The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

1967-28

Advanced School in High Performance and GRID Computing

3 - 14 November 2008

Rocks: what is and how to install it

ZULUAGA CALLEJAS Jorge Ivan
Universidad de Antioquia Fac. de Ciencias Exactas Y.Nat
Departamento de Fisica, Apartado Aereo 1226
Calle 67 No. 53-108
Medellin
COLOMBIA

Advanced School in High Performance and Grid Computing
ICTP, Trieste, November 3-14, 2008

Rocks and Rolls

HPC Clusters made easy

Jorge I. Zuluaga, Dr.
zuluagajorge@gmail.com
Regional Advanced Simulation and Computer Center, ***CRESCA***
Sede de Investigación Universitaria, SIU
Institute of Physics, FCEN, Universidad de Antioquia
Engineer School, Universidad de San Buenaventura

CRESCA
Centro Regional de Simulación y Cálculo Avanzado

SIU
Sede de Investigación Universitaria

UNIVERSIDAD DE ANTIOQUIA
1803

UNIVERSIDAD DE SAN BUENAVENTURA
MEDELLÍN

Outline

- Cluster installation: **the challenges**
- Presenting **Rocks and Rolls**
- The **Life Cycle** of a Rocks Cluster
- Rocks Clusters ***pros*** & ***cons***
- Let's **Rocks**!

Cluster Installation: The Challenges

- Installation of a cluster is a always a time consuming and challenging task

  … for the newbie and unexperienced admin. it could be also very intimidating

- You have to:
  - Get the proper hardware…

    … or conform yourself with what you already have!
  - Put all together: cluster assembly…

    … trying to not harm the hardware or yourself
  - Choose the right software…

    … operating system, distro, vesion, etc.
  - Install and configure the cluster “head”…

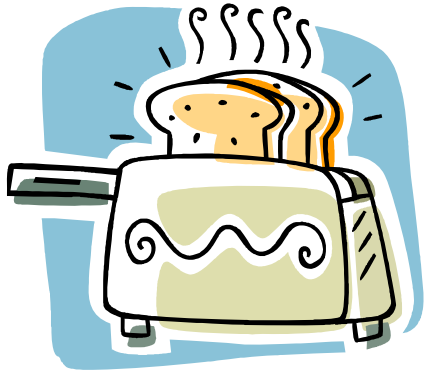
    … cluster frontend, core services
  - Install the compute nodes/storage components…

    … if you are lucky to have the latter!
  - Configure the whole system to exploit the best of it…

    … fine-tuning the “beast” (endless process!)

Cluster Management: The Challenges

- (*not so un*)common requirements:
  - Use **already available** machines and network resources
  - Use **state-of-the-art** compilers, libraries and field specific software
  - Make the Installation+configuration process as **fast** as possible
  - Design/build the installation+configuration process **scalable**
  - Build a platform able to really **speed up** your applications or increase your effective computer power



Cluster Management: The Challenges

- Cluster installation challenges:
  - **Challenge 1**: The base system

    ¿which is the proper distibution (kernel version, rpm/deb, ...)? ¿which tools should I install for HPC? ¿how can you install more components on the way?
  - **Challenge 2**: The “golden disk image” effect

    Golden disk image: OS image/installation description to “broadcast” in your cluster, ¿what about heterogeneity?, ¿what about transfer times? ¿what about (re)configuration?
  - **Challenge 3**: The Blind Installation

    “Interfaceless” nodes, ¿KVM? ¿manual switch? ¿installation scalability issues?

Cluster Management: The Challenges

- Cluster installation challenges (cont'd):
  - **Challenge 4**: Cluster nodes synchronization

    Version of libraries/applications, synchronization of configuration files, new packages installation, software and upgrades
  - **Challenge 5**: The Authentication Bottleneck

    How to avoid bottleneck in the centralization of configuration information
  - **Challenge 6**: Cluster Customization

    Dedicated clusters are unique, avoid the installation of unnecesary software and services

Presenting Rocks & Rolls

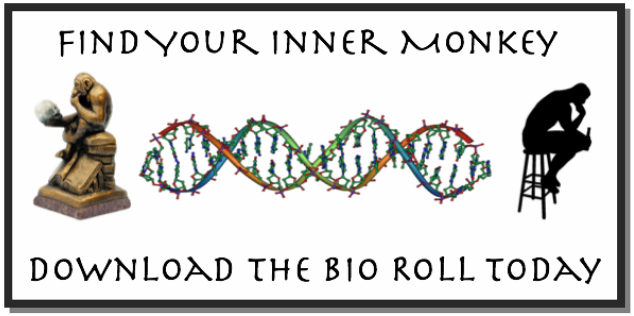
- Rocks:
  - In the surface Rocks is…
    - A Linux Distribution + HPC toolkit + scientific applications for HPC systems
    - Operating system is now Based on Redhat (Rocks 5.x – CentOS : RHEL 5.x), In the future: Solaris also
  - Rocks:
    - A toolkit to build clusters
    - Programmatic approach to build extensible HPC systems
  - It has been developed in the San Diego Super Computer Center (SDSD)
  - Main goal of Rocks: **“make clusters easy”**
    - Easy to deploy, manage, upgrade and scale
    - “Deliver computational power of clusters to a wide range of scientific users”

Presenting Rocks & Rolls

- Rolls:
  - In the surface:
    - The building blocks of the cluster software
    - The main customization mechanism of Rocks
    - Packages selected to create specialized HPC systems
  - What they actually are:
    - Set of binary packages +

      relationships between roll packages and rest of cluster pack.

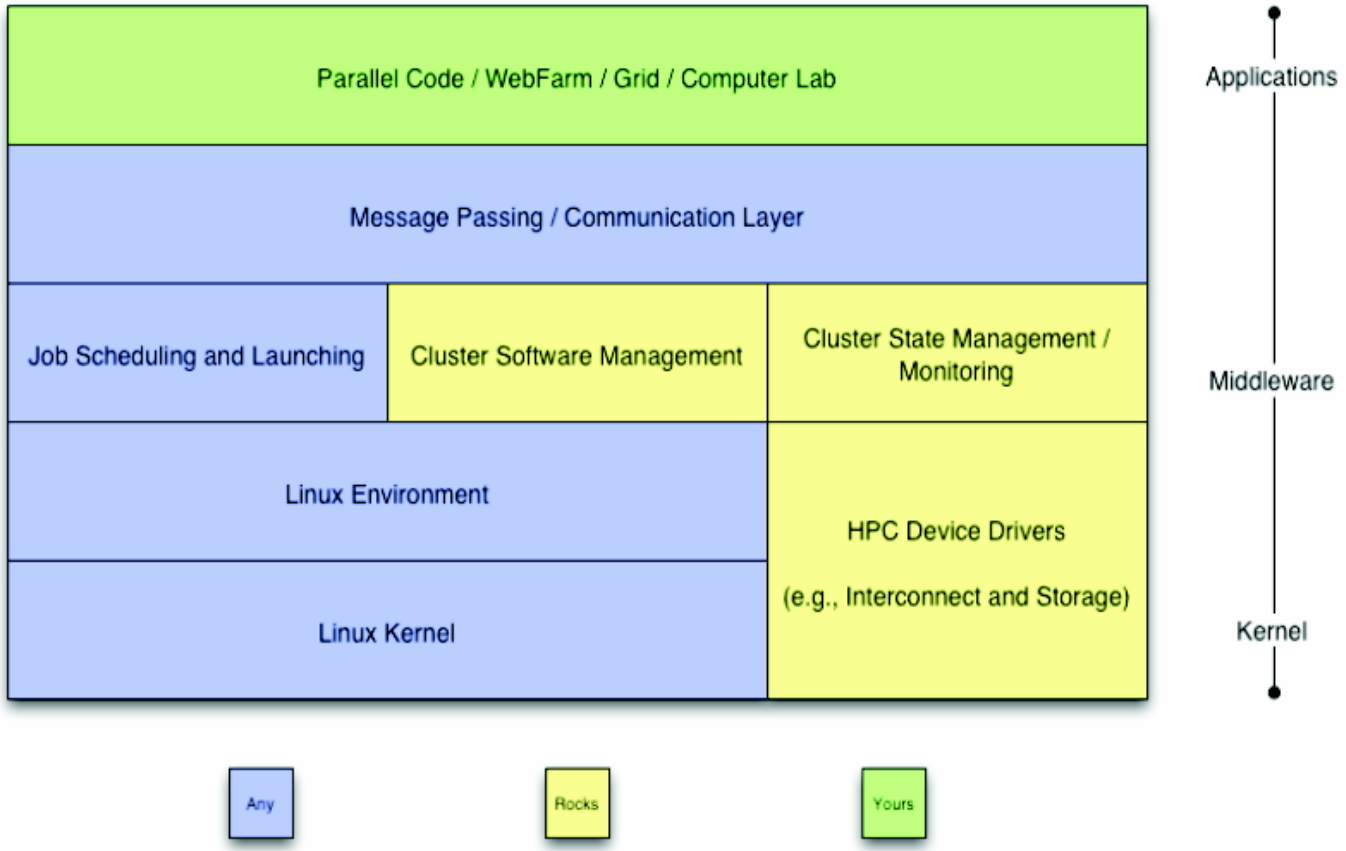
      … configuration subgraph
    - Provides the extensivility characteristic of Rocks
    - Everything is in a Roll*: from the OS to scientific applications*



Presenting Rocks & Rolls

¿What is provided with Rocks?

Taken from "Rocks-a-Palooza", 2005, Slides Sesion 2, IntroRocks.

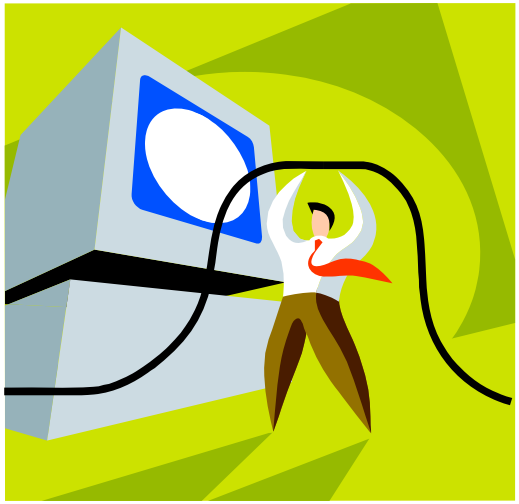


Presenting Rocks & Rolls

- Rocks & Rolls as a solution …
- **Challenge 1: The Base System**
  - Rocks kernel and OS rolls provides all the basic packages for the OS installation
  - HPC Roll contains a set of standard tools for HPC (PBS, Maui, SGE, MPICh, etc.)
  - The base system can be custom installed and configured by selecting the appropiate set of Rolls

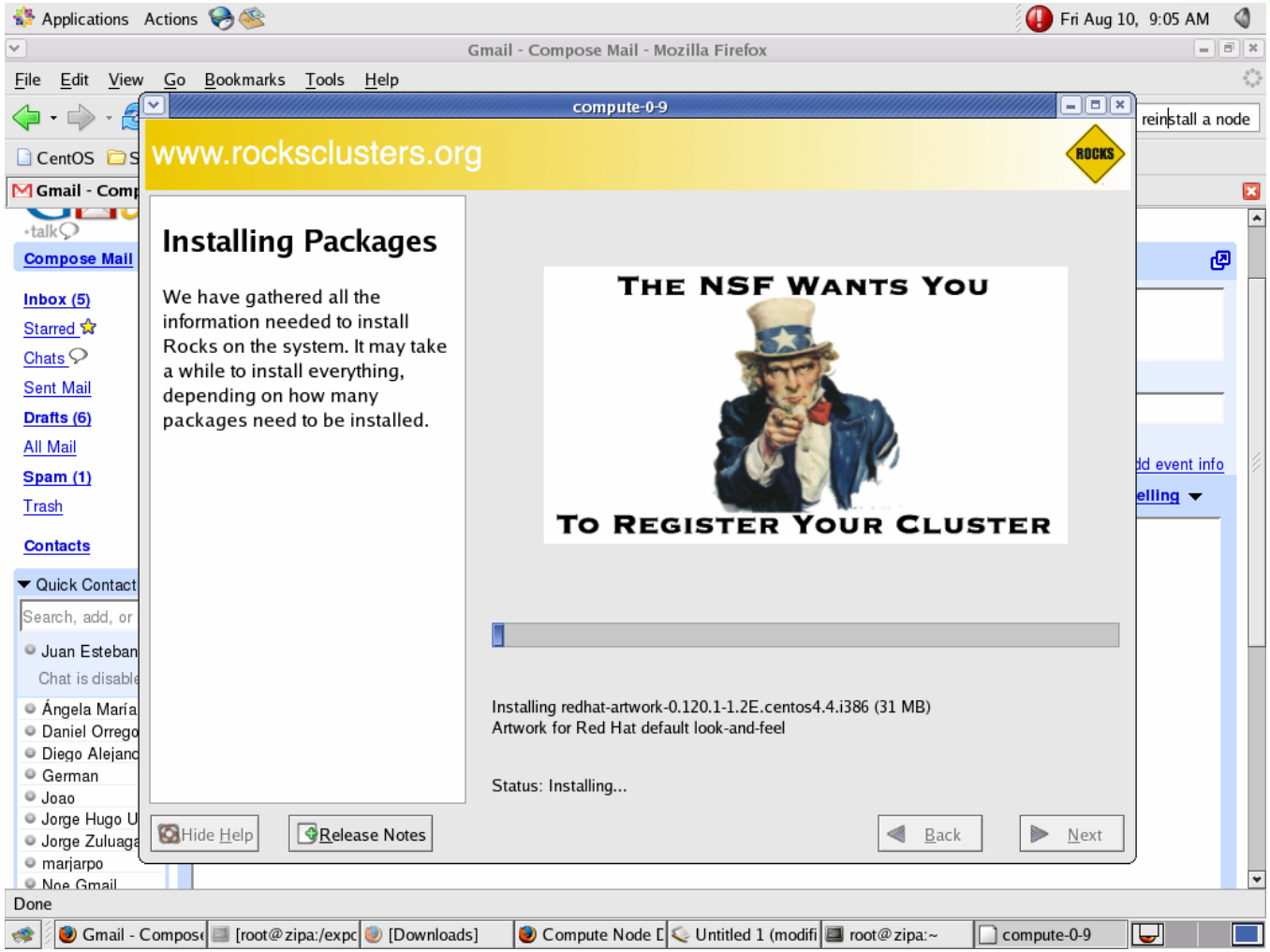
Presenting Rocks & Rolls

- **Challenge 2: The Golden Image**
  - No images at all
  - Redhat KickStart mechanism
  - Obtain packages from the frontend or other nodes while they are being installed (Avalanche Installer)
  - Hardware autodectection
  - Post-configuration scripts
  - Dynamic module loading



Presenting Rocks & Rolls

- **Challenge 3: The Blind Installation**
  - When an appliance is first booted the frontend is notified.
  - Each appliance (specially compute nodes) starts a VNC server during installation in order to monitor the process.
  - Installation of tens of nodes could be achieved at the same time.
  - No KVM or any interface at all.


Applications Actions
Fri Aug 10, 9:05 AM
Gmail - Compose Mail - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
compute-0-9
www.rocksclusters.org
ROCKS
reinstall a node
Installing Packages
We have gathered all the information needed to install Rocks on the system. It may take a while to install everything, depending on how many packages need to be installed.
THE NSF WANTS YOU
TO REGISTER YOUR CLUSTER
Installing redhat-artwork-0.120.1-1.2E.centos4.4.i386 (31 MB)
Artwork for Red Hat default look-and-feel
Status: Installing...
Hide Help
Release Notes
Back
Next
Compose Mail
Inbox (5)
Starred
Chats
Sent Mail
Drafts (6)
All Mail
Spam (1)
Trash
Contacts
Done

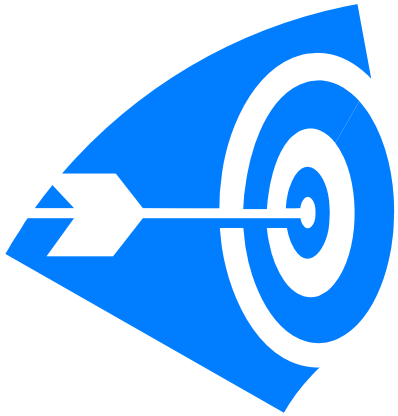
Presenting Rocks & Rolls

- **Challenge 4: Cluster Nodes Synchronization**
  - The *"fire and forget"* strategy…

    … when new rolls, updates and extensive reconfiguration has been performed just REINSTALL the appliances
  - Very fast, fault-tolerant and administrator-independent strategy
  - Rocks is provided with all the automatic mechanisms to reinstall appliances from a new state
  - Non-system data in appliances are preserved

Presenting Rocks & Rolls

- **Challenge 5: The Authentication Bottleneck**
  - 1000s of processes (e.g. MPI threads) trying to be authenticated to a central server poses a problem!
  - The 411 configuration system
    - Agressive replication of configuration information
    - Files are SSL/TSL encrypted
    - Web based system



- **Challenge 6: Cluster Customization**
  - You choose the rolls you need, add them when you want and reconstruct your system
  - Programmatic approach to clustering: system is a program (described in XML files) and appliances distribution is compiled
  - Basic aspects of the appliances can be also compiled

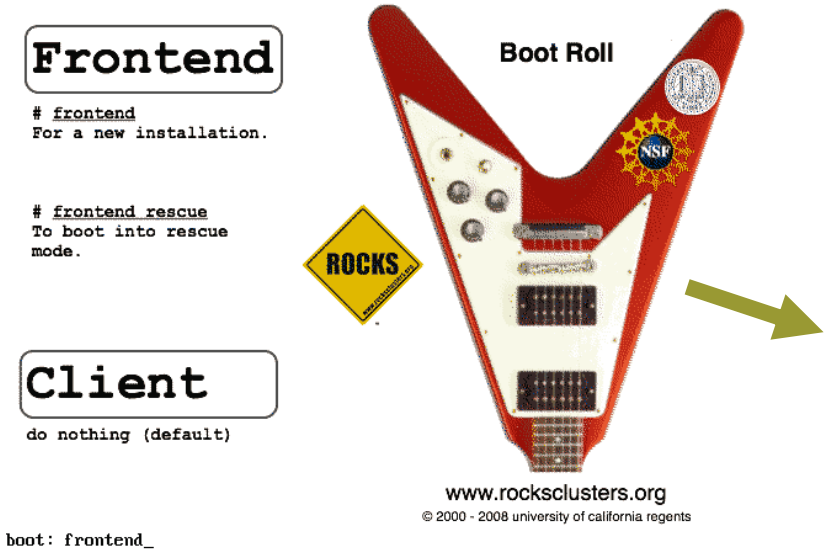
The Life Cycle of a Rocks Cluster

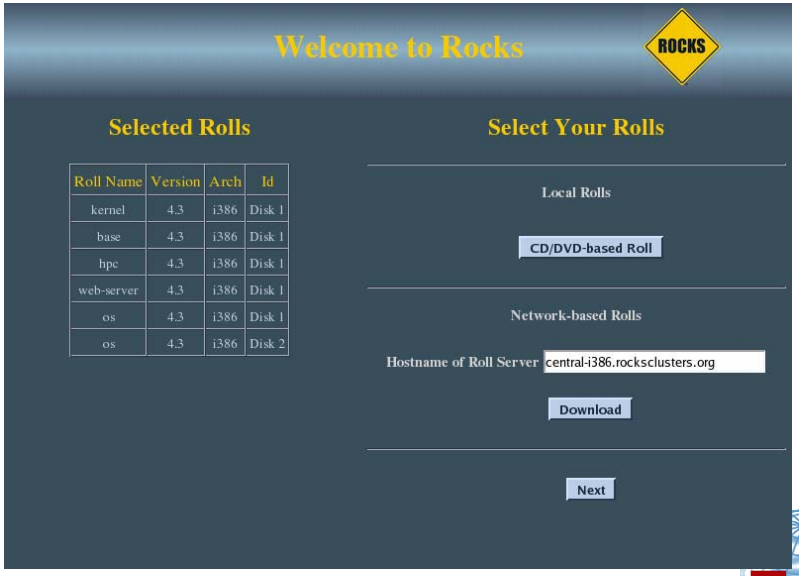
- Rocks and Rolls Installation:
  1. Frontend (appliance) installation
  2. Building appliances distribution
  3. Install appliances
  4. Configuration and fine tuning
  5. Extension

Detailed instruction can be found in :
http://www.rocksclusters.org

The Life Cycle of a Rocks Cluster

- Installation tips:
  - Frontend (appliance) installation:
    - Be prepared: frontend (2 NICs), WAN and LAN access, NIC ids (eth1 for WAN, eth0 for LAN)
    - Get the right media and check it: (pref.) Jumbo DVD

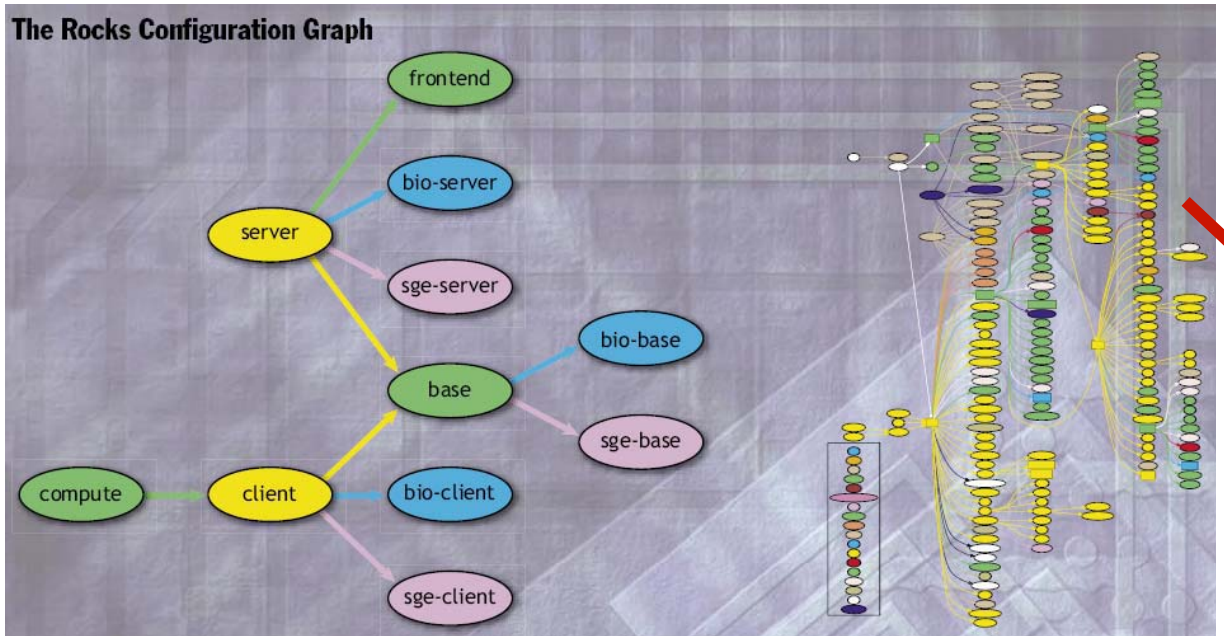




Taken from the Rocks User Guide

The Life Cycle of a Rocks Cluster

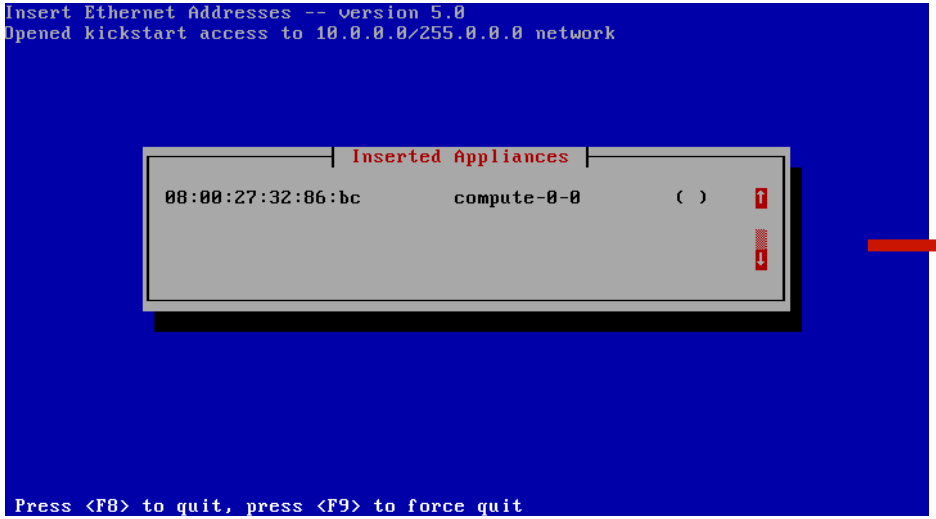
- Installation tips (cont'd):
  - Building appliance distribution
    - "Define" multiple types of compute nodes
    - Choose custom partition for appliances

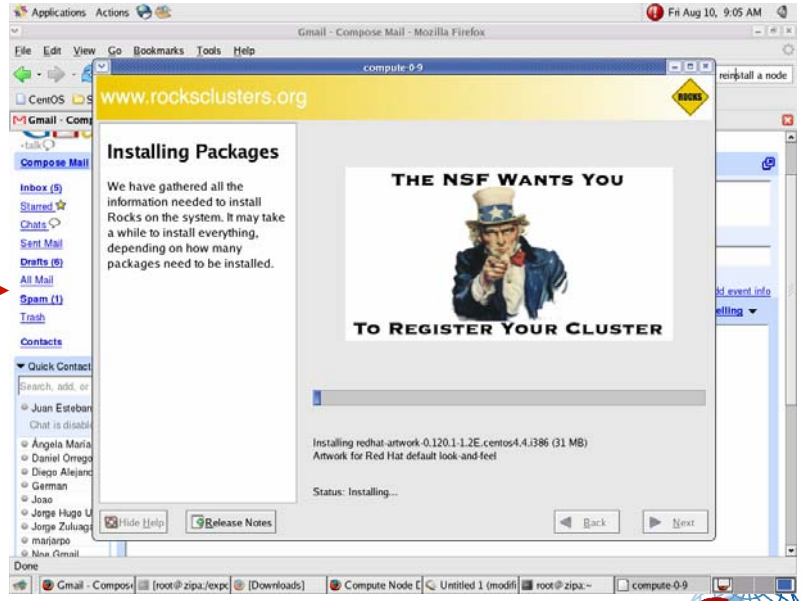


Kickstart conf. files
+
Web accesible repo.

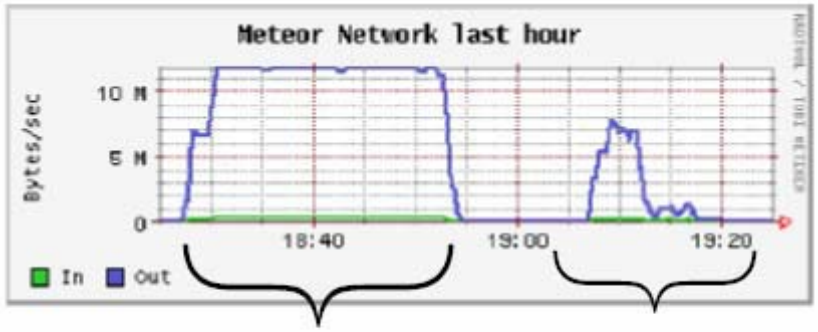
The Life Cycle of a Rocks Cluster

- Installation tips (cont'd):
  - Installation of the appliances:
    - Get PXE bootable appliances (compute/storage nodes)
    - Be careful with network configuration (firewalls other DHCP servers)

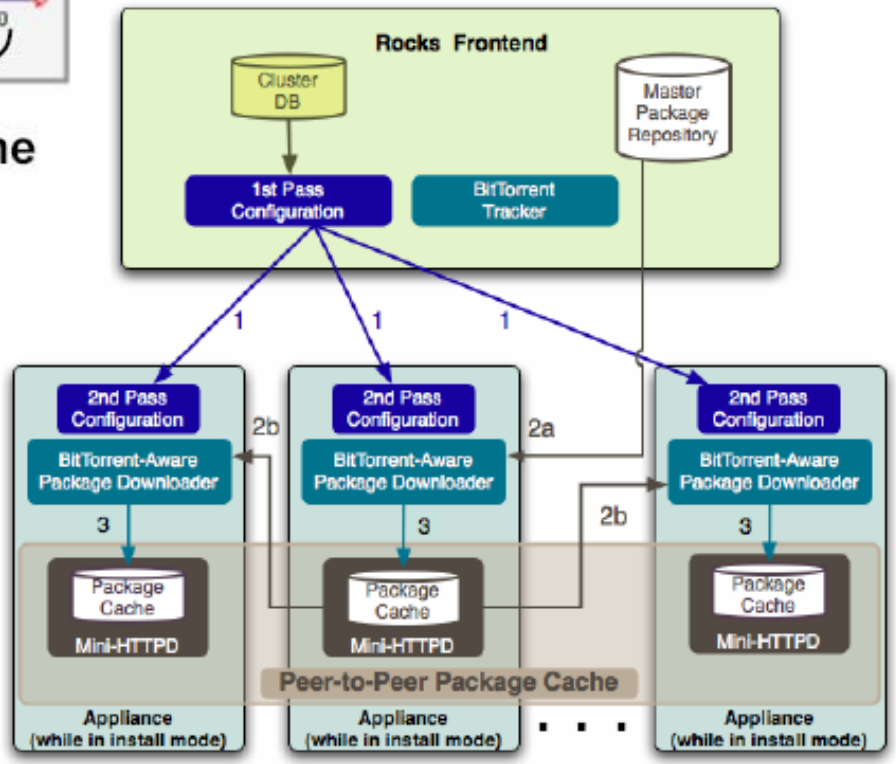




The Life Cycle of a Rocks Cluster



The Avalance Installer



The Life Cycle of a Rocks Cluster

- Installation tips (cont'd):
  - Configuration and fine tuning:
    - Don't forget to "fire and forget" when required
    - For the HPC software use the shared file system: /share/apps
    - Configure other networks (data network, managament network)
    - Do not forget to benchmark and fine tune always with your own applications

Rocks *pros*

- Rocks really rocks!:
  - **It's fast**

    Complete installation of frontend appliance and tens to 100s of compute nodes in less than one hour, nothing special should be done to get done!
  - **It's easy**

    Burn the DVD + turn on the frontend + … wait … + turn on the nodes and Bingo!, no administrative skills are required
  - **It's custom**

    You create the cluster you actually need, modular and extensible architecture

Rocks *pros*

- Rocks really rocks! (cont'd):
  - **"straight to your point"**

    If you are a scientist what are you doing installing a computer cluster? In few hours you can get science out of it
  - **Growing community of users**

    +1,000 sites configured around the world, +2,000 registered users, active mailing list, responsive developers
  - **Succesful experiences**

    Largest: Jaws (MHPCC) – Top16 [Jun2007], Tungsten2 @ NCSA – Top108 [Jun2007], communities and cyberinfrastractures (GEON Grid, Chile Grid, Thailand Grid, CAMERA – Marine Ecology Research, BIRN – Biomedical network)

Did I mention not to trust vendors

Rocks *cons*

- Rocks is not the *panacea*:
  - **Gray box**

    It is hard to fix bugs, ¿do you know python?, unpredictable behavior, installation issues
  - **RAM & Disk thresholds**

    There is thresholds for RAM size and disk size for frontend and compute nodes, errors are not properly catched up
  - **Diskless cluster**

    Compute nodes requires hard disk for distribution installation

✓ Diskless Roll

Rocks contra

- Rocks is not the *panacea (contd.)*:
  - **Limited documentation**

    Rocks user's guide and rolls guides, only schematic documentation, simple examples, poor programmer or system manager documentation
  - **Rocks Monotheism**

    In HPC and Grid computing you MUST be politheist (or even better, ***atheist***!)
  - **Kernel compilation**

    Kernel customization is even harder in Rocks, there is not simple kernel synchronization mechanisms among nodes

✓ Rocks 4.3

Let's Rocks

- The flexibility and extensibility of Rocks toolkit allows you as a developer, scientist or system administration to do very special things…

  … sticking to the way Rocks works!

- Normally you can…
  - Install Rolls after the system has been installed and configured
  - Create your own Rolls
  - Compile multiple kernels and install it in different appliances
  - Create a cluster of diskless nodes
  - Install virtual cluster and virtual compute nodes inside a real one

Let's Rocks

- You can also:
  - Install and configure public computer rooms using Rocks…

    … windows in partition, windows as virtual machines
  - Create an easily configurable and deployable HPC learning environment
  - Create storage and high availability clusters
  - Install and configure home made clusters

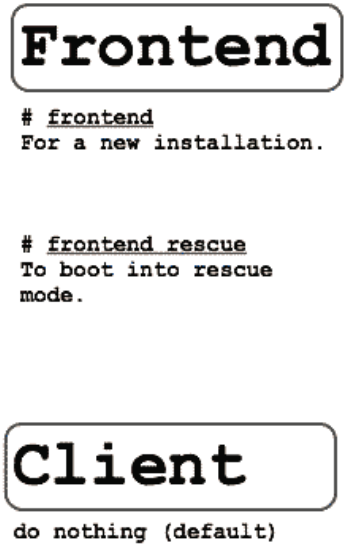
…

Gracias!

zuluagajorge@gmail.com

Rocks art

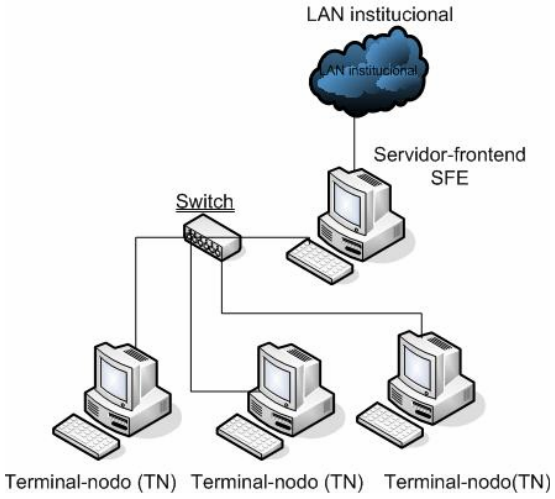
Frontend
frontend
For a new installation.
frontend rescue
To boot into rescue
mode.
Client
do nothing (default)
Boot Roll
v4.3 - Mars Hill
NSF
ROCKS
www.rocksclusters.org
© 2000 - 2007 university of california regents
Image courtesy of NASA

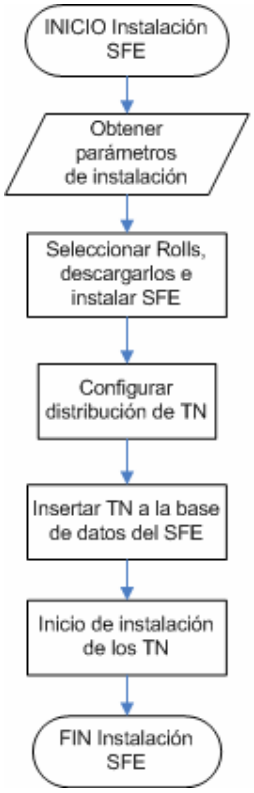

Frontend
frontend
For a new installation.
frontend rescue
To boot into rescue
mode.
Client
do nothing (default)
boot: frontend_

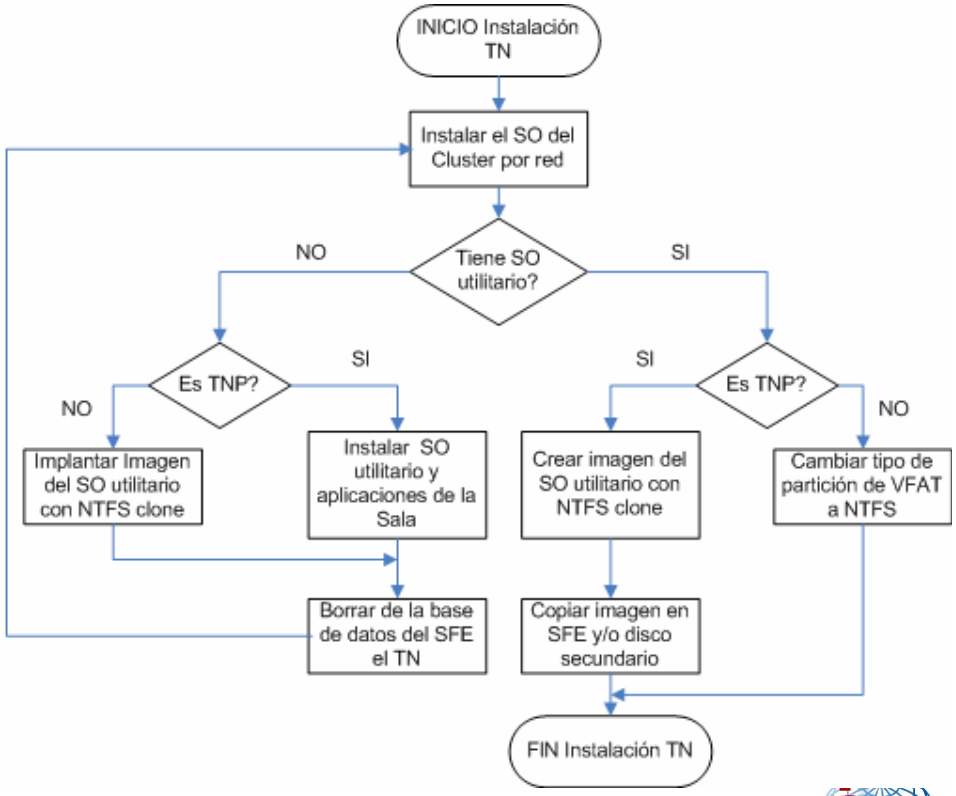
Boot Roll
NSF
ROCKS
www.rocksclusters.org
© 2000 - 2008 university of california regents

Let's Rocks

- Cluster-Room:







Let's Rocks