Overview

Configuring z/VM for Linux on zSeries

- Must configure z/VM – many defaults incorrect
- Linux must be configured for shared resource environment
- Many actions not intuitive

Infrastructure unknowns for “new” installations

- How to manage performance / capacity planning?
- What are the limits of a configuration and how to measure
- How to share resources to reduce ROI

Measurement and Tuning for z/VM IS Required

- Start with Proper Configurations
Configuration Options

General Storage Options

Linux Options
- Storage Sizes
- Swapping for Linux
- Linux virtual processors
- Network

z/VM Configuration
- Network, Virtual Switch, I/O, FTP Topics
- MDC
- Paging and Spooling for z/VM
- DASD/Cache/Channels
- z/VM System parameters
- Expanded Storage

Infrastructure
- Linux infrastructure – monitoring availability and performance

General Storage Requirements

Configuration requirements different for
- Small Infrastructure Servers – “Small” Systems
  - DNS, Apache, Samba
  - Low I/O rate
  - Real storage less than 2gb
  - Virtual servers sized 64mb to 256mb
- Medium (31bit) Application Servers - Small to Large Systems
  - Websphere, Domino, Oracle
  - z/VM Real storage greater than 2GB
  - High I/O rate potential
  - Typical 512MB to 2GB
- Large (64bit) Application Servers - Large Systems
  - Oracle, SAP
  - z/VM Real storage greater than 10GB
  - High I/O rate potential
  - Virtual Servers Typical 512MB to 16GB
Storage Considerations

z/VM is shared resource environment
- Over-committing storage improves costs per server
- Over-allocating storage reduces servers that can be supported
- QDROP IS QUITE IMPORTANT

Storage requirements of Linux very high
- Linux designed for dedicated storage, references all storage
- Linux is LRU, competing with VM's reference pattern
- High percent of referenced pages – what can z/VM page out?

Linux does not drop from queue –
- 100 timer pops per second was 1st problem, fixed.
- CP storage management bypassed, forces “emergency scan”
- Current release of IBM JDK (WAS) polls 10 ms

Listserv comment (about SAP): “One more thing, the ABAP stacks go dormant when no users are working, but the dual stack never goes dormant. Rob vd Heij checked for me that is due to a 10 millisecond timer” in Java that we can’t get rid of due to SAP Java version requirements.

Reed Mullen: “The IBM WebSphere team plans to further study the performance/scalability issues presented by Velocity Software. While there is no outcome to communicate at this time, it is IBM’s desire to address these issues.”

This now seems to impact Domino, WAS, DB2 and Tivoli applications.
Qdrop issues

Storage in use by LINUX07 has 166K pages

- Server was active 8 hours prior, but idle for 8 hours
  - never dropped from queue, never gave up storage
- Active server LINUX02 must compete for reduced storage

Guideline: Force Linux Servers to drop from queue

Linux Storage

z/VM Paging

- Over commitment of storage causes paging
- **Over commitment of storage reduces cost**
- Paging is common (manageable) performance problem

Linux Swapping

- Swapping result of over commitment of Linux storage
- Swapping to vdisk very fast, uses storage when it happens
- Swapping to dasd very slow, always noticeable
Linux Cache

- Linux avoids I/O by using cache
- Linux will cache gigabytes of data if allowed
- Oracle SGA MUST fit in cache Use this feature in z/VM environment
- Swap historically was slow SCSI device

Reduce size of Linux Virtual Machine MAJOR Knob.

- Reducing virtual machine size reduces caching of old data
- Define virtual disk for swap
- Virtual Disk paged out when not in use - Unlike “Real” memory
- Experiment with Linux server swapped 40,000 per second.
- Experiments with swap to dcss in z/VM 5.4 will start soon

---

Tailoring Linux Storage

Linux data shows

Real storage
Swap storage “cache”

Some Swapping is “good”

If not swapping, reduce vm size
Use CMM to reduce

---
**Linux Swapping**

Reducing virtual storage size may cause swap
- Linux does not swap until out of storage

Swapping to disk
- VERY VERY SLOW
- Other platforms increase storage size because disk is slow
- **Swap to disk if you want to penalize a server**
- Max swap rate maybe 200 on a very good day

Linux Swapping to Vdisk
- Not a performance degradation
- 40,000 / second is FAST

**Linux Swap to Virtual Disk**

Virtual disk for swap pros
- Very very fast swapping
- Very high bandwidth
- Requires very little resource if unused
- Swap I/O does not require below the 2gb line storage

Virtual disk cons
- Requires control block storage below the line
- Swapping requires CPU, must manage swap rate

Swap Guideline:
- Define 2 virtual disks, prioritized swap
- Use DIAG driver instead of FBA - Reduces I/O by factor of 8
VM Storage Overview, Paging Hierarchy

z/Linux Page / Swap Hierarchy

- Utilize features of z/VM – Virtual Disk
- Linux not limited in swap rate,
- z/VM manages storage, high band width
**z/VM Paging Hierarchy**

- z/VM paging bandwidth very high, multi-level

![Diagram of z/VM paging hierarchy]

**Linux Storage Case Study**

- First case study:
  - Process took hours, system paged significantly
  - Reduced size of Linux Virtual Machine, 128mb to 24mb
  - Defined 100MB Swap disk
  - Linux reduces storage requirement
  - Process took minutes

- Virtual Disk paged out when not in use
  - This works!!! Paging greatly reduced, Linux performance greatly improved!!!

- This research critical to using Collaborative Memory Mgmt (CMM)
LINUX Swapping to VDISK

Change 128MB Server to 24MB with 100MB Swap
Reduction of Overall Storage Requirements of 100MB

- Unused VDISK is paged out

Virtual Storage vs Virtual Disk tradeoffs

Virtual Disk I/O 838K / 900 seconds
- About 900 - 1,000 per second
- (NOTE MDISK HIT RATE!!!!!)

Report: ESAUSR3 User Resource Utilization - Part 2 Domino Redbook ESAMAP 3.4.0
Monitor initialized: on 2066 serial 71CE3 First record analyzed: 08/21/03 12:00:00

<table>
<thead>
<tr>
<th>Time</th>
<th>User ID</th>
<th>DASD Block</th>
<th>Cache Disk</th>
<th>Hit Prty</th>
<th>I/O Requests</th>
<th>I/O Hits</th>
<th>I/O Cons</th>
<th>U/R</th>
<th>CTCA</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>08/21/03</td>
<td>12:15:00</td>
<td>613K</td>
<td>240K</td>
<td>838K</td>
<td>74.8</td>
<td>0</td>
<td>1510</td>
<td>0</td>
<td>321</td>
<td>0</td>
</tr>
<tr>
<td><strong>Top User Analysis</strong></td>
<td>LINUXA</td>
<td>610K</td>
<td>240K</td>
<td>838K</td>
<td>74.8</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12:20:00</td>
<td>615K</td>
<td>250K</td>
<td>822K</td>
<td>74.6</td>
<td>0</td>
<td>1487</td>
<td>0</td>
<td>324</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Top User Analysis</strong></td>
<td>LINUXA</td>
<td>613K</td>
<td>240K</td>
<td>822K</td>
<td>74.6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12:45:00</td>
<td>631K</td>
<td>260K</td>
<td>884K</td>
<td>75.5</td>
<td>0</td>
<td>1634</td>
<td>0</td>
<td>321</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Top User Analysis</strong></td>
<td>LINUXA</td>
<td>628K</td>
<td>258K</td>
<td>884K</td>
<td>75.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Cost of Swap daemon about 10%

| Date     | Application | Total active | Running | ResWait | Loaded | Percent utilization | seconds | Avg
|----------|-------------|--------------|---------|---------|--------|---------------------|---------|-----
| 08/21/03 | LINUXA java | 15.0         | 15.0    | 2.0     | 13.0   | 10.3                | 92.6    | 0.7 |
|          | kswapd      | 1.0          | 1.0     | 0       | 1.0    | 9.1                 | 82.9    | 9.1 |
|          | router       | 11.0         | 11.0    | 0       | 11.0   | 10.6                | 98.4    | 1.0 |
|          | server       | 67.0         | 67.0    | 1.0     | 63.0   | 63.2                | 566.5   | 0.9 |
|          | snmpd        | 1.0          | 1.0     | 0       | 1.0    | 3.3                 | 29.3    | 3.3 |
|          | update       | 3.0          | 3.0     | 1.0     | 2.0    | 10.2                | 91.7    | 3.6 |

Cost of Swap daemon about 10%

Virtual Disk Analysis

On following report, all virtual disks shown,

- Size in pages, Resident pages
- ESAMON Alert should be set for high resident vdisk pages
- PAZXXT03 should be given more storage AND 2nd vdisk

<table>
<thead>
<tr>
<th>Owner</th>
<th>Spare Name</th>
<th>Space Name</th>
<th>AddSpc VDISK</th>
<th>Cre- Del-</th>
<th>or rate</th>
<th>Priv VIDO</th>
<th>size</th>
<th>pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:02:00</td>
<td>PAZXQ01</td>
<td>VDISPAXZXQ0100230010012 25856 205K</td>
<td>0</td>
<td>0</td>
<td>Shrd</td>
<td>0</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>PAZXQ01</td>
<td>VDISPAXZXQ0100230010010 25856 205K</td>
<td>0</td>
<td>0</td>
<td>Shrd</td>
<td>0</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>PAZXQ02</td>
<td>VDISPAXZXQ0100230010011 25856 205K</td>
<td>0</td>
<td>0</td>
<td>Shrd</td>
<td>0</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>PAZXQ02</td>
<td>VDISPAXZXQ0100230010013 25856 205K</td>
<td>0</td>
<td>0</td>
<td>Shrd</td>
<td>0</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>PAZXQ02</td>
<td>VDISPAXZXQ0100230010012 25856 205K</td>
<td>0</td>
<td>0</td>
<td>Shrd</td>
<td>0</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>PAZXQ02</td>
<td>VDISPAXZXQ0100230010010 25856 205K</td>
<td>0</td>
<td>0</td>
<td>Shrd</td>
<td>0</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>PAZXQ02</td>
<td>VDISPAXZXQ0100230010011 25856 205K</td>
<td>0</td>
<td>0</td>
<td>Shrd</td>
<td>0</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>PAZXQ02</td>
<td>VDISPAXZXQ0100230010013 25856 205K</td>
<td>0</td>
<td>0</td>
<td>Shrd</td>
<td>0</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>PAZXQ02</td>
<td>VDISPAXZXQ0100230010012 25856 205K</td>
<td>0</td>
<td>0</td>
<td>Shrd</td>
<td>0</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>PAZXQ02</td>
<td>VDISPAXZXQ0100230010010 25856 205K</td>
<td>0</td>
<td>0</td>
<td>Shrd</td>
<td>0</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>PAZXQ02</td>
<td>VDISPAXZXQ0100230010011 25856 205K</td>
<td>0</td>
<td>0</td>
<td>Shrd</td>
<td>0</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>PAZXQ02</td>
<td>VDISPAXZXQ0100230010013 25856 205K</td>
<td>0</td>
<td>0</td>
<td>Shrd</td>
<td>0</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>PAZXQ02</td>
<td>VDISPAXZXQ0100230010012 25856 205K</td>
<td>0</td>
<td>0</td>
<td>Shrd</td>
<td>0</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>PAZXQ02</td>
<td>VDISPAXZXQ0100230010010 25856 205K</td>
<td>0</td>
<td>0</td>
<td>Shrd</td>
<td>0</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>PAZXQ02</td>
<td>VDISPAXZXQ0100230010011 25856 205K</td>
<td>0</td>
<td>0</td>
<td>Shrd</td>
<td>0</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>PAZXQ02</td>
<td>VDISPAXZXQ0100230010013 25856 205K</td>
<td>0</td>
<td>0</td>
<td>Shrd</td>
<td>0</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>PAZXQ02</td>
<td>VDISPAXZXQ0100230010012 25856 205K</td>
<td>0</td>
<td>0</td>
<td>Shrd</td>
<td>0</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>PAZXQ02</td>
<td>VDISPAXZXQ0100230010010 25856 205K</td>
<td>0</td>
<td>0</td>
<td>Shrd</td>
<td>0</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>PAZXQ02</td>
<td>VDISPAXZXQ0100230010011 25856 205K</td>
<td>0</td>
<td>0</td>
<td>Shrd</td>
<td>0</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>PAZXQ02</td>
<td>VDISPAXZXQ0100230010013 25856 205K</td>
<td>0</td>
<td>0</td>
<td>Shrd</td>
<td>0</td>
<td>1</td>
<td>32</td>
</tr>
</tbody>
</table>
Additional Storage Performance

Named Saved System
  • Fast IPL, shared kernel storage
DCSS with XIP File System
  • Load all programs into shared DCSS,
CMM: Collaborative memory management
  • Dynamically manage storage size

How many Virtual Processors?

- Linux is multiprocessor capable
- Global lock is very large issue
  - One processor acquires lock
  - Other processors attempt to spin
  - On 390 – spin converted to Diagnose 44
- Problem easily detected
  - High Diagnose -> Instruction Simulation -> SIE
  - High TV ratio
  - Guideline: Minimize virtual processors

- Linux updates reduce dependency on global lock (253 kernel)
### How many Virtual Processors?

**Report: ESACPUA**  
**CPU Utilization Analysis**

<table>
<thead>
<tr>
<th>Time</th>
<th>CPU Util</th>
<th>Ttotl Ovrhead</th>
<th>Diag</th>
<th>Inst</th>
<th>SIE</th>
<th>Fast Path Rate</th>
<th>Sim intrcp</th>
<th>path fault</th>
<th>/sec</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16:01:00</td>
<td>0 66.6 12 25</td>
<td>60K</td>
<td>82K</td>
<td>83275</td>
<td>2108</td>
<td>0.1</td>
<td>350</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>67.6 12 25</td>
<td>89K</td>
<td>91K</td>
<td>91879</td>
<td>1051</td>
<td>0 332</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>62.3 12 24</td>
<td>83K</td>
<td>85K</td>
<td>85768</td>
<td>1219</td>
<td>0.1</td>
<td>383</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>62.7 11 25</td>
<td>77K</td>
<td>78K</td>
<td>79354</td>
<td>776</td>
<td>0 293</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>63.6 12 24</td>
<td>84K</td>
<td>85K</td>
<td>86175</td>
<td>1047</td>
<td>0 329</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>63.1 11 26</td>
<td>82K</td>
<td>84K</td>
<td>85064</td>
<td>1188</td>
<td>0 297</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>64.1 11 22</td>
<td>83K</td>
<td>84K</td>
<td>84874</td>
<td>1079</td>
<td>0 304</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>57.3 10 22</td>
<td>73K</td>
<td>75K</td>
<td>75481</td>
<td>1044</td>
<td>0 323</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>62.7 10 26</td>
<td>53K</td>
<td>57K</td>
<td>58761</td>
<td>1421</td>
<td>0.1</td>
<td>267</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**System:** 570 101 218 704K 723K 730630 11K 0.2 2879

- **CPU Performance typical of many Linux Apps:**
  - High Diagnose -> Instruction Simulation -> SIE
  - z/VM 5.2 modifies logic

### Network Storage Impact

**OSA Adapters:**
- Dedicated to Linux servers or virtual routers
- Require approximately 8MB of 2GB real storage per adapter
- z/VM 4.2, 4.3, 4.4 – no qdrop

**z/VM 4.2**
- Use Guest Lan – Reduce OSA Adapters,
- Requires virtual router (VM TCPIP or Linux)

**z/VM 4.4**
- Use Virtual Switch - Eliminate virtual router

**z/VM 5.1**
- Qdrop occurs with dedicated OSA

**Guideline:** Use Virtual Switch, Guest LAN
Mainframe I/O Expectation Issues

Mainframe I/O expectations OFTEN wrong
- I/O traditionally tuned to operate within limitations
- Separate I/O processors
- Competition not limited by ESCON channel speeds

Customer says “FTP on Linux under z/VM is slow”
- Benchmark was large FTP, problem NOT network
- Escon channels, 30ms “CONNECT” time
- 500K transfers

Questions:
- How fast are ESCON channels? FICON channels? Ficon Express?
- How fast are SCSI disks on other platforms?
- PAV?
  - What are options when high utilization on shared disks?
  - PAV Available z/VM 5.2 - Use for high activity shared devices ONLY

FTP Benchmarks: Results NOT intuitive

- Benchmark 1: (G5 processor)
  - FTP through Linux router with OSA dedicated to Virtual Router
  - FTP to Linux on single (ESCON) device
  - Throughput limited to 4mb / second: why?

- Benchmark 2:
  - Eliminate router, dedicate OSA,
  - Throughput increased to 8mb / second, why?
- Guideline: Use dedicated OSA or Virtual Switch

- Benchmark 3:
  - Switch to LVM striped over 2 devices
  - Throughput reduced to 7mb / second, why?
- Guideline: Evaluate carefully use of striped LVM

- Answer to all questions: CPU was limiter
## FTP Benchmarks: Results NOT intuitive

Compare Linux Asynchronous I/O vs synchronous I/O

- Asynchronous is default
- Synchronous writes data without buffering
- DASD response time
  - Asynchronous: 50ms (6 I/O / second, 512k / IO),
  - Synchronous: 1.5ms (300 I/O / second, 4k / IO)

Which is better throughput?

**Guideline:** Use Asynchronous?

DASD Response time rot don’t work

**Guideline:** Fight for FICON/(express)!!!!

---

### Benefit of MDC

<table>
<thead>
<tr>
<th>Time</th>
<th>Actv In</th>
<th>Seq Per</th>
<th>Par Per</th>
<th>Hit</th>
<th>Not Hit</th>
<th>Obj Stls</th>
<th>Delt Avg</th>
<th>MIN</th>
<th>MAX</th>
<th>Obj Stls</th>
<th>Rds Wrts</th>
</tr>
</thead>
<tbody>
<tr>
<td>08/24/03</td>
<td>20</td>
<td>7.9</td>
<td>4.0</td>
<td>48</td>
<td>374</td>
<td>162</td>
<td>178</td>
<td>6.4</td>
<td>0</td>
<td>373</td>
<td>2K</td>
</tr>
<tr>
<td>20:00:00</td>
<td>22</td>
<td>8.1</td>
<td>4.0</td>
<td>47</td>
<td>370</td>
<td>176</td>
<td>178</td>
<td>6.4</td>
<td>0</td>
<td>373</td>
<td>2K</td>
</tr>
<tr>
<td>21:00:00</td>
<td>20</td>
<td>7.7</td>
<td>3.9</td>
<td>48</td>
<td>377</td>
<td>179</td>
<td>180</td>
<td>6.4</td>
<td>0</td>
<td>375</td>
<td>2K</td>
</tr>
</tbody>
</table>

**Summary**

- MINIDISK Cache very good for Linux under z/VM
- As servers get smaller, MDC takes over caching
  - Hit rate (48%) Should be evaluated
  - Use diagnose driver, record cache to best utilize MDC
### MDC Pains

- CP algorithms VERY poor at sizing MDC Storage
- Control the size of MDC!

---

#### Monitor initialized: 02/07/07 at 00:00:05 on 2084 serial 447AA
First record analyzed: 02/07/07 00:00:05

<table>
<thead>
<tr>
<th>Time</th>
<th>Actv In Q</th>
<th>/sec</th>
<th>Pct</th>
<th>rds hits</th>
<th>Max</th>
<th>Min</th>
<th>AID</th>
<th>Avg</th>
<th>MIN</th>
<th>MAX</th>
<th>Obj</th>
<th>Stls</th>
<th>Delt</th>
<th>Avg</th>
<th>MIN</th>
<th>MAX</th>
<th>Obj</th>
<th>Rds</th>
<th>Wrts</th>
<th>Stls</th>
<th>Pages</th>
<th>DASD</th>
</tr>
</thead>
<tbody>
<tr>
<td>12:20:00</td>
<td>26</td>
<td>18.7</td>
<td>2.2</td>
<td>63</td>
<td>33</td>
<td>8K</td>
<td>7.5</td>
<td>0</td>
<td>2K</td>
<td>0</td>
<td>8K</td>
<td>2K</td>
<td>0.1</td>
<td>180</td>
<td>1K</td>
<td>0</td>
<td>3K</td>
<td>1K</td>
<td>55</td>
<td>0</td>
<td>0.1</td>
<td>253</td>
</tr>
<tr>
<td>12:35:00</td>
<td>26</td>
<td>19.1</td>
<td>2.1</td>
<td>63</td>
<td>5.4</td>
<td>10K</td>
<td>5.8</td>
<td>0</td>
<td>2K</td>
<td>0</td>
<td>8K</td>
<td>2K</td>
<td>0</td>
<td>69.9</td>
<td>1K</td>
<td>0</td>
<td>3K</td>
<td>1K</td>
<td>10</td>
<td>0</td>
<td>0.0</td>
<td>53</td>
</tr>
<tr>
<td>12:50:00</td>
<td>26</td>
<td>18.3</td>
<td>2.0</td>
<td>69</td>
<td>6.0</td>
<td>12K</td>
<td>4.7</td>
<td>0</td>
<td>1K</td>
<td>0</td>
<td>8K</td>
<td>2K</td>
<td>0</td>
<td>63.8</td>
<td>1K</td>
<td>0</td>
<td>3K</td>
<td>1K</td>
<td>12</td>
<td>0</td>
<td>0.0</td>
<td>32</td>
</tr>
<tr>
<td>13:05:00</td>
<td>27</td>
<td>19.5</td>
<td>2.2</td>
<td>38</td>
<td>29</td>
<td>11K</td>
<td>5.2</td>
<td>0</td>
<td>4K</td>
<td>0</td>
<td>8K</td>
<td>4K</td>
<td>1.2</td>
<td>1062</td>
<td>1K</td>
<td>0</td>
<td>3K</td>
<td>2K</td>
<td>3</td>
<td>0.0</td>
<td>1.3</td>
<td>571</td>
</tr>
<tr>
<td>13:26:00</td>
<td>31</td>
<td>17.4</td>
<td>1.7</td>
<td>28</td>
<td>28</td>
<td>14K</td>
<td>12</td>
<td>0.7</td>
<td>2K</td>
<td>0</td>
<td>8K</td>
<td>2K</td>
<td>2.9</td>
<td>1324</td>
<td>272</td>
<td>0</td>
<td>3K</td>
<td>2K</td>
<td>3.7</td>
<td>0.0</td>
<td>4.5</td>
<td>1090</td>
</tr>
<tr>
<td>13:41:00</td>
<td>25</td>
<td>19.9</td>
<td>2.9</td>
<td>63</td>
<td>60</td>
<td>41K</td>
<td>7.3</td>
<td>0</td>
<td>2K</td>
<td>0</td>
<td>8K</td>
<td>3K</td>
<td>2.5</td>
<td>483</td>
<td>227</td>
<td>0</td>
<td>3K</td>
<td>2K</td>
<td>2.0</td>
<td>0.2</td>
<td>742</td>
<td>42</td>
</tr>
</tbody>
</table>

**Guidelines:**
- SET MDC STORAGE 0M 128M
- SET MDC XSTORE 0M 0M

---

### z/VM Page Space

Overcommitting real storage is good, reduces cost
- Back up is Paging storage

If 40GB main storage
- Overcommit factor of 2 - How much paging storage needed?
- VM installations often very underconfigured
- **Guideline: Paging storage should still be 2 times requirement**

Number of paging devices? Number of channels?
- ROT not valid

Lack of page space planning is top reason for first installation z/VM outage
**Expanded Storage**

Expanded Storage required for paging performance
- True LRU
- Page the correct pages
- Page rates to disk drop when converting real storage only system to real+expanded

How much expanded?
- Enough for 30 second window
- Enough so STEAL does not page to disk
- 20% usually enough
- Measure on ESABLKP

**Infrastructure impact on CPU**

Performance Management Infrastructure:
- Performance analysis for current service
- Capacity planning for future service
- Accounting and Chargeback
- Operational alerts

Shared resource environment:
- Avoid unnecessary work,
- One agent or many agents?
- Avoid “waking up Linux”

Availability Monitoring – necessary?

Using Encryption - necessary if on virtual lan?

Measure your infrastructure and determine scalability!
Question:
- Why always hit every 15 minutes?

SOP: Standard Operating Procedure

Detect and alert looping processes

**Infrastructure: Alerts**

**Virtual and Total CPU Utilization**

<table>
<thead>
<tr>
<th>Node/Time</th>
<th>&lt;-----Software Name-------&gt;</th>
<th>Program Status</th>
<th>Total Intrval</th>
<th>Pct</th>
<th>StgSize</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:32:00</td>
<td>LINUXA</td>
<td>init</td>
<td>0.9</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>kjournal</td>
<td>2.5</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>db2find</td>
<td>0.3</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ssqd</td>
<td>0.4</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>event</td>
<td>19.5</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>smtpd</td>
<td>193.4</td>
<td>4.2</td>
<td>7.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>admind</td>
<td>58.5</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>server</td>
<td>1.0</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>server</td>
<td>4.3</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>server</td>
<td>44697</td>
<td>58.3</td>
<td>99.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>java</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>java</td>
<td>1.9</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>java</td>
<td>4.9</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>server</td>
<td>19.0</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>java</td>
<td>0.4</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ssqd</td>
<td>0.2</td>
<td>0.0</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Show process by ID
- Status
- Total CPU
- Percent CPU
- Storage
(Non-velocity mib)
**Performance Instrumentation**

Performance Management Instrumentation
- Cost of instrumentation often excessive
- "Native Linux" tools will not detect many problems
- Agents may take 5-10% of a processor
- One agent for availability, one for operational alerts, one for accounting, one for "performance"???

Cost of instrumentation should be < .1% per server
- Performance instrumentation should not change performance

Active agents vs Passive agents
- Active agent wakes up at constant interval and records data
- Passive agent only responds to external request

Dynamically turn off monitoring of idle servers!!!!
- If z/VM data shows server is idle, should agent wake up to find out what is running?
- Logon to a server to tell the agent to stop working?

---

**Agent (SNMP) Overview**

Simple Network Management Protocol
- "Passive agent"
- TCPIP application usually provided by TCPIP Vendor
- Requires operation of the SNMP Daemon

NET-SNMP is SNMP implementation for Linux
- Available on WWW.SourceForge.NET (net-snmp project)
- Supports Linux, Solaris, NT, HP-UX, others…
- Standard on SUSE and REDHAT

TOP, other agents "lie" when under z/VM
- Sample of factor of 10: HTTP://velocitysoftware.com/present/CpuAFS/

The new "monitor interface" does not help
- Misleading record update for CPU data makes problem worse

NETSNMP validated against top
- Processor, Storage; Processes: HTTP://velocitysoftware.com/present/TOPvSNMP/

Possible to prorate Linux process data against z/VM monitor data
**Infrastructure Requirements**

- Performance Management:
  - Performance analysis – z/VM AND Linux
  - Capacity Planning for Linux applications and consolidation
  - Chargeback and Accounting
  - Operational alerts

- Cost of infrastructure
  - Agent needs to support all servers
  - Operational cost should be .1%

- Systems Programming effort
  - Must be small
  - Don't waste your time building infrastructure

Use other's experience. Recent IBM zEXPO had 5 user experiences
- Nationwide, NCCI, Virginia Supreme Court, Mercedes, Transzap
- ALL of them said they use ESALPS or said they would (and have) installed.

---

**Linux Configuration Summary**

- Virtual machine size
  - Minimize until some swap

- Swapping
  - Swap to virtual disk
  - Define 2 virtual disks,
    - One to meet the average requirement
    - Second one for overflow
  - Use DIAG driver instead of FBA
    - Reduces I/O by factor of 8

- Virtual processors
  - Minimize to meet the workload/application requirement

- Infrastructure costs
  - Minimize – shared resource architecture
**z/VM Subsystem Configuration**

**DASD Channels**
- ESCON channels are 17MByte / second
- Ficon channels 100MB, Ficon Express 200MB
- Ficon compares to SCSI disks on other platforms

**Paging/Spooling**
- How much spooling is required to dump **40GB server**?
- How much paging is required to support 2 times over commitment of 40GB z/VM system?

**MDC**
- Caches data – read-ahead, often used data
- Does not require 2GB storage for I/O
- Default too high
- SET MDC STORAGE 0M 128M
- SET MDC XSTORE 0M 0M

---

**z/VM Expanded Storage**

Expanded storage to z/VM is like vdisk swap for Linux

2GB issues results in paging even when high storage available
- Expanded storage is very necessary for paging hierarchy

**Expanded Storage Requirement:**
- 25% if using “small infrastructure servers”
- More if using “large applications servers”:

If paging to disk, and available list high,
- convert more real storage to expanded storage
  - Evaluate real storage peak period “available list”
  - convert to expanded storage

**SET MDC XSTORE 0 0**
- (MDC in expanded storage has little value)
**z/VM System Parameters**

**SET SHARE**
- Use RELATIVE 100 for single virtual CPU
- Use RELATIVE 200 for two virtual CPU

**SET SRM STORBUF – allow overcommit**
- SET SRM STORBUF 300 300 300
- SET SRM LDUBUF 100 80 60

**SET QUICKDSP**
- Use for only absolutely critical servers

---

**Configuration Summary**

Configure z/VM to support Linux
Configure Linux to share
Manage performance with scaleable infrastructure