**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 [?ports](#) firewall configuration

**Corporate Firewall**

| 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
```

after that, the ESGF required RPM packages :

```
$ yum install autoconf automake bison file flex gcc gcc-c++ gettext-devel libtool libuuid-devel libxml2 libxml2-devel libx
```

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

**ESGF user configuration**

Fist, add a `esgf` user:

```
$ adduser esgf
...
```

After that, change the password:

```
$ passwd esgf
...
```

To finish, configure the `esgf` user with `sudoers` privileges. Add the following line to `/etc/sudoers` file:

```
esgf    ALL=(ALL)       ALL
```

## Install the ESGF data/compute node

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

Do it as `esgf` user

```
$ whoami
esgf
$ cd /usr/local/bin
$ wget -O esg-bootstrap  http://198.128.245.140/dist/esgf-installer/esg-bootstrap
$ diff <(md5sum esg-bootstrap | tr -s " " | cut -d " " -f 1) <(curl -s  http://198.128.245.140/dist/esgf-installer/esg-boo
$ chmod 555 esg-bootstrap
$ esg-bootstrap --devel
```

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

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

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? [data.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]:
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
What email address should notifications be sent as? []: meteo@unican.es
Is the database external to this node? [y/N]:
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]:
```

```
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] data.meteo.unican.es

You have selected: (Manual Entry)
ESGF-TEST -> data.meteo.unican.es

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:

```
$ sudo ./esg-node restart
```

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

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

## Index peer configuration

Do it as `root` user

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:

```
$ 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
..................................
..................................
```

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

```
$ 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`

Then rebuild the Tomcat's trustsore

```
$ esg-node --rebuild-truststore
```

## Data Publishing

### Configuring a new project for ESGF publication

See the [?ESGF](#) publication reference for details.

The configuration file is a text file, `/esg/config/esgcet/esg.ini`. For this propose, we are going to configure a new project called `cordex`:

```
[initialize]
log_level = INFO
initial_standard_name_table = /esg/config/esgcet/esgcet_models_table.txt

[DEFAULT]
thredds_dataset_roots =
      esg_dataroot | /datasets

project_options =
      cmip5 | CMIP5 | 1
      ipcc4 | IPCC Fourth Assessment Report | 2
      test | Test Project | 3
      cordex | CORDEX Output data | 4

[project:cordex]
categories =
      project               | enum | true  | true  | 0
      domain                | enum | true  | true  | 1
      institute             | enum | true  | true  | 2
      driving_model         | enum | false | true  | 3
      experiment            | enum | false | true  | 4
      ensemble              | enum | false | true  | 5
      model                 | enum | false | true  | 6
      time_frequency        | enum | false | true  | 7
      version               | enum | false | true  | 8
      rcm_model             | enum | false | true  | 9
      rcm_version           | enum | false | true  | 10
      description           | text | false | false | 99
category_defaults =
      domain | EUR-22
      institute | SMHI
      driving_model | ERAINT
      ensemble | r1i1p1
      model | RCA4-v1
      time_frequency | day
dataset_id = cordex.%(domain)s.%(institute)s.%(driving_model)s.%(experiment)s.%(ensemble)s.WRF331G_v02.%(time_frequency)s.
#/datasets/CORDEX/output/EUR-22/UCAN/ECMWF-ERAINT/evaluation/r1i1p1/UCAN-WRF311G/v02/day/pr/20131108/
directory_format = /datasets/CORDEX/output/%(domain)s/%(institute)s/%(driving_model)s/%(experiment)s/%(ensemble)s/%(rcm_mo
domain_map = map(project_id,domain : domain_description)
      cordex | SAM-44 | South America
```

```
       cordex | CAM-44  | Central America
       cordex | NAM-44  | North America
       cordex | EUR-44  | Europe
       cordex | EUR-22  | Europe
       cordex | AFR-44  | Africa
       cordex | WAS-44  | West Asia
       cordex | EAS-44  | East Asia
       cordex | CAS-44  | Central Asia
       cordex | AUS-44  | Australasia
       cordex | ANT-44  | Antarctica
       cordex | ARC-44  | The Arctic
       cordex | MED-44  | HYMEX Mediterranean
       cordex | EUR-11  | High-res. Europe
       cordex | SAM-44i | South America
       cordex | CAM-44i | Central America
       cordex | NAM-44i | North America
       cordex | EUR-44i | Europe
       cordex | AFR-44i | Africa
       cordex | WAS-44i | West Asia
       cordex | EAS-44i | East Asia
       cordex | CAS-44i | Central Asia
       cordex | AUS-44i | Australasia
       cordex | ANT-44i | Antarctica
       cordex | ARC-44i | The Arctic
       cordex | MED-44i | HYMEX Mediterranean
       cordex | EUR-11i | High-res. Europe
       cordex | MNA-44  | Middle East and North Africa
       cordex | MNA-44i | Middle East and North Africa
       cordex | MNA-22  | Middle East and North Africa
       cordex | MNA-22i | Middle East and North Africa
domain_options = SAM-44,CAM-44,NAM-44,EUR-44,EUR-22,EUR-44i,AFR-44,AFR-44i,WAS-44,EAS-44,CAS-44,AUS-44,ANT-44,ARC-44,MED-4
driving_model_options = ERAINT, ECMWF-ERAINT, CCCma-CanESM2, CNRM-CERFACS-CNRM-CM5, ICHEC-EC-EARTH, MIROC-MIROC5, MOHC-Had
ensemble_options = r1i1p1, r12i1p1, r0i0p0
experiment_options =
       cordex | evaluation | no description
       cordex | historical | no description
       cordex | rcp4 | no description
       cordex | rcp26 | no description
       cordex | rcp45 | no description
       cordex | rcp85 | no description
institute_map = map(project_id,model : institute)
       cordex | WRF331G-v02 | UCAN
institute_options = UCAN
las_configure = false
las_time_delta_map = map(time_frequency : las_time_delta)
       mon     | 1 month
       day     | 1 day
       fx      | fixed
       sem     | semi
maps = institute_map, las_time_delta_map, domain_map
model_options = WRF331G-v02
parent_id = wdcc2.cordex
project_handler_name = basic_builtin
rcm_model_options = UCAN-WRF331G
rcm_version_options = v1, v02
thredds_exclude_variables = a, a_bnds, alev1, alevel, alevhalf, alt40, b, b_bnds, basin, bnds, bounds_lat, bounds_lon, dbz
time_frequency_options = day,fx,mon,sem
variable_locate = ps,ps_
variable_per_file = true
version_options = 20131108
```

Then you have to add the project name to the `esgcet_models_table.txt` file

```
$ echo "  cordex | WRF331G-v02 | http://meteo.unican.es | UNICAN WRF3.3.1 Model version, 2.0" >> /esg/config/esgcet/esgce
```

After modifying `esgcet_models_table.txt` and `esg.ini` files, you have to update the data base by executing :

```
$ export ESGINI=/esg/config/esgcet/esg.ini
$ cd /usr/local/uvcdat/1.4.0/bin/
$ ./esginitialize -c
```

- `./esginitialize -d 0`: To remove all tables
- `./esginitialize -c` : Upgrade the database schema to the latest version, and initialize projects, models, experiments, and standard names from configuration files.

## Using the ESGF [?Publisher](#)

First, obtain a digital certificate from an ESGF trusted MyProxy server, and rename it to whatever path you have defined in esg.ini.

Remember, you have to log in a Federation to do it.

```
$ /usr/local/globus/bin/myproxy-logon -s vesgint-idx.ipsl.jussieu.fr -l blancojc -o ~/.globus/certificate-file
```

Finally, run the commands below to parse the cordex project on the local Data Node, ingest it in the local Postgres database, and send it for harvesting to the configured Index Node.

```
$ cd /usr/local/uvcdat/1.4.0/bin
$ ./esgscan_directory -i /esg/config/esgcet/esg.ini --project cordex -o ~/cordex.txt /datasets/CORDEX/output/EUR-22
$ ./esgpublish -i /esg/config/esgcet/esg.ini --project cordex --map ~/cordex.txt --service fileservice
$ ./esglist_datasets -i /esg/config/esgcet/esg.ini cordex
$ ./esgpublish -i /esg/config/esgcet/esg.ini --project cordex --map ~/cordex.txt --noscan --publish --thredds --service fi
```

```
$ cd /usr/local/uvcdat/1.4.0/bin
```