Large Virtualization Case Studies

(What happens when the honey moon is over?)

(suse 2014)

Barton Robinson

Velocity Software

Barton@VelocitySoftware.com

VELOCITY SOFTWARE

PROVEN PERFORMANCE

Planning for 1000 virtual servers

Objectives

- **Capacity Considerations**
- **Profiling possibilities?**
- **Case Studies**
  - What successful installations are doing
  - How installations “save boatloads of money”
- **Capacity Planning for:**
  - Consolidation
  - Workload growth
- **LPAR Configurations**
- **Storage ROTs (WAS, Oracle, SAP)**

What is large virtualization?

Resource Sharing

- Learn to share (hard for my kids when they were 2)
- Performance matters

Metrics – Processor overcommit

- Servers per “box” (100-1000)
- Servers per LPAR (80-200)
- Servers per IFL (5-20
- VCPU per IFL (10-40)

Metrics – Storage overcommit

People comittment

- Systems support per server (one per 200-800 servers)

Capacity Planning Processor Overview

Processor requirements

- CECs (DR, floor space)
- IFLs (CPU capacity)
- LPARs (Storage capacity, separation)

Considerations for a better TCO

- Software is licensed per core / IFL
- 95% IFL (Effective) utilization provides the lowest cost solution
- One installation replaced 30 “oracle servers” with one IFL
- One installation gets hardware & system software for “free”

Plan for:

- CPU Capacity
  - Know your target utilization (Ghz is Ghz, Mips is meaningless)
  - Know target capacity requirement
- Number of LPARs -

Capacity Planning Processor Considerations

Term: Processor Overcommit, higher is better

- Number of virtual cpus per IFL

Critical concepts (Objective is meeting SLA)

- z/VM on z196, z9, z10 has a LOW MP effect
- One CEC with 2 IFLs has MORE capacity than 2 CECs with one IFL
- One IFL runs 40-50%, 2 IFLs run 50-80%, 20 IFLs run 95%
- 95% IFL utilization lowest cost (TCO)
- An IFL over 3 years costs about $100,000
- Two IFLs at 30% cost $100,000 more than ONE IFL at 60%.

- **Processors at low utilization are MORE expensive**

Capacity Planning – The Job (quick digression)

Capacity Planning is an art and an OLD profession

- Doing it well saves the company money (high utilization)
- Not doing it saves the company money (people, tools)

Capacity Planning objective – High utilization

- Ensure enough resources (storage) are available
- Not a lot of capacity planning on distributed side

When is capacity planning justified?

- How much can be saved if targeting 1,000 servers?
- 10 IFLs – 10% savings is $100K
- 100 IFLs - ….

Bank in South Africa presentation – "Open Systems"

- Outages are not necessary very visible, impact is minimal
- Capacity is as a rule under utilized
- Resources not shared
- **"Monitoring tools only for recreational purposes"**

What is different under z/VM?

- Resources shared
- Resources utilized to the maximum

Status:

- 12 IFLs, 240 guests, 420 virtual CPUs
- 35 VCPU per IFL
- "http://www.velocitysoftware.com/present/zvpsnedb.pdf"

Configuration Topics - Processors

Higher processor utilization provides better TCO

Sharing processors with virtualization

- Multiple LPARs sharing IFLs
- Multiple servers within LPAR
- Multiple virtual CPUs within virtual server

Capacity planning questions:

What level of sharing – Linux virtual cpus per IFL

What is associated overhead?

What is the workload requirements?

Capacity Planning Storage Overview

Storage Consideration (to keep ifls at 95% busy)

- How much storage is required? (10-15 gb / IFL?)
- What storage tuning should / can be performed?
- What level of paging will be supported?

Storage requirements inputs

- Target Overcommit level (1:1, 2:1, 4:1?)
- (VDISK NOT PART OF OVERCOMMIT)
- Storage maximums (250GB per LPAR as of z/VM 6.2)
- Expanded Storage (20%)

Server Consolidation Planning

Replacements? (Distributed server to z server)

- 1 to 1, 2 to 1, 1 to 2 ? 1 to 10?

Processor sizing

- Gigahertz is gigahertz
- “Barton’s number”: 1 mip is 4-5 megahertz
- Z196: 5.0 Ghz
- EC12: 5.4Ghz (BC12 – 4.0 Ghz)

Server Storage sizing

- Smaller is better, tuning easier, managing easier
- Cost of extra servers small

Linux Internal overhead (mp effect)

- 5-10% reduction going from 2 to 1 vcpus

Common in large successful installations:

If I can't manage it, it is not going to happen

Management Infrastructure in place (ZVPS)

Infrastructure Requirements

- Performance Management
- Capacity Planning Requirements
  - Analysis by server, by application, by user
- Operations, Alerts
- Chargeback, Accounting

Infrastructure resource consumption serious planning issue and obstacle to scalability

Costs for 1,000 Servers:

- A 2% agent requires 20 IFLs just for management
- A .03% agent requires 30% of one IFL
- (Cost of 1,000 2% agents: 20 IFLs: $2M)

Ask the right questions!

- Data correct?
- Capture ratio?
- Cost of infrastructure?
- References…..

Performance Management Planning

```
Report: ESALNXP      LINUX HOST Process Statistics Report
Monitor initialized: 21/01/11 at 07:03:00 on
-------------------------------------------------------------
node/     <-Process Ident-> Nice <------CPU Percents---->
 Name      ID    PPID   GRP  Valu  Tot  sys user syst usrt
--------- ----- ----- ----- ---- ---- ---- ---- ---- ----
 snmpd     2706     1  2705  -10 0.07 0.02 0.05    0    0
 snmpd    24382     1 24381  -10 0.04 0.02 0.02    0    0
 snmpd     2350     1  2349  -10 0.04 0.02 0.02    0    0
 snmpd    28384     1 28383  -10 0.14 0.10 0.04    0    0
 snmpd    28794     1 28793  -10 0.09 0.09    0    0    0
 snmpd    31552     1 31551  -10 0.07 0.03 0.03    0    0
 snmpd    11606     1 11605  -10 0.04 0.02 0.02    0    0
 snmpd     2996     1  2995  -10 0.08 0.03 0.05    0    0
 snmpd    31589     1 31588  -10 0.05 0.03 0.02    0    0
 snmpd    15356     1 15355  -10 0.16    0 0.16    0    0
 snmpd    15413     1 15412  -10 0.10 0.08 0.02    0    0
 snmpd    30795     1 30794  -10 0.05    0 0.05    0    0
 snmpd     1339     1  1338  -10 0.05 0.04 0.02    0    0
 snmpd    30724     1 30723  -10 0.02 0.02    0    0    0
 snmpd    28885     1 28884  -10 0.06 0.02 0.04    0    0
 snmpd     2726     1  2725  -10 0.13 0.08 0.05    0    0
 snmpd    14632     1 14631  -10 0.02 0.02    0    0    0
```

SNMP on every server
Consumes < .1

NO spawned processes

Agent Overhead of z10EC

```
Report: ESALNXP      LINUX HOST Process Statistics Report
-------------------------------------------------------
node/     <-Process Ident-> Nice <------CPU Percents---->
 Name     ID    PPID   GRP  Valu  Tot  sys user syst usrt
--------- ----- ----- ----- ---- ---- ---- ---- ---- ----
 agent     8853     1  4390    0 2.24 0.01 0.02 1.38 0.83
 agent     9878     1  4657    0 1.98 0.01 0.02 1.15 0.80
 agent     6451     1  4392    0 5.68 0.03 5.59 0.03 0.02
 agent     9644     1  4392    0 2.14 0.01 0.01 1.34 0.78
 agent     7488     1  4379    0 1.42 0.01 0.01 0.84 0.56
 agent     9634     1  4362    0 1.92 0.01 0.01 1.14 0.75
 agent     5524     1  4414    0 5.22 0.04 5.14 0.03 0.02
 agent     7613     1  4525    0 1.44 0.01 0.02 0.88 0.53
 agent     7506     1  4388    0 1.41 0.01 0.02 0.83 0.55
 agent     6673     1  3725    0 1.41 0.01 0.02 0.83 0.55
 agent     6610     1  3680    0 1.44 0.01 0.02 0.89 0.52
 agent     6629     1  3680    0 1.51 0.01 0.01 0.90 0.59
 agent     6624     1  3677    0 1.39 0.01 0.02 0.82 0.54
 snmpd     1042     1  1041  -10 0.03 0.02 0.02    0    0
 snmpd      977     1   976   15 0.04 0.02 0.02    0    0
```

Note “agent” uses little CPU, same as “snmpd”

Spawned processes excessive – Need full picture

Capacity Planning for 1000 virtual servers

Company A: Consolidation project, 10,000 distributed servers

- 10 CECs (5 196, 5 ec12), 300 IFLs
- 2Q2011: 1,200 virtual servers (adding 200 per month)
- 1Q2012: 1,800 virtual servers (adding 200 per month)
- 3Q2012: 2,200 virtual servers, “ramping up soon”
- 1Q2013: 2,500 virtual servers

Company B: Consolidation and new workload

- 12 CECs, 60 LPARs, 183 IFLs
- 800 servers

Company C: Websphere

- 4 CECs (+2) , 16(+4) LPARs, 60 IFLs
- 800 servers, (+100)

Company M (Oracle)

- 1 CEC, 7 LPARS, 17 IFLs -> 50 IFLs->2 CECs, 120 IFLs
- 160 (LARGE) servers (july/2012)

Installation A – Server Consolidation

Consolidation source servers

- IBM HS21 (8GB),(2x4 core, 2.5Ghz)
- IBM X3550 (4GB) (2x4 core, 2.5Ghz)
- IBM X3655 (32GB) VM (2x4 core, 2.5Ghz)
- Sun M4000 (64GB) (4x4core, 2.4Ghz)
- Sun T5140 (32GB) (2x8 core, 1.2Ghz)
- Many others

Capacity planning process for consolidation:

- Inventory server counts (10,000+)
- Tally Gigahertz used (using native SAR)
  - By server, by application
- Spec processors based on GHz consumed by workload
- Spec storage on conservative basis

Installation A Highlights

Processors

- 1 z196 (R&D)
- 4 z196 (was z10) -> NOW 4 z196, 4 EC12….

IFLs

- 58 IFLs production -> 300 IFLs

Architecture

- Two data centers, High availability

Server counts

- 1800 servers (1Q12)
- 2200 servers (2Q12)
- 2500 servers (1Q13)

Processors (1Q,2012):

- Z196 – Lab, 18 IFLs, 2 LPARs, 4:1 Storage overcommit
- Z196(4) Production
  - 2 z/VM LPARs each, Production, Staging
  - 20-30 IFLs per CEC
  - (Some number of GP as well)
  - Disaster Recover available by shutting staging down

LPAR Sizes for Production

- 14-24 IFLs each (Shared)
- 256 GB Central each LPAR
- 24-72 GB Expanded (-> 128GB)

Installation A – Initial Project

Linux project started April, 2009

- 38 servers
- 3 IFLs

Small "traditional vm" system prior,

- skills available
- Hired one more
- Current staff including manager: 5

2500 servers now operational (1Q 2013)

Workloads: Websphere, Oracle

Users get 50 guests at a time,

- 25 on each datacenter

Growth

- Adding 200 servers per month for existing workload
  - Planned 3000 servers by 11/2012? (on target – not….)
- Last years “Next” application: New oracle workload,
  - replacing 400 cores (SUN)
  - 4 TB database (12 TB / cluster)
  - Sized at 32 IFLs (12:1) (Gigahertz sizing)
  - 1 TB real storage
- This year “next” 5 Petabytes

Project: Ground up resizing

- Jvms per server, heap sizes

Installation B – z Overview

Highlights of Z/VM LPARs (2012)

- 12 z10 / z196 (ramping up, 24 cecs currently)
- 183 IFLs (LPAR Overcommitt Level - 288 Logical processors - 1.5: 1)
- 3800 GB Cstore, 250 GB Xstore
- Five major data centers
- 800 servers (Websphere, Oracle)
  - Many servers in 30-40GB range
- 200 Servers per FTE is working number

Production LPARS

10-32 IFLs Each (Run out of 250GB storage with large servers)

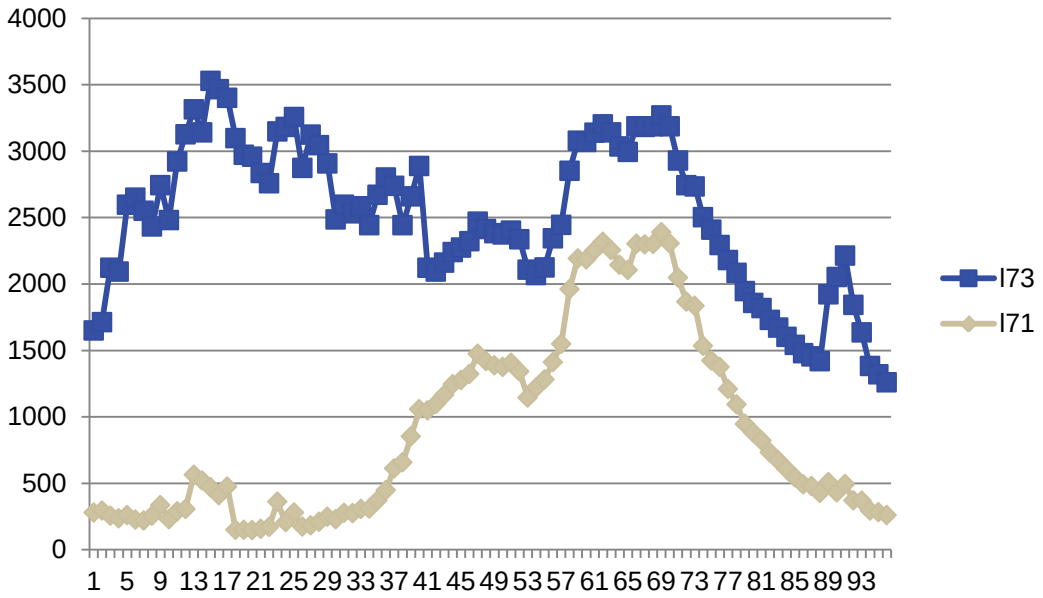
150GB – 250GB Central Storage

20-100 servers per LPAR

Installation B – z Overview (Big CPU Picture)

```
Report: ESALPARS     Logical Partition Summary
Monitor initialized: 11/06/10 at 16:07:10 on 2097 serial 374E: 11/0
-------------------------------------------------------------------
          <--Complex--> <-------Logical Partition--->  <-Assi Proce
          Phys Dispatch              Virt <%Assigned>  <---LP Type
Time      CPUs    Slice Name     Nbr CPUs Total  Ovhd  Weight
--------  ---- -------- -------- --- ---- -----  ----  ------ -----
16:09:00    37  Dynamic Totals:    0   50  3146  25.0    3000
                        L43       19    6 574.6   0.6     148 IFL   <-- 95%
                        C41       10    1 100.0   0.0     Ded ICF
                        C42       11    1  96.1   0.1     850 ICF
                        C43       14    1  99.7   0.0     Ded ICF
                        C44       15    1   0.8   0.1     150 ICF
                        P41        1    7 422.1   3.2     717 CP
                        P44        9    2  43.4   0.2      70 CP
                        T41        4    5 197.5   0.5     193 CP
                        T44        7    2   9.8     0      20 CP
                        L41       17   22  1557  19.6     777 IFL  <- 71%
                        L42       18    2  44.7   0.8      75 IFL
Totals by Processor type:
<---------CPU-------> <-Shared Processor busy>
Type Count Ded shared total assigned Ovhd Mgmt
---- ----- --- ------ ----- -------- ---- ----
CP       6   0      6 584.7    573.3  3.6  7.8
IFL     27   0     27  2220   2176.3 21.0 22.9  <-80% of IFLs
ICF      3   0      3 297.8    296.5  0.1  1.1
ZIIP     1   0      1  99.9     99.5  0.3  0.1
```

Installation B – z Overview



CEC “01” for one day, 38 IFLs

Storage overcommit: none

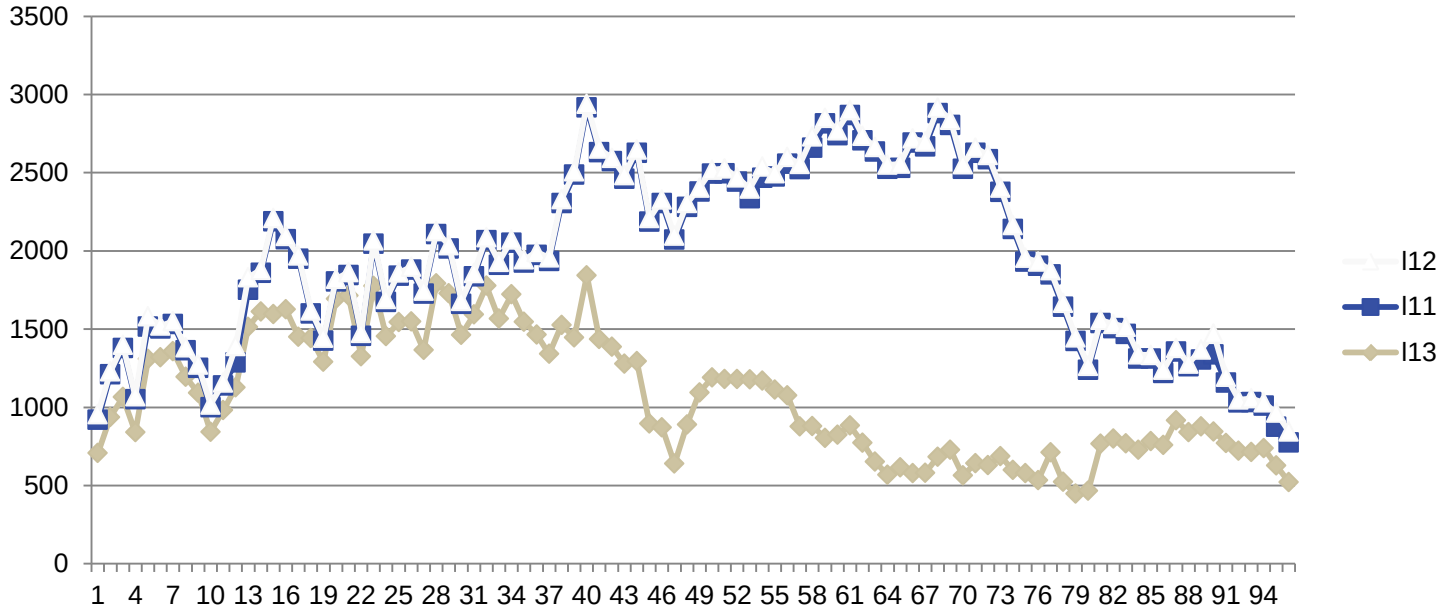
Processor overcommit: 5:1 (5 linux vcpu / IFL)

OPTIMAL WHITE SPACE!!!!

VELOCITY SOFTWARE

PROVEN PERFORMANCE

Installation B – z Overview



CEC “13” for one day, 38 IFLs

- 30 IFLs consumed is 80% busy

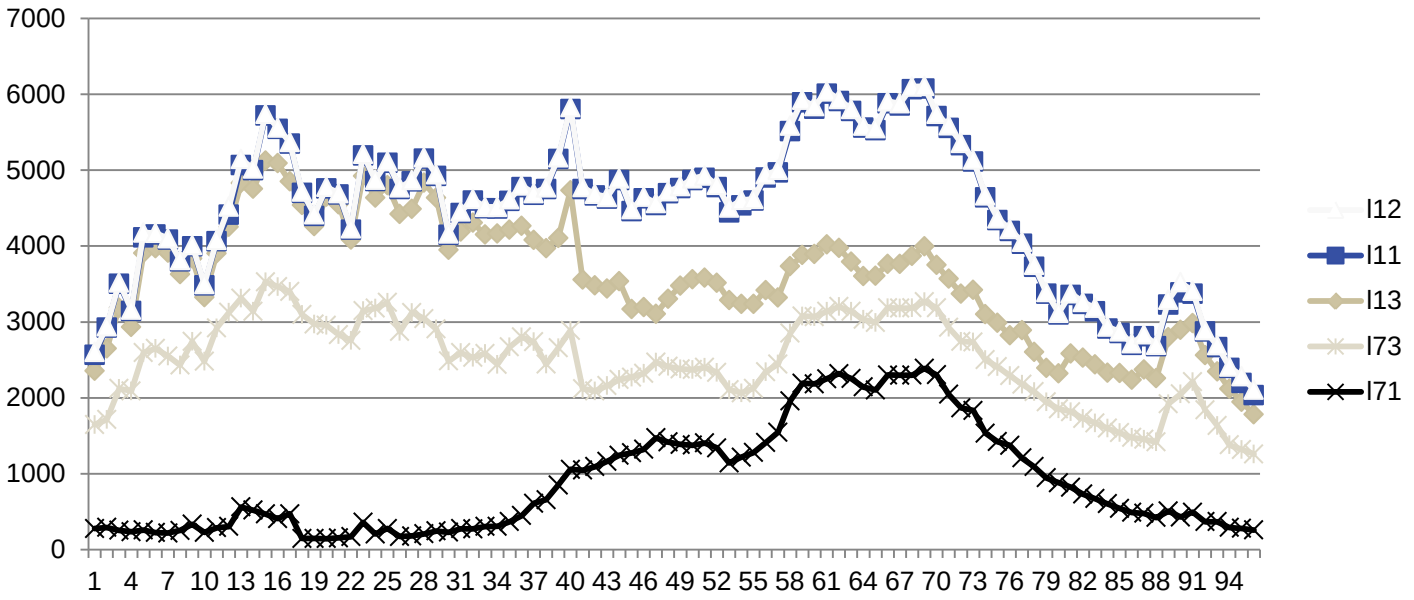
Storage overcommit: none

Processor overcommit: 5:1

VELOCITY SOFTWARE

PROVEN PERFORMANCE

Installation B – z Overview



Both CECs for one day, 76 IFLs

Room for growth or consolidation

- Balancing workload across CECs?

VELOCITY SOFTWARE

PROVEN PERFORMANCE

Installation C – z Overview

Highlights (POC 2005ish)

- 4 z196 (+1)**, in house DR**
- 60 IFLs
- 16 LPARS (+4 in 6 months)
- Two data centers, High availability
- 675 servers (Websphere) -> 800 servers
- Serious chargeback requirements

Production LPARS

4 production LPARs, 400GB / 90 GB ExStore (~ 20%)

Storage Overcommitt: 560gb / 490gb = 1.15

TEST/Dev LPARS

Installation C – Overcommit

IFLs: 55 (-5) (Went from z10 to z196)

675 servers (Websphere)

- 12 servers per IFL (was 10)
- 1030 Virtual CPU (25:1)

Storage

- 970 (+100) GB Central
- 184 GB Expanded (~20%) (IBM Recommendation 2-4GB BAD!)
- Virtual storage: 1600GB (+300)
- Overcommit (overall): 1.3 to 1

Installation M – Z Overview

4 Year project to date (2012)

- POC summer 2008
- Two VM/Linux Systems programmers

Processors:

- 1 z10 EC, 17 IFLs
- 7 lpars, 17 virtual cpus each (Worst Case)
- 560GB Real storage / 92 GB Expanded (~15%)
- DR site available

Storage – data on FCP (30TB), systems on ECKD

Linux Servers

- 120 servers (Big, ORACLE)
  - 7 servers per IFL
- 395 vcpus
  - (23:1 overcommit)
- 4gb-40gb
  - (1 / 2 size from original SUN servers)

- 974 GB Server storage
  - (1.5 : 1 overall overcommit)
  - 8GB per server???

Zones separated by LPAR

- Development
- Validation (Quality Assurance)
- Production (gets the resources when needed)

Workload zones (3 tier, by LPAR)

- Presentation
- Data (Oracle)
- Application (WAS)
- All heavy hitting (data, application) moved/moving to “z”

Installation M – Z Production LPAR Overview

LPAR “A” Development

- oracle,
- 110gb Central / 22gb Expanded, (~20%)
- 30 servers, 100 vcpus
- 30 page packs 3390-9

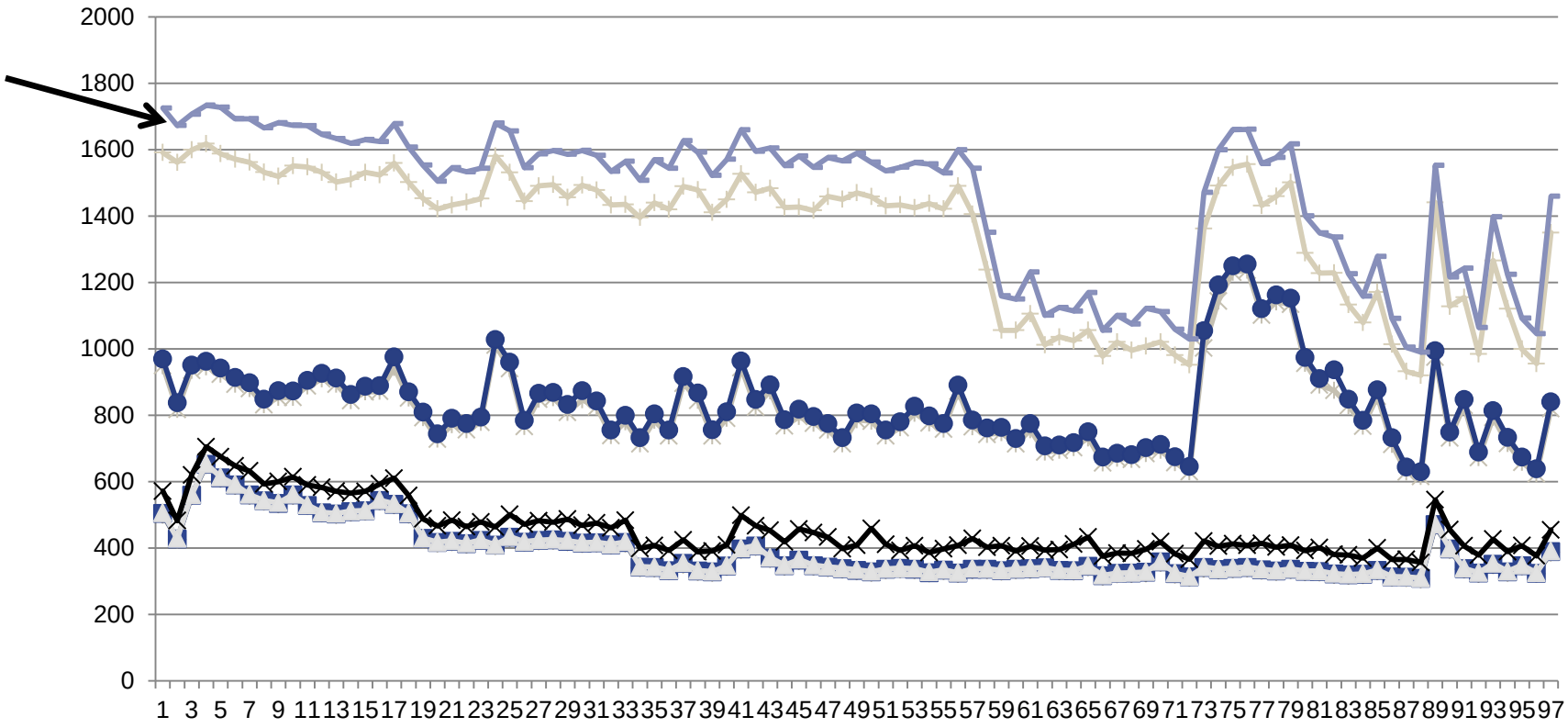
LPAR “1” Application

- WAS,
- 180gb Central / 40gb Expanded
- 20 servers, 80 vcpus
- 60 page packs 3390-9,

LPAR “4” Data

- Oracle
- 130gb Central / 24gb Expanded

Installation M – LPAR Sizing



- 17 IFLs, 7 lpars, 17 vcpus each, 7:1 overcommit
- Overhead significant from real processor overcommit

VELOCITY SOFTWARE

PROVEN PERFORMANCE

Installation M – Z Growth

Processors: Over 4 years

- Z9, 11 IFLs moved to z10 17 IFLs
- Moved to Z196, 25 IFLs (doubling capacity)
- Moved to 40 IFLs….
- Moved to 2 EC12s (50 IFLs)

Appl Developers see “pretty good performance”

- **Can we move too?**
- **Always issues on “other side”**

Workload Growth

- Adding 110 Oracle databases
- Replacing 32 Solaris Servers (120 cores)
  - “Server from Hell” had 30 databases on it

2011 status

- We have added a total of 154 z/Linux guests.
- We have turned a lot of these into Enterprise guests meaning in some cases we have multiple JVMs on a guest as well as multiple Oracle Data bases on a single guest.
- The majority of the guests are Oracle Data base guests ranging from 500MB to 15TB in size for a single Data base.
- We have also brought over multiple WAS servers. Other than using a lot of Memory and DASD storage things seem to be running well.

Velocity Software Performance Management

- **Instrumentation Requirements**
  - Performance Analysis
  - Operational Alerts
  - Capacity Planning
  - Accounting/Charge back
- **Correct data**
- **Capture ratios**
- **Instrumentation can NOT be the performance problem**

A scalable z/VM Performance Monitor

Traditional model (1989)

ZMON: Real time analysis

- Uses Standard CP Monitor

  Real Time Analysis

ZMAP: Performance Reporting

Post Processing

Creates Long Term PDB

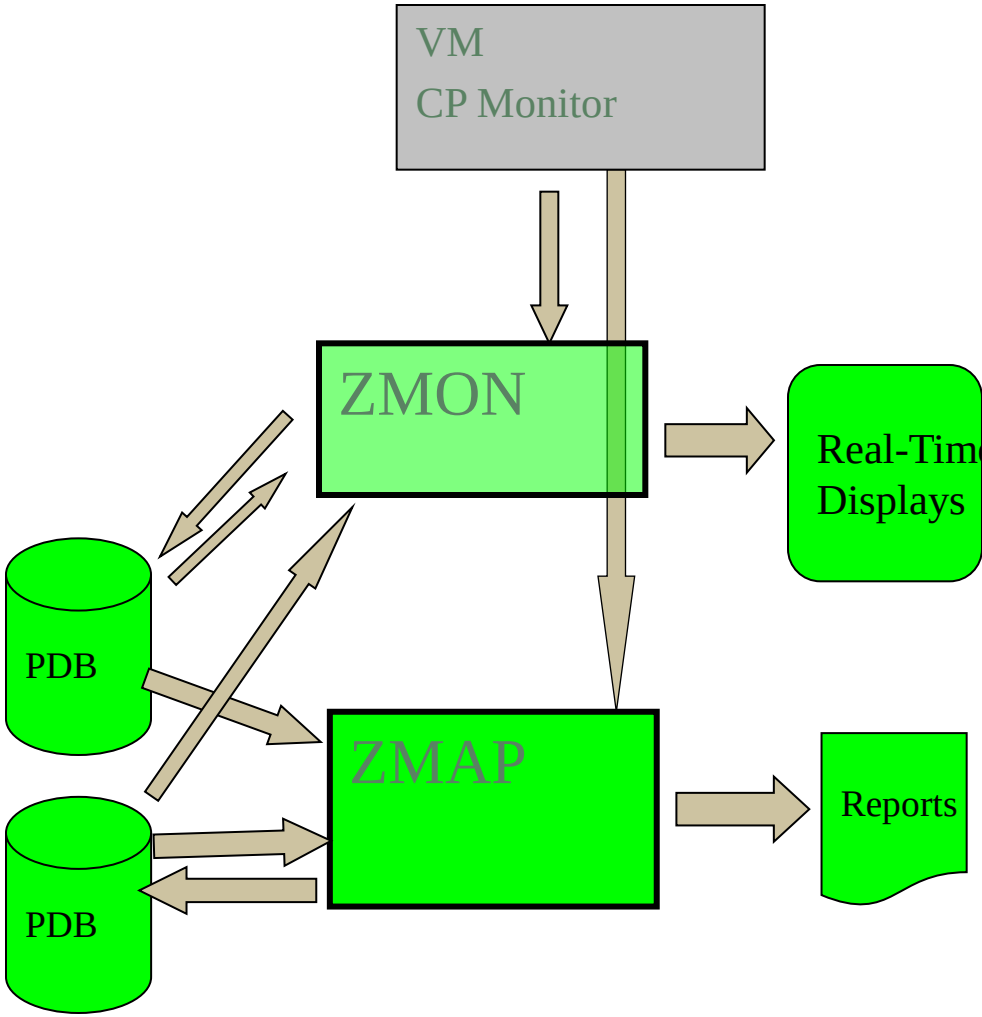
PDB or monwrite data input

PDB (Performance DataBase)

Complete data

By Minute, hour, day

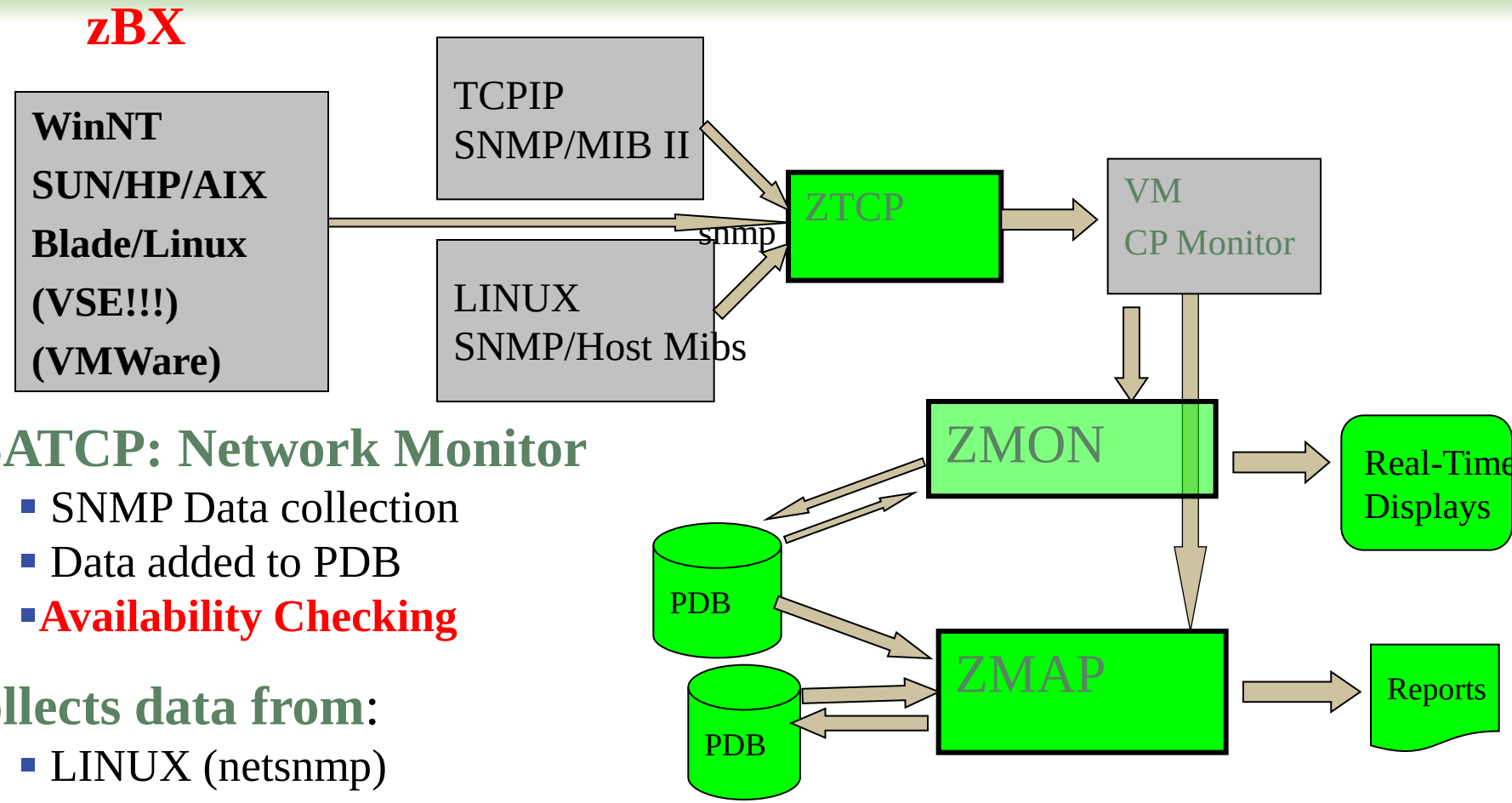
Monthly/Yearly Archive



VELOCITY SOFTWARE

PROVEN PERFORMANCE

Linux and Network Data Acquisition



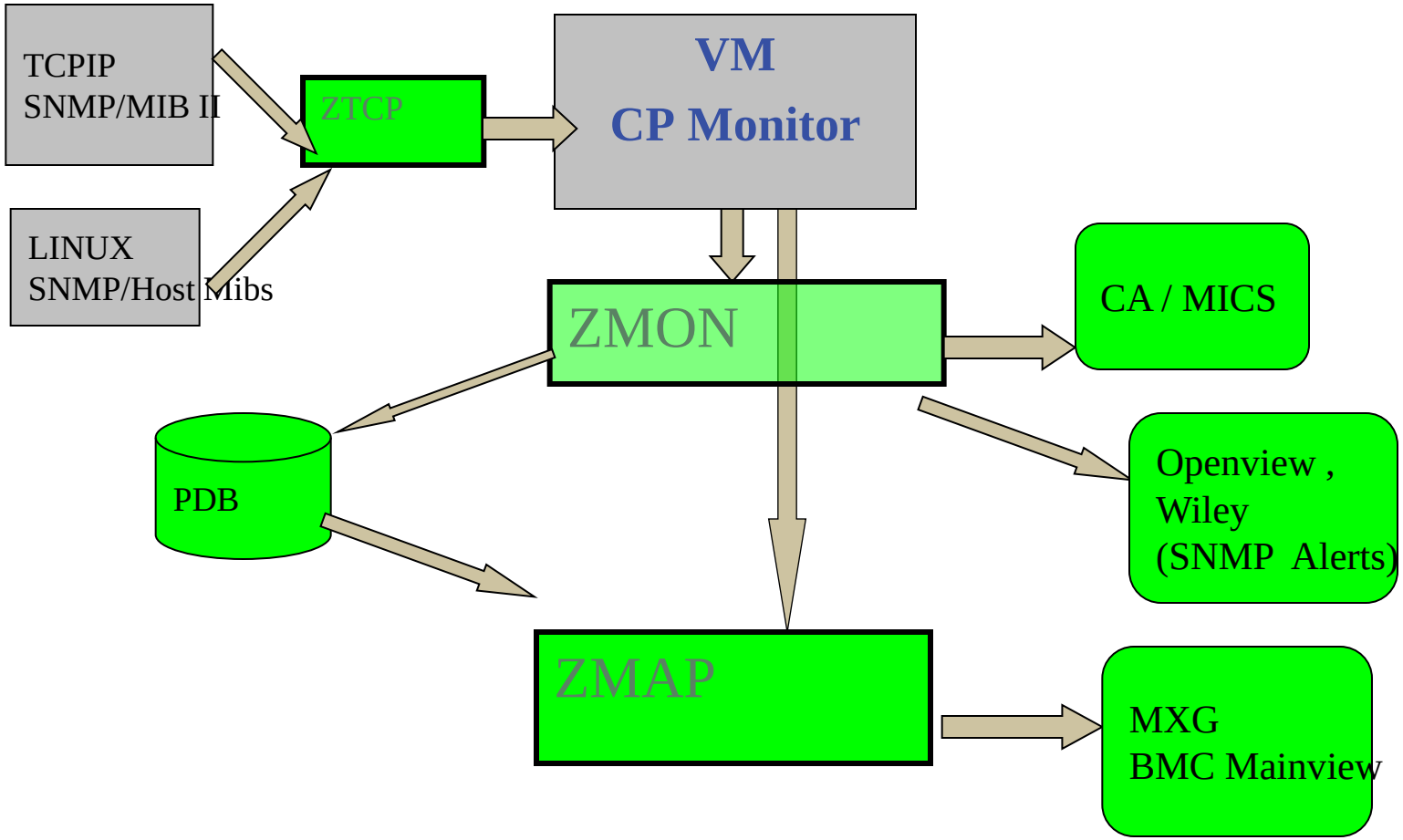
ESATCP: Network Monitor

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

Collects data from:

- LINUX (netsnmp)
- NT/SUN/HP (native snmp)
- Printers/Routers….

Add Enterprise Support for capacity planning tools

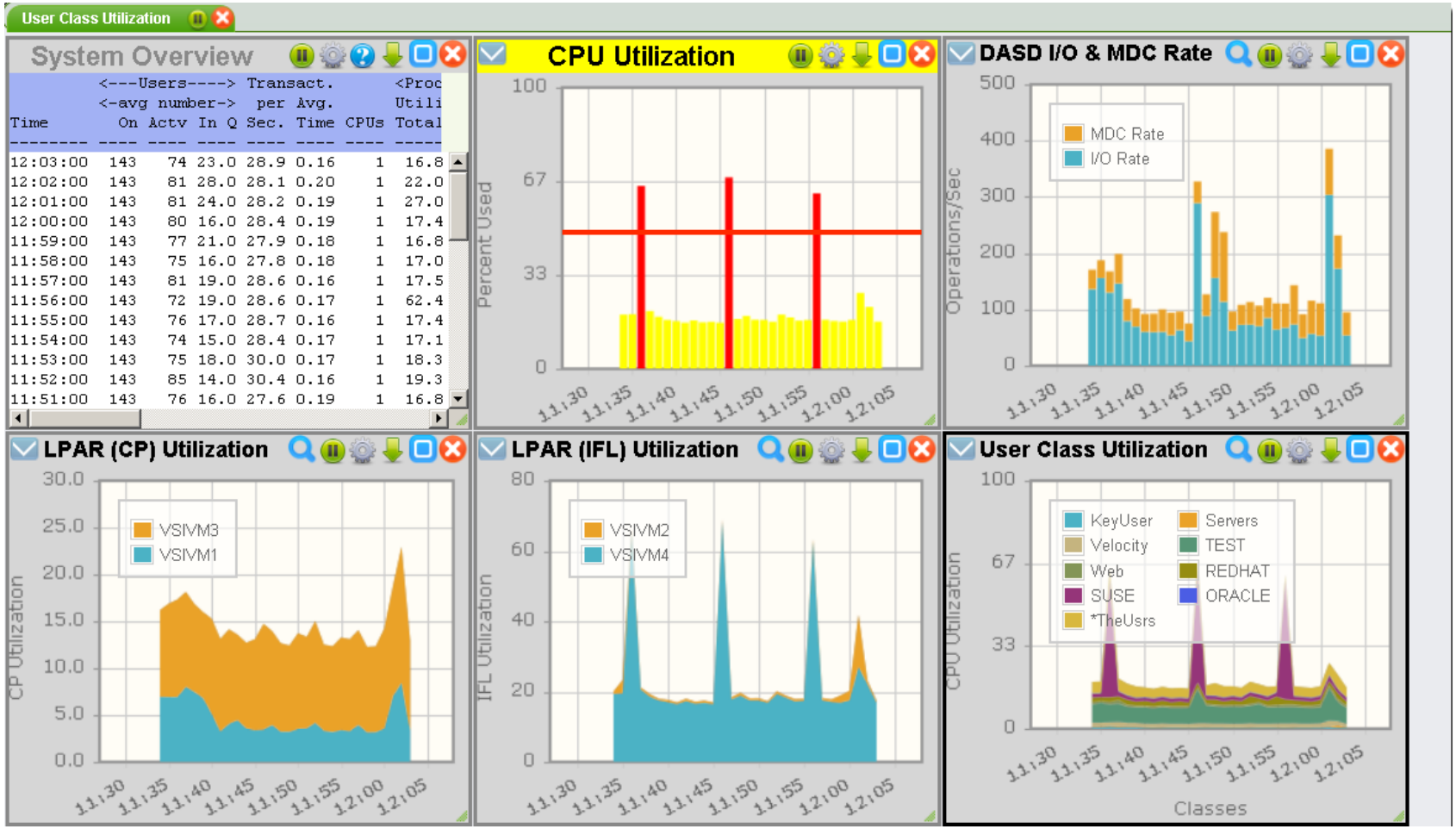


VELOCITY SOFTWARE

PROVEN PERFORMANCE

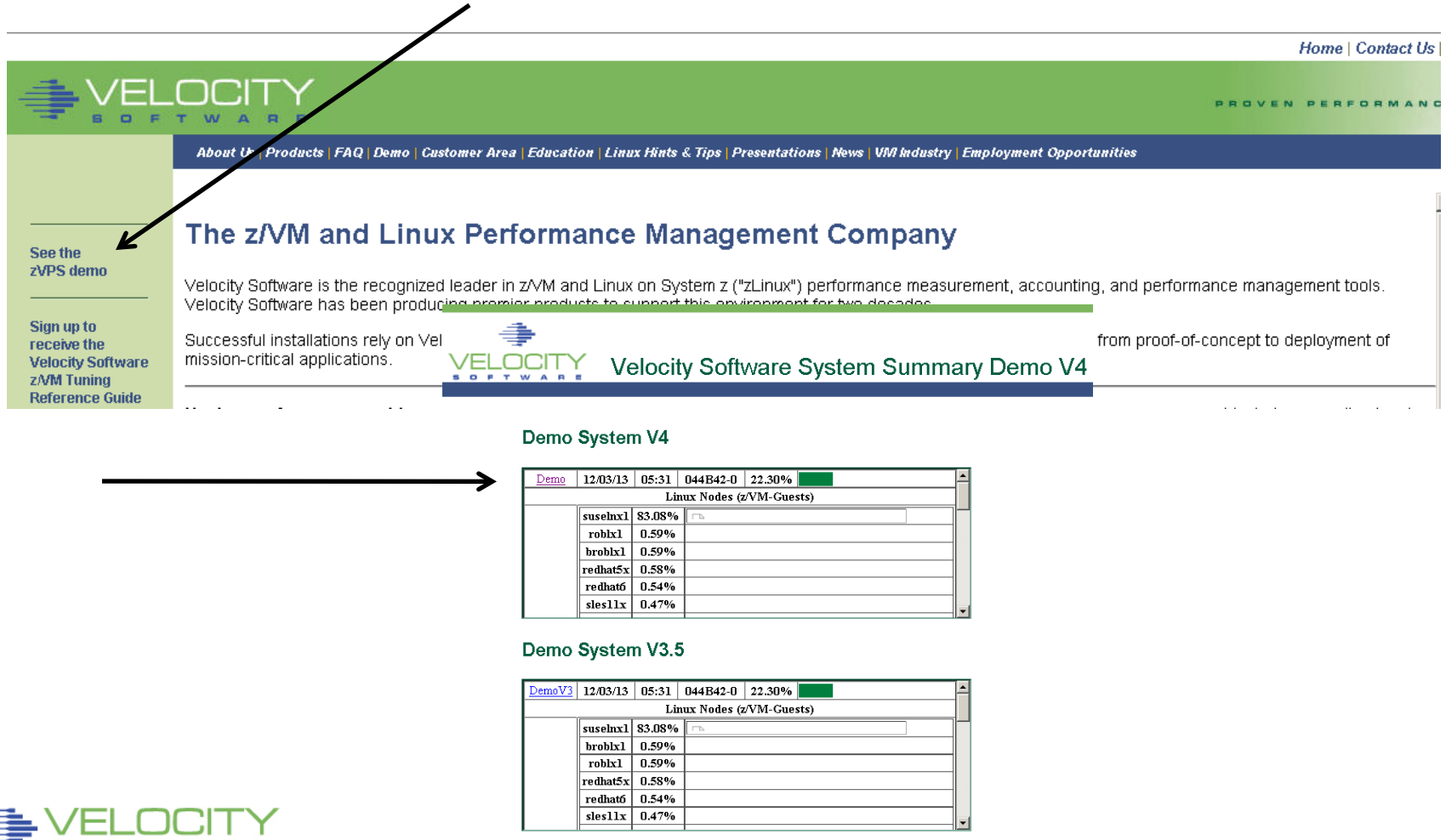
What we're doing for Capacity Planning

- CPU by lpar by Processor type
- CPU BY userclass



See what we're doing for Capacity Planning

- **VelocitySoftware.com**
- **See the demo**

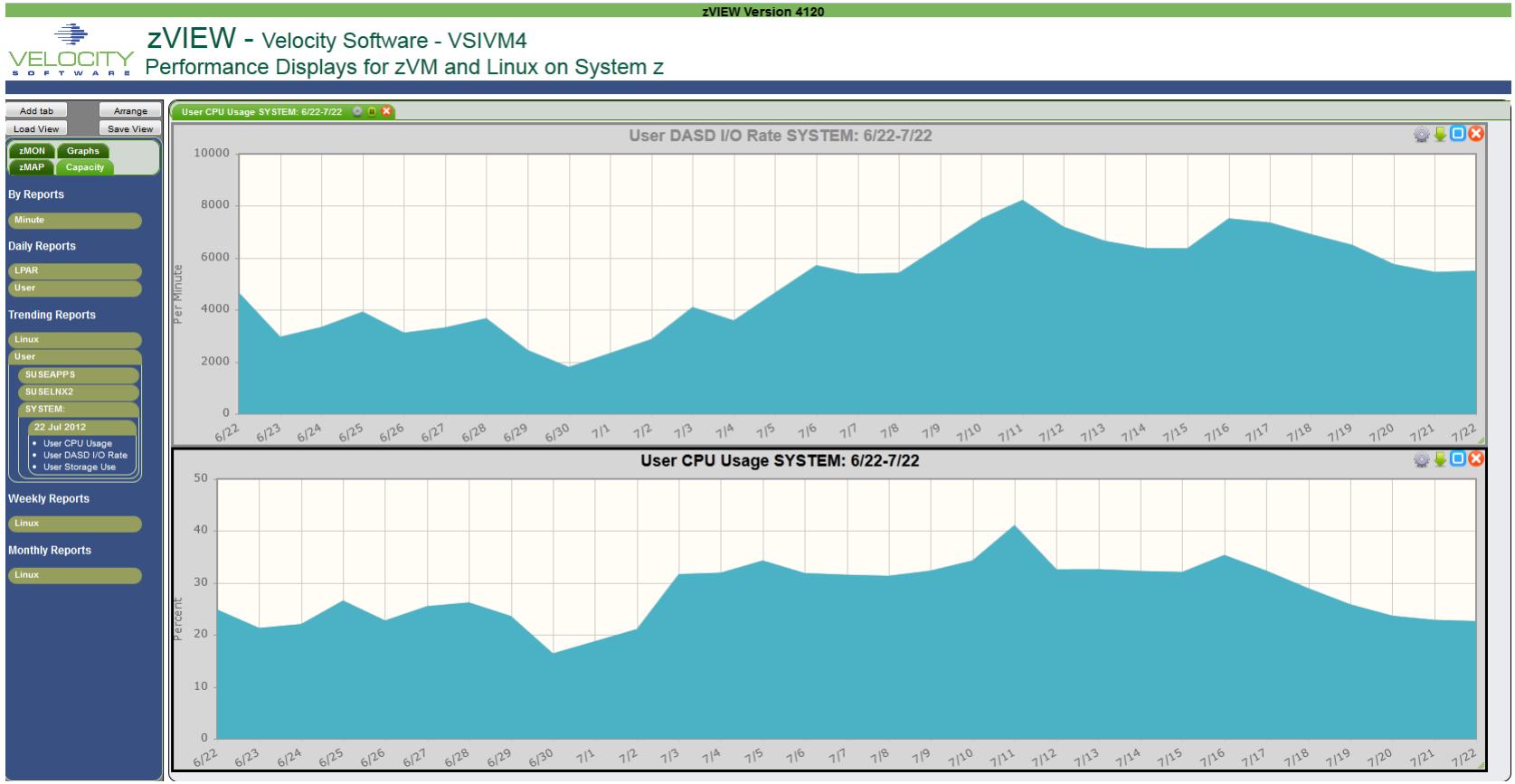


VELOCITY SOFTWARE

PROVEN PERFORMANCE

See what we're doing for Capacity Planning

- **Monthly charts now easily viewed**

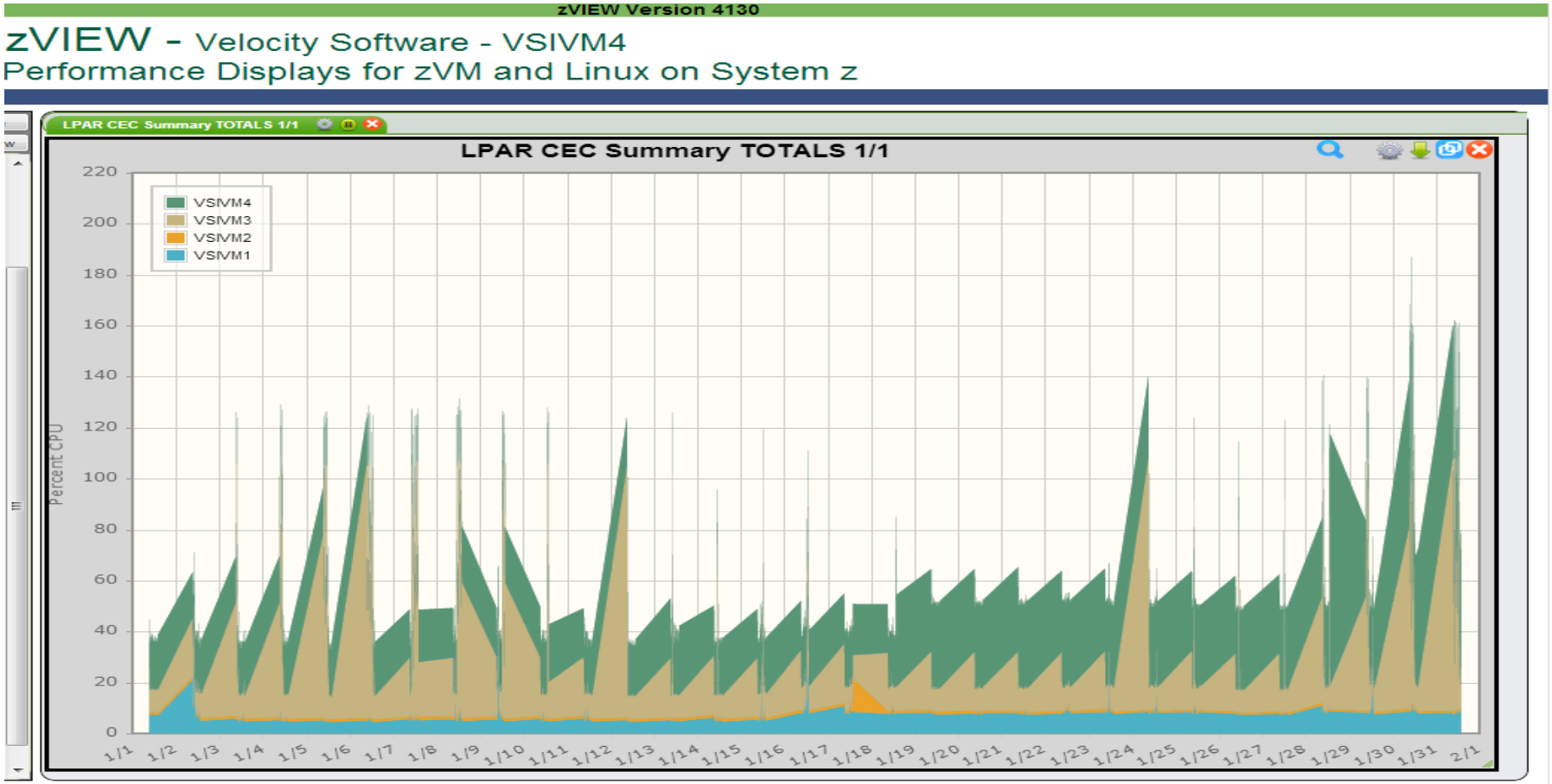


VELOCITY SOFTWARE

PROVEN PERFORMANCE

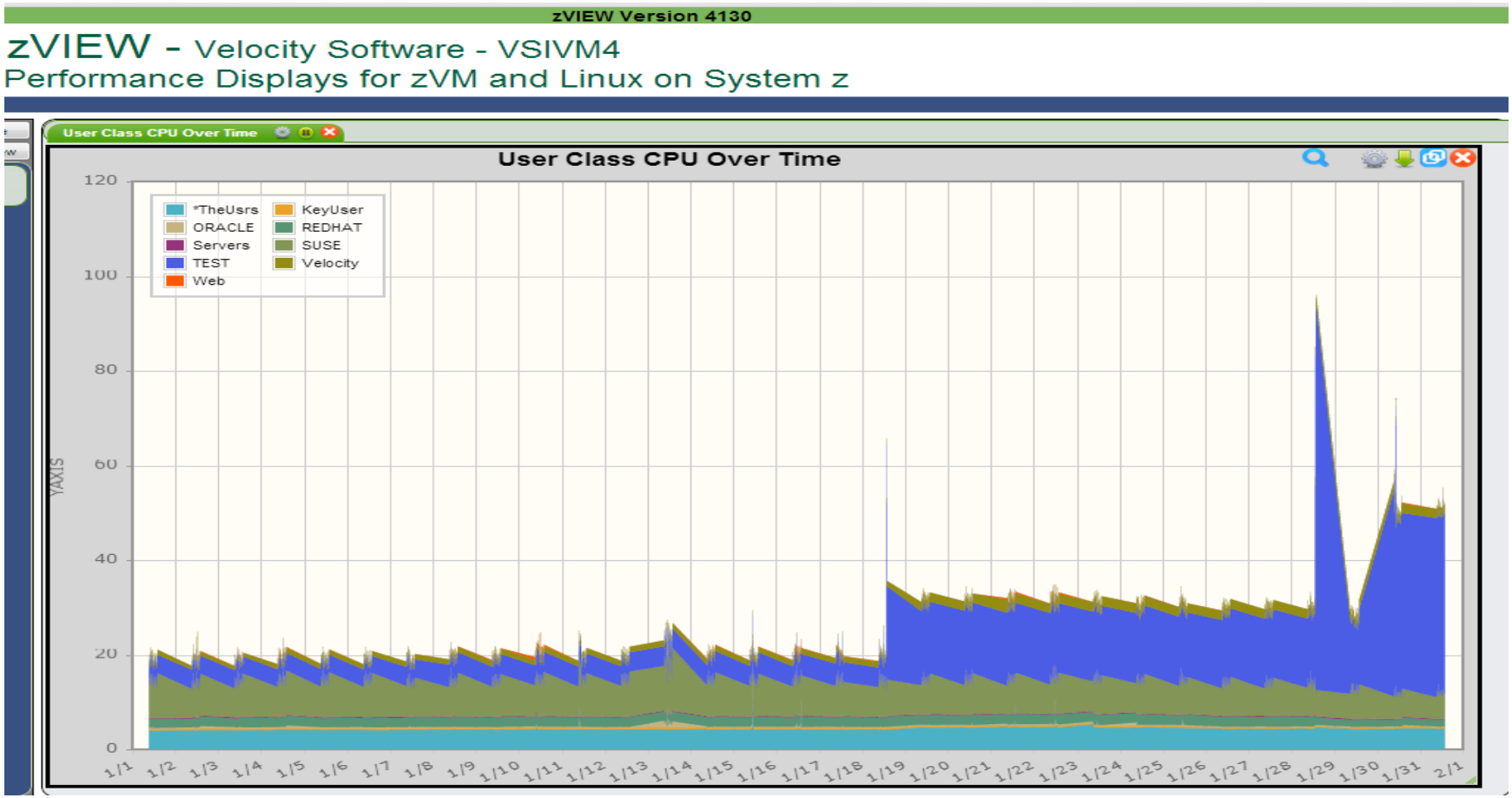
See what we're doing for Capacity Planning

- **CEC Utilization for January**



See what we're doing for Capacity Planning

- **DEMO LPAR Utilization for January**



VELOCITY SOFTWARE

PROVEN PERFORMANCE

Capacity Planning Metrics

Processor Ratios:

- LPAR logical processors per real processor (LPAR Overhead)
- Linux virtual processors per real (Linux overhead)

Storage ratios

- Storage per processor
- Expanded storage per Real storage
- Overcommit ratios

Servers per processor

- How many distributed servers replaced per IFL?

Capacity Planning Summary

1000 servers has been done

- **Management required.**
- **Issues are "driving too fast to stop for gas"**
  - Saving too much to figure out where we're at
  - Do a capacity plan, but don't have time to review accuracy (2 years later)

Processors:

- Gigahertz are gigahertz
- Processors highly utilized and shared save money

Storage: No good guidelines

- Oracle and SAP are usually larger than WAS
- Expanded storage should follow the "Velocity best practices"

"I don't have time to see any crazy salesman; I have a battle to fight."