IBM Technical Forum & Executive Briefing

17 al 21 Octubre 2011

Imagine PODER Imagine CAPACIDAD

Server Consolidation on Power Systems

Cesar Diniz Maciel
Executive IT Specialist – IBM Power Systems
Global Techline – Latin America
cmaciel@us.ibm.com

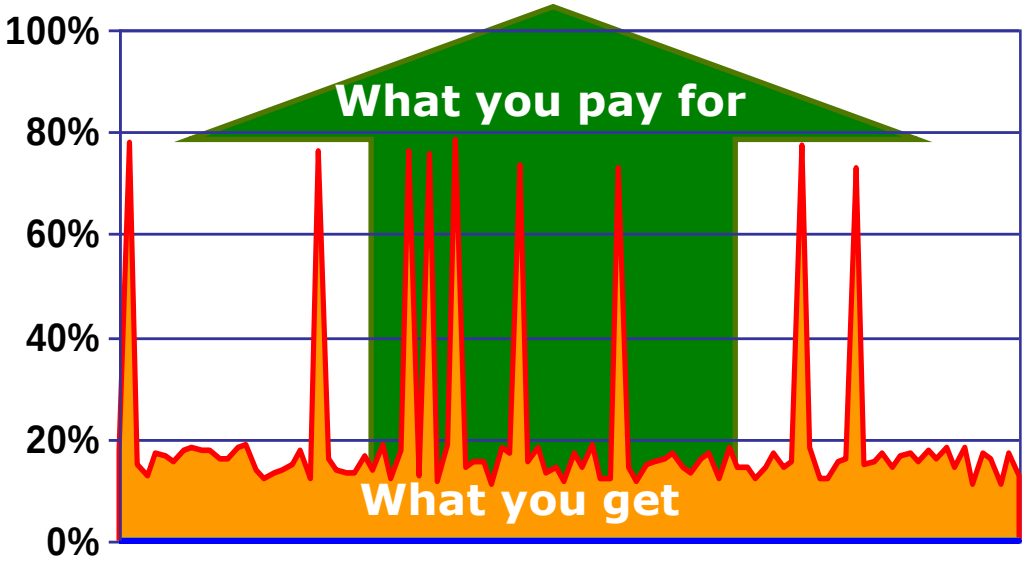
Agenda

- The reason for server consolidation
- Benefits of virtualizing a server infrastructure
- Tools and methods for collecting data for a server consolidation study

Low Utilization Drives Up Cost

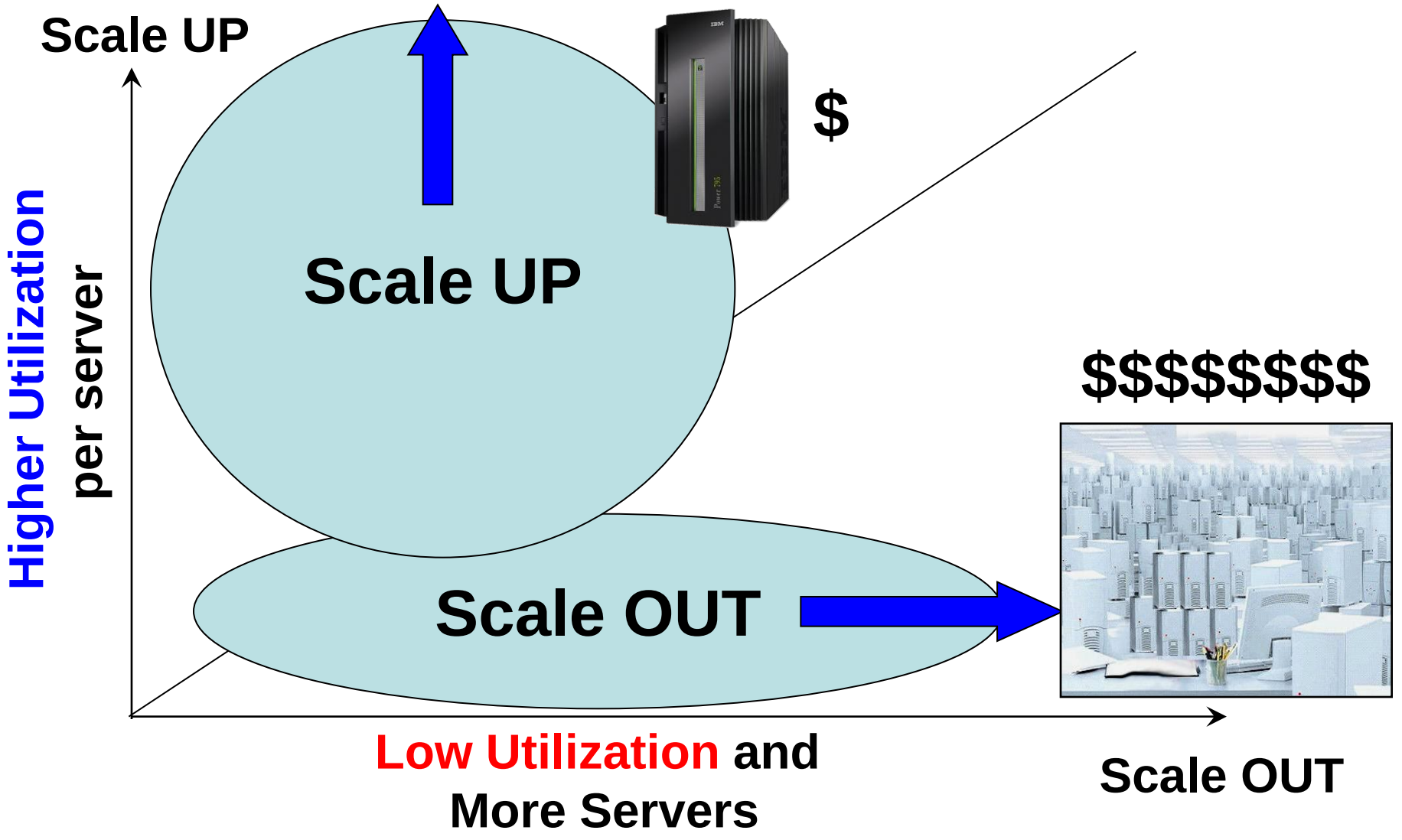
The typical UNIX or x86 serving running a single operating environment is only 10 - 20% utilized

- System waits for I/O and memory access even when it is working
- Configuration planned for peaks
- Configuration planned for growth

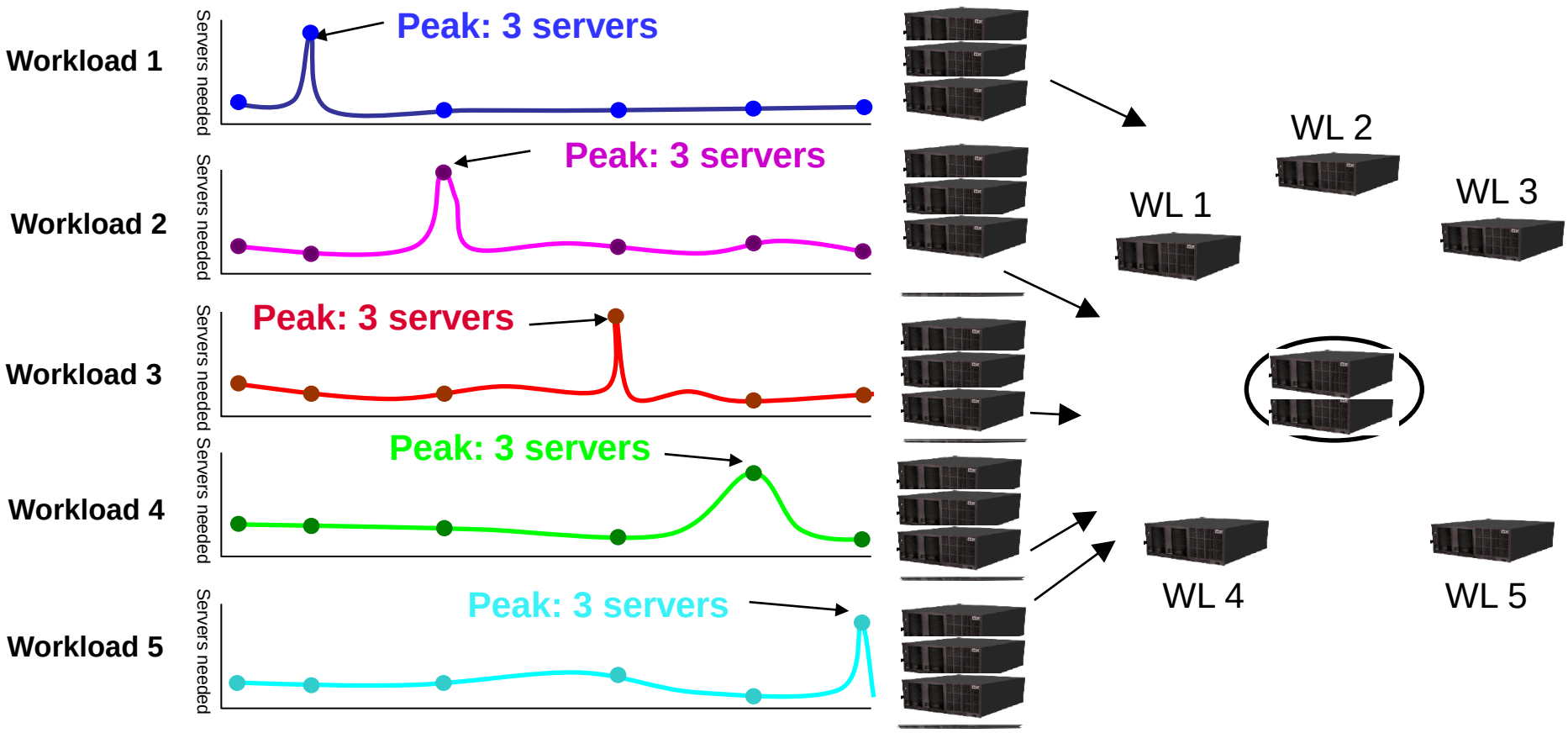


Result is that 80% of the hardware, software licenses, maintenance, floor space, and energy that YOU pay for, is wasted

Higher Utilization Servers are Key for Cutting Cost



How Virtualization Helps Cutting Cost Through Consolidation



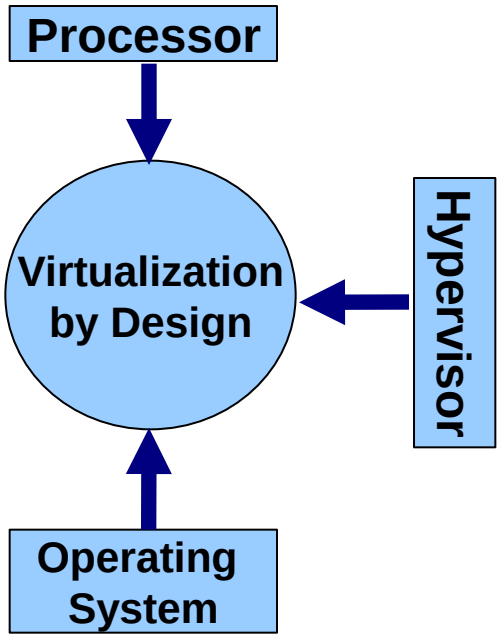
A robust hypervisor can do this type of resource sharing with CPU's, Memory, Networks and I/O

Power Systems Virtualization is Part of the Platform Design, and has been Since 2001

IBM Mainframe inspired hypervisor built into Power from the ground up *

PowerVM is combination of hardware and firmware that provides CPU, memory, network and disk virtualization

- **Best performance which means lowest cost per workload**
  - Hypervisor is integrated into POWER7 Hardware
  - No software overhead or “fix as you go” on the platform components
- **Richest set of capabilities for Flexibility**
  - All components (CPU, Memory, Network, I/O) are aware of virtualization environment and managed dynamically
- **Integrated Dynamic Management System**
  - Based on System Director
  - CPU, Memory, Network, I/O
- **Impenetrable Security and Reliability**
  - Addressed at design – not an add-on
  - Integrated into the firmware and hardware



*** News Flash:**

1967 - IBM develops the world’s first **hypervisor** called “VM” for S/360

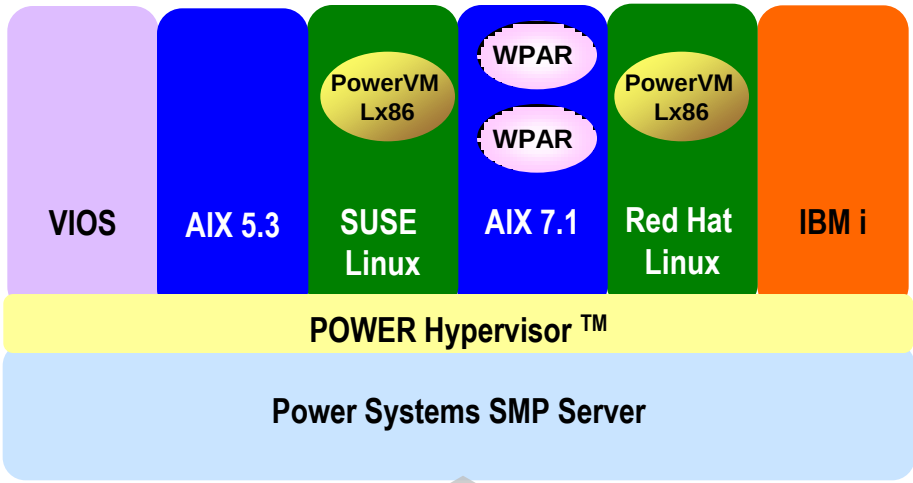
PowerVM Capabilities

The Virtual I/O server (VIOS) enables virtual servers to share I/O hardware.

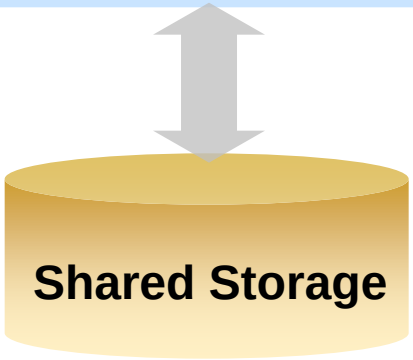
AIX allocates resources to Workload Partitions (WPARs).

The In-Memory Virtual Ethernet enables high speed memory to memory networking between partitions.

Logical Partitions (LPARs) are virtual servers that provide operating system and application isolation.

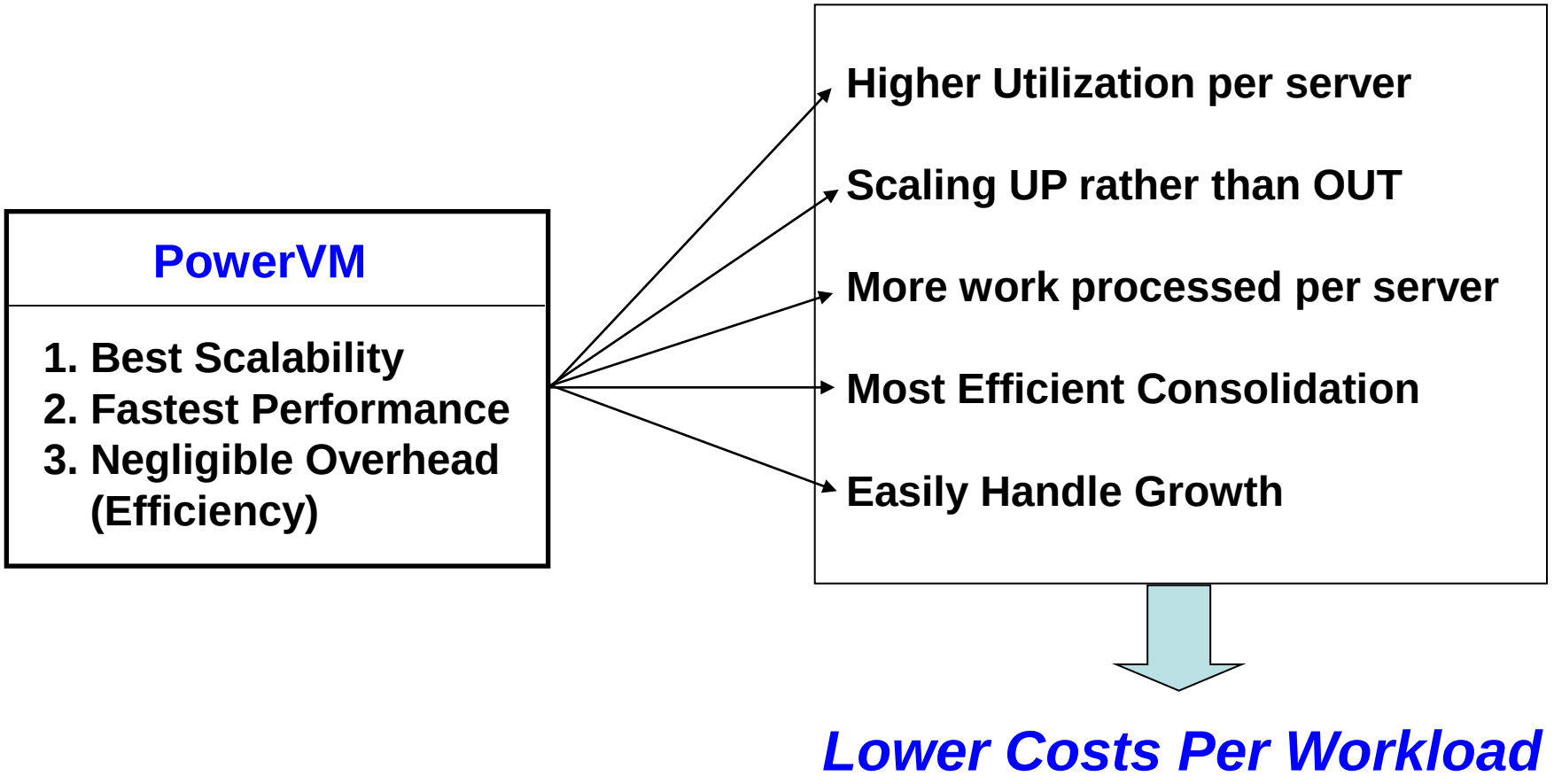


Shared access to disk storage and to external networks.



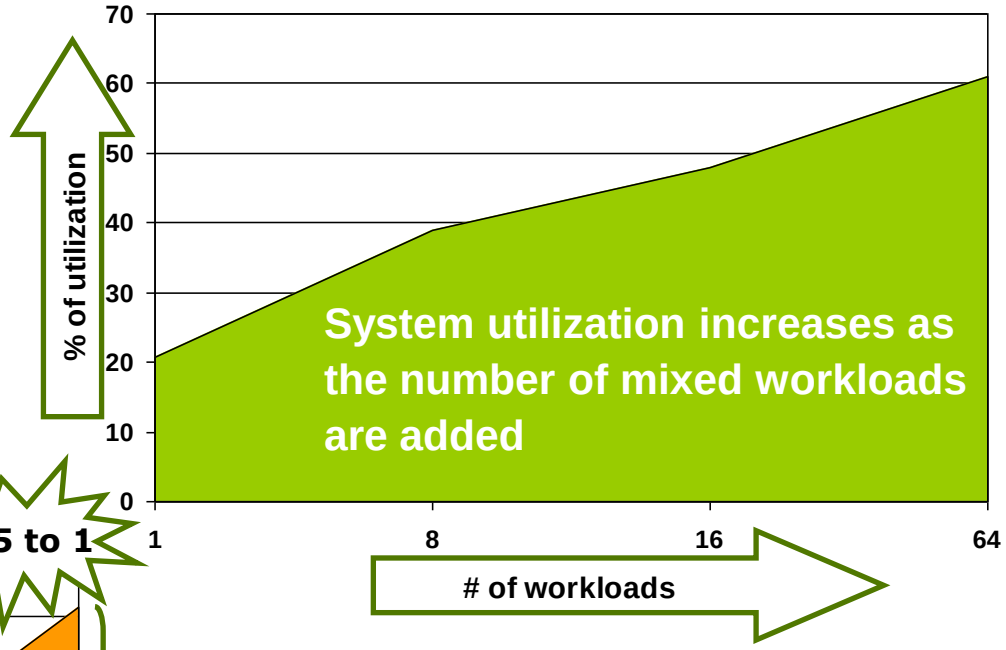
The Power Hypervisor shares processing resources among LPARs with up to 1024 dispatchable threads on 256 processors.

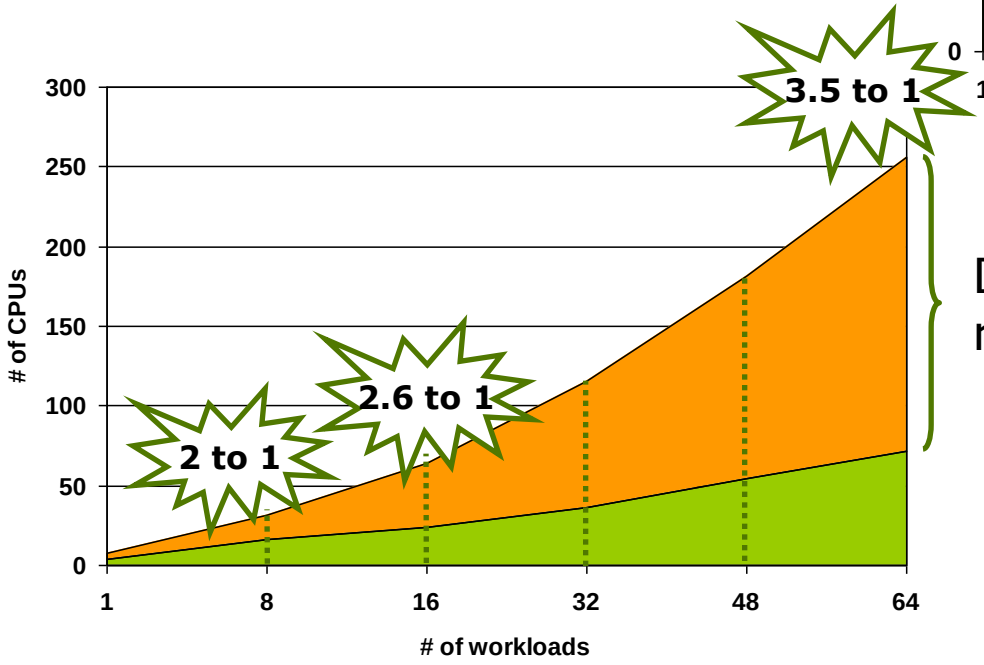
Why is it Important to Lead in Scalability, Performance and Efficiency?



Statistical Multiplex Causes Consolidation Ratios to Increase with Server Size

- The amount of “leverage” increases as the number of workloads are added
- The potential for savings increases with the amount of consolidation



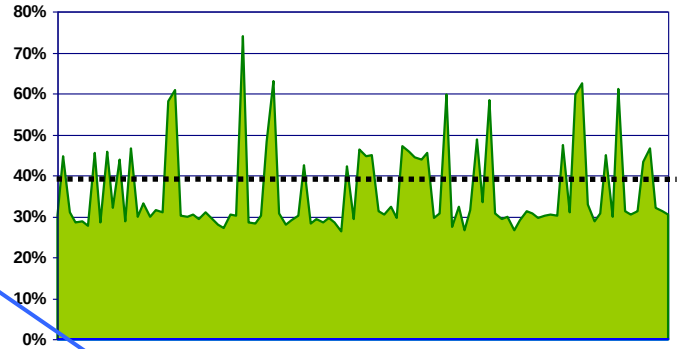


Delta in system capacity as a result of shared resources

Statistical Multiplexing Allows You to Drive System Utilization Up

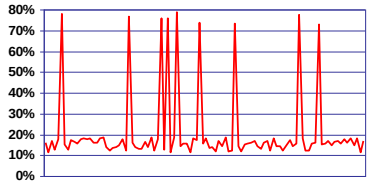
- **8** separate workloads on **8** identical systems
  - Average utilization is 17%
  - Peak is 6 times the average
- **8** separate workloads on one system*
  - **Average utilization is 36%**
  - Peaks is 2.76 times the average

8 to 1 Systems Consolidation (16 cores)

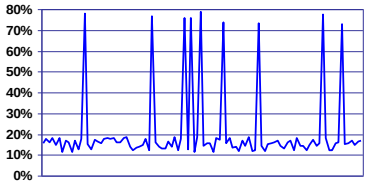


32 cores reduced to 16 cores (2 to 1)

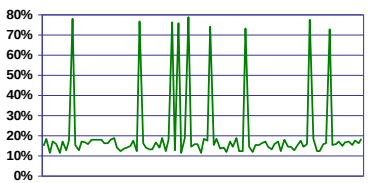
Single Application Server (4 cores)



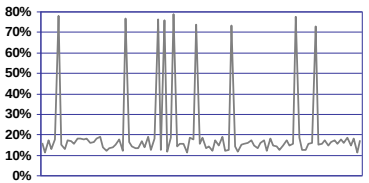
Single Application Server (4 cores)



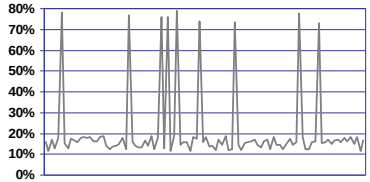
Single Application Server (4 cores)



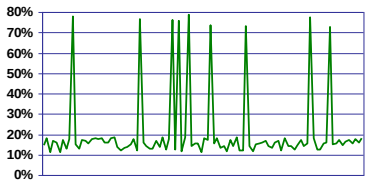
Single Application Server (4 cores)



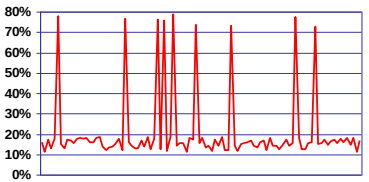
Single Application Server (4 cores)



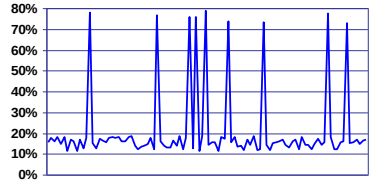
Single Application Server (4 cores)



Single Application Server (4 cores)



Single Application Server (4 cores)

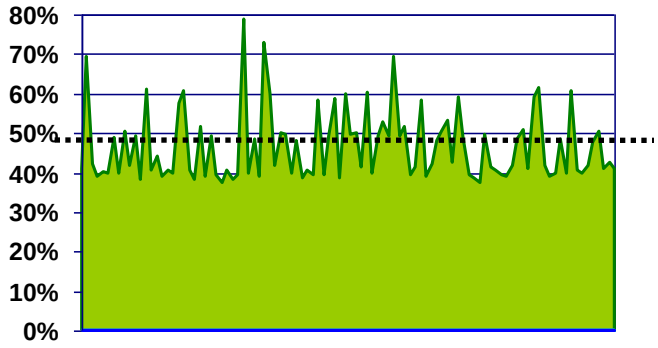


* Assumes independent workloads

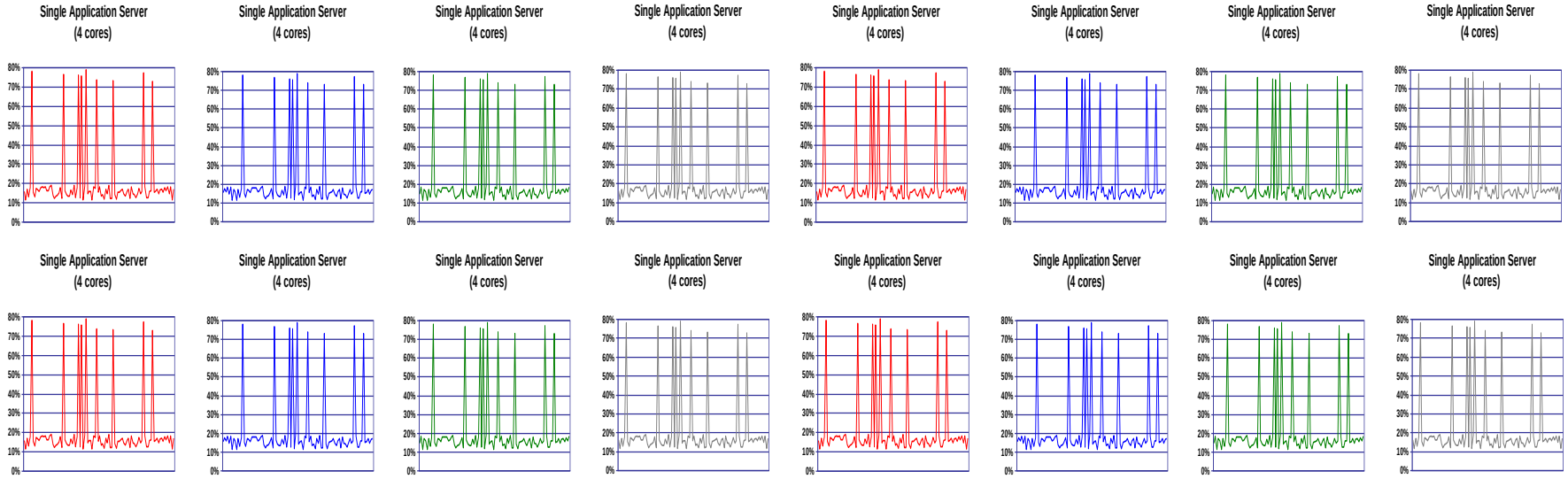
Statistical Multiplexing Shows that Larger Servers can Consolidate More

- **16** separate workloads on **16** identical systems
  - Average utilization is 17%
  - Peak is 6 times the average
- **16** separate workloads on one system*
  - **Average utilization is 44%**
  - Peaks is 2.25 times the average

16 to 1 Systems Consolidation (24 cores)



64 cores reduced to 24 cores (2.65 to 1)



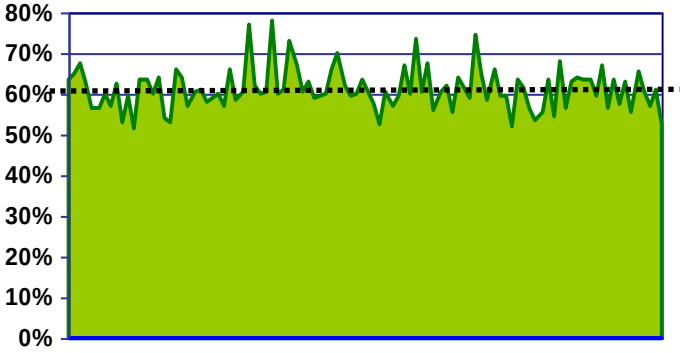
* Assumes independent workloads

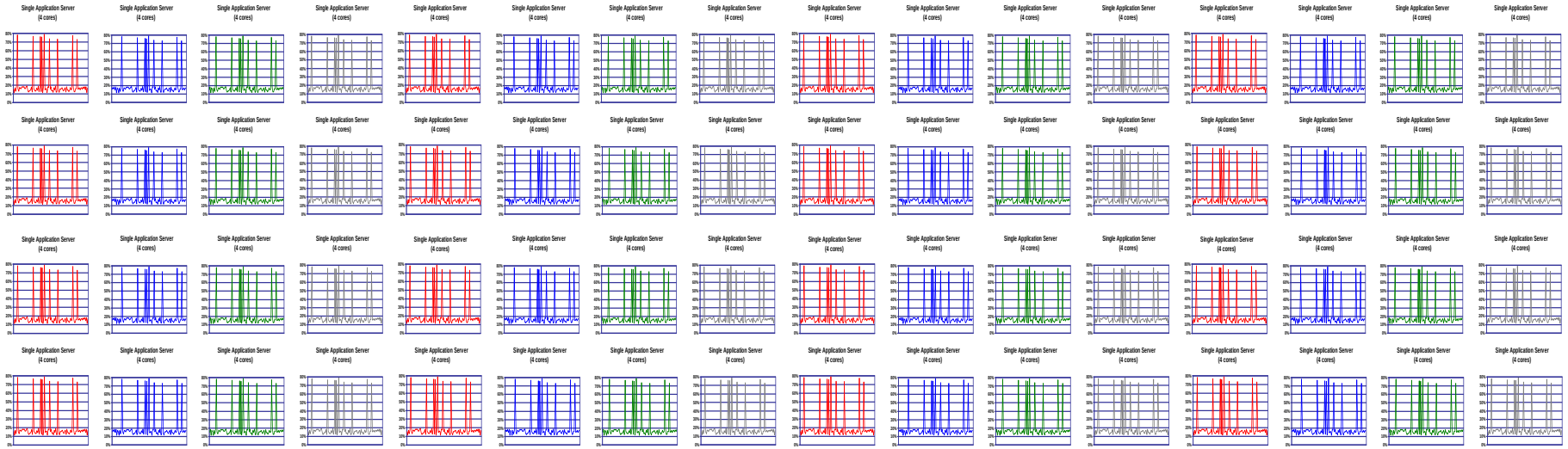
Very Large Scale Servers Consolidate Even More

- **64** separate workloads on **64** identical systems
  - Average utilization is 17%
  - Peak is 6 times the average
- **64** separate workloads on one system*
  - **Average utilization is 60%**
  - Peaks is 1.625 times the average

256 cores reduced to 72 cores (3.55 to 1)

64 to 1 Systems Consolidation (72 cores)





* Assumes independent workloads

Customer consolidation scenarios

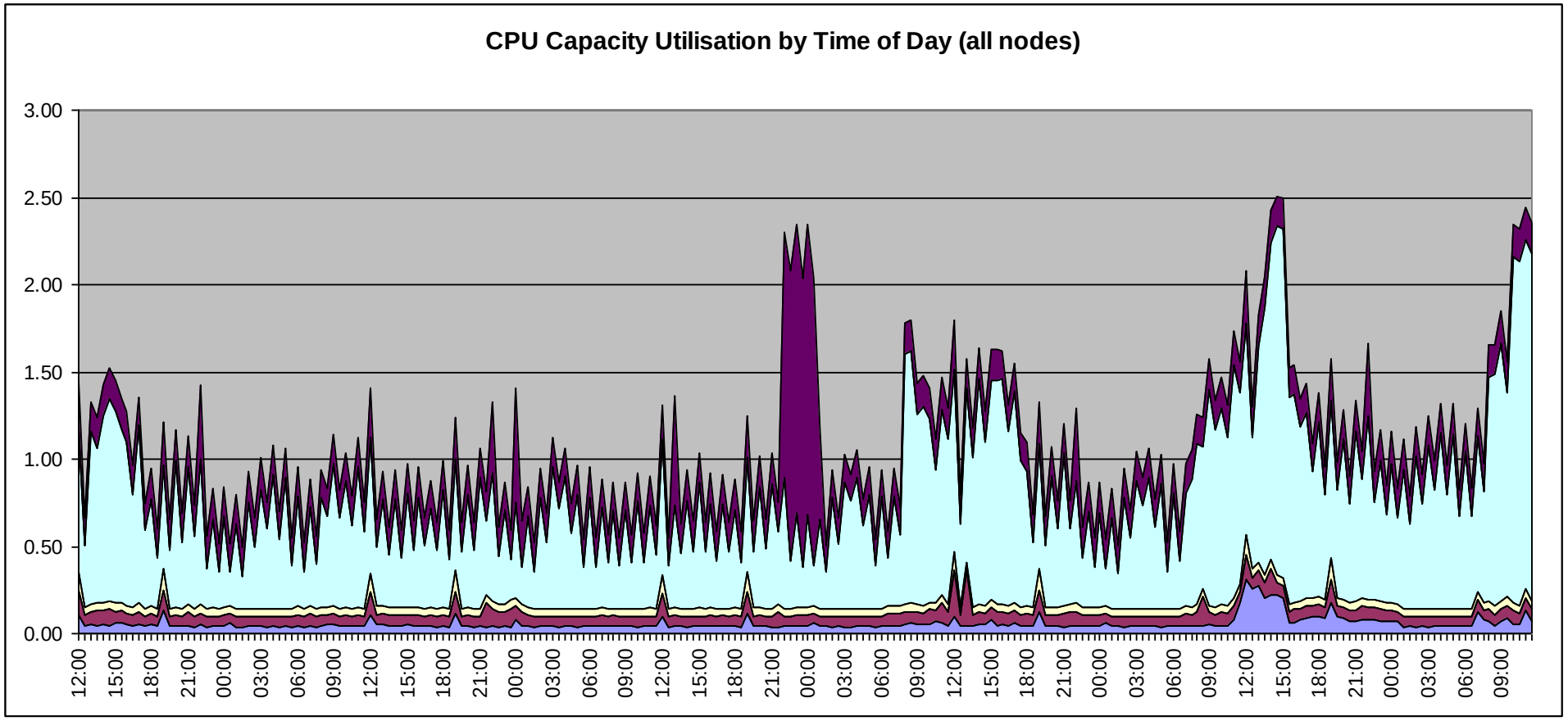
- Database servers and application servers on different physical systems, running different applications
- 3 consolidation scenarios
  - Only production databases
  - Production databases and application servers
  - Production and non-production environments

Scenario 1: Production databases

- 5 partitions, with the following capacities:

Partition	Capacity
1	1.5
2	1.6
3	0.4
4	1.7
5	2

Total capacity: 7.2 cores



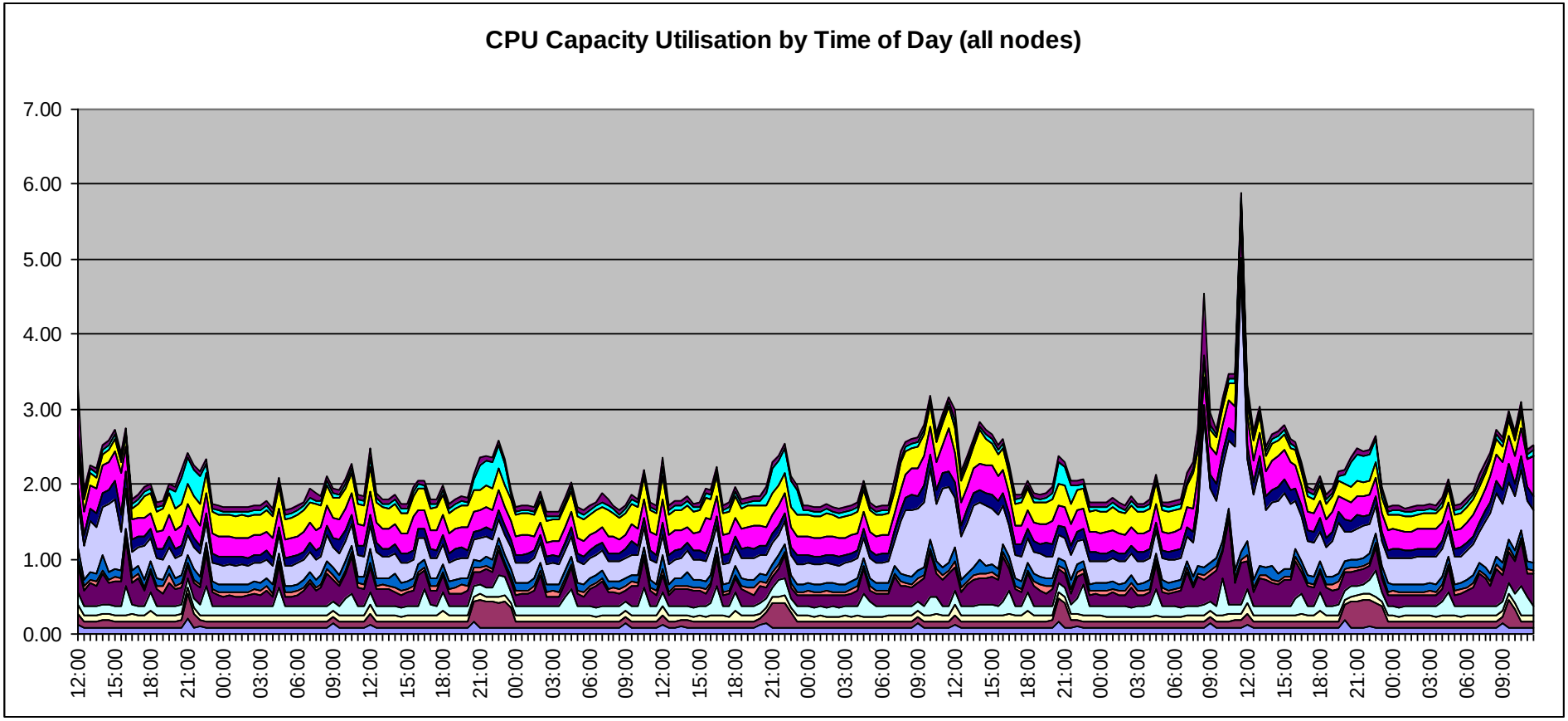
Max utilization: 2.5 cores – reduction of 2.88 to1

Scenario 2 – Production DB and App

- 5 partitions, with the following capacities:

Partition	Capacity
1	0.5
2	0.9
3	2.1
4	2.6
5	2.5
6	0.3
7	1
8	2
9	0.2
10	1.5
11	0.3
12	0.9
13	0.8

Total capacity: 15.6 cores



Max utilization: 5.9 cores – reduction of 2.64 to 1

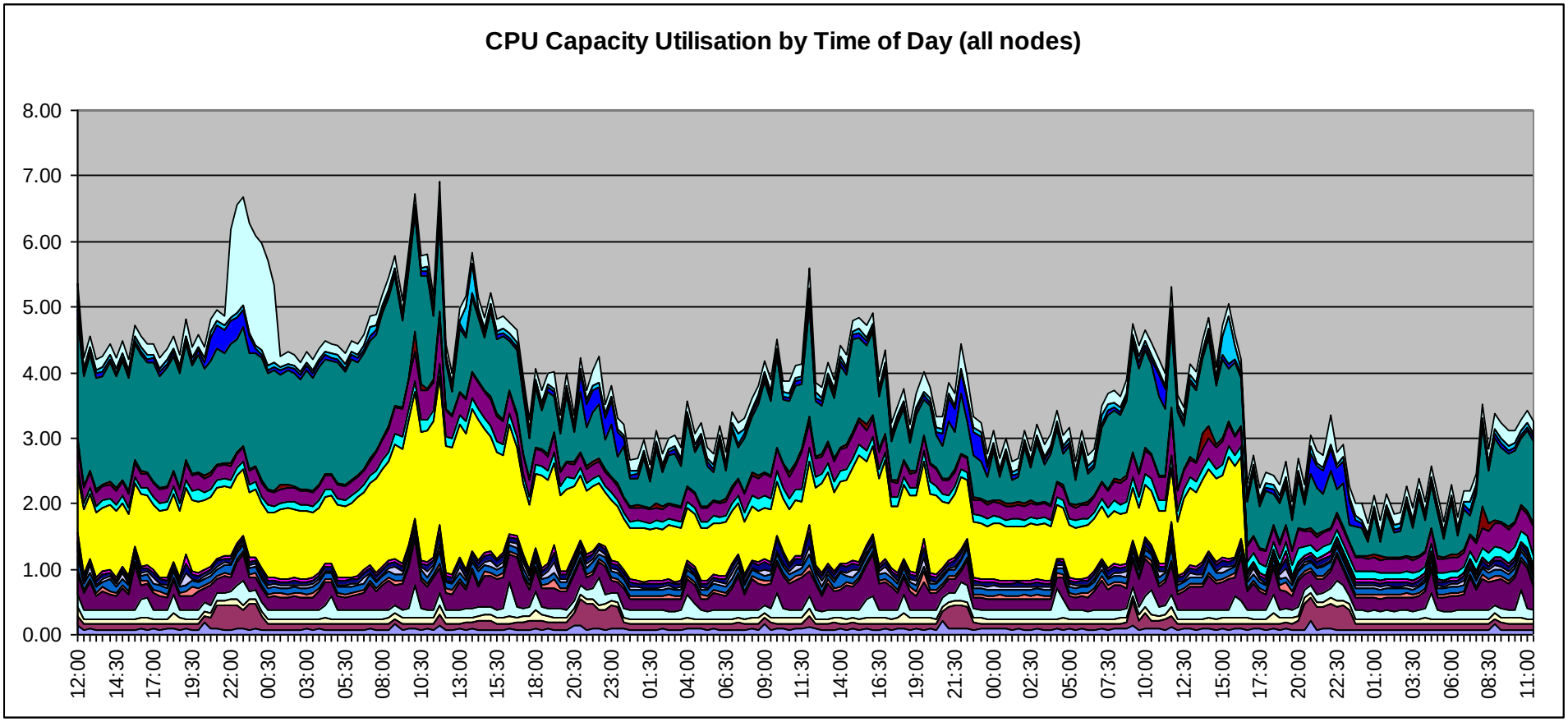
•If the large single peak is an abnormal situation (any activity not normally part of the processing) and is discarded, max utilization drops to 3.5 cores – reduction of 4.59 to 1

Scenario 3 – Production and non production

- 18 partitions, with the following capacities:

Partition	Capacity
1	0.5
2	0.9
3	2.1
4	2.6
5	2.5
6	0.3
7	1
8	1.5
9	1.6
10	0.4
11	2
12	0.2
13	1.5
14	0.3
15	1.7
16	0.9
17	0.8
18	2

Total capacity: 22.8 cores

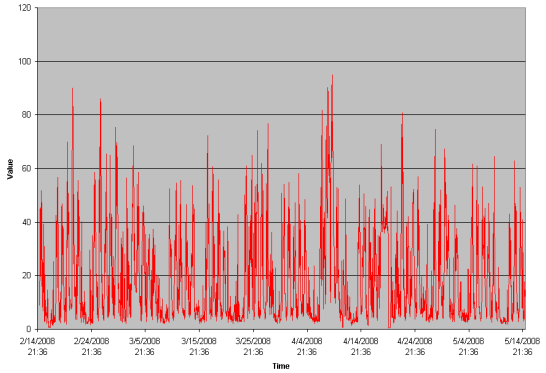


Max utilization: 6.9 cores – reduction of 3.3 to1

Customer X – SAP R/3

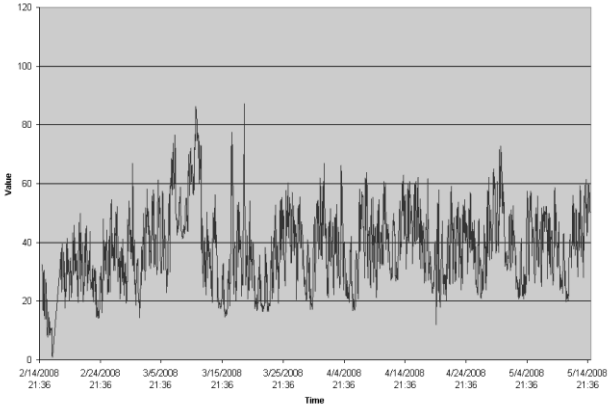
- Production R/3 environment running on POWER5 servers
- Dedicated partitions running on separate servers
- All partitions part of the same workload (i.e., peaks would happen on all partitions at the same time, since they are working together on the same transactions – Database and Application Server relationship)
  - This is not the optimal case for Server Consolidation, since peaks will overlap

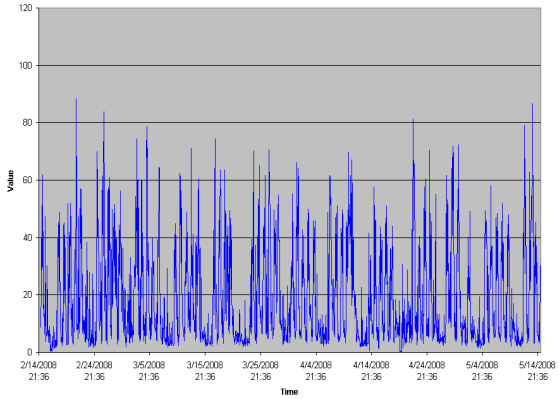
Customer X – R/3 production environment



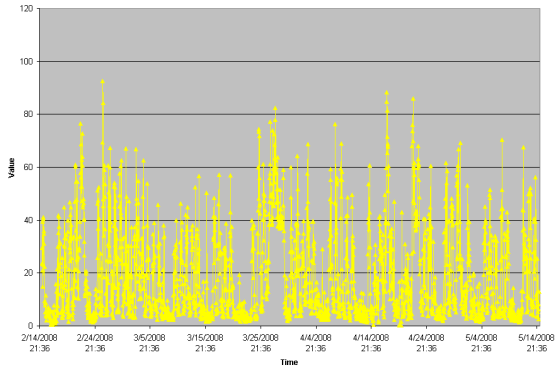
App Server 1
Average: 19%
Max: 95%
Ncores=8

DB Server
Average: 38%
Max: 87%
Ncores = 13





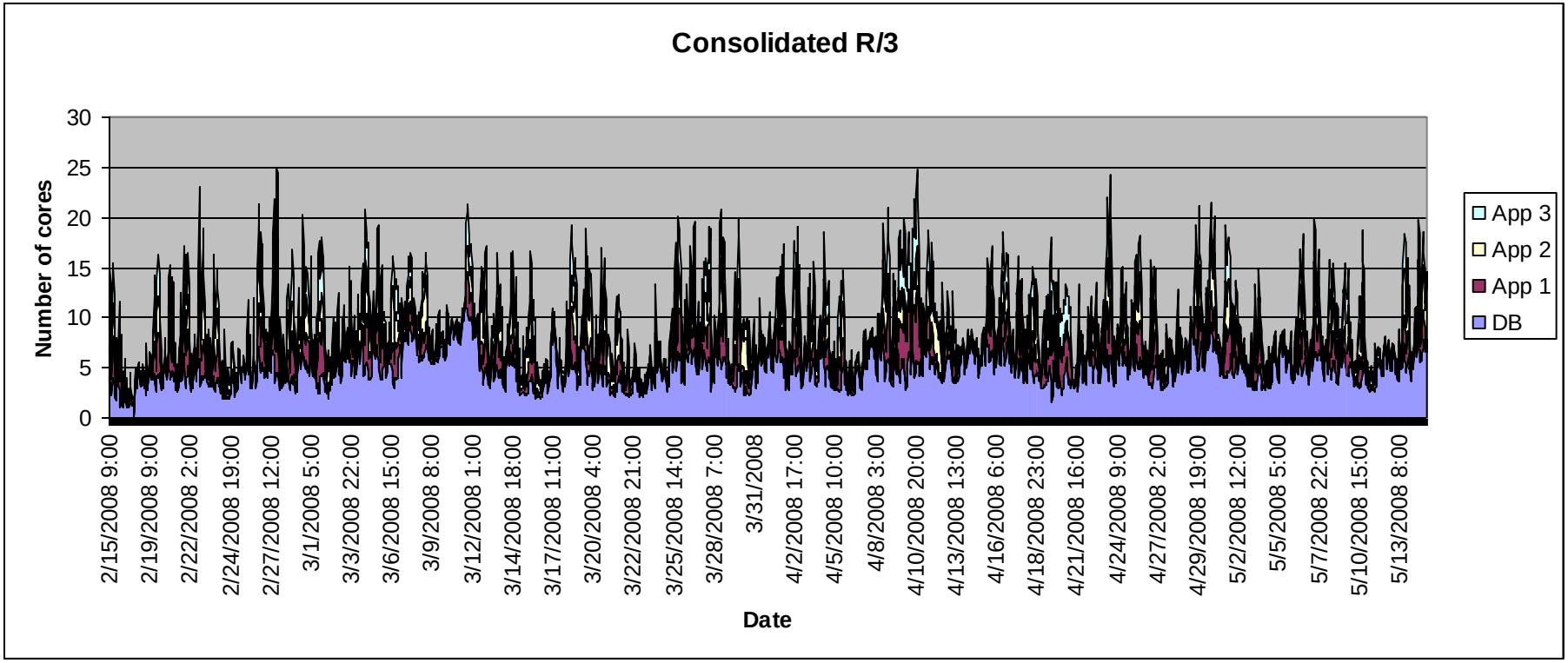
App Server 2
Average: 19%
Max: 87%
Ncores=8



App Server 3
Average: 19%
Max: 93%
Ncores=8

Average utilization for all systems together: 26.2%

Customer X – Production after consolidation



Instead of 37 cores, we need 25 cores (32.5% reduction)

* If we “smooth” the peaks, max number of cores is around 15 – reduction of about 3 to 1

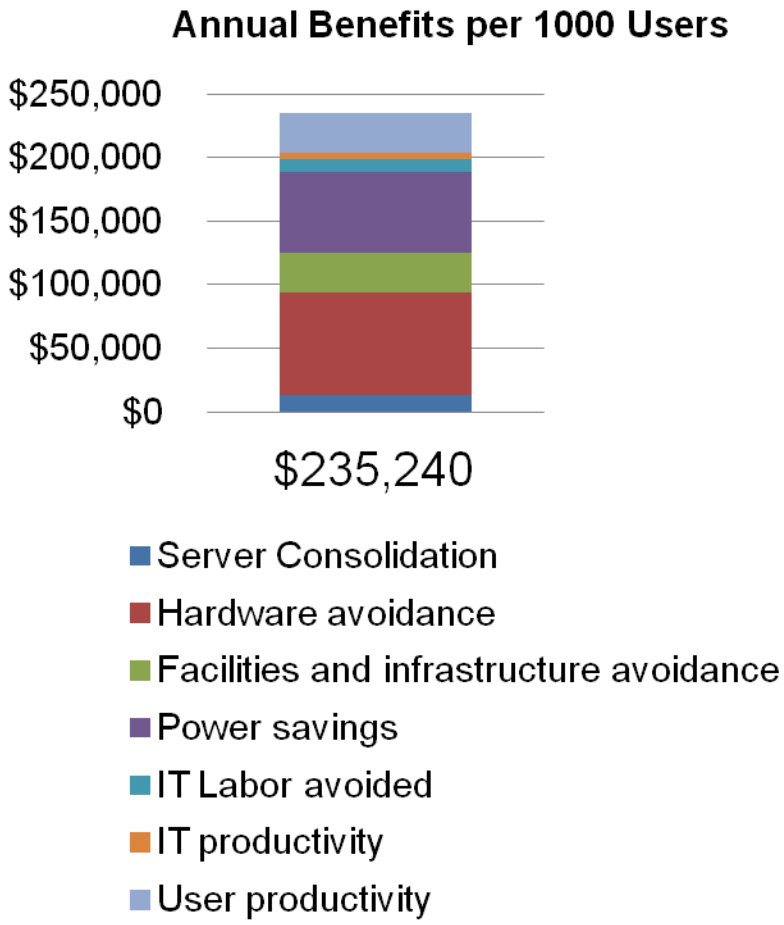
Average utilization is 9.5 cores (38% usage of 25 cores)

Max capacity is now 25 cores for all partitions – more room to grow

IDC Study Demonstrates the Business Value of Large-Scale Server Consolidation on POWER

Clearly illustrates that consolidation on Power Systems with PowerVM yields tremendous cost savings.

- Client results from consolidations onto enterprise Power Systems
- Up to 339% ROI, payback in as little as 7.8 months
- 20-40% increase in performance
- Up to 94% reduction in downtime



Source: The Business Value of Large-Scale Server Consolidation, IDC, August 2010; http://wwwibm.com/systems/power/idclargescale/

How to consolidate?

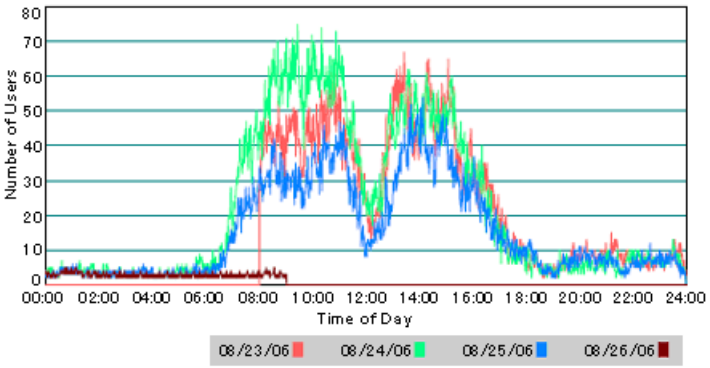
- Several tools and techniques
- Basic methodology is to gather information on resource utilization
  - Easier said than done…
- Depending on application type, server type, physical requirements (high availability, physical location) servers can be consolidated in different ways
- Some customers consolidate systems based on service type
  - One consolidation for databases, one for Java app servers, one for SAP, one for non-production, one for e-mail, etc
  - Although better than no consolidation at all, it reduces the efficiency because it prevents different workloads to share resources
    - Non-production environments are typically idle at night and weekends, therefore batch jobs from production could leverage the idle resources
- The most efficiente way to consolidate and utilize the resources is to have as many partitions as possible on a single server
  - Obviously considering adequate processing resources, and high availability requirements

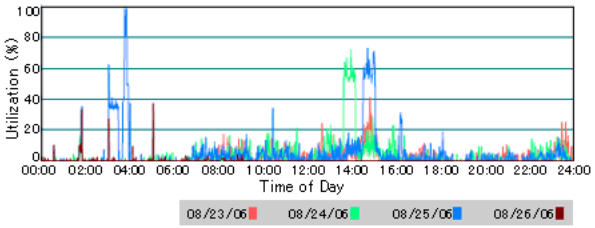
Data collection

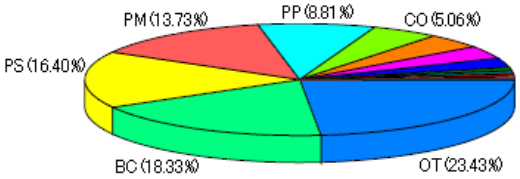
- Several commercial and free tools
  - Operating system tools
  - Application tools
  - Consolidation tools
- If the consolidation is about a specific application, try to use a tool that reports based on the application perspective
  - SAP: IBM Insight for SAP
  - Oracle: IBM Insight for databases
  - Tivoli Monitoring for Databases
- If there are multiple applications, or there is no specific reporting tool for the application, use the system utilization
  - Numerous ways to collect data, from OS tools to system monitors

IBM Insight for SAP

- http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS381
- Provides information on number of users, CPU and memory utilization, workload distribution per module and pe time of day
- Can be used as input data for an SAP sizing for replacement or upgrade
- Can be used as input data for VaSar – IBM virtualization and analysis tool for SAP

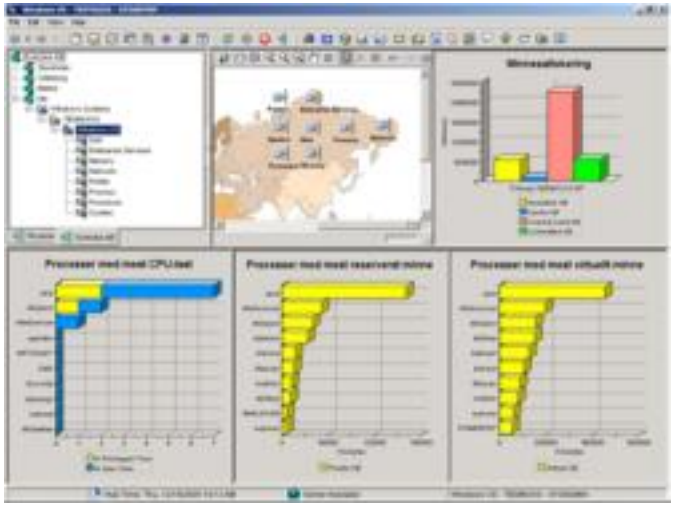


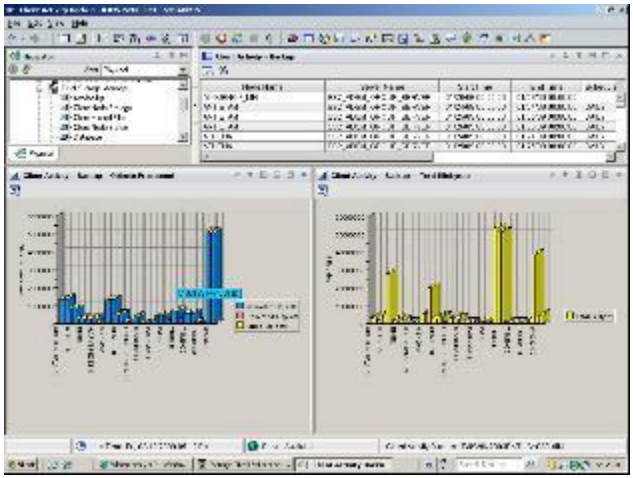




IBM Tivoli Monitoring

- http://www-01.ibm.com/software/tivoli/products/monitor/
- Provides system monitoring and utilization reporting, and also application monitoring (DB2 agent, Websphere agent, etc)
- Data stored in Tivoli Warehouse and used for capacity planning
- Can be used as input data for VaSar along with the Insight data



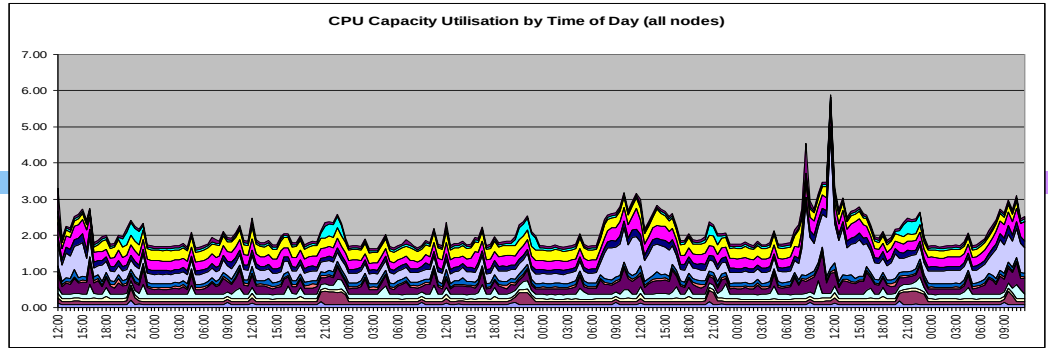


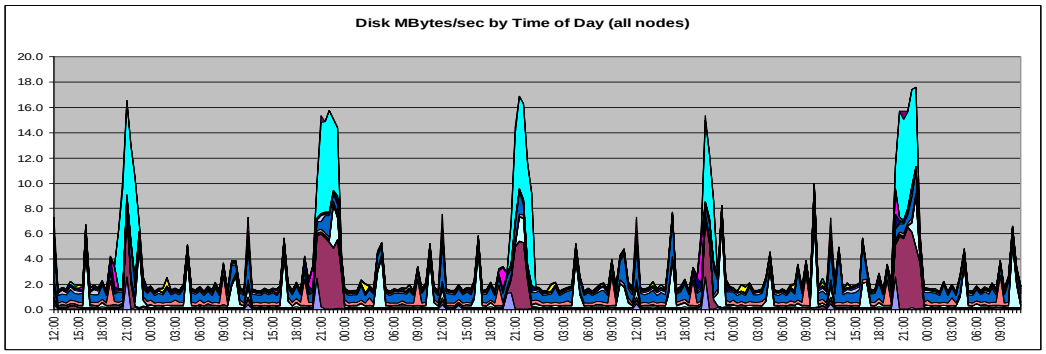
Nmon and nmon consolidator

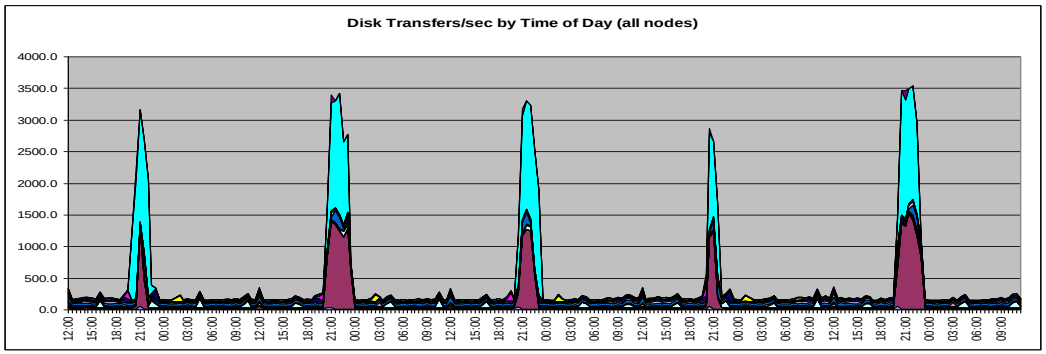
- Nmon is a freeware tool for AIX and Linux that provides realtime monitoring and performance data capture.
- It is now included as part of AIX
- Nmon standalone is used for real time analysis, and it is used with the Nmon Analysed and Nmon Consolidator for post-processing and capacity planning

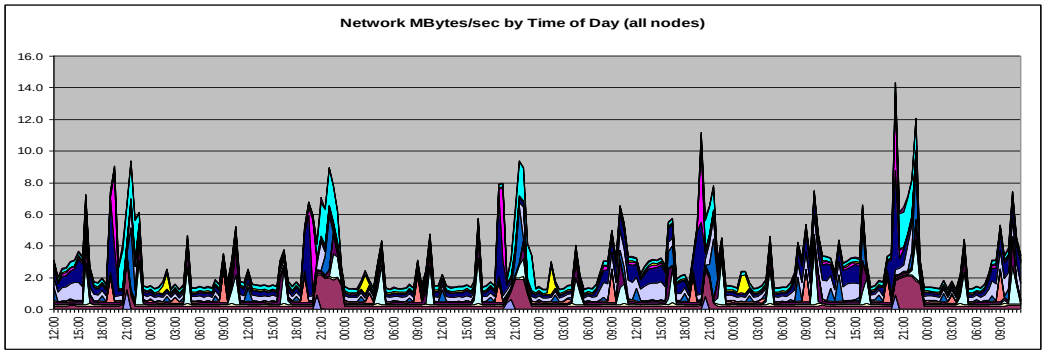
Nmon consolidator

- http://www.ibm.com/developerworks/wikis/display/wikiptype/nmonconsolidator
- Data is collected using nmon on each system/partition
- Nmon consolidator spreadsheet reads all the files and stacks the data
- By inspecting the charts, it is easy to obtain the required capacity for the consolidated server







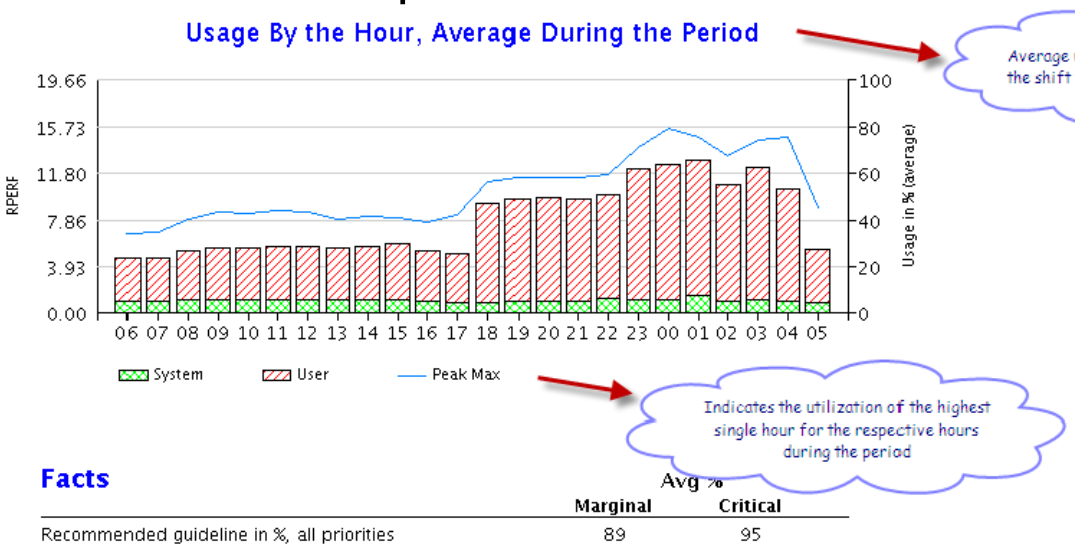


Topasrec/topasout (AIX commands)

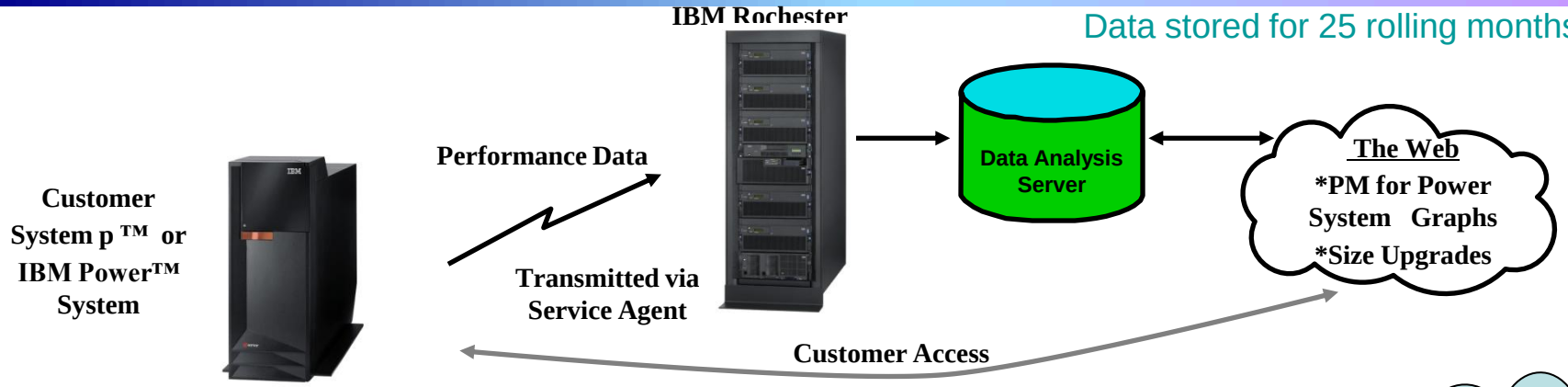
- The topasrec command collects a set of metrics from the AIX® partitions running on the same system.
- The topasrec command collects dedicated and shared partition data, and a set of aggregated values to provide an overview of the partition set on the same CEC.
- The topasout command is used to convert the binary recordings generated by thetopasrec utilitiy. It can also process nmon files.
- The output file from topasrec can be used as input data for the IBM Workload Estimator tool

Performance Management for Power Systems

- http://www-03.ibm.com/systems/power/support/perfmgmt/
- PM for Power Systems is an automated capacity planning and performance analysis report and graphs offering that is designed to help you plan for and manage the growth and performance of your system.
- Besides offering a report for system evaluation and capacity planning, it can also be used as input data for the IBM Workload Estimator tool.



IBM Performance Management for Power Systems – More Detail



- Automates Capacity Planning and performance analysis
- Integrated into AIX (version 5.3, 6.1 and 7.1 (recent TL's))
- Transfers Collected Performance Data to IBM via ESA, HMC options or IBM Systems Director (Service and Support Manager)
  - (customer retains the option to collect and transmit)
- Usage / growth information, reports & graphs via secure web access
- Option to size your next upgrade or replacement system via data integration with the IBM Systems Workload Estimator
- Customer password protected Web access to view/print/save graphs

PM for Power Systems helps enable
-integration
-simplification
-non disruptive growth

IBM Performance Management (PM) for Power Systems - - What are the options?

Two reporting offerings for customers to choose from
http://www.ibm.com/systems/power/support/perfmgmt

• No Charge service….. A no additional charge option that provides an Internet accessible one page summary graph showing key performance and growth data plus a projection of remaining growth of the system or partition

- System must be under IBM warranty or on an IBM hardware maintenance agreement to be eligible
- Available worldwide
- **Performance data can be merged with the Workload Estimator to size needed upgrades**

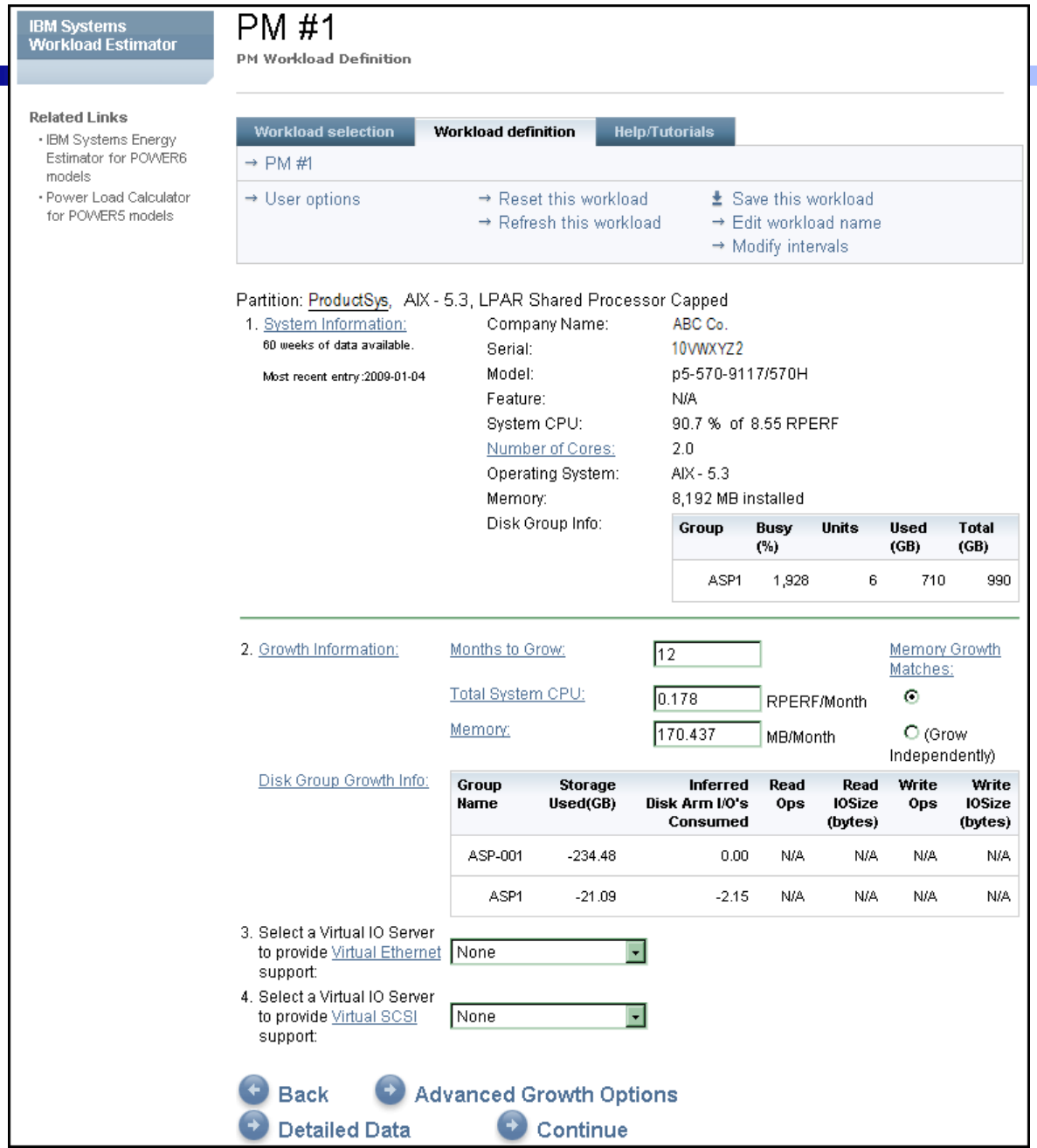
• Full function service….. An IBM Global Technology Services (billable) report set available via Internet access that provides multiple detail reports on an ongoing basis depicting the growth and performance of the system

- Available either as a stand alone fee offering or as part of an IBM premium service offering like the IBM Enhanced Technical Support offering
- Performance data can be merged with the IBM Workload Estimator to size needed upgrades
- Contact your IBM Representative to determine the availability, packaging, naming and pricing in your respective country

IBM Workload Estimator

- The IBM Systems Workload Estimator (WLE) is a web-based sizing tool for IBM Systems. Included are Power Systems, System x, and System z.
- WLE is available at http://www.ibm.com/systems/support/tools/estimator/. You can use this tool to size a new system, to size an upgrade to an existing system, or to size a consolidation of several systems. WLE will characterize your projected workload either with customer measurement data or by using one of the many workload plug-ins (a.k.a., sizing guides). Virtualization can be reflected in the sizing to yield a more robust solution, by using various types of partitioning and virtual I/O. WLE will provide current and growth recommendations for processor, memory, and disk (either internal or SAN) that satisfy the overall customer performance requirements.

WLE Definition Window



All of your IBM PM for Power System configuration and historical performance data will automatically be loaded for you into the WLE.

WLE Proposed Solution – Peak Weeks

The WLE recommends an appropriate upgrade with configuration suggestions at the total system and LPAR level

Remember: PM and the WLE can be used to size server consolidations, a capacity on demand processor, LPAR's, etc.

	The 750-8233-E8B can handle the defined workloads now.	It can also handle the specified growth.
rPerf:	86.99	86.99
Processor CPW:	47,800	47,800
Cores:	8	8
CPU Utilization:	34%	48%
5250 OLTP CPW:	47,800 (0% utilized)(1)	47,800 (0% utilized)(1)
Operating System:		
Software Pricing Tier:		
Memory (MB):	16,384 of 131,072	22,014 of 131,072
Int. Disk Drives (arms):	56 of 584	70 of 584
Capacity (GB):	5,987 of 261,000	6,531 of 261,000
Offering Family	IBM Power Systems	IBM Power Systems
Processor:	IBM® POWER7	IBM® POWER7

Random Illustration

The user may alter growth rates and time horizon

IBM Workload Estimator

- WLE calculates requirements based on utilization, annual growth (if specified), virtualization requirements, VIOS sizing, I/O requirements for disk sizing and memory requirements.
- It is available for Power Systems (i, AIX and Linux), System x and Mainframe.
- It **does not** allow for heterogeneous consolidation
  - You can not have Power Systems and x86 on the same study.
- It does an excellent job for existing Power Systems environments.

IBM ATS SCON Monitor

- The ***IBM ATS Server Consolidation Monitor*** *(ASCM)* is a FREE service offering, providing a report detailing the workload and utilization of production systems
- ASCM is a set of two Microsoft Windows 64-bit applications, that run on any version of Windows running the .NET 4.0 framework:
  - ASCM Collector: Runs on a dedicated PC. Captures system architecture information and **utilization statistics from Windows and UNIX targets**.
    - Recommend 1 core and 1GB RAM per 100 systems observed
  - ASCM Reporter: Generates the ASCM report and csv files (run by Techline specialist)
- Demonstrates current workload and resource consumption for server consolidation studies

38

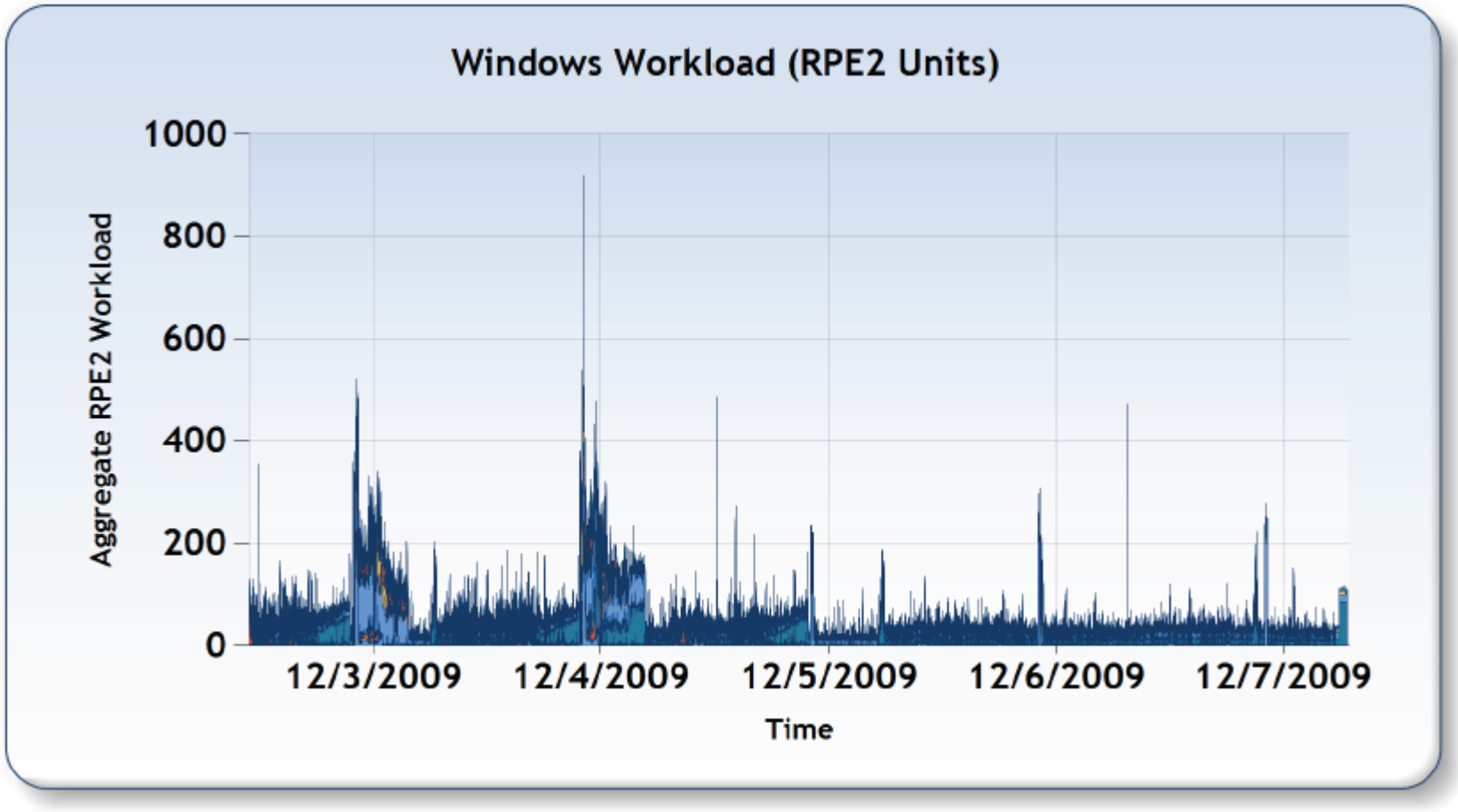
Supported Collection Methods

Method	Windows	AIX	Linux	HP-UX	Solaris	VMWare
WMI (Windows Management Instrumentation)	Yes	No	No	No	No	No
Ganglia	No	Yes	Yes	Yes	Yes	No
VMWare ESXi (bare metal hypervisor)	No	No	No	No	No	Yes
NMON	No	Yes	Yes	No	No	No

- Data collector must have TCP/IP access to all hosts being monitored

39

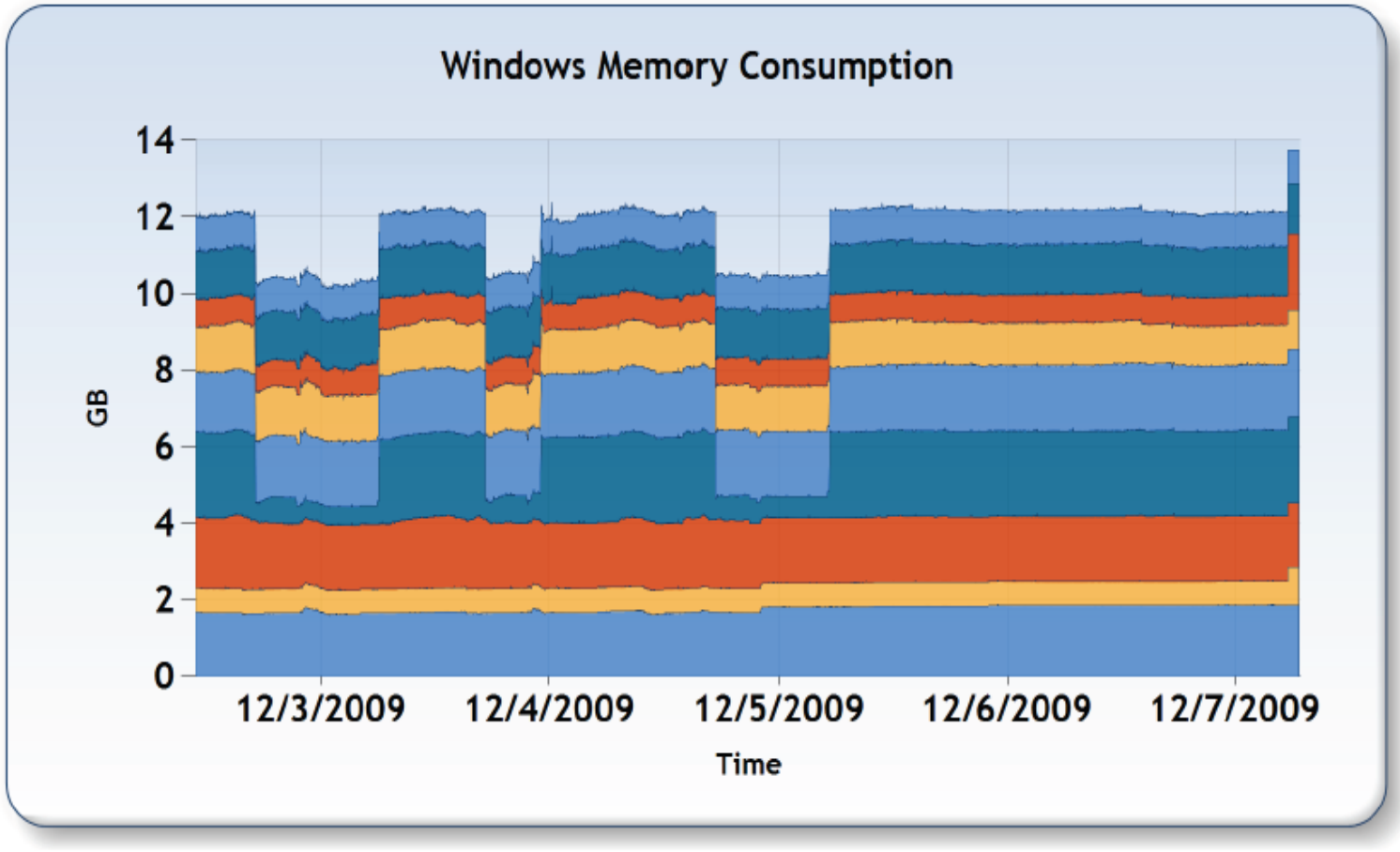
Consolidated CPU Utilization Normalized to RPE2



RPE2 Average: 64

40

Consolidated Statistics – Memory



Average: 11.68 GB

41

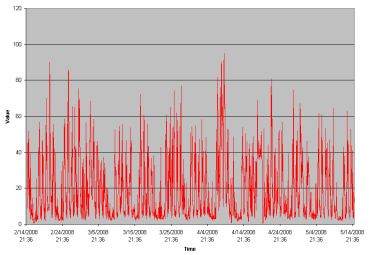
Considerations on the ATS SCON Monitor

- Unlike the WLE, it does not recommend a server to consolidate the workloads
  - It shows the capacity requirements for the server. With this information, a specialist can then select the appropriate server for consolidation, taking the existing data as well as additional requirements (such as growth)
- It allows heterogeneous data collection – Windows, Linux, UNIX – effectively enabling data center consolidation opportunities
- And it is FREE! Talk to your IBM Technical specialist, or your business partner!

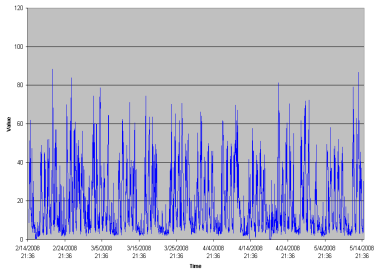
What if you don’t have data?

- Sometimes performance data for the servers to be consolidated is not available
  If performance data is not available…
  - Not good! Aren’t you monitoring your servers?? ☺
  - Different applications, operating systems, physical location, etc may create difficulties to collect data in a consistent way
- When utilization cannot be measured, the altenative is to estimate the actual usage
  - Some customers have no idea on usage, and just assume 100% utilization. It definitely works, but reduces the benefit of virtualization
  - Some customers have information on average utilization – while it is a measure of usage, it does not take peaks into consideration – this may lead to undersized systems if peaks happen at the same time.

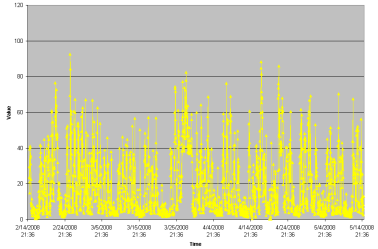
Back to Customer X – R/3 production environment (concurrent peaks)



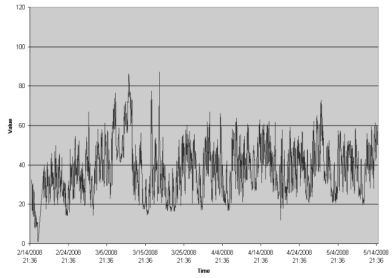
App Server 1
Average: 19%
Max: 95%
Ncores=8



App Server 2
Average: 19%
Max: 87%
Ncores=8



App Server 3
Average: 19%
Max: 93%
Ncores=8



DB Server
Average: 38%
Max: 87%
Ncores = 13

Average utilization for all systems together: 26.2%

- If we have no performance data, we would have to assume all cores in the systems – 37 cores.
- If we use average, we get 26.2% of the cores – 9.7 cores.
- From the performance data, we know that we need at least 15 cores
- Bottom line: having the performance data is **always** more accurate
  - If peaks are **not** at the same time, average will be closer to real utilization. If you know your environment, you can tell whether average is good enough for the consolidation.

IBM Factories Get You Started on the Road to Consolidation

- Free Proof of Concept and cost/benefit analysis
- Includes high level architecture
- Tools and techniques to provide the best consolidation solution

Our teams conduct data center interviews and run analysis tools to assess current efficiency and make consolidation recommendations.

- **Server Consolidation Factory**
- **x86 Server Consolidation Factory on Power Systems**
- **Availability Factory**
- **Migration Factory**

http://www-03.ibm.com/systems/migratetoibm/factory/

IBM

Technical Forum & Executive Briefing

17 al 21 Octubre 2011

Imagine **PODER** Imagine **CAPACIDAD**

Gracias!

cmaciel@us.ibm.com