If it can go wrong, it will

How not to bring a customer in production

7th European IT Technical University of users exploiting z/VSE, z/VM and Linux on System z.

30.09.-02.10.2013

Hotel Le Royal Méridien Hamburg

Hans Picht,
System z Technical Sales Manager
Central & Eastern Europe
hans@tr.ibm.com

Background

- The story told here is based on a real customer
- They are an existing z/OS customer
- We are Working with a Business Partner
- And the plan sounds easy
- No big deal and not very exotic
- We just want to consolidate a couple of x86 servers running Oracle Databases

Don't Try This at Home

An ongoing Journey

2010:

This might be a good prospect to talk about Server consolidation & IT Optimization

2012:
30 guests with various Oracle DB's running in Production (100 more to go...)

We need some charts for a customer Meeting

Add Some Charts Here

Transzap
Fuels competitive edge with increased application uptime from IBM System z

Business challenge:

Transzap offers its customers a comprehensive suite of financial software tools. As a small business with tens of billions of dollars in client transactions flowing through their systems each year, Transzap needed an economical, reliable platform to provide clients with high availability while enabling the capacity to accommodate growth within their software as a service business model.

Solution:

Transzap decided to consolidate on an IBM System z platform to provide the stability and scalability needed to accommodate triple digit volume growth, enabling them to focus on the business of software innovation. Transzap migrated to System z and virtualized its critical applications on Linux on System z, a platform that supports Transzap's dynamic Java™ and **Oracle** environments.

Benefits:

- *Long-term cost savings, including savings realized through the virtualization of Oracle licenses*
- *Transzap is now able to create new Oracle database instances over a period of two or three days.*
- *Provides higher levels of uptime for their customers*
- *Offers peace of mind through 24x7 world-class hardware support*

> *"We intend to deliver a 99.9% application uptime guarantee to our customer base, thanks to the availability characteristics of System z."*
>
> *— Peter Flanagan,*
> *CEO of Transzap, Inc.*

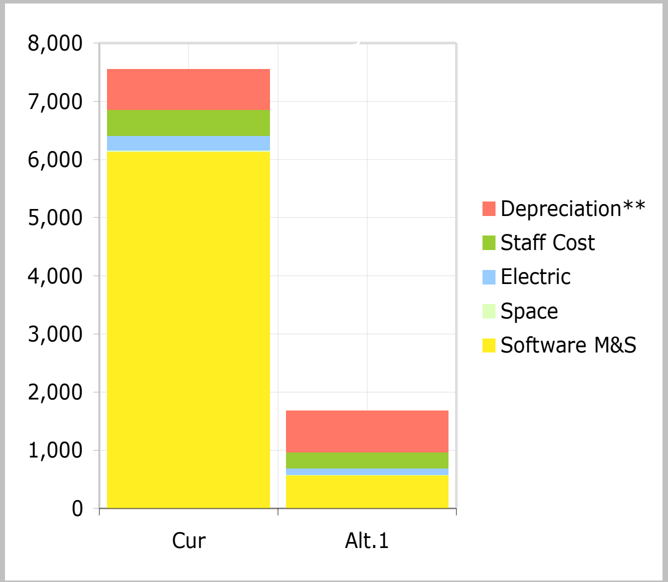
Solution components:

- IBM System z
- Linux on System z
- z/VM

IBM zLinux vs. x86 Consolidation Study – Save ~$6M over 5 Years (1)

Potential cost savings projections below are based on modeling a US Financial Institution's current state data for their Oracle DB environment running on x86/Linux vs. Linux on zEnterprise

Sizing 3	Current	AltCase1 9:1	Change
Server Type	Mixed - x86	**z196-ELS-1bk**	
Total Cores/ IFLs	352	6	-98%
Used Cores/ IFLs	352	6	-98%
Total Sockets/ IFLs	153	6	-96%
#Logical Servers	53	53	0%
#Physical Servers (or #IFLs)	51	6.00	-88%
Total RIP Capacity(installed)	275,129	27,464.6	-90%
Total RIP Workload(used)	22,233	22,233.1	0%
Ave %Utilization	8%	81%	
Estimate # Network Ports	103	4	
Annual Operating Costs (AOC)			
Software M&S	$1,226,324	$113,424	-91%
Hardware Maint*	$0	$0	0%
Space	$4,297	$1,543	-64%
Electric	$49,901	$21,574	-57%
Staff Cost	$90,167	$54,512	-40%
Depreciation**	$140,525	$144,309	3%
Total AOC	**$1,511,214**	**$335,362**	**-78%**
Est Potential Savings /Yr		$1,175,852	
5 Year Projection			
OTC + 5x AOC	**$7,556,070**	**$1,676,809**	
5 Yr Savings		$5,879,261	



(1) Notes:

- Existing server utilization based on customers reported distributed server utilization rates
- Financial results based on 5 year depreciation mode l and include IBM System z ELS bundle (including HW, HW maintenance and virtualization software costs)
- RIP = Relative Indicator of Performance (across platform) and is based upon 3rd party and IBM observed performance analysis

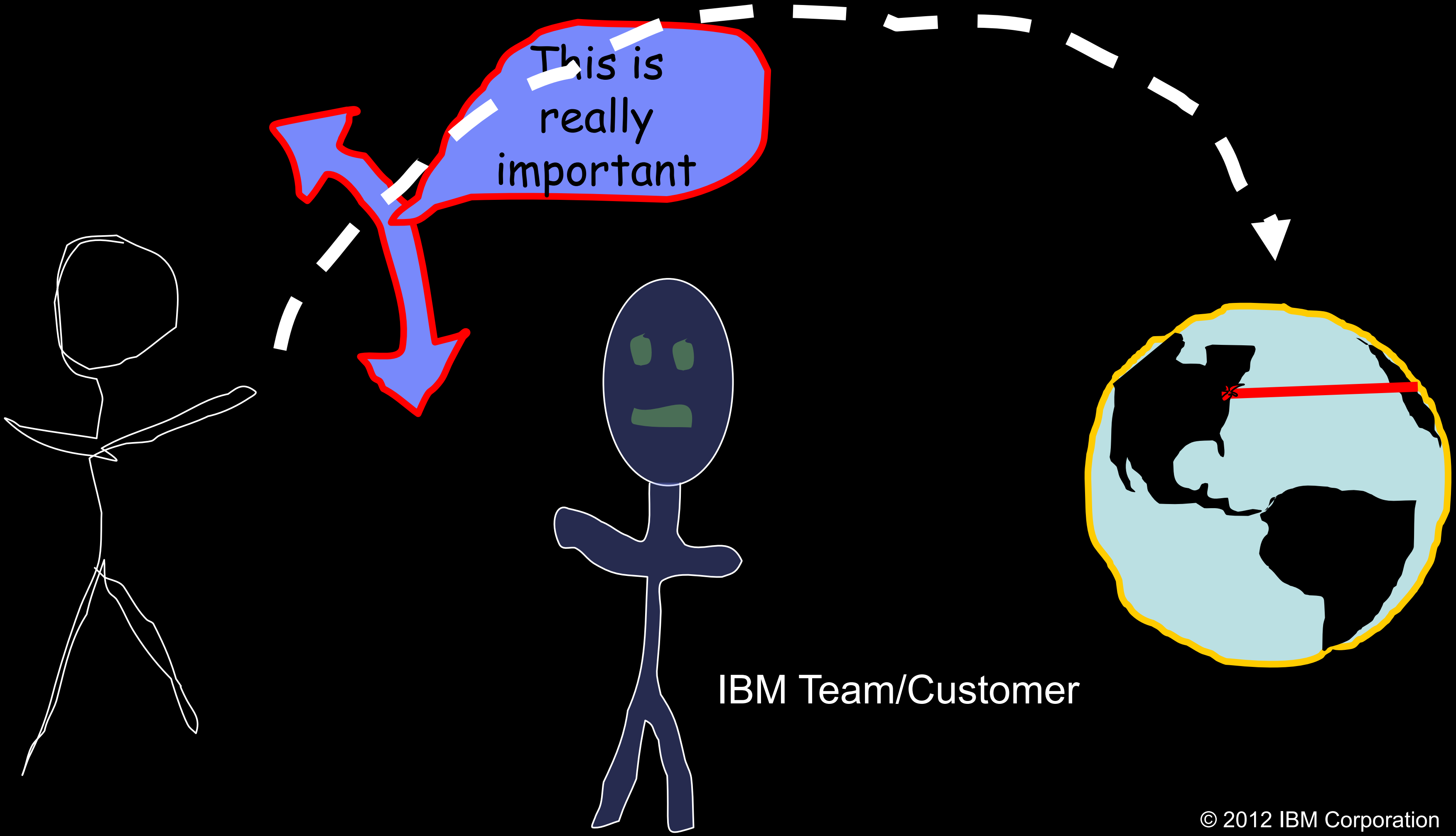
Project Progression: Q1 & Q2 2010

- It is the same Linux, just on a different architecture
- It is the same database just on a different architecture
- We have done this thousands of times
- No big deal: Export there, Import here and we are done
- Linux on System z is compatible with all major storage vendors
-

Customer: Proof that zLinux does not effect our z/Os Installation

- Objectives
  - Demonstrate the viability of the consolidation approach and prove that this will not have a detrimental impact on the business critical application workload in the z/OS environment.
  - Demonstrate the viability of the proposed consolidation of the selected Red Hat Linux based Oracle database servers.
- Basic PoC
  - Prove that the Linux on System z environment will not have detrimental impact on the z/OS system and application.
  - Drive utilization of Linux on System z LPAR to more than 90%.
  - Monitor z/OS and application environment to determine that no detrimental impact experienced.
  - Note: this PoC will not include Oracle. Its sole purpose is to demonstrate the superior workload isolation capabilities between different LPARs.

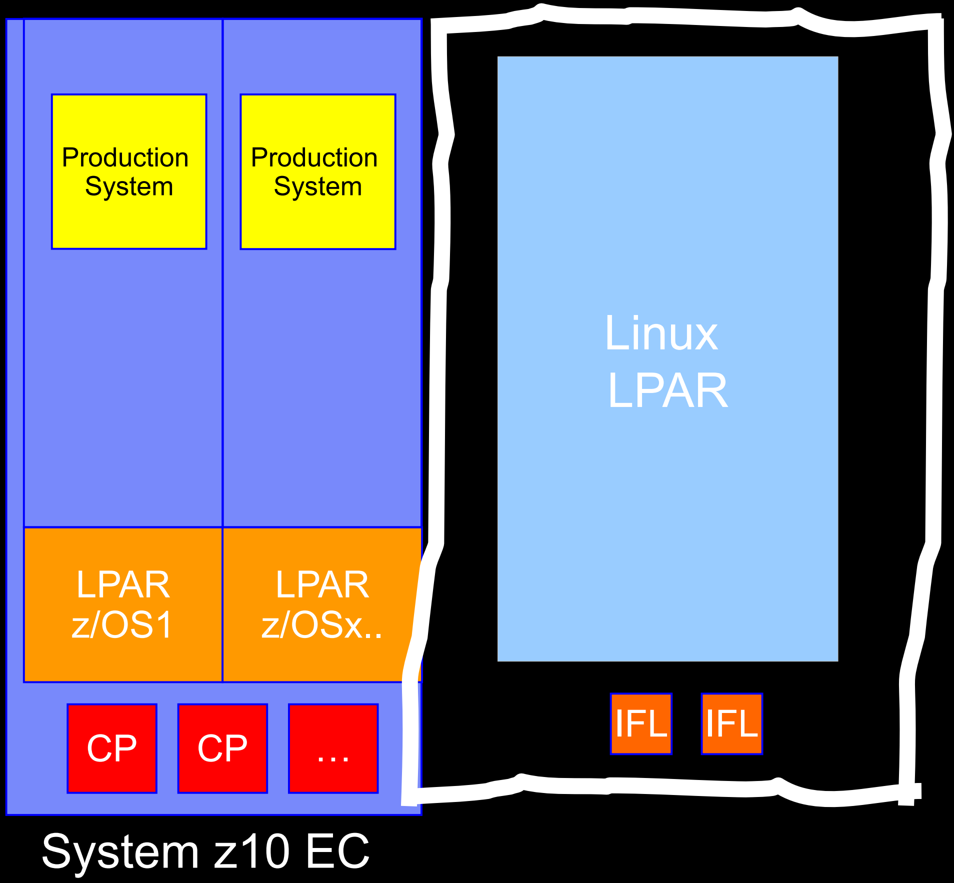
Let's start with the PoC on Monday...



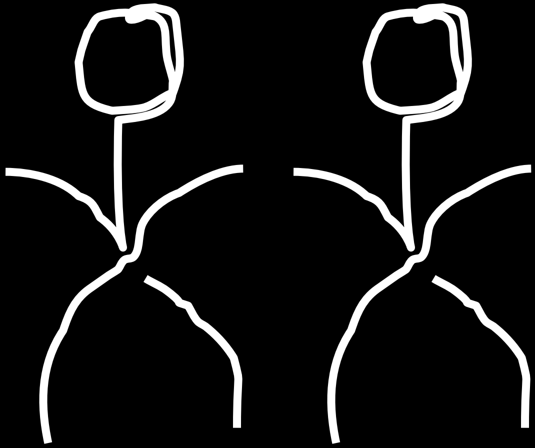
IBM Team/Customer

Basic PoC set-up, currently installed (ready for test)

- Additional memory and second IFL installed on temporary basis.



First Delays



Network	✗
Hardware	✓
Architects	✗
Administrators	✓
Storage	✓
Security	✗

(including network security)

More Delays

- Linux on System z installed and ready for Basic PoC test.
- Waiting for client (critical person was on vacation) to put test load on z/OS – LPAR to measure influence of loaded Linux LPAR.
- It is assumed that this part can be handled by client personnel, as all set up was done. Installation was performed by Hans-Joachim Picht.
- The customer knows how to put the Linux system to 100% IFL utilization and where to obtain the critical performance data.

How to burn some CPU cycles on the IFL

```
root@localhost:~# for i in `seq 1 100`;
do  cat /dev/zero > /dev/null & ; done
```

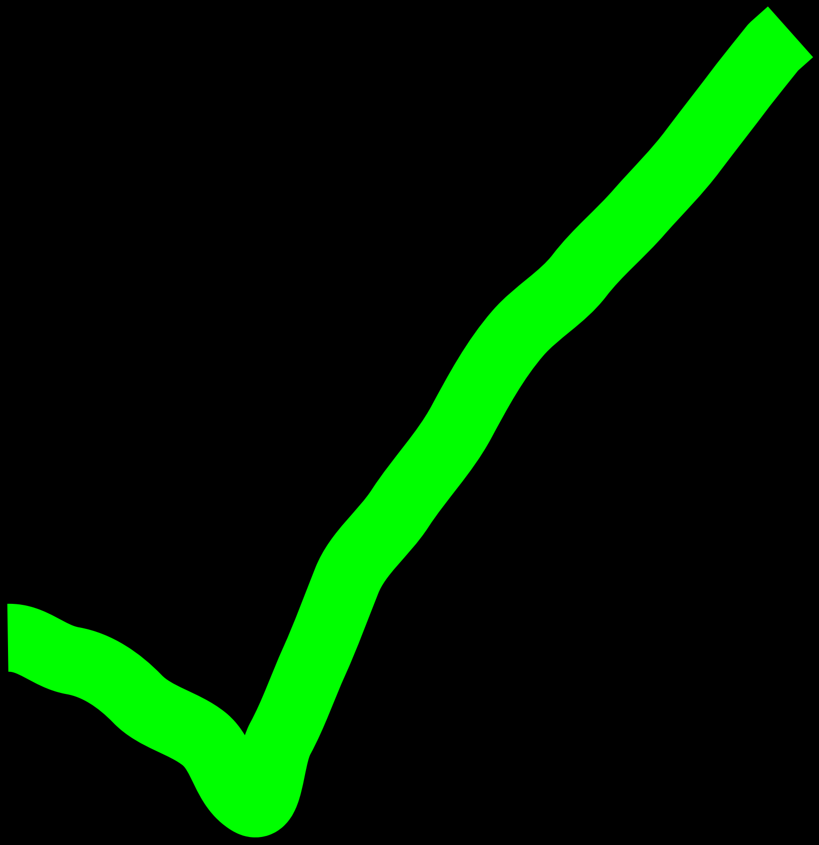
Test1

- During actual test/measurement periods, it was required to put load on the z/OS LPAR as well as to drive up utilization of the Linux LPAR.

- Appropriate tools for resource consumption analysis were deployed on the z/OS LPAR (for example: SMF and/or RMF on z/OS) to validate that there is no detrimental impact when Linux LPAR utilization is increased.

- In the first runs we could not see any results because the test where performed on the D/R z10 where the customer could only drive the z/OS application to 2-5% system utilization

Test1: Passed



So Linux runs in an LPAR! Now we want to see that is can also run under z/VM

Extended PoC

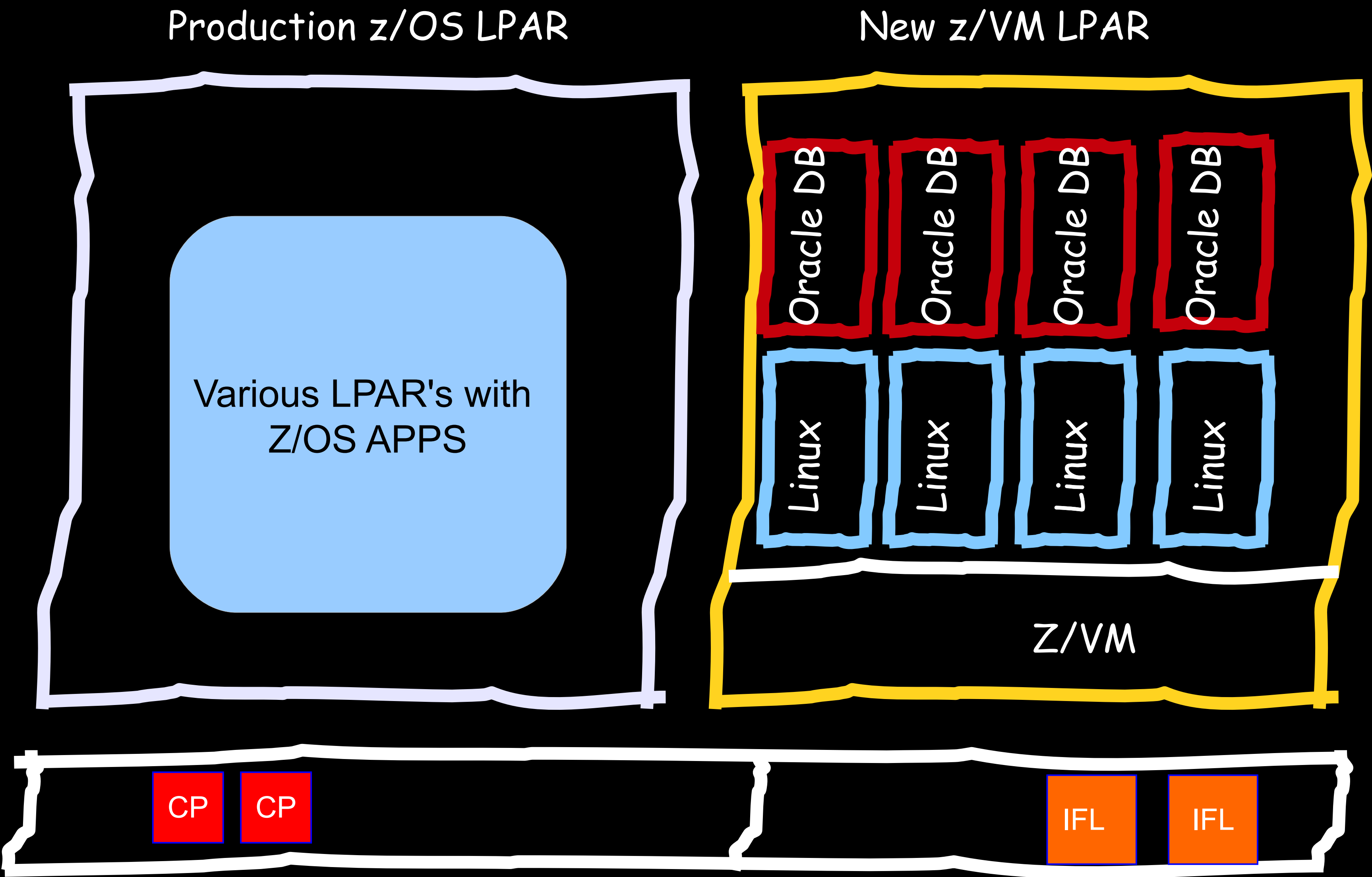
- Prove the viability of the proposed consolidation and reduce the risk of a later production implementation through proper testing with focus on Oracle.

- Provide basic functional verification of Oracle DB servers with Linux on System z.

- Demonstrate Oracle DB behaviour under load conditions.

- Demonstrate the viability of migration from (back-level) Oracle 9i DB and Oracle 10g on distributed (back-level) RHEL 4 platforms to a current and supported environment on System z.

Current Hardware Configuration (z10EC)



Installation Challenges

- Poc on the D/R z10 EC with DS8000 storage.
- 32 Linux images (RHEL 5.4) with different Oracle DBs (imported to 10gR2) have been installed under z/VM 6.1.
- During the porting process the customer experienced an ABEND when trying to import multiple DBs in parallel.
- The reason was no enough memory - only 8 GB were defined.

Installation Challenges (cont.)

- After assigning the 32 GB (available for this PoC) the system ran ok, but they see a very high paging rate (and needed to define add'l DASD for paging - now 104 GB in total) at only about 20% processor utilization.

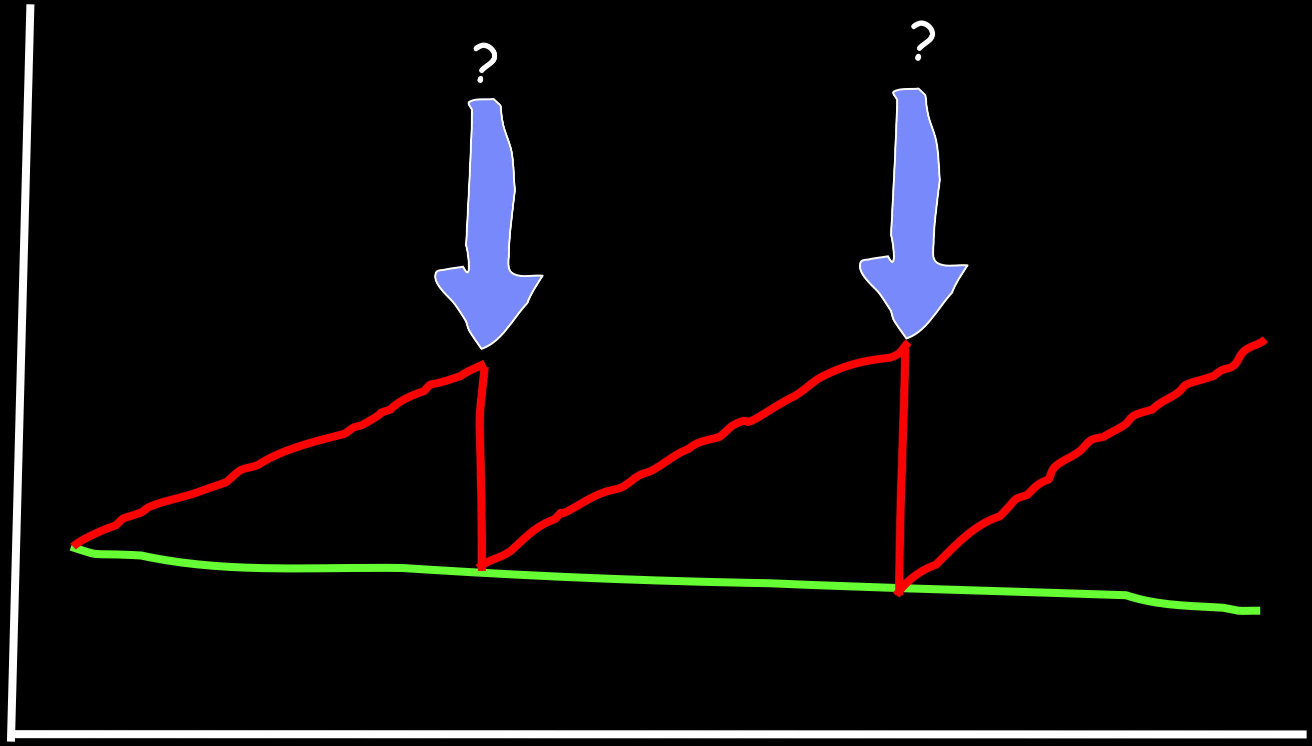
- Currently they have allocated 2 GB per Linux image, no Expanded memory defined.

- The customer is well aware that this is a PoC environment and that he will not be able to do realistic load testing as they are running in a test only environment.

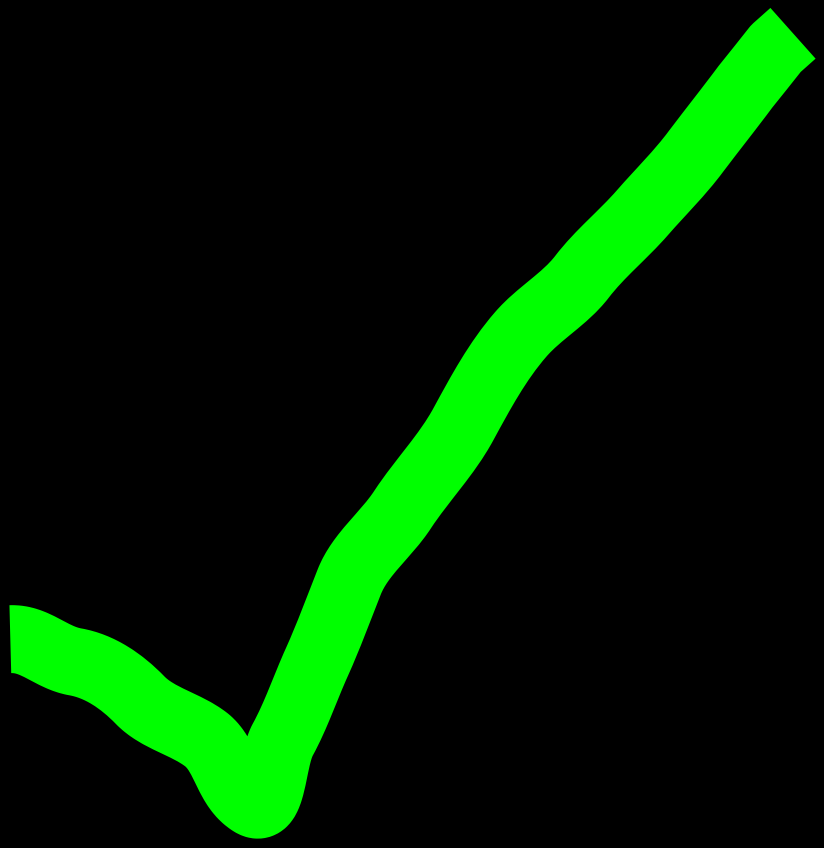
- Client is not concerned with the current performance, but - he wants to come up with a reasonable prediction of the needed memory size once he would start deploying DBs in production.

Test Plan?

- Just playing around with the system, no test or measurement criteria



Test2: Passed

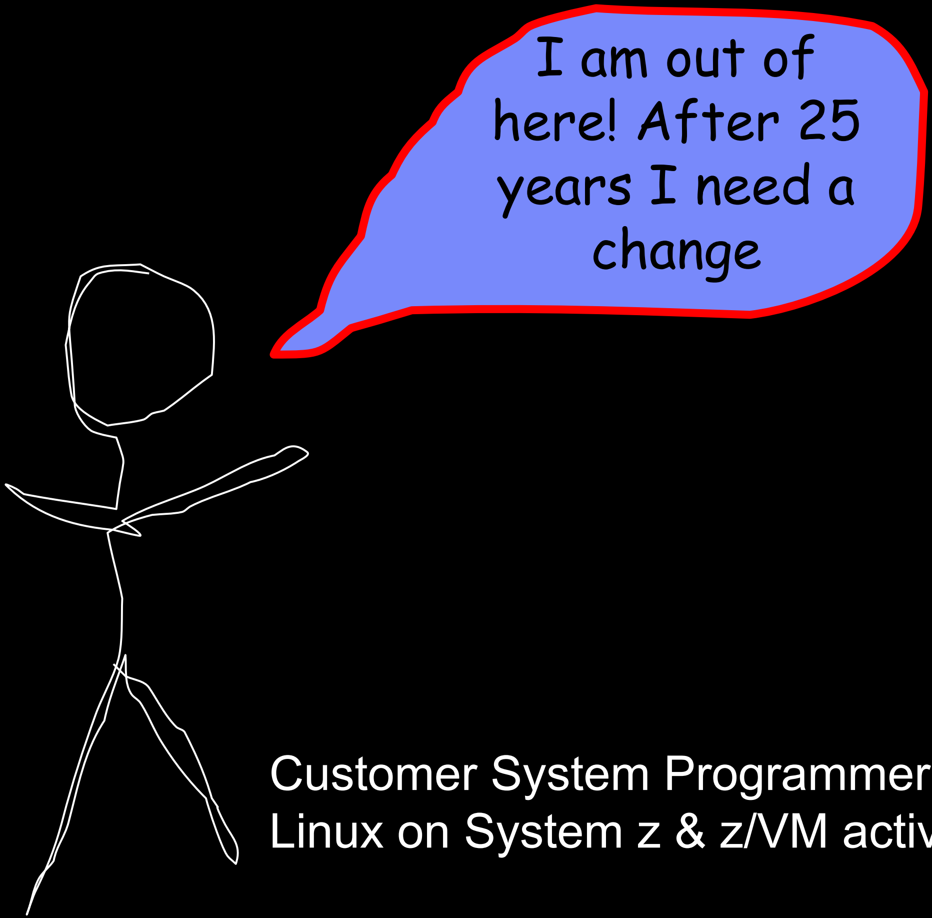


Statement of Work…..Ignored

Areas that are typically addressed by a SOW are as follows:

- Purpose
- Scope of Work
- Work
- Period of Performance
- Deliverables Schedule
- Applicable Standards
- Acceptance Criteria
- Special Requirements
- Type of Contract/Payment Schedule
- Miscellaneous

Sometimes people change jobs….

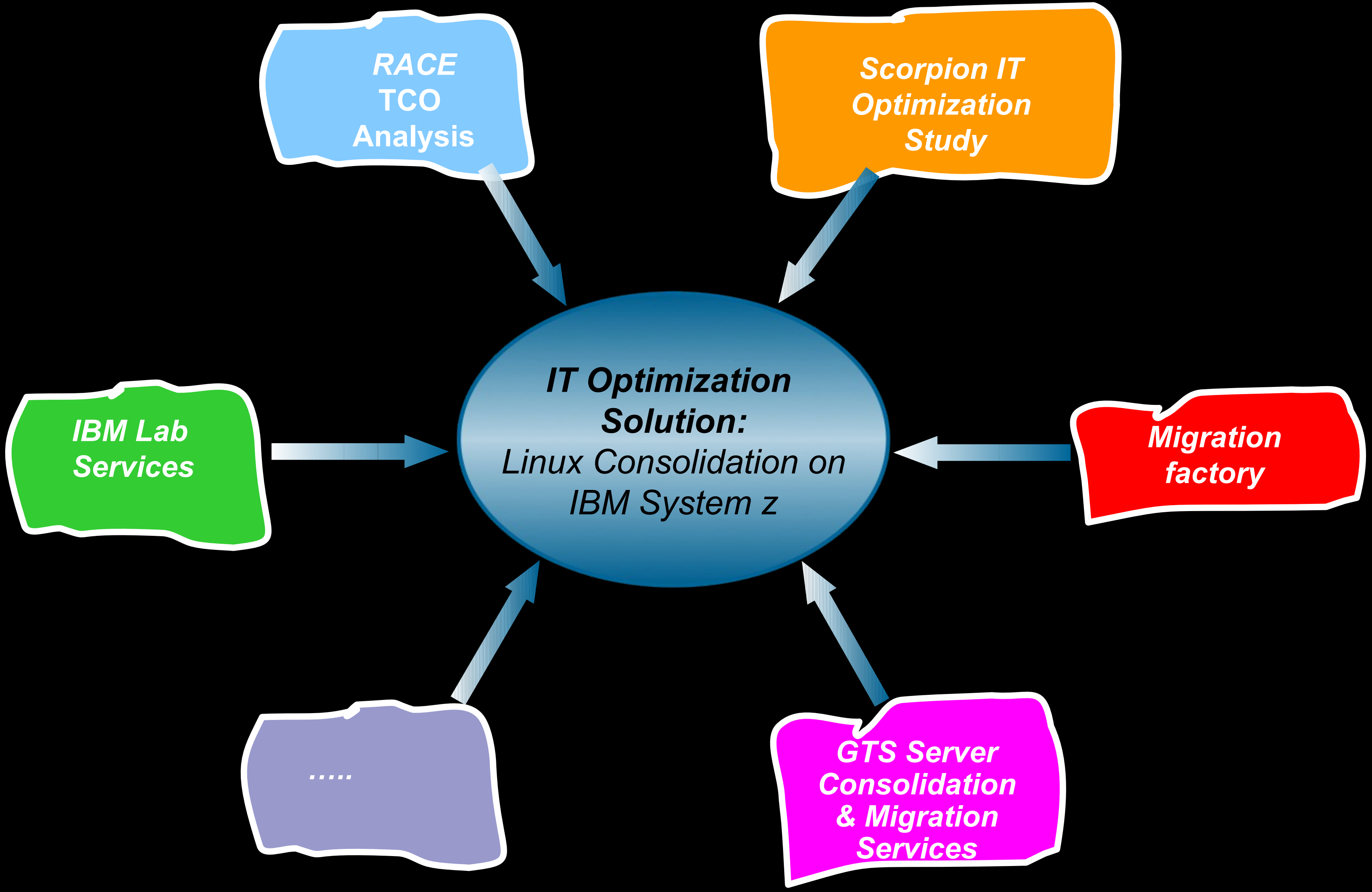


Customer System Programmer in charge of the Linux on System z & z/VM activities

Project Progression Stage II (Sep 2010)

- Meetings
- Workshops
- Studies
- "Foreign Clown from out of Town" Visits

Let's make a study



RACE TCO Analysis

Scorpion IT Optimization Study

IBM Lab Services

IT Optimization Solution:
Linux Consolidation on IBM System z

Migration factory

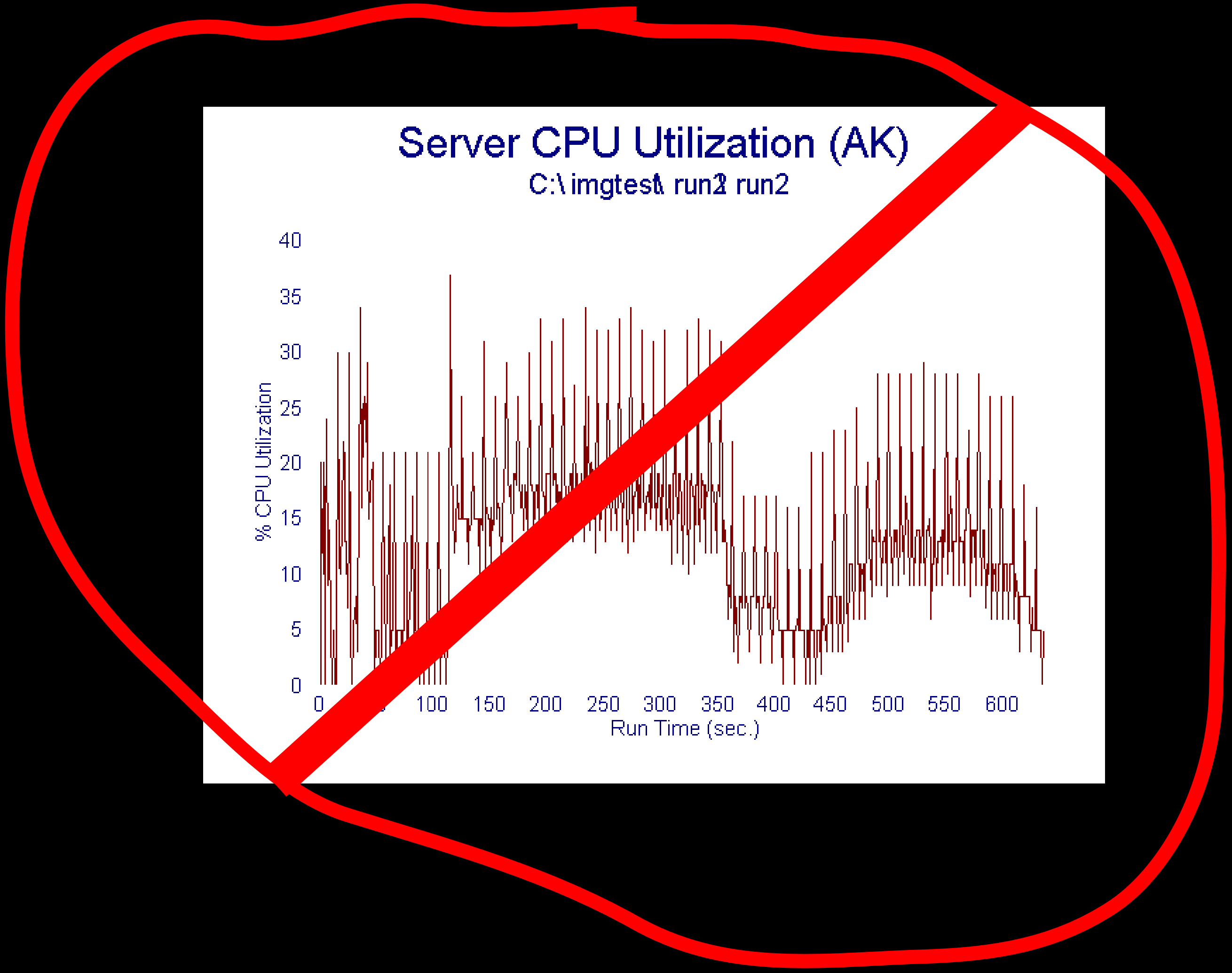
…..

GTS Server Consolidation & Migration Services

The Scorpion Results

- The IBM mainframe TCO study quantified potential for cost savings - in excess of **20 million dollars over five years.**
- 1. Customer has potential for savings in their existing environment without major new investments by:
  - Utilizing existing Integrated Facility for Linux*, IFL, into use for POC and later production.
  - Conducting the z/OS application fine-tuning exercise for potential longer term efficiencies.
- 2. The real business value comes through consolidating distributed servers onto mainframe.
  - The customer can be accomplished by adding capacity to the existing environment or by
  - updating onto newer technology.
  - Extended savings will be realized by utilizing the latest technology and has proposed

And of course.....no utilization data was available

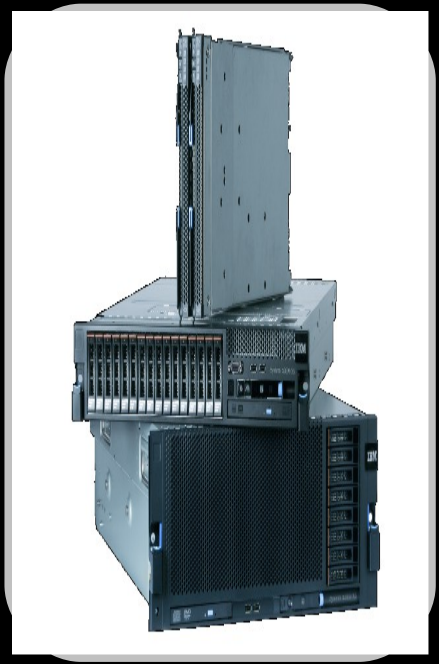


End of Q4/2010: Project Progression or how we compete with other IBM brands

End of Q4/2010: Project Progression or how we compete with other IBM brands

IBM Server

The
BAD



The
GOOD

The
UGLY

IBM Server

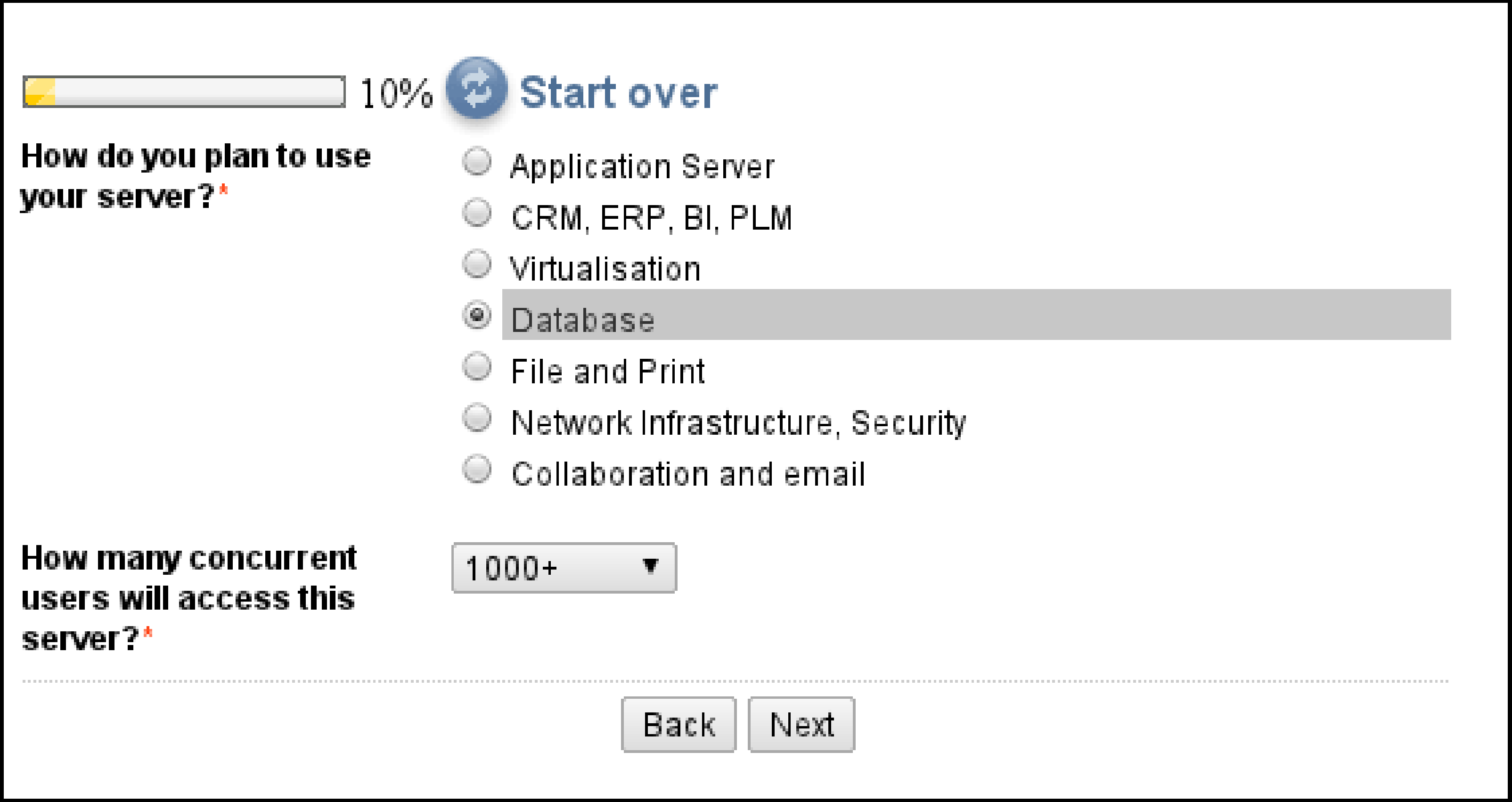
http://www-03.ibm.com/systems/express/sat/en_gb/index.html

IBM Systems Advisor Tool
Not sure which server or storage to choose? Find out here.

- Systems Advisor Tool will help midsize businesses find the right systems hardware while protecting their investments with flexible, scalable products that can grow as business grows.
- Not sure which server or storage to choose? By answering a few quick questions, we'll identify products that can help meet your business needs. Let's get started.

IBM Systems Advisor Tool

Not sure which server or storage to choose? Find out here.



10% Start over

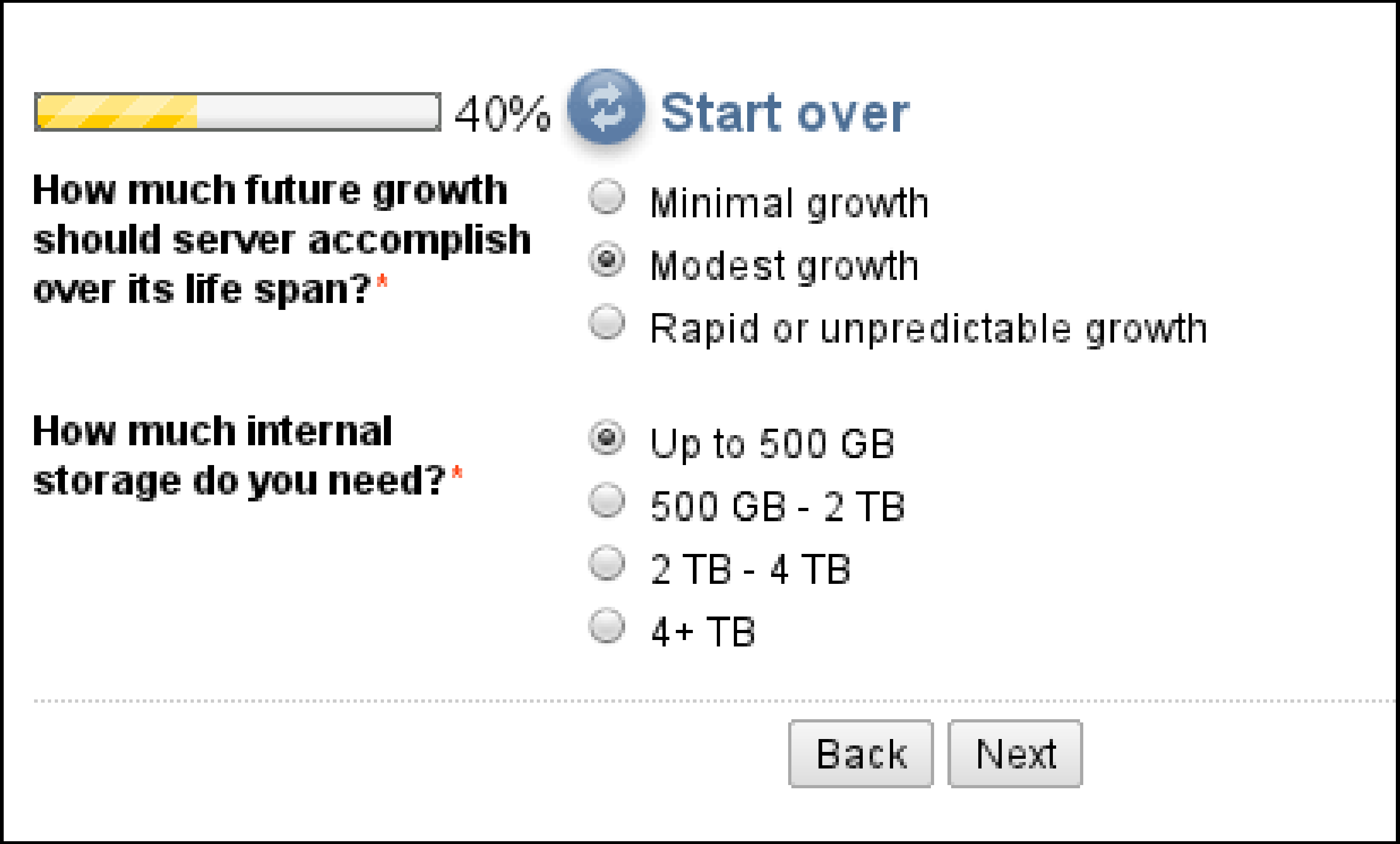
How do you plan to use your server?*

○ Application Server
○ CRM, ERP, BI, PLM
○ Virtualisation
● Database
○ File and Print
○ Network Infrastructure, Security
○ Collaboration and email

How many concurrent users will access this server?* 1000+

IBM Systems Advisor Tool

Not sure which server or storage to choose? Find out here.



How much future growth should server accomplish over its life span?*

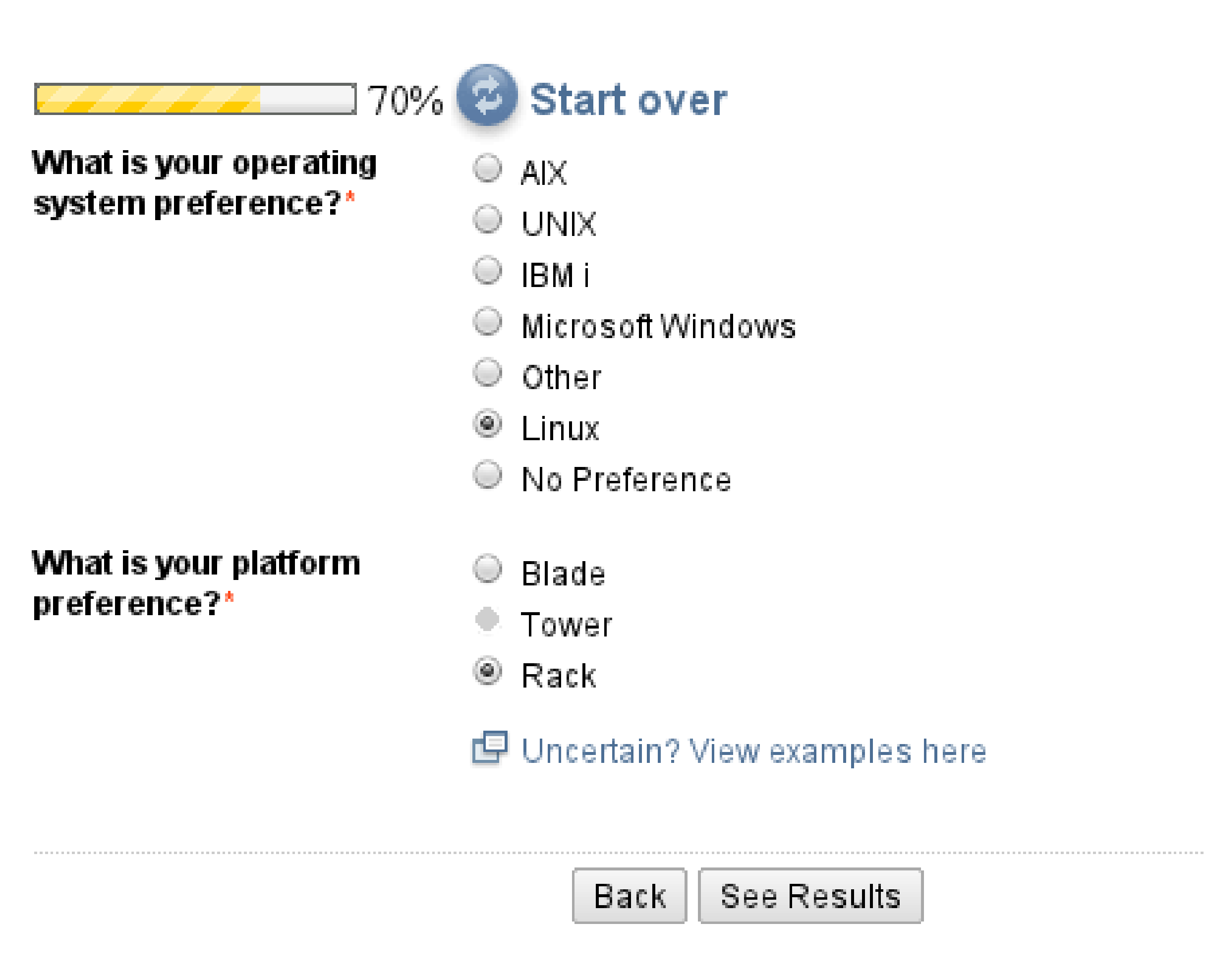
○ Minimal growth
● Modest growth
○ Rapid or unpredictable growth

How much internal storage do you need?*

● Up to 500 GB
○ 500 GB - 2 TB
○ 2 TB - 4 TB
○ 4+ TB

IBM Systems Advisor Tool

Not sure which server or storage to choose? Find out here.



70% Start over

What is your operating system preference?*

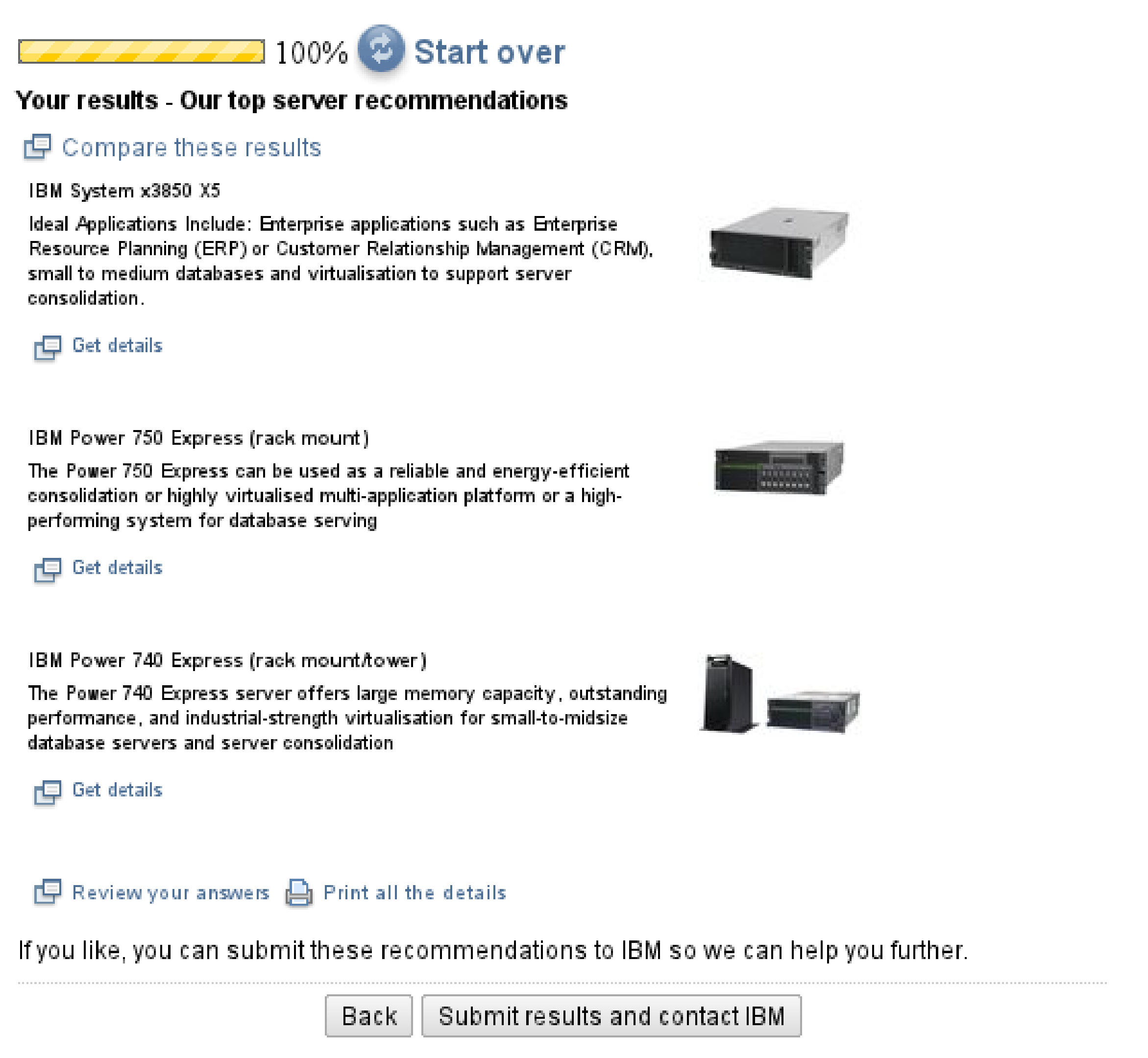
○ AIX
○ UNIX
○ IBM i
○ Microsoft Windows
○ Other
● Linux
○ No Preference

What is your platform preference?*

○ Blade
○ Tower
● Rack

Uncertain? View examples here

IBM Systems Advisor Tool
Not sure which server or storage to choose? Find out here.



100% Start over

Your results - Our top server recommendations

Compare these results

IBM System x3850 X5

Ideal Applications Include: Enterprise applications such as Enterprise Resource Planning (ERP) or Customer Relationship Management (CRM), small to medium databases and virtualisation to support server consolidation.

Get details

IBM Power 750 Express (rack mount)

The Power 750 Express can be used as a reliable and energy-efficient consolidation or highly virtualised multi-application platform or a high-performing system for database serving

Get details

IBM Power 740 Express (rack mount/tower)

The Power 740 Express server offers large memory capacity, outstanding performance, and industrial-strength virtualisation for small-to-midsize database servers and server consolidation

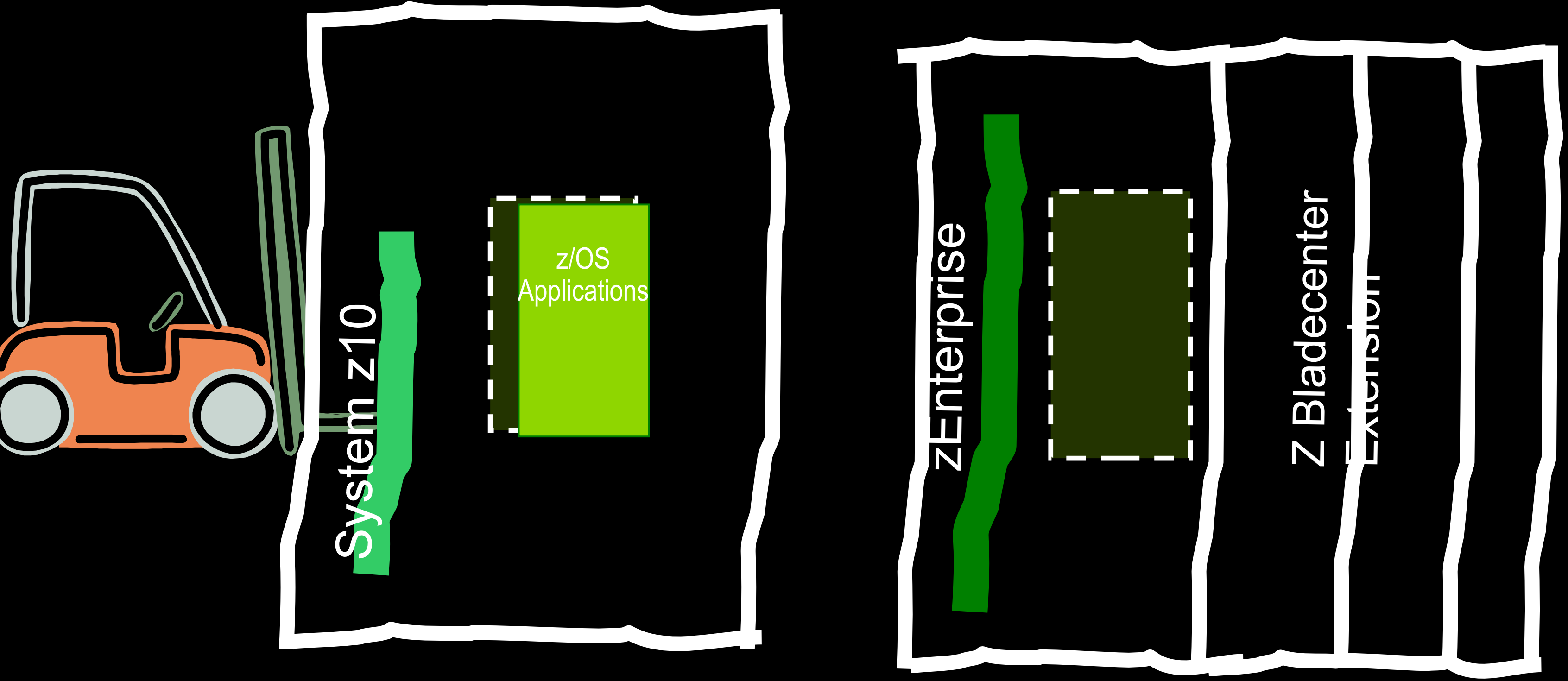
Get details

Review your answers Print all the details

If you like, you can submit these recommendations to IBM so we can help you further.

2010 Year End

Fork Lift Hardware Upgrade



IBM System z 10 EC

IBM zEnterprise

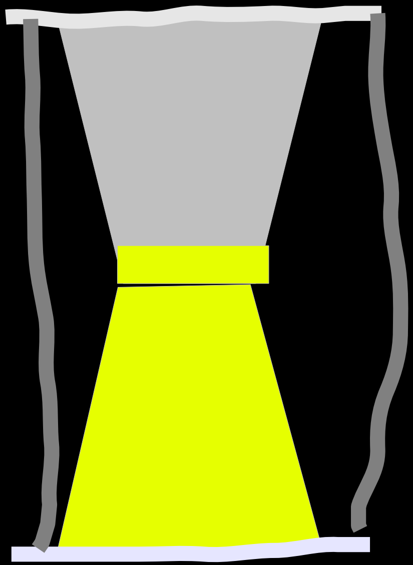
Some people are optimistic

- Mail from the account team/business partner on Feb2 2011:

> “We will start the Linux on System z Implementation Project in 2 Weeks from now”

Customer: We are moving our datacenter

Project Manager: Time is your Enemy



Just FYI: The data center move was postponed a couple of times and has still not happened today...

Project Progression: Architectural Workshop

Architectural Workshop : Executive Summary

- Following a TCO study, the customer showed an interest in a consolidation of 257 Oracle DB's running on distributed Intel servers on Linux on System z.
- First step was to run a Proof of Concept on site to demonstrate that Linux LPARs had no impact to their z/OS environment and that Oracle was running well on Linux on z
- This PoC, done in 2010, was successful, so the customer chose to carry on the project with a design workshop on high availability and to ask for an IBM proposal for the actual migration (out of scope of this document, dealt separately)
- The IBM Oracle Center in collaboration with Linux team from the IBM Lab in Böblingen, Germany, ran the design workshop with the customer IT team from 14th to 16th February

Existing environment : Overview

- 2 sites, with a z196 server in each (primary DC and D/R)
- New DC in construction 30miles distance
- 257 Oracle DB on 123 Intel servers to migrate, All single instance (no RAC is installed)
- Some DB are clustered with Veritas, on the 2 sites (with automated or manual failover)
- Back up strategy Oracle Recovery Manager (RMAN) + IBM Tivoli Storage Manager (TSM) cold & hot back up
- Backup policy: daily to monthly, Restore : ~ 1GB/min
- Most of the DB are 10g, the ones that are 9i should be migrated
- Disaster recovery : no Oracle Data Guard (DG) in the Linux DB (DG is used within the customer with other OS)
- Hitachi Storage sub system with no more free space available
- Network LAN -> 10Gb, SAN -> 2Gb / 4Gb

Existing environment :

- Datacenter 1
  - 38,8 TB Data + 6,1 TB replicated DB to 2nd site
  - 36 DB category 1
  - 67 DB category 2
  - 103 DB category 3
- Datacenter 2:
  - 12,8 TB Data + 23,8 TB replicated 1st site
  - 6 DB category 1
  - 16 DB category 2
  - 40 DB category 3

SLA (Service Level Agreements)

RPO is very important
Near zero data loss for most of the DB
We understand this is Oracle responsibility as it is Oracle DB (redo logs, archive logs, commit, partial commits…)

Category 1 DBs: Business operations
RTO 5 min (critical DB)
RTO 30 min (other DB)
Daily backup

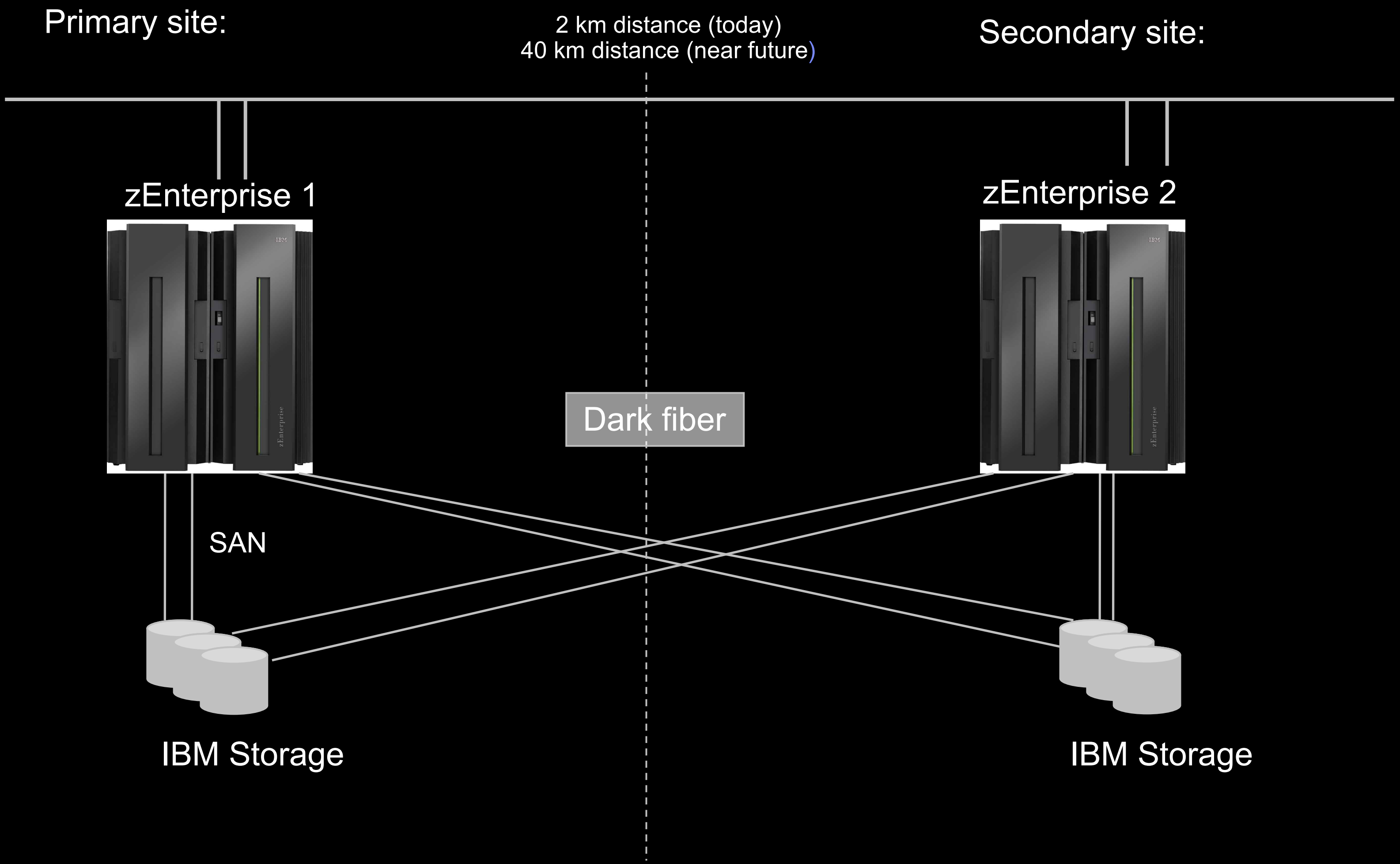
Category 2 DBs: Financial
RTO 30 min (critical DB)
RTO 3 h (other DB)
Daily backup, incremental for the bigger ones

Category 3 DBs: HR and DWH
RTO 1 day
Backup: Daily, weekly or monthly for big ones (depends on size)

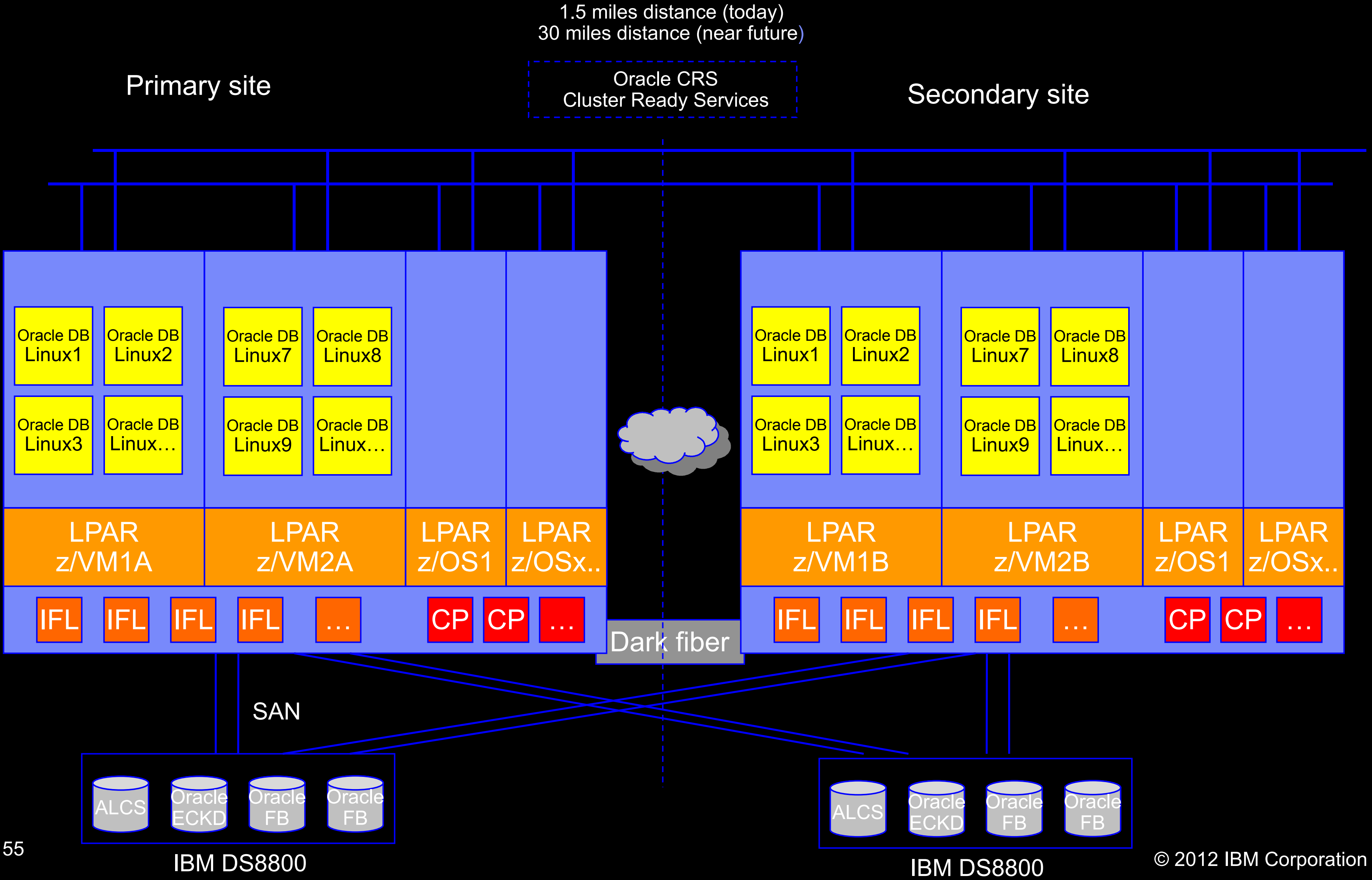
Requirements

- Oracle RAC (i.e. active / active clustering of databases) and Oracle Dataguard are excluded of the scope of this project
- DR is excluded of the scope of the project (not to mix between HA & DR)
- Regarding the storage sub system, there is an IBM proposition on going for replacement but for this exercise we should consider Hitachi
- **To be confirmed**
- There will be no database consolidation (no instances consolidation and no changes in the number of instances per OS)
- 123 physical servers will be transformed in 123 Linux virtual machines

Current Mainframe High level Overview

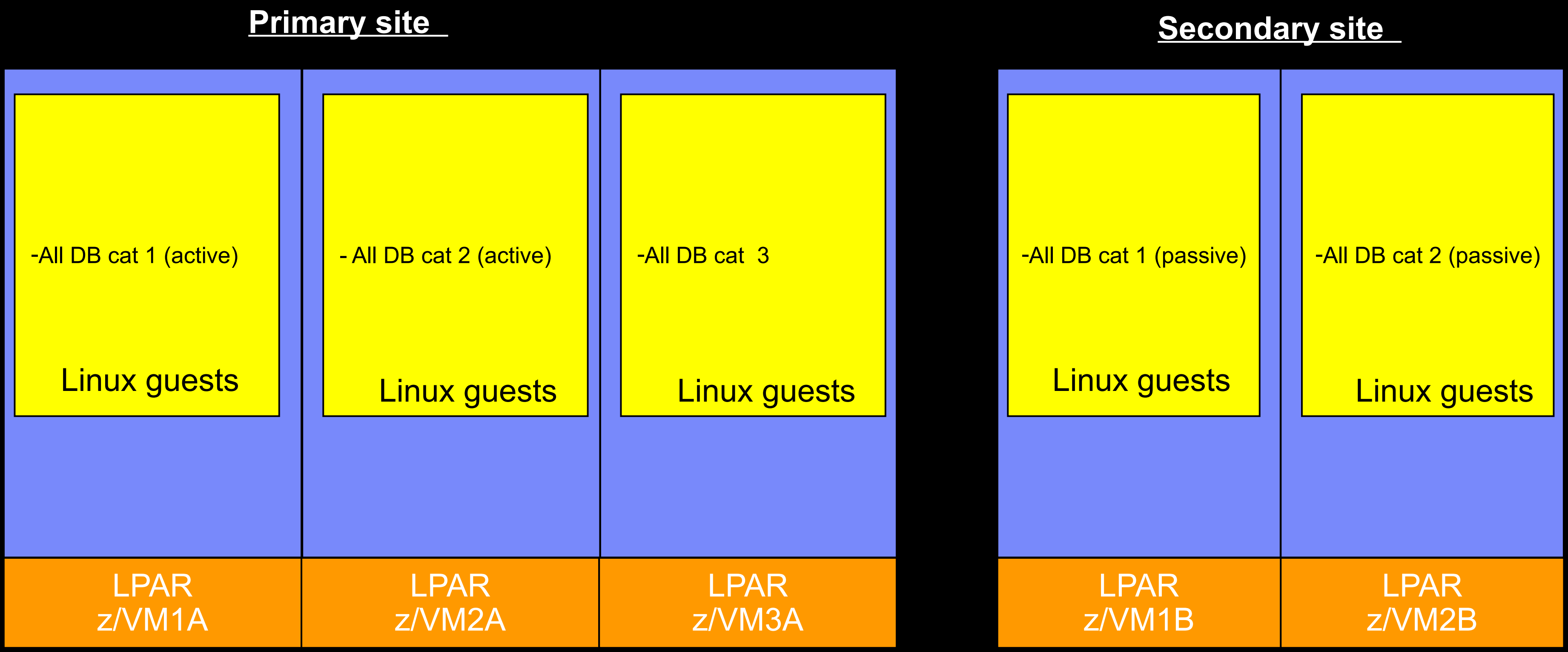


Target Architecture



Target Architecture DB repartition option 2 (active/passive mode) → Flavour 2: Changes in all tiers protections

One of the trends of the customer would be to transform the level of some DB:
For category 1 and 2
All the DB become protected DB
For tier 3 DB : all the protected DB become, regarding business needs :
Either protected tier 2 DB
Either single tier 3 DB



More Remarks

- No information was provided regarding the applicative landscape and architecture → out of scope
- Active / Passive Clustering options for Oracle DB workloads on Linux on z
  - Oracle CRS = high availability (Dataguard is more for disaster recovery)
  - RedHat cluster suite, not available on System z as of today
- 9i DB are not part of the scope if they can't be migrated (9i is not recommended on Linux on z)
- For the migration a large amount of additional storage is required (to be determined with Migration Factory team) and the customer will not have enough existing storage for this operation (no more free storage is available)
- Technical recommendations:
  - Performance: IBM recommends not to above
    - 10 IFL per z/VM partition
    - 200 GB Memory per z/VM partition
  - Storage: recommendation is a mixed configuration (possible in IBM storage):
  - Monitoring : To monitor the z/VM environment, recommendation is to use the Performance Tool Kit

Pricing, discounts, corefactors or why the oracle sales rep is not to keen to see his products running on his client's ifls....

The IBM Oracle Alliance

Pricing, discounts, corefactors or why the oracle sales rep is not to keen to see his products running on his client's ifls....

End of Q1

Now that the client purchased the IFL's...how many projects to we have right now?

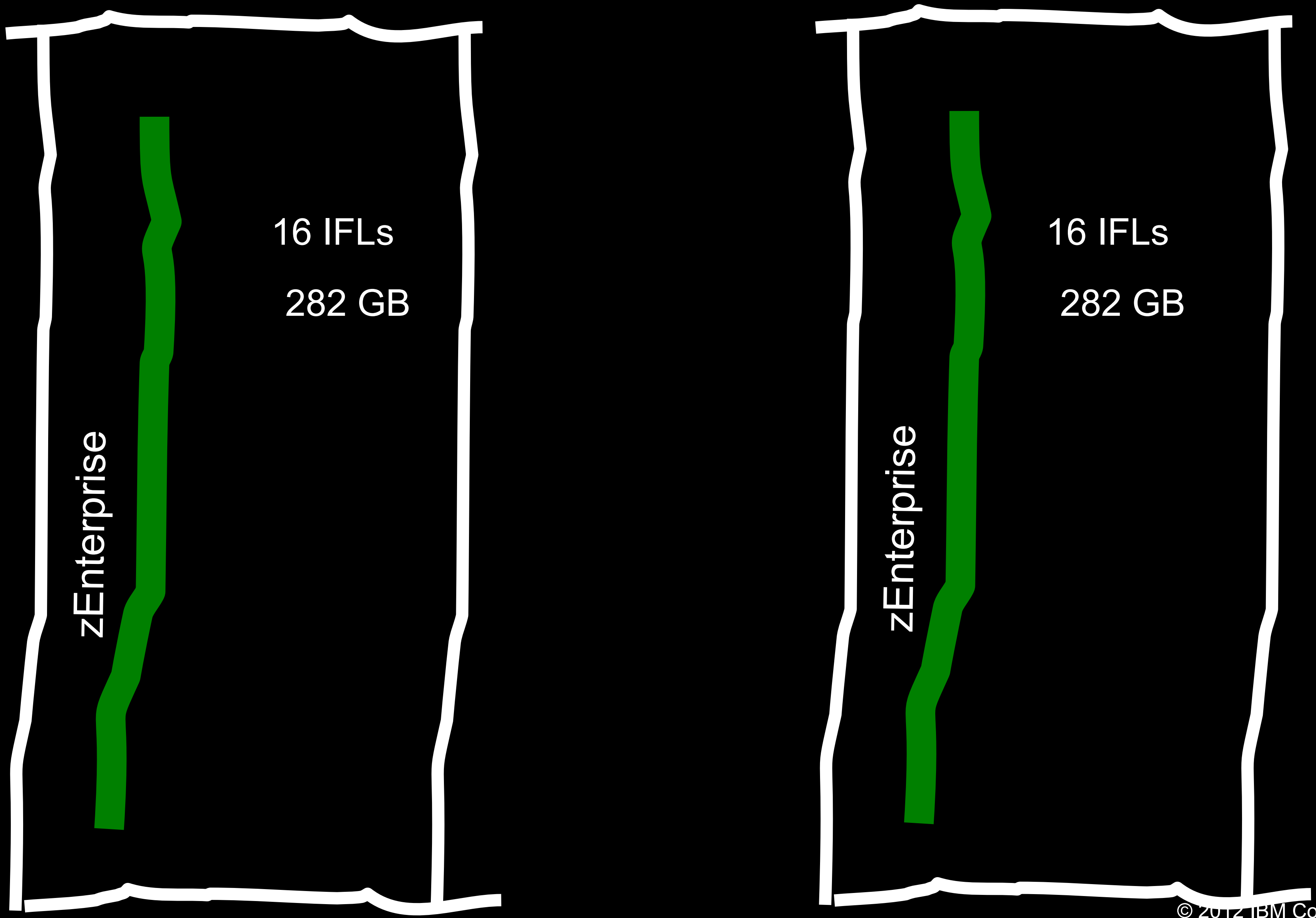
- Storage Migration
- Hardware Upgrade 2x z10EC → z196
- Linux on System z & z/VM Implementation (infrastructure)
- Oracle Migration Project

Migration Factory Workshop

A few little details

- Availability of the Migration Factory
- Working from Remote
- Time has to be scheduled in advance

This is the Hardware



zEnterprise

16 IFLs

282 GB

zEnterprise

16 IFLs

282 GB

z/VM Setup

LPAR1

28 zLinux Guests

(42 DB Category 1 DBs)

Business operations

RTO 5 min (critical DB) RTO 30 min (other DB)

Central Storage : 88GB

Expanded Storage :2GB

DASD PAGE : 360GB

IFL : 16 Shared

LPAR2

73 zLinux Guests

(116 Category 2 DBs: Financial)

RTO 30 min (critical DB) RTO 3 h (other DB)

Central Storage : 138GB

Expanded Storage :2GB

DASD PAGE : 560GB

IFL : 16 Shared

LPAR3

36 zLinux Guests

(76 DB Category 3 DBs)

HR and DWH

RTO 1 day

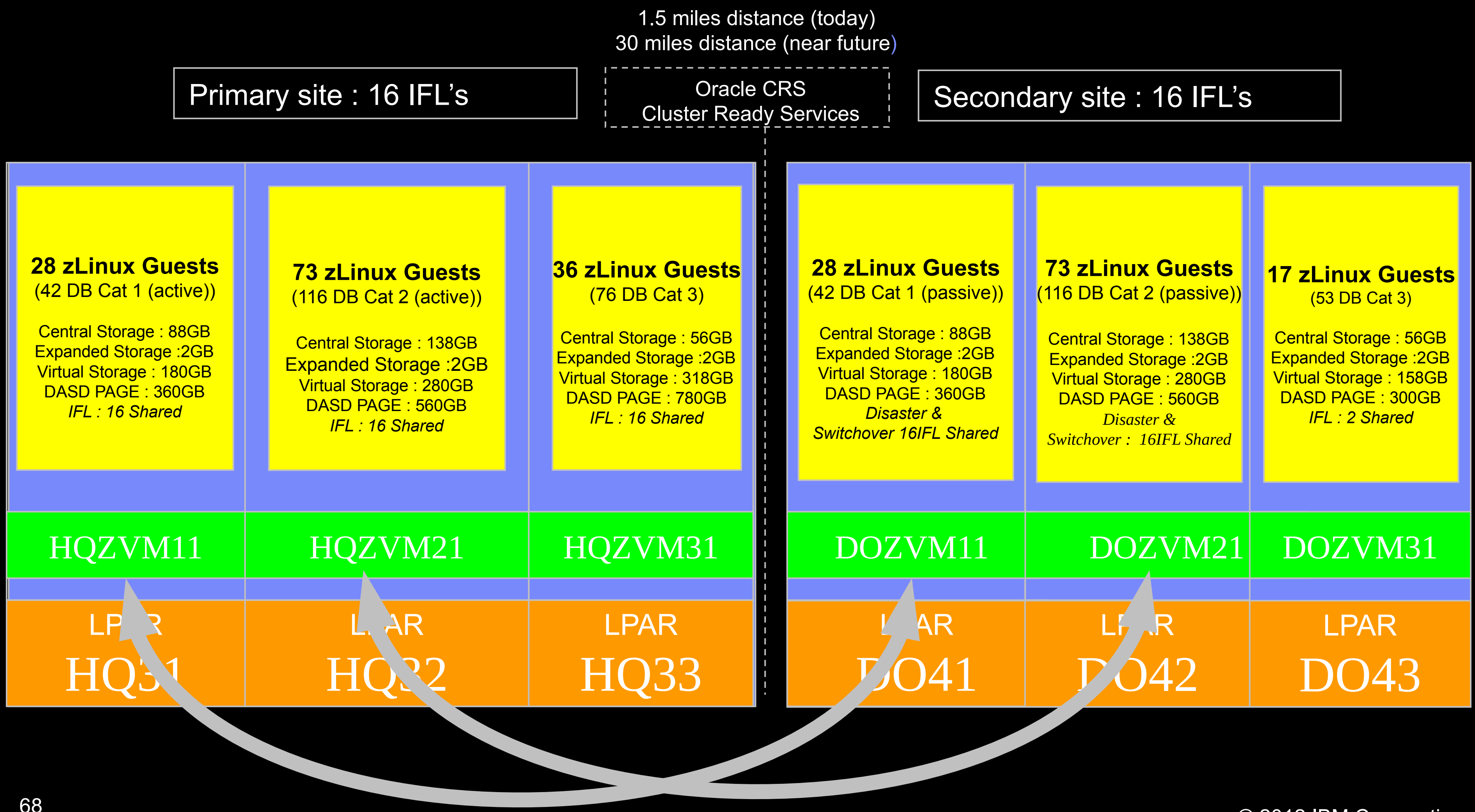
Central Storage : 56GB

Expanded Storage :2GB

DASD PAGE : 780GB

IFL : 16 Shared

And this is new the high level architecture



1.5 miles distance (today)
30 miles distance (near future)

Primary site : 16 IFL's

Oracle CRS
Cluster Ready Services

Secondary site : 16 IFL's

28 zLinux Guests
(42 DB Cat 1 (active))
Central Storage : 88GB
Expanded Storage :2GB
Virtual Storage : 180GB
DASD PAGE : 360GB
IFL : 16 Shared

HQZVM11

LPAR
HQ31

73 zLinux Guests
(116 DB Cat 2 (active))
Central Storage : 138GB
Expanded Storage :2GB
Virtual Storage : 280GB
DASD PAGE : 560GB
IFL : 16 Shared

HQZVM21

LPAR
HQ32

36 zLinux Guests
(76 DB Cat 3)
Central Storage : 56GB
Expanded Storage :2GB
Virtual Storage : 318GB
DASD PAGE : 780GB
IFL : 16 Shared

HQZVM31

LPAR
HQ33

28 zLinux Guests
(42 DB Cat 1 (passive))
Central Storage : 88GB
Expanded Storage :2GB
Virtual Storage : 180GB
DASD PAGE : 360GB
Disaster &
Switchover 16IFL Shared

DOZVM11

LPAR
DO41

73 zLinux Guests
(116 DB Cat 2 (passive))
Central Storage : 138GB
Expanded Storage :2GB
Virtual Storage : 280GB
DASD PAGE : 560GB
Disaster &
Switchover : 16IFL Shared

DOZVM21

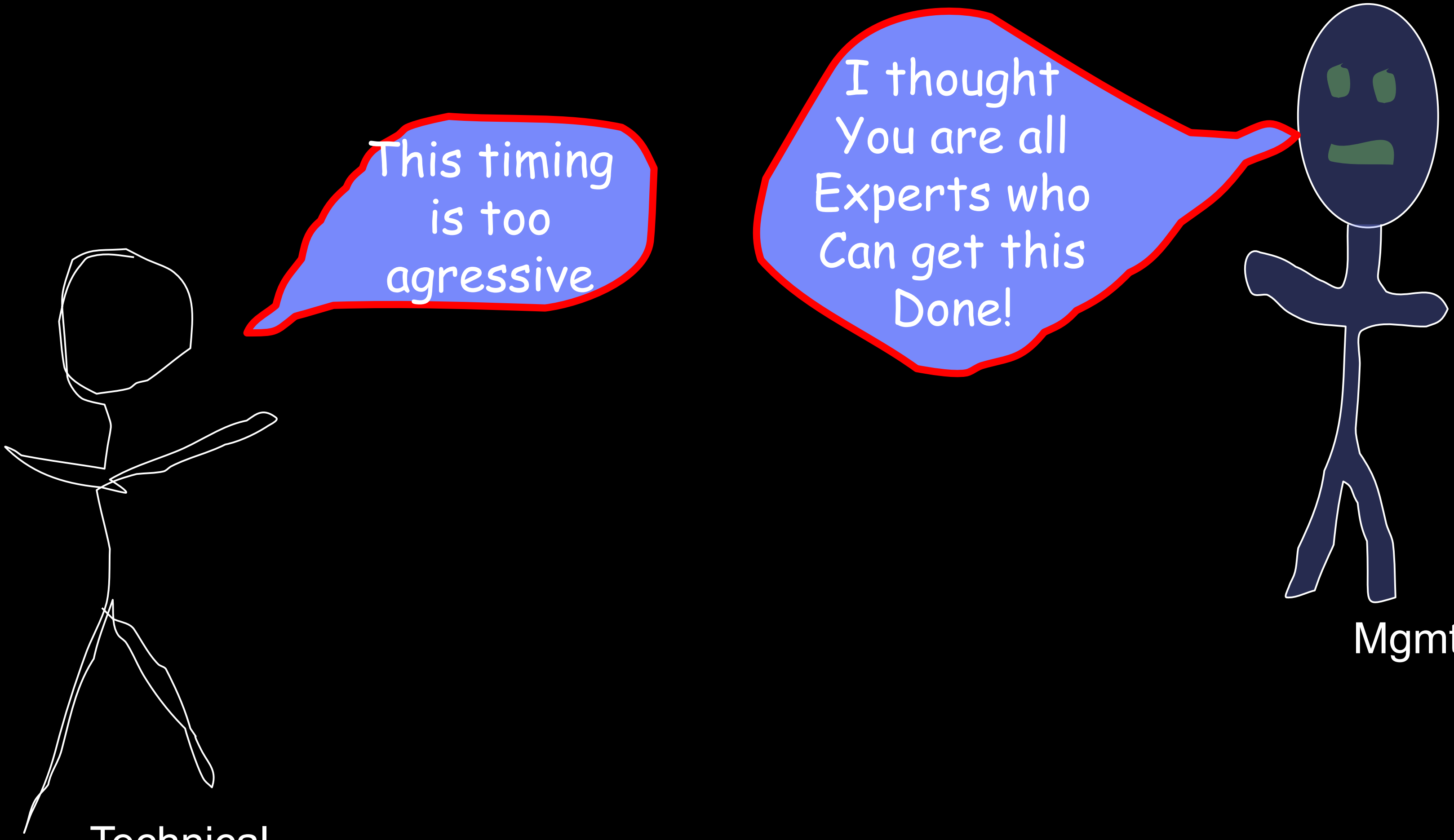
LPAR
DO42

17 zLinux Guests
(53 DB Cat 3)
Central Storage : 56GB
Expanded Storage :2GB
Virtual Storage : 158GB
DASD PAGE : 300GB
IFL : 2 Shared

DOZVM31

LPAR
DO43

We will move the Datacenter....by the end of the year!


This timing is too agressive
I thought You are all Experts who Can get this Done!
Technical
Mgmt

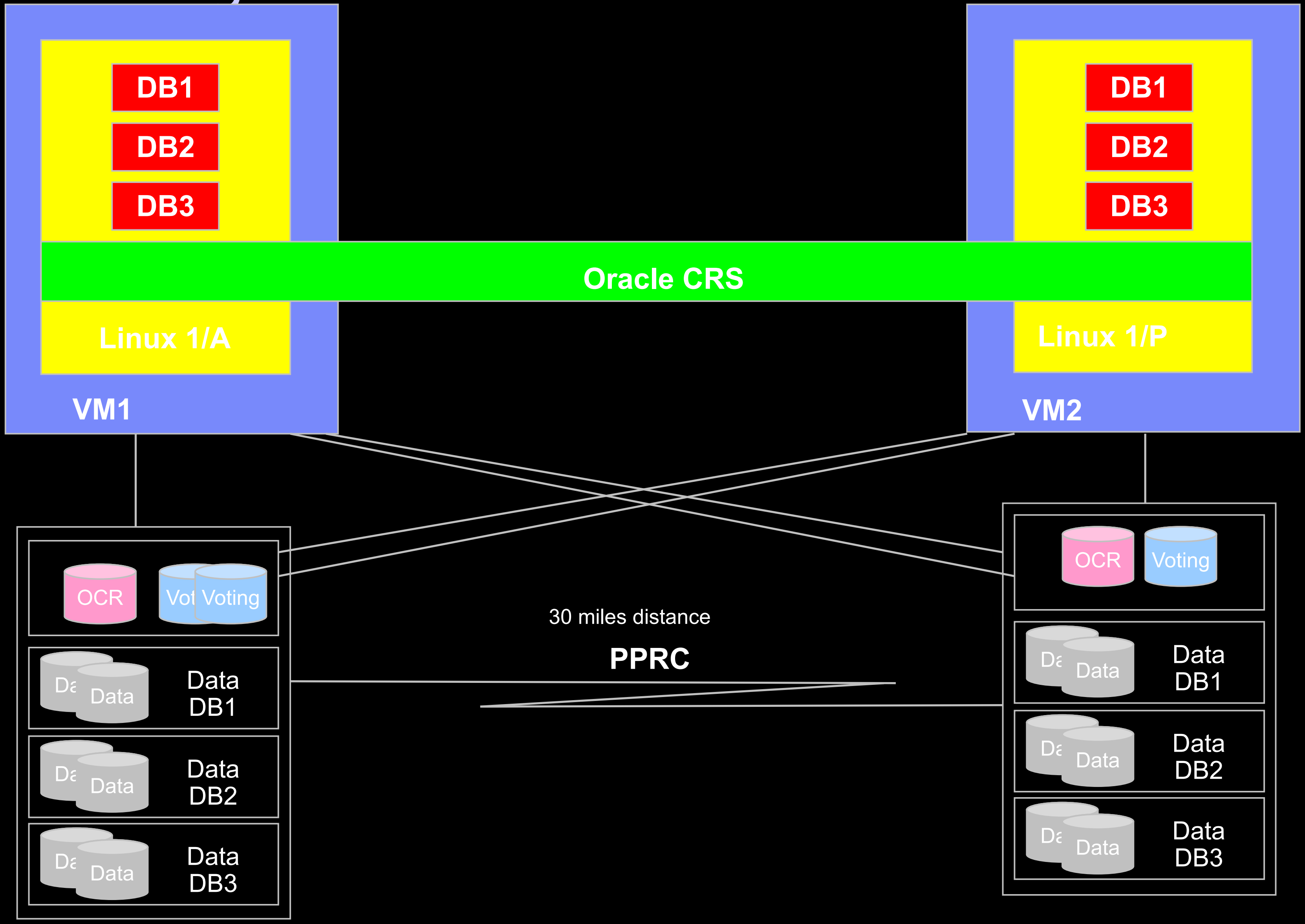
High Availability

- There are customers using CRS for Oracle
- But not on System z
- It is supposed to work (without a cluster filesystem)
- But let's see how we can actually get this to work
- And by the way: With Oracle 11 we can no longer work with RAW devices,.....then we need a cluster filesystem

Proof of Concept Oracle DB on Loz with CRS

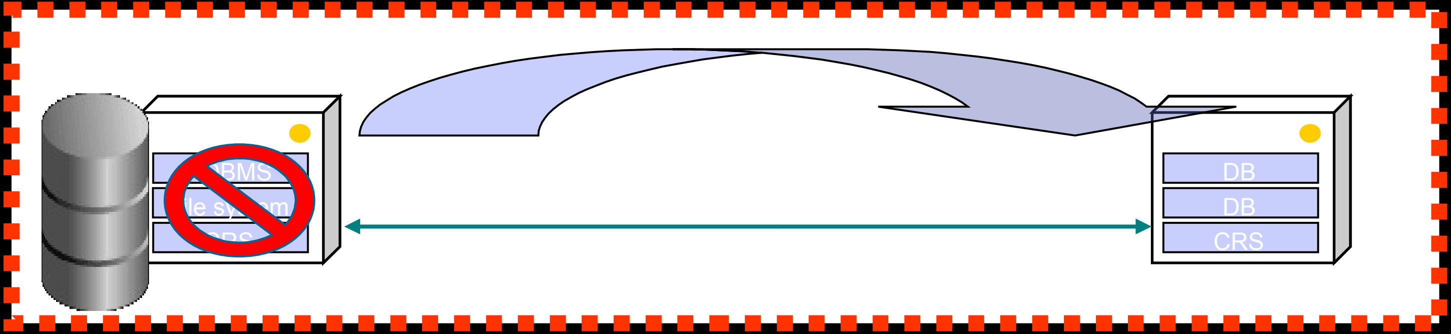
PSSC Montpellier – 05 September

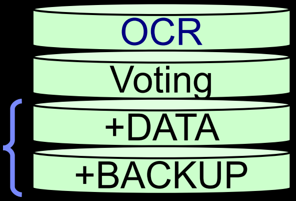
Clustering overview



IBM DS8800

CRS concepts



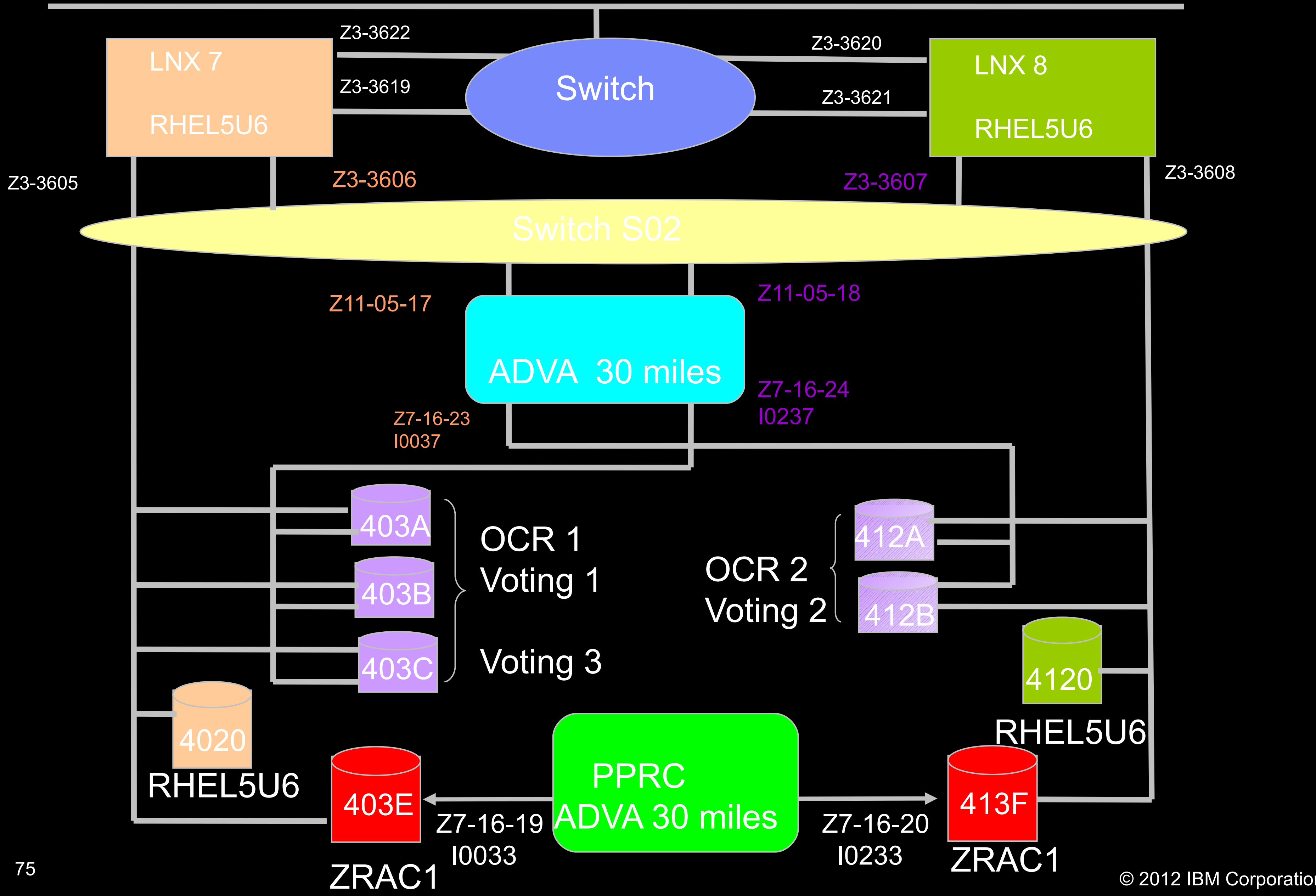


CRS concepts

- Oracle Clusterware is the software, which enables the nodes to communicate with each other, and forms the cluster
- Oracle Clusterware is run by Cluster Ready Services (CRS) using two key components

  Oracle Cluster Registry (OCR), which records and maintains the cluster and node membership information
- Voting disk which acts a tiebreaker during communication failures. Consistent heartbeat information from all the nodes is sent to voting disk when the cluster is running.
- CRS service has four components
  - OPROCd,
  - CRS Daemon (crsd),
  - Oracle Cluster Synchronization Service Daemon (OCSSD)
  - Event Volume Manager Daemon (evmd) and each handles a variety of functions
- Failure or death of the CRS daemon can cause the node failure and it automatically reboots the nodes to avoid the data corruption because of the possible communication failure between the nodes
- CRS is installed and run from a different oracle home known as ORA_CRS_HOME, which is independent from ORACLE_HOME.

OCR and Voting disks view



Scénarios description

Scenario 1: planned failover for 1 DB among 3

3 databases (ZRAC1, ZRAC2 and ZRAC3) are running into one Linux guest LNX7 on LPAR1. One of the database (ZRAC1) is manually relocated on the second Linux Guest LNX8 on LPAR2

Scenario 2: unplanned failover (for 1 database among 3)

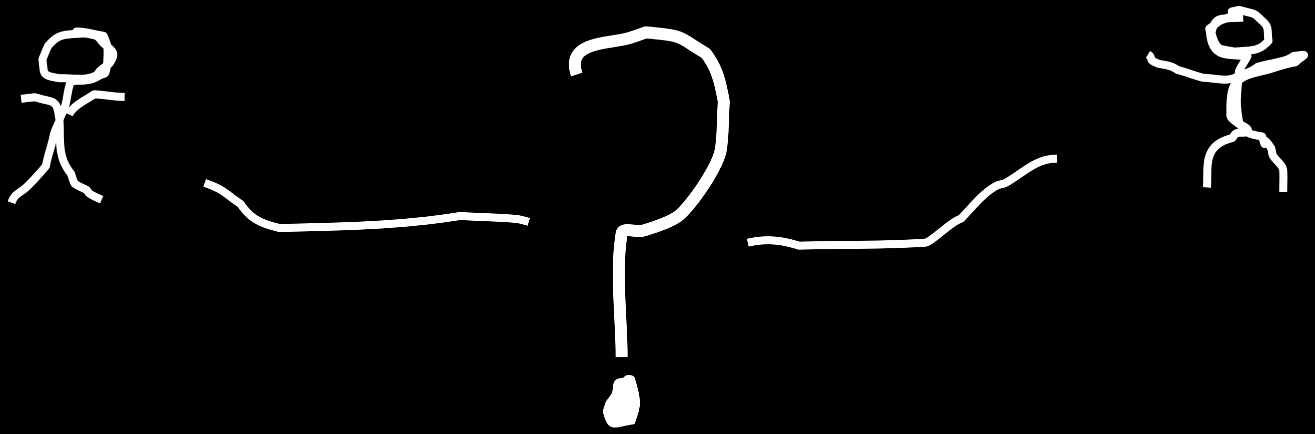
On Linux guest LNX7, 2 databases are running (ZRAC2 and ZRAC3), whereas ZRAC1 database is running on Linux guest LNX8. LNX8 is stopped, we want to check that ZRAC1 is going to be automatically relocated on LNX7.

Scenario 3: unplanned failover (for all the 3 databases)

On Linux guest LNX7, all the 3 databases are running (ZRAC1, ZRAC2 and ZRAC3), whereas no database is running on Linux guest LNX8. LNX7 is stopped, we want to check that all the databases (ZRAC1, ZRAC2 and ZRAC3) are going to be automatically relocated on LNX8.

Let's talk about Linux

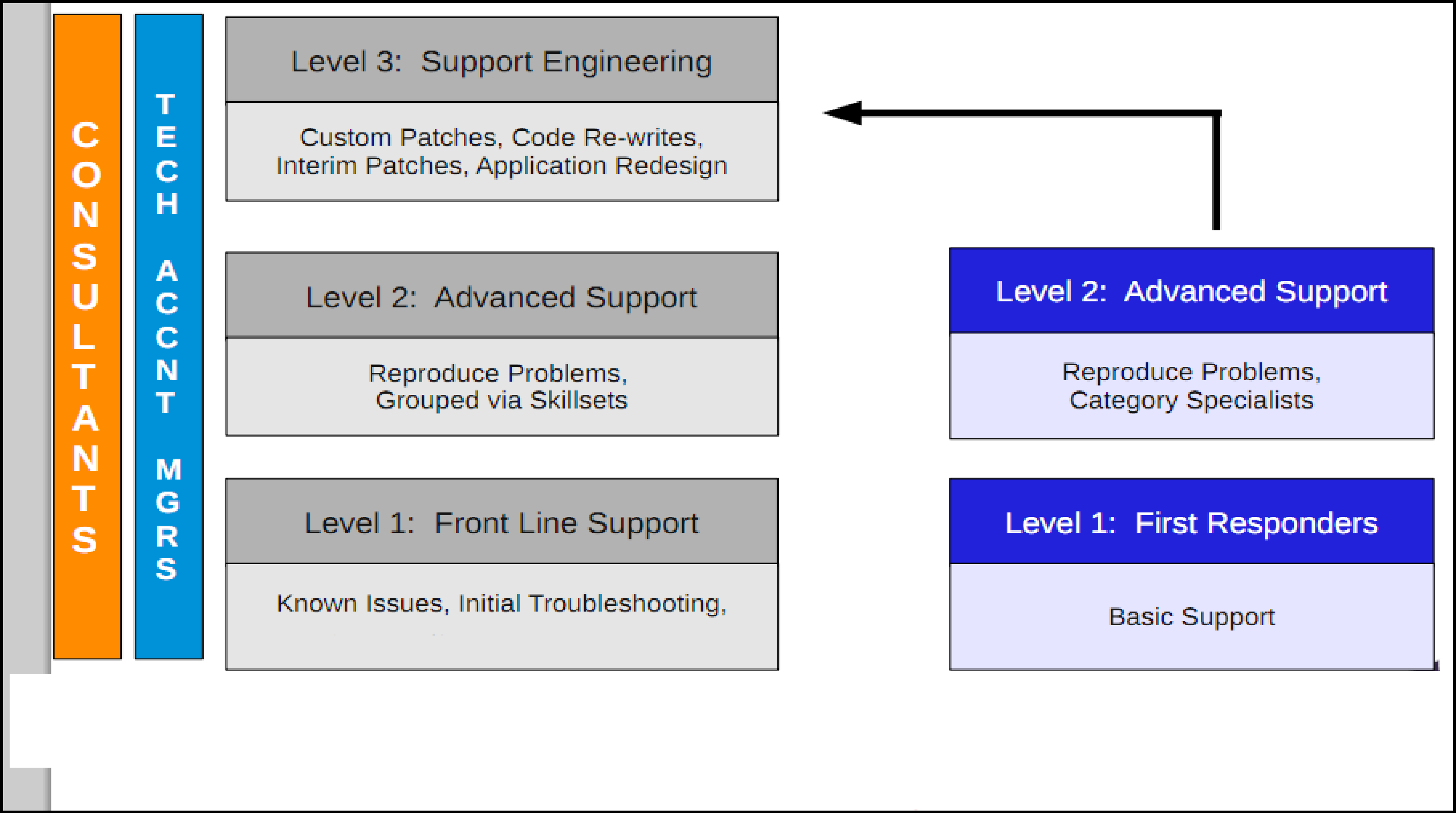
Let's talk about Linux



Novell / SuSE

Red Hat

How is the Linux Subscription delivered?
And what about z/VM support?



CONSULTANTS

TECH ACCNT MGRS

Level 3: Support Engineering

Custom Patches, Code Re-writes,
Interim Patches, Application Redesign

Level 2: Advanced Support

Reproduce Problems,
Grouped via Skillsets

Level 1: Front Line Support

Known Issues, Initial Troubleshooting,

Level 2: Advanced Support

Reproduce Problems,
Category Specialists

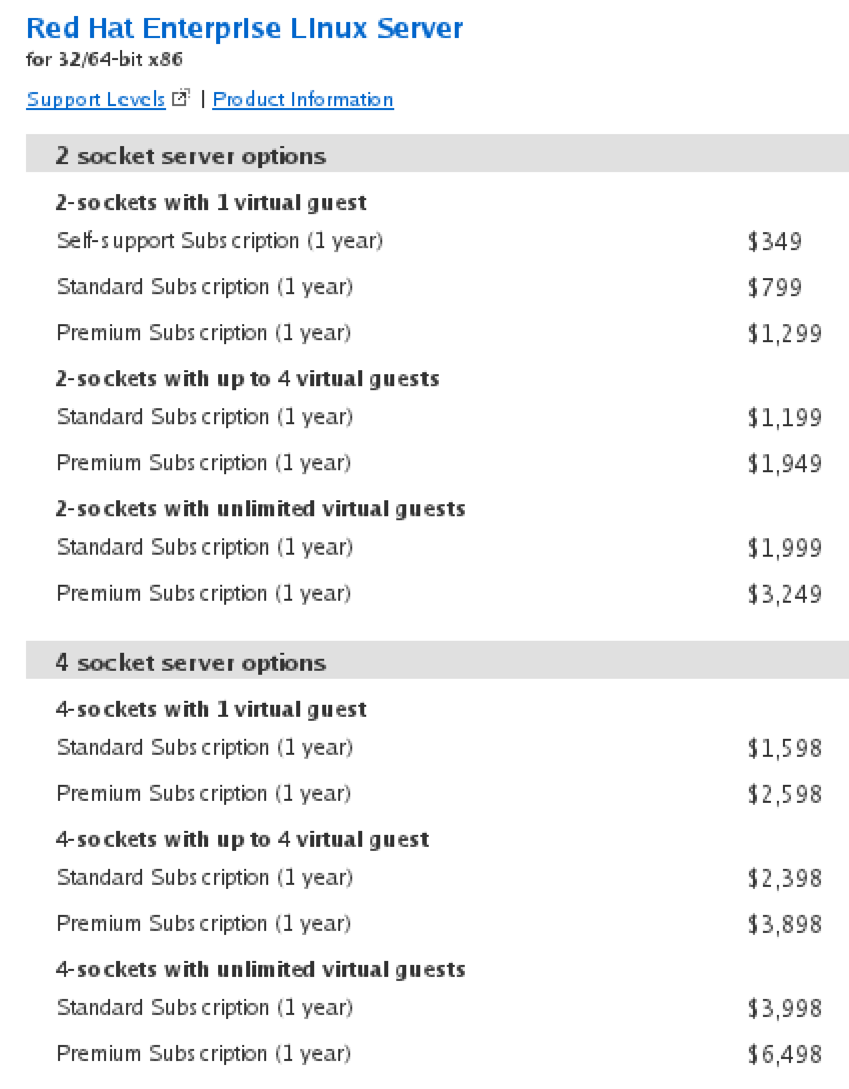
Level 1: First Responders

Basic Support

DISTRIBUTION

IBM GTS

The Linux price



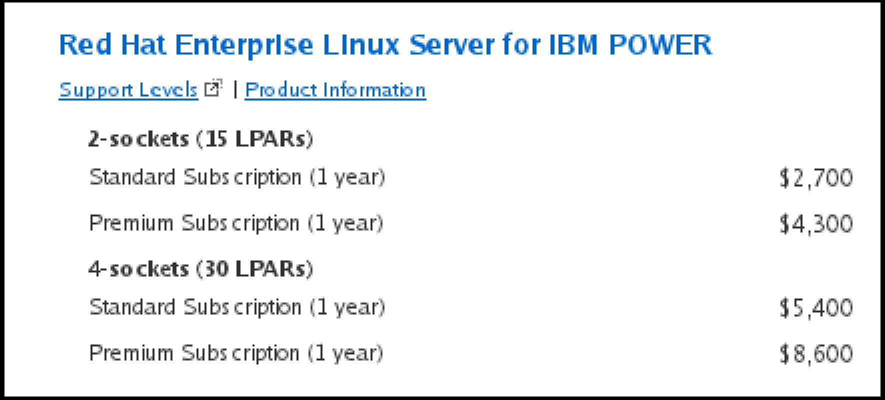
Red Hat Enterprise Linux Server

for 32/64-bit x86

Support Levels | Product Information

2 socket server options	
2-sockets with 1 virtual guest	
Self-support Subscription (1 year)	$349
Standard Subscription (1 year)	$799
Premium Subscription (1 year)	$1,299
2-sockets with up to 4 virtual guests	
Standard Subscription (1 year)	$1,199
Premium Subscription (1 year)	$1,949
2-sockets with unlimited virtual guests	
Standard Subscription (1 year)	$1,999
Premium Subscription (1 year)	$3,249

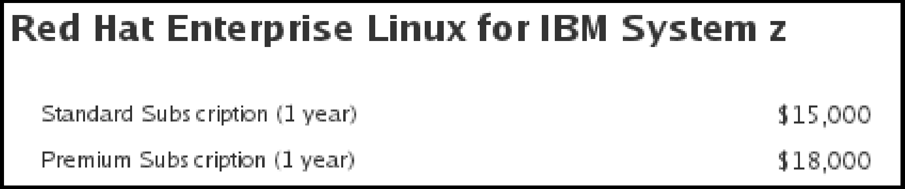
4 socket server options	
4-sockets with 1 virtual guest	
Standard Subscription (1 year)	$1,598
Premium Subscription (1 year)	$2,598
4-sockets with up to 4 virtual guest	
Standard Subscription (1 year)	$2,398
Premium Subscription (1 year)	$3,898
4-sockets with unlimited virtual guests	
Standard Subscription (1 year)	$3,998
Premium Subscription (1 year)	$6,498



Red Hat Enterprise Linux Server for IBM POWER

Support Levels | Product Information

2-sockets (15 LPARs)	
Standard Subscription (1 year)	$2,700
Premium Subscription (1 year)	$4,300
4-sockets (30 LPARs)	
Standard Subscription (1 year)	$5,400
Premium Subscription (1 year)	$8,600



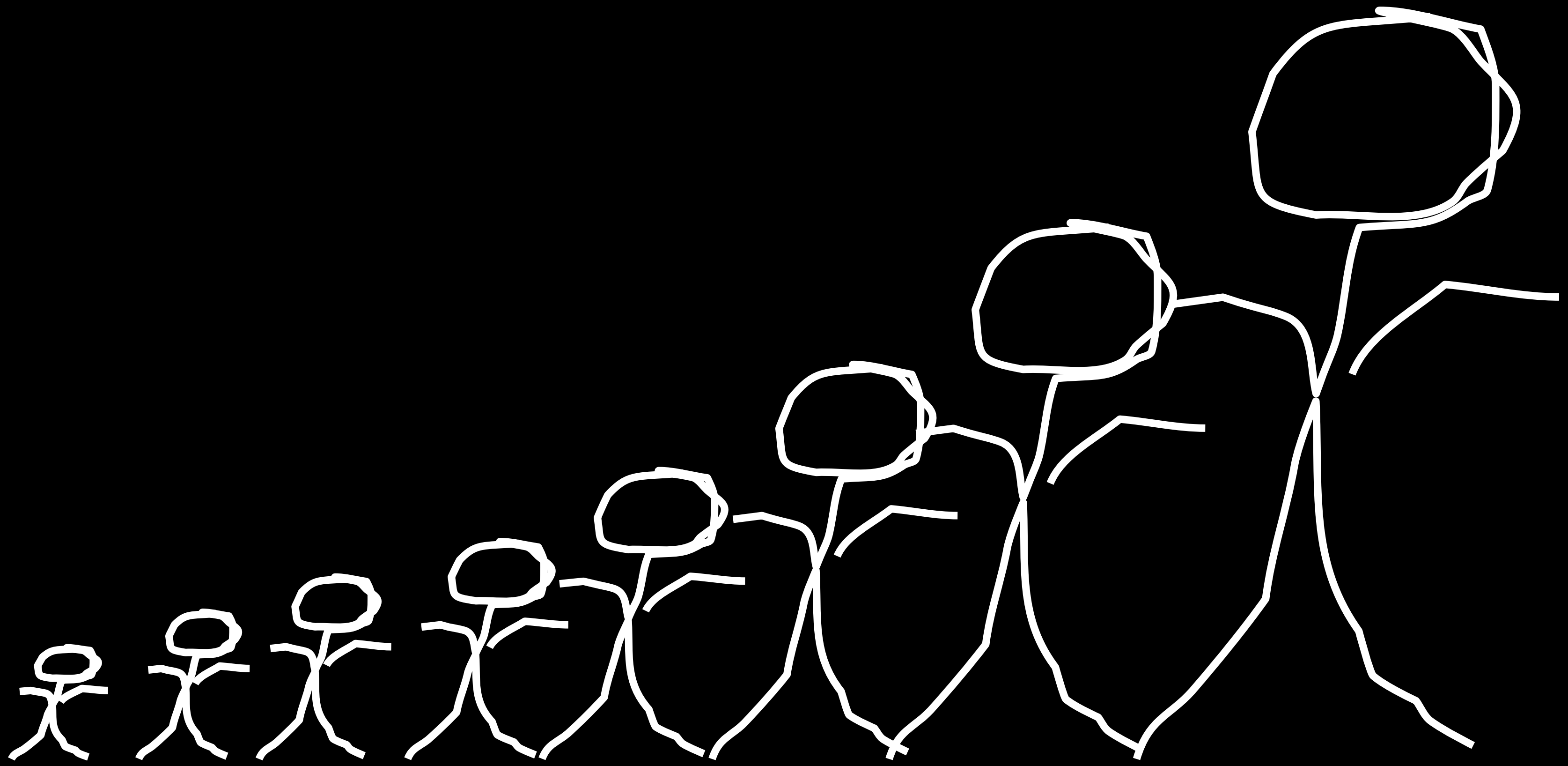
Red Hat Enterprise Linux for IBM System z

Standard Subscription (1 year)	$15,000
Premium Subscription (1 year)	$18,000

Implications from the Distribution

- Cluster Filesystem
- Support for PAV vs HyperPAV
- Future: Database Certification

If we add more people, we will be faster



We get the new hardware inventory

- Original Hardware Inventory: September 2010
- New Hardware Inventory: July

The Archive Log

- Each Oracle database has a redo log.
- This redo log records all changes made in datafiles.
- Purpose: The redo log makes it possible to replay SQL statements.

Filesystem Layout

Sr	File system	Size in GB	Remarks
1.	/	10 GB	(the root file system will also hold the usr which includes Linux executable and libraries)
2	/home	2 GB (should be LVM)	Home for user files and ordinary user home directories
3.	/tmp	5 GB	Managing temporary file system
4.	/opt	20 GB (should be LVM)	Oracle or third party software’s need to be installed

Filesystem Layout (cont)

<table>
<tr><th>Seq</th><th>File system</th><th>Size in GB</th><th>Remarks</th></tr>
<tr><td>5.</td><td>/var</td><td>5 GB</td><td>System log files and mail. This has to be a separate partition as there are occasions when log files and mails use up all space and could cause a file system full issue.</td></tr>
<tr><td>6.</td><td>Swap</td><td colspan="2">Should be equal to the physical memory. We prefer a minimum of 4 GB RAM at least.
(Red Hat recommendations)
4GB to 16GB of RAM a minimum of 4GB of swap space
16GB to 64GB of RAM a minimum of 8GB of swap space
64GB to 256GB of RAM a minimum of 16GB of swap space
256GB to 512GB of RAM a minimum of 32GB of swap space</td></tr>
</table>

Filesystem Layout

- 2 Stage SWAP Configuration
  - 256 MB VDISK
  - Between 1 to 4 GB (depending on the DB size) per Guest as emergency swap space on DASD

```
root@localhost:~> grep swap /etc/fstab
/dev/dasdb1 swap swap pri=-1 0 0
/dev/dasdc1 swap swap pri=-2 0 0
```

Storage Requirement Differences

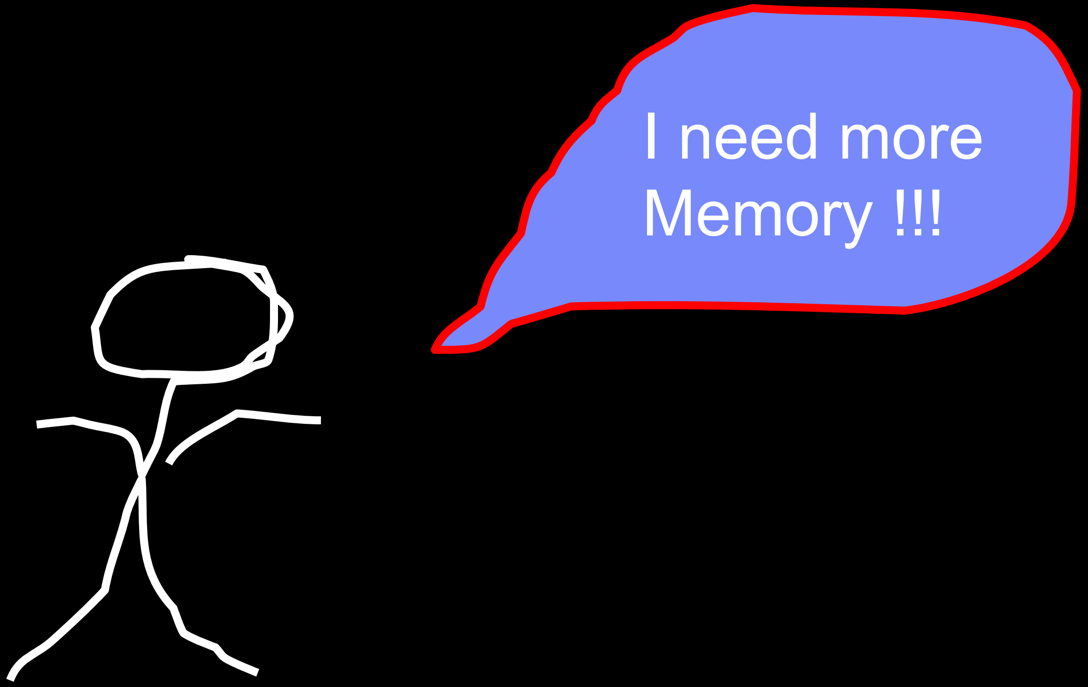
Categorie 1 DB	4,2 TB
Categorie 2 DB	18.8 TB
Categorie 3 DB	6.2 TB
Total	**31.6 TB**

Storage Requirement Differences

Categorie 1 DB	4,2 TB
Categorie 2 DB	18.8 TB
Categorie 3 DB	6.2 TB
Total	**31.6 TB**

29,2TB

The DBA

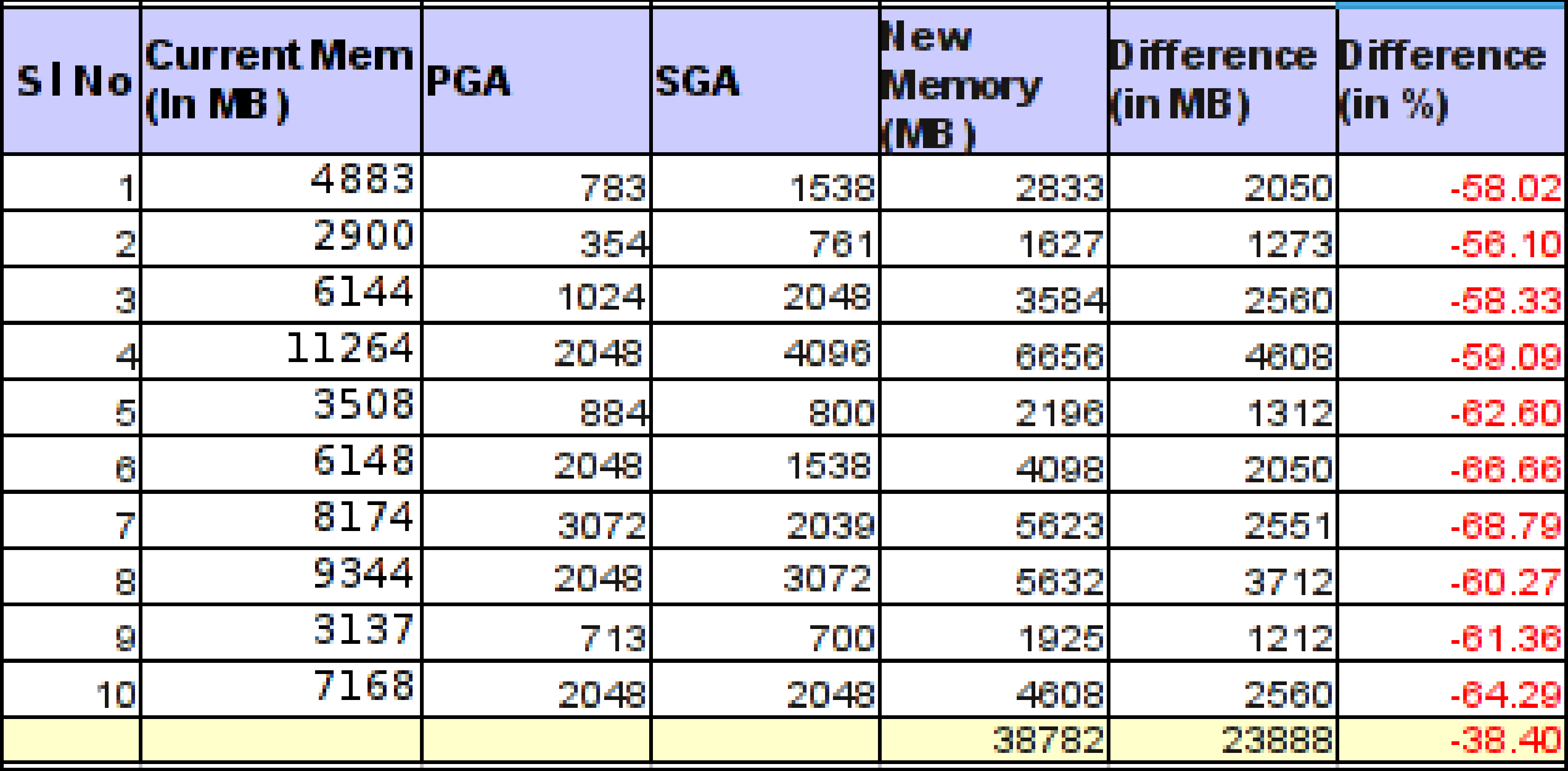


We want more memory.....and more swap

- For Oracle 10 G we use the following best practice calculation (per Database Instance).

SGA + PGA (per DB)
+ Linux (512MB)

We want more memory.....and more swap

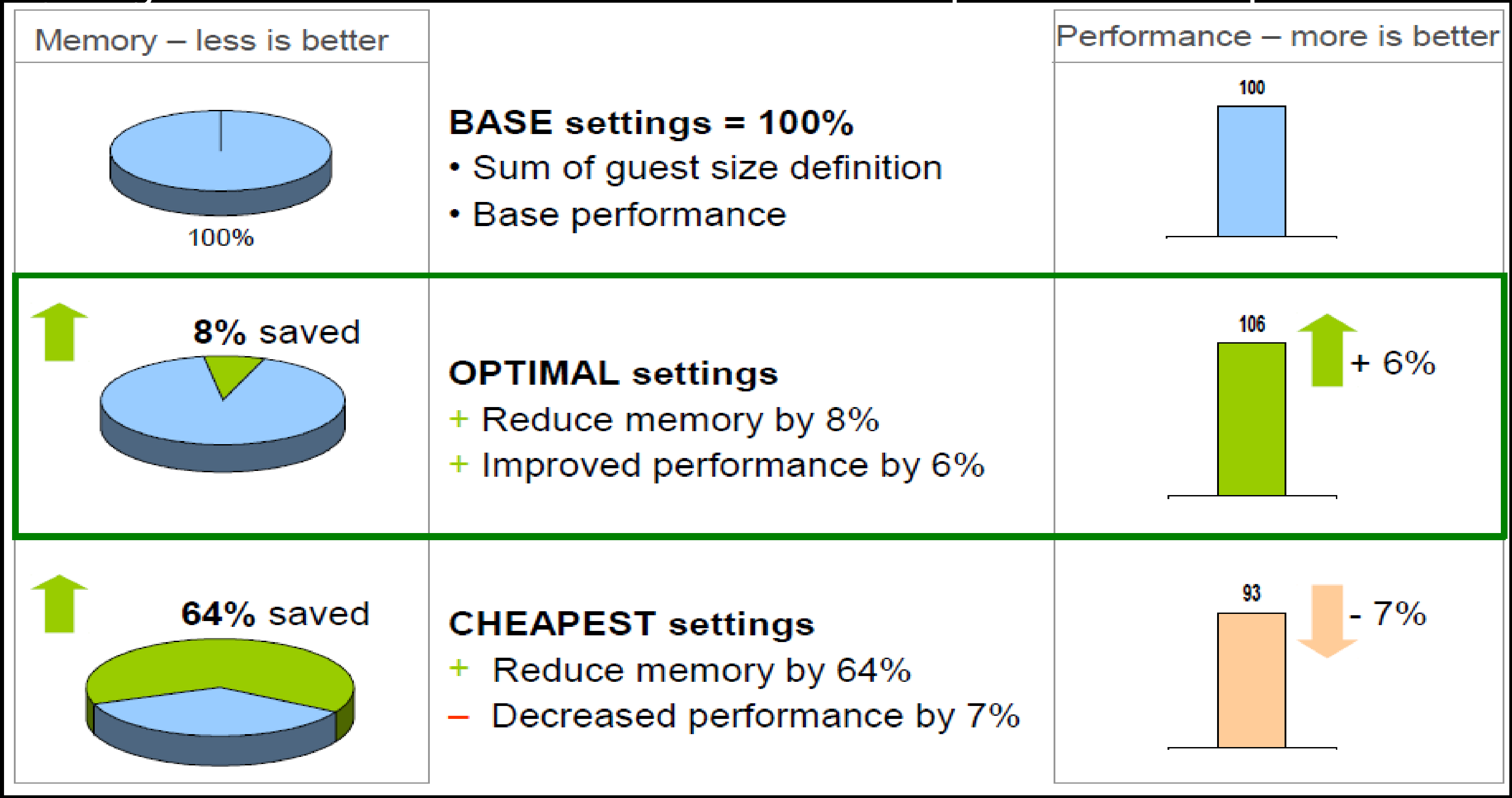


Sl No	Current Mem (In MB)	PGA	SGA	New Memory (MB)	Difference (in MB)	Difference (in %)
1	4883	783	1538	2833	2050	-58.02
2	2900	354	761	1627	1273	-56.10
3	6144	1024	2048	3584	2560	-58.33
4	11264	2048	4096	6656	4608	-59.09
5	3508	884	800	2196	1312	-62.60
6	6148	2048	1538	4098	2050	-66.66
7	8174	3072	2039	5623	2551	-68.79
8	9344	2048	3072	5632	3712	-60.27
9	3137	713	700	1925	1212	-61.36
10	7168	2048	2048	4608	2560	-64.29
				38782	23888	-38.40

DBA's dont want to “loose memory“

Test results

- Running a mix of server types as Linux guests on z/VM
  - LPAR with 28 GB central storage + 2 GB expanded storage
  - Guest workloads: WAS (13.5 GB), DB2 (12.0 GB), Tivoli Directory Server (1.5 GB), idling guest (1.0 GB)
- Leave guest size fixed – decrease LPAR size in predefined steps to scale level



BASE settings = 100%
- Sum of guest size definition
- Base performance

OPTIMAL settings
+ Reduce memory by 8%
+ Improved performance by 6%

CHEAPEST settings
+ Reduce memory by 64%
– Decreased performance by 7%

DISK Allocation

- Virtual guest memory calculation
  - SGA + PGA + Linux (512MB)
- Disk space calculation (per server)
  - OS size
  - Archive size
  - DB size
- Based on high-availability requirements, each category-1 & 2 server will have dedicated disk storage devices (none shared)
  - Mapping of disk storage space requirement to devices (3390 models)
  - Requires different sizes/3390 model types (approx. formatted space)
  - Mod-3 = 2.2GB; Mod-9 = 6GB; Mod-27 = 22GB; Mod-54 = 45GB; Mod-A = 180GB
  - Default = 256 MB vdisk per guest
  - CRS requires extra disks for OCR (Oracle Cluster Registry) and Voting disk(s) – Mod-3 and dedicated interconnect for heartbeat monitoring (low latency)

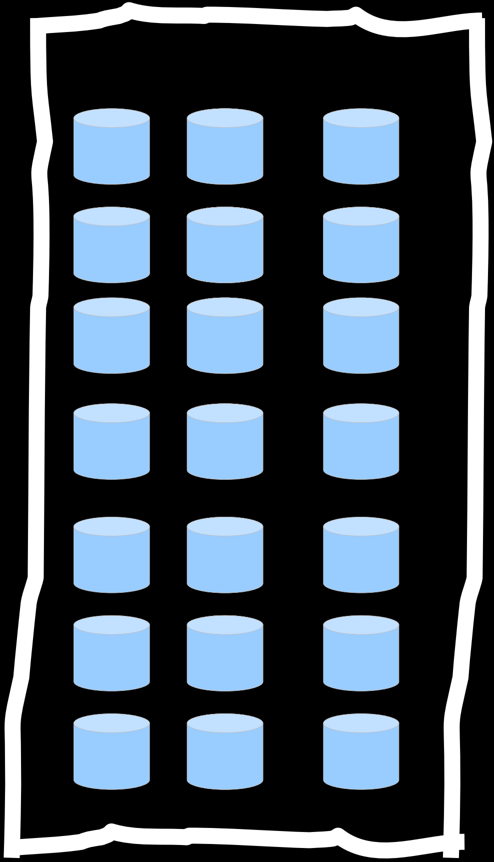
DISK Allocation

	HQZVM11	DOZVM11	HQZVM21	DOZVM21	HQZVM31	DOZVM31	TOTAL
MOD 3	94	94	184	184	38	38	**632**
MOD 9	60	60	94	94	130	50	**488**
MOD 27	tbd	tbd	tbd	tbd	40	15	**55**
MOD 54	tbd	tbd	tbd	tbd	14	6	**20**
MOD A	tbd		tbd	tbd	93	84	**177**

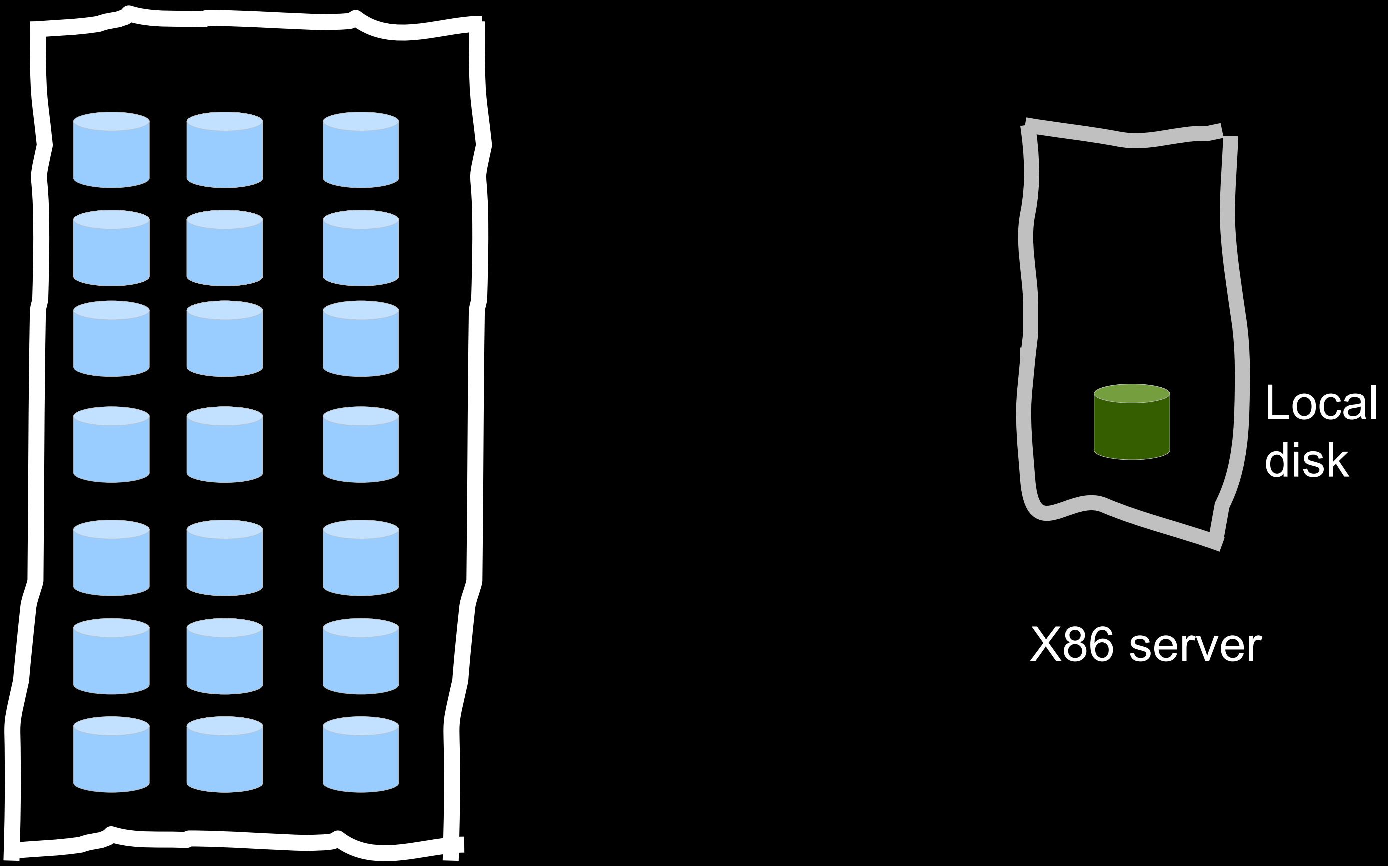
Connection & Configuration

- PAV – Parallel Access Volumes ((1 disk + 3 aliases)
- Storage Pool Striping
- 16 shared Ficon channels

This is how we choose our disks on x86….and we also want this on System z

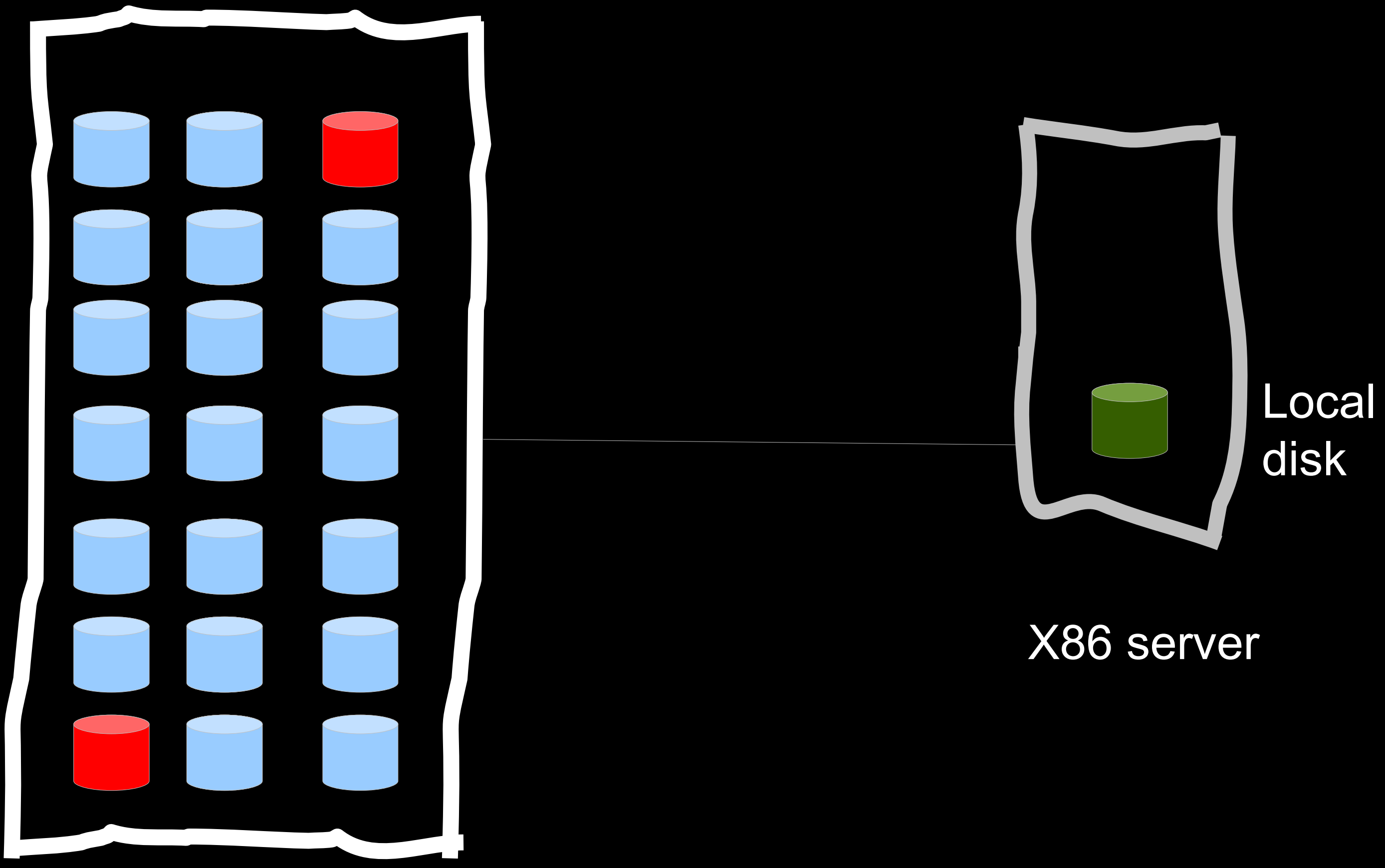


This is how we choose our disks on x86....and we also want this on System z



Storage Server

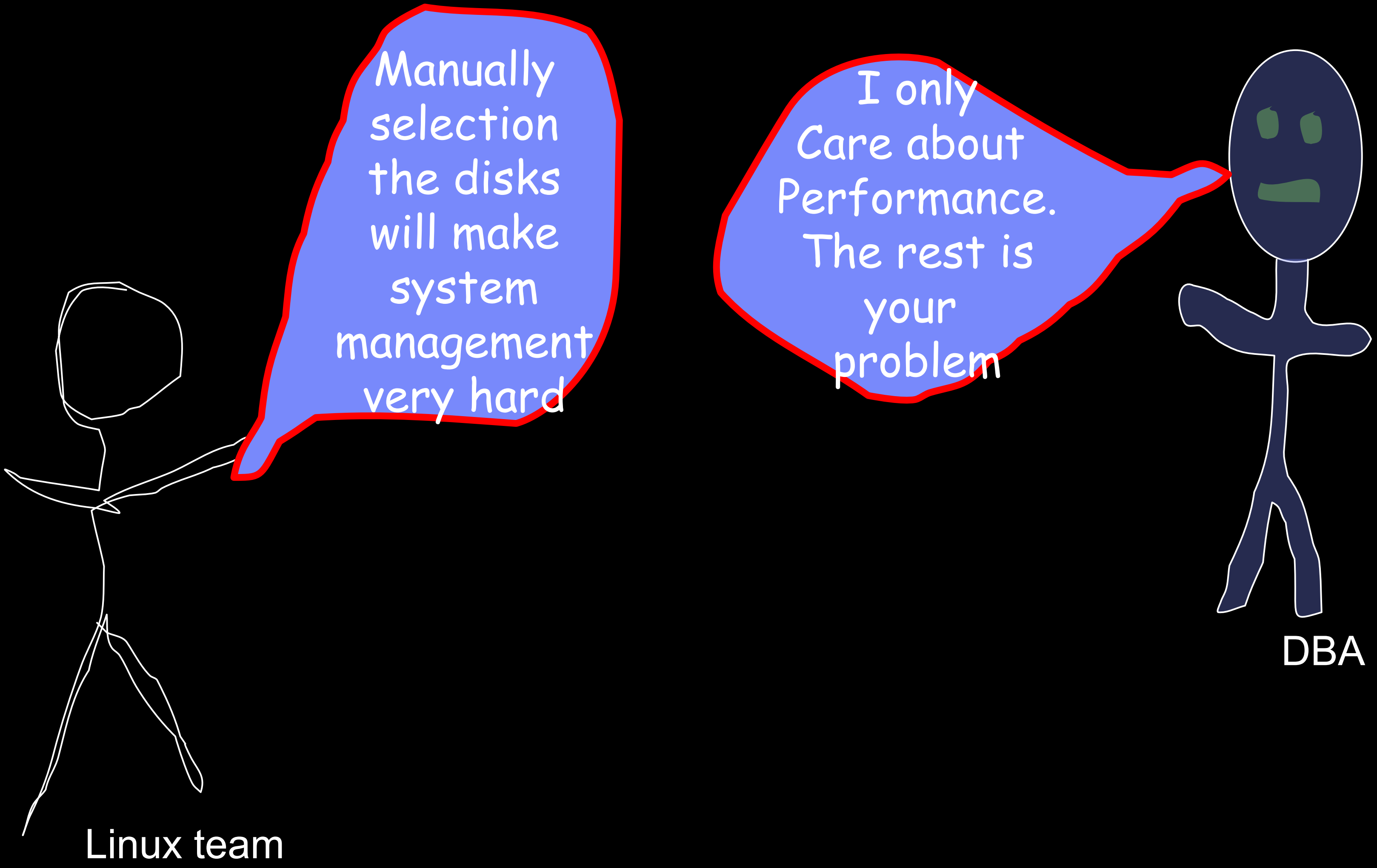
This is how we choose our disks on x86....and we also want this on System z



Storage Server

X86 server

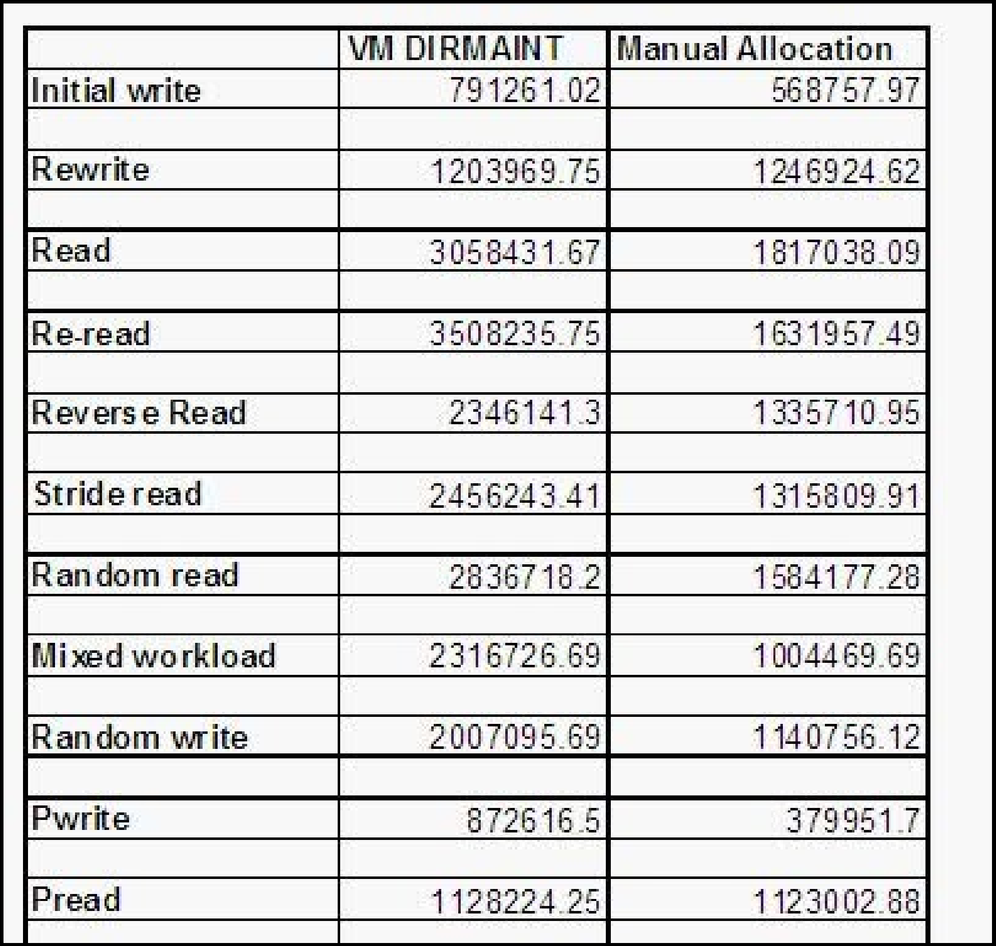
Inside the IT-Department


Manually selection the disks will make system management very hard
I only Care about Performance. The rest is your problem
DBA
Linux team

A quick benchmark removes this problem

- We configured 2x2 disks.
- 2 manually choosen
- 2 from our Storage Pool Striping + PAV setup
- Then we used IOZone

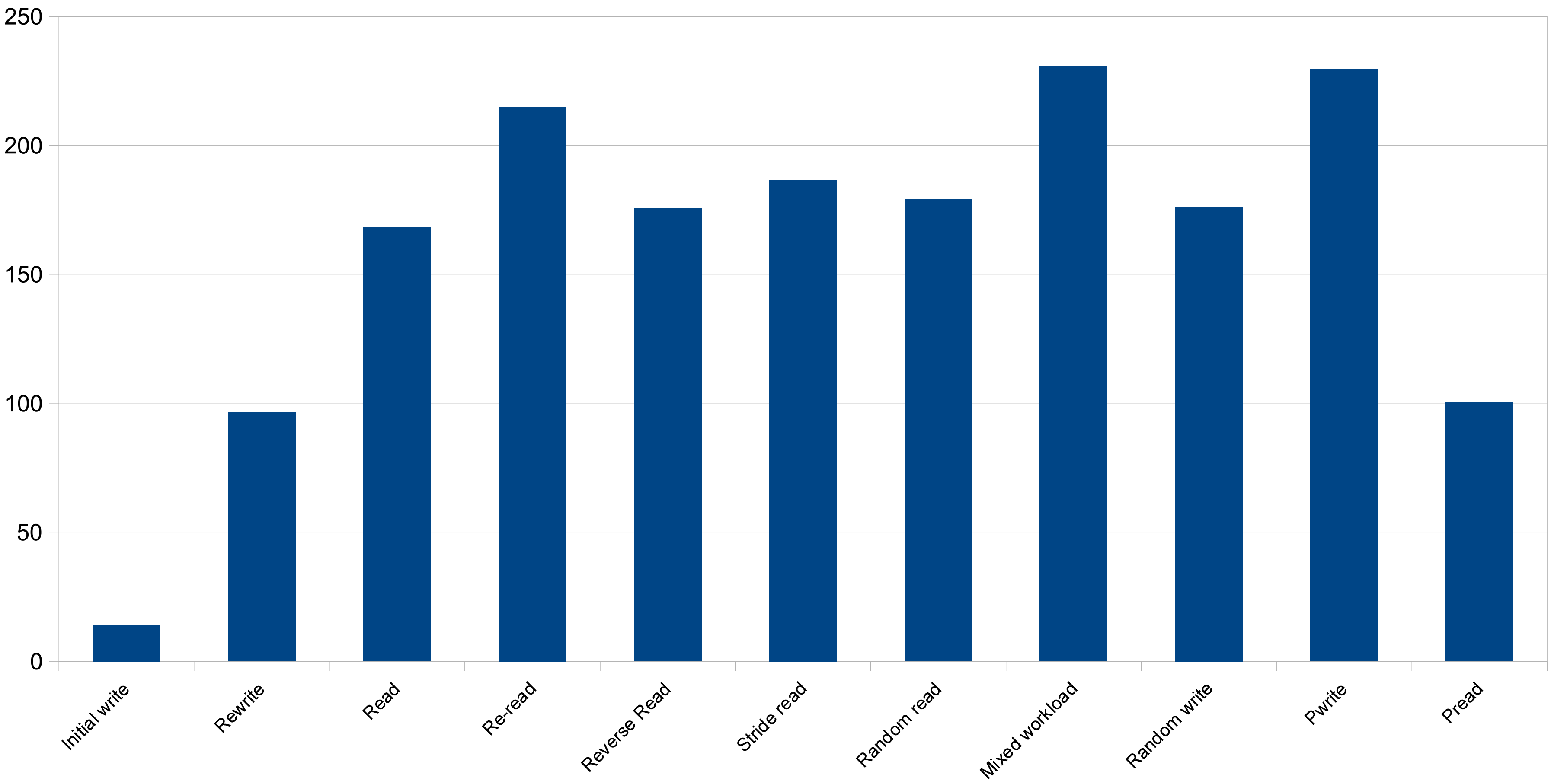
Benchmark Results



	VM DIRMAINT	Manual Allocation
Initial write	791261.02	568757.97
Rewrite	1203969.75	1246924.62
Read	3058431.67	1817038.09
Re-read	3508235.75	1631957.49
Revers e Read	2346141.3	1335710.95
Stride read	2456243.41	1315809.91
Random read	2836718.2	1584177.28
Mixed workload	2316726.69	1004469.69
Random write	2007095.69	1140756.12
Pwrite	872616.5	379951.7
Pread	1128224.25	1123002.88

Results are in Kbytes/second

People like pictures!



The implementation starts

START

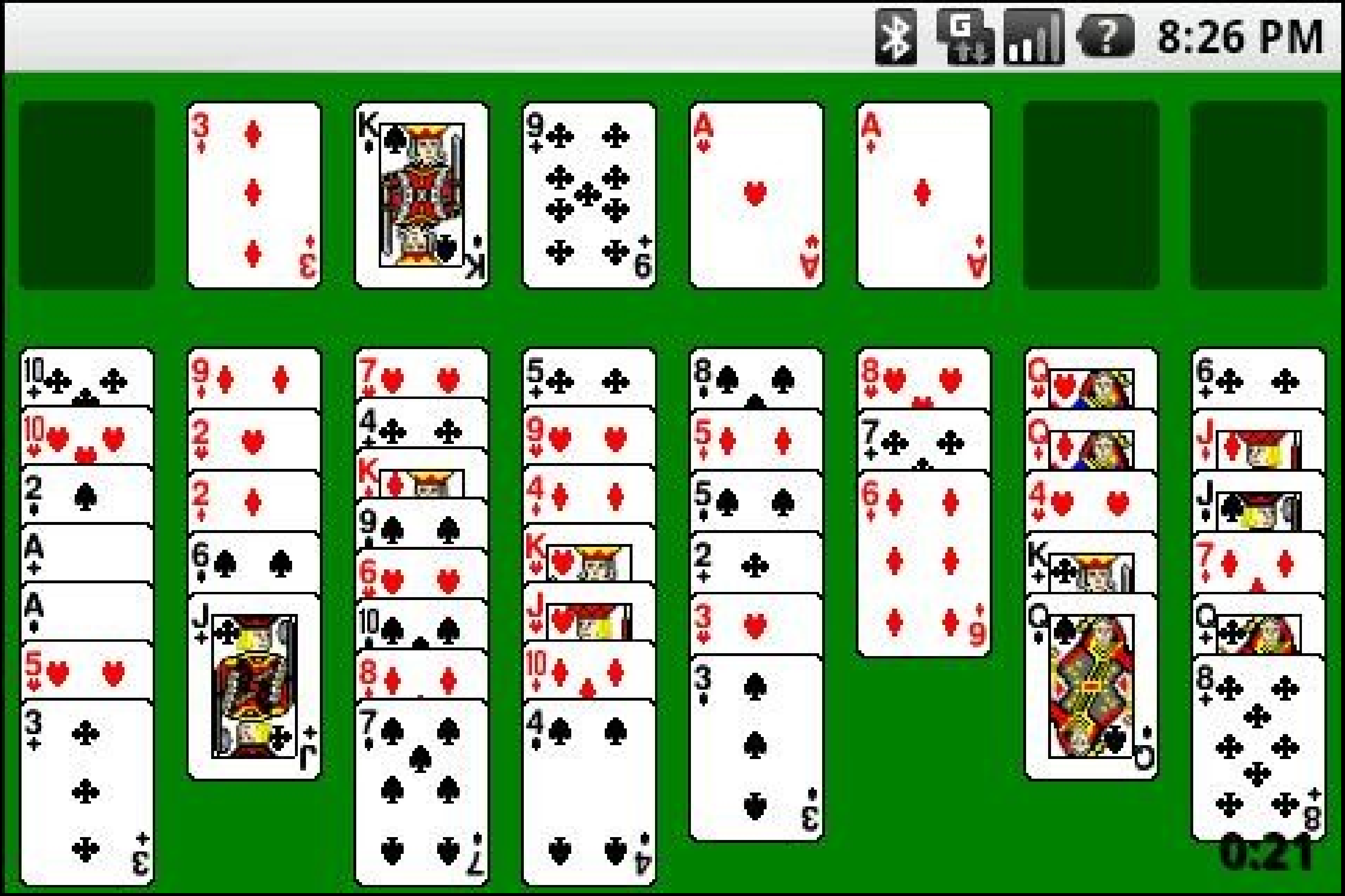
Some like it manually

- The local System programmer spend 1 week to low level format a couple of hundred dasd disks.....

How a little shell script removed our systems...

```
root@localhost:~> for i in `cat devices.txt`; \
do echo $i && chccwdev -e 0.0.$i
&& sleep 2 && dasdfmt  -f /dev/dasda -b 4096  \
-p /dev/disk/by-id/0.0.$i && fdasd -a  \
/dev/disk/by-id/ccw-0.0.$i ; done
```

When you pay 2 people for 5 weeks to play Solitaire



Multipath

- RedHat Level 3 Support Confirmed that multipath.conf userfriendly names are not supported in the ramdisk
  - This impacts our disk configuration
- We have to use the /dev/IBM4711... names instead

MOD-A Cylinders: setback in the project

- Today: Linux golden image ready, and 17 virtual machine cloned with disks attached and configure with LVM

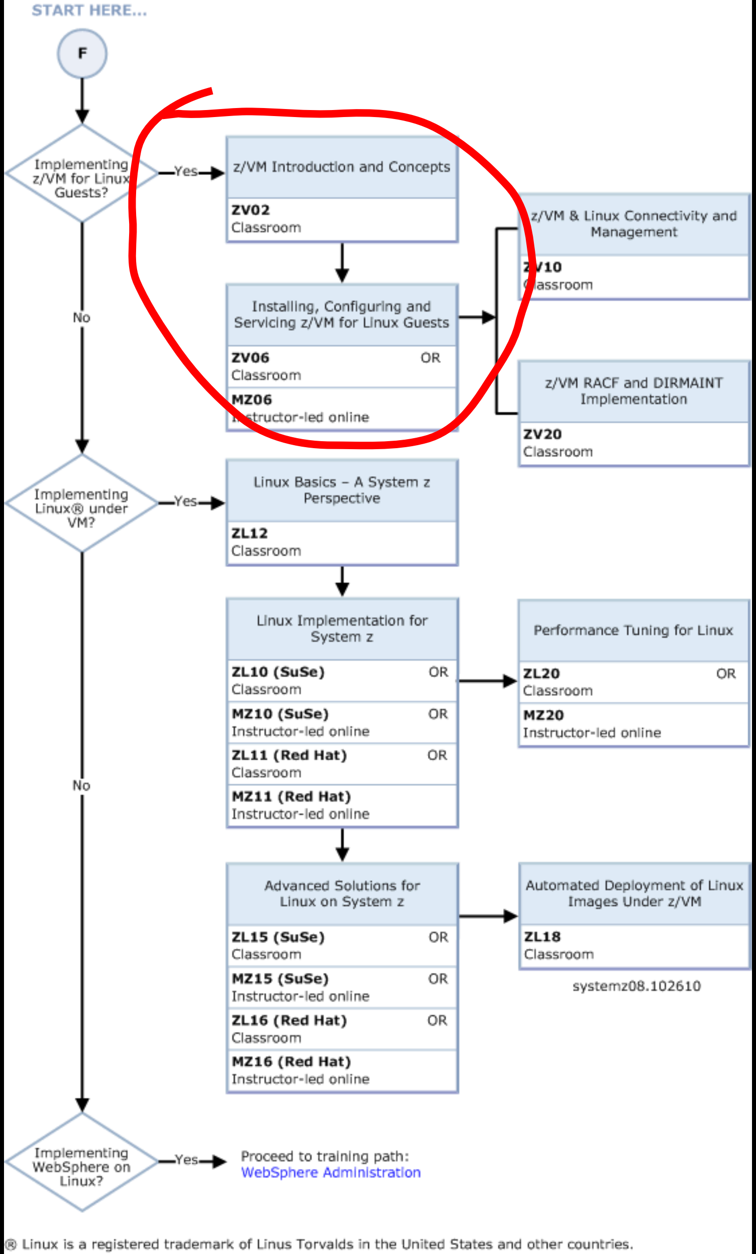
MOD-A Cylinders: setback in the project

- We realized all mod-a volumes given to us have less cylinders than we expected, thus making smaller disks (difference around 30 GB per volume).
- We expected them to have a size of approx. 180GB each with 262,668 cylinders.
- the MOD-A disks where created with 212,583 cylinders each
- At this moment we don't know where this specific cylinder size comes from
- We are proposing two solutions - a) add one mod-54 for each mod-a to compensate or b) resize volumes in DS8k.

How we fixed it

- Is possible to grow the volume in DS8k without reformatting it there.
- This is much easier from management point of view and we don;t need to waste another four device numbers to a new disk (1 disk + 3 aliases).
- We will need to reformat from Linux probably but that is fine.

Training



START HERE...

F

Implementing z/VM for Linux Guests?

Yes

z/VM Introduction and Concepts

ZV02 Classroom

Installing, Configuring and Servicing z/VM for Linux Guests

ZV06 Classroom OR

MZ06 Instructor-led online

z/VM & Linux Connectivity and Management

ZV10 Classroom

z/VM RACF and DIRMAINT Implementation

ZV20 Classroom

No

Implementing Linux® under VM?

Yes

Linux Basics – A System z Perspective

ZL12 Classroom

Linux Implementation for System z

ZL10 (SuSe) Classroom OR

MZ10 (SuSe) Instructor-led online OR

ZL11 (Red Hat) Classroom OR

MZ11 (Red Hat) Instructor-led online

Performance Tuning for Linux

ZL20 Classroom OR

MZ20 Instructor-led online

Advanced Solutions for Linux on System z

ZL15 (SuSe) Classroom OR

MZ15 (SuSe) Instructor-led online OR

ZL16 (Red Hat) Classroom OR

MZ16 (Red Hat) Instructor-led online

Automated Deployment of Linux Images Under z/VM

ZL18 Classroom

systemz08.102610

No

Implementing WebSphere on Linux?

Yes

Proceed to training path: WebSphere Administration

® Linux is a registered trademark of Linus Torvalds in the United States and other countries.

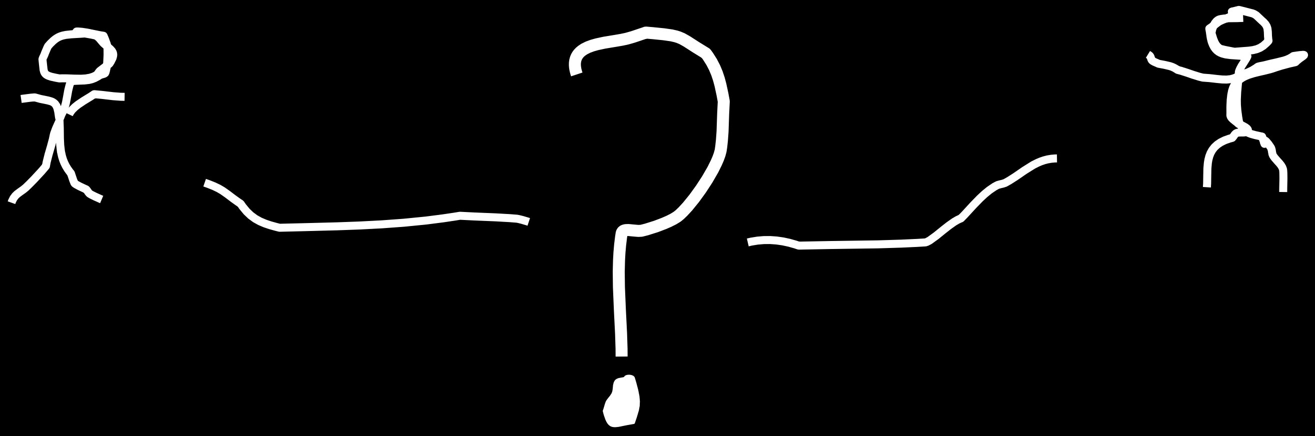
Network

- On Aug6 the network team says “we don't have any free ports”
- Next Monday they found some ports
- “It should be ready this week”
- It took 8 working days to set up the initial network cables
- During the whole time IBMer where onsite not beeing able to do much

Network (cont)

- We have 4 different networks
- But....
  - Even today not all cables are in place
  - We cannot connect to 12 out of 36 guests (via SSH)
  - In the second D/C we can ssh into 10 our of 27 systems
  - We don't have IP addresses for all network interfaces
  - The client commited to provide the infrastructure by early August
- Our CISCO switches dont support “Link Aggregation”

Deciding on a Distribution

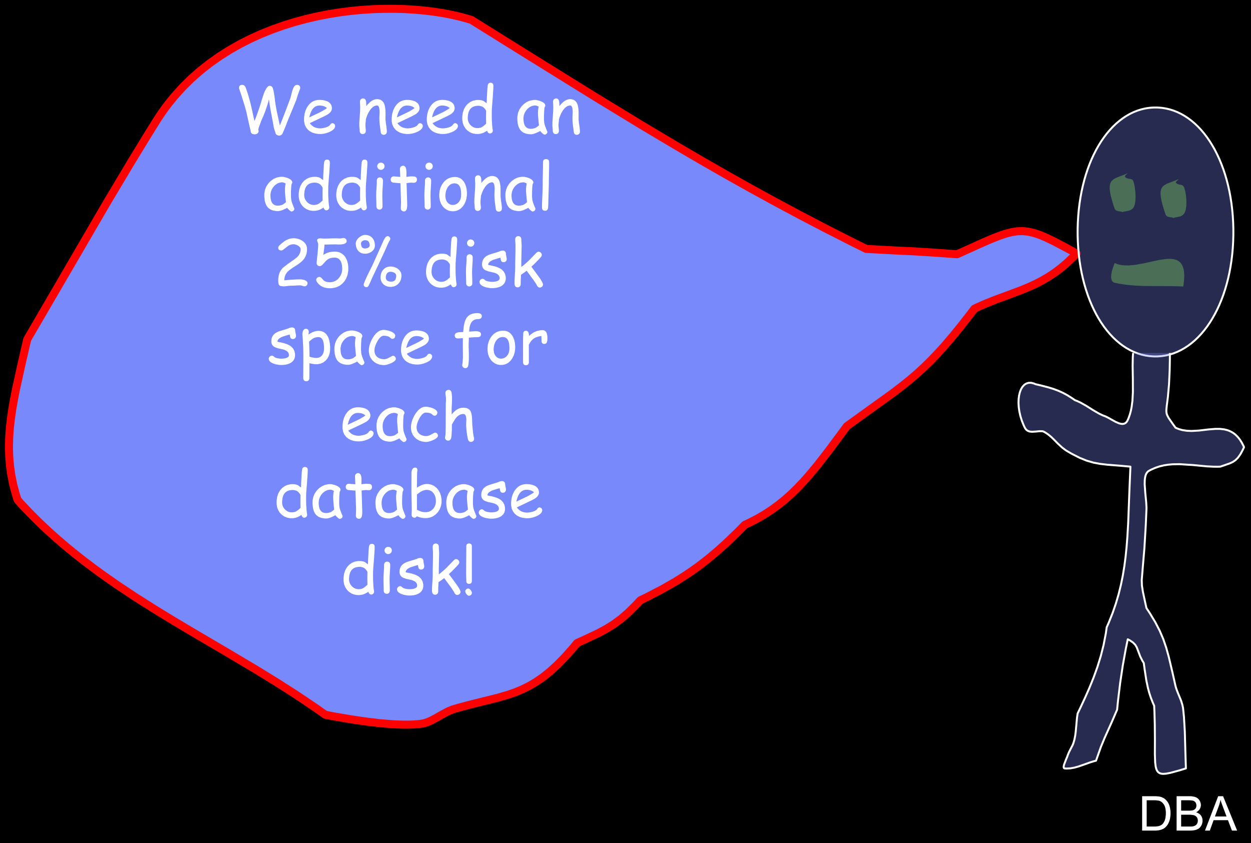


Novell / SuSE

Red Hat

Cabling & IUCV

And we redo our low level design a few more times


We need an additional 25% disk space for each database disk!
DBA

And we redo our low level design a few more times

From: userid@customer.com
Date: 07/09/2011 10:35
To IBM
Subject RE: Disk Layout

IBM, can you modify your excel sheet for the below databases related to archive log sizing. The below applications are going to grow in the near future and we need to size them efficiently in the IFL environment.

Hostname	Database	Oldsize(GB)	Newsize(GB)
linuxbl112	ABCD	76	120
linuxbl203	BCDE	36	60
linuxbl203	FHIJ	15	40
linuxbl268	KLMN	50	80
linuxbl326	OPQRRPSL	363	500
linuxbl49	STUV	28	40

Remote Access

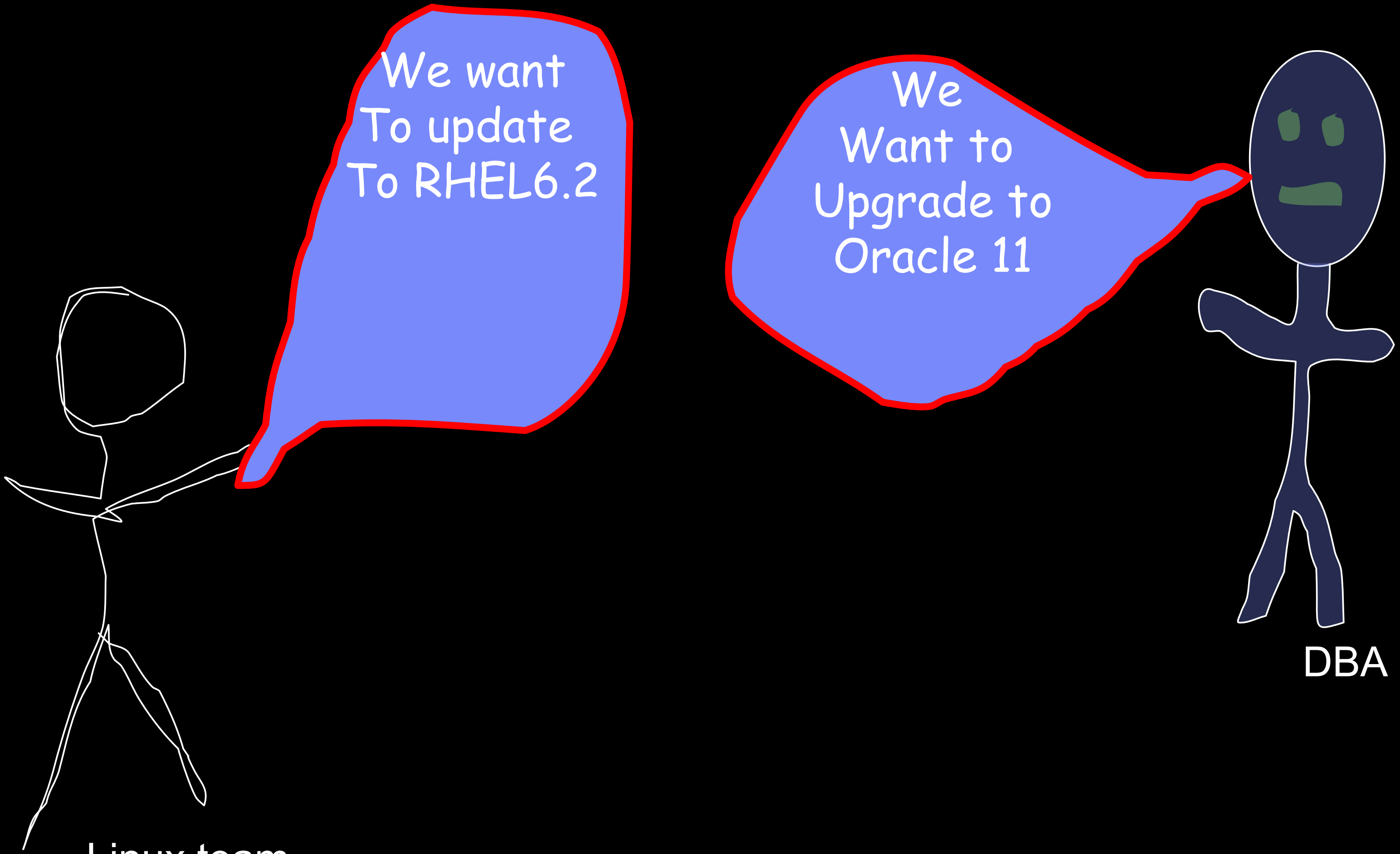
- In the Statement of Work we requested remote Access
- The first 4 weeks we spend with the layers
- 3 weeks ago the migrations factory was supposed to start to work (from remote)
- Last week the VPN Access was enabled....
- ...but we can only access the mainframe via ICMP (ping)
- We need to request some Firewall changes
- These should be implemented within 10 working days

Checkpoint: CRS/ASM

- During a checkpoint and review meeting the customer asked why we are not using Oracle Automatic Storage Management (ASM)
- We proposed this is the beginning
- But the customer did not want that
- By today they forgot about their decision

We might need a subscription and a RH SME (quarter end)

RHEL6 & Oracle



Omegamon XE & HP Openview Integration

- It is unclear who, how & when it will be implemented
- As of today 36 + 17 guests would need to be changed manually for this
- Currently the customer is using HP Openview
- Open Question: How can we integrate Omegamon into HP Openview
- Following up with Development

TSM Backup Performance

- Performance requirement: meet existing DB back-up volume of 20TB per day.

TSM server connection is 2x 1 GB Ethernet
Using link aggregation

20 TB per day
1 TB per hour
250 MB/sec → approx. 2 Gbits/second

Need to understand 20 TB requirement

Is this peak load or sustained requirement

What is th typical backup time (24 hours or less)?

What is the TSM server capability?

End of February 2012

- After having Linux systems and databases in Production since September 2011 – the customer finally purchased the Linux Distribution Subscription

Summary

- It could have been such a nice project
- Currently it is progressing
  - ...but much slower than it could
- The time of many people was wasted
- Most of the problems where "political"/organizational
  - We did not have a single bit technical Linux/VM problem which impacted this implementation

Questions?

Trademarks & Disclaimer

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries. For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml: AS/400, DB2, e-business logo, ESCON, eServer, FICON, IBM, IBM Logo, iSeries, MVS, OS/390, pSeries, RS/6000, S/390, System Storage, System z9, VM/ESA, VSE/ESA, WebSphere, xSeries, z/OS, zSeries, z/VM.

The following are trademarks or registered trademarks of other companies

Java and all Java-related trademarks and logos are trademarks of Sun Microsystems, Inc., in the United States and other countries. LINUX is a registered trademark of Linux Torvalds in the United States and other countries. UNIX is a registered trademark of The Open Group in the United States and other countries. Microsoft, Windows and Windows NT are registered trademarks of Microsoft Corporation. SET and Secure Electronic Transaction are trademarks owned by SET Secure Electronic Transaction LLC. Intel is a registered trademark of Intel Corporation. * All other products may be trademarks or registered trademarks of their respective companies.

NOTES: Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply. All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions. This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography. References in this document to IBM products or services do not imply that IBM intends to make them available in every country. Any proposed use of claims in this presentation outside of the United States must be reviewed by local IBM country counsel prior to such use. The information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice. Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.