



ISTANBUL KEMERBURGAZ UNIVERSITESI

Information Technology department

The project topic is

Beowulf and Linux clusters

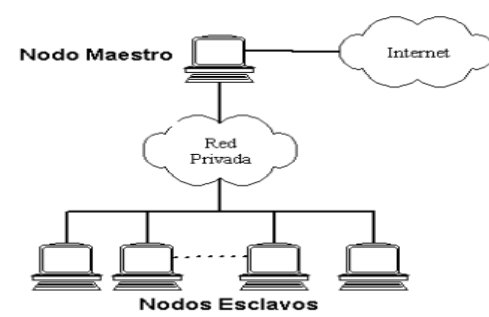
Supervisor: Prof. Dr. Hasan Huseyin Balik

Prepared by: Gaffala Isaa Ali Aborisha (IT) No. 163110448

Email : Gaffala_aborisha@yahoo.com

Outline

- **Introduction**
- **The consists of the Cluster**
- **Cluster specification**
- **Related work**
- **A Beowulf Cluster**
- **The Construction of Beowulf cluster**
- **The Cluster Computer Architecture and Parallel Environment**
- **Cluster Computer Architecture**
- **When we can use Beowulf and Linux clusters**
- **Conclusion**
- **References**



Introduction



- **A computer cluster** is a group of linked computers, working together closely so that in many respects they form a single computer. The components of a cluster are commonly, but not always, connected to each other through fast local area networks. Clusters are usually deployed to improve performance and/or availability over that provided by a single computer, while typically being much more cost-effective than single computers of comparable speed Or availability.
- **A Beowulf Cluster** is a computer design that uses parallel processing across multiple computers to create cheap and powerful supercomputers. In practice is usually a collection of generic computers, either stock systems or wholesale parts purchased independently and assembled, connected through an internal network.

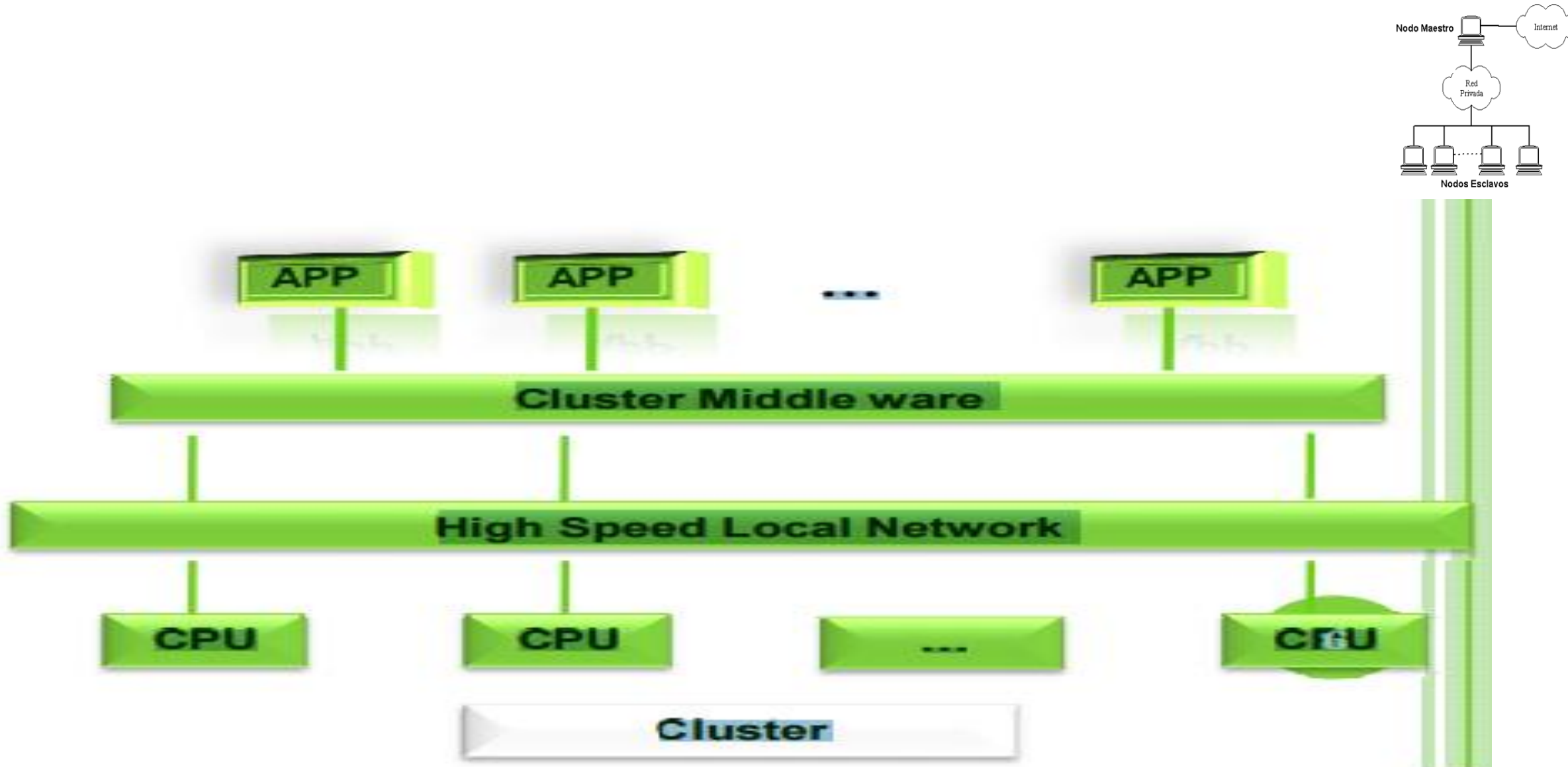


Figure 1. cluster elements

Cluster specification

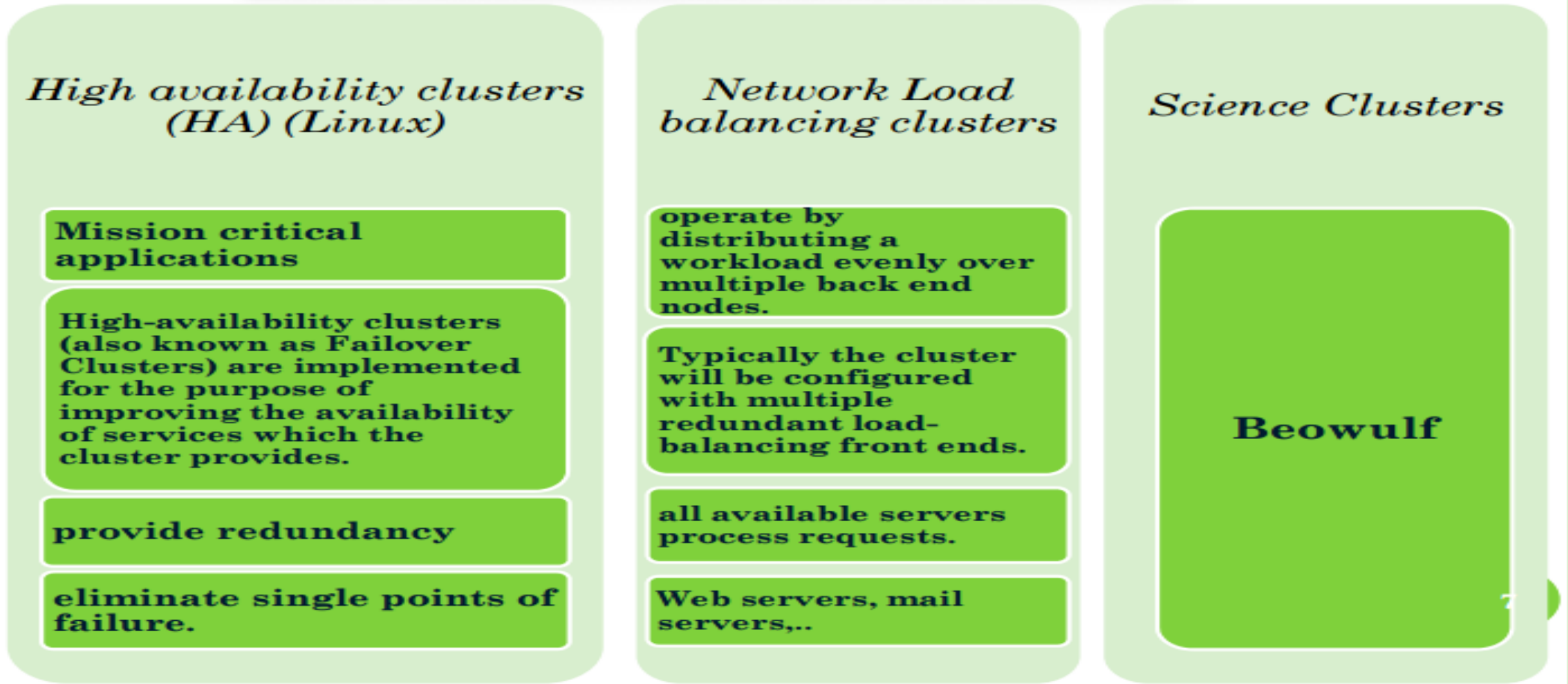


Figure 2. Cluster specification

Related work

- [Ng and Rogers, 2001](#) has declared that A cluster is an accumulation of PCs that coordinate with a specific end goal to take care of an extensive issue. In 1994 NASA started the Beowulf Project at the Goddard Space Flight Center, soon thereafter they delivered the main operational Beowulf group .
- [\(CHAVAN, 2012\)](#) presented that PCs utilized as nodes are typically less expensive and that's only the tip of the iceberg effortlessly updated, such many bunches depend on this design. The greater part of these are running Linux or, on the other hand some variation, as the real option, Windows NT, is excessively restricted and insecure for genuinely extensive scale groups.
- [Ngxande and Moorosi, \(2014\)](#) The High-Performance computing (HPC) parts are exceptionally costly yet with the utilization of business off-the-rack (COTS) equipment parts, the expenses can be dropped down. Beds give modest figuring other option to instructive foundations to play out their examination that need superior PCs.



Fig.3. the Beowulf and Linux clusters

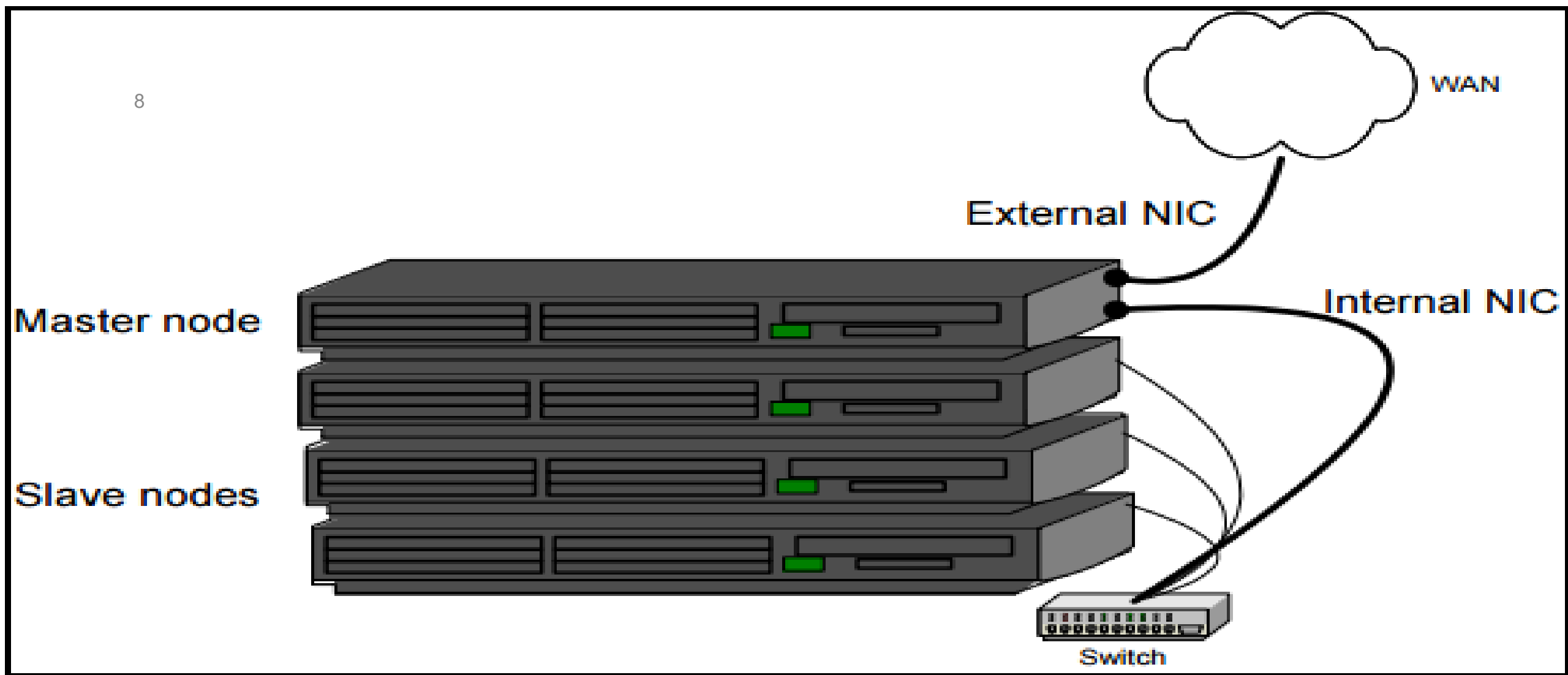


Figure.4. A Beowulf Cluster with Master and Slave Nodes and Network Connections. The master node with its two NIC is in this case the gateway between the two networks the internal and the external. The slave nodes are connected to the master node with a switch in the internal (private) network adapted from (Daillidis, 2004).

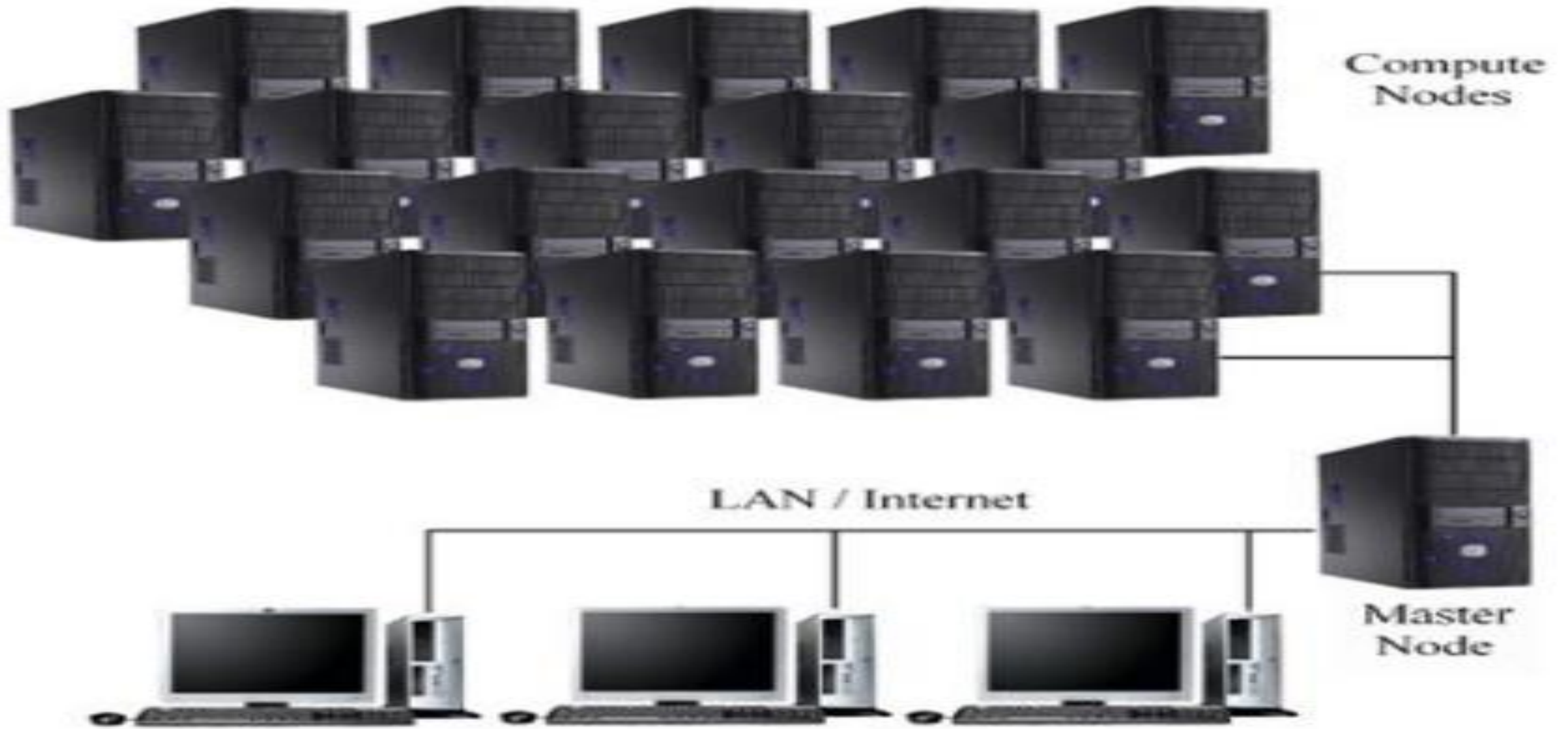
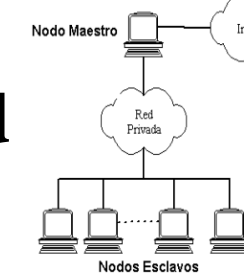


Figure.5.: High Level Beowulf cluster Architecture (Source: Ngxande and Moorosi, 2014).



- **Construction of Beowulf cluster there are four distinct but interrelated areas of consideration:**

Hard ware system structure

resource administration and management environment

Distributed Programming libraries and tools

Parallel algorithms

Fig.6. four distinct area of Beowulf cluster

- **The Construction of Beowulf cluster**

- **Brief Technical Parameters**

- OS: CentOS 5 managed by Rochs-cluster+ Service node: 1 (Intel P4 2.4 GHz)

- Computing nodes: 32 (Intel P4 2.4- 2.8 GHz) and System Memory: 1 GB per node

- Network Platforms: Gigabit Ethernet, 2 cards per node Myrinet 2 G and Language: C, C++, Fortran, java

- Compiler: GNU gcc, Intel compiler, sun Java compiler Parallel Environment: MPICH

- Tools: Ganglia (Monitoring) Pbs - Torque (Scheduler)

- Redhat Linux The most used OS for a Beowulf Cluster. provides High Performance and Scalability / High Reliability / Low Cost (get freely and uses inexpensive commodity hardware) .

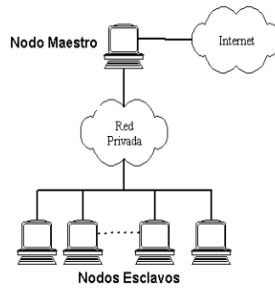
- Network Platform**

- Fast Ethernet (100Mbps)

- Gigabit Ethernet (1Gbps)

- Network structure Bus/Switched

- Maximum bandwidth and Minimum latency.



• Cluster Computer Architecture and Parallel Environment

Two of the most commonly used Parallel Interface Libraries:

- PVM (Parallel Virtual Machine)
- MPI (Message Passing Interface) Parallel Interface Libraries:
- provide a group of communication interface libraries that support message passing. Users can call these libraries directly in their Fortran and C programs.

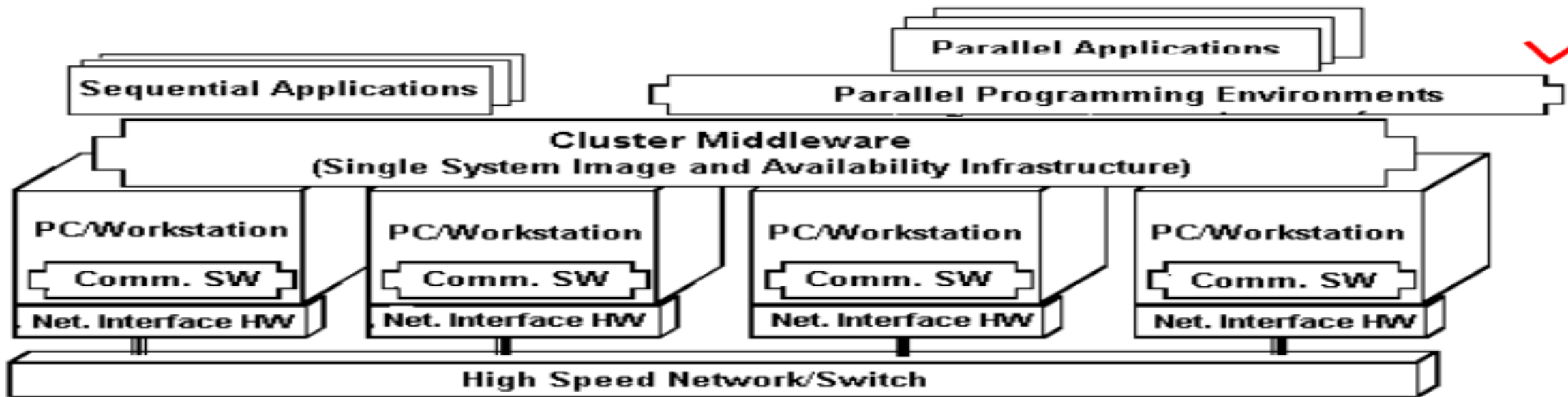
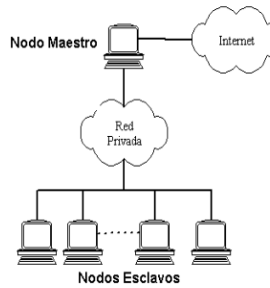
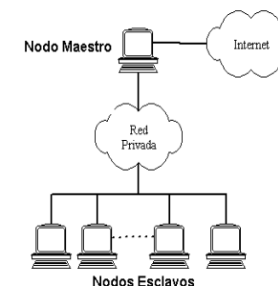


Fig.7. cluster computer architecture

- **The requirements of Building Beowulf cluster**
- A number of [Linux distributions](#), and at least one [BSD](#), are designed for building Beowulf clusters. These include (CHAVAN, 2012).:
- [MOSIX](#), geared toward computationally intensive, IO-low, applications
- [ClusterKnoppix](#), based on [Knoppix](#)
- [Kerrighed](#)
- [Rocks Cluster Distribution](#)
- [Scyld](#)
- [DragonFly BSD](#)
- [Quantian](#), a [live DVD](#) with scientific applications, based on Knoppix and ClusterKnoppix
- [KestrelHPC](#)
- ABC GNU/Linux based on [Ubuntu](#).
- [Kentucky Linux Athlon Testbed](#).
- [PelicanHPC](#), based on [Debian Live](#)



• Process Scheduler

In this case, it is important to compose the scripts and the supporting documents for the scheduler to work. The Scheduler approach is appeared in the following outline.

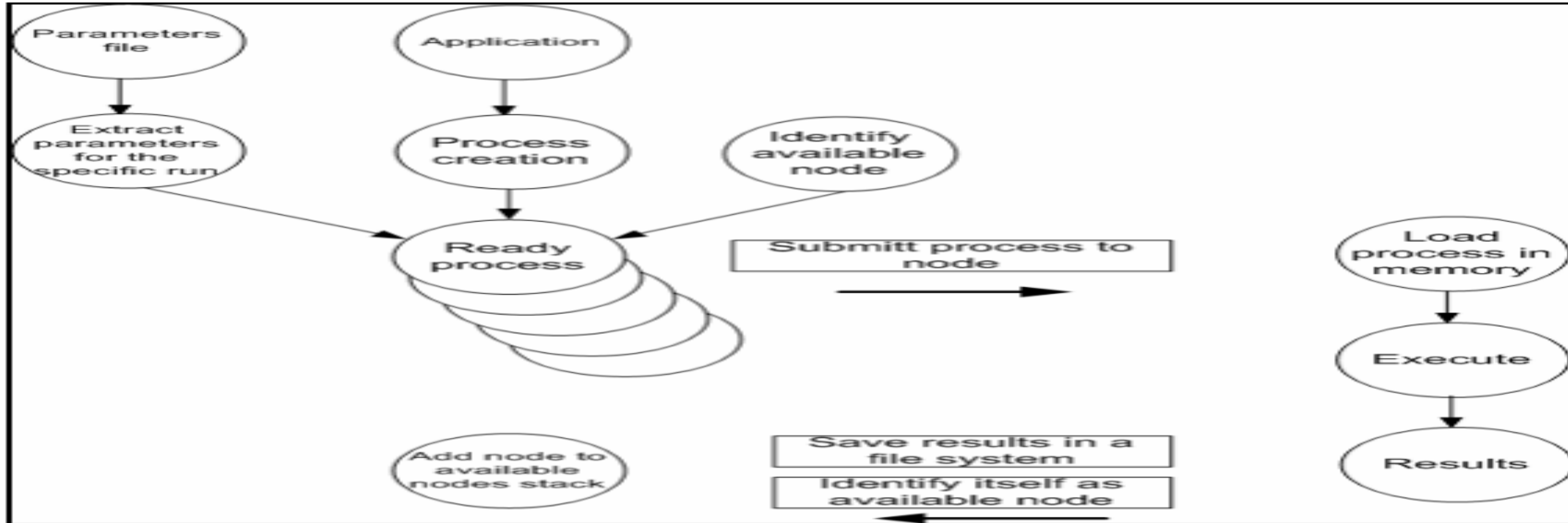
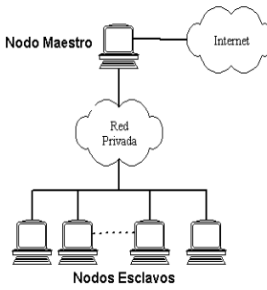


Figure 8. The Scheduler Functionality with the Submission of the Same Job with Different Parameters Every Time adapted from (Daillidis, 2004).

When we can use Beowulf and Linux clusters

- HPC platform for scientific applications.
- Storage and processing of large data.
- Satellites image processing.
- Information Retrieval, Data Mining.
- Computing systems in an academic environment.
- Geologists also use clusters to emulate and predict earthquakes and model the interior of the Earth and sea floor.
- Clusters are even used to render and manipulate high-resolution graphics in engineering.



EXAMPLE OF THE BIGGEST Beowulf cluster

- 1000 nodes Beowulf Cluster System Used for genetic algorithm research by John Coza, Stanford University.



The advantages of Beowulf clusters

17

18-Apr-17

- By using COTS PCs, networks, enables any group to buy and build its own supercomputer.
- Cluster systems offer an alternative to the technical and commercial computing market.
- Users have greater flexibility of configuration.
- Beowulf-class systems exploit mass-market components.
- A Beowulf cluster is that when building a HPC, the performance of the entire cluster is given priority over low cost.
- Let computers networked into a small [local area](#).
- Parallel computing is a way of dividing a large job (Ngxande and Moorosi, 2014).
- Solve larger problems.
- Superior price/performance.

- Faster turn-around time;
- Overcome limits
- Cheap components
- Superior price/performance.
- A Beowulf system's principle attribute is its scalability.
- Parallel computer structure.
- Flexibility of Configuration.
- High Availability.
- Personal Empowerment.
- Development Cost and Time.

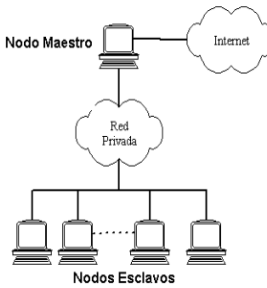
Disadvantages of Beowulf clusters

- Complex software models.
- Difficult development.

➤ CONCLUSION

- In this presentation , this study of Beowulf and Linux clusters can be found in screening the collected commodity clusters from the electronic databases.
- It presents the main points in this study and is essential for many of the organizations who have an interest in reaping the best benefits of using Beowulf and Linux clusters in their business.
- The main contribution of this is study the impact and the main benefit of using or mis using Beowulf and Linux clusters in different type of the organizations or project domains .
- Finally, Beowulf cluster computing is a fascinating microcosm of parallel processing, providing hands-on exposure and experience with all aspects of the field, from low-level hardware to high-level parallel algorithm design and everything in between. While many solutions are readily available to provide much of the necessary services required for effective use of Beowulf clusters in many roles and markets, many challenges remain to realizing the best of the potential of

References



- [1] James P Womack, Daniel T Jones, and Daniel Roos. A machine that changed the world. Simon and Schuster, 1990.
- [2] Poppendieck, M., & Poppendieck, T. (2006). Implementing Lean Software Development: From Concept to Cash. Pearson Education.
- [3] M. Poppendieck and T. Poppendieck, Lean Software Development: An Agile Toolkit. Addison-Wesley Professional, 2003.
- [4] Rodríguez, P., Markkula, J., Oivo, M., & Garbajosa, J. (2012). Analyzing the drivers of the combination of lean and agile in software development companies. Product Focused Software Process Improvement, 145-159. Springer Berlin Heidelberg.
- [5] Fabian Fagerholm , Marko Ikonen , Petri Kettunen , Jürgen Münch , Virpi Roto , Pekka Abrahamsson,” Performance Alignment Work How software developers experience the continuous adaptation of team performance in Lean and Agile environments”, 2015 Elsevier
- [6] Carsten Ruseng Jakobsen ,Tom Poppendieck,” Lean as a Scrum Troubleshooter”, 2011, 2011 Agile Conference, DOI 10.1109/AGILE.2011.11, IEEE
- [7] Hayata, T., Han, J., & Beheshti, M. (2012). Facilitating Agile Software Development with Lean Architecture in the DCI Paradigm. 2012 Ninth International Conference on Information Technology- New Generations. 343 – 348
- [8] Ebert, C., Abrahamsson, P. & Oza, N. (2012). Lean Software Development. IEEE computer society. 22 – 25

How to build Beowulf and Linux clusters



<https://www.youtube.com/watch?v=gvR1eQyxS9I&t=305s>

***Thank You for Your
Attention***

