Dynamic logical partitioning for Linux on POWER

Setting up and using dynamic LPAR

Clifford Spinac

September 29, 2005

A powerful feature of the POWER5™ architecture is its capability to support server consolidation through the use of logical partitioning (LPAR). Dynamic logical partitioning increases the flexibility of partitioned systems by enabling administrators to add, remove, or move system resources between partitions without the need to reboot the partition. This paper is intended for Linux® on POWER™ system administrators or application developers who want to learn how to enable dynamic LPAR on Linux, how it is used to control a partition’s resources, and how to determine the changes in these resources.

Introduction

With logical partitioning (LPAR), IBM servers that feature the POWER processor can support multiple system requirements in a single machine. LPAR provides the capability for server consolidation, which helps optimize system resources and reduce system complexity. On servers featuring the POWER5 processor, up to 10 logical partitions can be configured per processor. Each partition is treated as a separate server, with the potential of providing a more efficient overall solution. However, varying requirements from application environments, as well as a mix of production and test environments, can require additional control of resources across the partitions. Dynamic logical partitioning provides this necessary flexibility to control resource allocation.

Dynamic LPAR enables system administrators to add, remove, or move system resources, such as I/O adapters and CPU allocations, between partitions without the need to reboot the partitions after they are modified. Dynamic LPAR allows you to move not only virtual resources, but also physical resources. This allows the administrator to dynamically assign the resources to the partition where they are needed, when they are needed. This also gives the administrator the capability to adjust the LPAR hardware assignments without impacting the system’s availability.

There are various application examples where dynamic LPAR support can improve the utilization of a server’s resources. One example includes the movement of a PCI I/O adapter between multiple partitions that require the adapter. When the applications in a partition require the use of the adapter, the resource is moved to that partition. Another example is the movement of CPU resources between partitions based on workload requirements for the partitions. This is very
common for a system that experiences changes in partition utilization requirements from one work shift to another. For instance, a server that has a Web commerce application partition and a report generation application partition could have a majority of its CPU resources devoted to the Web commerce partition during the first and second shift, but changed to the report generation partition during the third shift.

Installing the IBM productivity tool packages for Linux on POWER is required for dynamic LPAR support. After you install these packages, you can dynamically add or remove processors or I/O slots from a running partition using an attached Hardware Management Console (HMC). This article illustrates how to set up the Linux environment for dynamic LPAR, how to use the dynamic functions, and how software can detect the changes in the resources.

**Working with dynamic LPAR**

This article assumes that the reader has a working knowledge of Linux, POWER5 processor-based server hardware, partitioning concepts, and the HMC. To set up and use dynamic LPAR, you will perform the following steps:

1. Set up a POWER5 processor-based server with the Advanced Virtualization Feature installed.
2. Attach and configure an HMC to the server and make sure that the HMC can connect to the Linux partition.
3. Create a logical partition and install SUSE LINUX Enterprise Server 9 (SLES9) or Red Hat Enterprise Linux AS 4 (RHEL4).
   **Note:** For the most current dynamic LPAR support, you must install SLES9 Service Pack 2.
   Always install the latest SLES9 Service Pack or RHEL4 Update.
4. Download and install the IBM productivity tools support for dynamic LPAR.
5. Use the HMC to make changes to the dynamic LPAR’s resources.
6. (Optional) Install the libATA module for dynamic IDE CD/DVD ROM support.
7. Create a hotplug script to detect dynamic changes to PCI adapters or processor resources in a partition.

Information about items one through three can be found in articles on the Linux on POWER ISV Resource Center and the Linux on POWER Architecture developer’s corner, as well as IBM Redbooks (See Resources). This article focuses on items four through seven.

**Installing Linux support for dynamic LPAR**

After installing Linux, you must install the IBM productivity tool packages to support the dynamic LPAR functions. These packages also include the Resource Monitoring and Control (RMC) daemon, which communicates with the HMC.

Dynamic LPAR support is available for POWER5 running SUSE Linux Enterprise Server 9 or Red Hat Enterprise Linux AS 4. The packages can be downloaded from the following Web pages:

• Service and productivity tools for Linux on POWER on HMC-managed servers

As mentioned earlier, SLES9 Service Pack 2 is also required for the full dynamic LPAR functionality. In addition, the rpm packages rdist and compat are prerequisites and should be installed from the SLES9 distribution. Check the service and productivity tools Web site for additional requirements for future dynamic LPAR tools support.

Always download and install the latest service and productivity tools from the Web site. Also, you should consider installing all of the packages from the site for your Linux distribution, which also include the reliability, availability, and serviceability (RAS) tools.

At the time of the writing of this article, the dynamic LPAR tools include the following productivity tools:

**Table 1. Dynamic LPAR productivity tools**

<table>
<thead>
<tr>
<th>Platform Enablement Library</th>
<th>librtas-1.2-1.ppc64.rpm</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC</td>
<td>src-1.2.2.1-05054.ppc.rpm</td>
</tr>
<tr>
<td>RSCT utilities</td>
<td>rsct.core.utils-2.4.1.2-05070.ppc.rpm</td>
</tr>
<tr>
<td>RSCT core</td>
<td>rsct.core-2.4.1.2-05070.ppc.rpm</td>
</tr>
<tr>
<td>CSM core</td>
<td>csm.core-1.4.0.12-98.ppc.rpm</td>
</tr>
<tr>
<td>CSM client</td>
<td>csm.client-1.4.0.12-98.ppc.rpm</td>
</tr>
<tr>
<td>ServiceRM</td>
<td>devices.chrp.base.ServiceRM-2.2.0.0-2.ppc.rpm</td>
</tr>
<tr>
<td>DynamicRM</td>
<td>DynamicRM-1.1-2.ppc.rpm</td>
</tr>
<tr>
<td>PCI Hotplug Tools</td>
<td>rpa-pci-hotplug-1.0-12.ppc64.rpm</td>
</tr>
<tr>
<td>Dynamic Reconfiguration Tools</td>
<td>rpa-dlpar-1.0-13.ppc64.rpm</td>
</tr>
</tbody>
</table>

After downloading the dynamic LPAR tools packages, use the *rpm* command to install the packages in the order listed in Table 1. For example:

```bash
# rpm -Uvh librtas-1.2-1.ppc64.rpm
```

RPM will install and start the dynamic LPAR services. Wait for the services to start, or reboot the system. To see if the services are running, use the *lssrc* -a command:

```bash
# lssrc -a
```

The following services should display an active status:

- ctrmc
- IBM.ERRM
• IBM.AuditRM
• IBM.DRM
• IBM.CSMAgentRM
• IBM.ServiceRM
• IBM.HostRM

After the services are active (running), the partition will be able to receive dynamic LPAR commands from the HMC. To successfully communicate, the Linux partition and the HMC must have access to the same network and must be able to connect to each other. Therefore, the HMC can neither be isolated from the partition nor have a firewall blocking the connection.

Modifying logical partition resources

Now that the dynamic LPAR tools are installed and the services are running, the HMC can be used to make dynamic changes to the CPU or I/O adapter slot resources. Currently, dynamic LPAR does not support memory changes because of limitations in the SLES9 and RHEL4 Linux kernel.

Using the HMC to change partition resources

To change the resources of a dynamic LPAR, perform the following steps on the HMC:

1. Open Server and Partition and select Server Management.
2. Open the server that contains the partition, and open the partitions.
3. Right-click the logical partition and select Dynamic Logical Partitioning.
4. Select the resource type that should be changed (Physical Adapter Resources, Processor Resources, or Virtual Adapter Resources).
5. Select Add, Remove, or Move.

Continue through the change-specific dialogs to complete the dynamic change.

Figure 1 shows an example of using Physical Adapter Resources - Add to add a PCI I/O adapter slot.
**Figure 1. Adding a physical adapter resource**

**Figure 2** shows the selection of a PCI 1Gbps Ethernet adapter for addition to the partition.

**Figure 2. Adding a PCI 1Gbps Ethernet adapter to the partition**
In a similar manner, you can change the CPU resources. Figure 3 shows the dialog for making a change to the number of virtual processors in a shared processor partition. This dialog can also be used to increase the processing units, or processor entitlement.

Figure 3. Adding a virtual processor to the partition

Verifying partition resource changes

After using the HMC to make dynamic LPAR changes, you can verify the partition resource changes to see the effect of the changes. To view the current status of the partition’s PCI devices, use the `lspci` command:

```
# lspci
PCI bridge: IBM RADS-X PCI-X to PCI-X Bridge
Ethernet controller: Intel Corporation 82545GM Gigabit
```

The content of the `/proc/ppc64/lparcfg` file provides the current status of the LPAR’s processor resources. Listing 1 shows the lparcfg file for the partition after adding the virtual processor. The number of virtual processors, or a dedicated partition’s number of processors, can be seen in `partition_active_processors`. The value of a shared processor partition’s processing units can be seen in `partition_entitled_capacity`. Also notice that `shared_processor_mode=1`, since this is a shared processor partition.

Listing 1. Example lparcfg file for a partition

```
# cat /proc/ppc64/lparcfg
lparcfg 1.6
serial_number=IBM,0210BE5ED
```
Additional information about using dynamic LPAR and the HMC to make resource changes, as well as information about installing the dynamic LPAR productivity tools, can be found in the IBM eServer Hardware Information Center. (See Related topics.)

Adding dynamic support for IDE CD/DVD ROM

The Linux IDE device interface for CD/DVD ROM does not support hot swapping. Therefore, it cannot be used for dynamic LPAR changes to an IDE CD/DVD ROM drive. However, recent support in the Linux Parallel ATA (PATA) and libATA modules can be used to support an IDE CD/DVD ROM drive as a dynamic LPAR device. With this support, the IDE CD/DVD ROM appears to Linux as an SCSI device. This support is included with SLES9 Service Pack 2.

To use the dynamic LPAR IDE CD/DVD ROM support, the pata_pdc2027x module must be installed as an INITRD module. In the file /etc/sysconfig/kernel, add pata_pdc2027x to the beginning of the INITRD_MODULES, as shown in Listing 2.

Listing 2. /etc/sysconfig/kernel with pata_pdc2027x initrd module

```bash
## Path: System/Kernel
## Description: 
## Type: string 
## Command:     /sbin/mkinitrd
#
# This variable contains the list of modules to be added to the
# initial ramdisk by calling the script "mk_initrd"
# (like drivers for scsi-controllers, for lvm or reiserfs)
#
INITRD_MODULES="pata_pdc2027x ibmvscsic"
```
Now use the `/sbin/mkinitrd` command to create initrd, and then reboot.

```
# mkinitrd
# shutdown && now
```

The loaded modules can be examined for the `pata_pdc2027x` and `libATA` modules with the `lsmod` command:

```
# lsmod
Module                  Size  Used by
evdev                  31416  0
joydev                 31520  0
st                     73464  0
ipv6                  478392  29
sg                    74560  0
ibmveth                44800  0
usbcoure               185404  1
subfs                30168  1
dm_mod                 112744  0
ibmvscsic               44552  2
pata_pdc2027x          33908  0
libata                 97304  1 pata_pdc2027x
sr_mod                 44508  0
sd_mod                 44736  3
scsi_mod              199424  6 st,sg,ibmvscsic,libata,sr_mod,sd_mod
```

Now that the libATA device stack is loaded, the HMC can be used to add the IDE CD/DVD ROM drive. The drive adapter is listed on HMC I/O adapter lists as an *Other Mass Storage Controller.*

**Figure 4** shows the selection of "Other Mass Storage Controller" as an I/O adapter to add to the partition.

**Figure 4. Adding an IDE CD/DVD ROM drive to the partition**
To examine the partition's PCI devices after adding the CD/DVD ROM drive adapter, use the `lspci` command:

```
# lspci
0000:00:02.3 PCI bridge: IBM EADS-X PCI-X to PCI-X Bridge (rev 03)
0000:cc:01.0 Unknown mass storage controller: Promise Technology, Inc. 20275 (rev 01)
```

To find the SCSI device name for the CD/DVD ROM drive, use the `lsscsi` command to list all SCSI devices, and use `grep` to find cd/dvd:

```
# lsscsi | grep cd
[1:0:0:0]    cd/dvd  IBM     RMB00020501    H106  /dev/sr0
```

Now use the `mount` command to mount the SCSI device with the file system type iso9660 as read-only to `/media/cdrom`:

```
# mount -t iso9660 -r /dev/sr0 /media/cdrom
```

When the DVD/CD ROM drive is no longer needed by the partition, unmount the `/media/cdrom` directory and use the HMC to remove the "Other Mass Storage Controller" I/O adapter.

**HOTPLUG scripts**

Hotplug support for devices became a standard feature on Linux starting with the kernel 2.4. It initially included support for USB and PCI devices, as well as support for automatically configuring some network interfaces. With hotplug support, you can often plug in a new device and use that device immediately because the system can use a hotplug event to auto-configure it.

A hotplug event can be used to invoke an agent task to configure an adapter, subsystem, or program. Agent tasks utilize scripts that are used to process and manage this configuration. The scripting framework is referred to as hotplug policy agents. The agents are tied together by a common entry point in the `/sbin/hotplug` script. This script can be used to debug and log event changes, as well as control the system's hotplug agents. The hotplug agents are script files placed in the directory `/etc/hotplug`. One or more agent types, with the name type.agent, can be invoked for a hotplug event.

**Using HOTPLUG scripts to detect resource changes**

POWER5 processor-based servers running SLES9 or RHEL4 with the IBM productivity tools support `/sbin/hotplug` event notification. These events include PCI I/O adapter slot changes.

For example, when an Ethernet adapter is dynamically added to a partition, the following agents are invoked to configure the adapter and network: `pci_bus.agent`, `pci.agent`, and `net.agent`. Some of these agents are invoked more than once to allow for capturing the event changes at different stages.

In addition to dynamic LPAR I/O changes, SLES9 with Service Pack 2 includes support for invoking a `cpu.agent` when dynamic LPAR changes are made to the number of virtual processors.
or dedicated processors. Therefore, a cpu.agent script can be added to the /etc/hotplug directory to enable monitoring CPU change events. This script agent can be used for scaling multi-threaded applications. It can also be used by license management/entitlement software and system management tools, such as performance monitoring tools. The script can be used to modify user space parameters or to invoke different applications based on the CPU change.

Processing units (entitlement capacity) changes are not reported as Linux hotplug events because these events are transparent to the operating system.

The support for a cpu.agent is not complete in RHEL4 at the time of the writing of this article. Check future updates to RHEL4 for the addition of this support.

Creating a CPU agent script
Dynamic LPAR event scripts can use standard hotplug event script programming techniques. Several event scripts can be found in /etc/hotplug and used as guides for event programming.

A CPU event agent for SLES9 SP2 can be created by using the example template in /etc/hotplug/generic_empty.agent. Listing 3 shows an example of a cpu.agent script. This script will be invoked for dynamic LPAR CPU change events, with the hotplug actions online and offline.

Listing 3. Example of a /etc/hotplug/cpu.agent script

```bash
#!/bin/bash
# $Id: cpu.agent
# Kernel hotplug params include:
# ACTION=%s [online or offline]
# DEVPATH=%s
#
cd /etc/hotplug
. ./hotplug.functions

case $ACTION in
  online)
    echo `date`":cpu.agent" add cpu >>/tmp/hotplug.txt
    ;;
  offline)
    echo `date"":cpu.agent" remove cpu >>/tmp/hotplug.txt
    ;;
  *)
    debug_mesg CPU $ACTION event not supported
    exit 1
    ;;
esac
```

Summary
Logical partitioning provides POWER processor-based servers with the capability to do server consolidation and optimize system resources. Dynamic logical partitioning enhances this capability by providing control of the allocation of the resources without impacting the logical partitions availability. Linux on POWER supports dynamic LPAR for changes to physical I/O, virtual I/O, and processor resources.

This article has introduced how dynamic LPAR support is enabled for Linux on POWER and controlled with the Hardware Management Console. It also identified how software can monitor for
changes in these dynamic LPAR resources. For more information about dynamic LPAR for Linux on POWER, refer to Related topics below.

**Acknowledgements**

I would like to acknowledge Linda Kinnunen and James Mathewson for their helpful editorial reviews. I would also like to thank John Rose, Doug Maxey, and Joel Schopp for their technical assistance with this article.
Related topics

- **Service and productivity tools for Linux on POWER** provides hardware service diagnostic aids and productivity tools for IBM servers running Linux operating systems on Power4 and Power5 processors.
- **The IBM eServer Hardware Information Center**: Your source for technical information about IBM eServer(TM) i5, eServer p5, and eServer OpenPower(TM) servers.
  - "**Dynamically managing Linux logical partition resources**" provides information so that you can dynamically add, remove, and move resources between partitions.
- **Linux Hotplugging** provides an overview of hotplugging and the Linux operating system.
- **Linux on POWER ISV Resource Center** gives you access to the various available for Linux on POWER systems.
- **Linux on Power Architecture Developer's Corner**: Learn more about Linux on Power. Find technical documentation, education, downloads, product information, and more.
- **Learn about Linux at IBM**: See what's going on with Linux within IBM.
- **IBM Redbooks**: One of the most comprehensive resources available to developers and administrators who want the best, detailed technical information on IBM products and development techniques that extend to all brands.

© Copyright IBM Corporation 2005

**Trademarks**
(www.ibm.com/developerworks/ibm/trademarks/)