Linux running on the IBM S/390

BIG BLUE PENGUIN

There must be few Linux developments that have excited more attention at the moment than the Linux/390 port. Ulrich Wolf shows us what is especially interesting about it is not just the power of the exotic architecture but also the number of new application options which the mainframe opens up for Linux.

Far from our humdrum IT world, far from the overhyped markets and the poorly written software that gets debugged by the paying customer and the "Broken by Design" hardware architecture lies the kingdom of the mainframe.

There, almost everything is different from what we are used to. So different, that for example hard disks are not called hard disks, but DASD, pronounced "Daz-dee", which stands for Digital Access Storage Device. There are computers living in this kingdom which are only ever run up once in their lives and whose address space can be occupied by several different operating systems. If, in the manner of the king in the fairytale, you ask who owns all the nice, stable hardware and software, you almost always get the same answer: it's IBM.

The Mainframe Market

The S/390 architecture from IBM and its predecessors S/370 and S/360 plays, in the field of mainframes, the role of a standard in the same way as the architecture of "IBM-compatible" PCs at the other end of the scale. Except that in choosing the operating system of the "big iron" IBM has not let itself be taken for a ride by a second-class garage firm, but has retained control over the system software. Not least for this reason: the mainframe division is regarded by many in the know as a cash cow for the company.

But even if IBM dominates in this market segment, it is not an autocrat. When it comes to S/390-compatible machines, Hitachi, HDS, Amdahl and others are active world-wide, while those with regional importance include Olivetti, Comparex and PQ Data. According to an analysis by the Gartner-Group in 1999, S/390-compatibles account for some 85% of the mainframe market. In Europe, Siemens with its BS/2000-based business server series is also important, even if this has reduced somewhat in recent times.

Apart from its extremely high reliability, there are two main advantages of a mainframe architecture. These are the I/O performance and the high number-crunching power of the processor. The former is achieved through the design of the architecture. Each device comes with its own high-performance controller which completely takes over the administration of physical addressing. For this reason, the S/390 could also be seen as an asymmetrical parallel computer.

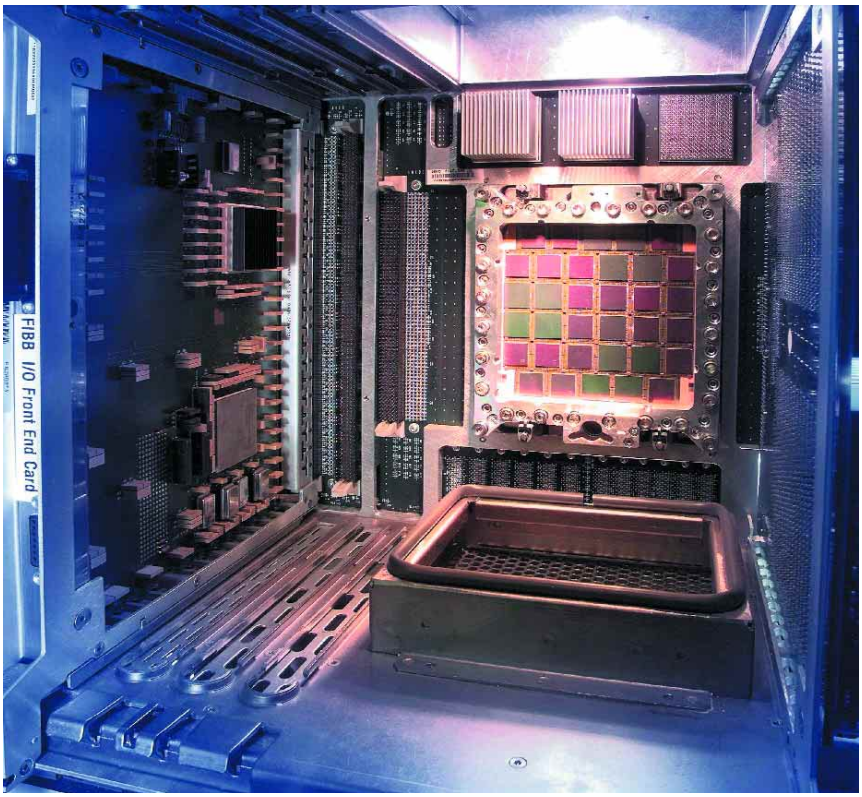
Management information is, to put it bluntly in a single phrase, an attempt to distil from the totality of the business data of a company, information that is valuable to its management. The process of data creation and the extraction of information and ultimately knowledge is described by terms such as Data Warehousing, Data Mining or Online Analytical Processing at a very high level of abstraction. These technologies are currently being implememted using databases and middleware on the S/390 architecture.

The second application area which IBM is hothousing for mainframes has now, unlike MIS, itself become a byword known even by the man in the street: "Dot.com-isation". Much of the data hoarded on the "big box" is, of course, also highly suitable for use as the basis of "e-commercialisation" of the company.

Rise or fall of the big iron?

The decline of the mainframe world has often been heralded, for various reasons, yet it keeps being postponed. In recent years it has become apparent that highly accessible Unix servers represent only a limited threat to mainframes. On the whole it is now being assumed that there will be a return to moderate medium-term growth of the mainframe sector.

In particular, in this respect, new tasks are being created for the "old iron" in the domain of e-business. This is why IBM is currently staking two principal claims, especially by means of marketing activities, for mainframes. Firstly, everything which comes under the heading of "Management Information Systems" (MIS). Many companies have been using S/3x0-systems for decades, on which their entire mission-critical data and processes are hoarded over a period during which the world of Microsoft, but to some extent also that of Unix, has been stumbling due to incompatibility with its neighbours. So what could be a better idea than letting loose MIS tools on these gigantic, consistent databases? And why should a Sun server or perhaps something Intel-based with Linux or NT be necessary for applications?



Inside S/390 Enterprise Server. The IBM flagships is available in 24 versions.

A few terms from the S/390 environment

Channel	Processor-controlled unit that permits the transfer of data between main memory and peripherals.
CMS	*Conversational Monitor System* The native operating system for virtual machines under VM CTC
DASD	*Direct Access Storage Device* In the broadest sense, any main memory, but more precisely hard disks including the controller unit
EBCDIC	*Extended binary-coded decimal interchange code* A set of symbols similar to ASCII, but coded differently.
ESCON	*Enterprise Systems Connection* A sort of IBM in-house network standard. There are special ESCON channels, processors etc.
IPL	*Initial Program Load* Synonym for the boot process under Unix/Linux, also used for loading a configuration file into the main memory in order to restore a working environment. Is often used as a verb: "to IPL".
LPAR	*logical partition* Logical partition of the complete hardware. Parts of all resources are assigned to an LPAR - CPU, RAM, I/O - and after that it is a fully autonomous computer inside the mainframe. Once installed, LPARs run for a long time.
MVS	*Multiple Virtual Storage* Operating system for the S/390. Predecessor of the OS/390. Also used as a name for the MVS part of the OS/390 or, not quite correctly, for the OS/390 as a whole.
Open Edition	The Unix Interface of OS/390
OS/390	Current standard operating system for S/390 mainframes
SNA	IBM-specific network architecture with layer structure. Defines own logic structures, protocols and formats.
VM	*Virtual Machine* The virtual CPU, virtual memory and virtual I/O channels available to the user of a guest operating system under VM – Thus, the virtual hardware.
VM/ESA	*Virtual Machine/Enterprise Systems Architecture* The software which makes it possible to create virtual machines. The "enterprise system"
Sysplex	*System Complex* A cluster of MVS- or OS/390 operating systems on one or more real machines.
Minidisk	under VM, a DASD or a logical part of a DASD with its own virtual device number and virtual cylinders.

But the Internet and the mainframe world are two universes which have long since been developing alongside each other. Internet technology is traditionally Unix technology and a stalwart MVS developer has little interest in Unix. On the other hand, in the Unix world there is an untold number of established applications, protocols and tools available which are fine-tuned to each other, and which are also mostly still free software. So it would be a shame not to be able to use them.

In fact, the new OS/390 does "talk" to a Posix-compatible Unix, but many developers are not really happy with this mixture from two worlds. It is in this connection that the initiative of IBM to make Linux available for the S/390 architecture must be seen. This will in fact make it possible to run all applications available for the "de facto standard" Unix, on extremely high availability hardware.

Nothing like home

In principle there are three ways in which an operating system can run on an S/390. The first way is direct on the hardware, in which case it takes control of the complete resources of the system.

More usually, however, the hardware available is partitioned. Note that this does not mean the same thing at all as partitioning a hard disk under Linux. On the S/390, all resources (such as CPUs, RAM, IO channels) can be assigned to different logical partitions (LPARs). This allocation may be static or dynamic, depending on the resource. This means, for example, that a CPU can belong to two partitions and can be made available as required to one or other of the partitions (floating CPU). A large S/390 machine from Generation 6 (G6) has 16 CPUs. Two of these are dedicated IO CPUs and are not directly available to the OSs. Another two CPUs are "reserves" in case one or more CPUs of the remaining 12 fail. The same technology which allows floating CPUs also makes the reserve CPUs available transparently for the operating systems (i.e. without the operating systems even noticing that a CPU has failed). Each LPAR represents a separate and complete system within the physical machine. The OS mounted on it has full control over the assigned resources. The S/390 architecture permits up to 15 such LPARs.

Guest system

The third option for operating a "foreign" OS, is as a "guest system" within the VM (Virtual Machine) operating system. VM multiplies the resources of a complete system or an LPAR by a time-slicing procedure almost as many times as it likes. The individual duplicates of the hardware are called VM guests. By means of a log-in process very similar to that used in the world of Unix, a user can now obtain access to the system console of this virtual 390 system. Using CP (Control Program) the user can configure the hardware of this virtual system and boot up operating systems located on the disks or tapes accessible to it. Booting, in the 390 world, is known as IPL-ing. IPL stands in this case for "Initial Program Load". In this way more than 40,000 different Linux kernels have been run in parallel on a ten-processor machine.

Originally VM/ESA was conceived by IBM as an interactive multi-user operating system. VM takes over the multi-user section, providing each user with their own little 390 machine. Each interactive session then runs a small single-user operating system specially written for it, which is started in the VM guest. This system is known as CMS, which stands for "Conversational Monitor System".

The predominant operating system on the S/390 is OS/390, which has a long line of direct predecessors going back 35 years. The last of these predecessor versions, MVS, is to a large extent compatible with the current OS/390. Developers and administrators are still fond of using MVS as a synonym for OS/390. On the other hand MVS can also currently be regarded as a subset and/or lower layer of OS/390. This operating system is best suited to the special requirements of mainframe hardware.

Among other things, it also contains UNIX Services, which provide a Posix-compatible interface to MVS functions. MVS was originally only intended to execute batch jobs entered using punched cards. These punched cards were still in use well into the 1980s. Batch processing is still one of the primary tasks of computers in this size class, and the JCL (Job Control Language) developed for this is one of the most polished languages available for non-interactive applications.

Besides this, VM is often run for interactive applications and for software development. VM is useful for software development because its simulation of S/390 hardware is so good that any operating system which runs on the S/390 architecture will also run there. This makes it possible to make available to every software developer his own S/390. For operating system development this is especially handy as the developer will not hurt anyone else if there is a system crash and it is not necessary to provide each developer with his own real 390 machine, which would probably be too expensive, even for IBM.

One of the great advantages of VM in development (including, of course, the development of Linux for S/390) is the debugging options available under CP. It is possible to track accesses to memory areas and to step through the program running in the guest instruction by instruction. By doing this it is possible to watch how contents of registers and the real or virtual memory alter.

Enter Linux

The idea of transferring Linux onto mainframes is not new, and nor was it born in the IBM lab. An earlier project going by the name of Bigfoot had the aim of supporting the earlier version S/370 too.

Bigfoot was (or is) a normal free software project, whose initiator Linas Vepstas worked for IBM. Linas Vepstas is also famous for such things as GnuCash and is involved in various projects in the field of Linux Enterprise Computing. Most readers will, however, certainly know him as the author of Linux Software RAID Howtos. For various reasons, the S/370 project he initiated was at first put on ice, and rumours say that IBM did not exactly welcome the Bigfoot people. The background, from the subjective viewpoint of Linas Vepstas, can be read on http://linas.org/linux/i370.html. Had it come about, this project would have made it possible to make Linux run on older hardware, too. However, it is doubtful whether this would have been useful, since older hardware designed using bipolar technology is a power-guzzler par excellence and is therefore now scarcely used.

The developers at IBM did, however, ensure that the IBM port uses instructions which are only available on relatively new machines (G2 onwards). These instructions make it possible for a new compiler to function very much more simply in some points than that of Linas Vepstas and in addition to generate faster code. This difference is, however, so fundamental that no part of Vepstas' code could be included.

The competing project with the official name of "Linux for S/390" was started at IBM Germany in Böblingen and was developed in secret until the kernel and Binutils were finished. On 15th December 1999 the almost-complete port was demonstrated for the first time. The result can be seen: A Linux which runs on "bare iron", on LPARs and under VM.

The latter is probably the most potentially useful method of operation since with this it is possible to start literally hundreds, even thousands of kernels, which run in completely separate address areas. But running it in an LPAR is also useful because of the cost advantage. In August IBM offered licensing models with dramatically reduced costs for Linux, compared to OS/390, both for Logical Partitions and Virtual Machines.

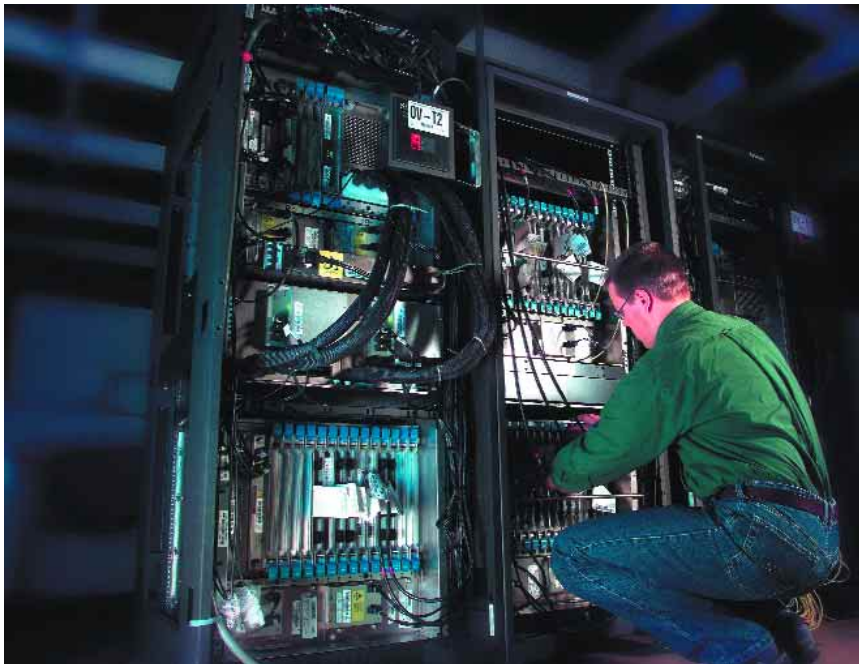
VM is the fastest

The installation of a Linux for S/390 under VM requires only a little knowledge of VM and a VM guest account. The account must have access to at least two minidisks of a specific size and two dedi-

One of the S/390's processors. This multi chip carrier is the most complex one in the whole world produced in series.

Info

Official Linux/390 page: *http://oss.software.ibm.com/developerworks/opensource/linux390/index.html*
S/390-Hardware: *http://www.s390.ibm.com/hardware/ Linas Vepstas' pages on* Linux for mainframes: *http://linas.org/linux/i370.html*
Think Blue Linux, the distribution for Linux/390: *http://linux.s390.org/*
Hercules-Homepage: *http://www.snipix.freeserve.co.uk/hercules.htm*
Linux/S390 under Hercules: *http://penguinvm.princeton.edu/hercules/index.html*
Private Homepage with gigantic collection of links: *http://os390-mvs.hypermart.net/homepage.htm*



A little too big for the desktop, but just fine for big business

cated device addresses for the network connection. All this has to be provided by the VM administrator. Once these requirements have been fulfilled the installation process proceeds in a manner familiar to anyone who has installed Linux.

One interesting feature under VM is the so-called Virtual Reader. (Reader in this case stands for punched card reader.) This behaves in a similar way to magnetic tape. When Linux is booted up for the first time the Linux kernel, the parameter line and the image for the RAM disk in the virtual reader are copied under VM. Then you give the instruction to make an IPL (initial program load = boot up) from the virtual reader. In this way the content of the reader is loaded into the main memory and executed from a defined address.

When Linux has been successfully run up using a file system in the RAM disk, the hard disks (minidisks or dedicated DASDs) can be mounted, formatted and provided with the necessary content (root file system). Now the system is in a condition to allowing booting without a RAM disk.

After this there are two options for booting up Linux for S/390 under VM. The first is to continue loading and IPL-ing the kernel and the parameter line in the Virtual Reader. The second is to cancel the IPL instruction for a specified hard disk. In this case, a bootloader must be installed on the corresponding disk. Similar to the LILO loader used to boot Linux on Intel hardware, the bootloader for Linux for S/390 is known as SILO.

Naturally there are some differences of principle in an architecture which deviates so far from the PC, the "home planet" of Linux, which are visible from the outside. Since the DASDs of the S/390 are a special kind of hard disks, they are not addressed via the device names for SCSI or IDE disks. The device names for DASDs are */dev/dasda, /dev/dasdb*, etc. (In the previous versions of Linux for S/390 the DASDs were still referred to as */dev/dd<letter>*, which was changed on the advice of Alan Cox). The problem that, due to this naming system, the DASDs can only address 26 disks will presumably be resolved by the device file system in 2.4, because up to 65536 devices can be connected to an S/390 machine.

There are now several Linux distributions based on this port.. One of the first is mounted on the server of Marist College, which co-operates closely with IBM. On the basis of this, the German company Thinking Objects has developed its own distribution based on RPM packages named "Think Blue" But soon SuSE entered the bandwagon with their own mainframe distribution. Red Hat CEO Matt Szulik, however, told "Linux Magazine" that his company had no such plans.

Hercules - a "giant emulator"

Hercules was around even before Linux/390 was born. This is an emulation of the S/390 instruction set including the channel program under Linux 2.2x. Together with Linux/390 it enables, on a home PC (with at least a Pentium processor) one of what must be the wildest "emulation orgies" currently possible. When Hercules is up and running (it is reported that it takes some time before the first prompt appears), it is possible to install Linux/390 on this emulated mainframe. So everyone whose appetite has been whetted by this article for the world of the mainframe can at least get a taste of this on the PC.

Anyone who would rather have a "real" mainframe OS can also use the since-released OS/360. For anything else a licence is needed. VM, in principle, does not run with Hercules: the ultimate goal of running several Linux kernels in the virtual area of an emulated mainframe thus remains a dream.

The emulator is released under a special licence, which the author refers to as the Hercules Public Licence. This allows use only for "educational and hobby use" and prohibits among other things the distribution of modified versions or the use of parts of the code in other programs. But anyone who is simply curious about how an S/390 feels should be able to live with this. ■