

# Capacity Planning for 1000 virtual servers (What happens when the honey moon is over?) (**SHARE SESSION** 10334)

Barton Robinson  
Velocity Software  
Barton@VelocitySoftware.com

# Capacity Planning for 1000 virtual servers

## Objectives

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

# Capacity Planning Processor Overview

## Processor requirements

- CECs
- IFLs
- LPARs

## LPAR Processor Overhead

- LPAR vcpu ratio to real IFL
- (1/2 % Physical overhead for each)

## Considerations

- Software paid per IFL
- **95% IFL utilization lowest cost**
- One installation replaced 30 “oracle servers” with one IFL
- One installation gets hardware & system software for free

# Capacity Planning Processor Considerations

## Term: Processor Overcommit

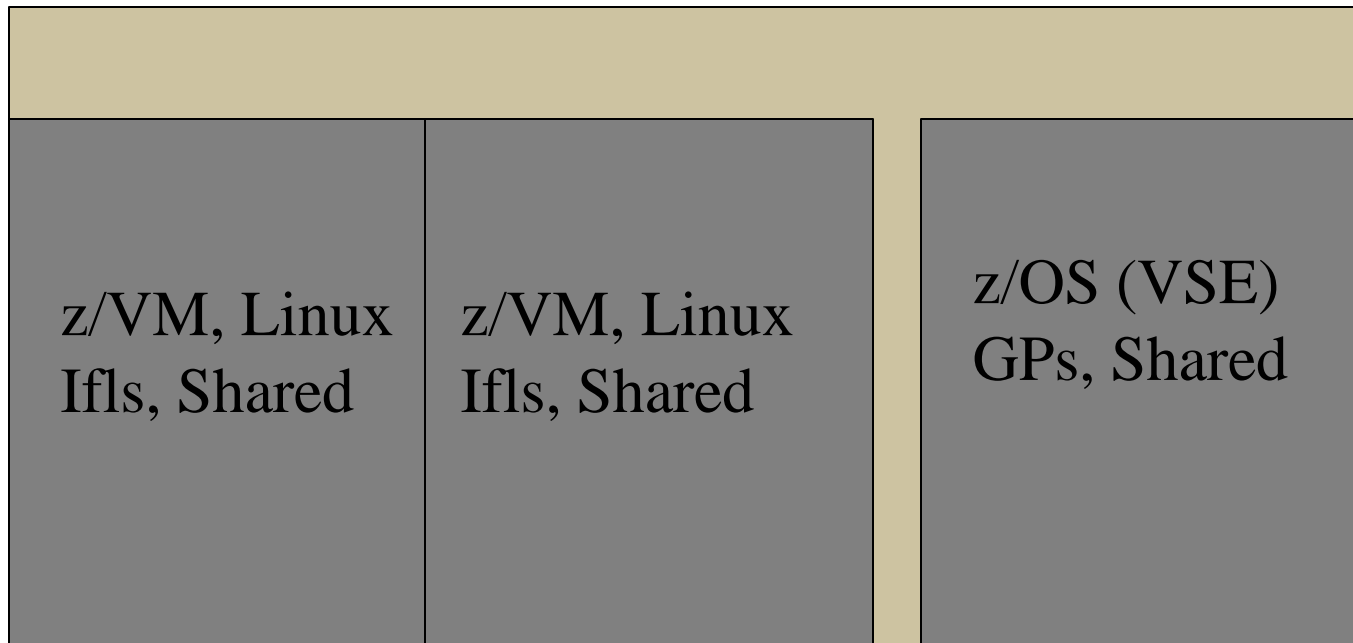
- Number of virtual cpus per IFL
- Software licensed per “cpu”

## Critical concept

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

# Configuration Topics - Processors

- **Processor CEC (z196, z10, z9)**
- **Configured multiple LPARs,**
  - IFLs (1-96), shared or dedicated
  - General Purpose Processors (1-96), shared or dedicated



## Storage requirements

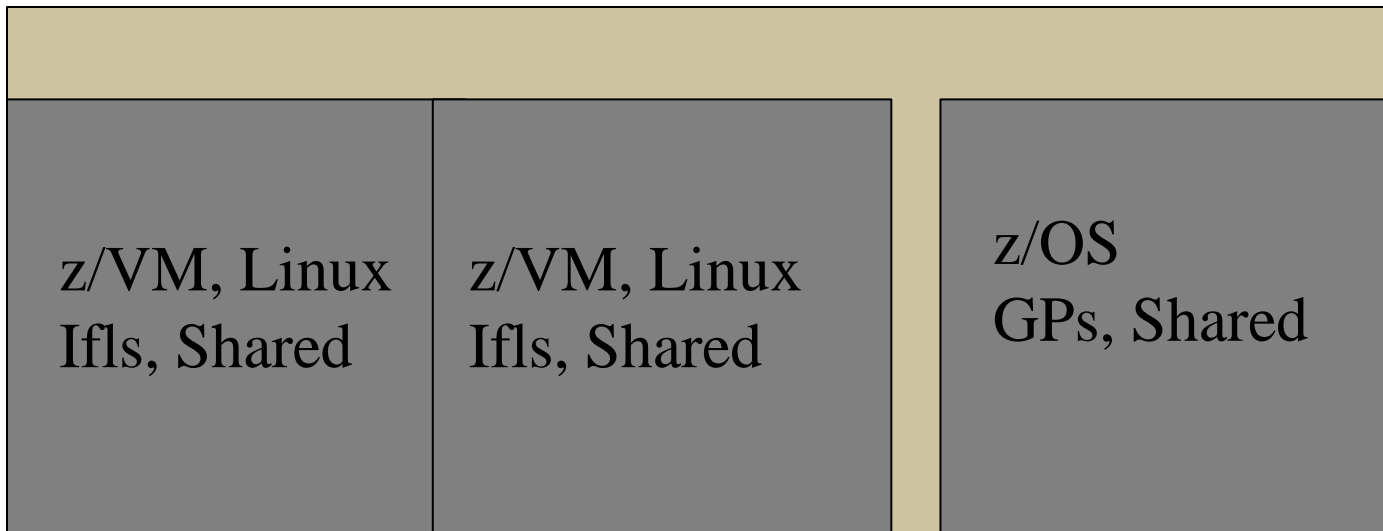
- Target Overcommit level
- Storage maximums (250GB per LPAR as of z/VM 6.2)
- Expanded Storage (20%)

## Storage consideration (to keep ifls at 95% busy)

- How much storage is required?
- What storage tuning should be performed?

## Configured multiple LPARs,

- Storage dedicated (256GB per LPAR)
- Terabytes central storage on the machine
- Expanded Storage for paging buffer (20% of real)
- Overcommit Ratio within LPAR? (1:1, 2:1, 4:1?)



## Replacements?

- 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: 5.0-5.2 Ghz

## Server Storage sizing

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

## Linux Internal overhead

- Linux vcpu ratio to real IFL (20:1 ?)
- 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 – Velocity Software Performance Suite)**

## Infrastructure Requirements

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

## Management 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 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/ Name	<-Process ID	Ident-> PPID	GRP	Nice Valu	<-----CPU Percents-----> 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 is on every server  
Consumes < .1

Note, NO spawned processes

# Agent Overhead of z10EC

Report: ESALNXP          LINUX HOST Process Statistics Report  
Monitor initialized: 04/15/11 at 10:00:00 on 2097 serial

```
-----  
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

- 4 CECs, 120 IFLs
- Currently (2Q2011) 1,200 virtual servers (adding 200 per month)
- Currently (1Q2012) 1,800 virtual servers (adding 200 per month)

## 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
- 675 servers, (+75)

## Company M (Oracle)

- 1 CEC, 7 LPARS, 17 IFLs
- 120 (LARGE) servers

# 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 used
- Spec storage on conservative basis

## Processors

- 1 z196 (R&D)
- 4 z196 (was z10)

## IFLs

- 58 IFLs production

## Architecture

- Two data centers, High availability

## Server counts (1Q11)

- 1800 servers (+600)

## Processors (1Q,2011):

- 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)



## Linux project started April, 2009

- 38 servers
- 3 IFLs

## Small “traditional vm” system prior,

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

## 1800 servers now operational (March, 2012)

## Workloads: Websphere

## Users get 50 guests at a time,

- 25 on each datacenter

## Growth

- Adding 200 servers per month for existing workload
  - 3000 servers by 11/2012?
- Last year “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

- 12 z10 / z196 (ramping up, 24 ceecs currently)
- 183 IFLs (288 Logical processors (1.5: 1))
- 3800 GB Cstore, 250 GB Xstore
- Five 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

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

```

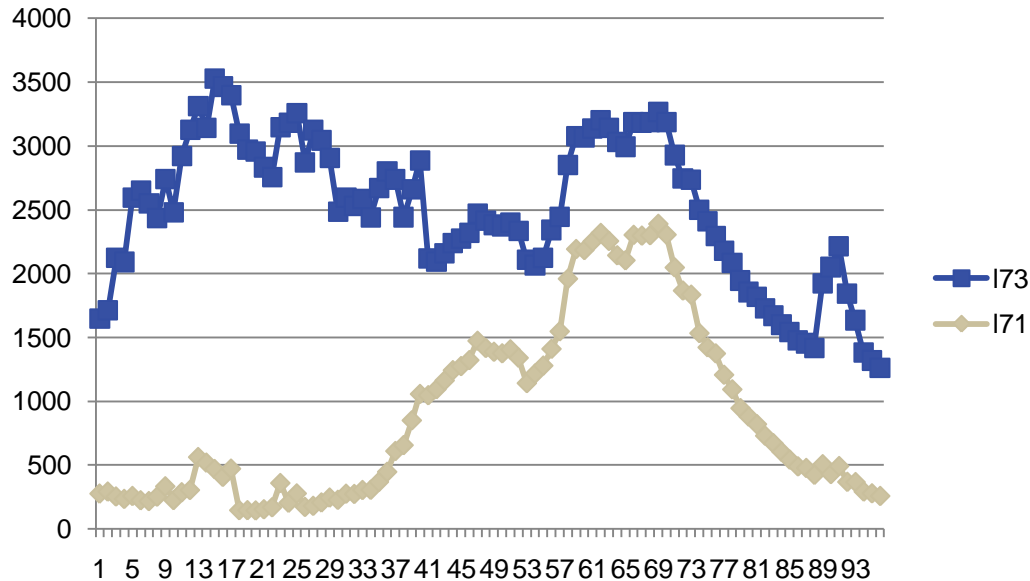
-----
      <--Complex--> <-----Logical Partition---> <-Assi Proce
      Phys Dispatch          Virt <%Assigned> <---LP Type
Time      CPUs      Slice Name      Nbr CPUs Total  Ovhd  Weight
-----
16:09:00   37  Dynamic Totals:      0   50  3146  25.0   3000
          <b>L43</b>      <b>19</b>   <b>6 574.6</b>  <b>0.6</b>   <b>148 IFL</b>  <b><-- 95%</b>
          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
          <b>P44</b>      <b>9</b>   <b>2 43.4</b>  <b>0.2</b>   <b>70 CP</b>
          T41         4     5 197.5   0.5     193 CP
          T44         7     2   9.8    0       20 CP
          <b>L41</b>      <b>17</b>  <b>22 1557</b>  <b>19.6</b>  <b>777 IFL</b>  <b><-- 71%</b>
          <b>L42</b>      <b>18</b>   <b>2 44.7</b>  <b>0.8</b>   <b>75 IFL</b>
  
```

Totals by Processor type:

```

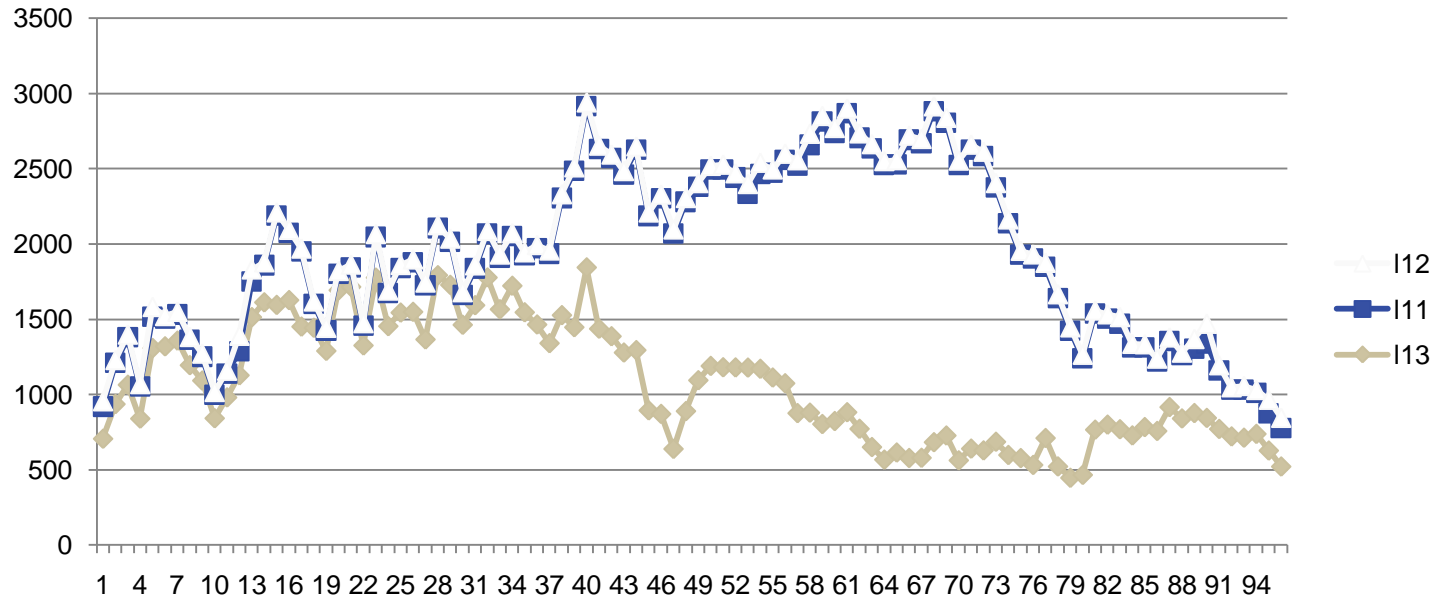
<-----CPU-----> <-Shared Processor busy>
Type Count Ded shared total assigned Ovhd Mgmt
-----
CP      6   0     6 584.7   573.3  3.6  7.8
<b>IFL</b>  <b>27</b> <b>0</b>   <b>27 2220</b>  <b>2176.3</b>  <b>21.0</b>  <b>22.9</b>  <b><--80% of IFLs</b>
ICF     3   0     3 297.8   296.5  0.1  1.1
ZIIP    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**

# Installation B – z Overview



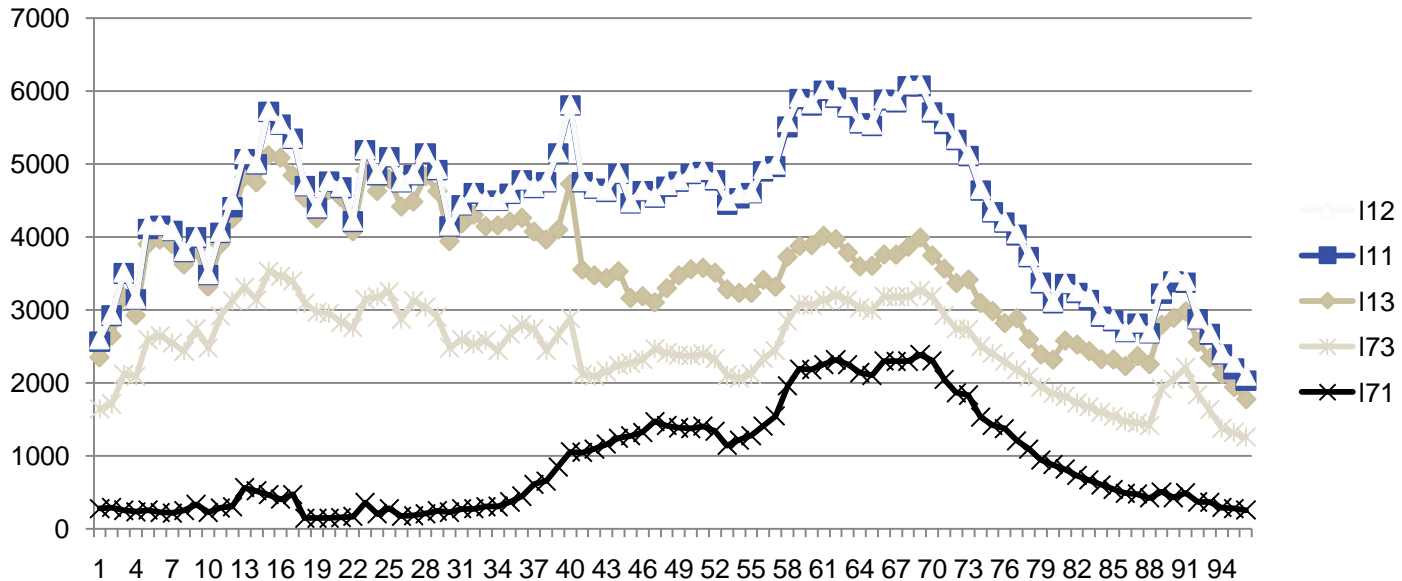
## CEC “13” for one day, 38 IFLs

- 30 IFLs consumed is 80% busy

**Storage overcommit: none**

**Processor overcommit: 5:1**

# Installation B – z Overview



## Both CECs for one day, 76 IFLs Room for growth or consolidation

- Balancing workload across CECs?

# Installation C – z Overview

## Highlights (POC 2005ish)

- 4 z196 (+1), **2 production, inhouse DR**
- 60 IFLs
- 16 LPARS (+4 in 6 months)
- Two data centers, High availability
- 675 servers (Websphere)
- Serious chargeback requirements

## Production LPARS

4 production LPARs, 400GB / 90 GB ExStore

Overcommitt: 560gb / 490gb = 1.15

## TEST/Dev LPARS



**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
- Virtual storage: 1600GB (+300)
- Overcommit (overall): 1.3 to 1

# Installation M – Z Overview

## 3 Year project to date (2011)

- POC summer 2008
- Two VM/Linux Systems programmers

## Processors:

- 1 z10 EC, 17 IFLs
- 7 lpars, 17 virtual cpus each
- 560GB Real storage / 92 GB Expanded
- DR site available

## Storage – 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,
- 30 servers, 100 vcpus
- 30 page packs 3390-6

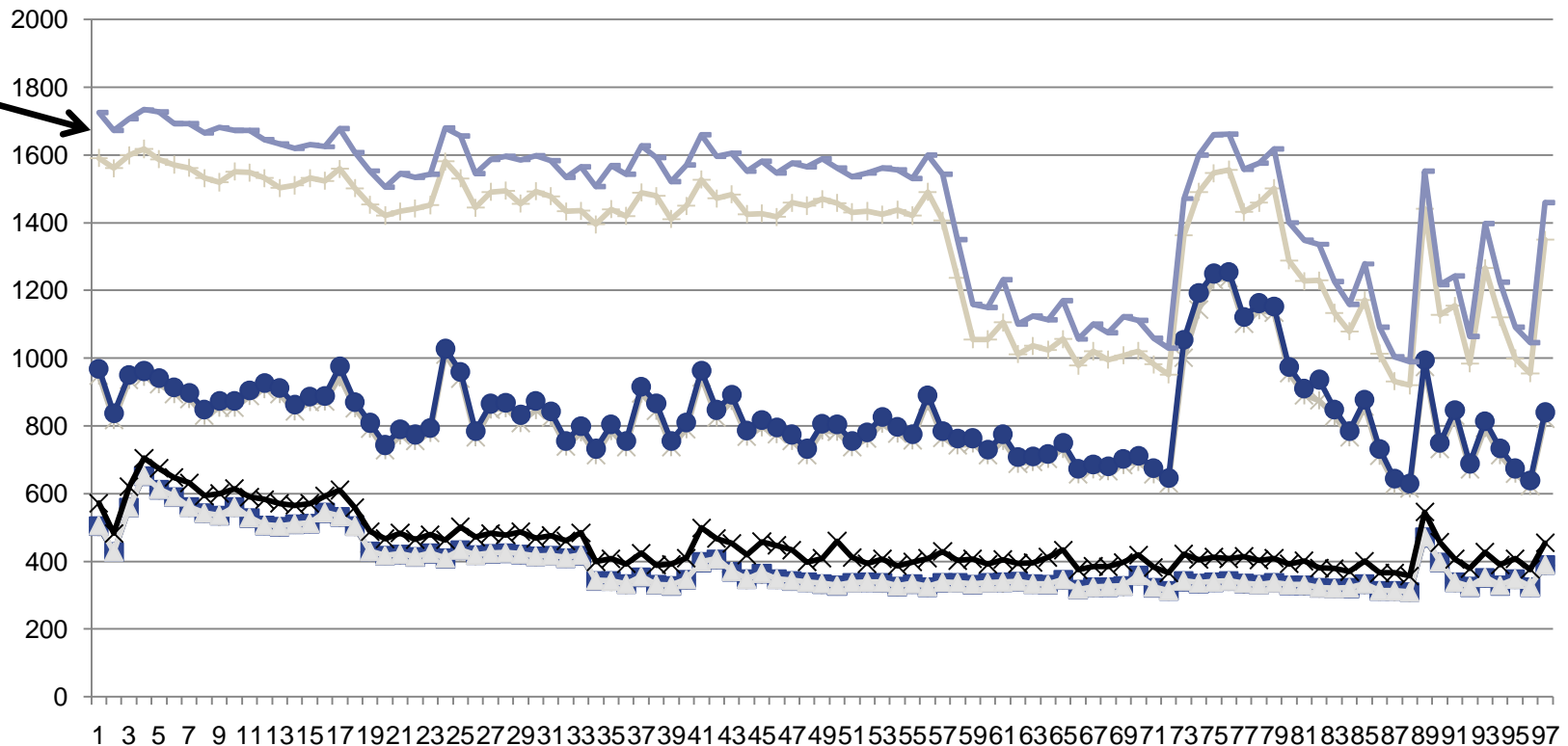
## LPAR “1” Application

- 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

## Processors: Over 3 years

- Z9, 11 IFLs moved to z10 17 IFLs
- Moving to Z196, 25 IFLs (doubling capacity)

## 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

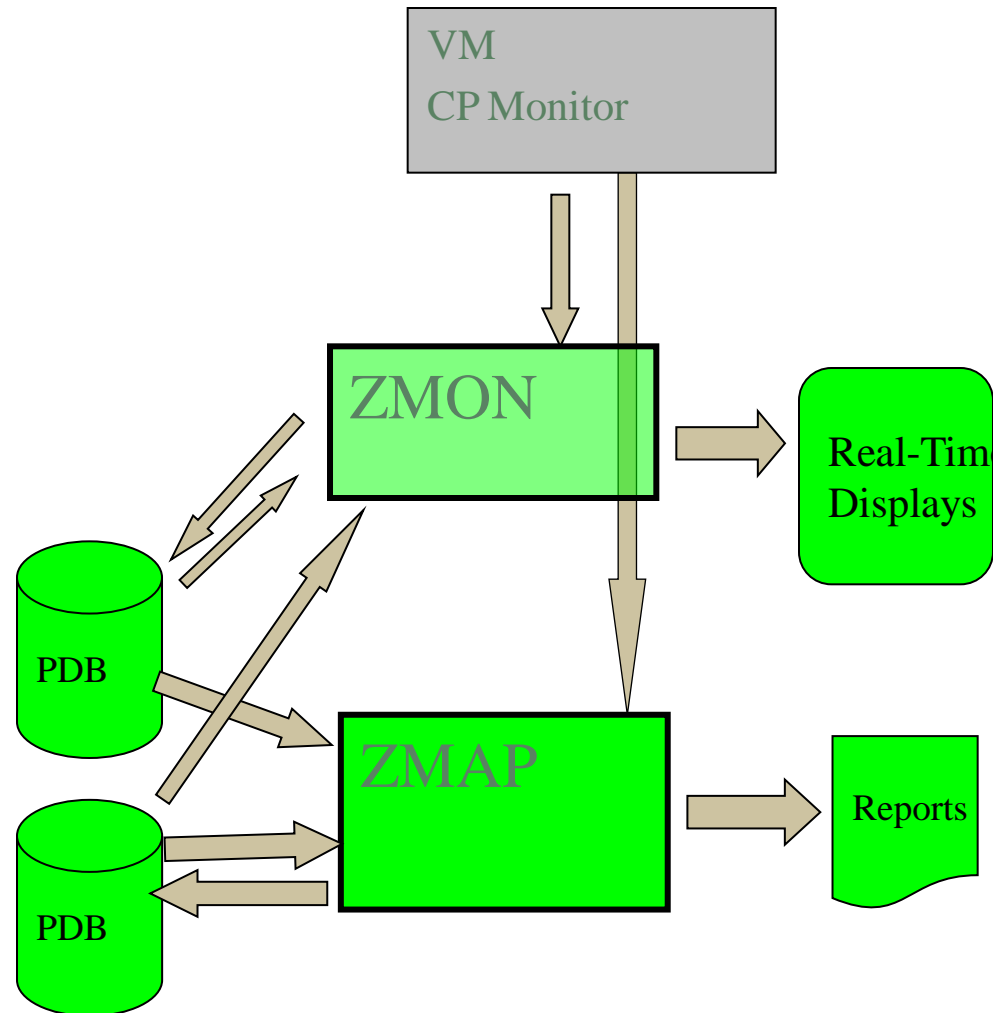
- Uses Standard CP Monitor  
Real Time Analysis

ZMAP: Performance Reporting

Post Processing  
Creates Long Term PDB  
PDB or monwrite data input

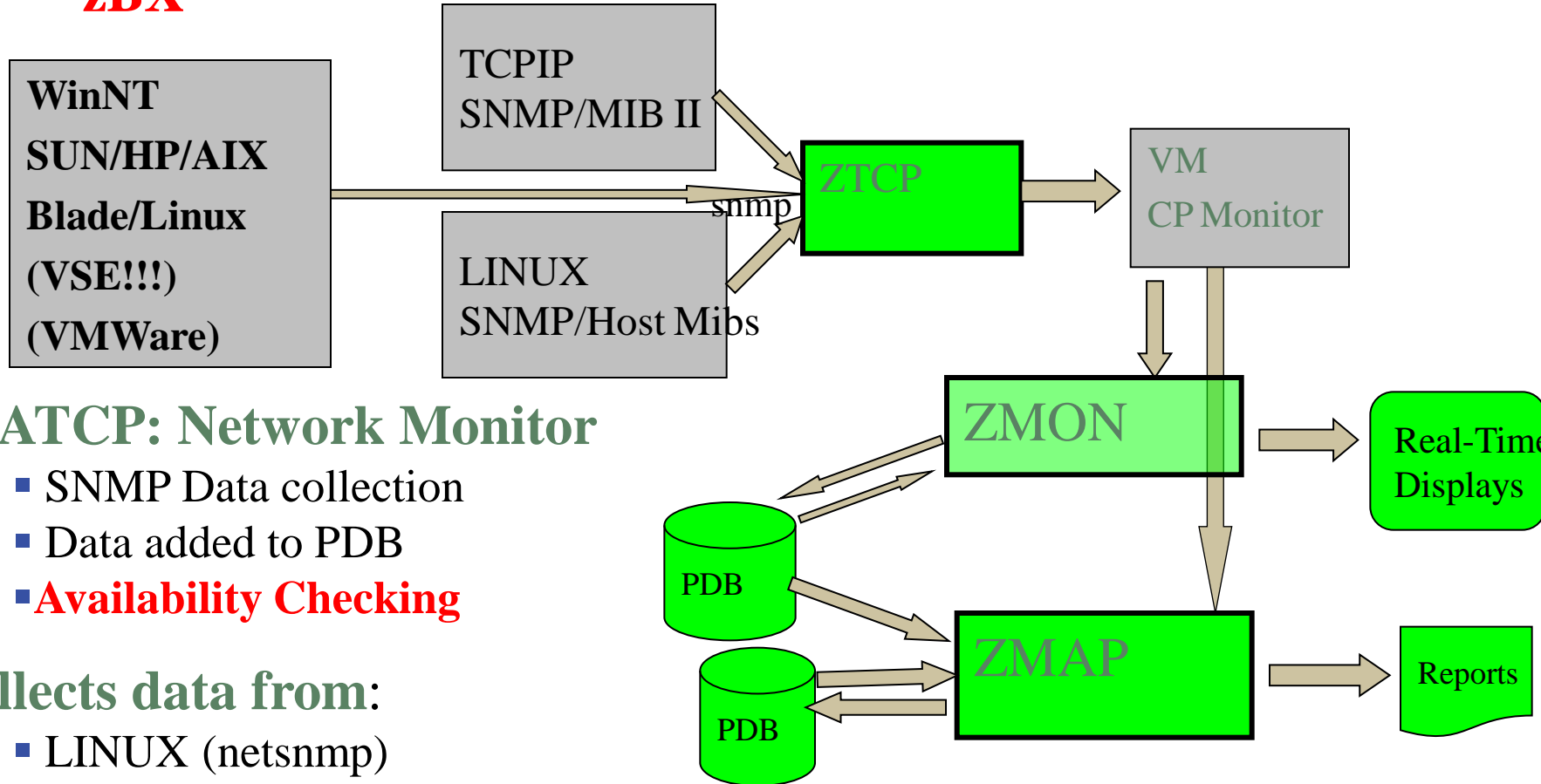
PDB (Performance DataBase)

Complete data  
By Minute, hour, day  
Monthly/Yearly Archive



# Linux and Network Data Acquisition

**zBX**



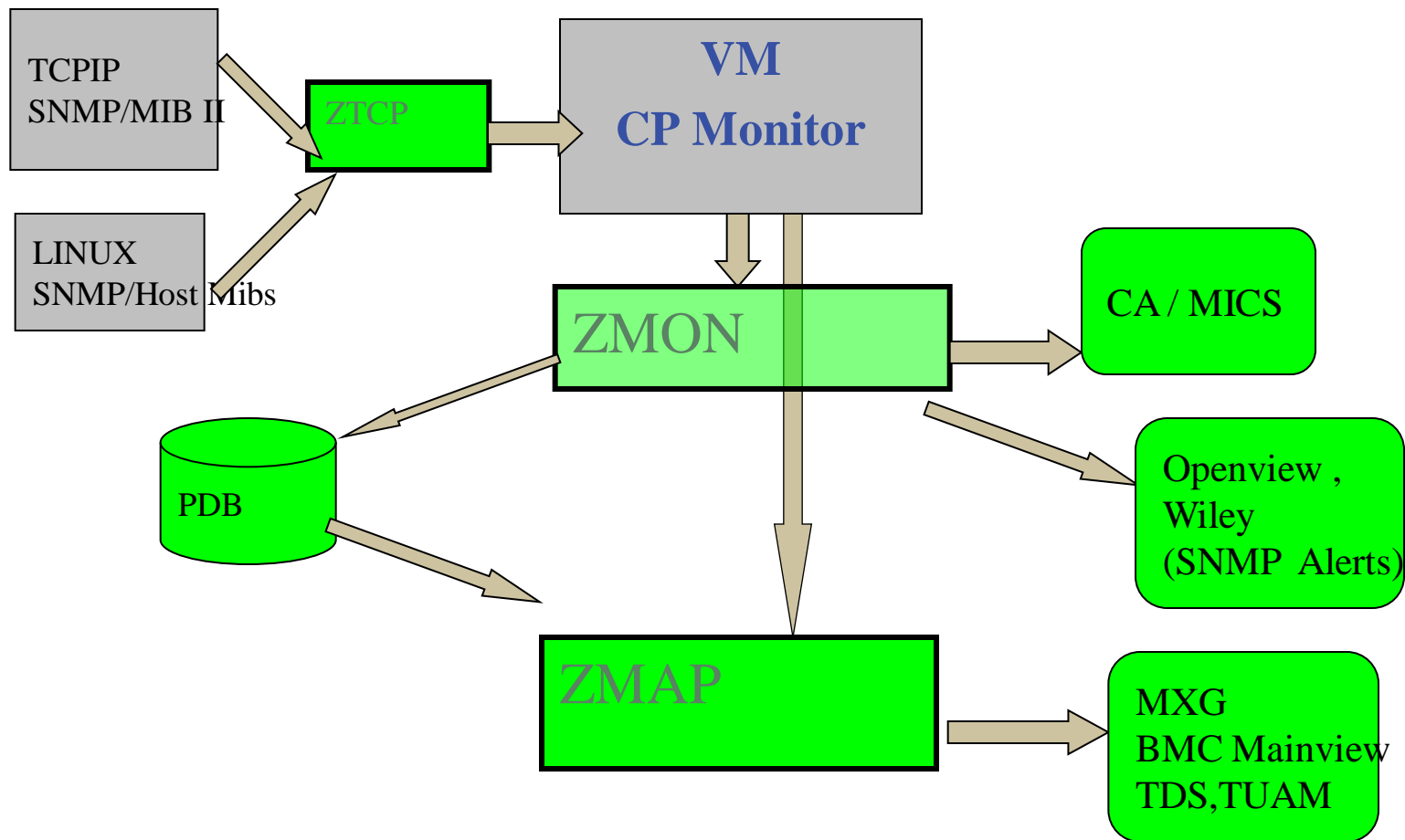
## ESATCP: Network Monitor

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

## Collects data from:

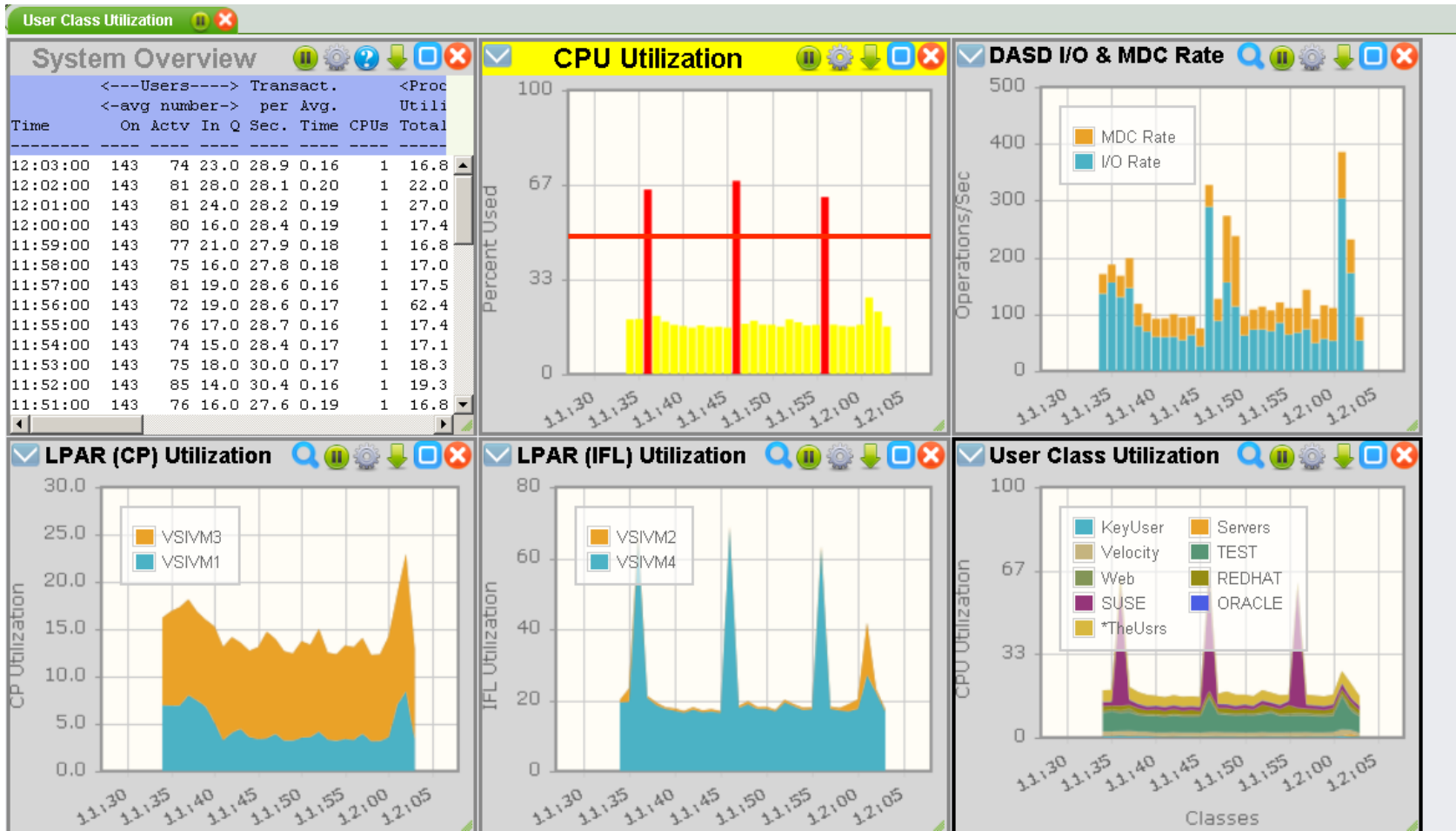
- LINUX (netsnmp)
- NT/SUN/HP (native snmp)
- Printers/Routers....

# Add Enterprise Support for capacity planning tools



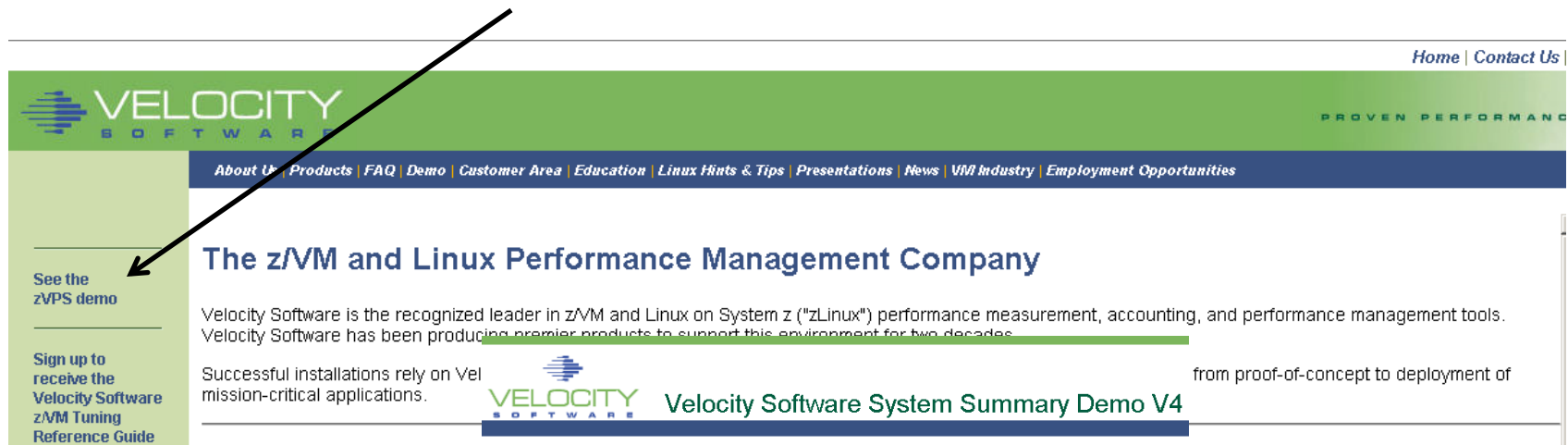
# What we're doing for Capacity Planning

- CPU by lpar by Processor type
- CPU BY userclass



# See what we're doing for Capacity Planning

- [VelocitySoftware.com](http://VelocitySoftware.com)
- See the demo



Home | Contact Us

VELOCITY SOFTWARE

PROVEN PERFORMANCE

About Us | Products | FAQ | Demo | Customer Area | Education | Linux Hints & Tips | Presentations | News | VIM Industry | Employment Opportunities

## The z/VM and Linux Performance Management Company

Velocity Software is the recognized leader in z/VM and Linux on System z ("zLinux") performance measurement, accounting, and performance management tools. Velocity Software has been producing premier products to support this environment for two decades.

Successful installations rely on Velocity Software from proof-of-concept to deployment of mission-critical applications.

Sign up to receive the Velocity Software z/VM Tuning Reference Guide

See the zVPS demo

VELOCITY SOFTWARE

### Velocity Software System Summary Demo V4

#### Demo System V4

Demo	12/03/13	05:31	044B42-0	22.30%	
Linux Nodes (z/VM-Guests)					
suselnx1	83.08%				
roblx1	0.59%				
broblx1	0.59%				
redhat5x	0.58%				
redhat6	0.54%				
sles11x	0.47%				

#### Demo System V3.5

Demo V3	12/03/13	05:31	044B42-0	22.30%	
Linux Nodes (z/VM-Guests)					
suselnx1	83.08%				
broblx1	0.59%				
roblx1	0.59%				
redhat5x	0.58%				
redhat6	0.54%				
sles11x	0.47%				

# 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 saves money

## Storage: No good guidelines

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