The Linux PC Farms at DESY Zeuthen

Thorsten Kleinwort
Computer Center
DESY Zeuthen
tkleinw@ifh.de
http://www.ifh.de/computing

October 15, 1998

1 Introduction

In the end of 97 the experimental groups at DESY Zeuthen wanted to increase their compute power. Although the different groups had different legacy platforms, like IBM, HP or IRIX, it came out, that all the computing needs could be fulfilled by providing commodity PC hardware with the operating system Linux, because PC hardware has a much better price/performance ratio, the operating system Linux is free and some of the groups had already gained some experience with Linux. In comparison to other HEP sites, all the computers, except some Desktops and MACs, are provided and supported by the Computer Center in Zeuthen. Up to then we had no support for Linux, in fact it was a new operating system for us. We decided to start some investigations about Linux, especially which distribution to use. In addition we wanted to have some automatic installation for setting up large clusters (farms) of Linux PCs, because we realized in a very early stage that otherwise it is not possible to set up a large number of equal PCs with little amount of manpower and to keep track with the rapidly changing versions. The decision was made for the Linux by SuSE [1], a German Linux distributor, and SuSE made us a 'hands off' version of their installation. In addition to this automatic installation we wanted to integrate Linux as a new operating system into our existing environment, although some new features started with Linux. In the end we also want to set up Linux WorkGroupServers (WGS) as common login machines and Linux Desktop PCs.

2 The Hardware

The farms are equipped with commodity hardware:

- Pentium II PC (266, 333, 400 MHz)

- ASUS (P2L97) Motherboard
- 128 MB RAM (+ 256 SWAP space)
- 2 SCSI disks:
  - 2 GB for the system, Linux, scratch area
  - 4 GB for local experimental data
- Fast Ethernet (100MBit)
- Floppy disk for installation
- Cheap graphic card, because no console is connected to the PC
- Console manager over serial port
- located in racks in the Computer Center
- For the network: 3COM Fast Ethernet switch

Some remarks: Because PC hardware develops fast and we tried to have the best CPU power which we could afford at that time, the CPU frequency increased within a period of three months from 266 MHz to 400 MHz. The ASUS motherboard has problems with the memory used by us, that doesn't allow more than 128 MB of memory. For the first farm we bought Kingston KNE100TX Fast Ethernet cards. These cards cannot negotiate with the 3COM network switch in auto sense mode, though we had to fix the switch to 100MBit on each port. The console management on the serial port is done by a DEC Alpha, with the program Polycenter Console Manager. Unfortunately, the Linux kernel cannot be configured to send all boot and shutdown messages to this port.

3 The Setup

At the time we started to investigate the SuSE Linux, the current version was Linux 5.1, with a kernel version 2.0.31. Because we decided to have the file system AFS on the Linux machines and it was no AFS client available for the 2.0.31 kernel, we felt back to the kernel 2.0.30, where the necessary daemons and modules were there. The SuSE Linux had no automatic installation, also kickstart from RedHat was not available. We decided to ask SuSE to modify their installation for us in a way, that the installation procedure can be done with a boot disk and a server without a keystroke. A special program on the boot disk starts the installation tool, called YaST, which sets up the file system, loads all necessary packages from the server and makes the necessary configurations. This installation of Linux takes some 15 minutes over Fast Ethernet. We install around 250 packages from the Linux distribution, which takes 380 MB on the system disk. Maybe not all of these packages are needed on a farm PC, but it could not be foreseen what would be needed on the different farms and there is anyway enough disk space for this 'big' distribution.

After this, what we call BasicInstallation, is done, the PostInstallation starts. With the PostInstallation we configure the Zeuthen specific environment. As a last package during BasicInstallation a startup script is installed. The PostInstallation is started by removing the boot disk and rebooting the system. After the system is coming up again, the startup script for the PostInstallation is executed. It covers the following features:

- AFS (Andrew File System): AFS is used as binary server and, what was new in Zeuthen, as file system for the home directories of the users. This is very convenient: It is much more reliable than NFS home directories and the file system is available immediately after installing AFS on the farm node, no server configuration has to be done and it is more secure. In addition the central AFS files are accessible from everywhere and the users don't have to handle different home directories on different farms. Also no local backup of the farm PCs is necessary.

- Shell profiles & printing: Because we integrated the Linux platform in our existing system, the common shell environment and all DESY printers are available on Linux.

- Security: We have the newest Secure SHell, with integrated AFS login and token forwarding, as on all platforms. Most of the network services of each farm PC are disabled, or configured in a restricted way, e.g. telnet and FTP to the farms are only possible inside our domain.

- NFS & automounter: The second (data) disks are shared within a farm, and the data disks from our file servers are accessible via NFS & automounter.

- Kerberos4: In our domain we use Kerberos4 for user authentication. This has two advantages: The users have the same passwords and there are no (crypted) passwords in the NIS maps.

- Batch system: On demand, a batch system is installed on a farm. We use CODINE [2] , a commercial product.

- PATROL: PATROL [3] is a tool, originally written at SLAC, enhanced by us, for watching the running system. It checks for running daemons, disk space, load, and several other things, and can start activities, if some unusual conditions occur, e.g. it restarts daemons, sends mail if a user program consumes too much CPU time, or informs the system administrator if a disk runs full. The status of the machines can be checked with a web browser.

The PostInstallation is done with the tool CFENGINE, which we also use for updates and patches.

In conclusion we can say that the setup of the farm PCs is not much different from our existing Unix machines, the special name farm it deserves for two reasons: The farms are clusters of PCs dedicated to special groups (experiments) and they are used for their special applications.

4 The Farms

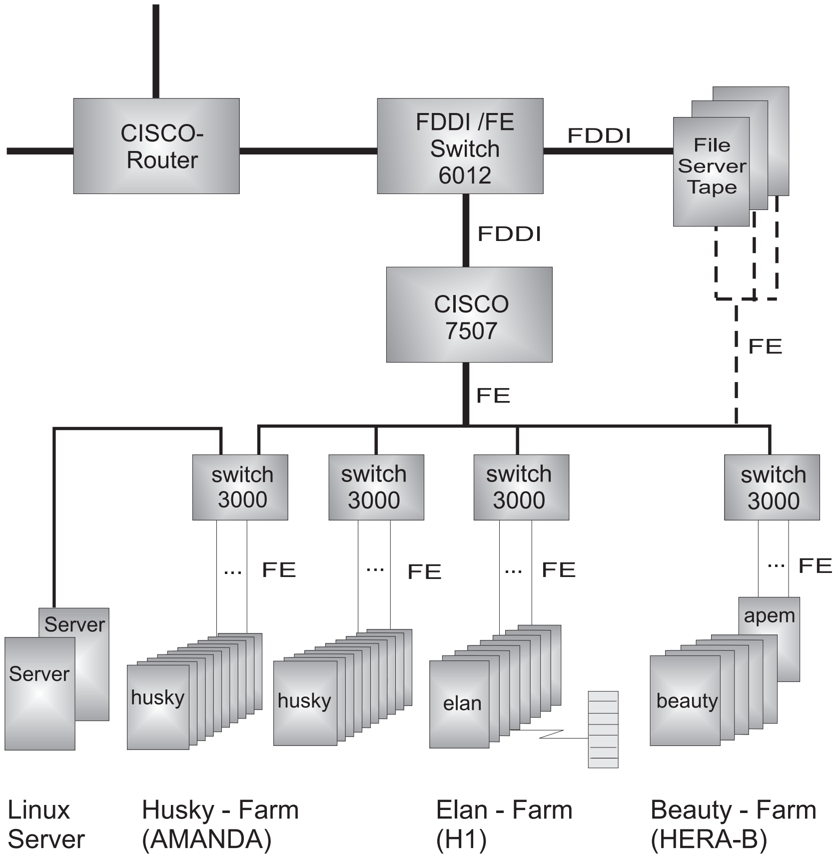

Figure 1:

So far, we set up three farms for three different experiments in Zeuthen:

- The Husky Farm (Experiment: AMANDA)
  This was the first farm we set up. It consists of 20 PCs with 266 MHz

CPUs. This farm is mainly used for Monte Carlo production and reconstruction, where the experiment has its own tools (SiEGMuND, Korsika and RECOOS). This farm has a batch system installed.

- The Elan Farm (Experiment: H1)
  This farm consists of eight PCs running with 333 MHz CPUs. In addition the first farm node is used as a file server with eight 9 GB SCSI disks. This was done because in the beginning the network connections to the other file servers were slow, and we wanted to avoid additional traffic by keeping the data on the farm subnet. We have not detected any problems with a Linux PC running as a file server, the experiences are good, but in the future we want to have extra file servers for the farms. The main use for this farm is Monte Carlo production with GEANT (with some special implementation for H1) and data analysis with PAW.

- The Beauty Farm (Experiment: HERA-B)
  This farm was set up with five PCs with 333 MHz CPUs, and was extended by a sixth one with 400 MHz CPU. In the beginning this farm was also used for the main tasks, Monte Carlo production and reconstruction (PYTHIA, Fritiof and ARTE, a special HERA-B software), but it is now also used for program development and event display. In addition it is a kind of “test farm” for the HERA-B Linux PC farm in Hamburg. They have installed a CAN–bus Control System, which is a System independent controlling system for the PCs. It checks the hardware by measuring temperatures and voltage, and can switch off a computer in case of a hardware problem. There are two talks about the HERA-B PC farm in Hamburg at CHEP ‘98.

In addition to these farms we have two Linux PC servers, for the main services that are needed:

- Installation & Linux: These two servers provide the necessary files for the automatic installation, in particular the Linux distribution, via NFS.
- Bootp-Server: This service is needed for setting up the farm nodes from scratch, because no IP information is on the installation disk, all information is provided via Bootp.
- NIS (YP): One of these servers is the master NIS server. For a more stable NIS, also the first node of each farm is a NIS slave server.
- Batch: One of the servers is set up as the server for the batch system.
- SYSLOG: These servers collect the SYSLOG messages of all farm nodes.

In fact, only one server is needed for these tasks, but it is good to have two, especially if you want to test new things, here in particular to install a new

Linux version. At the moment we also use these servers to set up other Linux hosts, like Desktop PCs.

The farm network is shown in figure 1. All farm nodes are connected via Fast Ethernet and are located in a separate subnet. The connection to the file servers and the rest of the world is done by a CISCO router. The drawback still is the poor connection to the file servers, but it is planned to equip these file servers with an additional network interface, which is in the farm subnet.

In general we can say that the three farms we have so far, have increased our compute power for the experiments by more than a factor of two. This is also a great relief for our other Unix workstations, because they also act as login machines. In comparison to our other Unix machines one farm PC CPU is up to two times faster. The load of the farm PCs is good, they are used all the time and the different experiments plan to upgrade or extend the number of PCs. Beside some problems in the beginning with buggy drivers and kernels, the uptime of the PCs is good (around 60 days) and they are only shut down for regular maintenance.

5 Conclusions

Because we decided from the beginning to integrate Linux as a new Unix platform into our existing computing environment, it was a full time job to set up the farm PCs. It was a good decision to have some installation tool to set up equal farm PCs, but that is flexible enough to make individual configurations. The drawback of a "free" operating system like Linux is that you rely on public domain software. Even if the Linux programmers are doing a good job, new software is sometimes buggy. So you have to keep track with the newest releases of the Linux software, and sometimes you have to solve problems on your own. In comparison of a Linux PC farm with 20 nodes to one big Unix workstation with 20 CPUs you have to give more support. On the other hand with commodity PC hardware running Linux you can increase your compute power with low hardware costs.

For us the farms were an entry to the operating system Linux, and with the gained experience we now start to upgrade our farms and to set up Linux PCs also as common login machines and as Desktop PCs.

References

[1] SuSE: http://www.suse.de or http://www.suse.com

[2] CODINE is available from GENIAS: http://www.genias.de/

[3] From ftp://ftp.ifh.de/pub/unix/sysadmin/patrol/, a snapshot of PATROL can be loaded