Wikiprint Book

Title: Pre-requisites

Subject: TracMeteo - ESGFNodeInstallation

Version: 135

Date: 10/28/2021 09:25:08 AM

Table of Contents

This installation guide will provide instructions about how to install an ESGF data/compute node. In order to do it, the VM should have 1 core, 2GB of RAM memory and 20GB of Hard Disk.

For the installation process, it is highly recommendable to provide more than 1 core

Pre-requisites

TCP and UDP firewall configuration

Corporate Firewall

For a detailed info follow ?this link.

Port	Direction	Type	Application	Description
80	in	tcp	Tomcat	Web server access
443	in	tcp	Tomcat	SSL - Secure Web Server Access.
5432	in	tcp	Postgres	Postgres Access. ***(not external: by default bound ONLY TO LOCAL INTERFACE)***
2811	in	tcp	GridFTP	user-configured GridFTP Server control channel
[60000-61000]	in/out	tcp	GridFTP	user-configured GridFTP Server data channel (or as defined in the global variable GLOBUS_TCP_PORT_RANGE)
2812	in	tcp	GridFTP	BDM-configured GridFTP Server control channel. May run together with the user-configured one though **not** recommended - system resource intensive!
[60000-61000]	in/out	tcp	GridFTP	BDM-configured GridFTP Server data channel. May run together with the user-configured one though **not** recommended - system resource intensive!
7512	out	tcp	MyProxy	MyProxy client access to the certificate repository
8984	-	tcp	esgf-search (Tomcat)	local connection to the Solr master instance ***(not external!)***
8983	in/out	tcp	esgf-search (Tomcat)	Connection to remotes Solr slave instance. Used in distributed search (shard).
80	out	tcp	esg-publisher	Local connection to THREDDS server (e.g., to check catalogs) and other nodes (node-manager)
443	out	tcp	esg-publisher	Local secure connection to THREDDS server (e.g., to restart the application) and to the idp

IPTables configuration

Add the rules below to the IPTables configuration file, i.e. `/etc/sysconfig/iptables`

```
-A INPUT -m state --state NEW -m tcp -p tcp --dport 22 -j ACCEPT
-A INPUT -m state --state NEW -m tcp -p tcp --dport 80 -j ACCEPT
-A INPUT -m state --state NEW -m tcp -p tcp --dport 443 -j ACCEPT
-A INPUT -m state --state NEW -m tcp -p tcp --dport 2811 -j ACCEPT
-A INPUT -m state --state NEW -m tcp -p tcp --dport 2812 -j ACCEPT
-A INPUT -m state --state NEW -m tcp -p tcp --dport 8984 -j ACCEPT
-A INPUT -m state --state NEW -m tcp -p tcp --dport 8983 -j ACCEPT
-A INPUT -m state --state NEW -m tcp -p tcp --dport 60000:61000 -j ACCEPT
```

then, restart the IPTables services

```
$ services iptables restart
```

Install RPM packages

First, install the sourceforge RPM repository for the `*ExtUtils*` packages:

```
$ rpm -iv http://dag.wieers.com/packages/rpmforge-release/rpmforge-release-0.3.6-1.el4.rf.x86_64.rpm
```

If `epel` repository doesn't exist in `/etc/yum.repos.d` then:

```
$ wget http://epel.mirror.net.in/epel/6/i386/epel-release-6-8.noarch.rpm
$ rpm -Uvh epel-release-6-8.noarch.rpm
```

Now you can list out the installed repositories using command:

```
$ yum repolist
```

For making sure epel repository is enabled and rpmforge is disabled, run):

```
yum --disablerepo=rpmforge --enablerepo=epel install autoconf automake bison file flex gcc gcc-c++ gettext-devel libtool u
```

Please make sure that the `ntp` package is installed `$ rpm -qa | grep ntp`, otherwise instal it `$ yum install ntp`

Install the ESGF data/compute node

Before running the ESGF Installer script, it is recommended that some directories are backed up. Backup your node before installation: ?https://github.com/ESGF/esgf-installer/wiki/ESGF-Pre-Installation-Backup

The instructions have been provided by the ESGF[1].

First, download and Execute the Bootstrap Script

```
$ cd /usr/local/bin
$ wget -O esg-bootstrap  https://raw.github.com/ESGF/esgf-dist/master/installer/master/esg-bootstrap --no-check-certificat
$ chmod 555 esg-bootstrap
$ esg-bootstrap
```

Second, check installer script version, with:

```
$ esg-node --version
```

The output:

```
Version: v1.7.0-phoenix-release-master
Release: phoenix
Earth Systems Grid Federation (http://esgf.org)
ESGF Node Installation Script
```

In our case, we are going to configure only `data`, `compute` and `index` types:

```
$ script -a -c "esg-node --type data compute index --install" /tmp/esgf_install.log
-------------------------------------------------------
```

You must continue and yes in answers. Then during the installation, you will have to fill in several questionnaires:

```
Welcome to the ESGF Node installation program! :-)

What is the fully qualified domain name of this node? [spock.meteo.unican.es]:
What is the admin password to use for this installation? (alpha-numeric only) []:
Please re-enter password:
What is the name of your organization? [unican]:
Please give this node a "short" name: []: data-unican
Please give this node a more descriptive "long" name []: data-unican
What is the namespace to use for this node? (set to your reverse fqdn - Ex: "gov.llnl") [es.unican.meteo]:
What peer group(s) will this node participate in? (if not sure, use default) [esgf-test]: #important that use "esgf-test"
What is the default peer to this node? [esgf-node1.llnl.gov]: data.meteo.unican.es
What is the hostname of the node do you plan to publish to? [esgf-node1.llnl.gov]: vesgdev-idx.ipsl.jussieu.fr #important
What email address should notifications be sent as? []: meteo@unican.es
Is the database external to this node? [y/N]: #yes
Please enter the database connection string...
(form: postgresql://[username]@[host]:[port]/esgcet)
What is the database connection string? [postgresql://dbsuper@localhost:5432/esgcet]: postgresql://
entered: postgresql://dbsuper@localhost:5432/esgcet
What is the (low priv) db account for publisher? [esgcet]:
What is the pass of db account?:
```

```
Finished processing dependencies for esgcet==2.12.1
Would you like a "system" or "user" publisher configuration:
       ------------------------------------------
       *[1] : System
        [2] : User
       ------------------------------------------
        [C] : (Custom)
       ------------------------------------------
select [1] >

You have selected: 1
Publisher configuration file -> [/esg/config/esgcet/esg.ini]

Is this correct? [Y/n]
```

```
Looking for keystore [/esg/config/tomcat/keystore-tomcat]... (don't see one)...
Keystore setup:
Launching Java's keytool:
store_password = ******
Would you like to use the DN: (OU=ESGF.ORG, O=ESGF) ? [Y/n]:
Using keystore DN = CN=data.meteo.unican.es, OU=ESGF.ORG, O=ESGF
Enter key password for <my_esgf_node>
       (RETURN if same as keystore password):
Re-enter new password:
Do you wish to generate a Certificate Signing Request at this time? [Y/n]
```

```
Please enter the password for this keystore   :
Please re-enter the password for this keystore:
```

```
Create user credentials
Please enter username for tomcat [dnode_user]:
dnode_user
Please enter password for user, "dnode_user" [********]:   73769edbd97410aacfb3560ebb817f882d141517
Would you like to add another user? [y/N]:
```

```
Please Enter the IP address of this host [134.157.179.48]:>

Using IP: 134.157.179.48
Please select the IDP Peer for this node:
       ----------------------------------------
       *[1] : ESGF-PCMDI-9 -> pcmdi9.llnl.gov
        [2] : ESGF-PCMDI   -> pcmdi3.llnl.gov
        [3] : ESGF-JPL     -> esg-gateway.jpl.nasa.gov
        [4] : ESGF-ORNL    -> esg2-gw.ccs.ornl.gov
        [5] : ESGF-BADC    -> cmip-gw.badc.rl.ac.uk
        [6] : ESGF-DKRZ    -> ipcc-ar5.dkrz.de
        [7] : ESGF-PNNL    -> esg1-gw.pnl.gov
        [8] : ESGF-ANL     -> dev.esg.anl.gov
        [9] : ESGF-PCMDI-TEST3 -> esgf-node3.llnl.gov
       ----------------------------------------
        [C] : (Manual Entry)
       ----------------------------------------
select [1] > C
Please enter the IDP Peer's name [ESGF-PCMDI-9] ESGF-TEST
Please enter the IDP Peer's hostname [pcmdi9.llnl.gov] vesgdev-idx.ipsl.jussieu.fr

You have selected: (Manual Entry)
ESGF-TEST -> vesgdev-idx.ipsl.jussieu.fr

Is this correct? [Y/n] Y
```

```
Creating directory /esg/content/thredds/esgcet
INFO       2013-08-02 16:48:46,144 Writing THREDDS ESG master catalog /esg/content/thredds/esgcet/catalog.xml
INFO       2013-08-02 16:48:46,173 Writing THREDDS root catalog /esg/content/thredds/catalog.xml
THREDDS dataset root directories (option=thredds_dataset_roots)
Each entry has the form 'path_identifier | absolute_directory_path':
Current value is:

esg_dataroot | /esg/data

Enter lines, or <RETURN> to end
Add new line:
```

```
# ESGF cronjob BEGIN ###
35 0,12 * * * ESG_USAGE_PARSER_CONF=/esg/config/gridftp/esg-bdm-usage-gridftp.conf /esg/tools/esg_usage_parser
# ESGF cronjob END ###
Is this ok ? [Y/n]Y
```

```
# ESGF cronjob BEGIN ###
35 0,12 * * * ESG_USAGE_PARSER_CONF=/esg/config/gridftp/esg-bdm-usage-gridftp.conf /esg/tools/esg_usage_parser
5 0,12 * * * ESG_USAGE_PARSER_CONF=/esg/config/gridftp/esg-server-usage-gridftp.conf /esg/tools/esg_usage_parser
# ESGF cronjob END ###
Is this ok ? [Y/n]Y
```

```
Server sent 2 certificate(s):

1 Subject CN=vesgdev-idx.ipsl.fr, OU=simpleCA-vesgdev-idx.ipsl.fr, OU=GlobusTest, O=Grid
  Issuer  CN=Globus Simple CA, OU=simpleCA-vesgdev-idx.ipsl.fr, OU=GlobusTest, O=Grid
  sha1    cf f9 20 2b ce a6 bc b0 5d b4 a7 bb 0c 08 18 99 14 47 a6 86
  md5     bd 6d ab cb 0b 75 58 fb 54 52 89 60 8e 1b 44 b8

2 Subject CN=Globus Simple CA, OU=simpleCA-vesgdev-idx.ipsl.fr, OU=GlobusTest, O=Grid
  Issuer  CN=Globus Simple CA, OU=simpleCA-vesgdev-idx.ipsl.fr, OU=GlobusTest, O=Grid
  sha1    06 09 9b cc b6 70 6f 3e 59 00 34 b9 fa 0a ba 87 0b f1 16 10
  md5     0b b0 a3 56 f6 a7 c7 32 7e 35 b5 b9 e3 bb cd 26

Enter certificate to add to trusted keystore or 'q' to quit: [1] > 1
```

After that, you should restart the esg-node:

```
$esg-node restart
```

If you want to re-install it, you have to use the `force` option :

```
$ ./esg-node --type data compute --install --force
```

Generating CSR certificate

based in: ?http://forge.ipsl.jussieu.fr/prodiguer/wiki/ESGF-FR%3Acerts

1) Make a directory named workdir in `/etc/grid-security` and copy `hostkey.pem` of tomcat in this directory

```
$ mkdir /etc/grid-security/workdir
$ cp /esg/config/tomcat/hostkey.pem /etc/grid-security/workdir
```

2) In `workdir` generate the CSR with this configuration file localopenssl.cnf?

```
$ cd /etc/grid-security/workdir
$ openssl req -config localopenssl.cnf -new -subj /O=Grid/OU=GlobusTest/OU=simpleCA-vesgint-idx.ipsl.jussieu.fr/CN=spock.m
```

3)Check if CSR is well generated

```
$ openssl req -text -noout -in spock.meteo.unican.es.csr
```

Index peer configuration

In order to configure the host certificate and CA public key, you have to send the csr file located under `/esg/config/tomcat/` directory to the CA.

```
$/esg/config/tomcat/data.meteo.unican.es-esg-node.csr
```

Then you should put the signed csr under the `/etc/grid-security/` directory.

```
$ /etc/grid-security/data.meteo.unican.es-esg-node-globus.csr.signed.pem
```

And, if the tomcat key is not in /etc/grid-security directory, copy it inside:

```
$ cd /etc/grid-security
$ cp /esg/conf/tomcat/hostkey.pem ./
```

Install the key pair in tomcat. You will be prompted to enter the cacert file; enter the url to the index node cacert.pem:

```
$ cd /usr/local/bin
$ ./esg-node --install-keypair data.meteo.unican.es-esg-node-globus.csr.signed hostkey.pem
Please enter your Certificate Athority's certificate chain file(s):
[enter each cert file/url press return, press return with blank entry when done]
certfile> http://vesgint-idx.ipsl.jussieu.fr/cacert.pem
.................................
.................................
```

Predeterminates password of esg-truststore.ts is 'changeit'. For list the entrys in truststore:

```
 $ /usr/local/java/bin/keytool -v -list -keystore /esg/config/tomcat/esg-truststore.ts -storepass changeit
```

Set auto fetch certs false, otherwise `/etc/grid-security/certificates/*` will be overwritten by `esgf-prod` peer groups certificates

```
$ cd /usr/local/bin
$ ./esg-node --set-auto-fetch-certs false
$ ./esg-node restart
```

Register connects to desired node, fetches and stores their certificate to enable ingress SSL connections

```
$ ./esg-node --register vesgint-idx.ipsl.jussieu.fr
```

```
$ cd /etc/grid-security/certificates/
$ grep vesgint-idx.ipsl.jussieu.fr *
373bd876.signing_policy: access_id_CA      X509         '/O=ESGF/OU=ESGF.ORG/OU=ESGF-vesgint-idx.ipsl.jussieu.fr/CN=ESGF C
373bd876.signing_policy: cond_subjects     globus       '"/O=ESGF/OU=ESGF.ORG/OU=ESGF-vesgint-idx.ipsl.jussieu.fr/*"'
```

This process should fetch the CA cert to `/etc/grid-security/certificates`, If this process don't work you need to manually copy the CA cert. You can get them from ?https://vesgint-idx.ipsl.jussieu.fr/cacert.pem or generate the certificates from esg-truststore.ts, for more see here: Generating Certs From Truststore

Then rebuild the Tomcat's trustsore

```
$ cd /usr/local/bin
$ ./esg-node --rebuild-truststore
```

Finally, you have to update some files as this manual said, which is available here:
?https://github.com/snic-nsc/datanode-mgr-doc/raw/master/ro/Datanodemgr-doc.pdf

Install TERENA certificate

1. Check if checksum are coincident. The two checksum must match.

```
$ openssl x509 -noout -modulus -in <pub_cer>| openssl md5
$ openssl rsa -noout -modulus -in <priv_key> | openssl md5
```

Install it in the esgf node. (password: the same as in THREDDS, certfile:chain.pem?)

```
$ esg-node --install-ssl-keypair <pub_cert> <priv_key>
.................................
```

```
.................................
Please enter your Certificate Athoritys certificate chain file(s):
certfile> chain.pem
.................................
.................................
```

You can check if the chain file is complete with this command:

```
openssl verify -verbose -purpose sslserver -CAfile <full path to chain file> <full path to host cert file>
```

The output should only be the name of the file followed by 'OK'

1. Restart esgf node

```
$ esg-node --restart
```

1. Test it

```
$ openssl s_client -connect data.meteo.unican.es:443 </dev/null 2>/dev/null |openssl x509 -noout -subject -issuer_hash -
subject= /OU=Domain Control Validated/CN=data.meteo.unican.es
d9be2151
issuer= /C=NL/O=TERENA/CN=TERENA SSL CA
notBefore=Jun 12 00:00:00 2014 GMT
notAfter=Jun 11 23:59:59 2017 GMT
```

Statistics for ENES data nodes

Prerequisites

1. Install postgres client

```
$ yum install postgresql
```

1. Install bind utils

```
$ yum install bind-utils
```

Procedure

The procedure (including all the queries) is reported in the is-enes2_statistics.txt? file

1) Create a 'stats' directory (if not exists)

```
$ mkdir stats
$ cd stats
```

1.1.

```
$ psql esgcet -U dbsuper -c "select distinct remote_addr from esgf_node_manager.access_logging" > ip_addresses.txt
```

1.2. Open the file ip_addresses.txt and cleanup the first two rows and the last one at the end of the file

2) Execute geoiplookup0.2.sh? and generate `ip_addr_country.txt`

```
$ ./geoiplookup0.2.sh > ip_addr_country.txt
```

3) Creation of a support table in the esgf_dashboard namespace
3.1.

```
$ psql esgcet -U dbsuper -c "drop table esgf_dashboard.stats_support_table;"
```

3.2.

```
$ psql esgcet -U dbsuper -c "create table esgf_dashboard.stats_support_table(ip varchar(20), country varchar(10), eunoteu
```

4)

```
$ psql esgcet -U dbsuper -e < ip_addr_country.txt
```

5)

```
$ psql esgcet -U dbsuper -c "update esgf_dashboard.stats_support_table set eunoteu='EU' where country in ('FR','GB','SI','
```

6)

```
$ psql esgcet -U dbsuper -c "update esgf_dashboard.stats_support_table set eunoteu='NotEU' where eunoteu is NULL;"
```

7)

```
$ psql esgcet -U dbsuper -c "create table esgf_dashboard.euips as select distinct ip from esgf_dashboard.stats_support_tab
```

8) Execute query stats: Generate a query_stats_[NUMBER].sql for each statistics queries in is-enes2_statistics.txt? and execute:

```
$ psql esgcet -U dbsuper -e < query_stats_1.sql > query_stats_1.result
```

9) Generate a zip

```
$ zip -9r node_stats_data.meteo.unican.es.zip stats
```

Notes:

3.1 and 3.2 aim at providing the same results. 3.1 is the original query for the statistics inherited from IS-ENES1. Since the results were not as we expected, we defined the query 3.2 3.2 is the new query that basically replaces the regexpression with a simple "joining" condition between the access_logging table (node_manager namespace) and the file_version table (publisher)

The same for 4.1 and 4.2

Both at CMCC and DKRZ 3.2 and 4.2 ran fine. On the contrary 3.1 and 4.1 did not provide the expected statistics.

Concerning query number 5:

Get distinct remote locations (by remote address) who have performed at least 1 download from the data node select distinct(remote_addr) from esgf_node_manager.access_logging;

please store the result (list of IPs) into a different file and send it to us. Actually, this is already provided by step n°1 ("ip_addresses.txt"). Please send the "ip_addresses.txt" file to us. We'll then try to get a global picture about all the remote locations connecting to the IS-ENES data nodes.

See Also

- ESGF Data Visibility API
- ESGF Cordex Data Publication
- ESGF-Security