APPENDIX A

Typical Hardware Requirements for a Linux Development Workstation

There are many factors that go into selecting an appropriate hardware specification for any given task; and the final specification will vary greatly depending on a particular environment, budget and time available to accomplish the task. In general, however, the things that need to be considered are: the speed of the processor, the amount of memory, size of the local hard drives, network connectivity, the video card and the display. Depending on the development projects to be undertaken, additional devices such as modems, sounds cards and the like will also need to be considered.

The three situations that we will consider are a single developer working on a project alone, a distributed environment in which multiple developers are accessing a single machine via remote log-on (telnet, ssh or X-Windows), and a distributed environment utilizing a software repository.

A.1 Parts of a Workstation

Computers consist of many parts working in conjunction with each other to turn fluctuations in current into meaningful output on the display screen. Much of the work of the computer takes place using the motherboard, the processor(s) and the computer's memory. These components provide the "horsepower" that largely determines how fast the computer runs.

Working in conjunction with these parts are things like the hard disk drive that provides the internal long-term storage. There is also the video card that is used to convert the information in the computer into something that is readily usable to us mere mortals and the modem or network interface card that permit computers to communicate with one another.

Externally there are things such as the keyboard and the mouse that permit the user to enter data into the system, and the monitor which enables the computer to present information in a way that the operator can easily understand.

When evaluating a new computer system, remember that the motherboard, the processor and the memory determine *how fast* your computer will be able to perform the tasks required of it.

The hard disk drives, the network and video cards, etc. determine *what and how much* your computer will be able to do. However, speed of hard drive is also a major factor in overall speed of compilation in case your computer is low on memory.

Lastly, the input and output devices determine *how much style* your computer will have while it goes about doing whatever it was told to do.

While there is some overlap in function, these definitions provide a general outline of the parts of the computer and can give us a basis for determining a base-line system.

A.2 Section 1—The Processor and Memory

Systems are available with one or multiple processors. Random Access Memory (RAM) is presently cheap and easy to find. Processor speed seems to double every eighteen months and almost as soon as a new system is purchased, it becomes obsolete.

The speed and the number of processors will determine the sheer raw power that a computer has. Speed is measured in megahertz (MHz) and more is always better.

The memory that a system has installed will also effect the overall speed of the computer. If the system runs out of physical memory, it will, if the space has been set aside for it, start paging memory to disk. This is the process of writing a portion of memory that has not been used recently (in computer terms that could mean milliseconds) to the disk in order to free the memory up for more immediate usage. When the information on the disk is required again, the computer frees up physical memory and reads the data from the disk into that area.

As the hard disk drive is several orders of magnitude slower than simply reading the information from memory, this slows down the entire computer as it waits for the information to be written and read from the drive. If your computer sits idle while there are large amounts of access to the disks, it is possible that the computer is low on memory and swapping to the disk. Adding more memory to the system should speed it up.

A.3 Section 2—The Disk Drive

At present, there are two main types of drives available on the market for personal computers: IDE and SCSI (pronounced “scuzzy”). The first and most common is the IDE drive. This is the type sold with most off-the-shelf workstations. A standard computer will be able to support up to four IDE drives at a time. These drives may be hard drives, CD-ROMs, DVD drives or even tape drives. Once the maximum of four devices has been installed into the computer, no more may be added until one is removed. This and the fact that IDE drives have traditionally been slower than

SCSI drives are the main limitations in using this type of device. They are however, cheaper and somewhat easier to install than SCSI drives.

SCSI devices may also be used for mass storage, CD-ROM and DVD drives, but SCSI devices can also include scanners, printers and other devices as well.

Up to seven devices may be attached to a single SCSI interface card inside the computer and a computer may have more than one interface card. The total number of SCSI devices installed on one computer is limited to the number of available interface slots on the motherboard, times seven. On most modern workstations and servers this will allow a total number of SCSI devices ranging from seven to over thirty. New SCSI standards allow more devices to be connected to a single SCSI adapter.

SCSI drives are often configured in a Redundant Array of Inexpensive Disks (RAID) that can be used to increase the reliability or the speed of the system. The trade-off in this configuration is a loss of available drive space. The decrease in available space may be anywhere from negligible to 50% of the total space on the drives depending on the RAID level used.

A.4 Sizing the Task

When developing requirements for any computer system, it is important to understand exactly what the system will be used for and how it will be used. Once the purpose of the computer is understood, potential bottlenecks can be located and the potential for these to cause problems can be minimized or eliminated.

A.5 Stand Alone Development Workstation

In this configuration a workstation will be used to develop applications locally. The source code will be kept on the local hard drive and the editing and compiling will take place on the same machine.

The main concern with this configuration would be in the area of processor and memory, and in having a backup system in place to ensure that data is not lost in the even of a hardware failure.

A typical developer's machine will have a processor in the 800+ MHz range and, depending on the scope of the application being developed, may have more than one processor. The available memory will again depend on the type of application be written with typical machines having in excess of 256 MB of memory and, in the event that large amounts of data have to be manipulated, more than a gigabyte of memory.

The Linux OS, with development tools, can be installed utilizing approximately 1 gigabyte (GB) of disk space. The source code for the OS, which represents a decade of work from people all over the globe, can be installed in less than 50 megabytes of space. A single developer working full-time would have to be very productive in order to fill up most drives available today as these typically come in 60GB sizes. Hard drive space is more often used for development libraries and tools, separate runtime environments, etc.

Again, disk size may have to be increased if large volumes of data are being used in the testing of the application.

These estimates for simply generic in nature and would be a good starting place for new equipment purchases. It is entirely possible for productive development to take place on older, used equipment in the sub 300 MHz range. It depends entirely upon the application being developed and the time available to accomplish the task.

However, with the cost of equipment constantly falling, it is often well worth the investment in newer equipment in order to reduce the compile time of the application.

A.6 Distributed Development—Remote Access

This example supposes that multiple developers will be accessing a remote server and developing on that platform. The source code will reside on the server and the server will be used to compile the programs.

In this instance, the requirements for the workstations will be modest compared to the previous example, but the need for central server is added. This server will need to be far more powerful than the previously specified workstation.

The server should be as fast as the budget and circumstances permit, and the number of simultaneous developers using the platform should be considered. You will almost certainly want a server with several processors.

It will require more disk space, preferably with RAID enabled and the most utilized file systems spread out over several physical disks. A complete backup strategy should be in place and tested on a regular basis in order to ensure that the data is secure and can be restored in the event that any problems should arise.

This server will require a large amount of memory as well. Several gigabytes is not unreasonable for a small number of developers. Several developers trying to compile at once can quickly drain the system's resources.

If the remote access strategy includes using a secure shell (ssh) to connect to the server, additional processing power will be required due to the encryption overhead.

On the other hand, the actual workstation in this example would be comparatively modest. Most workstations capable of handling word processing and e-mail tasks should be up to handling this type of load with little difficulty. All of the storage, access and intensive processor use will take place on the server.

A.7 Distributed Development—Source Code Repository

This last example is a combination of the other two designs, and has probably supported more Open Source developers than any other model. In this design, a relatively small-scale server is used as a software repository and individual workstations similar to the one in the first example are used to provide the actual processing power.

The main requirement for the server is a large disk storage space and a modest amount of memory. CVS, the source code control application used in this book, as well as others, can use

quite a bit of memory while comparing old and new versions of files with one another. As a result, the server should have at least 256 MB of memory and potentially much more if you expect to use large source code files.

The server will not be responsible for compiling the application, so a fast processor is not necessary.

What would be helpful, however, would be to ensure that the server has a well-tuned networking subsystem, and possibly some redundancy in that area as well. Additionally, if required by a particularly active source code repository, multiple NICs (Network Interface Cards) may be set up on different subnets to help reduce any network congestion that might arise.

As previously mentioned, individual workstations will resemble the ones listed in the first example in this section—the stand-alone developer workstation.

A.8 Summing Up

While there is no way to accurately give minimal and optimal recommendations for any and all software development projects, these guidelines should give you an understanding of some of the issues that arise in selecting the servers and workstations necessary to provide development support.