

NOTES:
- The installer will abort in case of any configuration errors. Correct the errors before you re-run.
- If you choose not to start Trafodion after the installation (that is, if you enter **N** for **Start Trafodion after install (Y/N)**), you will need to manually start and initialize Trafodion after **db_install.py** completes. See notes below.
- For details on manually enabling security in EsgynDB, refer to the section Enabling Security Features.

The installation process will
- Prompt for a license key

```
Add a new license file or license string (Y/N) [N]: Y
Enter full path to license file or the license string [NONE]:
```

- Install necessary RPMs on CentOS/Redhat Linux systems
- create the `trafodion` user ID
- set up passwordless ssh for the `trafodion` user ID
- copy the EsgynDB distribution files across the cluster, and
- generate startup files
- start EsgynDB, Database Connectivity Services (DCS) to access the system, and EsgynDB Manager services.

5. Once installation completes successfully, you will see this message:
```
**************************
  Installation Complete
**************************
```

6. Your EsgynDB system should now be up and running. Log on to the system as the `trafodion` user. If you chose not to start Trafodion after the installation, start and initialize Trafodion as follows
```
cds
sqstart

[trafodion@nap001 ~]$ sqlci
EsgynDB Enterprise Conversational Interface 2.4.0
Copyright (c) 2015-2018 Esgyn Corporation
>>initialize trafodion;
```

Manage

You use the `trafodion` user ID to perform EsgynDB management operations.

The following table provides an overview of the subsystem management scripts.

Component	Start	Stop	Status
EsgynDB core	`sqstart`	`sqstop`	`sqcheck`
RMS Server	`rmsstart`	`rmsstop`	`rmscheck`

REST Server	reststart	reststop	-
LOB Server	lobstart	lobstop	-
DCS (Database Connectivity Services)	dcsstart	dcsstop	dcscheck

Example: Start EsgynDB

```
$ cd $TRAF_HOME/sql/scripts
$ sqstart
$ sqcheck
```

Cloudera Manager Installer

This installer leverages Cloudera Manager for the installation, and is applicable only in cases where the Cloudera Hadoop distribution (CDH) is installed, as a Cloudera parcel.

The EsgynDB software is comprised of a Custom Service Descriptor (CSD), two Parcels, and a parcel manifest. A configuration helper script (`cm_settings.sh`) is also provided. Use of the helper script is optional, as all of the needed configurations are described in this document (see Hadoop Configuration).

EsgynDB software is provided in a single compressed tar-file download, such as **`esgyndb-CM-`*`version`*`.tgz`**.

Installing CSD (Cloudera Service Descriptor)

The CSD is a jar file, named **`ESGYNDB-`*`version`*`.jar`**, such as `ESGYNDB-2.4.0-1.jar`. This file must be placed on the Cloudera Manager Server host in the Local Descriptor Repository directory, which defaults to `/opt/cloudera/csd`.

If Cloudera Manager is not yet installed, the directory may be pre-created and the EsgynDB CSD placed there. When Cloudera Manager is installed, it will be automatically picked up.

If Cloudera Manager is already installed, then the server process (cloudera-scm-server) must be re-started to pick up the CSD (e.g., `sudo service cloudera-scm-server restart`).

Installing Parcels

The ESGYNDB_TRX parcel is a HBase plug-in for transaction processing. The ESGYNDB parcel is the main service providing the SQL engine and other database services.

The parcel files and parcel manifest file must be placed in a remote parcel repository. This may be any regular web-server location or a temporary web-server location.

To configure the parcel location, click the parcel icon in the Cloudera Manager menu bar, and then the Configuration button in the upper right. Add a new Remote Parcel Repository URL, referencing your EsgynDB parcels location.
On the main Parcels screen, click the Check for New Parcels button.

Now that Cloudera Manager knows about the EsgynDB service and where to find EsgynDB parcels, the EsgynDB service is available to be installed.

Adding EsgynDB to an Existing Cluster

For an existing cluster, follow these steps:

1. Go to the Parcels screen (package icon on title menu), select the cluster.
2. For both the ESGYNDB_TRX and ESGYNDB parcels,
   a. Download
   b. Distribute
   c. Activate
3. Modify the required Hadoop Configuration settings, and restart services as necessary (at least HBase).
4. Use the cluster Actions menu to Add Service, then select EsgynDB.

Note: ESGYNDB_TRX must be activated on all HBase Regionserver hosts.

Creating a New EsgynDB Cluster

EsgynDB can also be installed when adding a new cluster.
When selecting parcels, select CDH, ESGYNDB, and ESGYNDB_TRX.
When selecting services, select EsgynDB from the Custom Service list. The dependent services, such as HBase, will automatically be included.

In this scenario, EsgynDB service will give an error and fail to start until HBase required configuration settings are deployed and HBase is restarted, as indicated in Hadoop Configuration.

When EsgynDB start-up gives the TRX configuration error, Cloudera Manager will give options to Retry or go Back. Instead, click on the title menu bar to go to the main screen. From there you can configure Hadoop and restart services as necessary.
That should also start EsgynDB, but you must then use the EsgynDB Actions menu to select Initialize EsgynDB MetaData.

Host Selection

EsgynDB Node

This is the main worker role, which runs database queries and interfaces with other Hadoop components (HDFS, HBase, Hive). EsgynDB nodes are usually placed on HBase Regionserver nodes, but may be placed on any HBase Gateway, Regionserver, or Master nodes. EsgynDB nodes also act as Hive clients, so need to be co-located with Hive Gateway or Server roles.

Distributed Connectivity Server

This role routes outside requests (JDBC/ODBC connections) to specific nodes. For availability, more than one node should be selected. These roles must be co-located with EsgynDB Nodes.

If configured, a floating IP address can be used to refer to the active DCS master node. Use of this feature will use the "sudo" command for a couple of network administration commands. Otherwise, the client driver (JDBC/ODBC) needs to be configured with a list of the server nodes instead of a single floating IP address. A third option is to maintain a floating IP address using the keepalived package. EsgynDB will provide a sample configuration file, but keepalived must be configured separately, since Cloudera Manager does not manage it.

EsgynDB Manager

This role provides an administrative interface to monitor and interact with your EsgynDB cluster. This role must be colocated with an EsgynDB Node.

Hadoop Configuration

EsgynDB depends on certain configuration settings for HDFS, Zookeeper, and HBase. They are all important for proper operation, but EsgynDB will not start at all without the required settings. The cluster administrator must change these values.
They can be modified manually via the Cloudera Manager web interface, or via the provided tool.

Required

These HBase settings are required before starting EsgynDB service:
• Region Server Advanced Configuration Snippet for hbase-site.xml

Name	Value
hbase.hregion.impl	org.apache.hadoop.hbase.regionserver.transactional.TransactionalRegion
hbase.regionserver.region.split.policy	org.apache.hadoop.hbase.regionserver.ConstantSizeRegionSplitPolicy

Note: The ESGYNDB_TRX parcel must be activated on all HBase Regionserver nodes, or these settings will cause HBase errors.

Important

These HBase settings are recommended:

Name	Value
hbase.regionserver.lease.period	**1 (hour)**
hbase.hregion.memstore.flush.size	512 (MiB)
hbase.hregion.memstore.block.multiplier	7
hbase.hstore.blockingStoreFiles	200
hbase.rootdir.perms	750
hbase.snapshot.master.timeoutMillis	10 (minutes)
hbase.superuser	trafodion

These HDFS settings are recommended:

dfs.namenode.acls.enabled	true

These Zookeeper settings are recommended:

maxClientCnxns	0

If Sentry service is enabled:

sentry.service.admin.groups	trafodion
sentry.service.allow.connect	**trafodion**

Automated Configuration

The required and important Hadoop settings can be set via the cm_settings.sh tool. The tool can be run on any linux system with network access to the cluster manager host. Use the -h option for a full list of options.
The tool presents the current and proposed settings of each parameter and prompts for confirmation unless a force option is given to skip the prompts.

Example Usage

```
cm_settings.sh -s https://myhost.domain.net:7180 -c "Cluster5" -u joe_admin
```

This tool requires you to authenticate with Cloudera Manager. After initial authentication, a session cookie is used for the remainder of the script and then deleted. Three options are available to provide the password:

1. Default option allows the underlying "curl" command to prompt for your password.
2. The -n option allows lookup of the password in a standard ~/.netrc file.
3. The -p option allows a password to be passed in via the command line. This option is not recommended as it is less secure.

Use the Cloudera Manager web interface to restart services as needed after using this tool.

EsgynDB Configuration

License Key

When installing the EsgynDB service, you will be prompted for an EsgynDB License Key.
In 2.4.x release, a valid 2.3.x license may be used.
For initial installation, use the license key provided by Esgyn Corporation if this is a non-production use of EsgynDB. For initial installation for production usage, leave the license key blank and continue with installation. A short grace period is allowed to obtain a license key.
To obtain a license key, you need to send a copy of the file `/etc/trafodion/esgyndb_id` file to Esgyn. It is a very small binary file which provides a unique cluster identity.

Others

Other features may be configured after initial installation, including LDAP connection and Connectivity HA. If Hive is Sentry enabled, be sure to indicate this by setting the "Hive Sentry enabled" (traf.sentry.enabled) parameter to true via the checkbox.

Meta-Data Initialization

After initial installation or after installing a new version of EsgynDB, meta-data needs to be initialized. In Cloudera Manager, go to the EsgynDB service, then use the Actions menu to select Initialize EsgynDB MetaData.

It does not hurt to re-run this command. If initialization steps have already been run, then nothing is done. If some ODBC/JDBC connections have been made (DBmanager, for example), they may continue to report that the database has not been initialized, even after a successful initialization. This may be cleared by re-starting EsgynDB service, or use the EsgynDB Actions menu to “Force Connections Restart”.

EsgynDB Start Up

Even after Cloudera Manager has reported that EsgynDB roles have started successfully, it may take a couple of minutes for all the EsgynDB subsystems to start up and accept requests.
Detailed status may be seen using the Check EsgynDB Status command, found on the EsgynDB service Actions menu

Validate

Perform basic sanity checks.

1. Using the EsgynDB Enterprise Conversational Interface (sqlci). Create a table with a few records. For example:

```
[trafodion@edb001 ~]$ sqlci
EsgynDB Enterprise Conversational Interface 2.4.0
Copyright (c) 2015-2018 Esgyn Corporation
>>create table test1 (f1 int, f2 int);

--- SQL operation complete.
>>insert into test1 values(1,1);

--- 1 row(s) inserted.
>>insert into test1 values(2,2);

--- 1 row(s) inserted.
>>select * from test1;

F1            F2
-----------   -----------
          1             1
          2             2

--- 2 row(s) selected.
>>get tables;

Tables in Schema TRAFODION.SEABASE
==================================

TEST1

--- SQL operation complete.
>>exit;
```

2. Run the `mgblty_check` tool to verify all the manageability components (EsgynDB Manager, TSD, TCollector and Bosun) are running

```
[trafodion@edb001 ~]$ mgblty_check

Status of OpenTSD...(Expect 1 per node)
Total count of TSD process: 12

Status of Tcollector...(Expect 1 per node)
Total count of tcollector process:  12

Status of DBMgr...(Expect 1 per DB Manager node)
Total count of dbmgr process: 1

Status of Bosun...(Expect 1 per DB Manager node)
Total count of bosun process: 1
```

3. Open a web browser and connect to http://localhost:4205 to verify that EsgynDB Manager is started and you can successfully login. Refer to the EsgynDB Manager User Guide for more details.

4. Download and install the EsgynDB JDBC and/or ODBC drivers on your client workstation to be able to connect to EsgynDB from a client application. For instructions, refer to the Trafodion Client

Installation Guide which explains how to install the JDBC and ODBC drivers, connect to an EsgynDB system, and run sample programs to test the connection. The package specific to EsgynDB is named **esgynDB_clients-2.4-RH6-x86_64.tar.gz**

9. Uninstall

Use the Trafodion Provisioning User id for these instructions.
Run the commands from the first node of the cluster. Do not run them from a machine that is not part of the EsgynDB cluster.

You do not need to use the `trafodion_uninstaller` script if upgrading EsgynDB.

Stop EsgynDB

Do the following

```
$ su trafodion
$ cd $TRAF_HOME/sql/scripts
$ sqstop
$ exit
```

Uninstall

The `trafodion_uninstaller` script completely removes EsgynDB.

10. Troubleshooting

If you are not able to start up the environment or if there are problems running sqlci or trafci, then verify that all the EsgynDB processes are up and running.

`trafcheck` should indicate all processes are running.

If processes are not running as expected, then:

- `sqstop` to shut down EsgynDB. If some EsgynDB processes do not terminate cleanly, then run `ckillall`.
- `sqstart` to restart EsgynDB.

If problems persist please review logs:

$TRAF_HOME/logs: EsgynDB logs.

11. Enabling Security Features

EsgynDB supports user authentication with LDAP, integrates with Hadoop's Kerberos environment and supports authorization through database grant and revoke requests (privileges).

LDAP is optional and can be configured by running the EsgynDB installer. If Kerberos is enabled in Hadoop, then the EsgynDB installer will ask questions needed to configure Kerberos for EsgynDB

- If Hadoop has enabled Kerberos, then EsgynDB must also enable Kerberos.
- If Kerberos is enabled, then LDAP must be enabled.
- If LDAP is enabled, then database authorization (privilege support) is automatically enabled.
- If Kerberos is not enabled, then enabling LDAP is optional.
- If multi-tenancy is enabled, then database authorization is automatically enabled.

Configuring EsgynDB for Kerberos

Kerberos is a protocol for authenticating a request for a service or operation. It uses the notion of a ticket to verify accessibility. The ticket is proof of identity encrypted with a secret key for the particular requested service. Tickets exist for a short time and then expire. Therefore, you can use the service as long as your ticket is valid (i.e. not expired). Hadoop uses Kerberos to provide security for its services, as such EsgynDB needs to function properly with Hadoop systems that have Kerberos enabled.

Kerberos configuration file

It is assumed that Kerberos has already been set up on all the nodes by the time EsgynDB is installed. This section briefly discusses the Kerberos configuration file for reference.

The Kerberos configuration file defaults to /etc/krb5.conf and contains, among other attributes:

```
* log location: location where Kerberos errors and other information are logged
* KDC location: host location where the KDC (Key Distribution Center) is located
* admin server location: host location where the Kerberos admin server is located
* realm: the set of nodes that share a Kerberos database
* ticket defaults: contains defaults for ticket lifetimes, encoding, and other attributes
```

You need to have access to a Kerberos administrator account to enable Kerberos for EsgynDB. The following is an example request that lists principals defined in the Kerberos database that can be used to test connectivity:

```
kadmin -p 'kdcadmin/admin' -w 'kdcadmin123' -s 'kdc.server' -q 'listprincs'
* -p (principal): please replace 'kdcadmin/admin' with your admin principal
* -w (password): please replace 'kdadmin123' with the password for the admin principal
* -s (server location): please replace 'kdc.server' with your KDC admin server location
* -q (command): defines the command to run, in this case principals are returned
```

Ticket Management

When Kerberos is enabled in EsgynDB, the security installation process:

- Adds a Trafodion principal in Kerberos, one per node with the name trafodion/hostname@realm.
- Creates a keytab for each principal and distributes the keytab to each node. The keytab name is the same for all nodes and defaults to a value based on the distribution, for example: etc/trafodion/keytabs/trafodion.service.keytab.

- Performs a "kinit" on all nodes in the cluster for the trafodion user.
- Adds commands to perform "kinit" and starts the ticket renewal procedure on each node.

The Hadoop renewal service, renews the Kerberos TGT (ticket granting ticket) up until the maximum number of renewals allowed. So if your ticket lifetime is one day and the number of renewals is seven days, your ticket is good for 7 days. After the number of renewals is expired, the ESGYNDB renewal service re-initializes the ticket. Therefore, the ticket should never expire. The EsgynDB renewal service manages the trafodion user only.

Kerberos installation

The EsgynDB installation script automatically determines if Kerberos is enabled on the node. If it is enabled, then the environment variable SECURE_HADOOP is set to "Y". The installer then gathers the following information needed to integrate Kerberos with EsgynDB:

```
KDC server address
KDC admin principal
KDC admin password
Max lifetime for the trafodion user
Max renew lifetime for the trafodion user
```

KDC admin password will be saved only in configuration file **db_config.bakYYMMDD_HHMM** in the installer folder when installation is completed. You can delete this file for secure perspective.
NOTE: Keytab files are auto detected by installer in CDH/HDP cluster.

Configuring LDAP

EsgynDB does not manage usernames and passwords internally but does support authentication via directory servers that use the OpenLDAP protocol, also known as LDAP servers. You can configure the LDAP servers during installation by answering the installer's prompts. For more information, see Configuring LDAP Servers for details. Installing LDAP also enables database authorization (privilege support).

Once authentication and authorization are enabled, EsgynDB allows users to be registered in the database and allows privileges on objects to be granted to users and roles (which are granted to users). EsgynDB also supports component-level (or system-level) privileges, such as MANAGE_USERS, which can be granted to users and roles. See Managing Users.

If you do not enable LDAP in EsgynDB, then a client interface to EsgynDB may request a user name and password, but Trafodion ignores the user name and password entered in the client interface, and the session runs as the database **root** user, **DB__ROOT**, without restrictions. If you want to restrict users, restrict access to certain users only, or restrict access to an object or operation, then you must enable security, which enforces authentication and authorization.

Configuring LDAP Servers

The EsgynDB installer sets up and propagates the LDAP configuration file called `.traf_authentication_config` located in `$TRAF_HOME/sql/scripts`. This file is a flat file, organized as a series of attribute/value pairs. Details on all the attributes and values accepted in the authentication configuration file and how to configure alternate locations can be found in Appendix A.

A sample template file is located in $TRAF_HOME/sql/scripts/traf_authentication_config.

The installer then gathers the following information needed to integrate Kerberos with EsgynDB:

Specification of your directory server(s) requires at a minimum:

1. LDAP Host name(s)
   One or more names of hosts that support the OpenLDAP protocol must be specified. EsgynDB will attempt to connect to all provided host names during the authentication process. The set of usernames and passwords should be identical on all hosts to avoid unpredictable results. The attribute name is LDAPHostName.

   Example:
   ```
   LDAPHostName:  ldap.company.com
   ```

2. LDAP Port number
   Port number of the LDAP server. Typically this is 389 for servers using no encryption or TLS, and 636 for servers using SSL. The attribute name is LDAPPort.

   Example:
   ```
   LDAPPort:  389
   ```

3. LDAP Unique Identifier
   Attribute(s) used by the directory server that uniquely identifies the username. You may provide one or more unique identifier specifiers. The attribute name is UniqueIdentifier.

   Example:
   ```
   UniqueIdentifier:  uid=,ou=users,dc=com
   ```

4. Encryption level
   A numeric value indicating the encryption scheme used by your LDAP server. The attribute name is LDAPSSL and values are:
   0: Encryption not used
   1: SSL
   2: TLS

   Example:
   ```
   LDAPSSL: 2
   ```

   If your LDAP server uses TLS you must specify a file containing the certificate used to encrypt the password. By default, the EsgynDB software looks for this file in $TRAF_HOME/cacerts, but you must specify a fully qualified filename, or set the environment variable CACERTS_DIR to another directory. To specify the file containing the certificate, you set the value of the attribute TLS_CACERTFilename, located in the Defaults section.

   Examples:
   ```
   TLS_CACERTFilename: mycert.pem
   TLS_CACertFilename: /usr/etc/cert.pem
   ```

5. Encryption level
   Some LDAP servers require a known username and password to search the directory of usernames. If your environment has that requirement, provide these "search" values.

   Examples:
   ```
   LDAPSearchDN: lookup@company.com
   LDAPSearchPwd: Lookup123
   ```

There are additional optional attributes that can be used to customize EsgynDB authentication.

You can test the authentication configuration file for syntactic errors using the utility `ldapconfigcheck`. If you have loaded the EsgynDB environment (`sqenv.sh`), then the utility will automatically check the file at `$TRAF_HOME/sql/scripts/.traf_authentication_config`. If not, you can specify the file to be checked.

Example:
```
ldapconfigcheck –file myconfigfile
File myconfigfile is valid.
```

If an error is found, the line number with the error is displayed along with the error. More information on the `ldapconfigcheck` utility can be found in Appendix B.

Note: The authentication configuration file needs to be propagated to all nodes, but there is a script that will do that for you described later. For now, you can test your changes on the local node.

You can test the LDAP connection using the utility `ldapcheck`. To use this utility the EsgynDB environment must be loaded (`sqenv.sh`), but the EsgynDB instance does not need to be running. To test the connection only, you can specify any username, and a name lookup will be performed using the attributes in `.traf_authentication_config`.

Example:
```
ldapcheck  --username=fakename@company.com
User fakename@company.com not found
```

If `ldapcheck` reports either that the user was found or the user was not found, the connection was successful. However, if an error is reported, either the configuration file is not setup correctly, or there is a problem either with your LDAP server or the connection to the server. You can get additional error detail by including the `--verbose` option. To learn more about `ldapcheck`, see Appendix C.

If you supply a password, `ldapcheck` will attempt to authenticate the specified username and password. The example below shows the password for illustrative purposes, but to avoid typing the password on the command line, leave the password blank (`--password=`) and the utility will prompt for the password with no echo.

Example:
```
ldapcheck –username=realuser@company.com  --password=StrongPassword
Authentication successful
```

Generating a Server Certificate

EsgynDB clients such as `trafci` encrypt the password before sending it to EsgynDB. By default, a self-signed certificate is used to encrypt the password. The certificate and key should be generated when the `sqgen` script is invoked. By default, the files `server.key` and `server.crt` are located in `$HOME/sqcert`. If the files are not present, or if you want to use your own certificate, you can manually generate the files or obtain a certificate from a Certificate Authority (CA). To manually generate a new certificate, run the script `sqcertgen` located in `$TRAF_HOME/sql/scripts`. The script runs `openssl` to generate the certificate and key.
To run `openssl` manually, follow the example:

```
openssl req –x509 –nodes –days 365 –subj
‘/C=US/ST=California/L=Milpitas/CN=host.domain.com/O=Some Company/OU=Service
Connection’ –newkey rsa:2048 –keyout server.key –out server.crt
```

Option	Description
-x509	Generate a self-signed certificate
-days <validity of certificate>	Make the certificate valid for the days specified
-newkey rsa:<bytes>	Generate a new private key of type RSA of length 1024 or 2048 bytes.
-subj <certificateinfo>	Specify the information that will be incorporated in the certificate. Each instance in a cluster should have a unique common name(CN)
-keyout <filename>	Write the newly generated RSA private key to the file specified
-nodes	It is an optional parameter that specifies NOT to encrypt the private key. If you encrypt the private key, then you must enter the password every time the private key is used by an application
-out <filename>	Write the self-signed certificate to the specified file

Both the public (`server.crt`) and private (`server.key`) files should be placed in the directory `$HOME/sqcert`. If you do not want to use the HOME directory or if you want to use different names for the private and/or public key files, see Alternate Trafodion Certificate Locations.

Alternate Trafodion Certificate Locations

By default, the private and public key files/certificate used to connect to EsgynDB are located in `$HOME/sqcert` and names `server.key` and `server.crt`. If you want to store the files in a different location and/or use different filenames, EsgynDB supports environment variables to specific the alternate locations or names.

EsgynDB first checks the environment variables SQCERT_PRIVKEY and SQCERT_PUBKEY. If they are set, EsgynDB uses the fully qualified filename value of the environment variable.

You can specify either one filename environment variable or both.

If at least one filename environment variable is not set, EsgynDB checks the value of the environment variable SQCERT_DIR. If set, the default filename `server.key` or `server.crt` is appended to the value of the environment variable SQCERT_DIR.

If the filename environment variable is not set and the directory environment variable is not set, EsgynDB uses the default location (`$HOME/sqcert`) and the default filename.

Managing Users

Kerberos is enabled for installations that require a secure Hadoop environment. LDAP is enabled to enforce authentication for any user connecting to Trafodion. The Trafodion database enforces privileges on the database, database schemas, database objects (table, views, etc) and database operations. Privileges are enforced when authorization is enabled. When LDAP or Kerberos is enabled, authorization is automatically enabled.

To determine the status of authentication and authorization, bring up sqlci and perform "env;".

```
>>env;
---------------------------------
Current Environment
---------------------------------
AUTHENTICATION     enabled
AUTHORIZATION      enabled
CURRENT DIRECTORY  /opt/trafodion/esgynDB-2.3.0
. . .
```

Once authorization is enabled, there are two predefined database users called DB__ROOT and DB__ADMIN. These users are associated with your specified LDAP username that was set up during install. Please connect to the database as one of these users setup required schemas, users, roles, and privileges.

To learn more about how to register users, grant object and component privileges, and manage users and roles, please see the EsgynDB SQL Reference Manual.

12. Securing the installation

The following optional steps can enhance the security of your EsgynDB installation.

Secure Linux

Refer to instructions from Red Hat or CentOS to secure your Linux installation.

OS	Version	Link
RedHat Enterprise Linux	7.x	https://access.redhat.com/documentation/en-us/red_hat_enterprise_linux/7/html/security_guide/index
RedHat Enterprise Linux	6.x	https://access.redhat.com/documentation/en-us/red_hat_enterprise_linux/6/html/security_guide/index
CentOS	Generic	https://wiki.centos.org/HowTos/OS_Protection#head-f80d332aeea03f57d34d7a5c09493a7d69cce177

Secure Hadoop

Refer to instructions from your Hadoop distribution vendor – Cloudera or Hortonworks.

Hadoop Distribution	Link
Hortonworks' HDP 2.x	https://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.4.0/bk_Security_Guide/content/ch_hdp-security-guide-overview.html
Cloudera CDH 5.x	https://www.cloudera.com/documentation/cdh/5-0-x/CDH5-Security-Guide/CDH5-Security-Guide.html

Secure Jetty Server

A number of components in EsgynDB serve web pages using the Jetty web server module. This server is configured to use HTTPS and enabled to use strong SSL ciphers. Refer to the Jetty security documentation.

- Configure Jetty Connectors
- Configure Security

Upgrade passwords

Change the default passwords with stronger passwords wherever applicable. Refer to this section on user-ids and passwords for specific users that are used by EsgynDB.

Secure ports

The following ports are typically required to be opened to external applications

Application	Port Range	Description
DCS Master	23400	Open a range of consecutive ports depending on the number of configured MXOSRVRs

DB Manager	4206

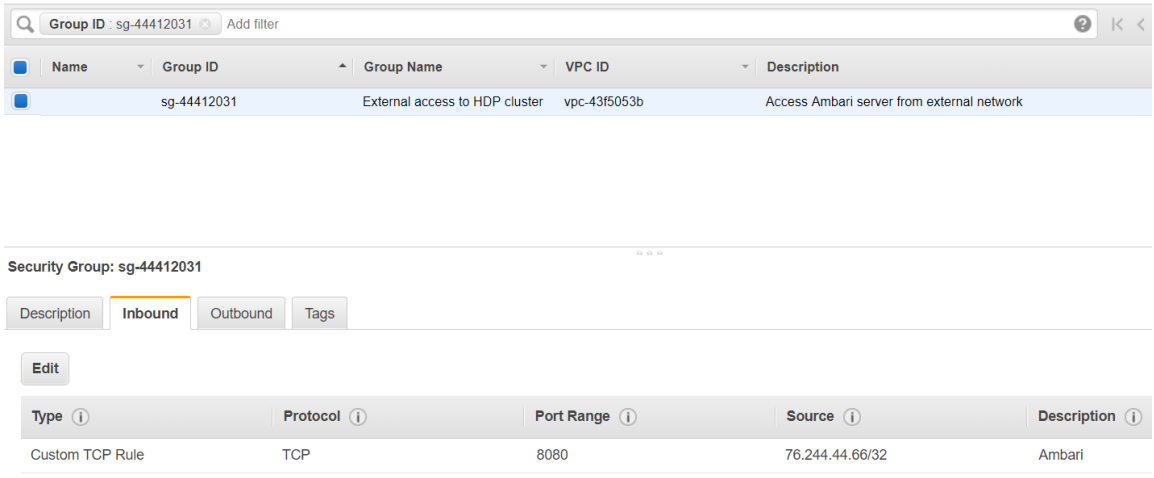
Secure AWS Installation

You can secure your EsgynDB installation on AWS by implementing the following recommendations.

Restrict access to Ambari or Cloudera Manager

Ensure the Hadoop manageability tools such as Ambari or Cloudera Manager are only accessible from a defined set of IP addresses. This may be your client machine or machines from within your corporate network.

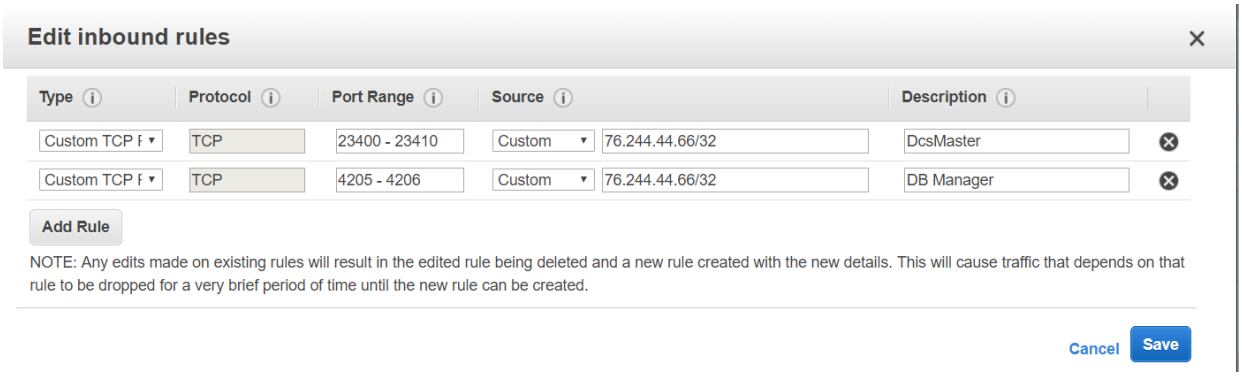
For example, this configuration grants access to Ambari from client 76.244.44.66, and restricts access from other IP addresses.



Restrict access to EsgynDB components

Configure AWS security rules to restrict access to the DCS subsystem and EsgynDB Manager.

For example, this configuration grants access only from client 76.244.44.66



Access within EsgynDB instance

EsgynDB components need to communicate across nodes within the instance, while certain components such as DCSMaster and MXOSRVR processes, need to be available for access from external clients. Configure the rules accordingly, using the following example as a template.

Description | **Inbound** | Outbound | Tags

Edit

Type	Protocol	Port Range	Source	Description
All TCP	TCP	0 - 65535	172.31.0.0/16	
Custom TCP Rule	TCP	23400 - 23410	34.204.77.209/32	saturn001 external...
Custom TCP Rule	TCP	23400 - 23410	54.158.104.187/32	saturn002 external...

Summary

The example screenshot below captures the summary of all inbound rules as set up on an EsgynDB cluster running Hortonworks' HDP. Access to this cluster is available only from the source address and applicable ports, but not from any other IP addresses.

Security Groups associated with i-0066f6deaf70a008a

Ports	Protocol	Source	SSH access	External access to HDP cluster	External access for EsgynDB	Internal subnet access
22	tcp	76.244.44.66/32	✓			
5901	tcp	76.244.44.66/32	✓			
8080	tcp	76.244.44.66/32		✓		
23400-23410	tcp	76.244.44.66/32			✓	
4205-4206	tcp	76.244.44.66/32			✓	
0-65535	tcp	172.31.0.0/16				✓
23400-23410	tcp	34.204.77.209/32, 54.158.104.187/32				✓

Final Steps

Using the AWS console,

- create the Elastic IP
- create the network interface; check the **Allow Re-association** button
- Associate the Elastic IP with network interface
- Create the user ID, and generate the access keys
- Create the policy and attach the policy to the user

Best Practices

Guidance for best practices in setting up IAM, user policies and roles for accessing Amazon's EC2 cloud is available here: http://docs.aws.amazon.com/IAM/latest/UserGuide/best-practices.html

Appendix A: Authentication Configuration File

By default the EsgynDB authentication configuration file is located at `$TRAF_HOME/sql/scripts/.traf_authentication_config`.

Attributes Supported in .traf_authentication_config

This is a list of the attributes supported in the file `.traf_authentication_config`. For each attribute, a description and example is included.

Attribute Name	Purpose	Example Value	Notes
LDAPHostName	Host name of the local LDAP server	ldap.master.com	If more than one LDAPHostName entry is provided, Trafodion will attempt to connect with each LDAP server before returning an authentication error. Also see the description related to RetryCount and RetryDelay entries.
LDAPPort	Port number of the local LDAP server	345	Must be numeric value. Related to LDAPSSL entry. Standard port numbers for OpenLDAP are as follows: Non-secure: 389 SSL: 636 TLS: 389
LDAPSearchDN	If a search user is needed, the search user distinguished name is specified here.	cn=aaabbb, dc=demo, dc=net	If anonymous search is allowed on the local server, this attribute does not need to be specified or can be specified with no value (blank). To date, anonymous search is the normal approach used.
LDAPSearchPWD	Password for the LDAPSearchDN value. See that entry for details.	welcome	None
LDAPSSL	A numeric value specifying whether the local LDAP server interface is unencrypted or TLS or SSL. Legal values are 0 for	0	None

	unencrypted, 1 for SSL, and 2 for TLS. For SSL/TLS, see the section below on Encryption Support.		
UniqueIdentifier	The directory attribute that contains the user's unique identifier.	uid=,ou=Users,dc=demo, dc=net	To account for the multiple forms of DN supported by a given LDAP server, specify the UniqueIdentifier parameter multiple times with different values. During a search, each UniqueIdentifier is tried in the order it is listed in the configuration file.
LDAPNetworkTimeout	Specifies the timeout (in seconds) after which the next LDAPHostName entry will be tried, in case of no response for a connection request. This parameter is similar to NETWORK_TIMEOUT in ldap_conf(5). Default value is 30 seconds.	20	The value must be a positive number or -1. Setting this to -1 results in an infinite timeout.
LDAPTimelimit	Specifies the time to wait when performing a search on the LDAP server for the username. The number must be a positive integer. This parameter is similar to TIMELIMIT in ldap_conf(5). Default value is 30 seconds.	15	The server may still apply a lower server-side limit on the duration of a search operation.
LDAPTimeout	Specifies a timeout (in seconds) after which calls to synchronous LDAP APIs will abort if no response is received. This parameter is similar to TIMEOUT in ldap_conf(5). Default value is 30 seconds.	15	The value must be a positive number or -1. Setting this to -1 results in an infinite timeout.
RetryCount	Number of attempts to establish a successful LDAP connection. Default is 5 retries before returning an error.	10	When a failed operation is retried, it will be attempted with each configured LDAP server, until the

			operation is successful or the number of configured retries is exceeded.
RetryDelay	Specifies the number of seconds to delay between retries. Default value is 2 seconds. See description of RetryCount.	1	None
PreserveConnection	Specifies whether the connection to LDAP server will be maintained (YES) or closed (NO) once the operation finishes. Default value is NO.	YES	None
RefreshTime	Specifies the number of seconds that must have elapsed before the configuration file is reread. Default is 1800 (30 minutes).	3600	If set to zero, the configuration file is never read. The connectivity servers must be restarted for changes to take effect if this value is zero. This attribute is not specific to either configuration and must be defined in the DEFAULTS section.
TLS_CACERTFilename	Specifies the location of the certificate file for the LDAP server(s). Filename can either be fully qualified or relative to $CACERTS_DIR.	cert.pem	This attribute applies to both configurations. If a configuration does not require a certificate, this attribute is ignored. This attribute must be defined in the DEFAULTS section.
DefaultSectionName	Specifies the configuration type that will be assigned to a user by the REGISTER USER command if no authentication type is specified. In the initial Trafodion release, only one configuration is supported.	LOCAL	This attribute must be defined in the DEFAULTS section. If the DefaultSectionName attribute is specified, a section by that name (or equivalent) must be defined in .traf_ldapconfig. Legal values are LOCAL and ENTERPRISE. This

	syntax is likely to change.

Appendix B: ldapconfigcheck Utility

The utility `ldapconfigcheck` validates the syntactic correctness of a EsgynDB authentication configuration file. EsgynDB does not need to be running to run the utility.

```
ldapconfigcheck  [<option>]...
<option> ::= --help|-h : display usage information
             -file <config-filename>
```

Considerations

If the configuration filename is not specified, the tool will look for a file using environment variables. Those environment variables and the search order are:

1. TRAFAUTH_CONFIGFILE
   A fully qualified name is expected.
2. TRAFAUTH_CONFIGDIR
   Filename `.traf_authentication_config/` is appended to the specified directory
3. TRAF_HOME
   `/sql/scripts/.traf_authentication_config` is appended to the value of TRAF_HOME.

Errors

One of the following is output when the tool is run. Only the first error encountered is reported.

Code	Text
0	File *filename* is valid.
1	File *filename* not found
2	File: *filename* Invalid attribute name on line *line-number*
3	File: *filename* Missing required value on line *line-number*
4	File: *filename* Value out of range on line *line-number*
5	File: *filename* Open of traf_authentication_config file failed
6	File: *filename* Read of traf_authentication_config file failed
7	No file provided. Either specify a file parameter or verify environment variables.
8	TLS was requested in at least one section, but TLS_CACERTFilename was not provided
9	Missing host name in at least one section. Each LDAP connection configuration section must provide at least one hostname.
10	Missing unique identifier in at least one section. Each LDAP connection configuration section must provide at least one unique identifier.
11	At least one LDAP connection configuration section must be specified.
12	Internal error parsing . traf_authentication_config.

Appendix C: ldapcheck Utility

The utility `ldapcheck` can be used to validate the EsgynDB authentication configuration and attempt to connect to a configured LDAP server.

```
ldapcheck  [<option>]...
<option> ::= --help|-h            display usage information
             --username=<LDAP-username>
             --password[=<password>]
             --primary            Use first configuration
             --local              Use first configuration
             --enterprise         Use first configuration
             --secondary          Use second configuration
             --remote             Use second configuration
             --cluster            Use second configuration
             --verbose           Display non-zero retry counts and LDAP errors
```

Considerations

- Aliases for primary include enterprise and local. Aliases for secondary include cluster and remote. If no configuration is specified, primary is assumed.
- The equals sign is required when supplying a value to username or password.
- To be prompted for a password value with no echo, specify the password argument but omit the equals sign and value.
- Passwords that contain special characters may need to be escaped if the password is specified on the command line or within a script file.
- If the password keyword is not specified, only the username will be checked. The tool can therefore be used to test the LDAP configuration and connection to the configured LDAP server(s) without knowing a valid username or password.