Implementation of an affordable computing cluster for pharmacometric analysis

Coen van Hasselt, Ron Keizer, Michel van Benten, Alwin Huitema

Department of Clinical Pharmacology, Netherlands Cancer Institute, The Netherlands

Introduction

This poster describes the development of a dedicated computing cluster for our population analysis group.

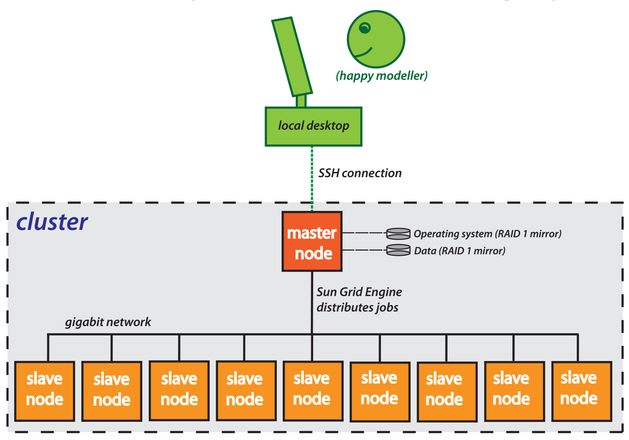
Objectives

The following requirements of the cluster were defined:

- **Central infrastructure**: All modelling software should be installed in a centralized controlled environment, to increase integrity and reproducibility of performed analyses.
- **Computing power**: Sufficient computing power should be available to support and accelerate computationally intensive analyses.
- **Extendibility**: The cluster system should be easily extendible with additional nodes.
- **Costs**: Affordable consumer hardware and preferably open-source software should be used to limit costs.
- **Access**: The modelling environment should be accesible from internal and external networks, preferably over SSH.
- **Robustness**: The system should be robust in terms of up-time and data storage.

Hardware

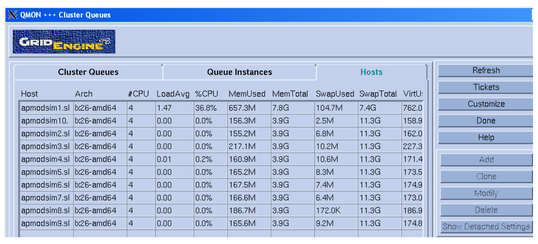
- The computing cluster consists of 1 master node and 9 computing nodes, with a total of 40 CPU cores.
- Node configuration:
  - Intel QuadCore CPU 2.66 GHz (4 cores)
  - 8 GB RAM (master node) or 4 GB RAM (slave nodes)
  - RAID 1 mirrored file-system for data integrity



Schematic representation of cluster setup

Cluster software

- Ubuntu Linux 9.10 Server edition
- Sun Grid Engine 6.2
  - distribution of computing tasks over cluster nodes
  - easy to setup and manage
  - extensive features for cluster, queue and user managment



Sun Grid Engine monitor displaying load of cluster nodes

Cluster development

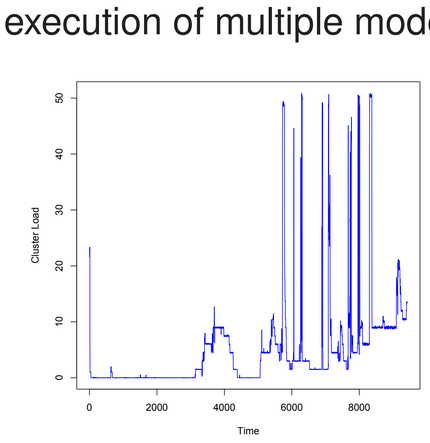
- Total hardware costs of this 40-CPU system was approximately € 4000,-.
- Installation of Ubuntu and Sun Grid Engine software is relatively straight-forward.

Pharmacometric software

- The in-house developed and freely available modelling environment Piraña [1] was installed to allow easy access to the pharmacometric software, to offer integrated access to study data, and for processing of results.
- A range of applications for pharmacometric data analysis were installed.
  - Multiple installations of NONMEM VI and VII [2]
  - PsN [3]
  - R [4]
  - Matlab
- The cluster server also hosts a wiki-based pharmacometrics knowledgebase.

Cluster load

- Sufficient computing power is available to support the modelling work of all users in our group, including computationally intensive tasks (e.g. bootstrap) and parallel execution of multiple models.



Change in clusterload over time

Physical location of cluster nodes

User interaction with the cluster

- Modellers access the cluster through SSH tunneling.
- Pharmacometric applications with a graphical interface, such as Piraña and Emacs/R can also be accessed remotely on the local desktop using SSH-X-tunneling.
- User data can be accessed locally using Samba or S-FTP.
- Piraña was used as interface to various pharmacometric software, but also managment of the cluster queue.
- No Linux-terminal interaction is necessery for end-users.

Conclusion

- The developed computing cluster offers a dedicated and reliable solution for the computational resources needed within our modelling group.
- This project demonstrates the feasibility of the setup of an affordable and scalable cluster in the pharmacometric setting.
- The centralized environment in which applications are installed, controlled and executed, increase regulatory compliance.

References

1. Keizer e*t al*, Comput Meth Program Biomed. (2010)
2. Beal *et al*, NONMEM Users Guides, (1989-98), Icon Development Solutions, Ellicott City, Maryland, USA.
3. Lindbom *et al*, Comp Meth Pr Biomed (2004), 75 (2): 85-94
4. R Development Core Team (2009), URL http://www.R-project.org.