

InGRID: Installing GRID

F. M. Taurino*, INFM, Naples, Italy
R. Esposito, G. Tortone, INFN, Naples, Italy

Abstract

The "gridification" of a computing farm is usually a complex and time consuming task. Operating system installation, grid specific software, configuration files customization can turn into a large problem for site managers.

This paper introduces InGRID, a solution used to install and maintain grid software on small/medium size computing farms.

GRID FARM IN THREE STEPS

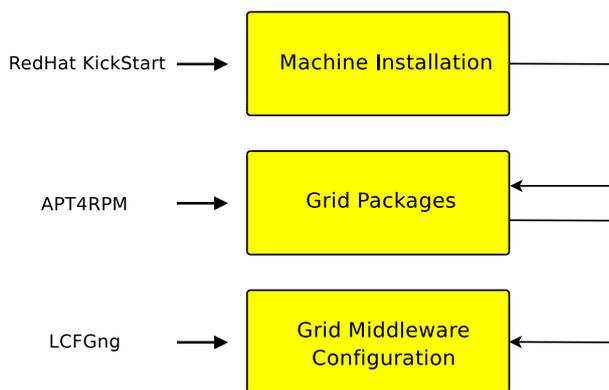


Figure 1: InGRID 3-steps installation.

Grid elements installation with InGRID consists of three steps.

1. In the first step nodes are installed using RedHat [1] Kickstart [2], an installation method that automates most of a Linux distribution installation, including disk partitioning, boot loader configuration, network configuration, base package selection.
2. Grid specific software is then integrated using apt4rpm [3], a package management wrapper over the rpm commands. Apt automatically manages packages dependencies, and is able to download, install and upgrade RPMs from a central software repository.
3. Finally, grid middleware configuration files are customized through LCFGng [4], a system to setup and maintain Unix machines, that can configure many system files, execute scripts, create users, etc.

Machine installation: Kickstart

Many system administrators would prefer to use an automated installation method to install Red Hat Linux on their machines. To answer this need, Red Hat created the kickstart installation method. Using kickstart, a system administrator can create a single file containing the answers to all the questions that would normally be asked during a typical Red Hat Linux installation.

Kickstart files can be kept on single server system and read by individual computers during the installation. This installation method can support the use of a single kickstart file to install Red Hat Linux on multiple machines, making it ideal for network and system administrators. Kickstart installations can be performed using a local CD-ROM, a local hard drive, or via NFS, FTP, or HTTP.

Kickstart lets you automate a Red Hat Linux installation. Using kickstart, a system administrator can easily manage hard disk partitions, base software installation, services to be started/stopped at reboot, and can also execute some simple pre and post installation scripts.

GRID packages installation: APT4RPM

APT is a system created in the Debian [5] community to automatically manage the packages dependencies. It can install, remove and upgrade packages, managing dependencies and downloading the packages. It is a frontend to other tools, and it uses the underlying package management system, like the rpm commands on a RedHat distribution. It is able to fetch packages from several media (cdrom, ftp, http, nfs), and it can be used to create ad-hoc software repositories.

APT has been ported from Debian to the rpm based distribution of Conectiva. The project APT4RPM now develops apt versions for all the other rpm based distributions.

An APT repository consists in a directory containing rpm packages and a file which describes the dependencies between them. This file is generated with a specific APT command (genbasedir). The repository can then be exported to the apt clients via nfs, http or ftp.

With InGRID [6], using a simple script (few shell lines), a system administrator can install any type of grid element, each one with a different set of rpms. For example, after the operating system installation, it is possible to execute the command "ingrid CE 2.2.0" to install all the packages required to setting up a computing element with the GRID-IT release version 2.2.0.

* Francesco.Taurino@na.infn.it

GRID middleware configuration: LCFGng

LCFGng (Local ConFiGuration system Next Generation), is a tool originally designed in the Department of Computer Science (now the School of Informatics) at Edinburgh University.

LCFG provides a configuration language and a central repository of configuration specifications, from which individual Unix machines can be automatically configured. Changes to the central specification automatically trigger corresponding changes in the actual configuration of individual nodes. A number of modular "component" scripts on the client are responsible for different subsystems, such as "globus configuration" or a "user configuration". The components are notified when their resources change and are responsible for translating the resources into the appropriate configuration files, and reconfiguring any associated daemons.

With InGRID, the LCFGng components on the installation server are used only to setup GRID middleware specific files, while the base clients subsystems are configured using Kickstart.

In the near future, we are going to study Cfengine [7] as a replacement for LCFGng to configure Grid middleware. Cfengine, or the configuration engine, is an autonomous agent and a middle to high level policy language for building expert systems which administrate and configure large computer networks. Cfengine uses the idea of classes and a primitive intelligence to define and automate the configuration and maintenance of system state, for small to huge configurations. It is ideal for cluster management and has been adopted for use all over the world in small and huge organizations alike.

THE INGRID SERVER

In this project, we have set up a Central Distribution Server (CDS), currently located in Naples, and some Local Distribution Servers (LDS), used to install GRID farms.

Setting up an LDS is an easy task. A system administrator can install a standard RedHat 7.3 distribution on a pc, with some extra space for the apt repository, and with no special software. On this machine an nfs and an http server must be installed, to let clients access the apt repository and the LCFGng profiles. Then the software repository, with the RedHat 7.3 installation tree, the GRID-IT [8] rpms and some extra programs, are mirrored from the CDS via rsync. After the mirroring procedure, the apt rpm and the LCFGng utilities and configuration files must be installed on the local InGRID server and customized to match the site requirements.

The server is then ready to kickstart the GRID farm.

CONCLUSIONS

This system is actually used to install, configure and maintain three small GRID farms in the Naples INFN site, for a total of 30 nodes. InGRID has proven to be a very

simple method to enable a site in setting up a GRID computing farm, with all the base grid elements: computing element, storage element, worker nodes, user interface.

Kickstart and apt4rpm are more suitable than the installation procedure supplied by LCFGng in setting up small and medium size farms. Moreover many system administrators already know the kickstart installation procedure and the apt functionalities.

Using kickstart as the installation method and apt4rpm as the base package manager, this procedure can also be adapted to newer RedHat versions, and its derivatives like Scientific Linux [9], with little or no effort.



Figure 2: The Naples INFN GRID-IT farm

ACKNOWLEDGEMENTS

The InGRID team would like to thank the Naples INFN and INFN research institutes and the GRID-IT Central Management Team for their support.

REFERENCES

- [1] <http://www.redhat.com>
- [2] <http://www.tldp.org/HOWTO/KickStart-HOWTO.html>
- [3] <http://apt4rpm.sourceforge.net>
- [4] <http://www.lcfg.org>
- [5] <http://www.debian.org>
- [6] <http://ingrid.na.infn.it>
- [7] <http://www.cfengine.org>
- [8] <http://grid-it.cnaf.infn.it>
- [9] <https://www.scientificlinux.org>