

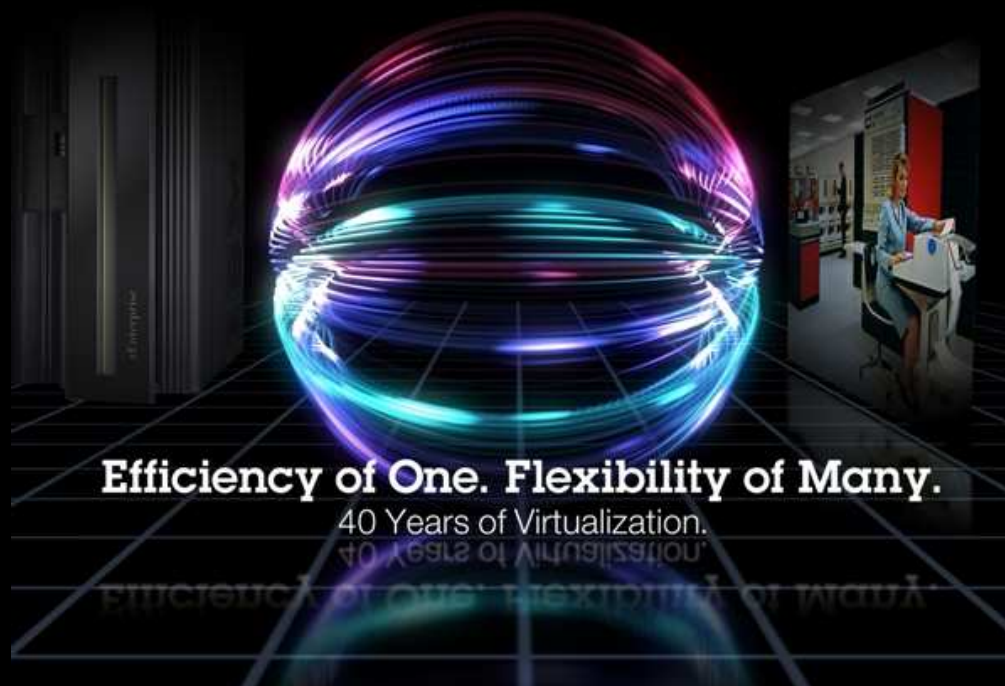


Efficiency of one. Flexibility of Many. 40 years of virtualization.



z/VM Capacity Planning Overview

SHARE 120 – San Francisco – Session 12476



Bill Bitner
z/VM Customer Care and Focus
bitnerb@us.ibm.com



Efficiency of one. Flexibility of Many. 40 years of virtualization.



Introduction

- Objectives:
 - Provoke thought and ultimately action about Capacity Planning
 - Concepts and approaches will be covered, not so much the mechanics
- Companion Piece:
 - z/VM Performance Metrics – also available from the author
- Time permitting – dialogue on what can IBM do to help in this space?



Companion Piece – z/VM Performance Metrics

- Lists the top 50+ metrics that we find useful, along with descriptions on them
- Where appropriate, includes which:
 - monitor record contains the information
 - Performance Toolkit report displays the information
 - OMEGAMON XE workspace for managing the information
- Rules of Thumb given in some cases

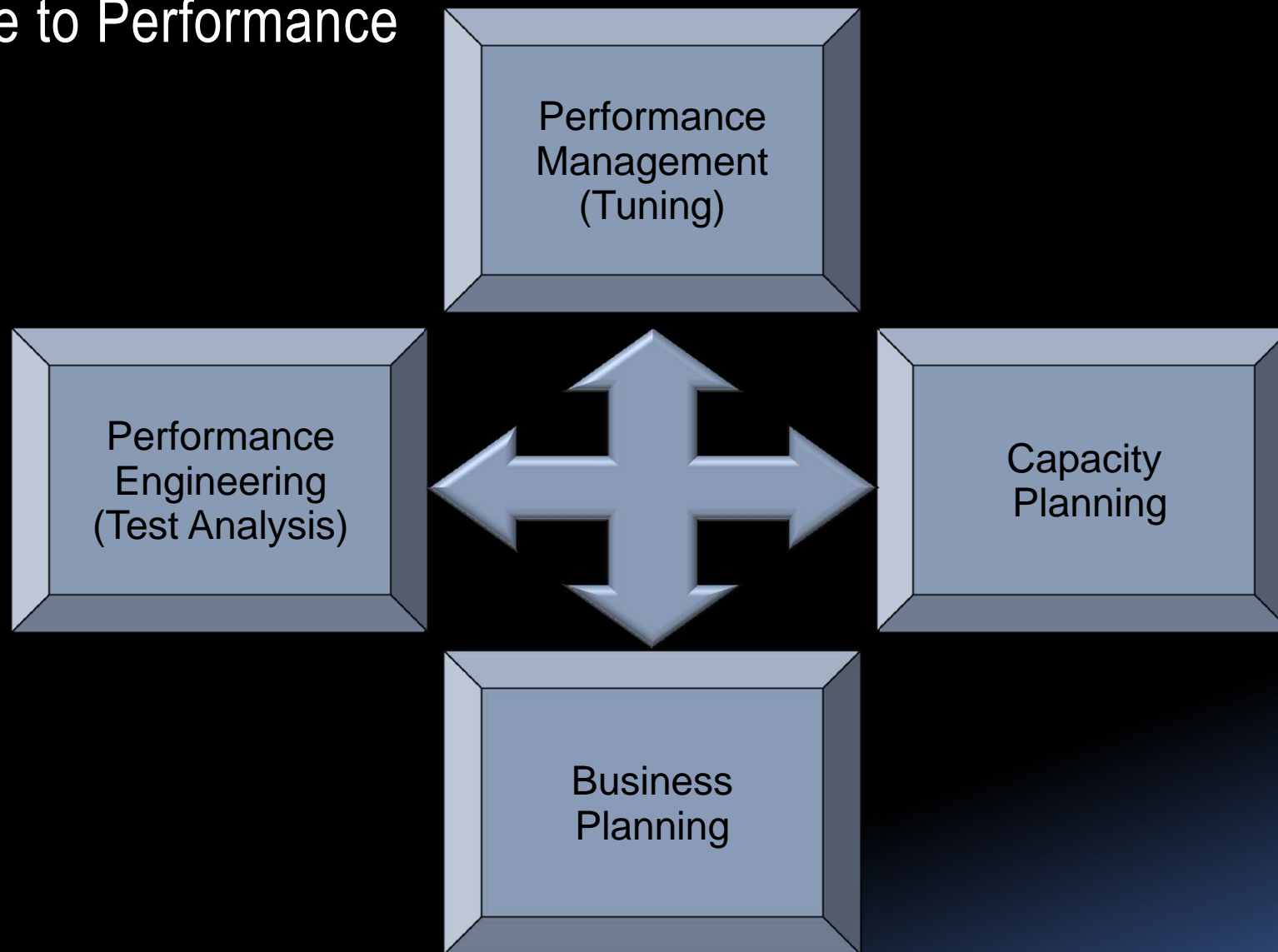
- **Total Processor Utilization:** (Monitor: D0/R2; Toolkit: FCX100 CPU; OMEGAMON: System workspace under headings of Percent CPU). This is the processor utilization from the VM perspective and includes CP, VM System, and Virtual CPU time. It is often beneficial to break this down into the three components:
 - **System Time:** This is the processor time used by the VM control program for system functions that are not directly related to anyone virtual machine. This should be less than 10% of the total processor time for the z/VM LPAR.
 - **CP Processor Time:** This is the processor time used by the VM control program in support of individual virtual machines.
 - **Virtual Processor Time: (Emulation Time):** This is processor time consumed by the virtual machine and the applications within it.



Efficiency of one. Flexibility of Many. 40 years of virtualization.



More to Performance





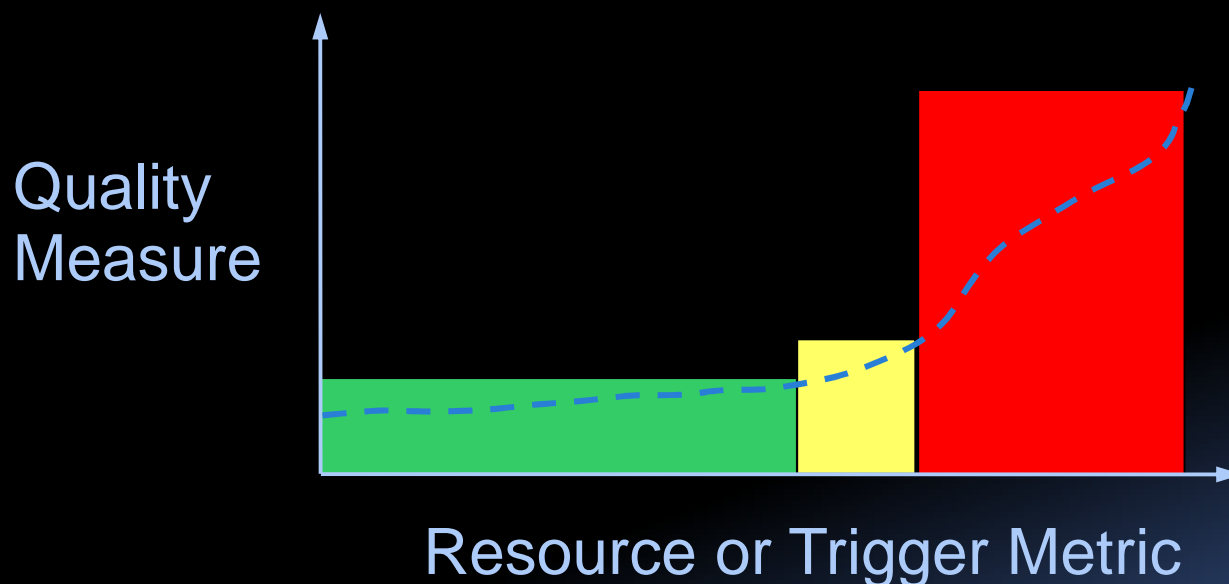
Real vs. Virtual Planning

- Capacity planning for 'real' resources is one thing, but how do we incorporate 'virtual' resources?
- Need to address “Overhead” or management costs
 - Define 'overhead'
 - z/VM Control Program processor time?
 - System Management virtual machines?
- Virtual \neq Free
- Need to address peaks
 - Averages alone are not sufficient
- Define what is acceptable – “capacity lag” or impact to SLA
 - Acceptable overcommitment of resources is very dependent on:
 - Workload
 - Environment
 - SLA



Looking at Resources

- Utilization Metrics: metrics of interest to determine utilization & distribution of resources
- Indicator Metrics: metrics that relate to thresholds or the degree of constraint and pain the system is expressing
- Quality Measures: something that indicates workload and response time and whatever is important to the business
 - Well defined and defined throughout all the disciplines
 - Something that can be mapped to other metrics to indicate a sweet spot





Efficiency of one. Flexibility of Many. 40 years of virtualization.



Real Processor Resources

- Real Processor Resources are perhaps easiest to measure and manage
- Utilization Metrics:
 - LPAR Overhead Time
 - System CPU Time
 - CP CPU time associated with virtual machines
 - Virtual CPU time associated with virtual machines
- Indicator Metrics:
 - System Spin Time (Wall clock, not processor measure)
 - LPAR Suspend Time
 - CPU Wait
- Need to handle Specialty Engines
 - Measure each type
 - Mixed speeds?
 - Changing speeds?
- Keep in mind processor resource limits can pop up elsewhere
 - e.g. Both ends of a HiperSockets connection require processor
- Compare or prorate based on workload (transaction rates).



Efficiency of one. Flexibility of Many. 40 years of virtualization.



Virtual Processor Resources

- Utilization Metrics:
 - CP CPU Time
 - Virtual CPU Time
 - Total CPU Time
- Potentially also include:
 - Processes within Linux
 - Linux Steal time
 - At very least have the above available from Performance Engineering for comparison if z/VM totals look abnormal.
- Again, make accommodations for Specialty Engines
 - Real and Virtual
- Indicator Metrics:
 - CPU Wait
 - Diagnose x'44'
 - Diagnose x'9C'



Linux View: %Steal

- Current Linux distributions (RHEL 5 & SLES 10) report %Steal as well as pct User and pct System
- %Steal: Linux view of percent of time that it had work to run but was unable to run.
 - z/VM was dispatching other virtual machines, compare to %CPU Wait in z/VM state sampling.
 - z/VM was executing on behalf of the Linux virtual processor, compare to CP CPU usage of the Linux virtual machine
 - Linux yielded its time slice to z/VM via diagnose 0x9C instead of spinning on a formal spin lock. Examine diagnose rates.
 - The z/VM partition was unable to run due to another logical partition being dispatched at the LPAR level.



Other Processor Planning Thoughts

- Need to have some measure of work or throughput
- Best to determine cost per <something meaningful to everyone>
 - Performance Engineering
 - Business Planning
 - Performance Management
 - Capacity Planning
- Establish in Performance Engineering testing what the target cost / transaction
- Bring in Business Planning to determine target or range of transaction load
- Capacity Planning projects requirements based on above two
- Performance Management folks can help identify problems when things do not track.
- This is a continuous process
- I prefer computing CPU seconds, but if you want to convert to some “MIPS” number or “IFLS” or “Computing Units” feel free. Just make sure everyone uses the same conversion numbers.



Real Memory

- Utilization Metrics:
 - NonPageable
 - Pageable
 - Minidisk Cache
 - Misc
- Don't forget Expanded Storage
- Indicator Metrics:
 - Emergency Scan on Demand Scan
 - Emergency Scan failures
 - Available List(s) going empty



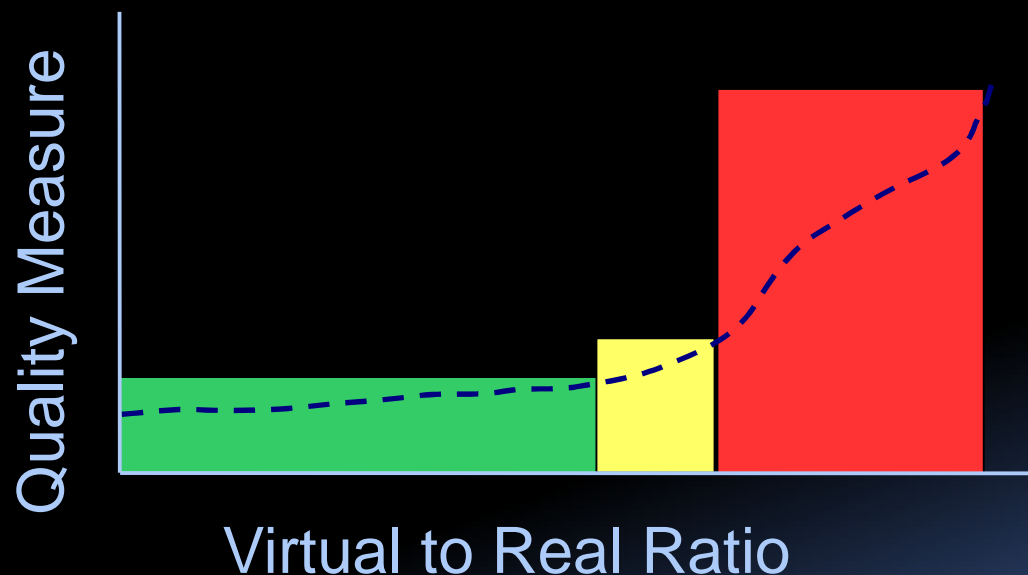
Virtual Memory

- Types of Virtual Memory:
 - Virtual Machines
 - NSS/DCSS
 - Virtual Disks in Storage
 - PTRM and other System Utility Spaces
- Virtual Machine Utilization Metrics:
 - Defined virtual memory
 - Backed virtual memory
 - Resident virtual memory
 - Estimated WSS
- Indicator Metrics:
 - Paging Rates (Reads and Writes)
 - Loading User
 - Page Wait (Asynchronous and Synchronous)
- A guest pages may exist on both DASD and Real Memory
- Private DCSSs are considered part of the virtual machine for most metrics, while Shared DCSSs have their own metrics.



Memory Overcommitment

- Gather data to determine a curve such as below
 - Performance Engineering
 - Tracking Production
 - Artificially limiting amount of real memory
- Result is a Virtual to Real ratio for your workload that is edge of green/yellow.
 - For example, lets say it is 1.8. If you are going to increase workload by adding 30GB of virtual, then you need to add real memory to keep the ratio at 1.8 or lower.





Other Thoughts on Memory Planning

- What is the right 'over commitment' number? It depends.
 - Definition of virtual and real in the question.
 - Constraints of SLAs
 - All transactions sub-second vs. 99% of transactions sub-second
 - How much can performance features improve things
 - Workloads that are sized poorly at start have more room for improvement
 - Paging configuration
- See <http://www.ibm.com/vm/perf/tips/memory.html> for additional information



Network

- Session of its own
- Real Level
- Virtual Level
- Link Aggregation
 - Aggregates total sessions across multiple OSD chpids
 - Does not spread load of a single TCP/IP application session across those chpids.
- Beware of measuring a single session as multiple sessions is where the bandwidth value is.
- Limits/Thresholds/Quality Measures



Real I/O

- Utilization Metrics:
 - Channel Utilization
 - Device I/O Rates
 - System I/Os
 - User Driven I/Os
 - Device Utilization
 - Access Density (I/Os per GB space)
- Indicator Metrics:
 - Device Queuing
 - Error Rates
 - IOP Statistics



Virtual I/O

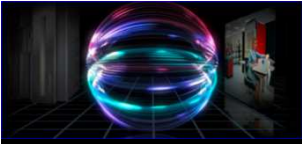
- Utilization Metrics
 - Virtual I/O per Guest
 - I/Os avoided due to MDC or VDisk
- Indicator Metrics
 - Various levels in software stack where I/O queuing can occur
 - Virtual I/O to Virtual CPU Ratio
- Caution:
 - %IOA (Asynchronous I/O Wait) in Performance Toolkit and similar field in Linux includes time of I/O processing.
 - I/O is relatively slow
 - %IOA may only show up when there is CPU activity or wait on CPU, so a high %IOA isn't necessarily bad.



SSI: Capacity Planning

- Great flexibility in managing multiple LPARs
 - Previously, if you split work across LPARs and had an imbalance, it was more difficult to rebalance
 - With SSI, virtual machines can run anywhere in the cluster without a lot of additional work
- Greater responsibility in planning, at two levels
 - Individual members
 - Need to ensure sufficient capacity and resources for the workload on each member
 - Track growth in requirements to limits of the member
 - Cluster-wide
 - Track growth in requirements of overall cluster to the limits of that cluster
 - Need to ensure sufficient white space for planned outages where LGR will be used to move workload out of a given member.

The “Getting Started With Linux” book has been updated with SSI and LGR planning tips.



SSI & LGR: Planning Relocations

- Need white space for planned outages where you move work off of a given member.
- How will work move off the member?
 - Use existing HA solutions to redirect work to existing servers on other members or elsewhere in enterprise.
 - Use LGR to move to another member.
 - Log off and then logon to another member.
 - Shutdown non-critical virtual machine for duration of the planned outage.
- To where do you move the virtual machines?
 - To a single member or multiple members?
 - To a member on same CEC or another CEC?
 - To a member held in reserve (such as a DR LPAR)?
 - It's not just one z/VM image anymore



Other Considerations for Planning

- “The bucket gets heavier as you add water.”
 - Destination system may become more constrained as you continue to relocate virtual machines to it.
- “Get the big rocks in first.”
 - In general, it is better to move the virtual machines generating the greatest memory load first.
 - Larger virtual machines
 - Virtual machines with higher page change rate



SSI & LGR: Planning White Space

- CPU
 - Shared logical processors?
 - Adjust LPAR weight settings?
 - Vary on additional engines?
- I/O
 - Ensure sufficient resources at all levels:
 - Channel, switch, control unit, device
 - Shared channels?
- Memory white space is not as easy to manage
 - Ensure sufficient paging space and concurrency or data rate capability
 - Increase real memory over commitment?
 - Temporarily decrease size of some virtual machines?
 - Use Dynamic Memory Upgrade?
 - No downgrade available



Data Collection Considerations

- Keep all groups in mind and in agreement
 - The value of your data increases when it can be combined with other data.
- Volume of data
- Retention time
- Granularity or interval of data
- Correlation with other data
- Time zone considerations
- Terminology



Efficiency of one. Flexibility of Many. 40 years of virtualization.



Methods of data collection

- Do It Yourself
- Performance Toolkit Summary/Trend/Histlog
- IBM Tivoli OMEGAMON XE on z/VM and Linux
- Shipping to z/OS



Efficiency of one. Flexibility of Many. 40 years of virtualization.



IBM System z



z10 Capacity Planning in a nutshell



Don't use "single-number tables" for capacity comparisons!

Use zPCR and/or zCP3000 to model before and after configurations
Work with IBM technical support for capacity planning!
Customers can now use zPCR

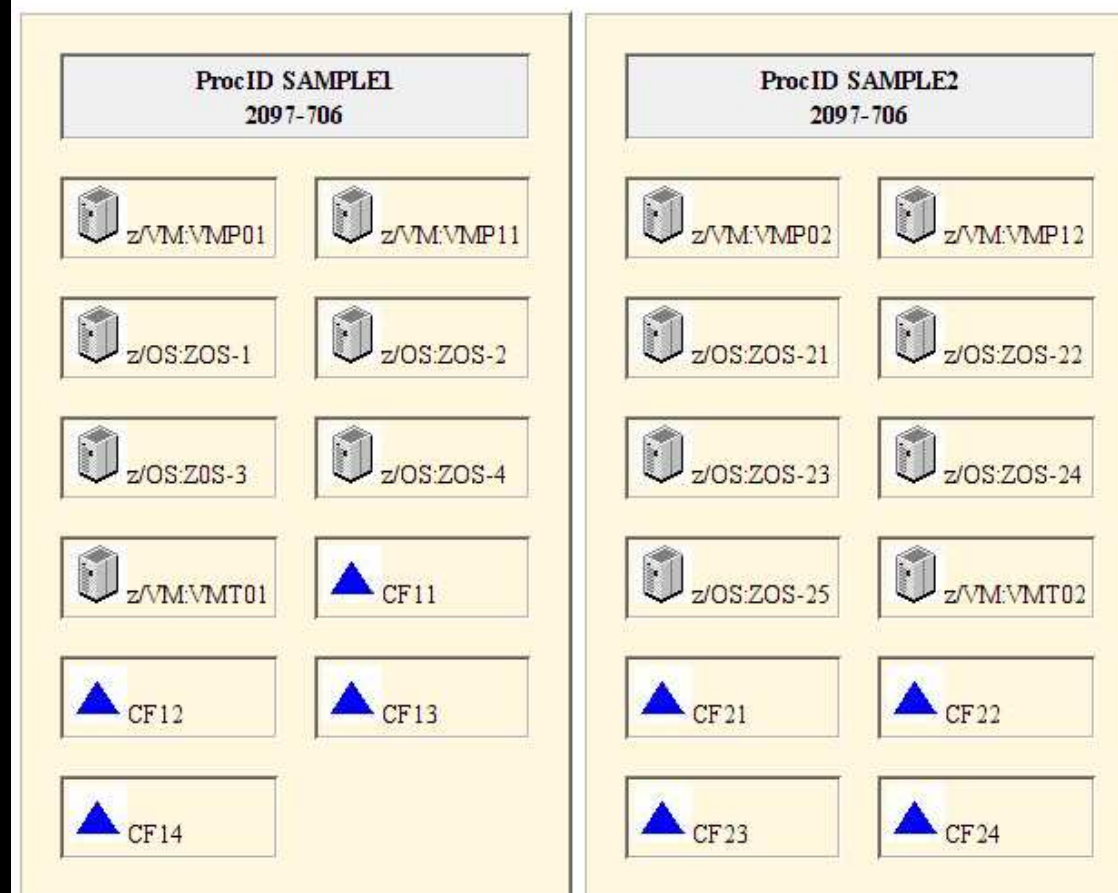


Efficiency of one. Flexibility of Many. 40 years of virtualization.



IBM Techline: Complete topology from z/VM data

Processor View for Generic Customer

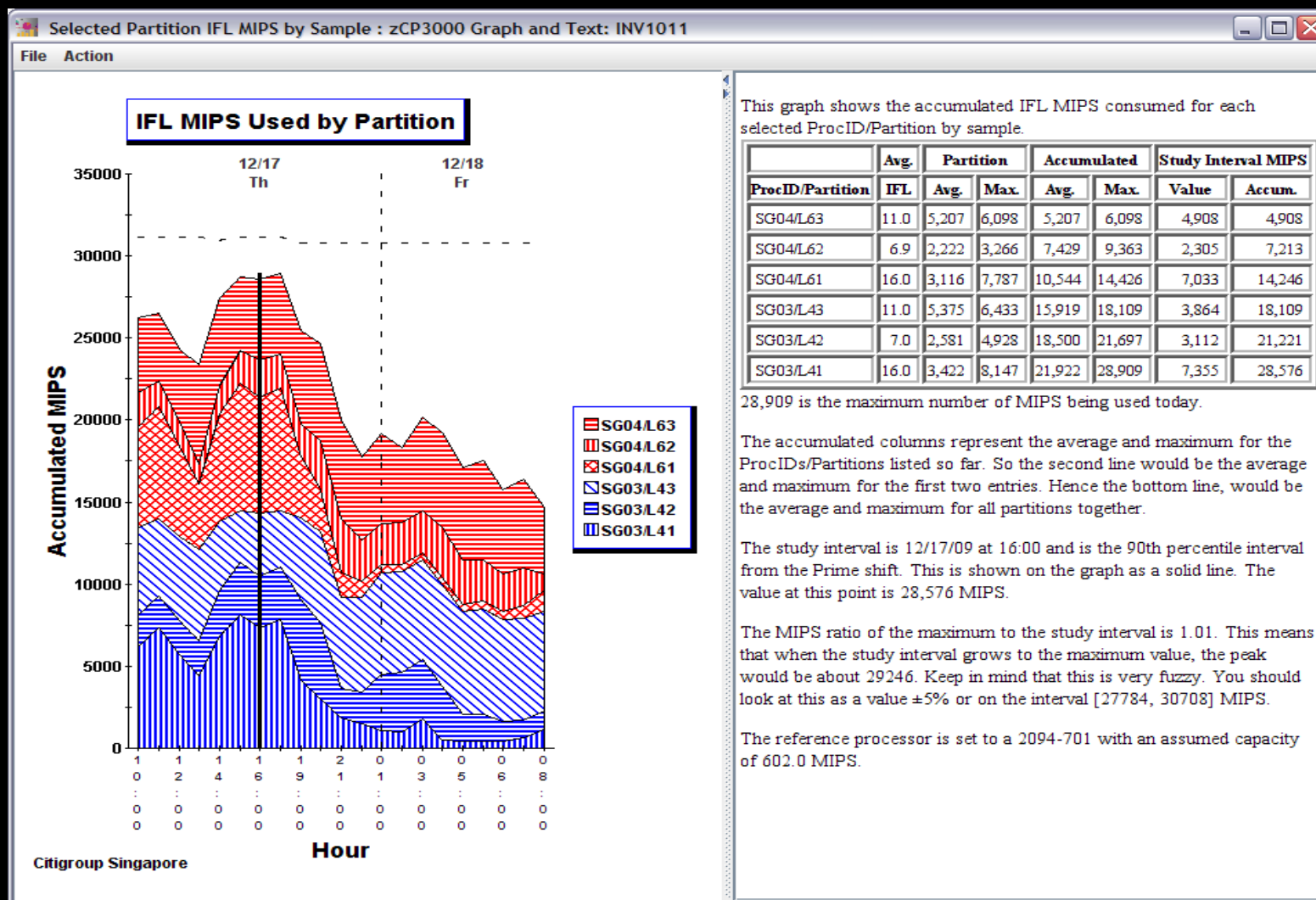




Efficiency of one. Flexibility of Many. 40 years of virtualization.



IBM Techline: Example of CPU Analysis





Efficiency of one. Flexibility of Many. 40 years of virtualization.



IBM Techline: Memory Summary Example

File
Analysis

System
Workloads
Memory
Performance

Express Memory Sizes As:
☒ MB
☐ Pages

Description	Virtual Memory	CMMA Active	WSS Intv	WSS Min	WSS Max	Memory Used Intv	Memory Used Min	Memory Used Max
LXPS3061	1,707	0	1,491	409	1,707	1,259	346	1,504
LXPS3095	4,608	0	4,096	3,772	4,291	3,406	3,145	3,581
LXPS3093	11,750	1	11,428	11,423	11,750	9,533	9,522	9,877
LXPS3033	4,736	0	4,736	4,341	4,736	4,662	3,031	4,708
LXPS3139	5,120	1	5,120	5,120	5,120	4,605	4,443	4,891

	Interval	Min	Max
Virtual Memory Sum	53,521	53,521	53,521
WSS Total	52,472	44,576	52,862
DPA	142,090	142,082	142,090
Memory Utilization %	36.9%	31.4%	37.2%
Memory Overcommit	0.4	0.4	0.4
Available Queue	9	6	39,017
CS<->ES Page Rate	603	2	603

LPAR ES 20,480 MB
LPAR CS 143,360 MB

Cancel
Apply



Efficiency of one. Flexibility of Many. 40 years of virtualization.



IBM Techline Support

- **IBM Techline Support – z/VM Capacity Planning**
 - <http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS2875>
- **Contact your IBMer for in-depth analysis**
- **See free tools such as zPCR for processor sizing**
 - <http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS1381>
- **Thanks to following for info on Techline and ATS Offerings:**
 - **Gretchen Frye**
 - **Liz Holland**



Efficiency of one. Flexibility of Many. 40 years of virtualization.



Summary

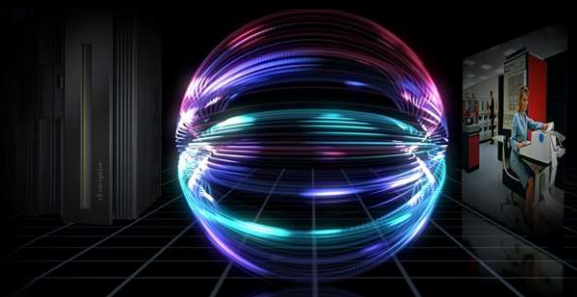
- You have to Plan to do Capacity Planning if you want to do it successfully
- Otherwise, it becomes Capacity Scrambling
- Lots of resources available to help from IBM and Others



Efficiency of one. Flexibility of Many. 40 years of virtualization.



Efficiency of one.
Flexibility of many.
40 years of virtualization.



Contact Information:

Bill Bitner

IBM Endicott

bitnerb@us.ibm.com

+1.607.429.3286



SESSION 12476



Efficiency of one. Flexibility of Many. 40 years of virtualization.



Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

IBM*	System z10*	System z196
IBM Logo*	Tivoli*	System z114
DB2*	z10 BC	System zEC12
Dynamic Infrastructure*	z9*	
GDPS*	z/OS*	
HiperSockets	z/VM*	
Parallel Sysplex*	z/VSE	
RACF*	zEnterprise*	
System z*		

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

OpenSolaris, Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

INFINIBAND, InfiniBand Trade Association and the INFINIBAND design marks are trademarks and/or service marks of the INFINIBAND Trade Association.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

All other products may be trademarks or registered trademarks of their respective companies.

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.