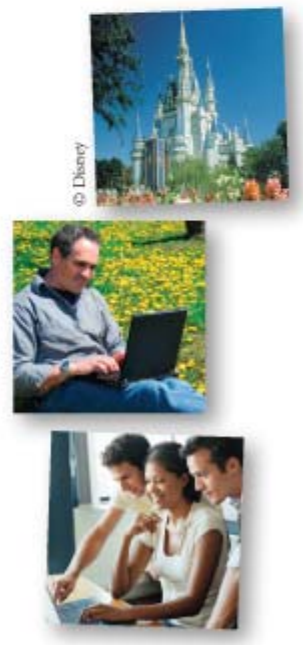
End to End Quality of Service (QoS) in FICON Environments

Dr. Steve Guendert, Brocade
Dennis Ng, IBM

February 29, 2008
Session Number 2575



Abstract

- **This session will discuss QoS in your FICON connected z/OS environment. It will review some basics of QoS, and look at QoS solutions. Next, it will explain why the only true QoS is one that is end to end from host to control unit. It will then review QoS enabling ESCON features that are part of WLM and IRD such as Dynamic Channel Path Management and Channel Priority I/O queuing. We will then explain how and why these features could be used in FICON environments to provide the best end to end QoS performance, availability, functionality, manageability and best practices.**

Definition of QoS

- Service: the expected behavior or outcome from a system
- QoS: the degree to which the expected outcome is realized
- Class of Service (CoS):
  - Defines the type of service but does not indicate how well the service is performed
  - Fibre Channel CoS defines message delivery guarantees which is far different from guaranteeing response times, bandwidths, etc.

Why QoS is a concern now ?

- When FICON was introduced to the world, the significant improvements in bandwidth, IOPS, and interleaving of I/O basically made QoS a non-issue.
- As bandwidth has continued to increase, QoS remained a non-issue as “bandwidth could solve the world’s problems”.
- Or, if a problem arose, we’d just install more hardware to combat the issue.
- Inability support our love ones effectively
  - Over capacity / utilization
  - High availability issues
- Limited resources, $$$, and time to react
- Asked to do more with less
- As architectures have grown more complex, the need to prioritize traffic has grown. Bandwidth is no longer enough. “Loved ones” must be protected. Throwing HW at the problem is not the long term answer

Storage Networking Industry Association (SNIA)'s Definition of QoS

- SNIA mission: **Lead the storage industry worldwide in developing and promoting standards, technologies, and educational services to empower organizations in the management of information.**
- "QoS is a technique for managing computer system resources such as bandwidth, by specifying user visible parameters such as message delivery time. Policy rules are used to describe the operation of network elements to make these guarantees."
- In a nutshell-in a storage network, QoS is a set of metrics that predict the behavior, reliability, speed, and latency of a path in order to protect "loved ones" and meet SLAs.

Predictions

- Based on
  - Algorithms
  - Proactive monitoring
  - Simulation modeling
  - Threshold values
  - Trend analysis techniques
- The goal is to offer performance management for the FICON infrastructure that is more proactive and forward looking, rather than the current reactive and backward looking methods in place.

So, What's the Problem?

- QoS in storage and storage networks is an optimization problem:
  - Achieve optimization by trading one performance metric for another (trade throughput for response time)
- In general, storage subsystems cannot make QoS guarantees:
  - They are constructed to accept, and queue all arriving I/O commands
- Certain performance tuning techniques help achieve partial QoS:
  - Queue prioritization, assigning more buffers/cache, DASD spindles, etc. to high-priority jobs
- **Partial QoS is merely a best effort service level that cannot give QoS guarantees**
- Most existing implementations are partial QoS, that only address performance monitoring and configuration issues in their own little segment of the data center sysplex
- These partial QoS implementations are not true end-to end QoS, which is what the mainframe environment really requires!

Everybody is only partial QoS!

- Nobody is truly end to end on QoS
  - In a FICON environment
- No guarantee on service levels achievement on various segments
- No single orchestrating/management vehicle for QoS
  - Ideal would be to manage from host

Some Ways FICON QoS Is Being Addressed

- The failed experiment with Fibre Channel Class 4 CoS
- InfiniBand's Virtual Lanes
- Virtual Channels
- Director-based software driven mechanisms

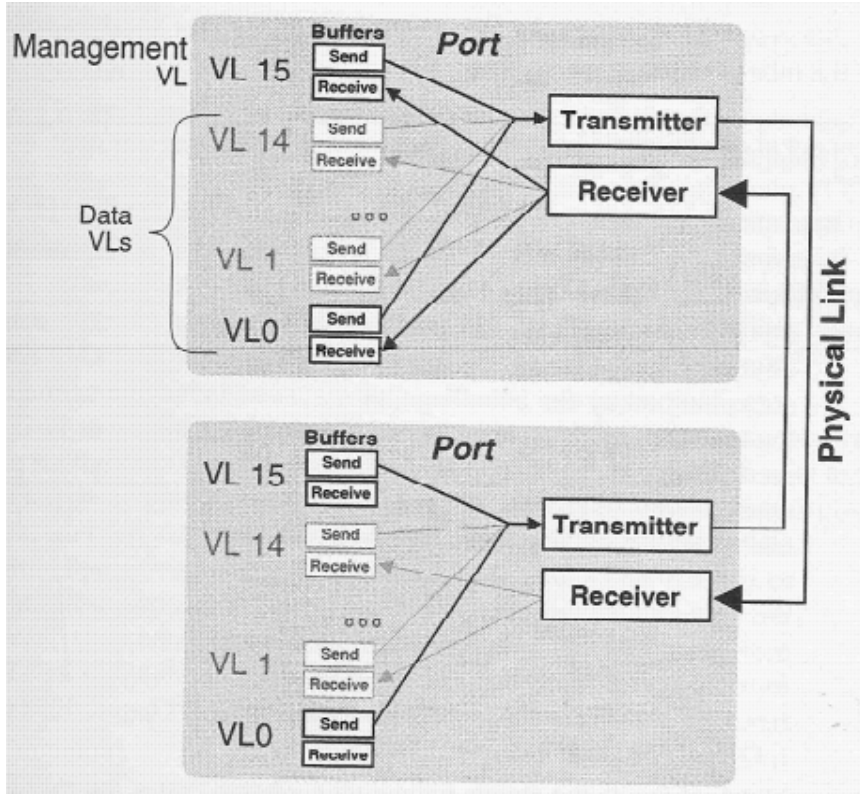
Some Ways FICON Is Being Addressed

- The failed experiment with Fibre Channel Class 4 CoS
- InfiniBand's Virtual Lanes
- Virtual Channels
- Director-based software driven mechanisms

- All are partial QoS mechanisms that only drive QoS for the director, not end to end from host to CU!

Infiniband Virtual Lanes example

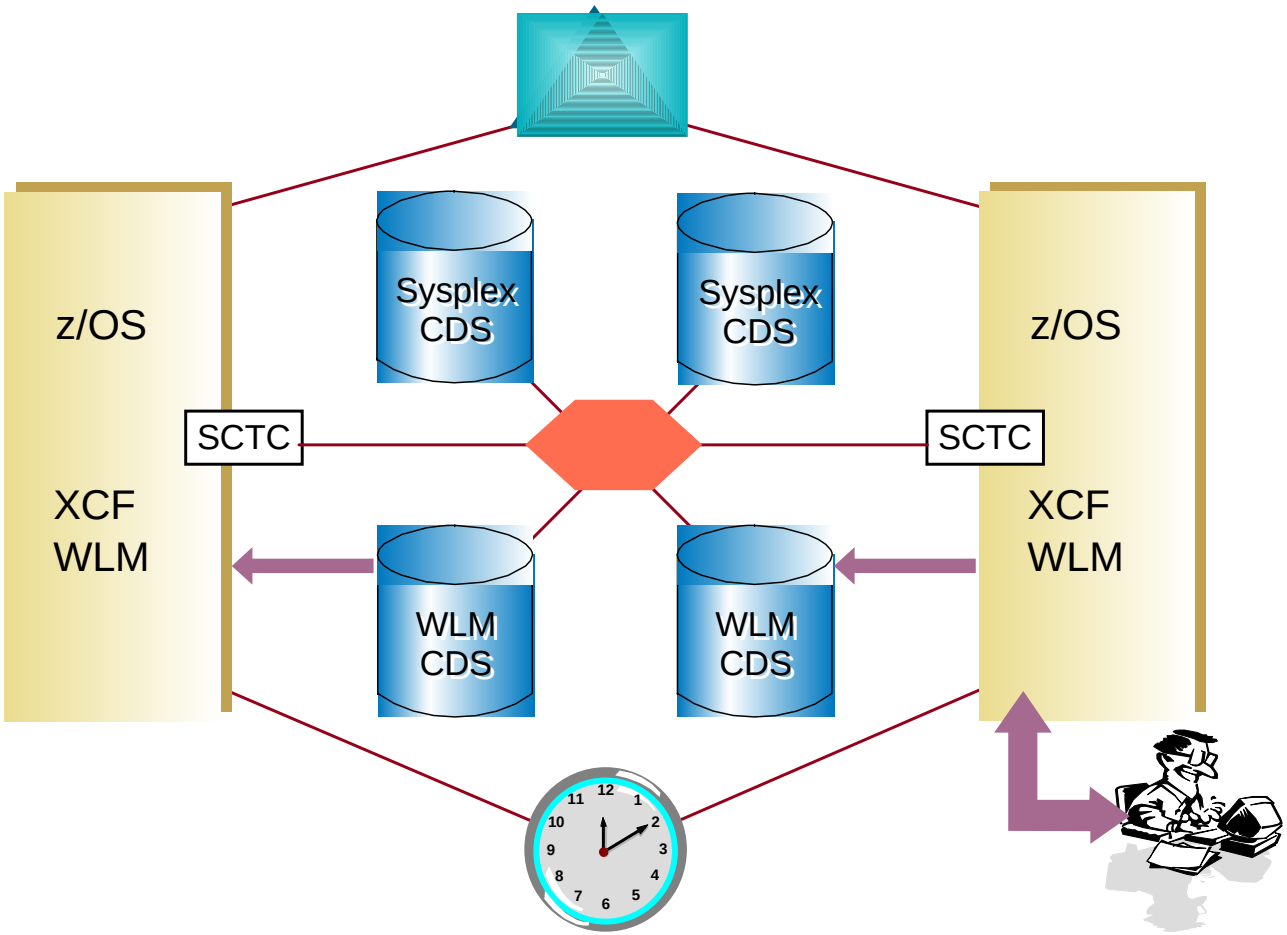
- Logically divide a physical link into multiple virtual links to enable different QoS guarantees across an Infiniband fabric.
  - Priority
  - Latency guarantees
  - Bandwidth guarantees
  - Up to 16 virtual lanes per physical link.



End-to-End QoS

- Workload Manager
- Intelligent Resource Director:
  - Dynamic Channel Path Management
  - Channel Subsystem Priority I/O queuing

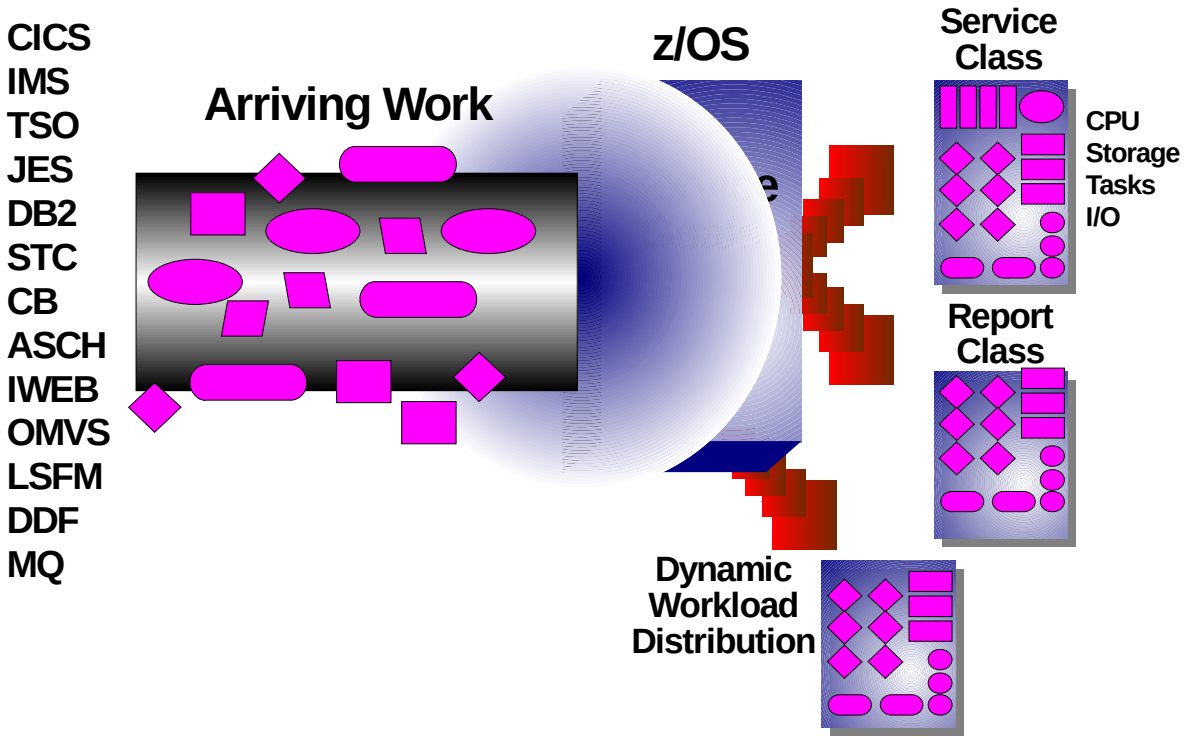
Complete WLM View z/OS



z/VM Resource Manager (VMRM) WLM and VM

- Dynamically tune a system
- Manage workloads to CPU and DASD I/O velocity goals
- Allow I/O priority queuing to be exploited on behalf of VM-based workloads
- Provide an infrastructure for more extensive workload and resource management for future releases of z/VM

z/OS Workload Management

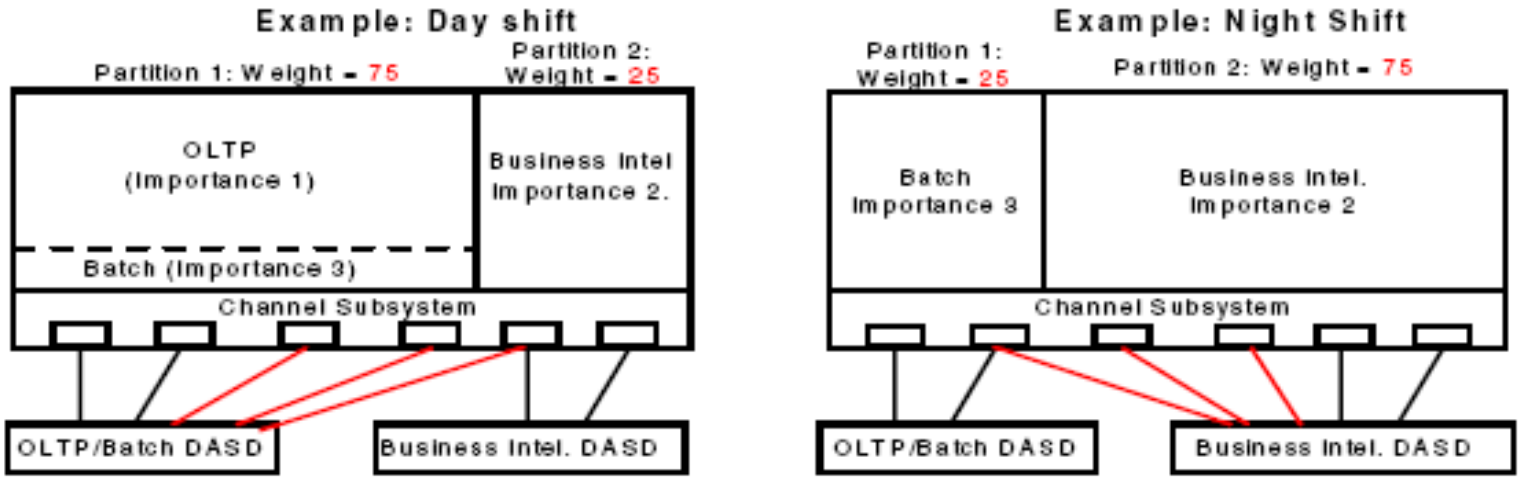


WLM Implementation

- Service class restrictions:
  - Can not be used by another subsystem
  - Velocity goal only
  - Single period
  - Storage protection not allowed
  - Transaction region management not allowed
  - Resource group not allowed
  - CPU critical not allowed
- Hardware restrictions:
  - Shared CP’s only
  - No hard capping
  - Partition enabled for weight management
  - General purpose CP’s only (no IFLs)

IRD

Intelligent Resource Director Benefits



Using WLM, PR/SM, and Parallel Sysplex, resources are dynamically moved to where they are needed

IRD and native FICON

- TBD

IRD

- **Intelligent Resource Director is not actually a product or a system component; rather it is three separate but mutually supportive functions**
  - **WLM LPAR CPU Management**
  - **Dynamic Channel-path Management (DCM)**
  - **Channel Subsystem I/O Priority Queuing (CSS IOPQ)**

What is WLM LPAR CPU Management ?

- `Two parts to WLM LPAR CPU Management:
  - **WLM LPAR Weight Management**
    - **Automatically change the weight of a Logical Partition**
    - **Based on analysis of the current workloads by WLM**
  - **WLM Vary CPU Management**
    - **Automatically vary a logical CP online or offline in an LP**
    - **Based on analysis and requests from WLM**
- Software managing hardware resources:
  - **Software - WLM Goal mode**
  - **Hardware - Shared CPs and Logical Partition weights**
  - **Parallel Sysplex - Used to share WLM information between the systems**

Dynamic Channel Management DCM

- **DCM provides the ability to have the system automatically manage the number of ESCON and FICON Bridge (FCV) I/O paths available to supported DASD subsystems.**
- **DCM currently does “not” support native FICON Express channels**

To be eligible for DCM, control units must:

- **Be attached to ESCON channels or FICON Bridge (FCV) channels via an ESCON Director bridge card**
- **Be attached via a switch**
- **Support multiple paths**
- **==> Potential candidates are disk, tape**
- **==> As disk is most critical in terms of performance and availability , DCM currently supports most current DASD.**

Channel Subsystem Priority Queuing

- Channel Subsystem Priority Queuing is an extension of I/O priority queuing. In an LPAR cluster, if important work is missing its goals due to I/O contention on channels shared with other work, it will be given a higher channel subsystem I/O priority than the less important work. This function works together with Dynamic Channel Path Management — as additional channels are moved to the partition running the important work, Channel Subsystem Priority Queuing is designed so that the important work that really needs it receives the additional I/O resource, not necessarily the other work that just so happens to be running in the same LPAR cluster.

I/O Priority

- The following conditions must be met to implement WLM CSS I/O Priority Management:
  - **IBM 2064 processor running z/OS in z/Architecture mode**
  - **Basic or LPAR mode**
  - **z/OS running in WLM Goal mode**
  - **I/O priority management set to YES in WLM policy**
  - **CSS I/O priority management enabled in CPC Reset Profile**
  - **Valid range of CSS I/O priorities specified in Image Profile**

I/O priority queuing

- WLM Goal mode role in I/O priority queuing
- WLM assigns UCB and CU I/O priority so that:
  - **System-related SCP is assigned highest priority**
  - **An SCP missing its goal because of I/O delay gets helped**
  - **An SCP competing for the same devices with a more important, and not happy (PI > 1), SCP is the donor**
- W LM assigns CS S I/O priorities so that:
  - **System-related SCP is assigned highest priority**
  - **High Importance SCP missing goals has next highest**
  - **SCPs meeting goals are managed so that light I/O SCPs have a higher I/O priority than heavy I/O users**
  - **Discretionary work has the lowest priority**
- I/O weight is determined as the ratio o f Co n n e c t Time to Elapsed Time

Summary of WLM

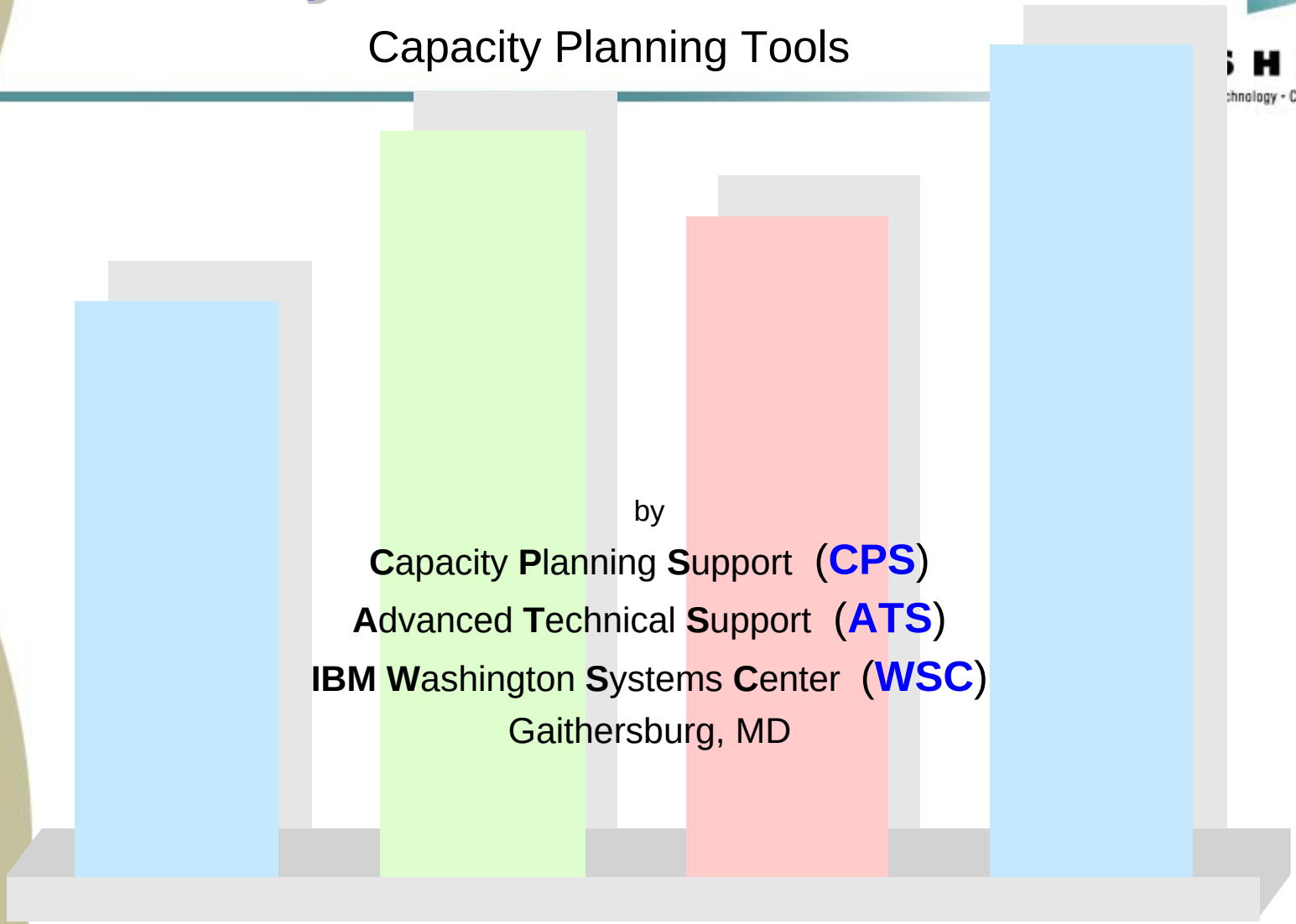
- IRD is the next step in having the manage itself
  - Brings the resources to where the work is
  - Reduces human intervention
- LPAR Weight Management
  - Adjusts the weights of partitions within the LPAR cluster to bring CPU capacity to high importance work
- CPU Varying Management
  - Adjusts the number of online CPs to each partition in the LPAR cluster
  - Optimizes the number of online logical CPs for the dynamic weight of the partition

What does capacity tools mean to QoS ?

- If you don’t understand your environment limitations, it’s very difficult to develop a baseline for evaluation.
- Understanding of the weakest link in your fabric infrastructure (ex. ISL links or trunk)
- Being able to enhance other “traffic cops” with the known limitations

IBM System z10 & zSeries

Capacity Planning Tools



by

Capacity **P**lanning **S**upport (**CPS**)

Advanced **T**echnical **S**upport (**ATS**)

IBM Washington **S**ystems **C**enter (**WSC**)

Gaithersburg, MD

Capacity Planning Tools

zPCR Processor Capacity Reference for IBM System z

zCP3000 Performance Analysis and Capacity Planning for IBM System z

zPSG Processor Selection Guide for IBM System z

zMCAT Migration Capacity Analysis Tool for IBM System z

zTPM Tivoli Performance Modeler

zVM-Planner z/VM Planner for Linux Guests on System z

SoftCap Software Migration Capacity Planning Aid

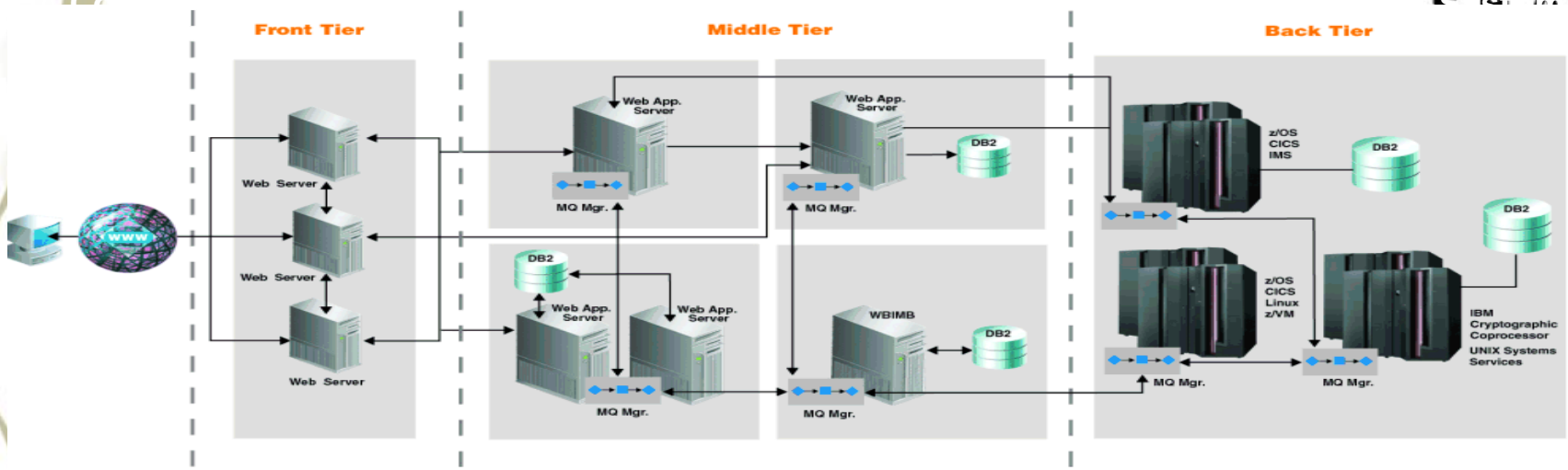
BWATOOL Batch Workload Analysis Tool

zTPM

Tivoli Performance Modeler

- Model z/OS based Systems
  - Uses RMF or CMF reports as input
  - Summary charts showing current LPAR and workload utilizations
  - Uses simulation techniques to model the impact of changes
    - Change workload volumes or hardware configuration
    - Combine multiple system images onto a single processor
    - Workload growth
    - Change LPAR definitions
  - Results are presented in the form of graphs

The World of Composite Application Management



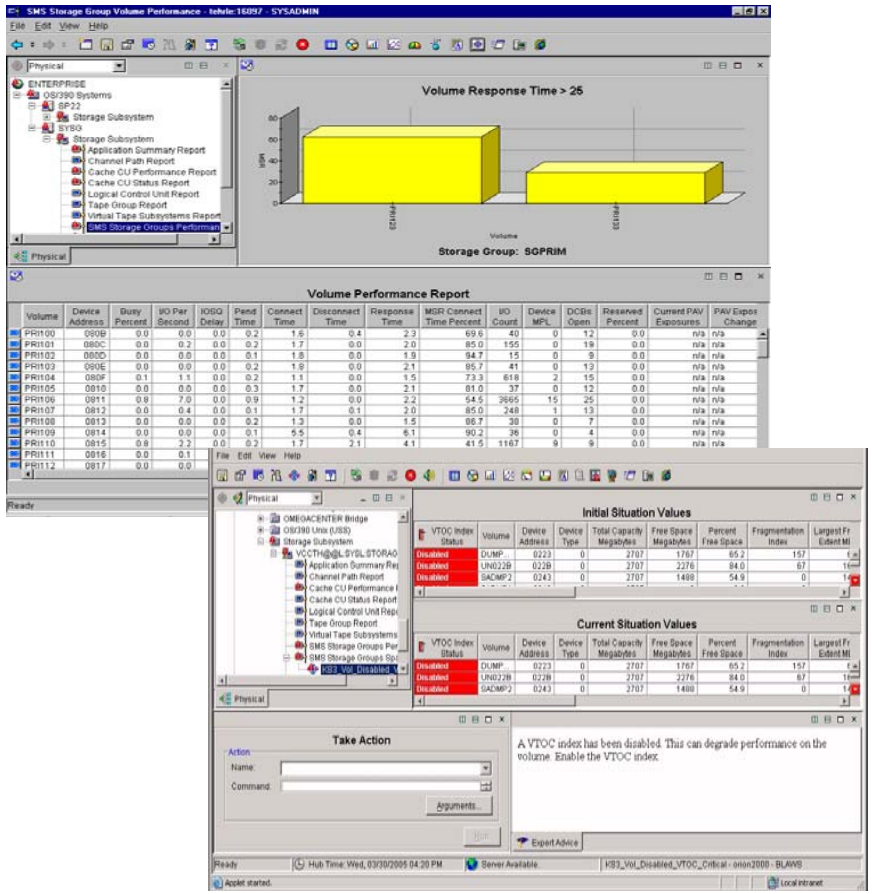
- Business processes increasingly depend on multi-tier composite applications
- Challenging to design, build, test, and manage for high performance and availability
- Traditional stovepiped management processes and tools exacerbate problem

Proactive tools and others

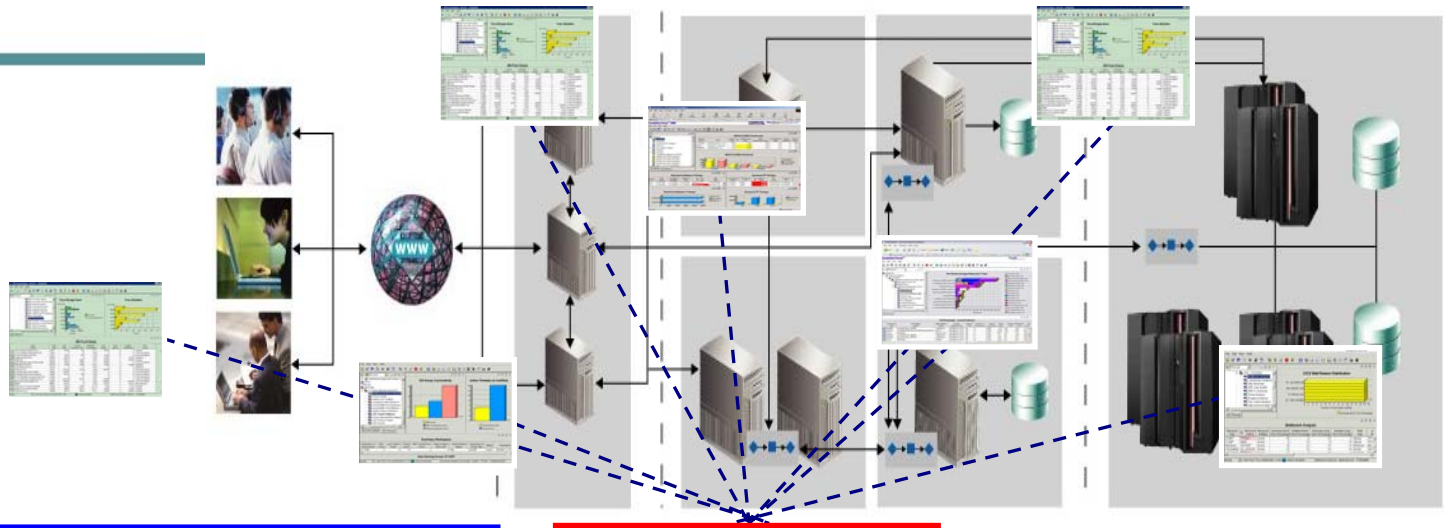
- Tivoli Enterprise Portal (TEP)
- Tivoli Omegamon XE
- ITCAM
- Brocade QoS ASIC
- Cisco QoS VOQ
- Integrated Removable Media Manager (IRMM)

The Tivoli Enterprise Portal (TEP)

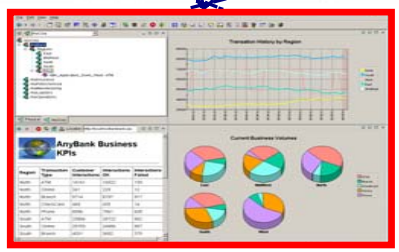
- Advanced interoperability via workspace linkages
  - Enhances and improves problem resolution
  - TN3270 and Browser interface
- Leverage the information from multiple sources
- Capability of creating alerts using multi-level logic as opposed to just threshold setting
- Ability to see alerts and associated information about the problem from the same interface
- Reflex automation capabilities (take action)
- Provide information about a situation through the Expert Advise feature
- Real-time, near term and long term history



Tivoli Enterprise Portal (TEP)



- TEP provides a dashboard to IT Service Management
  - a single point of control for monitoring and managing resources and applications
- TEP consolidates multiple components into a single customizable view



Tivoli Enterprise Portal

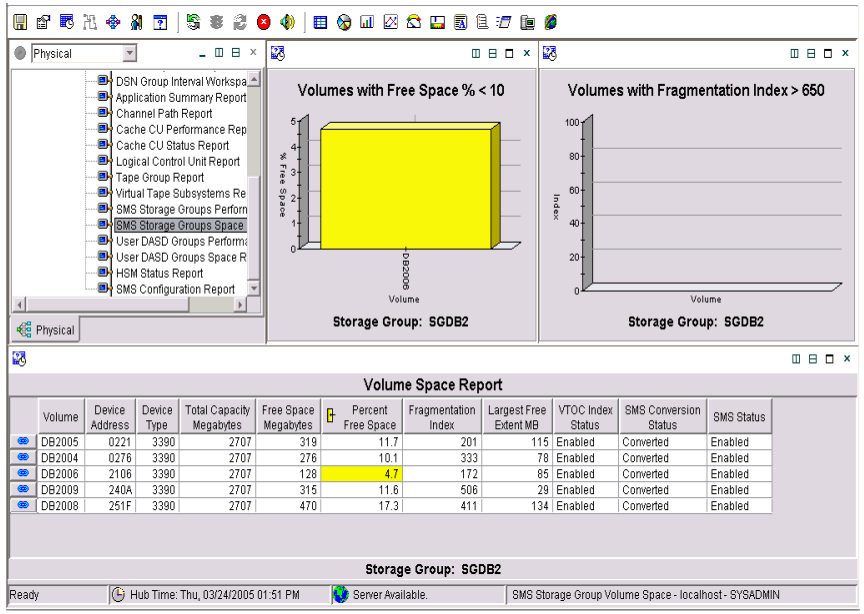
Resource Monitoring:
- WebSphere Servers
- zSeries Subsystems
- Distributed OS
- DB2
- MQ messaging
- Netview

Integrated Portal:
- ITCAM
- Tivoli Distributed Monitoring
- OMEGAMON XE for zSeries

Monitor the overall health the infrastructure through one single portal

IBM Tivoli OMEGAMON XE for Storage on z/OS

- A mainframe STORAGE monitor, real-time and historical
- XE user interface, comes with the CUA UI component
- A wide breadth of mainframe storage information:
  - **Space (storage groups or user groups … define your own)**
  - **Performance (storage groups or user groups … define your own)**
  - **Tape / VTS**
  - **CACHE**
  - **Channels (FICON)**
  - **Control Units**
  - **DFSMShsm (View your HSM queues, control Datasets, etc.)**
  - **DFSMShsm/DFSMSdss online toolkit**
  - **SMS constructs**
  - **DS8000 support**
  - **Ability to see all logical volumes on a physical disk**
  - **Powerful applications view**
  - **Powerful dataset view and action capability**
  - **Integration capabilities from TEP interface (Launch to TPC in v4.1.0)**



ITCAM for RTT Identifies the Problems

SHARE
Technology · Connections · Results

Identify and Prioritize

Isolate and Assign

Diagnose and Config

Fix

Dashboard [Next refresh: 3:54]

Filter

Events over the past 8 hours

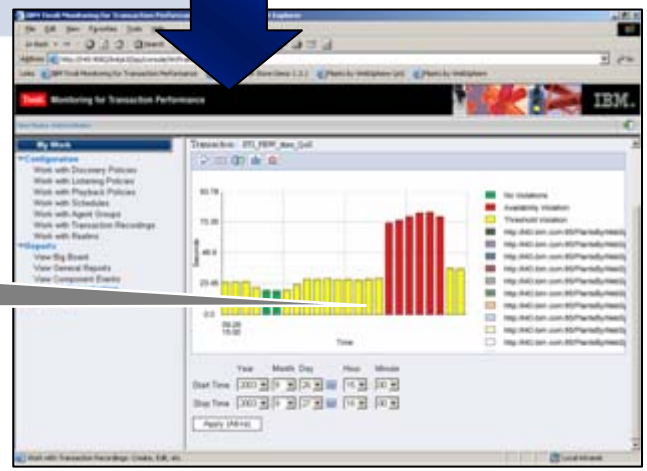
Please make a selection | Go Alt+g

Policy Group	Status	Violating/Total Policies	Agent Availability	Performance Violation Events	Availability Violation Events	Time Since Last Violation Event (HH:MM)	Total Transactions	Number of Performance Violations	Number of Availability Violations
Texas	Warning	2/2		0	0	>8	788	0	788
Arizona	Normal	0/1		0	0	>8			
Arm	Normal	0/3		0	0	>8			

Page 1 of 1 | 1 | Go Alt+g

Dashboard shows transactions with a availability or response time problem

Click to see response time trend



ITCAM for RTT Isolates and Assigns the Problem

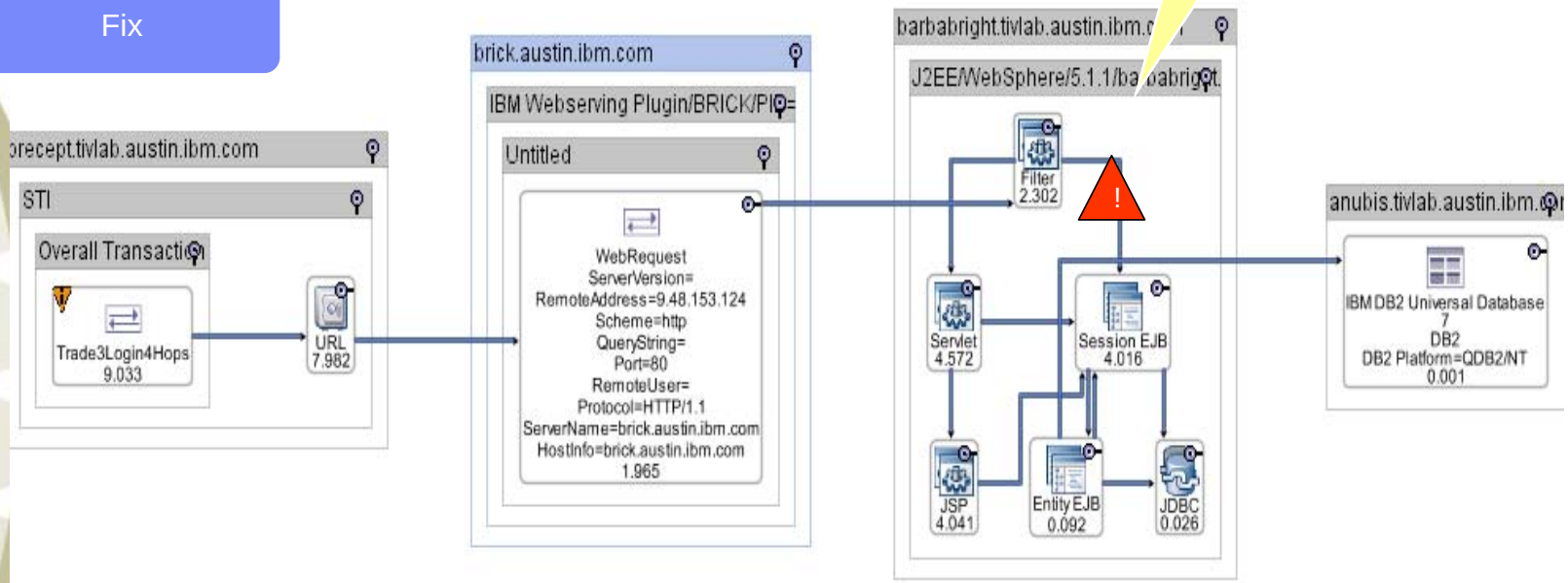
Identify and Prioritize

Isolate and Assign

Diagnose and Config

Fix

Topology shows problem system so that it is assigned to correct SME



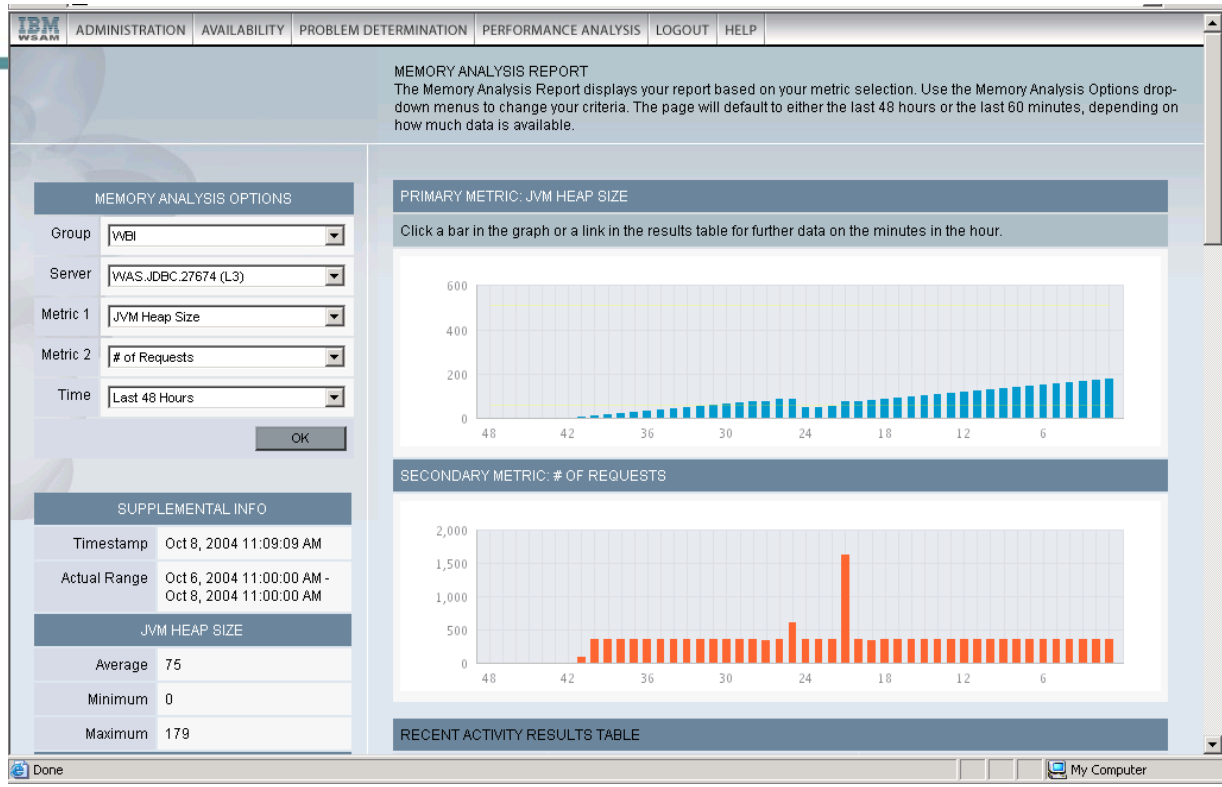
ITCAM for WebSphere Diagnoses and Corrects Problems

Identify and Prioritize

Isolate and Assign

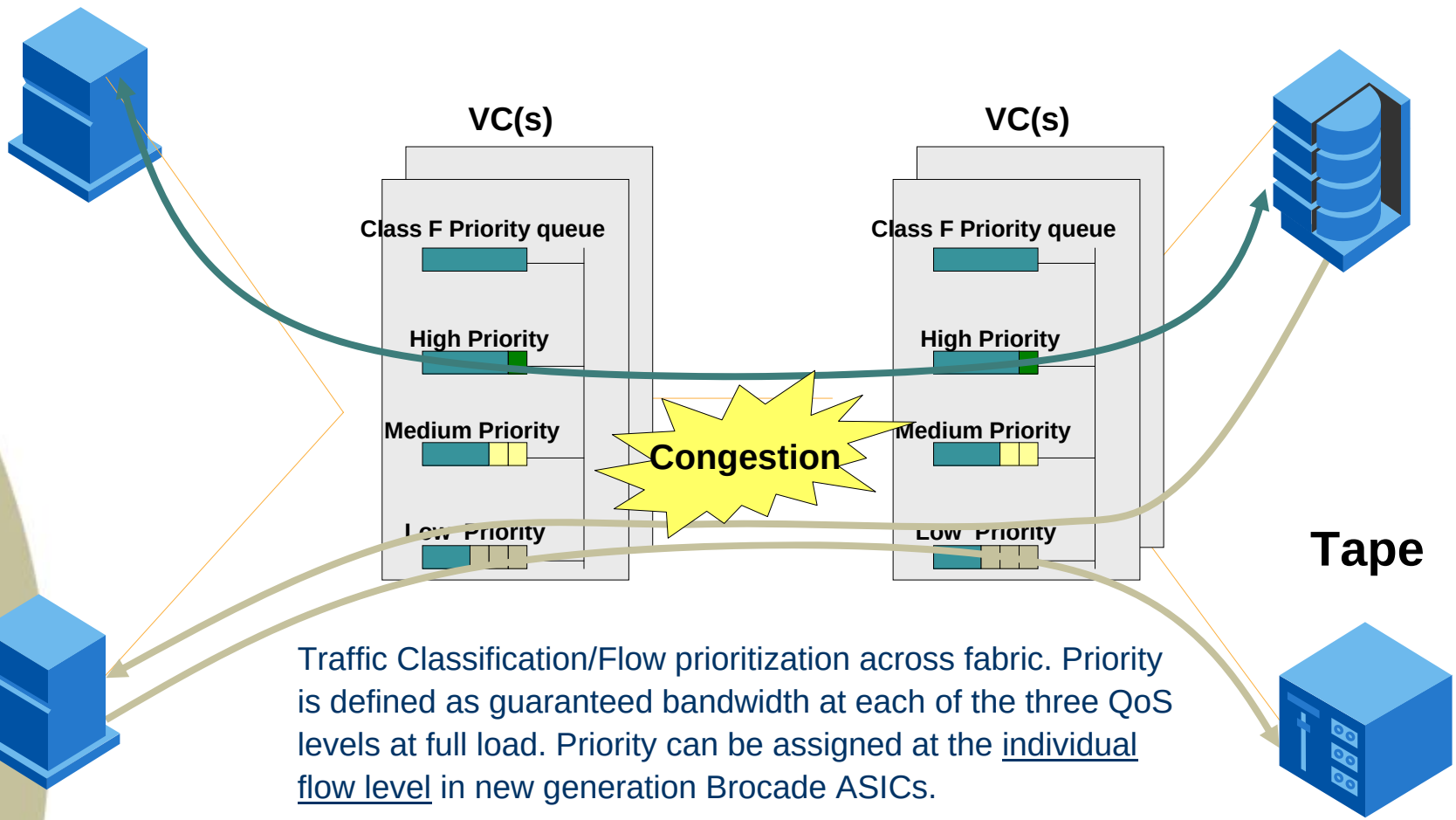
Diagnose and Config

Fix



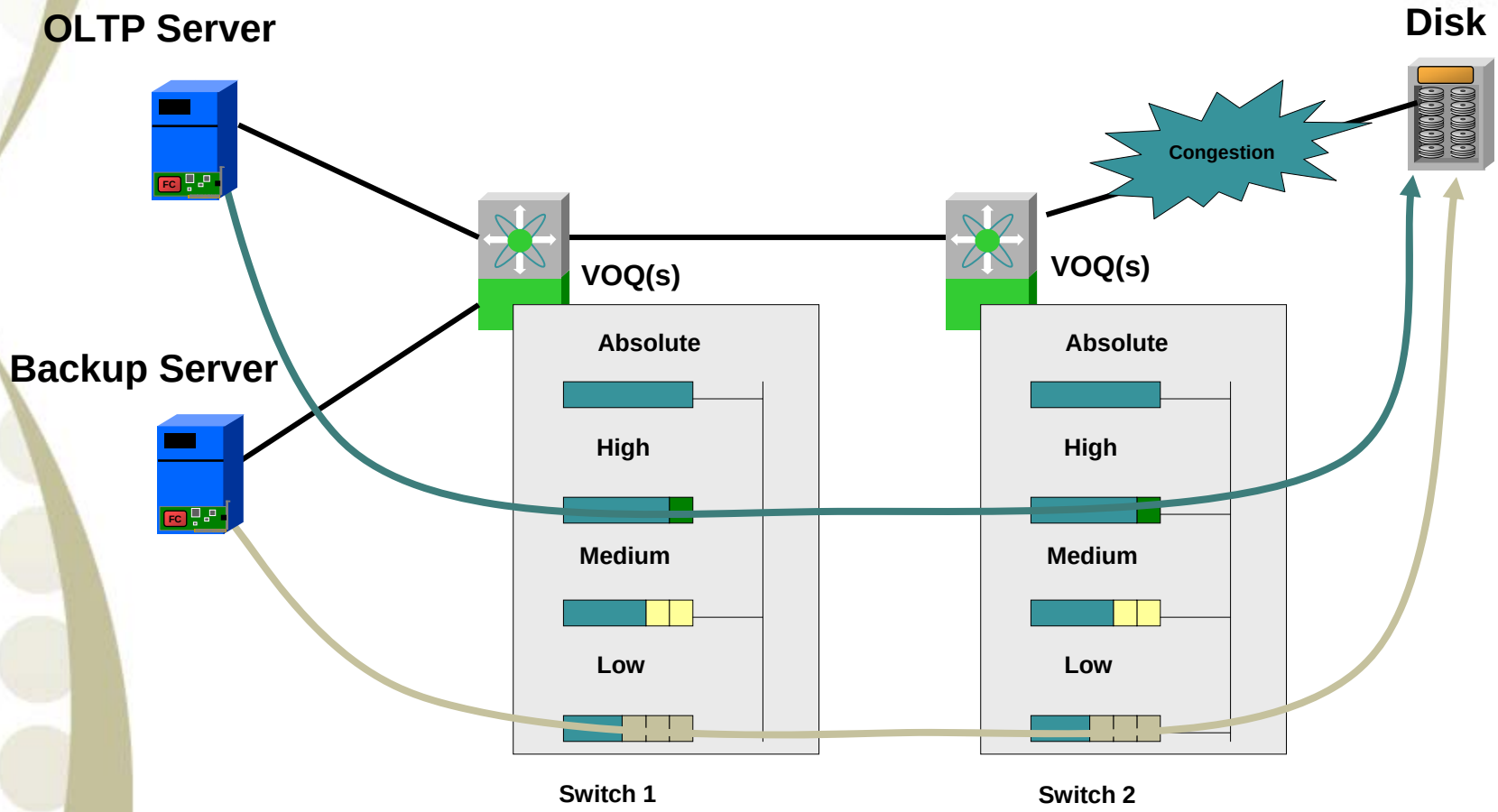
- Are memory heap sizes indicating a leak?
- Is performance trending towards a failure?
- Can I easily identify if the problem is with WAS or with the application?

Brocade QoS – Basic Overview



Traffic Classification/Flow prioritization across fabric. Priority is defined as guaranteed bandwidth at each of the three QoS levels at full load. Priority can be assigned at the individual flow level in new generation Brocade ASICs.

Cisco QoS for Data Traffic



Queuing and Scheduling:

- 4 VOQ per egress port
- Scheduling algorthim is based on DWRR

IRMM in a Nutshell

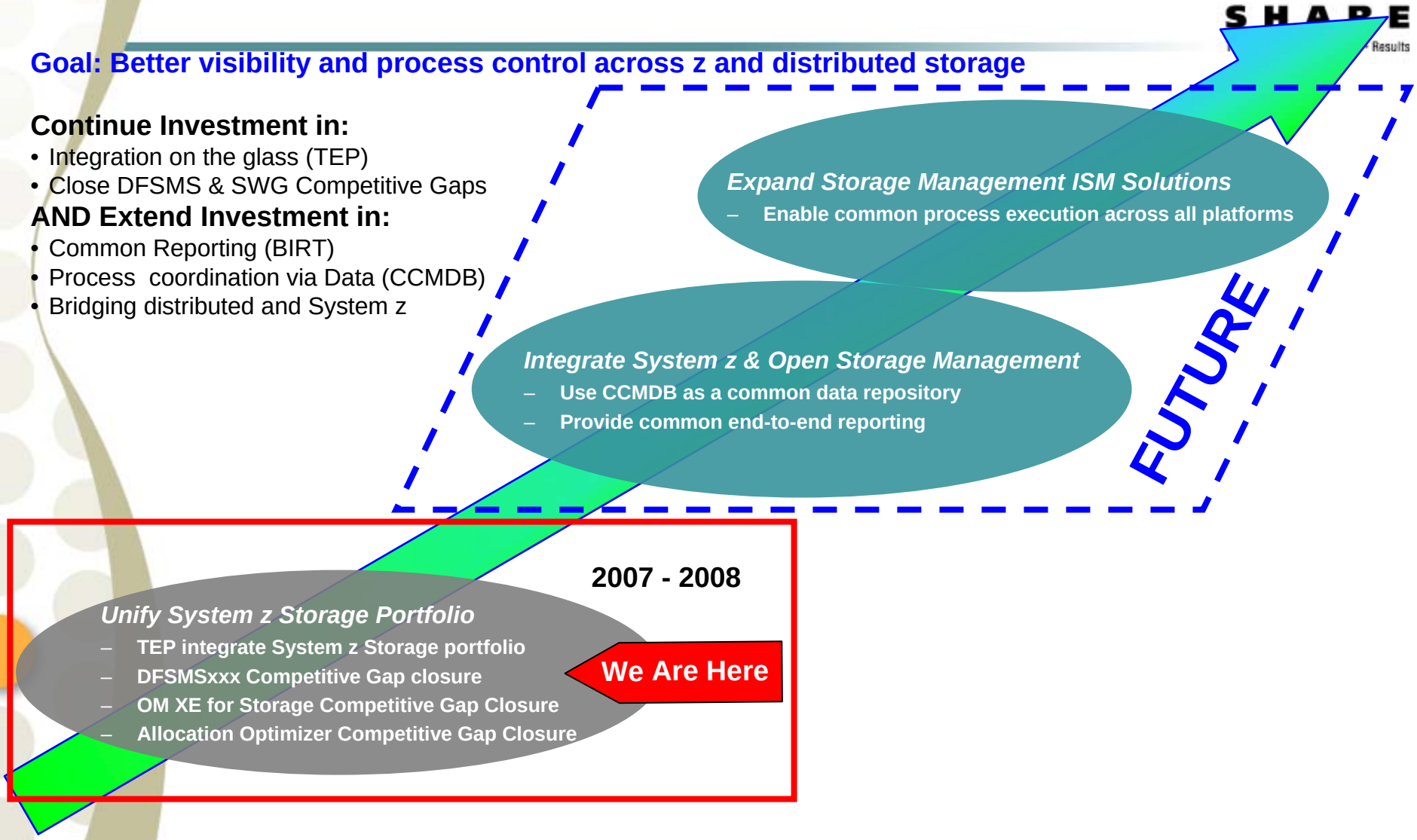
- Integrated Removable Media Manager (IRMM) is:
  - **A new robust systems management product for Linux® on IBM System z™ that manages open system media in heterogeneous distributed environments and virtualizes physical tape libraries, thus combining the capacity of multiple heterogeneous libraries into a single reservoir of tape storage that can be managed from a central point**
- IRMM is designed to provide:
  - **Centralized media and device management**
  - **Dynamic resource sharing**
- IRMM extends IBM's virtualization strategy to tape library resources (drives and cartridge pools)
- IRMM complements Linux on System z consolidation efforts

IBM System z Storage Management Future Direction

Goal: Better visibility and process control across z and distributed storage

Continue Investment in:
- Integration on the glass (TEP)
- Close DFSMS & SWG Competitive Gaps

AND Extend Investment in:
- Common Reporting (BIRT)
- Process coordination via Data (CCMDB)
- Bridging distributed and System z



Future considerations for QoS

- Tivoli applications
  - Tivoli workload automation
- Cisco QoS products and applications
- Brocade QoS products and applications
- Flexibility=additional complexity

- **Host managed, pro-active end to end QoS for mainframe-mainframe storage environments.**

Conclusion: 3 things in life

- Death
- Taxes
- Change

Be Pro-active!

- “Setting the goal is not the main thing, it is deciding how you will go about achieving it and staying with that plan”.
  - Tom Landry

- “Some people want it to happen, some people wish it would happen, some people make it happen.”
  - Michael Jordan

SHARE
Technology · Connections · Results

Thank you