Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.
Enterprise Storage Server
ESCON*
FICON
FICON Express
HiperSockets
IBM*
IBM logo*
IBM eServer
Netfinity*
S/390*
VM/ESA*
WebSphere*
z/VM
zSeries
* Registered trademarks of IBM Corporation
The following are trademarks or registered trademarks of other companies.
Intel is a trademark of the Intel Corporation in the United States and other countries.
Java and all Java-related trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc., in the United States and other countries.
Lotus, Notes, and Domino are trademarks or registered trademarks of Lotus Development Corporation.
Linux is a registered trademark of Linus Torvalds.
Microsoft, Windows and Windows NT are registered trademarks of Microsoft Corporation.
Penguin (Tux) compliments of Larry Ewing.
SET and Secure Electronic Transaction are trademarks owned by SET Secure Electronic Transaction LLC.
UNIX is a registered trademark of The Open Group in the United States and other countries.
* All other products may be trademarks or registered trademarks of their respective companies.
Performance Measuring and Tuning

- Resource profiles
- Linux and z/VM
- Tools
- Storage
- Networking
- Future enhancements
Tracking down performance problems

Questions to be answered

- What do I expect?
- Do I have numbers for comparison?
- Do I really have a problem?
- Where do I suspect the problem?
- What data do I want to collect?
- Which tools can I use?
- What do the numbers tell me?
- What measures evolve from the numbers?
resource profiles

- Know your applications resource profile
- Know your systems resource profile

What your application needs
What your system provides

Does it really match?
know your setup...

1. know your applications  
2. know your environment (hardware, software, network)  
3. know your resource limits (memory, cpu, IO, shares)  
4. use tools to identify or ask your administrator  
5. know your Linux
Linux and z/VM
If your Linux is a VM guest have in mind:

- Linux tools do only see the share they got in VM
  - CPU might only be a part of a \textit{physical} CPU
- use VM tools to monitor performance for your guests

Sharing of one physical CPU
on-demand timer patch

- Linux image gets external timer interrupt (tick) every 10 ms
- several idle Linux images running in z/VM cause significant load
- solution: use larger timer intervals on idle Linux guests
- timer behaviour may be changed during run time by writing a '0' or a '1' to /proc/sys/kernel/hz_timer

but:

- older versions of the on-demand timer patch could lead to a misaccounting of Linux CPU load for certain workloads (high volume network traffic)
Linux and z/VM : VM monitor data

- In case you want to report a performance problem with z/Linux running as a VM guest, you may be asked to collect monitor data:
- data set with information about CPU load, storage, IO and guest activity
- prepare your VM system to collect monitor data. For a detailed description on how to do that, see [http://www.vm.ibm.com/perf/tips/collect.html](http://www.vm.ibm.com/perf/tips/collect.html)
- be sure to enable collection for all domains
- collect the data
- send the **compressed** file to us (eMail or ftp)
Linux and z/VM: monitor data collection

- logon to userid

  cp monitor start
  monwrite MONDCSS *monitor DISK mon dat a
  monwstop
  cp monitor stop

- send result file mon dat a to user FCONX for analysis
Tools
A wide range of tools is available:

- System Tools (top, vmstat)
- rmfpms
- sysstat Package (collect a large set of Linux system data)
- VM Tools (FCON, Monitor...) discussed in later sessions
- OSA SNMP (capture OSA Express card data)
- Kernel Profiler (set of facilities for profiling the Linux kernel)
- Lockmeter (capture spin lock activity)
system tools: vmstat

- most important Linux system data at a glance
- low performance impact
- no setup necessary

- example:

```
[wolf@wolf wolf]$ vmstat 3
procs     memory     swap     io     system
       .cpu
  r  b  w  swpd  free  buff  cache  si  so  bi  bo  in  cs  us  sy
 id
  0  0  0  0 663248  24204 196816  0  0  66  32  582  663  4  1
  95
  0  0  0  0 663248  24220 196816  0  0  0  56  555  493  2  1
  97
  0  0  0  0 663248  24220 196816  0  0  0  0  548  482  1  0
  99
  0  0  0  0 663248  24228 196816  0  0  0  4  549  493  1  1
  98
```
system tools: top

- even more system data at a glance
- data on a per-process basis
- high performance impact
- no setup necessary

example:

[wolf@wolf wolf]$ top
1:40pm up 3:17, 11 users, load average: 0.00, 0.02, 0.03
134 processes: 127 sleeping, 6 running, 1 zombie, 0 stopped
CPU states: 1.9% user, 1.7% system, 0.0% nice, 96.2% idle
Mem: 1030464K av, 675028K used, 355436K free, 0K shrd, 156256K buff
Swap: 1663160K av, 0K used, 1663160K free 296068K cached

PID USER     PRI  NI  SIZE  RSS SHARE STAT %CPU %MEM   TIME COMMAND
1746 wolf      15   0 54772 14M 11600 R   1.1  1.4   0:00 kdeinit
6147 wolf      15   0 1080 1080   840 R   0.9  0.1   0:00 top
rmfpms

- long term data gathering
- xml over http interface
- independent from z/os; with z/os, you can also have an ldap interface to linux performance data
- modular architecture
- low performance impact
- see http://www.ibm.com/servers/eserver/zseries/zos/rmf/ rmfhtmls/pmweb/pmlin.htm for more info
sysstat Package

- collection of Linux tools to collect system data
- available as open source package at [http://perso.wanadoo.fr/sebastien.godard/](http://perso.wanadoo.fr/sebastien.godard/)
- on Linux for zSeries only recompile needed
- latest stable version is 5.0.1
- contains multiple components:

  - sadc: Data gatherer
    - stores data in binary file
  - sar: reporting tool
    - reads binary file and converts it to readable output
  - mpstat: Processor utilization
  - iostat: IO utilization
Storage
DASD statistics

- contained in kernel since SUSE SLES8
- statistics collected by dasd driver
- can be easily switched on/off in proc filesystem

```
echo set on > /proc/dasd/statistics
echo set off > /proc/dasd/statistics
```

- setting off and back on resets all counters
DASD statistics : example

/proc/dasd/statistics – Example

cat /proc/dasd/statistics

56881 dasd I/O requests
with 5270816 sectors (512B each)

Histogram of sizes (512B secs)

Histogram of I/O times (microseconds)

Histogram of I/O times per sector

Histogram of I/O time till ssch

Histogram of I/O time between ssch and irq

Histogram of I/O time between ssch and irq per sector

Histogram of I/O time between irq and end

# of req in chanq at enqueuing (1..32)
DASD statistics : the details

- collects statistics (mostly processing times) of IO operations
- each line represents a histogram of times for a certain operation
- operations split up into the following:

  - Start: Build channel program
  - Wait till subchannel is free
  - End: Tell block dev layer
  - Data has arrived

  - Histogram of I/O till ssch
  - Histogram of I/O between ssch and IRQ
  - Histogram between I/O and End
  - Histogram of I/O times
ext2, 8 Processes
Histogram of I/O times (microseconds)
ext3, 16 Processes
Histogram of I/O time before SSCH (IOSQ)
Ext3, 16 Processes
Histogram of I/O time between SSCH and IRQ
Ext3, 16 Processes
number of requests in subchannel-queue at enqueuing
Logical Volume Manager (LVM)

- Linux software raid with raid levels 0, 1, 4 and 5
- excellent performance
- excellent flexibility (resizing, adding/removing disks)
- available in SLES7, SLES8, and RedHat RHEL 3
- on zSeries, support multipath and PAV (under z/VM)
- [http://www.sistina.com/products_lvm.htm](http://www.sistina.com/products_lvm.htm)
LVM system structure

- (journaled) file system
- Raw Logical Volume
- logical volume
- logical volume
- volume group
- physical volume
- physical volume
- physical volume
- block device driver
- RAID adapter
- RAID array
- physical disk
- physical disk
Improving disk performance with LVM

With LVM and striping parallelism is achieved.
ESS Architecture

Scenario: four CHPIDs

- **FCP Switch 2109**
- **FCP CHPID**
- **FCP CHPID**
- **FCP CHPID**
- **z900 2064**

- **FCP CHPID**
- **FCP CHPID**
- **FCP CHPID**
- **FCP CHPID**

- **Cluster Processor Complex** - 4 way SMP RISC system

- **Host Adapter (HA) supporting FCP (FCP port)**
  - 16 Host Adapters, organized in 4 bays, 4 ports each

- **Device Adapter Pairs (DA)**
  - each one supports two loops

- **Disks are organized in ranks**
  - each rank (8 physical disks) implements one RAID 5 array (with logical disks)
ESS setup rules

- spread your accesses over as much chpids as possible
- use as much host adapter bays as possible
- spread disks equally over as much ranks as possible

⇒ maximize parallel access to disks
Networking
OSA SNMP

- provides means to readout a lot of useful information from an OSA Express Card
- distributed as part of s390-tools
- grab the latest MIB from www.ibm.com/servers/resourcelink (needs resourcelink sign-in)
- open Library → Open Systems Adapter (OSA) Library → OSA-Express Direct SNMP MIB module
OSA SNMP : example

root@g73vm8:~# cat /proc/qeth

<table>
<thead>
<tr>
<th>devnos (hex)</th>
<th>CHPID</th>
<th>device</th>
<th>cardtype</th>
<th>port</th>
<th>chksum</th>
<th>prio-q'ing</th>
<th>rtr</th>
<th>fsz</th>
<th>C</th>
<th>cnt</th>
</tr>
</thead>
<tbody>
<tr>
<td>F118/F119/F11A</td>
<td>x6A</td>
<td>eth0</td>
<td>OSD_100</td>
<td>0</td>
<td>no</td>
<td>always_q_2</td>
<td>no</td>
<td>64k</td>
<td>128</td>
<td></td>
</tr>
<tr>
<td>FA0C/FA0D/FA0E</td>
<td>x7A</td>
<td>eth1</td>
<td>OSD_1000</td>
<td>0</td>
<td>no</td>
<td>always_q_2</td>
<td>no</td>
<td>64k</td>
<td>128</td>
<td></td>
</tr>
<tr>
<td>F006/F007/F008</td>
<td>x7C</td>
<td>eth2</td>
<td>OSD_1000</td>
<td>0</td>
<td>no</td>
<td>always_q_2</td>
<td>no</td>
<td>64k</td>
<td>128</td>
<td></td>
</tr>
<tr>
<td>8209/820A/820B</td>
<td>xFD</td>
<td>hsi5</td>
<td>HiperSockets</td>
<td>0</td>
<td>no</td>
<td>always_q_2</td>
<td>no</td>
<td>40k</td>
<td>128</td>
<td></td>
</tr>
<tr>
<td>7000/7001/7002</td>
<td>x03</td>
<td>hsi12</td>
<td>GuestLAN</td>
<td>Hiper</td>
<td>0</td>
<td>no always_q_2</td>
<td>no</td>
<td>40k</td>
<td>128</td>
<td></td>
</tr>
<tr>
<td>9000/9001/9002</td>
<td>x05</td>
<td>eth16</td>
<td>GuestLAN</td>
<td>QDIO</td>
<td>0</td>
<td>no always_q_2</td>
<td>no</td>
<td>64k</td>
<td>128</td>
<td></td>
</tr>
</tbody>
</table>

root@g73vm8:~# /usr/bin/snmpwalk -Os localhost private ibmOSAExpChannelNumber

ibmOSAExpChannelNumber.6 = Hex: 00 7C
ibmOSAExpChannelNumber.7 = Hex: 00 7A
ibmOSAExpChannelNumber.8 = Hex: 00 6A
OSA SNMP : example

root@g73vm8:~# /usr/bin/snmpwalk -Os localhost private ibmOSAExpChannelPCIBusUtil1Min
ibmOSAExpChannelPCIBusUtil1Min.6 = 7
ibmOSAExpChannelPCIBusUtil1Min.7 = 7
ibmOSAExpChannelPCIBusUtil1Min.8 = 12

root@g73vm8:~# /usr/bin/snmpwalk -Os localhost private ibmOSAExpChannelProcUtil1Min
ibmOSAExpChannelProcUtil1Min.6 = 2
ibmOSAExpChannelProcUtil1Min.7 = 1
ibmOSAExpChannelProcUtil1Min.8 = 2
OSA SNMP: example

```
root@g73vm8:~# /usr/bin/snmpwalk -Os localhost private ibmOSAExpChannelProcUtil1Min
ibmOSAExpChannelProcUtil1Min.6 = 2
ibmOSAExpChannelProcUtil1Min.7 = 1
ibmOSAExpChannelProcUtil1Min.8 = 20

root@g73vm8:~# /usr/bin/snmpwalk -Os localhost private ibmOSAExpChannelPCIBusUtil1Min
ibmOSAExpChannelPCIBusUtil1Min.6 = 7
ibmOSAExpChannelPCIBusUtil1Min.7 = 7
ibmOSAExpChannelPCIBusUtil1Min.8 = 21
```
Application performance
Adjust Java Performance

• Check if Just In Timer Compiler is enabled
  JIT is not enabled if compat library is not available or Environment variable java_compiler is set to none

  `java -version`

  java version "1.4.0"
  Java(TM) 2 Runtime Environment, Standard Edition (build 1.4.0)
  Classic VM (build 1.4.0, J2RE 1.4.0 IBM build cxia32140-20020917a (JIT enabled: jitc))

• Check if garbage collector adjustments could improve your performance

  `java -Xgcpolicy:optthruput <java class>`
  `java -Xgcpolicy:optavgpause <java class>`

• To monitor garbage collector (show statistics)

  `java -verbosegc <java class>`
New features

- Channel measurement blocks
- z/VM monitor stream stage
  - Linux guest exports performance data into “APPLDATA monitor records”
  - performance data may be collected or display by z/VM performance monitoring tools
- virtual CPU timers
web resources

• Linux on zSeries Performance website

• z/VM Performance website
  http://www.vm.ibm.com/perf

• Linux on zSeries Performance redbook
Questions?