The Truth Behind IBM AIX LPAR Performance

Yann Guernion, VP Technology
ABSTRACT

Partitioning can simplify IBM server management and reduce overall operating costs. However, despite these advantages, performance tracking and problem solving in a virtual environment can become difficult or even impossible to handle.

This white-paper introduces and explains different concepts linked to active processor resource allocation for LPARs and sheds light on the complications that result from the abundance of available configuration possibilities.

Virtualization, more specifically, running numerous operating systems simultaneously, introduces a new level of abstraction that increases the risk of physical resource saturation tenfold. Performance management has once again become a strategic issue to contend with the need to maximize server usage rates.

Dedicated solutions, such as SP Analyst for AIX on pSeries, serve as a logical add-on to the system administration capabilities already in place. These solutions provide the granularity of data required to objectively manage system resources, and to monitor the impact of applications running in the logical partitions.

TABLE OF CONTENT

1. THE POWER VM HYPERVISOR ........................................................................................... 3
2. LOGICAL PARTITION (LPAR) .............................................................................................. 4
2.1 LOGICAL PARTITION ......................................................................................................... 4
2.2 MICRO-PARTITION ......................................................................................................... 4
2.3 CAPPED AND UNCAPPED MODE .................................................................................... 4
2.4 LOGICAL AND VIRTUAL PROCESSORS ............................................................................. 5
2.5 IBM AIX PERFORMANCE MANAGEMENT ......................................................................... 6
2.6 SYSLOAD FOR IBM AIX AND PSERIES ......................................................................... 12
3. ABOUT ORSYP .............................................................................................................. 15
1. The Power VM Hypervisor

The PowerVM Hypervisor, which facilitates separating a server’s material resources logically into partitions, constitutes the foundation of the IBM System P series’ virtualization concept.

It is comprised of a layer of firmware that is situated between the operating system and the physical server.

The hypervisor is predominantly responsible for redistributing the logical partition burden on the server’s available physical resources.

Though the hypervisor is installed and active regardless of the physical configuration, it is not bound to dedicated processors or specific resources.

It is possible to outline the hypervisor’s principle objectives in a simple manner:

- Supply the partitions with a layer of abstraction from the physical resources.
- Guarantee partition integrity by logically isolating them from one another.
- Handle physical resource sharing between the partitions.

Because of this level of abstraction, a logical partition cannot access the physical processors directly; the hypervisor presents processing resources in the form of virtual processors.

Current hypervisor implementation authorizes a physical processor to support up to 10 virtual processors. Therefore, there is not necessarily a unit relationship between the
number of physical processors and the number of virtual processors seen in the partitions.

2. Logical Partition (LPAR)

2.1 Logical Partition

A logical partition can be seen as a logical server in which it is possible to launch an operating system and run programs on it. The partitioning technology was made available with the POWER4 processor. This platform made it possible to dedicate processing resources to a logical partition. However, the idea of multiple partitions sharing the same physical processor was unimaginable.

2.2 Micro-Partition

On POWER5 or POWER6 processors, it is still possible to dedicate a certain number of processors to a logical partition. But thanks to virtualization technology enhancements, multiple partitions can share a pool of physical processors (shared pool).

These logical partitions, using the shared physical processors pool, are called micro-partitions.

Micro-partition capacity can be finely adjusted in processing units down to a granularity of one hundredth (0.01) of a physical processor. The minimum quantity is fixed at one tenth (0.10) of a physical processor.

For example, a system comprised of 2 physical processors would be able to allocate up to 2.00 processing units and create a maximum of 20 micro-partitions.

The quantity of processing units reserved for a micro-partition is called entitled capacity. This capacity is guaranteed to the micro-partition regardless of the physical system operating conditions.

2.3 Capped and Uncapped Mode

Micro-partitions can be configured using two different methods of setting the available capacity usage mode in the shared pool: Capped or Uncapped.

**Capped Mode**

A micro-partition in capped mode can only use the entitled capacity that was initially allocated to it.

**Uncapped Mode**

A micro-partition in uncapped mode can use resources in addition to what was initially allocated by the entitled capacity if more resources are available in the shared pool of physical processors.

If several micro-partitions are competing for access to the supplementary resources, the hypervisor initiates arbitration by leveraging the weight configured by the user for each.
A partition’s weight is represented by an integer between 0 and 255 with a default value of 128. The available capacity is distributed to competing uncapped partitions using a quota equal to the ratio of the partition’s weight to the sum of weight of all partitions.

For example, a partition weighing 128 will receive only 20% of the unused available capacity if it is competing with two partitions weighing 225 (128/(128+225+225)*100). However, a partition weighing 0 will never receive additional capacity.

2.4 Logical and Virtual Processors

The capacity allocated to a micro-partition is presented to the operating system in the form of logical processors, which rely on virtual processors.

The correspondence mechanism between the logical and virtual processors depends on the simultaneous multithreading (SMT) mode configured.

Virtual Processors

The entitled capacity of the partition is equally shared between all virtual processors. The number of virtual processors is user configurable. For example, if a partition has a 1.2 entitled capacity with 3 virtual processors, each virtual processor will have a 0.4 physical processor capacity.

The maximum capacity for a virtual processor cannot however exceed 1.0, meaning one (whole) physical processor. This concept can be easily understood if we consider that the hypervisor must associate one virtual processor to one physical processor in a 1 to 1 relationship at any given moment.

The result is that a micro-partition must be configured with a sufficient number of virtual processors to utilize all of its entitled capacity.

An example would be a micro-partition with a 2.5 capacity yet configured with only 2 virtual processors: each virtual processor would be allocated a 1.25 capacity – which
is impossible considering the previously stated principles. The hypervisor cannot allocate 1 virtual processor to 1.25 physical processors. It is thus necessary to configure 3 virtual processors for a 2.5 capacity – each one having a 0.83 capacity.

The necessity of granting a micro-partition the right number of virtual processors according to its maximum capacity poses a problem for uncapped partitions as their maximum capacity is not known in advance.

For example, an uncapped partition with an entitled capacity of 2.0 and 3 virtual processors cannot use more than a 1.0 additional capacity, even if other processing units are available in the shared pool.

This situation makes it necessary to statically configure a number of virtual processors greater than the entitled capacity of an uncapped partition even if a capacity increase is only needed for a short period of time.

However, keeping unused virtual processors online means costly management that impacts the operating system’s performance. It is highly recommended that administrators define this parameter by long term monitoring of the partition and eventually using the virtual processor folding features to dynamically put inactive processors offline.

**Logical Processors and SMT**

Available on POWER5 and 6 processors, the simultaneous multithreading (SMT) feature, not to be confused with symmetric multiprocessing (SMP), enhances throughput by allowing more than one thread to run at the same time on a single micro-processor.

In reality, the performance gain is obtained by reallocating the wait I/O times of blocked threads to other threads.

SMT use does not make the micro-processor more powerful, but makes it run multi-threaded applications more efficiently by reducing the latency time and implicitly boosting its throughput.

Not all workload types benefit from SMT’s contribution. Notably mono-threaded programs or particularly well optimized applications that are subject to a very low level of wait I/O. For this reason, POWER processors support a single threaded (ST) mode, which dedicates the whole processor to the active thread.

In practice, it is possible to dynamically select the execution mode (ST or SMT) for each logical partition. The operating system running in the partition then distributes the threads to the different logical processors, knowing that each virtual processor presents one or two logical processors to the operating system depending on whether the partition is in ST or SMT mode.

### 2.5 IBM AIX Performance Management

#### 2.5.1 Logical Processor Usage

Logical processor usage monitoring in an LPAR AIX can be carried out the same way as on a traditional physical system, using common metrics:

- %user : percentage of time spent running user code (applications)
- %sys : percentage of time spent running system code (kernel)
- %wait : percentage of time spent awaiting I/O termination
- **%idle**: percentage of time not used
- **lbusy**: logical processor usage rate

It is however, important to note that the first four indicators give data that is relative to the partition’s entitled capacity and not the entire physical machine. This can make reading and interpreting difficult if ever the capacity is changed or the partition is in uncapped mode.

An example to study is a partition with 0.5 entitled capacity which has its logical processor activity at 50% in user mode (%user). If the partition is capped, we can reasonably assume that the physical capacity consumed in user mode can be calculated as 0.5*50%, meaning 0.25. However, if the partition is uncapped, estimating the correct capacity consumed is not possible by reading only the logical processor usage metrics.

The lbusy metric shows the logical processor usage rate and thus implicitly, the virtual processor usage rate in an LPAR.

- **An lbusy rate close to 100% corresponds to a complete utilization of SMP or SMT capabilities of the logical partition. In this situation, adding additional virtual processors can further improve program execution performance.**
- **A low lbusy rate sheds light on applications’ inability to take part in SMP or SMT parallelism; in this case it is possible and recommended to reduce the number of virtual processors.**

For example, an LPAR that has two logical processors will show an lbusy of around 50% if a mono-threaded program is run on it. If we run the same program twice, the lbusy rate will jump to a value close to 100%.
2.5.2 Physical Processor Usage

Physical processor usage monitoring in a POWER system can be carried out in each LPAR looking at specific metrics:

- physc: quantity of processing units used by the logical partition (in hundredths of units)
- ent: quantity of processing units allocated to the logical partition (entitled capacity in hundredths of units)
- pool size: quantity of processing units in the shared pool to which the logical partition is attached

With an uncapped partition, the quantity of processing units consumed (physc) can be greater than the partition’s entitled capacity (ent). Although, this consumption will always be lower than the shared pool (pool size) capacity.
Physical processor usage by the logical partitions can also be displayed as a ratio using the following metrics:

- **%app**: percentage of unused processing units (available) in the shared pool, relative to the capacity of the shared pool (pool size).
- **entc**: percentage of processing units used by an LPAR, relative to its entitled capacity.

It is important to note that these two indicators are determined relatively to different quantities (shared pool capacity and entitled capacity), thus their evolution is not directly proportional.

The usage rate (entc) can therefore exceed 100% if an uncapped partition is using a greater capacity than its configured entitled capacity.
2.5.3 Processor Saturation

Running multiple operating systems simultaneously inside logical partitions statistically raises the risk of resource contention and notably processor saturation.

Processor saturation can be characterized by an increase in the number of processors awaiting execution because there is a lack of available execution cycles. Saturation levels thus depend on the number of active processes (or threads) and the number of execution cycles to be distributed. Applications and users bound to processes that cannot gain processors see their response times increasing proportionally with the size of the run queue.

With logical partitions, saturation levels can be observed in the same manner as within a physical server through metrics relative to number of processes waiting for a processor (run queue length or load average).

- **When the run queue length is greater than the number of logical processors allocated to the partition, the system most likely does not have sufficient execution cycles available and hence finds itself in a contentious situation.**

Nevertheless, it is important to confirm this situation by monitoring the usage rates: %user, %sys and %wait, in order to ensure that the contention is not due to I/O problems, which will be indicated by %wait > %user + %sys.
Sysload SP Analyst chart showing the processor queue was full during a few hours; it then takes about one hour to come back to normal.

Sysload SP Analyst chart showing the load average on the 3 logical partitions of a frame; processes in the VIO server sometimes have to queue.

If we consider a traditional physical server, most of the time a processor contention problem calls for an upgrade or even going as far as to change the machine.
With a virtualized server, it is easier to address a logical partition processor contention problem. As shown previously, saturation is caused by an execution cycle deficit.

*It is very simple to grant a certain number of additional execution cycles by increasing the entitled capacity and the number of virtual processors in the partition. Overweighting an uncapped partition can also reduce processor saturation by allocating it execution cycles with a higher priority.*

### 2.6 Sysload for IBM AIX and pSeries

To address Data Center heterogeneity and complex virtual environment monitoring issues, Sysload has developed a high-performance data acquisition technology that facilitates essential indicator collection and analysis for virtualization project management: from outset to production.

Sysload provides solutions for IBM AIX and pSeries environments facilitating precise and continuous performance tracking of usage and available capacity on the physical machine and in the different partitions.

- Consolidated data center view (physical and virtual servers)
- Unrivalled granularity of analysis
- Automatic micro-saturation detection
- Negligible system impact
- Interactive and flexible console
Example showing both server status with events and resources utilization

**Performance indicators on LPARS and applications**

The SP Analyst agent for AIX collects precise data on LPAR performance thus adding to overall performance management. The domains that are monitored: processor, memory, disk, file systems, processes (including those in the workload partitions), etc. – permit system administrators to analyze problems and better define the parameters of their systems.

In order to provide LPAR and application visibility that corresponds to IBM technology, the Sysload agent produces AIX-specific metrics: ‘ent’, ‘entc’, ‘physc’, ‘lbusy’, CPU pool size, etc.

**Physical Machine Visibility**

SP Analyst is able to recreate physical server visibility by reconstructing the data collected by each LPAR. This visibility facilitates maximizing the ROI on the physical machine by measuring if the LPAR is using the machine’s full capacity or if there is capacity available to support additional LPARs.

**Consolidated Data Center Visibility**

Sysload offers a complete range of multi-platform solutions supporting a wide variety of hardware, operating system and specific virtualization environment (VMware, SUN, HP, IBM) options.

Sysload’s data acquisition technology uses collection agents that are loaded directly onto the system, allowing them to objectively monitor all physical and virtual servers.

**Unrivalled Granularity of Analysis**
The Sysload agents discretely collect up to 300 metrics from the system (requiring less than 1% CPU utilization with no induced network load) at very high frequency (down to 1 second intervals).

This fine level of granularity guarantees precise usage monitoring and facilitates identifying malfunctions due to logical partition micro-saturation that would otherwise go undetected by most monitoring tools.

**Automatic Micro-saturation Detection**

By collecting data directly from the system kernel, Sysload is able to monitor and identify the specific application processes that are responsible for the quality of service experienced by end users. Thus, Sysload agents can automatically detect saturation-related malfunctions and report them to the respective application and system teams.

**Support for planning server deployment and upgrades**

Virtual servers and resources have completely changed the traditional server resource management cycle. Sysload helps IT administrators identify the exact utilization and capacity levels for all resources necessary to maintain service levels.

Offering precise historical data down to 5 minute intervals, Sysload facilitates profiling machine activity (typical day, typical week, etc.) and determining their compatibility within a resource pool.

**Homogeneous Metrics**

In order to have a homogeneous indicator, processor consumption is converted into physical processors, the unit common to all logical partitions on a frame. This makes it simple to directly determine the impact of applications on a resource pool.
Negligible Impact on IBM Infrastructure

Sysload agents’ extremely light impact on an OS (less than 1% of CPU) minimizes the physical resources required to achieve optimal server performance management regardless of how many partitions are being monitored.

Interactive and Flexible Console

Thanks to its high level of interactivity, the Sysload SP Analyst console provides the ability to shift effortlessly between real-time and historical views, from performance, trend analysis and reporting, to troubleshooting, with 2-3 intuitive mouse clicks. Global mosaics and/or Enterprise table views provide a high-level starting point for quick, easy and efficient drill down capabilities to pinpoint performance issues.

3. About ORSYP

ORSYP is an independent IT Operations Management solutions provider helping customers assure that IT services are delivered on-time. Headquartered in Paris, France, Boston, USA, Hong Kong, China, ORSYP has more than 20 years of growth and over 1400 blue chip customers. ORSYP software, including Enterprise Job Scheduling, IT Automation, Performance and Capacity Management and also ITSM consulting services, are trusted and proven in some of the world’s most demanding physical and virtual environments. We strive to provide customers with the assurance that time to delivery of IT Operations services is properly and effectively managed today, and tomorrow.