Atlas LCG 3D Oracle cluster migration strategy at BNL

Carlos Fernando Gamboa

On behalf of database group

Grid Group, RACF Facility, Brookhaven National Lab

WLCG Collaboration Workshop,

CERN Geneva, April 2008.

Table of Contents

- Motivation for upgrade from 32 to 64 bits
- General description of BNL RAC database architecture
- Plan executed
- Results
- Conclusions

Motivation for upgrade from 32 to 64 bits Oracle

Take advantage of cluster installed memory resources:

Current configuration the SGA can not be set up beyond 2.3 GB. Could be improved by implementing (high effort though):

-Hugemem kernel: The hugemem kernel allows for a 3.42 GB SGA with a VLM. This kernel does come with a performance overhead of probably 5-15% due to address space switching.

In 64-bits memory addressing is the improved, 64 bits words or 18 billion GB memory compared to 32 bits size or 2^32=4GB of memory.

More data can be held in memory, reducing I/O to disks and thereby increasing throughput.

Better performance by carrying out 64-bit integer and floating point integer arithmetic operations.

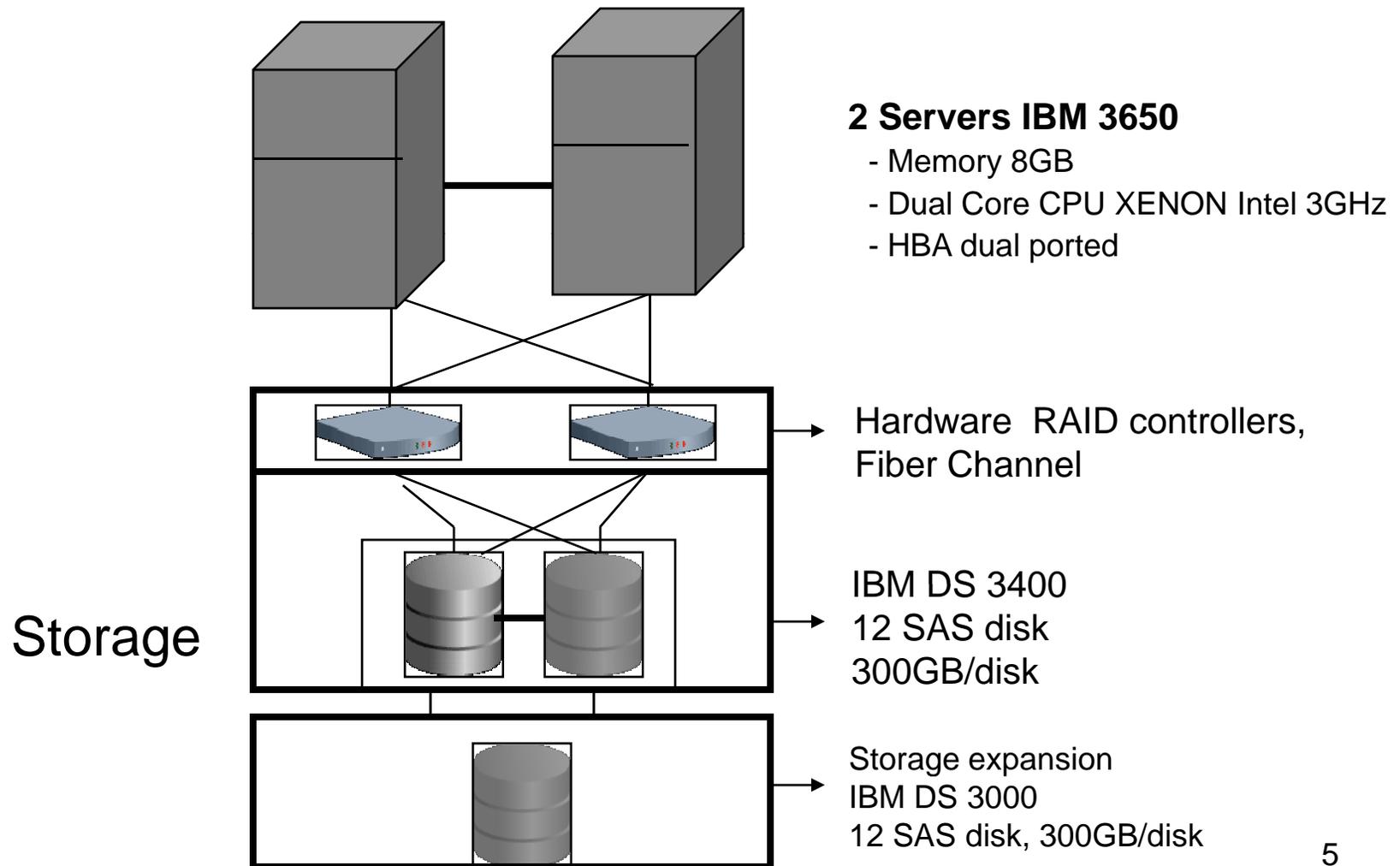# Motivation for upgrade from 32 to 64 bits Oracle

Increase cluster utilization

Upgrade would provide ability to host more services without interfering with deployed database services currently deployed
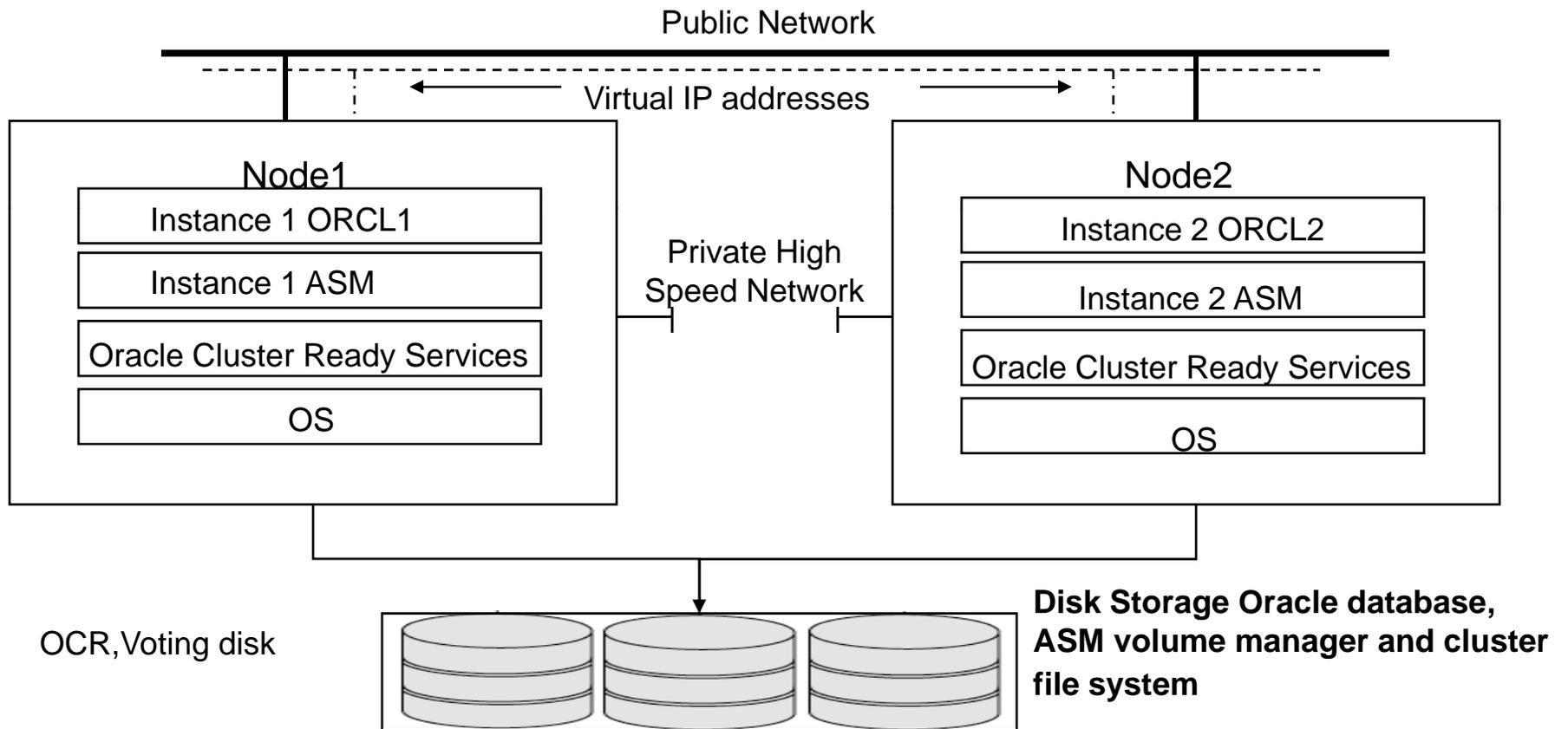- Stream apply requires 900 MB on memory

- 64 bit enable the extension of SGA memory allocation beyond 2.3GB, which would enable us, to host databases that requires a high buffer cache memory allocation, such as the TAGS database. This database requires at least 1GB of memory allocated buffer cache to achieve optimal configuration and performance.

LCG- 3D BNL Cluster
Hardware Specification

2 Servers IBM 3650
  - Memory 8GB
  - Dual Core CPU XENON Intel 3GHz
  - HBA dual ported

Hardware RAID controllers,
Fiber Channel

Storage

IBM DS 3400
12 SAS disk
300GB/disk

Storage expansion
IBM DS 3000
12 SAS disk, 300GB/disk

5

Oracle RAC architecture

Public Network

Virtual IP addresses

Node1
Instance 1 ORCL1

Instance 1 ASM

Oracle Cluster Ready Services

OS

Private High Speed Network

Node2
Instance 2 ORCL2

Instance 2 ASM

Oracle Cluster Ready Services

OS

OCR,Voting disk

Disk Storage Oracle database, ASM volume manager and cluster file system

Service	3D Conditions Database	TAGS test area	Backup/Recovery Area
Disk Group / size	+DG_DATA1 / 1.4TB	+DG_DATA2 / 700GB	+Flash Recovery Area / 700GB
Current data size	168GB	168TB	188GB

6

Oracle RAC Configuration

-Homogeneous node configuration

-Oracle homes are installed in every node on local ext 3 file system

-ASMlib is used to label partitions that are formatted the LUNs presented into the system (persistency across reboots and storage reorganizations)

32 bits install	64 bits install
System RHEL4 U4 WS.	System RHEL4 U4 ES.
Kernel 2.6.9-42.0.10.ELsmp #1 SMP	Kernel 2.6.9-67.0.7.ELsmp #1 SMP
oracleasmlibs-2.0.2-1.I386	oracleasmlib-2.0.2-1.x86_64

Migration process
-preparing backups-

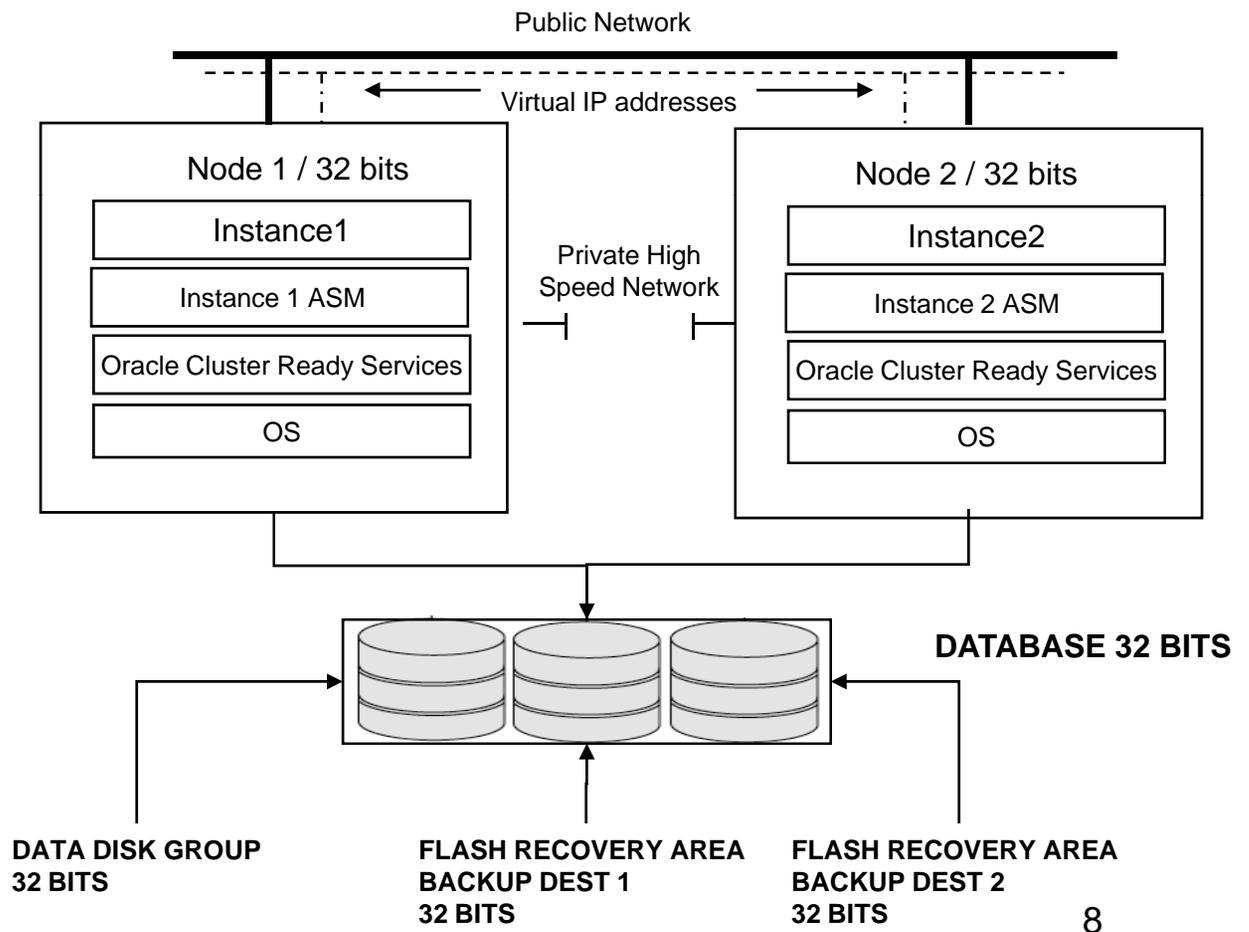**-VERIFY DATABASE RECOVERAVILITY**

Script can be found in the 3D twiki docs LCG 3D DBA meeting at CNAF 2007 "Hands on exercises"

-ENABLE SECOND BACKUP AREA
Don't forget to backup to a secure place:

-DB and ASM admin OS directories
-DBS directory
-database pfile
-TNSNAMES.ora
-Listener.ora (two nodes)
-sqlnet.ora

Database information needed in case of database recovery:
-Spfile location
-DB recovery file destination
-Dbid

Public Network

Virtual IP addresses

Node 1 / 32 bits

Instance1

Instance 1 ASM

Oracle Cluster Ready Services

OS

Private High Speed Network

Node 2 / 32 bits

Instance2

Instance 2 ASM

Oracle Cluster Ready Services

OS

DATABASE 32 BITS

DATA DISK GROUP 32 BITS

FLASH RECOVERY AREA BACKUP DEST 1 32 BITS

FLASH RECOVERY AREA BACKUP DEST 2 32 BITS

8

Migration strategy used
-Conservative approach-

- Since OCR and voting disk are shared on both nodes the entire service need to be stop

    - Oracle Cluster Registry (OCR): records cluster configuration information
    - Voting Disk: records node membership information
    - Needs to be reinstalled not compatible 32 with 64 bits

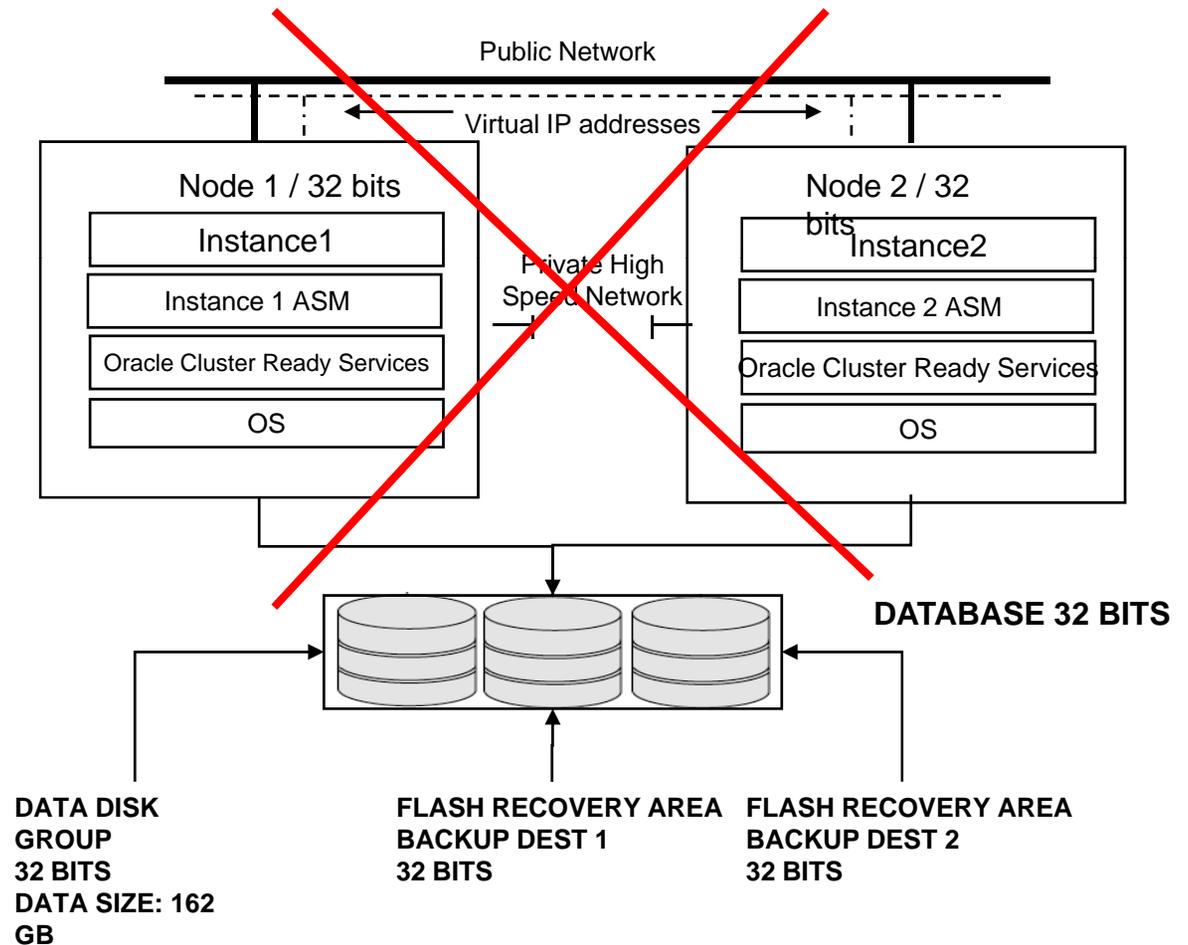- Stop database services

- Backup Database

Migration plan executed
-OS 64 bits install-

Nodes on the cluster reinstalled at the same time

-STOP DATABASE SERVICES

-BACKUP DATABASE

-REINSTALLATION OF OS 64 BITS

Public Network

Virtual IP addresses

Node 1 / 32 bits

Instance1

Instance 1 ASM

Oracle Cluster Ready Services

OS

Private High Speed Network

Node 2 / 32 bits

Instance2

Instance 2 ASM

Oracle Cluster Ready Services

OS

DATABASE 32 BITS

DATA DISK GROUP 32 BITS DATA SIZE: 162 GB

FLASH RECOVERY AREA BACKUP DEST 1 32 BITS

FLASH RECOVERY AREA BACKUP DEST 2 32 BITS

10

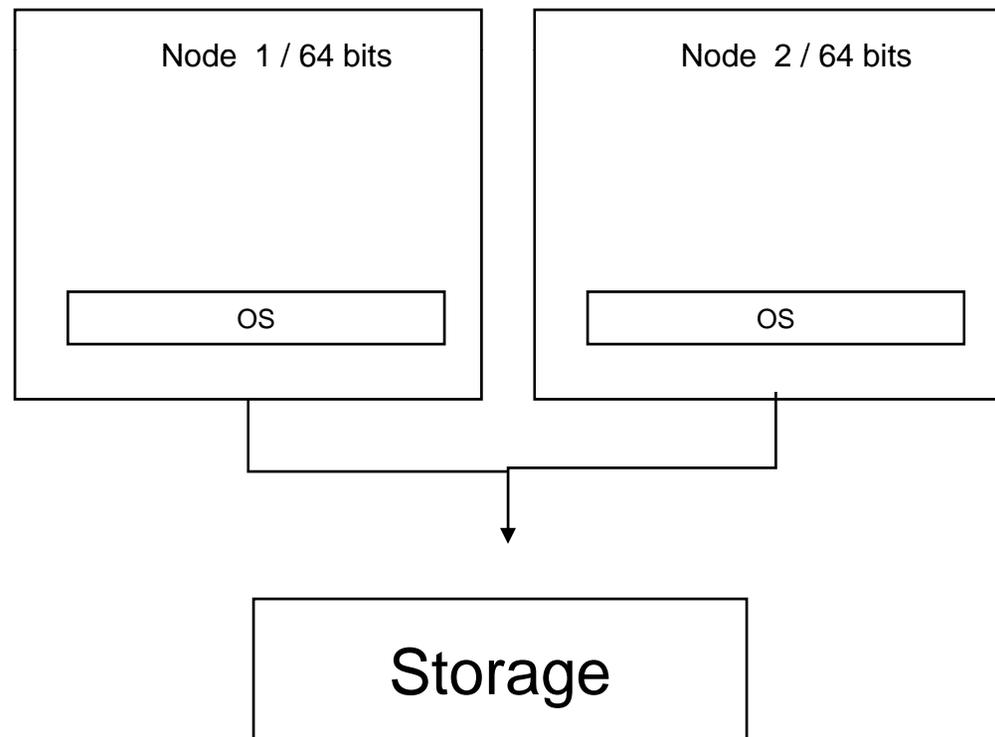# 3D Cluster Migration Intervention
-OS 64 bits install-

Migrate OS to 64 bits on two nodes

Preparing nodes for 64 bits deployment OS / Oracle binaries

Red Hat Enterprise Linux ES release 4

Installation drivers for Storage Access IBM DS 3400

Installation by GCE group

Node 1 / 64 bits	Node 2 / 64 bits
OS	OS

Storage

Data headers 32 bits

11

Migration plan executed
-Oracle 64 bits install-

Install Clusterware and Oracle database

   -Upgrade it to 10.2.0.3 version

   -Enable ASM instance

   (mount ASM disk groups)

   -Apply patches (CPU Jan 2008)

3D Cluster Migration Intervention -recompiling objects-

1. Alternative

Applying migration scripts directly, without restoring the database

On node 1

Start database in restricted mode
Oracle's recommended steps
1. Shutdown instance node 2

2.Change the cluster specific parameter
cluster_database=false

3. startup upgrade mode

4.Run the following scripts:
SQL> @ ?/rdbms/admin/utlirp.sq
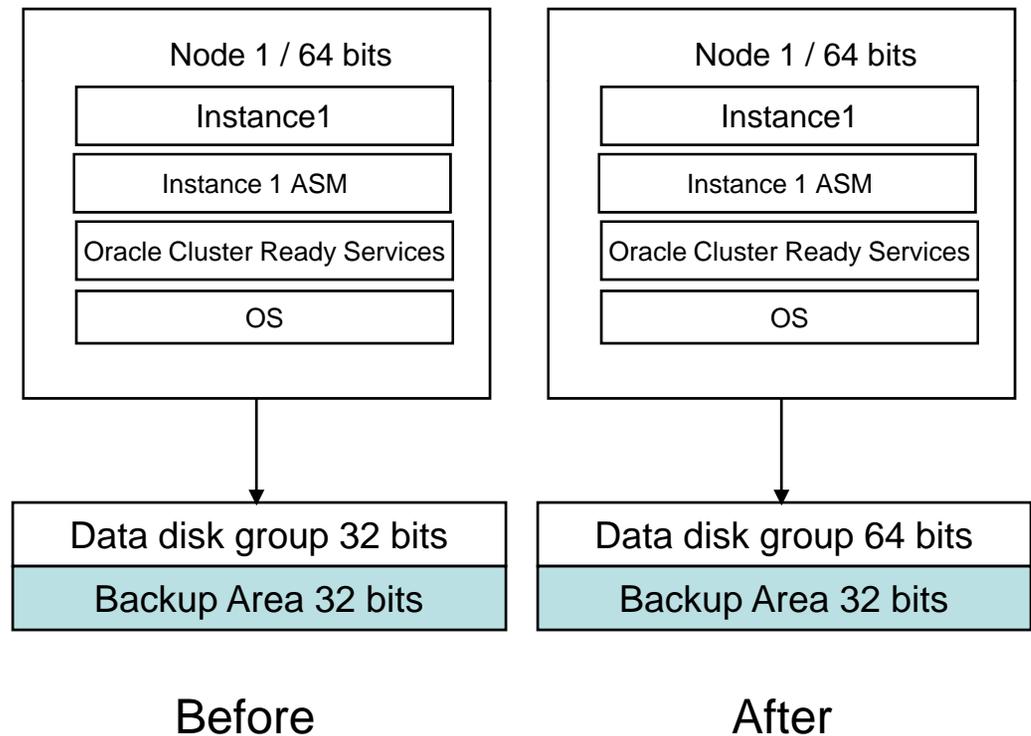SQL> @ ?/rdbms/admin/utlrp.sql
SQL> shutdown immediate;

5. Change back
 cluster_database=true

6. STARTUP DATABASE AND ALL NODES

Node 1 / 64 bits
Instance1
Instance 1 ASM
Oracle Cluster Ready Services
OS

Node 1 / 64 bits
Instance1
Instance 1 ASM
Oracle Cluster Ready Services
OS

Data disk group 32 bits
Backup Area 32 bits

Data disk group 64 bits
Backup Area 32 bits

Before

After

13

3D Cluster Migration Intervention

2. Alternative (in case of total data lost)

Restoring database a 32 bit on new 64 bits cluster installation

On node 1
Follow Oracle's metalink document
Note:467676.1

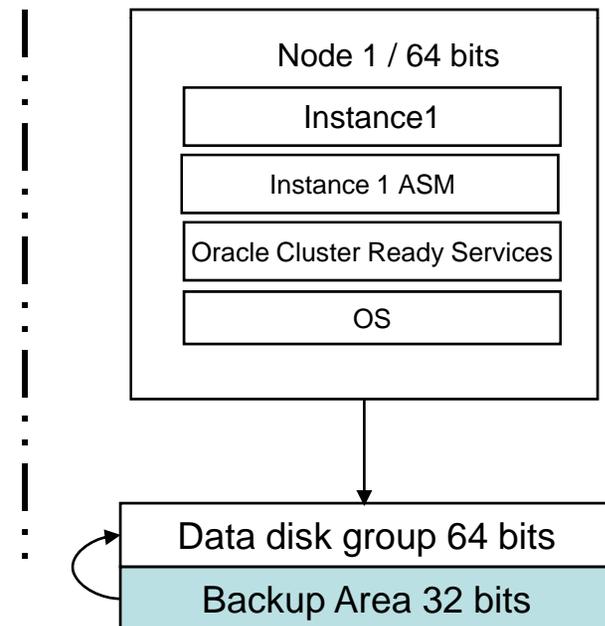After restoring the database and open it in restricted mode

-Migrating to 64 bit

SQL> @ ?/rdbms/admin/utlirp.sql
SQL> @ ?/rdbms/admin/utlrp.sql
SQL> shutdown immediate;

SQL> startup

Node 1 / 64 bits

Instance1

Instance 1 ASM

Oracle Cluster Ready Services

OS

Data disk group 64 bits

Backup Area 32 bits

3D Cluster Migration Intervention
-the scripts used-

UTILRP.sql

First invalidates and then recompiles PL/SQL modules in the

format required by the new database by:

    1. Alters certain dictionary tables

    2. Reloads STANDAR and DBMS_STANDAR necessary for using PL/SQL

    3. Recompiles of all PL/SQL modules (procedures, functions, packages, types, triggers, views)

No other DDL on the database while running the script

Primarily used for word size conversion

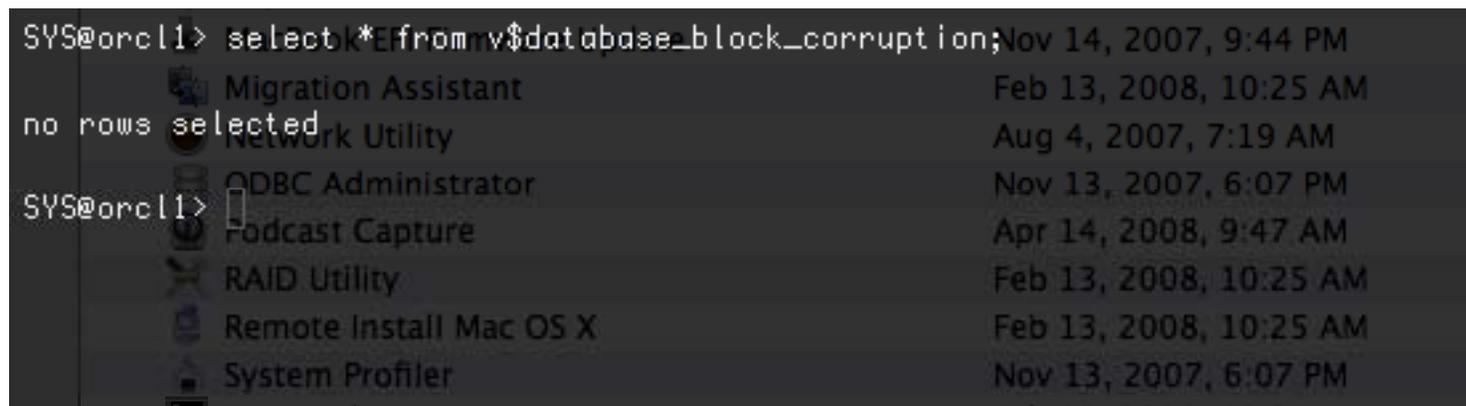# 3D Cluster Migration Intervention
-the scripts used-

UTLRP.sql

- Recompiles all invalid PL/SQL objects in the database.

- Runs a component validation procedure for each component in the database.

  Oracle recommends to run it after, upgrades, downgrades and patches to minimize latencies cause by on demand recompilation. (objects are automatically re-validated when used).

-5500 objects were recompiled on production

3D Cluster Migration Intervention

- Use Recovery Manager to search logical or physical data corruption after migration.

Migration plan executed
-enable database services-

- Startup all instances and verify database is open

- Start stream apply process

Conclusion and comments

- Migration procedure at BNL 3D atlas conditions database was presented.

- Upgrading the 3D LCG Conditions database cluster to 64 bits will take advantage of the hardware resources.

- An alternative 1 direct migration procedure was applied
    - Production system 5500 objects 25 minutes

- Recovery Manager oracle tool did not find any logical or physical data corruption

- The two alternatives presented used the same migration procedure

Bibliography

Oracle Database 10g Real Application Clusters Handbook, McGraw Hill Osborne Media; 1 edition (November 22, 2006)

Oracle Database 10g RMAN Backup & Recovery (Paperback)
McGraw-Hill Osborne Media; 1 edition (November 14, 2006)

Online documentation
> **www.metalink.com**
> **DOCUMENTS** Note: 467676.1
> 104457.1
> 272322.1

> **Oracle database concepts 10.2**
> http://download.oracle.com/docs/cd/B19306_01/server.102/b14220/toc.hm
> **3D Twiki documentation**
> https://twiki.cern.ch/twiki/bin/view/PSSGroupStreamsConfigurationChecklist

Acknowledgment

Special thanks to:

BNL GCE RACF facility group
      Dr.Jason Smith
      Robert Petkus

                CERN ITD PSS group

                   Dawid Wocjik

                   Luca Canali
                   Jacek Wojcieszuk
                   Eva Dafonte Perez
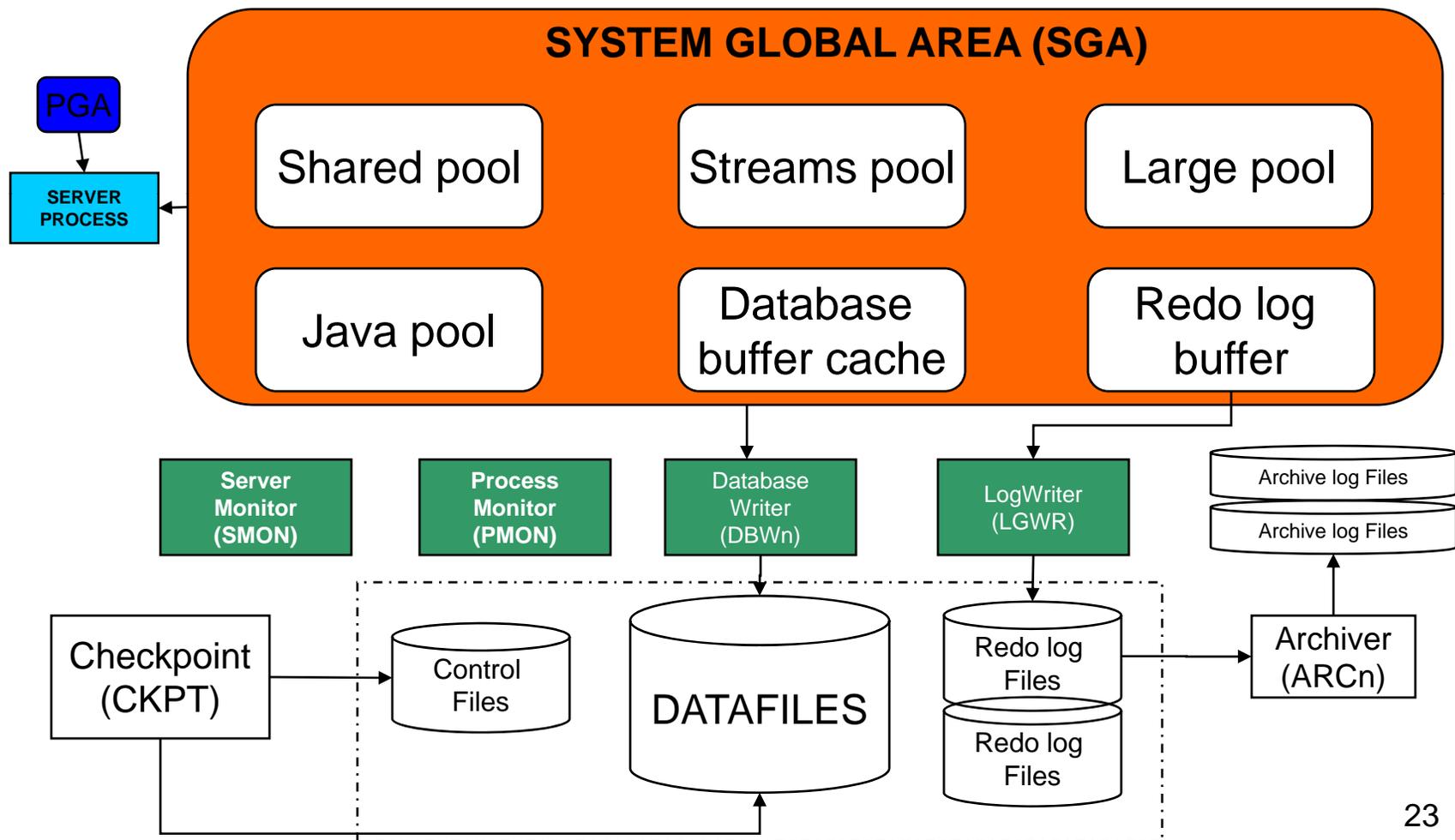   Dirk Duellmann
                   Maria Girone

Atlas DBAsPH/ATP-CO Group
    Gancho Dimitrov

Backup slides

Oracle single instance manager

SYSTEM GLOBAL AREA (SGA)

PGA

SERVER PROCESS

Shared pool

Streams pool

Large pool

Java pool

Database buffer cache

Redo log buffer

Server Monitor (SMON)

Process Monitor (PMON)

Database Writer (DBWn)

LogWriter (LGWR)

Archive log Files

Archive log Files

Checkpoint (CKPT)

Control Files

DATAFILES

Redo log Files

Redo log Files

Archiver (ARCn)

23

Oracle cluster architecture

Node1

Node 2

Cluster Manager

SGA

High Speed Interconnect

SGA

LogWriter (LGWR)

Database Writer (DBWn)

GLOBAL CACHE SERVICE (GCS)

GLOBAL CACHE SERVICE (GCS)

Database Writer (DBWn)

LogWriter (LGWR)

Redo log Files

Redo log Files

DATAFILES

Redo log Files

Redo log Files

24
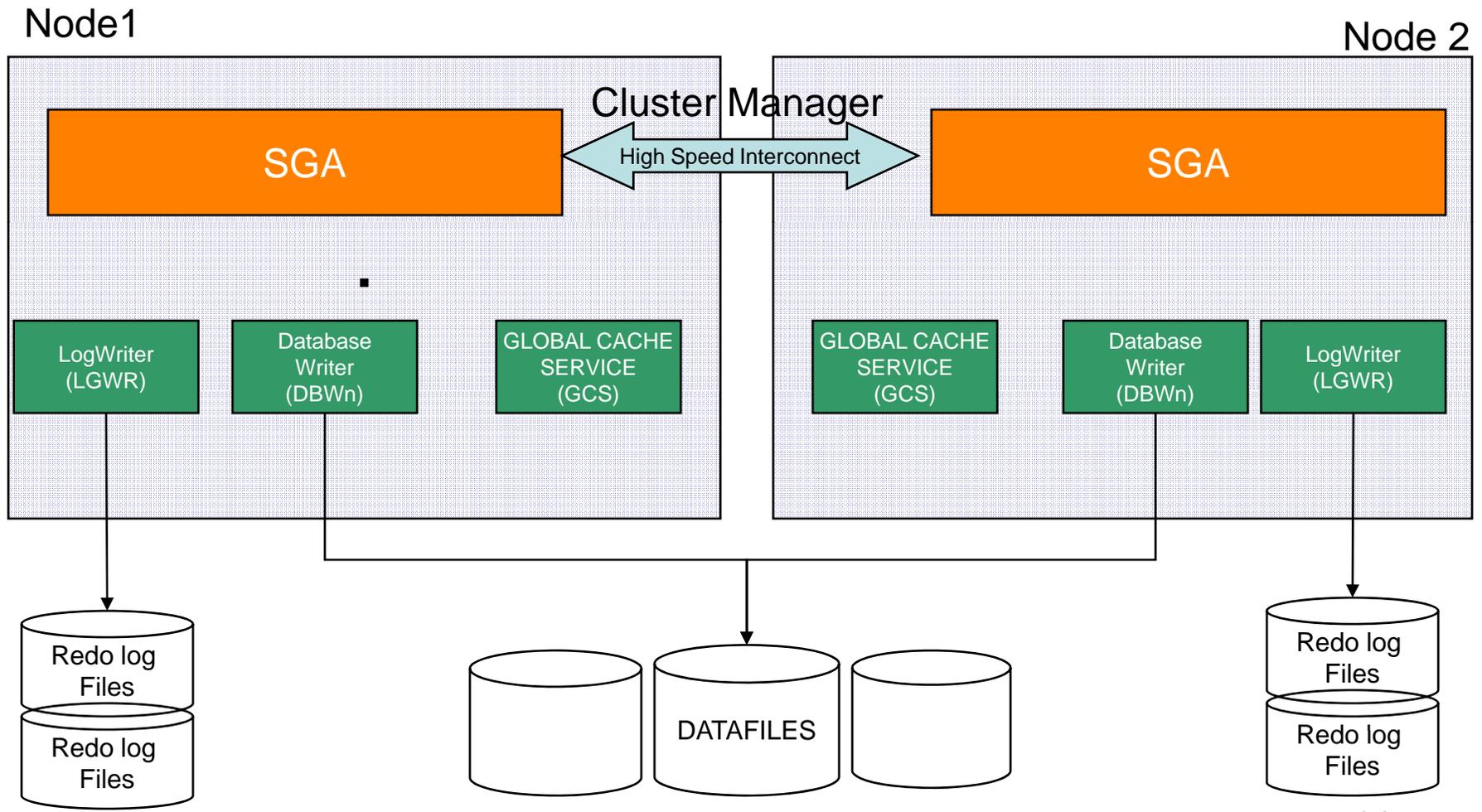
Migration plan another approach
-recovering database-

Restore database on new 64 bits cluster installation
  -Follow 467676.1 Oracle Metalink document

Time: 30 - 45 minutes
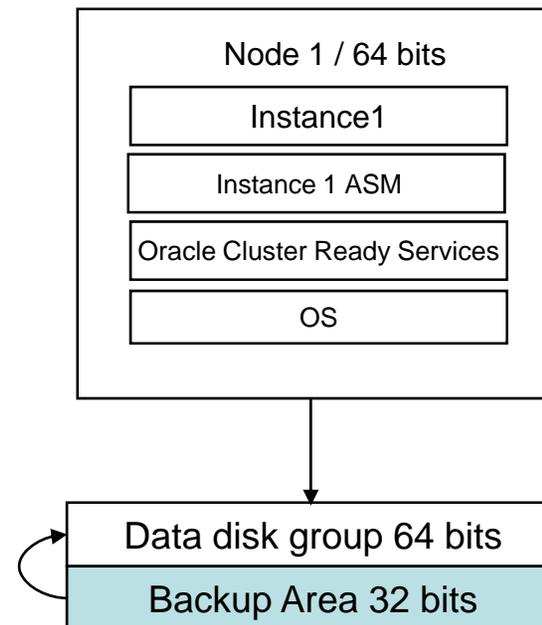
On node 1

-Restore database using RMAN

   Restore control file
   Restore database

-Migrating to 64 bit

SQL> recover database until cancel using backup controlfile;
SQL> alter database open resetlogs migrate;
SQL> @ ?/rdbms/admin/utlirp.sql
SQL> @ ?/rdbms/admin/utlrp.sql
SQL> shutdown immediate;

SQL> startup

```
Node 1 / 64 bits

Instance1

Instance 1 ASM

Oracle Cluster Ready Services

OS
```

```
Data disk group 64 bits

Backup Area 32 bits
```

Migration plan
-OS 64 bits installation, Alternative 2-

Remove one node 1 from cluster, database service open.
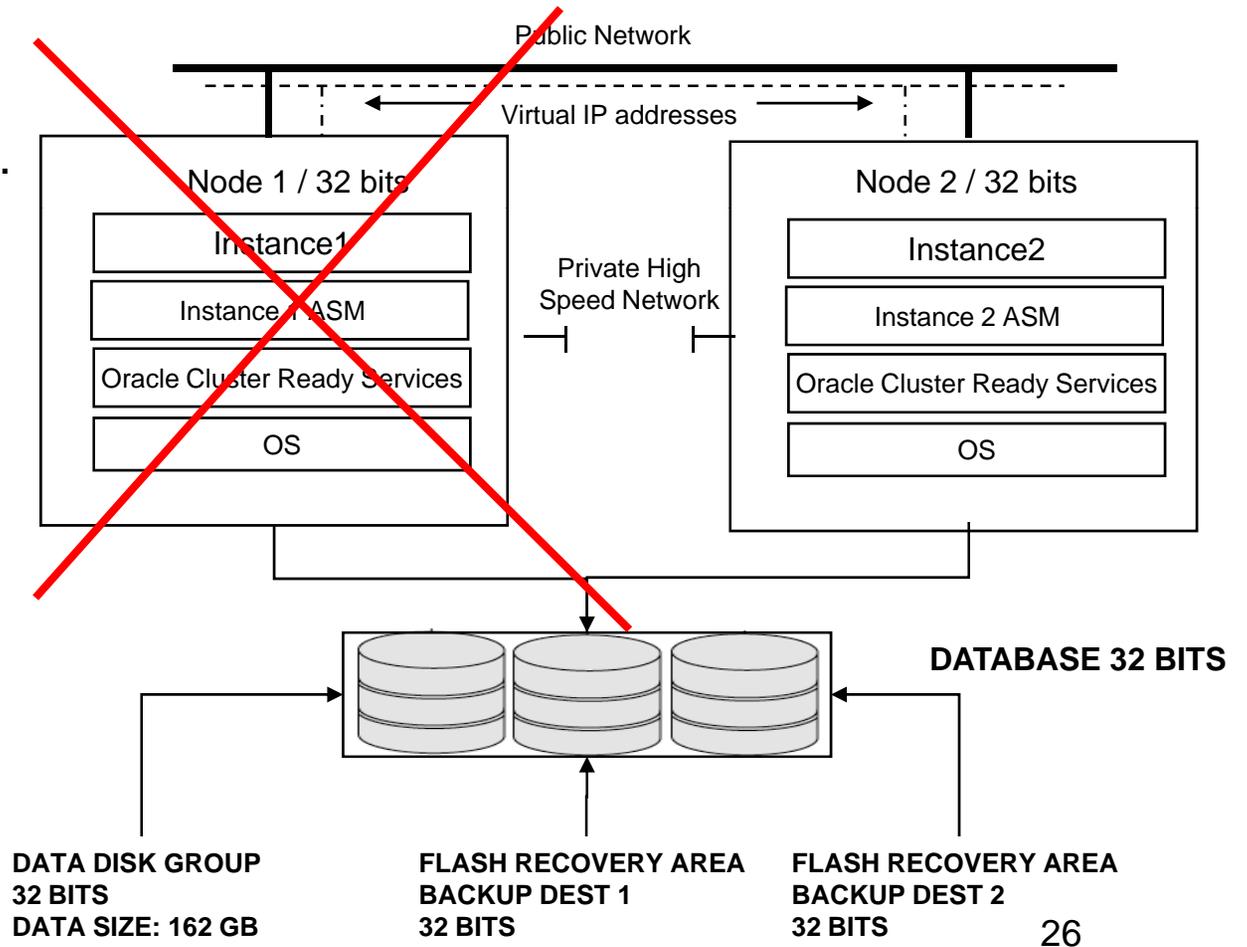
Preparing first node for upgrade

-SHUTDOWN INSTANCE1
(NO STREAM PROCESS RUNNING NODE 1)

-VERIFY STATUS DATABASE
OPEN AND RUNNING ON INSTANCE 1 and stream process is running

-To remove NODE 1 from cluster

Apply procedure presented last WLCG workshop by Jacek Wojcieszuk

http://indico.cern.ch/getFile.py/access?contribId=14&sessionId=7&resId=1&materialId=slides&confId=20080

Public Network

Virtual IP addresses

Node 1 / 32 bits
Instance1
Instance 1 ASM
Oracle Cluster Ready Services
OS

Private High Speed Network

Node 2 / 32 bits
Instance2
Instance 2 ASM
Oracle Cluster Ready Services
OS

DATABASE 32 BITS

DATA DISK GROUP
32 BITS
DATA SIZE: 162 GB

FLASH RECOVERY AREA
BACKUP DEST 1
32 BITS

FLASH RECOVERY AREA
BACKUP DEST 2
32 BITS

26

What if does not work?

- In case of loosing the entire data or data totally corrupted on migration process
    - Recover from second backup destination
    - Cluster will be taking out of production and data resyncronized separately from CERN. Then will be included on production.

  Time intervention: 1 to 2 days

- Installation OS/Oracle 64 bits fails

  Time intervention = 1 to 2 days
    - Cluster will be taking out of production and Oracle will be reinstalled and the data resyncronized separately from CERN. Then will be included on production.

- **User impact: None, will get data from closest Tier 1 3D site**