BOINC - an approach to grid (distributed) computing

Centre de Recherche Public Henri Tudor

Merging Computing Resources
Basic concepts and history

LinuxDays 2007 - marc.seil@tudor.lu

1

Parallel computing

Roots

The need for parallel computing was initially motivated by scientist and the research domain.

Why

- demanding and never satisfied performance and computing hunger

2

Parallel computing

Solution to satisfy

- Use faster dedicated hardware (vector computers, mainframes)
- Use improved algorithms
- Use available idle resources

Where to find idle resources?

- used 40h and 128h idle a week (5.3 days)
- idle 75 % (calculate the ROI :( )
- during the working time slices the load is below 1

3

BOINC - an approach to grid (distributed) computing

Centre de Recherche Public Henri Tudor

Parallel computing

Approaches to merge computing resources

- cluster
- distributed computing

[grid, Grids, Grid-Computing, GRID...]

4

Cluster solutions

Fail-Over Clusters

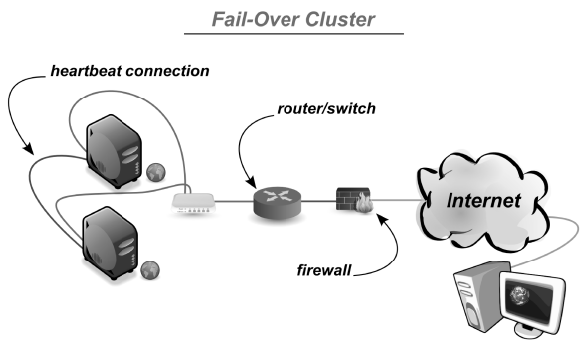
This clusters are used to guarantee services if one of the server fails. (web, email, spam-filters,....)

Basic Set Up

- two machines which are connected to each other by a heart-beat connection (monitoring, watchdog)
- An interface to provide the service.

5

Cluster solutions



6

marc.seil@tudor.lu

BOINC - an approach to grid (distributed) computing

Centre de Recherche Public Henri Tudor

Cluster solutions

Load-Balancing Clusters

This cluster type distributes running processes to the nodes of the cluster with an appropriate spare load level. This distribution of processes is called load balancing.

Basic Set Up

- multiple machines which are connected to each other (eg. switch or hub).
- one front node which has the user interface and where the applications are started (SSI - Clusters)

example : openmosix cluster

7

Cluster solutions

High Performance Computing (HPC) Clusters

This clusters are optimized towards a single application which uses parallel processing libraries (MPI, PVM). The application is split up to the cluster nodes where they are processed in parallel (SMP)

Basic Set Up

- multiple machines which are connected to each other (e.g. dedicated network).
- one front node which has the user interface and where the application is started

example : beowulf class cluster

8

Cluster Solutions

openMosix

an SSI cluster

www.openmosix.org

9

marc.seil@tudor.lu

BOINC - an approach to grid (distributed) computing

Centre de Recherche Public Henri Tudor

openMosix Cluster

History

- the idea was born in 1977
- the first mosix x86 implementation in 1999
- Mosh Bar started in February 2002 the openMosix project (GPL)
- actual version is based on a modified 2.4 linux kernel
- development is still in progress

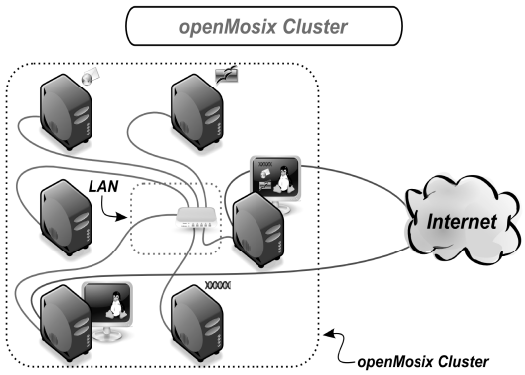
Basic set up

- multiple off the shelf PCs which are connected to each other through network
- each node (PC) of the cluster has a linux distribution running with a patched openMosic 2.4 Linux kernel and a shared filesytem
- one front node provides a user interface.

ref http://www.openmosix.org

10

openMosix Cluster



11

Cluster Solutions

Beowulf Class Cluster

www.beowulf.org

12

marc.seil@tudor.lu

BOINC - an approach to grid (distributed) computing

Centre de Recherche Public Henri Tudor

Beowulf Class Cluster

History

- the name “beowulf” was firstly used in the NASA group
- in 1993 Donald Becker and Thomas Sterling began to outline a cluster-system based on off the shelf hardware
- first milestones were the MPI and PVM library to name only some of them
- actual versions are based on the Linux kernel (channel bonding)
- development is still in progress

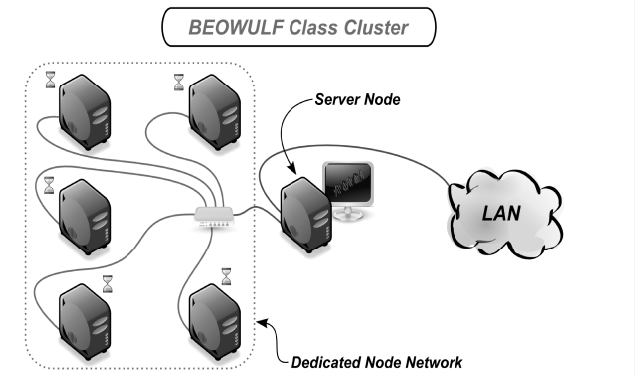
Basic set up

- multiple off the shelf x86 based PCs are connected to each other through a dedicated network
- the different nodes are in most cases based on identical hardware and os
- one front node where the cluster dedicated application is started

ref http://www.beowulf.org/

13

Beowulf Class Cluster



14

distributed computing

About

In this method the different parts of a computing job are distributed to computers (or other processing nodes). After processing the assigned tasks (work units), the nodes will send back the results.

Basic set up

- different hardware and software architectures exist
  - p2p
  - central dispatching unit

15

marc.seil@tudor.lu

BOINC - an approach to grid (distributed) computing

Centre de Recherche Public Henri Tudor

Distributed computing vs cluster

Characteristics

distributed computing	cluster environment
* loosely coupled nodes	* tightly coupled nodes
* no SSI	* SSI
* decentralized job manag.	* centralized job management
* decentralized scheduling	* centralized scheduling

Distributed computing vs cluster

Application domain

distributed computing	cluster environment
* segmentation of the dataset	* no segmentation of dataset
* no node communication	* heavy node communication
* no quality of service	* quality of service

grid, Grid, GRID ....

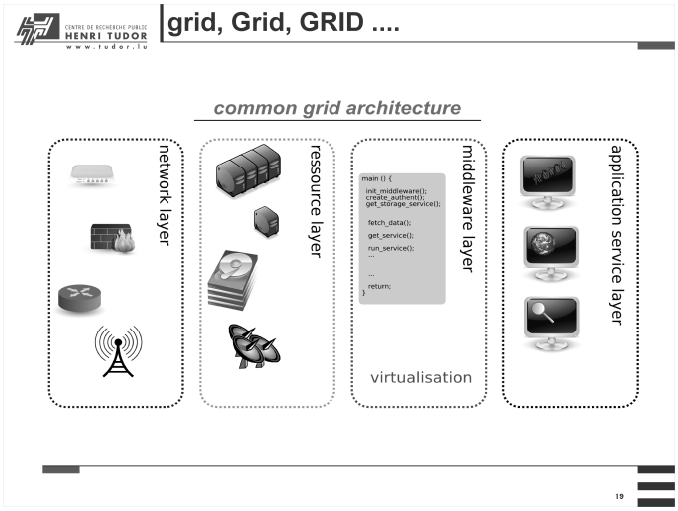
Basic Outline grid, Grid, ...

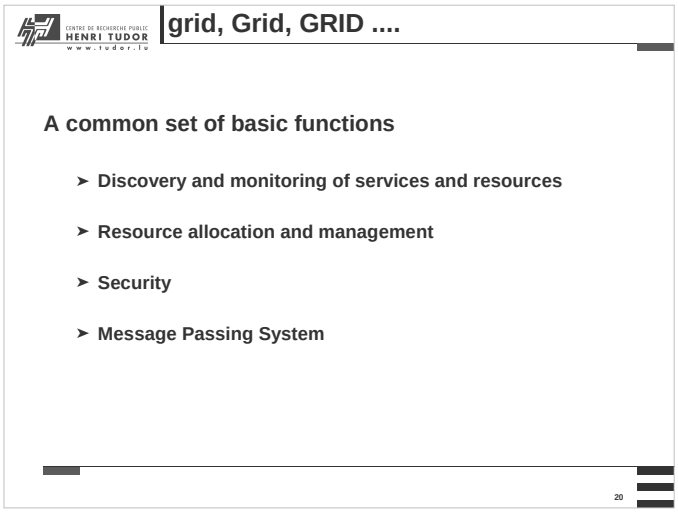
- Grid computing is a form of distributed computing where the physical location of the resources are spread. (storage, computing, applications ...)
- in most cases these resources can be allocated, accounted and used through services provided by a middle ware layer. In this context resource virtualisation plays an important role.
- the basic idea behind Grid approaches is analogue to the power line grids.

LinuxDays 2007
marc.seil@tudor.lu

BOINC - an approach to grid (distributed) computing

Centre de Recherche Public Henri Tudor





marc.seil@tudor.lu