

Linux on z/VM Performance Analysis

- Barton@VelocitySoftware.com
- [HTTP://VelocitySoftware.com](http://VelocitySoftware.com)

“If you can’t Measure it,
I am Just Not Interested™”

Configuring z/VM for Linux on zSeries

- Must configure z/VM – many defaults incorrect **or out of date**
- Linux must be configured for a shared resource environment
- Many actions not intuitive
- “Best Practices”

Infrastructure unknowns for “new” installations

- How to manage performance / capacity planning?
- Is chargeback important?
- Operational support for 1,000 servers?
- What are the limits of a configuration and how to measure it
- How to share resources to INCREASE the Return On Investment (ROI)

Measurement and Tuning for z/VM IS Required

- Start with Proper Configurations

Linux Requirements Summary

Linux (and networks) adds requirement

- **Correct data**
- **Complete data**
- **Low cost data**

Support requirements:

- z/VM 3.x, 4.x, 5.x, 6.x, 7.1, 7.2, 7.3....
- SLES 7, 8, 9, 10, 11 (Installations still have 7 and 8)
- RHEL 3, 4, 5, 6
- Other distributions
- Other platforms (VSE, VMWare, SUN, P, **MicroSoft**)

Must support:

- Performance tuning
- Capacity planning
- Operational alerts
- Chargeback/Accounting

Correct Linux Performance Data?

Valid and Correct?

- **Process data from Linux under z/VM is wrong**
- All process accounting based on timer ticks
- Corrected in SLES10, RHEL5 (“steal time” underreports)
- TOP, ALL other agents “lie” when under z/VM
- Sample of factor of 10 prior to SLES10
- Well known issue since 2001
- SMT adds to the problem

Leads to solving performance problems?

- z/VM owns the shared resources
- **“Native” tools will not detect many problems**
- **“Performance was unexplainably bad so we abandoned the project”**
- Skills, experience and education help...

Operational cost of agents

- Does your agent use 2%? 5%? 95%? of a processor per Linux server? (splunk for example)
- Does this matter on distributed servers where agents were created?
- Will local data collection fill up your file system?
- Does turning off performance monitoring solve the performance problem?
- Do you only turn on your agent when you have a problem???

Customer quote: “An agent that costs 1% of a processor will cost me 10 IFLs”

Agents must provide correct data

- Is your data correct? Or wrong by order of magnitude?
- Prior to SLES10/RHEL5, all “Virtual” agents provide wrong data
- **Why collect bad data?**

Network - Linux Instrumentation

Performance Data infrastructure existed (ESAMON/ESAMAP)

- Performance Data Base (PDB) already existed for performance analysis and Capacity Planning
- Data presentation tools existed

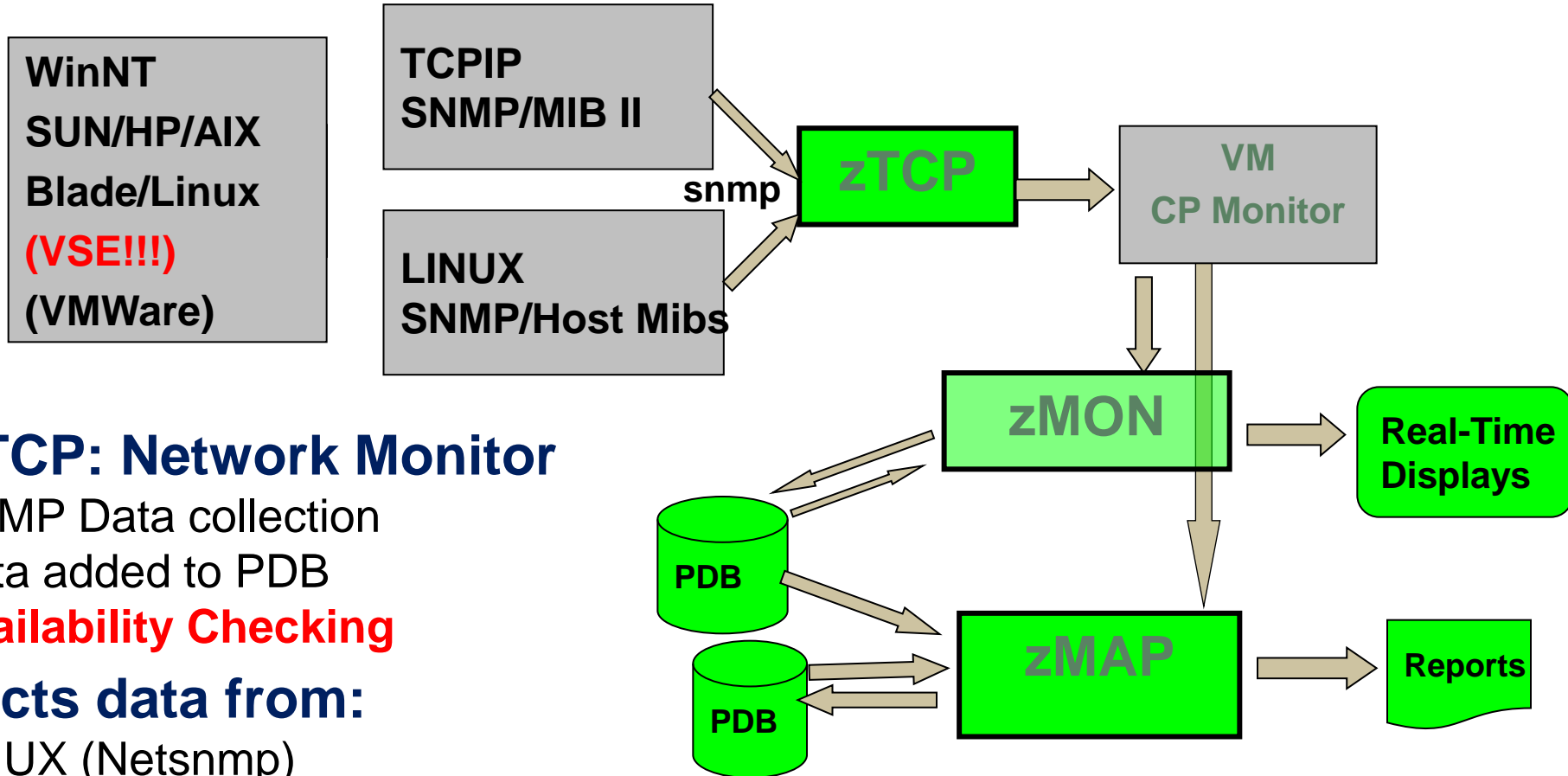
Data source needed for Linux and Network:

- **Low overhead (want to monitor 100 / 1000 servers under z/VM?)**
- **Agents developed for Intel and Distributed servers did not care about overhead**
- Standard interface

SNMP: Standard interface

- TCPIP application provided by TCPIP Vendor
- Used to collect network, host data from NT, SUN, HP
- NETSNMP available for Linux - Meets all requirements
- (Distributed with RHEL 3,4,5 SLES 7,8,9,10,11)
- **Platform independent (Intel, P-series, Microsoft, Linux)**
- **.1% of ONE IFL (z10) per server, ONE MINUTE COLLECTION**

Linux and Network Data Acquisition



ESATCP: Network Monitor

- SNMP Data collection
- Data added to PDB
- **Availability Checking**

Collects data from:

- LINUX (Netsnmp)
- NT/SUN/HP (Native SNMP)
- Printers/Routers....

Public MIBs:

- MIB-II (Network MIB, ESATCPx)
- Host MIB (Server data, ESAHSTx)
- UCD MIB (Linux specific, ESAUCDx)
- GPFS (IBM)
- Velocity Software
- Linux (ESALNXx)
- Oracle (ESAORAx)
- WAS (ESAJVMx)
- Openshift (ESAK8Sx, ESADOCKx)

Operational Costs

Low cost agent - Cost of SNMPD very low (.03%-.1%)

(Objective - Determine what process spikes at 1am Monday morning)

See "<http://velocitysoftware.com/applic.html>" for full listing (24 Linux servers)

```
Report: ESALNXA          LINUX HOST Application Report
-----
Node/   Process/   ID   <---Processor Percent--->
Date   Application
Time   name      Total sys  user  syst  usrt
-----
00:15:57
LINUX16 *Totals*   0   16.9  2.5  11.6  1.9  1.1
        amqpcsea  674  0.4  0.1  0.3   0   0
        amqzxa0 600  0.8  0.1  0.7  0.0  0.0
        cron   473  2.1  0.2  0.2  1.7  0.0
        dsmc   938  0.1  0.0  0.0  0.0  0.0
        httpd 31993 2.8  0.2  2.5  0.0  0.1
        java  32066 8.0  1.3  6.7   0   0
        kjournal 85  0.1  0.1  0   0   0
        kswapd 6  0.1  0.1  0   0   0
        qpea  4642 0.1  0.0  0.1  0   0
        qpmon  4674 0.8  0.1  0.7  0.0  0
        snmpd  361  0.1  0.1  0.0  0   0 <-----
        sshd   370  1.0  0.0  0   0.1  0.9
LINUX13 *Totals*   0   2.7  0.8  0.3  0.6  1.0
        cron   421  1.2  0.0  0.0  0.5  0.7
        init    1  0.2  0.0  0.0  0.0  0.1
        master 394  0.3  0.0  0.1  0.0  0.1
        ntpd   453  0.8  0.6  0.2   0   0
LINUX15 *Totals*   0   1.8  0.3  0.5  1.1  0.0
        amqzxa0 844  0.2  0.0  0.1  0   0
        cron   457  1.1  0.0  0.0  1.1  0.0
        qpmon  4726 0.1  0.0  0.1  0   0
        snmpd  354  0.4  0.2  0.2   0   0 <-----
```

Process Capture Ratio

High CPU capture ratio:

```
Report: ESALNXV          LINUX Virtual Processor Analysis Report
-----
```

Node/ Name	VM ServerID	<Linux Pct CPU> Total	<Process Data> Syst	<Process Data> User	<Process Data> Total	<Process Data> Syst	<Process Data> User	Capture Ratio	Prorate Factor
10:03:00									
NEALE1	LNEALE1	100.0	11.4	88.6	100.2	11.5	88.7	1.002	1.000

```
Report: ESALNXP          LINUX HOST Process Statistics Report
-----
```

node/ Name	<-Process ID	Ident-> PPID	GRP	Nice Valu	<-----CPU Tot	Percents-- sys	user	syst	usrt
10:03:00									
NEALE1	0	0	0	0	100	0.43	3.35	11.0	85.4
kswapd0	100	1	1	0	0.12	0.12	0	0	0
snmpd	1013	1	1012	-10	0.13	0.03	0.10	0	0
sh	3653	3652	30124	0	52.7	0	0	9.37	43.3
gmake	9751	9750	30124	0	43.4	0.02	0.02	1.37	42.0
sh	10129	9751	30124	0	0.02	0.02	0	0	0
sh	10130	10129	30124	0	0.63	0.03	0.23	0.28	0.08
cc1	10307	10306	30124	0	3.12	0.18	2.93	0	0
rpmbuild	30124	16382	30124	0	0.07	0.03	0.03	0	0
sh	30125	30124	30124	0	0.02	0	0.02	0	0
gmake	30126	30125	30124	0	0.02	0	0.02	0	0

```
Report: ESALNXC          LINUX Process Conf
-----
```

Node/ Name	<-Process ID	Ident-> PPID	GRP	<-----Pr Path
NEALE1				
init	1	0	0	init [3]
migratio	2	1	0	migratio
ksoftirq	3	1	0	ksoftirq
events/0	4	1	0	events/0
khelper	5	4	0	khelper
kblockd/	6	4	0	kblockd/
cio	41	4	0	cio
cio_noti	42	4	0	cio_noti
kslowcrw	43	4	0	kslowcrw
apldata	96	4	0	apldata
aio/0	101	4	0	aio/0
pdflush	5266	4	0	pdflush
pdflush	26647	4	0	pdflush
kswapd0	100	1	1	kswapd0
kmcheck	158	1	1	kmcheck
syslogd	976	1	976	/sbin/sy
klogd	979	1	979	/sbin/kl
snmpd	1013	1	1012	snmpd
portmap	1030	1	1030	/sbin/po
rpciod	1034	1	1	rpciod
lockd	1035	1	1	lockd
sshd	1072	1	1072	/usr/sbi
sshd	16272	1072	16272	sshd: bu
sshd	16288	1072	16288	sshd: bu
sshd	16290	16288	16288	sshd: bu
bash	16291	16290	16291	bash
python	16312	16291	16291	python
do-bui	16313	16312	16291	/bin/sh
bb_do	16382	16313	16291	/usr/bin
rpmb	16415	16382	16415	rpmbuild
rpmb	30124	16382	30124	rpmbuild

Distributed Systems Analysis Windows NT

```
Screen: ESAHST1  NT Data                      ESAMON V3.2  07/30 14:56-14:57
1 of 1  LINUX HOST Software Analysis Report    NODE * LIMIT 500      2066 11FF6
      <--Software Program-----> <CPU Seconds> CPU   Storage(K)
Time      Node      Name      ID  Type  Status  Total Intrval Pct   Current
-----
14:57:00 ENTWDB  NetTime.  2648   4    1   4259   0.68  1.12   1320
          NetTime.  2452   4    1   982    0.57  0.94   1040
          sqlagent  2408   4    1   100    0.03  0.05   3724
          snmp.exe  2268   4    1    73    0.07  0.12   3888
          taskmgr. 2224   4    1 21076   0.28  0.46   2524
          sqlservr 2136   4    1 50038   9.53 15.72  511624
          NetTime. 1808   4    1 10481   1.47  2.42   1092
          sqlmangr 1660   4    1   189    0.01  0.02   3664
          DLLHOST. 1648   4    1   102    0.02  0.03   4684
          liccheck 1352   4    1  1272   0.04  0.07   1584
          DLLHOST. 1284   4    1  2158   0.09  0.15   6660
          inetinfo 1208   4    1  3063   0.10  0.16   9708
          WinVNC.e 1160   4    1 20742   0.56  0.92   3536
          explorer  788   4    1  2252   0.14  0.23   5336
          SERVICES  272   4    1  6892   1.50  2.47   7480
          msdtc.ex  164   4    1    71    0.02  0.03   5108
```

Linux Infrastructure Requirements

Performance Management Requirements

- Performance Analysis
- Operational Alerts
- Capacity Planning
- Accounting/Charge back

Correct data (Virtual Linux CPU data wrong)

Capture ratios (Is the data valid?)

Instrumentation can NOT be the performance problem

Analyzing Linux Disks

Report: **ESAHST2** LINUX HOST Storage Analysis Report
 Monitor initialized: 02/05/07 at 10:41:41 on 2084 serial 55BAF

```

-----
NODE/          <-Utilization->          <-----Storage----->
Time/          <MegaByte> Pct          Alloc
Date    Index  Size  Used  Full  Errors  Units  Description
-----
10:43:00
acme
      1    495  14.2  2.9      0    1024  Memory Buffers
      2    495   487 98.4      0    1024  Real Memory
      3   2031  12.8  0.6      0    1024  Swap Space
      4   2310   775 33.6      0    4096  /
      6   2310  1293 56.0      0    4096  /usr

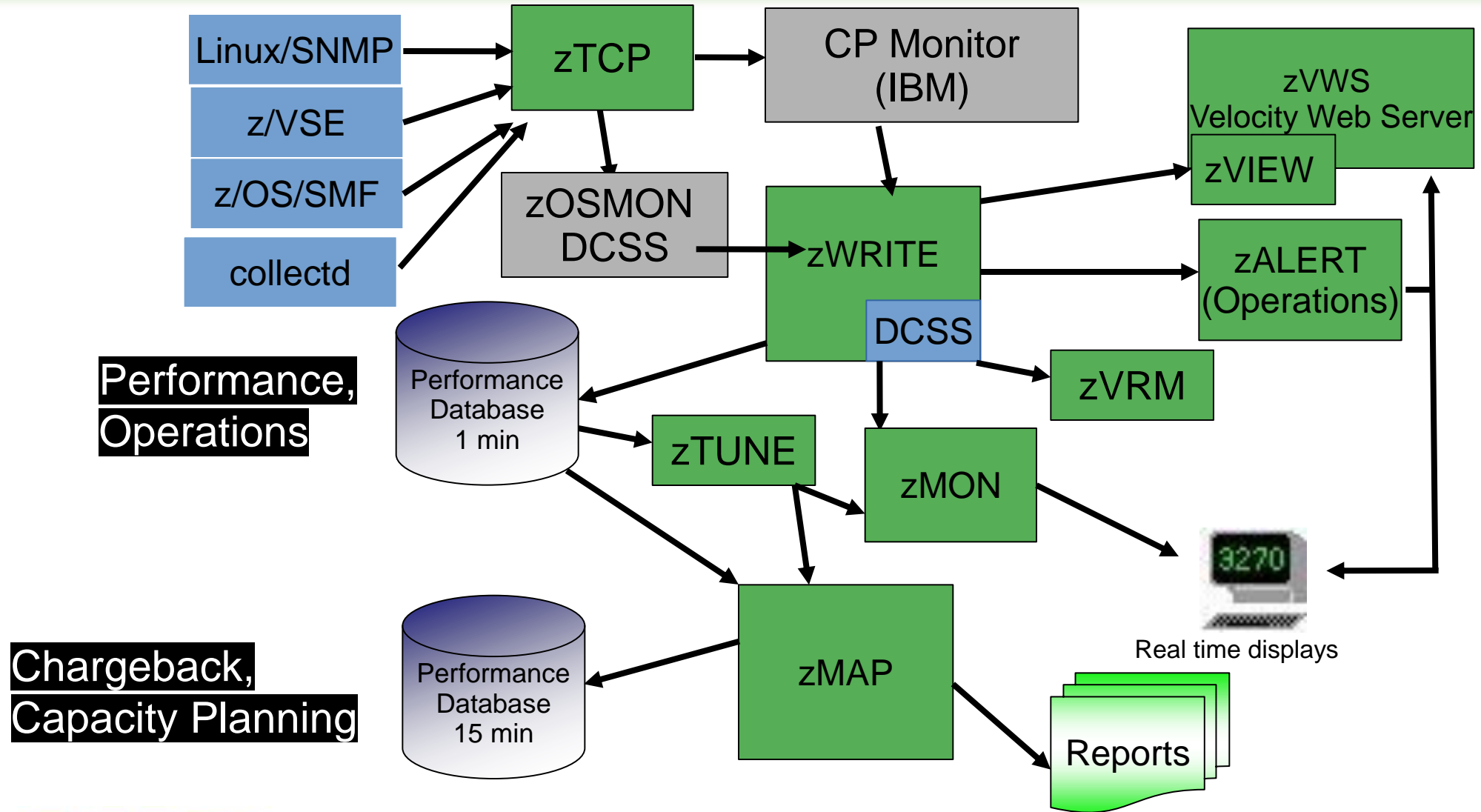
dominoz1
      1   2002  38.5  1.9      0    1024  Memory Buffers
      2   2002  1994 100      0    1024  Real Memory
      3   2031  97.4  4.8      0    1024  Swap Space
      4   2310  1556 67.4      0    4096  /
      6   2310  1398 60.5      0    4096  /usr
      7  984K   238K 24.2      0    4096  /notesdata

ebiz1
      1    997   9.0  0.9      0    1024  Memory Buffers
      2    997   992 99.5      0    1024  Real Memory
      3   2031   514 25.3      0    1024  Swap Space
      4   2310  1607 69.6      0    4096  /
      6   2310  1451 62.8      0    4096  /usr
      7  101K   10K 10.3      0    4096  /notesdata
  
```

HOST MIB data:

- Provides disk data
- Percent full
- Supports WinNT, Unix
- Alerts by disk full

zVPS Infrastructure



**Performance,
Operations**

**Chargeback,
Capacity Planning**