Differences Between IBM eServer iSeries and pSeries Servers Running Linux

Kay Tate

April 09, 2004

Since the inception of running Linux on an IBM eServer iSeries server, it has changed the computing environment and traditional iSeries workloads. It has been said that Linux and OS/400 are among the two most stable operating environments available. But it has truly been the combination of Linux running on iSeries hardware and OS/400 itself that has created this exciting environment. Each component brings its strengths to the equation creating greater possibilities for solution developers and end-users alike.

Introduction

This paper describes and compares the evolving technologies that provide support for Linux operating systems on IBM pSeries and iSeries systems. While much of this technology is common to both the iSeries and the pSeries platforms, this paper adds perspective on some specific details of these two systems that can help ISVs exploit series-specific advantages. The server systems described here run current operating system releases AIX 5L Release 5.2 and OS/400 V5R2 in addition to Linux. The advantages of using Linux in conjunction with this other software will also be discussed.

Since increasing numbers of ISVs are delivering products in consolidated environments, this paper focuses on comparing the partitioning capabilities and related technologies of iSeries and pSeries systems. To assist the reader, these comparisons are organized by topic with "iSeries" or "pSeries" denoted in boldface type. The web sites and white papers referenced in the text can be found at the end of this paper.

In addition to the partitioned environment, some pSeries systems can run Linux as a stand-alone operating system. This may be particularly desirable on smaller systems used as workstations or dedicated servers. The stand-alone pSeries Linux systems operate much like other Linux single-image systems and are addressed in more depth in Chapter 5 of the "Linux Applications on pSeries" Red Book.

Terminology and Acronyms

GUI
Graphical User Interface
Hardware Management Console (HMC)
Management interface for pSeries LPAR

Hypervisor
software system layer that facilitates the management of LPAR resources

Integrated xSeries Server for iSeries (IXS)
Intel card that can be plugged in "under the covers" of an iSeries system to provide Windows integration with other iSeries workloads

iSeries Navigator
Graphical interface for OS/400

Logical Partitioning (LPAR)
splitting system resources into multiple system images with their own assignments of hardware resources

Network Server Description (NWSD)
Concept that is used by OS/400 to manage guest operating systems

Network Storage Space (NWSSTG)
OS/400 disk space that is packaged for guest partitions on iSeries as their virtual disk space

PAM
Pluggable Authentication Modules

Historical Perspective on Partitioned Systems
The zSeries can be viewed as the originator of IBM eServer system partitioning and virtualization of system resources. pSeries and iSeries system partitioning on the POWER hardware architecture are its technological descendants. iSeries systems, being focused on integration (the "i" in iSeries) and workload consolidation, began delivering logical partitioning technologies more than five years ago, prior to partitioning being used for Linux support. iSeries object-based architecture has also provided flexibility and long-standing integration with other "guest OS" technologies via the Integrated xSeries Servers (IXS) for iSeries that run Windows operating systems. With LPAR and guest OS technologies in the iSeries repertoire already, it was a fairly easy step to enable Linux on iSeries in a partition using the established guest OS model.

Over the past several years partitioning capabilities have also been integrated into models in the pSeries product line. This has facilitated the delivery of independent Linux and AIX partitions successfully working side by side on the same system. In effect, pSeries has been able to leverage some of the iSeries technology that was initially derived from the zSeries systems. The use of the POWER chip architecture in both iSeries and pSeries servers has made it easier for them to share a variety of technologies including partitioning.

As a result, the first thing to note about iSeries and pSeries support for Linux distributions is how similar the support is on both platforms. The general overview of iSeries and pSeries commonality can be found in the "Commonality Across IBM eServer iSeries and pSeries Servers Running Linux" paper. Because of this, the great majority of code for each of the current Linux distributions that run on iSeries and pSeries servers is the same for both systems.

Logical Partitioning Architecture
Partitions on iSeries and pSeries systems each have their own allocations of processor cycles, memory and I/O. They are specifically defined hardware partitions, not fully virtualized machines
like the zSeries VM facility. Each partition has its own operating system image. Resources for a partition are assigned and managed by low-level software called a hypervisor. The operating systems images sit above the hypervisor in the system software stack with the hypervisor managing their access to the hardware below.

The underlying differences that exist in partition configuration and management on iSeries and pSeries are due to the use of different hypervisors. The iSeries hypervisor has grown out of OS/400 technology. The pSeries hypervisor underlies AIX and Linux partitions and is independent of either OS. The Linux kernels that interact with these two hypervisors differ in a few very low-level aspects like their page table management and interrupt handling, but the great majority of the kernel code is the same for iSeries and pSeries. The iSeries and pSeries specific diagrams shown here illustrate the slightly different hypervisor structures.

Linux distributions on both iSeries and pSeries systems run in a logically partitioned environment (LPAR). Both systems support coexistence of Linux with the other traditional operating system, OS/400 or AIX, respectively. pSeries systems also support using only Linux partitions. The partitioning is conceptually quite similar, though the interfaces for the partition management are different as shown in by the different console diagrams.

**Partitioning Basics**

A partition on iSeries or pSeries has its own processor allocation, memory allocation and I/O assignment. (In actuality, for I/O it is the slots that are assigned to particular partitions, but it is useful to discuss the assignment of I/O resources in end usage terms.) The hardware resources directly assigned to each partition are isolated from the other partitions. There are operational opportunities to move these resources around from one partition to another under sets of rules for each system. As a result, careful planning of workloads and resource sharing and movement is an important consideration in setting up application suites so that the customer gets the maximum use
of the hardware available. Where workloads differ over time, optimizing the resource scheduling across partitions on a single large system can achieve significantly greater hardware efficiencies compared to using a set of smaller stand-alone servers doing the same set of jobs. Because the iSeries hypervisor grows out of OS/400 technology, iSeries systems require a "primary" OS/400 partition. The other partitions on the systems are called "secondary" partitions. Linux distributions run in a secondary partition on iSeries and are given, at a minimum, their own processor resources and memory. A primary or a secondary OS/400 partition may "host" one or more Linux partitions, or the OS/400 partition may be used simply for initial installation of a non-hosted Linux partition on iSeries. When hosting a Linux partition, an OS/400 partition provides I/O resources and management for the Linux partition. If a Linux partition has all of its own I/O devices directly assigned to the Linux partition, it may be a secondary partition that is not hosted. More discussion of iSeries hosting and virtual I/O devices follows in the sections below.

pSeries servers support partitions, running either Linux or AIX in any combination. Each partition has its own full complement of hardware CPU, memory and I/O devices. On pSeries, there is no concept of a "primary" partition or of hosting. All partitions are on an equal footing. As mentioned earlier, some smaller pSeries systems can also run single stand-alone Linux images without using partitioning.

**Processor Partitioning Granularity**

A single iSeries processor of the appropriate feature code (located on the web site listed below) can, with OS/400 V5R2, be split into a maximum of 10 partitions. As noted earlier, each partition on iSeries is explicitly given its own processor and memory resources. Each partition requires a minimum of a tenth of a processor allocated to it whether active or not. The maximum number of partitions on an iSeries system is 32 which can be realized with four or more processors. Since the primary partition must run OS/400, there can be a maximum of 31 Linux partitions on this iSeries system. OS/400 and Linux partitions both can have processor resources moved dynamically from one partition to another without requiring reboot, giving substantial "on demand" capability to the hardware resources of the system. When new processor resources are made available by a Capacity on Demand processor activation, they can likewise be dynamically assigned to either OS/400 or Linux without disruption of the current workload.

Earlier V5R1 iSeries hardware models that provided Linux support were generally announced with the ability to be split into four partitions. Under V5R1, processor resources were dynamically movable for OS/400 without reboot, but Linux did require rebooting in order to take advantage of the resource change. Even earlier V4R5 models that provided Linux support required whole processor assignments to Linux partitions instead of partial processors, and Linux needed to be rebooted. All of these models and their levels of Linux support are explained at http://www.ibm.com/servers/eserver/iseries/linux/support.html

With the AIX 5.2 operating system, pSeries processor resources for AIX or Linux are allocated as whole processors. There is a maximum of 32 partitions per pSeries system with at least 32 processors on it. Processor allocations for AIX partitions can be changed dynamically without rebooting so that it can take advantage of Capacity Upgrade on Demand transparently. Linux needs to be rebooted to take advantage of changes in processor resource. More specific
information on individual pSeries models and Linux support is available at http://www.ibm.com/servers/eserver/pseries/hardware/factsfeatures.html.

Partition Management

pSeries partitioning is defined and managed using the Hardware Management Console (HMC). The HMC can provide a graphical user interface for partition resource management, stopping and starting partitions and displaying status. The HMC is used to load and manage firmware as well as to assign the processor, memory and I/O resources needed for each partition. The HMC is separate from AIX and Linux OS consoles.

The HMC runs on a dedicated Linux desktop workstation that is commonly connected to the pSeries system using a control cable from the HMC serial port. Remote connectivity options are available also. The fact that the HMC is a Linux system is incidental; it is really viewed as an appliance for managing the system partitioning. It has its own system login.

Up to 64 partitions residing on up to 16 non-clustered systems can be managed on one HMC. Redundant HMCs can also be configured. A pSeries partition is booted using a partition profile that identifies its hardware resources. Several partition profiles can be combined into a system profile that allows the defined partition set to be started at the same time. This capability facilitates defining a general workload with several parts that operate at the same time, yet need to be isolated from each other in the way that separate partitioning allows. A good example of this use would be in an ASP or service provider that offered a highly tailored set of firewall capabilities as well as back office operations for a client. The firewall could be in a separate partition, fully protecting the back office.

iSeries partitioning is managed on an OS/400 console using either text interfaces (5250) or, in V5R2, a graphical partitioning wizard provided in the iSeries Navigator operator interface that runs on a Windows system. The commands or GUI interact with OS/400 service tools that are accessing the primary partition where all of the partitioning information is maintained. The management software talks to the OS/400 hypervisor and requires the operator to use a protected dedicated service tool (DST) login. Processor and memory resources are allocated to each partition. If directly connected I/O devices (discussed below) are used, these are also managed through these interfaces.

Because there can be a maximum of 31 Linux partitions on an iSeries server with appropriate hardware, OS/400 can connect to that many local Linux partitions. OS/400 also has extensive remote systems management that can be used to work with remote iSeries systems for potentially a much larger span of control from one console. Booting Linux on iSeries can be done via OS/400 operator commands if the Linux partition is hosted (described below in the Linux Integration section) or by Linux commands.

I/O for a Linux Partition

I/O support for Linux is very straightforward on partitioned pSeries systems. Each Linux partition is assigned, owns and manages its own devices. In a partitioned system, device assignment is done using the HMC. The list of supported pSeries Linux devices is included in the earlier pSeries Linux
hardware reference. Many pSeries devices use the typical device drivers for their device type. For some additional devices, the ibmsis driver is used to access Linux-owned devices on pSeries. This driver has been included in some distributions in the past, but also can be downloaded from the University of Illinois web pages at ftp://lpatch.ncsa.uiuc.edu/redhat/rhel-3/pseries/ibmsis/

If you are working with an existing pSeries system configuration and adding Linux capability to it, be sure to check device support prior to installing Linux. Linux does not recognize some specific graphics adapters found in pSeries configurations, so these would need to be reconfigured for use in an AIX partition.

On iSeries, there are two types of supported I/O. One is what the iSeries team refers to as "direct" and it is like pSeries Linux I/O. The I/O device is assigned specifically to the Linux partition and is fully managed by the Linux OS. The lists of devices are linked from the Linux on iSeries web pages at http://www.ibm.com/servers/eserver/iseries/linux/support.html. There is significant overlap between the supported iSeries and pSeries devices, but there are some differences that are referenced in the two series web sites. As with pSeries, the ibmsis driver is used to attach some Linux-owned direct devices on iSeries and is available from ftp://lpatch.ncsa.uiuc.edu/redhat/rhel-3/pseries/ibmsis/ for download. A good example of the use of direct I/O is seen when configuring an iSeries partition for a Linux firewall. It is more advantageous to have the Linux partition directly connected to the "outside world" and to have it route the appropriate network traffic from and then to the OS/400 environments on the system since the Linux partition can protect the OS/400 system from external access.

The other type of iSeries I/O is referred to as "virtual." This is the most prevalent use of I/O on iSeries Linux partitions today. The Linux partition shares the I/O device that is being managed and serviced by OS/400. Virtual disk, tape, CD or DVD and communications adapters are supported. These virtual devices can minimize the cost of Linux partitions on iSeries since most Linux partitions can share OS/400 devices very effectively. They also benefit from the OS/400 first failure data capture and diagnostics if there are any problems with the devices.

The virtual disk for a Linux partition is called a Network Server Storage Space (NWSSTG). The majority of current iSeries Linux customers use virtual disk, and this is expected to continue. Using a single OS/400 command or Navigator action, a NWSSTG space is carved out of the iSeries file space as a single object and then associated with or "linked to" a Network Server Description (NWSD) that is used for a Linux partition, explained in the next section. A single NWSSTG can range from 1MB to 64GB. Up to 64 NWSSTGs can be linked to a single NWSD at one time, so the space potential is quite large for Linux partitions. In V5R2, read-only shared (between partitions) NWSSTGs are also supported. In addition, in V5R2, a NWSSTG can be added (linked) dynamically to a running Linux partition. These virtual disk devices can provide disaster recovery or efficient distribution mechanisms for the Linux partitions since they contain the entire system image when the Linux partition is not running. They can be backed up as individual SAV files by OS/400 and moved to other systems or quickly restored, then tailored to the networking environment.

Differences Between IBM eServer iSeries and pSeries Servers

Running Linux
Another key virtual I/O facility involves OS/400 support for up to 16 virtual LANs (VLANs) that can provide communications traffic across the iSeries bus between Linux partitions and OS/400. They act like a gigabit ethernet connection. These virtual LAN connections are defined with typical networking interfaces to virtual network devices with their own drivers in the Linux system software.

Types of I/O can be mixed and matched for Linux partitions on iSeries. A typical configuration for the firewall mentioned above is for the Linux partition to own the communications card that connects to the outside world while using virtual disks.

**Linux Console**

On the larger pSeries partitioned systems, there may be a limited number of serial ports, so virtual terminal interfaces on the HMC are typically used. It is generally not desirable to move a serial port assignment between Linux partitions because this would require a reboot of the affected partitions. The hardware password for systems management services (SMS) is required to establish the virtual terminal window. These virtual terminals are very useful for a centralized operations staff that is managing different partitions for different groups in a service provider environment. A console session can also be established via SSH so that a Linux GUI can be used for administration. On smaller stand-alone pSeries Linux systems, it may be desirable just to connect an ASCII terminal or terminal server to the serial port.

iSeries provides a virtual Linux console for each Linux partition through a unique mechanism. The console user accesses port 2301 on the primary iSeries partition via telnet. The resulting display allows the operator to select which partition to attach to. To complete the attachment, the user must enter the dedicated service tools (DST) password for the user defined with privileges to administer Linux. After satisfying the protected login for the partition, the user can log into the virtual Linux console as root or another user. This type of facility is used to good advantage in cases like a university environment where the Linux partition needs to be maintained by the operations staff, yet needs to be accessed independently by another group of Linux users; there may not even be a TCP/IP connection between the OS/400 management environment and the Linux partition so that each is completely isolated from the other except for the OS/400 ability to access the Linux console for management purposes.

**Linux Integration and Management Concepts**

It is not only possible but also likely that current IBM customers will mix partitions running the AIX operating system with partitions running the Linux operating systems on pSeries systems. However customers can also run a set of Linux partitions without AIX. The ability to run Linux on pSeries without AIX or other operating systems on POWER hardware is a capability of pSeries systems, but not iSeries systems.

In addition to supporting Linux in a partition, the new pSeries "Linux ready" systems can be used to run stand-alone Linux images. These systems do not need an HMC since they are not partitioned. They are just "business as usual" Linux systems. There is not currently an equivalent stand-alone Linux capability on iSeries.
There are several pSeries aspects of integration between AIX and Linux partitions to help customers maintain the system more easily and cost-effectively. The AIX and Linux security models are coordinated more easily because AIX has support for Pluggable Authentication Modules (PAM) to help align security on both systems. File sharing using Samba also assists in integrating use of the different environments.

For iSeries servers, the Linux partition management paradigm uses the concept of Network Server Descriptions (NWSDs) that represent the Linux partition operationally to OS/400. iSeries servers have been "hosting" other operating systems (Microsoft on the Integrated xSeries Server and its predecessors) for more than a dozen years using this model and virtual devices with it. When iSeries Linux support was designed, it was natural to introduce the Linux OS by reusing a honed, stable technology that was already familiar to iSeries customers. The NWSD is an entity that is recognized by OS/400 and can be managed by it. Linux partitions can be booted and shutdown from OS/400 operator interfaces by "varying on" and "varying off" the NWSD associated with the partition. The normal Linux shutdown and reboot commands also work, but since the OS/400 operator interface has capability for additional scripting and remote access, Linux partition management can be integrated into the same context as the other commercial workload management, like system backup, on OS/400.

An additional feature of the NWSD is the specification of where the kernel resides when Linux is booted using the OS/400 commands. There is an OS/400 concept of an "A side" and a "B side" that contain bootable OS/400 images; these are used by OS/400 customers to install and test fixes, yet have the previous production system available to return if production must be resumed before testing is completed. This concept has been extended to Linux support on iSeries servers as well, so there are two opportunities, a Linux "A side" and Linux "B side," to store bootable kernels. There is also the ability to specify that an OS/400 stream file contains a kernel to boot from, or the ability to specify that the kernel resides within the Linux virtual disk image described earlier. These options can give customers the ability to test with patched Linux kernels prior to putting them into production. These options are also very useful for developers who may want to try a variety of kernels during a day in their development cycle. Developers can write error logs from their experimental kernels prior to a crash, then link and boot up from a working NWSSTG and link to the NWSSTG with the error log for post mortem examination.

Other Details for Linux Tuning

For both iSeries and pSeries servers, the typical Linux kernel tuning parameters apply for building kernels for POWER systems and for optimizing kernels for a specific workload.

There are a few additional iSeries-specific details that may help with tuning Linux performance, depending on the specific iSeries hardware being used. The V5R1 iSeries hardware and the 800/810 V5R2 hardware have a "hardware multi-threading" parameter that OS/400 can set to "on" by the command

```
CHGSYSVAL SYSVAL(QPRCMLTTSK) VALUE(1)
```
to increase throughput for workload mixes that use a typically commercial level of I/O. This parameter is not used in POWER4 systems.

For other workloads that may generate a high rate of interrupts, either for communications traffic or other types of transactions, a Linux kernel IPL parameter specified on the NWSD lp_spread_events=n where n is the number of processors in the shared processor pool, may deliver faster responses to a high interrupt level by ensure that multiple processors have the chance to service the interrupts.

ISVs looking at application tuning should also remember to check their compiler options. iSeries and pSeries servers can use the VisualAge compilers from IBM, to generate well-optimized code for these systems. These can be found at http://www.ibm.com/servers/enable/linux/pdfs/vacpp.pdf.

Other iSeries and pSeries Software

The availability of IBM middleware for iSeries and pSeries POWER systems is covered in general in the "Commonality Across IBM eServer iSeries and pSeries Servers Running Linux" paper. Because the two systems use the same chip set, the binaries for many applications can be the same for Linux running on both. To see what is currently available, go to the Speed Start site for information and downloads at http://www-106.ibm.com/developerworks/linux/power/.

Summary

In summary, iSeries and pSeries systems have a lot in common in their support for Linux software in partitioned systems. pSeries also supports a stand-alone Linux image on some servers.

However, there are some areas where both iSeries and pSeries systems provide specific features and benefits to their Linux partitions with some, generally small, differences that can be used by ISVs to gain added value for their solutions on each of these server systems. Many ISVs will not need to use these differences, but some applications may benefit from specific platform tuning in specific cases. A summary of iSeries and pSeries facilities is shown in shown below.

<table>
<thead>
<tr>
<th>Facility</th>
<th>iSeries Facilities</th>
<th>pSeries Facilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linux coexistence with other environments</td>
<td>Partitioned environment, Linux coexists with OS/400; OS/400 primary partition required</td>
<td>Partitioned environment, Linux coexists with AIX and/or other Linux; standalone Linux environment on some models</td>
</tr>
<tr>
<td>LPAR operations interface</td>
<td>OS/400 console, local or remote for DST user</td>
<td>Hardware Management Console (HMC); local or remote</td>
</tr>
<tr>
<td>Linux system management</td>
<td>OS/400 operations interfaces (Navigator or OS/400 command line) for starting and stopping, I/O assignment; Linux virtual console via OS/400 port 2301</td>
<td>Hardware Management Console (HMC) virtual consoles; serial console option for stand-alone</td>
</tr>
<tr>
<td>Processor assignment</td>
<td>Sub-processor assignment out of processor pool, dynamic for OS/400 and Linux</td>
<td>Whole processor assignment, dynamic for AIX, Linux requires reboot</td>
</tr>
<tr>
<td>I/O device attachment</td>
<td>Virtual devices (shared with and owned by OS/400); I/O devices owned by the Linux partition; can mix and match</td>
<td>I/O devices owned by the Linux partition</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>LAN attachments</strong></td>
<td>Virtual LAN or direct I/O with communications adapter</td>
<td>Directly attached communications adapter</td>
</tr>
<tr>
<td><strong>Linux console</strong></td>
<td>Protected login to port 2301 on OS/400</td>
<td>Linux console via HMC virtual terminal; serial connection for stand-alone systems</td>
</tr>
<tr>
<td><strong>Database integration</strong></td>
<td>ODBC driver for DB2 UDB on OS/400 download, DB2 Connect on Linux can access data in DB2 UDB on OS/400, Java toolkit including JDBC</td>
<td>DB2 Connect on Linux can access data in DB2 UDB for AIX</td>
</tr>
<tr>
<td><strong>File Sharing</strong></td>
<td>Samba support in OS/400 NetServer, NFS</td>
<td>Samba and NFS support in AIX</td>
</tr>
<tr>
<td><strong>Additional Linux system facilities integration with other environments on the same system</strong></td>
<td>Starting and stopping Linux partition from OS/400 operator interface, disaster recovery for Linux using virtual disk</td>
<td>Enhanced common user management via PAM integration with AIX</td>
</tr>
<tr>
<td><strong>Hypervisor</strong></td>
<td>Hypervisor is part of OS/400</td>
<td>pSeries has OS-independent hypervisor; not used for stand-alone Linux systems</td>
</tr>
</tbody>
</table>
Related topics

- "Linux Applications on pSeries" (Red Book) at http://www.redbooks.ibm.com/abstracts/sg246033.html

© Copyright IBM Corporation 2004
Trademarks
(www.ibm.com/developerworks/ibm/trademarks/)