The LOFAR Cluster

This page describes the LOFAR cluster for (potential) users. The cluster is used to store the data from the BlueGene and to run the pipelines that will do the "standard" reduction and calibration of this data. After this pipeline processing, the results are stored in the Lofar Export (staging) Archive.

We welcome authorized users on the cluster. For the time being the users are bound to only a part of the cluster. This part is called a *subcluster*. Various users groups have been appointed to use a specific subcluster.

User Access

Access through portal to Cluster frontend

You can access the Lofar cluster through the portal: "**ssh -X portal.lofar.eu**" We maintain a ssh whitelist, so only known institutes are able to login. Please send an email to grit at astron.nl or h.paas at rug.nl to add your institute or personal IP number. Once logged in, you'll find an empty home directory. You'll have to login at one of the two frontends "lfe001" of "lfe002" using "ssh -X". When you don't know which one, please use "lfe001" by default.

LOGIN environment

See [The Lofar Login Environment](#) page for information on your default login environment. Only the default environment is supported, so if you defer from this, you're on your own..!

Do and Don'ts

DON'T:

- Store data on the `portal.lofar.eu` system where you enter the network; this system has a ridiculously small disk and filling it will prevent other users from doing their work!
- Pump around large amounts of data
- Leave unused data on the systems. Clean it up, so others can make use of the available disk space.

DO:

- Help improve the system by sending suggestions for improvements or new packages to the administrators

LOFAR Cluster layout (brief)

The Lofar cluster is devided into 8 subclusters. Each subcluster is a processing cluster for a specific

commsissioning group. There are 72 compute nodes (named `lcexxx`) and 24 storage nodes (named `lsexxx`) in total. Each subcluster has 9 compute nodes and 3 storage nodes of 24TB raw capacity. The storage nodes have 4 RAID5 partitions of 5.1TB each. Each partition holds a single XFS filesystem. Each filesystem is NFS mounted on all 9 compute nodes in the subcluster (but NOT in other subclusters). So 1 compute node has 12 NFS data volumes of approx. 5.1 Tbyte mounted.

Frontend

A frontend has 2 Intel Xeon L5420 quad core processors, 16GB internal meomory, 2 GbE interfaces and 2TB disks in RAID5 configuration. There are actually two identical frontends: `lfe001` and `lfe002`. Both of them serve a specific group of subclusters. The frontends are used to build the software and regulate the workload on the subclusters.

Processing units

The compute elements have 2 Intel Xeon L5420 quad core processors, 16GB internal meomory, 2 GbE interfaces and 1TB disks in RAID0 configuration. They can be accessed by secure shell and they are grouped.

Storage Units

The storage nodes are HP DL180G5 boxes, having 2 Intel Xeon L5420 quad core processors, 16GB internal memory, 6 GbE network interfaces and 24TB disks. The disks are devided into 4 partitions of 6 disks each, set up in RAID5 configuration. The XFS filesystems are called "/data1" till "/data4".

Available Subclusters

The current subcluster assignment is:

```
lfe001:~> showsub
```

```
This script shows the subcluster definitions
```

```
sub    lce-nodes       lse-nodes       cexec-lce       cexec-lse       In use for:
====   =========       =========       =========       =========       ===========
sub1   lce001-lce009   lse001-lse003   lce:0-8         lse:0-2         production
group
sub2   lce010-lce018   lse004-lse006   lce:9-17        lse:3-5         no power
sub3   lce019-lce027   lse007-lse009   lce:18-26       lse:6-8         imaging group
sub4   lce028-lce036   lse010-lse012   lce:27-35       lse:9-11        no power
sub5   lce037-lce045   lse013-lse015   lce:36-44       lse:12-14       pulsar group
sub6   lce046-lce054   lse016-lse018   lce:45-53       lse:15-17       no power
sub7   lce055-lce063   lse019-lse021   lce:54-62       lse:18-20       developers
group
```

```
sub8   lce064-lce072   lse022-lse024   lce:63-71    lse:21-23    no power
```

The lce-nodes ar ethe compute/processing nodes, the lse-nodes are the associated storage nodes. Only use the storage nodes that are associated with your processing node!

Old Obsoleted Stuff (only here for completeness)

The lioffen processing cluster

Hardware setup

The 10-node post processing cluster has `lioffen` as frontend node. The cluster nodes are called `lioff021` - `lioff030`.

Each node has 2-Gbyte of physical RAM, and another 2-Gbyte swap space. Each node contains two AMD Opteron 2-GHz CPUs with 1 Mbyte cache. They are currently running Ubuntu 7.10 (Gutsy).

Usage

The offline nodes are used for processing the CS1 data products that are stored on the lifs-nodes. To this purpose, all disks of all lifs-nodes are NFS mounted on all offline nodes.

Environment

There is a login environment with some startup-scripts that make it easier to use several available tools. Please see this description with more information.

Disks and NFS mounts

The big disks on the off-line storage nodes (`lifs001` - `lifs012`) are NFS auto-mounted when access is required. See this section below.

On each node is a 360 Gbyte `/data` partition to store and process local data.

Finally, there is the NFS mounted `/home` partition, where the user's home directories are, and the `/app`-mount on `lioffen`, where the centrally stored applications reside.

Applications

Most applications can be found in directory `/app`.

- See the login environment for how to start using any of the software packages installed.

listfen 32-bit cluster

The old 32 bit offline cluster has been removed. Only frontend "listfen" and the 10 lifs storagenodes are still available.

LIFS storage nodes

There are 12 old 1TB storage nodes, lifs001 till lifs012. They are still in use for development and testing. Lifs001 till lifs008 are used by the Observatory to store recent observations. Lifs009 til lifs012 hold datasets that are a result of some kind of processing. We are already observing with the new cluster hardware. The lifs nodes will be removed the end of November.

Offline processing clusters

The disks of the offline storage nodes are NFS mounted on all these offline nodes. These systems are in use for offline processing (i.e. when data has been transfered from the central storage facility).

Offline storage cluster

This cluster acts as a temporary data buffer. Data is copied from the online storage systems to these systems, so that the online systems can be used to store new data. These systems are accessable from the offline processing clusters, allowing post-processing of the data. More information can be found on this page.

Bootleg mechanism

The new cluster is subject to the "bootleg" deployment regime. This service and installation facility was developed by Harm Paas of the CIT department of the RUG University in Groningen. He already had many years of experience with this system at the Computer Science Faculty. One server in the CEP domain is capable of installing many cluster nodes at once in less than 10 minutes. The administration of all nodes is done on this single Bootleg server.

Central administration

The way the CEP cluster has been set up is as follows: We are using Ubuntu Linux LTS-8.04 as Operating System within the whole cluster. Bootleg takes care of all of the server management in the cluster by creating new images, administering servers, keeping track of updates and correct start up of servers by sending them new images when it finds outdated versions at boot time.

Profiles

Bootleg also administers machine profiles: Depending on the function of the server in the cluster it needs other settings and profiles. This management strategy implies that we have to keep track of all programs installed in the Linux system itself, so we will be able to add them to the image. We plan to build fresh images every month (at the maintenance days).

Instantanious changes

We are also able to respond quickly on extra-program demands for system programs. Within bootleg there is a mechanism to roll-out extra programs through the cluster from a central administration point (CAP). From the CAP we order to deploy a program or a settings-change and after 1 minute the whole cluster will be updated with the change. So if you need extra system programs, simply ask, and you get it almost instantly on every server in the cluster!

Application programs

For an application program change or addition there is another method. Of course you develop your programs in your own user space and/or svn repositories. So you have completely control over all of this. Program releases for the cluster itself reside on a central disk. Normally we (as system administrators) are not involved in the application program deployment. Only if a completely new package is added under a new directory name we need to make this available in the cluster by connecting it in the O.S. via a link to the newly created package name. So there is a clear interface between system programs and applications: The connection between them is at a well defined point, the /opt directory is intended for this. The same goes for database content (mysql, postgresql, …). Although it might involve OS program installation and services, the database content itself will never be on the Linux system image but stored on a separate data server.

Again: Simply ask, we can connect your program and make it available cluster-wide.

From:
https://www.astron.nl/lofarwiki/ - **LOFAR** *Wiki*

Permanent link:
https://www.astron.nl/lofarwiki/doku.php?id=public:lofar_cluster&rev=1268138290

Last update: **2010-03-09 12:38**