

Para Cluster

- [Torque implementation at Leiden Observatory](#)
- [Current user groups](#)
- [Things to know](#)
- [Example torque script](#)
- [Usage rules and requests for access](#)



Latest News

The current cluster consists out of para10-para13 and para33-para37. To log into the new cluster, go to **paracluster** or **para** (which will bring you to the head node of the cluster, currently para33, but users should not count on that number to be the same forever).

Currently para10-para13 + para34 are LOFAR specific nodes, while para33, para35-para37 are general purpose nodes.

These nodes are only accessible from the head node using the system-supplied TORQUE + MPI implementation. DO NOT use your own mpi version, don't even have it included in your PATH or LD_LIBRARY_PATH, unless you want strange and unpredictable results! You can however use the [environment modules](#) to switch between MPICH and OpenMPI.

Introduction

Over the past few years the speed of CPU's has not increased according to Moore's law. In fact we are at a hold on about 3 GHz for at least 10 years. To cope with the ever larger requirements for computing power the only option is to increase the quantity of CPU's. This is achieved by the implementation of multiple CPU's on one chip (dual core, quad core, etc) and by the availability of

multi-processor units as centralized resources. In addition, Hyperthreading can be available, which makes it (almost) possible to execute two threads simultaneously on one core.

The STRW computer department anticipated this trend by making available a parallel computer cluster some x years ago. At that time we bought a set of cheap machines to see if the need for a parallel cluster in the Sterrewacht was high enough. This was indeed the case. However, due to the low cost hardware of the current cluster the reliability was low and over time almost half of the machines died.

A newer cluster, based on Opteron 246 (64Bit) 2.6 GHz machines (2 per case) was bought in 2006 and extended with quadcore machines in 2008. That's ancient history, and so, in 2014 a whole new cluster was bought, consisting of "only" 5 machines, each with 4 8-core CPUs with hyperthreading, resulting on 64 (virtual) cores per node.

Dell PowerEdge R820

8x Intel(R) Xeon(R) CPU E5-4620 0 @ 2.20GHz

512 GB RAM

140 GB OS disk, of which 50 GB available for users (/data1)

10 TB RAID Data disk (/data2)

10 Gb network connections

The cluster has been named with computer names that are pine para#, where # is a number. Each of the nodes has two data disks, available as /net/para#/data1 and /net/para#/data2. Please make sure to use those "network names" for the disks, since the other nodes will not be able to find your data if you refer to it as /data1 or /data2. Nodes para33 - para37 have these two disks, nodes para10 - para13 only have one data disk: /net/para#/data1. Putting your data there during calculations will certainly work faster than having data on a remote disk (like on your desktop pc or on the ijssemeer disk server). You can copy data from/to every other machine in the STRW by logging in to the para machines. Each of the para machines 'sees' the normal STRW environment similar to your desktop. Actually, accessing your local desktop disks may slow down your job enormously, and it may even crash the nfs service on your desktop, which is just not configured for the combined access load of more than a few fast cpus. So don't do that!

Use

The cluster used to be free for use. But the time has come (according to your representatives on the computer committee) to implement a queuing system. Our implementation uses the Torque (OpenPBS) queuing system. Most of the para machines are no longer directly accessible. In stead, you log in to paracluster and submit your job to one of the queues from there. More details on [the torque pages](#).

To see the current use of the cluster have a look at [the ganglia WEB page](#).

Our installation of torque is linked to a single installation of MPICH, and is not supposed to cooperate with other MPI installations (or even your own MPICH compilation). if you run into problems, check your PATH and LD_LIBRARY_PATH and remove any references to your own MPI implementations you may use on other computers. (Actually, you may use this opportunity to reconsider whether maintaining your own copy of MPI is worth it, on any of the systems)

Interactive use

There is no node available for interactive jobs, but the `qsub -l` option is now available to submit an interactive job. This may make you wait until the requested resources become available, and inside the interactive login you get on one of the cluster nodes, you will have to use `mpiexec` like in a torque job script. So you are not allowed to login to para34 - para37 and start running jobs outside of the queue. Doing so will mess up the resource tracking of the queueing system and may thus hang up a para cluster node, for instance by over-allocation of RAM. If this is detected, your (non-queue) job will be killed mercilessly

LOFAR

The LOFAR group has payed for one standard node and four additional nodes with more local disk, and therefore they have a separate queue `lofarq`.

Dell PowerEdge R820

8x Intel(R) Xeon(R) CPU E5-4620 0 @ 2.20GHz

512 GB RAM

140 GB OS disk, of which 50 GB available for users (/data)

50 TB RAID Data disk (/data)1

10 Gb network connections

Other groups interested in paying for such a setup can contact the computer group.

From:

<https://helpdesk.strw.leidenuniv.nl/wiki/> - **Computer Documentation Wiki**

Permanent link:

<https://helpdesk.strw.leidenuniv.nl/wiki/doku.php?id=paracluster&rev=1454407331>

Last update: **2016/02/02 10:02**

