Setting up a virtual testbed for ESGF

Author: Prashanth Dwarakanath, NSC, Sweden

Version: 1.03
February 9, 2018

Contents

1 About this document

This document was created during the ESGF codesprint conducted by NSC/Linköping University, with support from IS-ENES2, and has been revised for use during subsequent workshops. This document and accompanying scripts and sample configuration files etc can be cloned from the following github repo:
https://github.com/snic-nsc/esgfcodesprint2015.git.

Preinstalled virtual machines are made available for testing use and are provided without any guarantee/warranty or liability whatsoever and are **NOT** meant to be used for production deployment. The virtual machines have been setup in a manner that would allow for easy deployment on a laptop/workstation, on a private local network. If the host machine (laptop/workstation) is connected to the internet, these vms would be able to access the internet, but would themselves not be reachable from the internet. This is to allow for testing ESGF deployments, without the need to have the resource on open internet.

2 Prerequisites

Instructions provided in this document describe setting up of a virtual testbed, using VirtualBox, on machines running Linux/Mac OS. Deployments on Windows are also possible, but require the use of the commandline utility to interact with VirtualBox, instructions for which are not provided here.

1. You will need a minimum of 10 GiB of free space on your harddisk, in order to be able to initialize one vm and install all of the ESGF roles on it.
2. You will need 15+ GiB of space in order to be able to setup a standalone virtual ESGF federation.
3. Ensure you have VirtualBox version 4.3 or higher, installed on your machine. You can check this by executing the command `vboxmanage -version`.
4. You'll need VT-x option (Intel Virtualization Technology) enabled in the BIOS.

3 Virtualbox configuration and vm import

1. Execute the following commands

```
VBoxManage hostonlyif create
VBoxManage hostonlyif ipconfig vboxnet0 --ip 192.168.56.1 --netmask 255.255.255.0
VboxManage natnetwork add --netname natnet1 --network "10.1.1.0/24" --enable
VBoxManage natnetwork modify --netname natnet1 --dhcp off
```

2. Now, when you execute the following commands, the outputs should look like the following:

```
[pchengi@direwolf ~]$ vboxmanage list hostonlyifs
Name:            vboxnet0
GUID:            786f6276-656e-4074-8000-0a0027000000
DHCP:            Disabled
IPAddress:       192.168.56.1
NetworkMask:     255.255.255.0
IPV6Address:     fe80:0000:0000:0000:0800:27ff:fe00:0000
IPV6NetworkMaskPrefixLength: 64
HardwareAddress: 0a:00:27:00:00:00
MediumType:      Ethernet
Status:          Up
VBoxNetworkName: HostInterfaceNetworking-vboxnet0

[pchengi@direwolf ~]$ vboxmanage list natnets
NetworkName:    natnet1
IP:             10.1.1.1
Network:        10.1.1.0/24
IPv6 Enabled:   No
IPv6 Prefix:
DHCP Enabled:   No
Enabled:        Yes
loopback mappings (ipv4)
        127.0.0.1=2
```

3. You now need to download the exported vm 'ova' files, which can be imported into VirtualBox. If you can deploy only one vm due to harddisk space constraints, it should be 'esg-idx.demonet.local.ova' you need. You can download the others if you have adequate disk space. A complete installation of ESGF on a freshly imported vm is approximately 9 GiB in size.

4. If you are a currently participating in the NICEST ESGF workshop in NSC, Sweden, the download url is ftp://192.168.2.1/virtualtestbed.
If you are viewing this document outside the workshop, the url is
http://esg-dn2.nsc.liu.se/virtualtestbed

5. Import the virtual machine(s) in this manner

```
[pchengi@direwolf workshop]$ vboxmanage import --options keepallmacs esg-idx.demonet.local.ova
0%...10%...20%...30%...40%...50%...60%...70%...80%...90%...100%
Interpreting /srv/ftp/workshop/esg-idx.demonet.local.ova...
OK.
Disks:
  vmdisk1 27033026560  -1  http://www.vmware.com/interfaces/specifications/vmdk.html#streamOptimized
  esg-idx.demonet.local-disk1.vmdk  -1  -1

Virtual system 0:
 0: Suggested OS type: "Linux_64"
    (change with "--vsys 0 --ostype <type>"; use "list ostypes" to list all possible values)
 1: Suggested VM name "esg-idx.demonet.local"
    (change with "--vsys 0 --vmname <name>")
 2: Number of CPUs: 2
    (change with "--vsys 0 --cpus <n>")
 3: Guest memory: 1024 MB
    (change with "--vsys 0 --memory <MB>")
 4: Sound card (appliance expects "", can change on import)
    (disable with "--vsys 0 --unit 4 --ignore")
 5: USB controller
    (disable with "--vsys 0 --unit 5 --ignore")
 6: Network adapter: orig HostOnly, config 3, extra slot=0;type=HostOnly
 7: Network adapter: orig NATNetwork, config 3, extra slot=1;type=NATNetwork
 8: CD-ROM
    (disable with "--vsys 0 --unit 8 --ignore")
 9: IDE controller, type PIIX4
    (disable with "--vsys 0 --unit 9 --ignore")
10: IDE controller, type PIIX4
    (disable with "--vsys 0 --unit 10 --ignore")
11: Hard disk image: source image=esg-idx.demonet.local-disk1.vmdk,
target path=/home/pchengi/VirtualBox VMs/esg-idx.demonet.local/esg-idx.demonet.local-disk1.vmdk,
controller=9;channel=0
    (change target path with "--vsys 0 --unit 11 --disk path";
    disable with "--vsys 0 --unit 11 --ignore")
0%...10%...20%...30%...40%...50%...60%...70%...80%...90%...100%
Successfully imported the appliance.
```

6. You have now (hopefully!) successfully imported the vm(s) which are clean installs of Centos 6 (minimal), with ESGF prerequisite packages preinstalled.

7. As root, on your laptop/workstation, add the following entries to your `/etc/hosts` file:

```
192.168.56.52   esg-idx.demonet.local   esg-idx
192.168.56.53   esg-idx2.demonet.local  esg-idx2
192.168.56.54   esg-data.demonet.local   esg-data
```

8. Update packages on the vms to the latest available versions, by doing the following.

   a) Power on the vm. Refer to Section 4 for details.

   b) `ssh -l root <vmname>` (password is 'esgftestvm123')

   c) Execute `yum update` to install updates, if any are available.

   d) Update the esgfcodesprint2015 repo:
   `cd /root/esgfcodesprint2015 && git pull`

   e) `shutdown -h now` //on the vm!

Contents

9. Obtaining a snapshot of the vm(s) at this point will allow you to return to the clean 'esgf-ready' state, at any later time. This is very useful if you have to/want to redo installations from scratch. Note that snapshots are best taken/restored with the vm in powered-off state. Obtain a snapshot like this:

```
[pchengi@direwolf ~]$ VBoxManage snapshot esg-idx2.demonet.local take 'esgf-ready'
0%...10%...20%...30%...40%...50%...60%...70%...80%...90%...100%
Snapshot taken. UUID: 22473e9c-dccc-46c5-9c0b-a43fd2e90af6
```

10. Restoring a vm to the state captured by the snapshot is done like this:

```
VBoxManage snapshot esg-idx2.demonet.local restore 'esgf-ready'
Restoring snapshot 22473e9c-dccc-46c5-9c0b-a43fd2e90af6
0%...10%...20%...30%...40%...50%...60%...70%...80%...90%...100%
```

11. The vm(s) are now ready for use.

4 Powering on the vm(s)

1. Simply execute `VboxManage startvm -type headless <vm fullname>`. It should start up within a few seconds.
2. You can do `ping <vm fullname>` to confirm startup.
3. Now, you should be able to **ssh** to it with `ssh -l root <vm fullname>`.

5 Prior to setting up ESGF

1. Create a globus-online account. You should sign-up here: https://www.globusid.org/create
2. Before proceeding further, here's a checklist to check your system environment for basic compliance. This is particularly useful to follow, prior to setting up a production server with ESGF.

5.1 Presetup checklist

1. Node installed with Centos 6 final release, with minimal packages, and all updates installed. nmap should be installed
2. `hostname --fqdn` should be identical to the name that would be visible on the internet, and proper forward and reverse name resolution, i.e. `nslookup` should produce the same value of `hostname --fqdn` and `nslookup` should produce their ip address. **This is key to a clean setup in the real-world, and may be ignored during this workshop.**

3. Clean http/https egress without proxies requiring authentication etc. Failure to achieve this will result in broken installations and wasted time. Simple test:
   `curl https://esg-dn1.nsc.liu.se` should not produce any output. If there is some output indicating requirement for authentication/login etc, the egress is NOT clean, and the situation would need to be resolved with your network admins.

4. nmap should show the following result

```
nmap -p T:80,443,2811,7512 esg-dn1.nsc.liu.se

Nmap scan report for esg-dn1.nsc.liu.se (130.236.100.123)
Host is up (0.00045s latency).
PORT     STATE SERVICE
80/tcp   open  http
443/tcp  open  https
2811/tcp open  gsiftp
7512/tcp open  unknown
```

5. If any port is not open (and `esg-dn1.nsc.liu.se` is up!), there's probably a firewall of some kind, interfering with the network egress, and you need to have it fixed.
6. Ensure that ports 80, 443, and 2811 from their node, should not be firewalled, i.e. should be accessible to the world. An easy way to check this, even before setting up any services is to do the following, to start a listening service on port 443 on the machine, on all network interfaces:

   ```
   ncat -l -p 443
   ```

   Now, on a different machine, preferably one outside your organization do this:

   ```
   ncat <FQDN of your machine> 443
   ```

   It should connect without any error messages, and wait. Typing something into the prompt should result in the same text being displayed on the listening service terminal on the server. Passing a ctrl-c from the client stops the listening server.
7. If you are using the virtual machines, you can run ncat on your vm and test connecting to it from your laptop.
8. Repeat for **80** and **2811** ports. If in the real-world, you are unable to access these ports from outside your organizational firewall, those ports are probably being filtered by the firewall, and you may need to contact your network administrator, to get them opened up.

6 Setting up ESGF

Prior to setting up ESGF, you need to know the version of ESGF you should install. This information is available on the following pages:
https://github.com/ESGF/esgf-installer/wiki/ESGF-Installation-From-scratch

Contents

https://github.com/ESGF/esgf-installer/wiki/ESGF-Installation-Using-Autoinstaller

Use the appropriate links from the links mentioned above, and fetch the bootstrap files for the installation.

```
wget -O esg-bootstrap --no-check-certificate http://distrib-coffee.ipsl.jussieu.fr/pub/esgf/dist/devel/2.6.4/esgf-installer/esg-bootstrap
chmod 555 esg-bootstrap
./esg-bootstrap
```

The execution of the boostrap script sets up key scripts that are required as prerequisites to perform an installation. It writes a file called `/usr/local/etc/esg-autoinstall.conf` which is an expect script. You must edit and populate with valid entries, before executing it. Below are the complete list of options and suggestions on how to fill it, to install the latest stable software, but without federating it with the ESGF federation. **It is very important to not federate a node unless it is a production deployment**. All testing should be done when the software has been configured as a standalone test installation. Only questions with options that need to be changed are mentioned below.

7 Answers for esg-autoinstall

7.1 esg-autoinstall answers for [data index idp compute] installation

These are the answers you'd need in your esg-autoinstall.conf file, for an all-role installation (the type you perform on the vm **esg-idx.demonet.local** while testing. Change the answers as appropriate.

1. FQDN: "esg-idx.demonet.local"
2. SHORTNAME "short" # you could use any other value you please.
3. LONGNAME "long" # you could use any other value you please.
4. ADMINPASS 'testpass1234' # please use something else!
5. ADMINIP '192.168.56.1'# don't change when testing with the virtual testbed. In the real world, you'd need to specify the ip address of the machine used to access certain IP-restricted admin pages
6. DBLOWPASS 'pubtest1234' # passwd for publisher database account (alphanumeric only)
7. GLOBUSUSER 'yourglobus id' #the one you signed up for, at the beginning of this document.
8. GLOBUSPASS 'your globus passwd'

9. ORGNAME ‘orgname’ # for your production node, you can use a more meaninful value.
10. NAMESPACE ‘local.demonet’ # don’t change during the workshop/or while testing with the virtual testbed.
11. DEFAULTPEER ‘esg-idx.demonet.local’
12. PUBLISHNODE ‘esg-idx.demonet.local’
13. EXTERNALIDP ‘n’
14. IDPPEER ‘esg-idx.demonet.local’
15. ADMINEMAIL “your email address”
16. GLOBUS “Y” # set t Y for fresh install, and N for upgrade
17. THREDDS “Y”

7.2 esg-autoinstall answers for [data] installation

These are the answers you’d need in your esg-autoinstall.conf file, for a data-only installation (the type you perform on the vm **esg-data.demonet.local** while testing. Change the answers as appropriate.

1. FQDN: “esg-data.demonet.local”
2. SHORTNAME “short” # you could use any other value you please.
3. LONGNAME “long” # you could use any other value you please.
4. ADMINPASS ‘testpass1234’ # please use something else!
5. ADMINIP ‘192.168.56.1’# don’t change when testing with the virtual testbed. In the real world, you’d need to specify the ip address of the machine used to access certain IP-restricted admin pages
6. DBLOWPASS ‘pubtest1234’ # passwd for publisher database account (alphanumeric only)
7. GLOBUSUSER ‘yourglobus id’ #the one you signed up for, at the beginning of this document.
8. GLOBUSPASS ‘your globus passwd’
9. ORGNAME ‘orgname’ # for your production node, you can use a more meaninful value.
10. NAMESPACE ‘local.demonet’ # don’t change during the workshop/or while testing with the virtual testbed.

11. DEFAULTPEER ‘esg-idx.demonet.local’
12. PUBLISHNODE ‘esg-idx.demonet.local’
13. EXTERNALIDP ‘y’
14. IDPPEER ‘esg-idx.demonet.local’
15. ADMINEMAIL “your email address”
16. GLOBUS “Y” # set t Y for fresh install, and N for upgrade
17. THREDDS “Y”

7.3 esg-autoinstall answers for [data index compute] installation

These are the answers you’d need in your esg-autoinstall.conf file, for an an index node with external-idp (the type you perform on the vm **esg-idx2.demonet.local** while testing. Change the answers as appropriate.

1. FQDN: “esg-idx2.demonet.local”
2. SHORTNAME “short” # you could use any other value you please.
3. LONGNAME “long” # you could use any other value you please.
4. ADMINPASS ‘testpass1234’ # please use something else!
5. ADMINIP ‘192.168.56.1’# don’t change when testing with the virtual testbed. In the real world, you’d need to specify the ip address of the machine used to access certain IP-restricted admin pages
6. DBLOWPASS ‘pubtest1234’ # passwd for publisher database account (alphanumeric only)
7. GLOBUSUSER ‘yourglobus id’ #the one you signed up for, at the beginning of this document.
8. GLOBUSPASS ‘your globus passwd’
9. ORGNAME ‘orgname’ # for your production node, you can use a more meaninful value.
10. NAMESPACE ‘local.demonet’ # don’t change during the workshop/or while testing with the virtual testbed.
11. DEFAULTPEER ‘esg-idx2.demonet.local’
12. PUBLISHNODE ‘esg-idx2.demonet.local’
13. EXTERNALIDP ‘y’

14. IDPPEER 'esg-idx.demonet.local'
15. ADMINEMAIL "your email address"
16. GLOBUS "Y" # set t Y for fresh install, and N for upgrade
17. THREDDS "Y"

8 Script for easy wiping/reinstalling installations during testing

In the `esgfcodesprint2015` repo, under the `directory`, there's a script called `rinseandrepeat.sh`, which blows away all ESGF files and refetches the bootstrap files, to make it easy to quickly obtain a clean system, ready for performing an ESGF installation.

9 Start the installation process

```
script -c '/usr/local/bin/esg-autoinstall' installation.log
```

After this, you'll not be prompted to do anything else, and if there haven't been any transient network issues or other gremlins, your installation should have completed. The installation screen would look like this

```
Finished!...
In order to see if this node has been installed properly you may direct your browser to:
http://esg-idx.demonet.local/thredds
http://esg-idx.demonet.local/esg-orp
http://esg-idx.demonet.local/
http://esg-idx.demonet.local/las

Your peer group membership -- : [esgf-test]
Your specified "default" peer : [esg-idx.demonet.local]
Your specified "index" peer - : [esg-idx.demonet.local] (url = http://esg-idx.demonet.lo
Your specified "idp" peer --- : [esg-idx.demonet.local] (name = ESG-IDX.DEMONET.LOCAL)
Your temporary certificates have been placed in /etc/tempcerts
You can install them by executing this : esg-node --install-keypair /etc/tempcerts/hostc
When promped for the chainfile, specify: /etc/tempcerts/cacert.pem

[Note: Use UNIX group permissions on /esg/content/thredds/esgcet to enable users to be a
 %> chgrp -R <appropriate unix group for publishing users> /esg/content/thredds

        ------------------------------------------------------
        Administrators of this node should subscribe to the
        esgf-node-admins@lists.llnl.gov by sending email to: majordomo@lists.llnl.gov
```

Contents

```
        with the body: subscribe esgf-node-admins
        ---------------------------------------------------

v2.5.17-master-release

Writing additional settings to db. If these settings already exist, psql will report an
INSERT 0 1
Node installation is complete.
Script done, file is installation.log
```

10 Post Installation Configuration