Oracle for administrative, technical and Tier-0 mass storage services

openlab Major Review Meeting 2009

29 January 2009

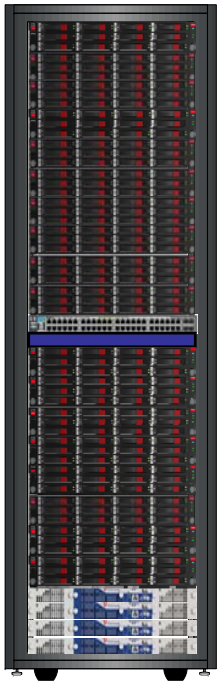
Andrei Dumitru, Anton Topurov, Chris Lambert, Eric Grancher, Lucia Moreno Lopez, Pablo Martinez Pedreira

Oracle Exadata

JST - Exadata Storage Server

Exadata Storage Server

Racked Exadata Storage Servers



- **Building block of massively parallel Exadata Storage Grid**
  - Up to 1GB/sec data bandwidth per cell
- HP DL180 G5
  - 2 Intel quad-core processors
  - 8GB RAM
  - Dual-port 4X DDR InfiniBand card
  - 12 SAS or SATA disks
- Software pre-installed
  - Oracle Exadata Storage Server Software
  - Oracle Enterprise Linux
  - HP Management Software
- Hardware Warranty
  - 3 YR Parts/3 YR Labor/3 YR On-site
  - 24X7, 4 Hour response

Exadata Important Features

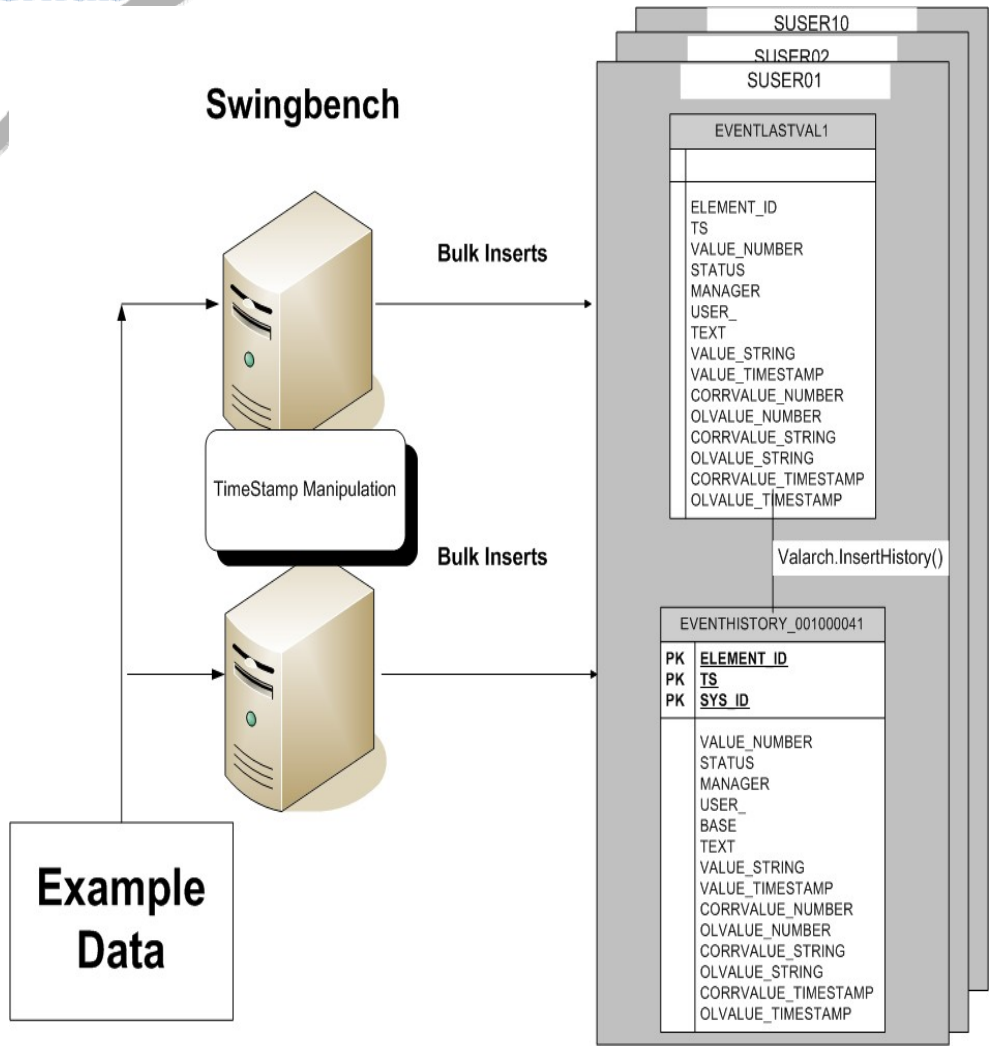
Database aware storage – does:

- Predicate filtering
- Column projection filtering
- Join processing (star-joins for DWH)
- ***Tablespace creation***
  - eliminates the I/O associated with the creation and writing of tablespace blocks
- I/O resource management – inter and intra database

JST – Oracle Exadata

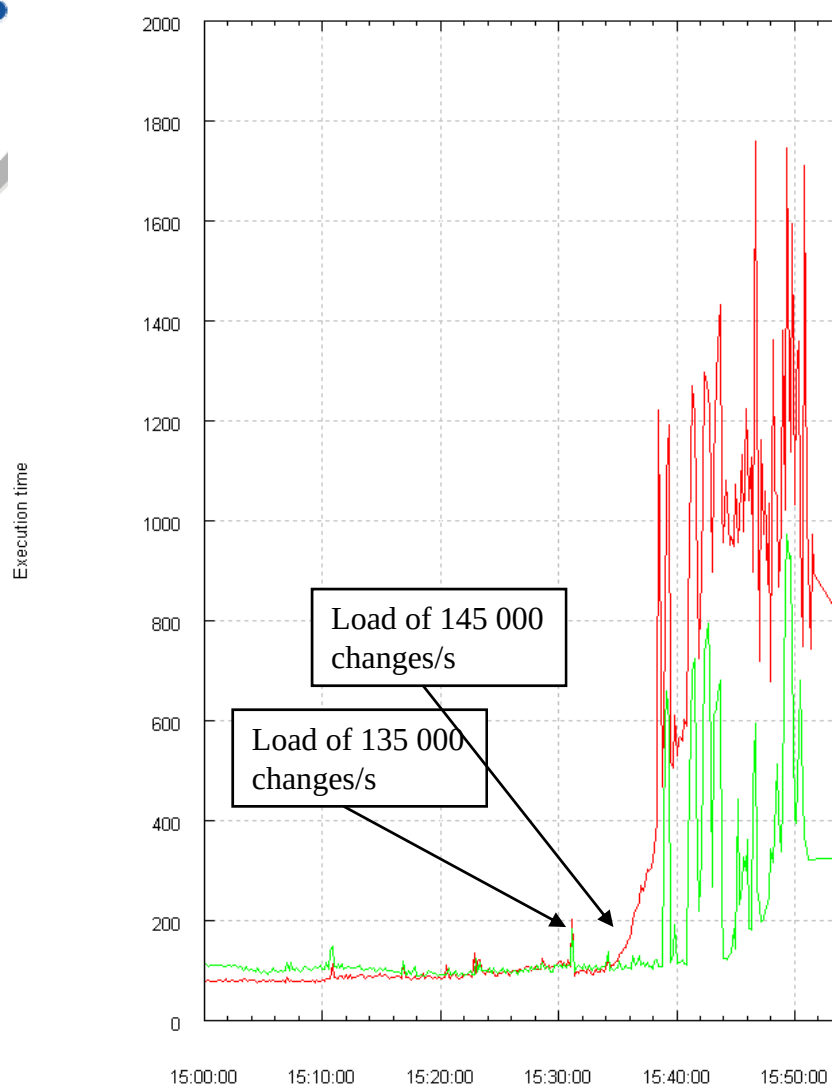
- Involved in the project since April 2008
  - Kick-off event in Reading, Apr'08
  - 1st phase of testing, Sept'08
  - 2nd phase of testing, Oct'08
- Exadata features could be of benefit for highly intensive data loading applications (PVSS, ACCMEAS/ACCLOG…)
- Database level PVSS workload simulator was developed to help testing this feature

PVSS Workload Generator



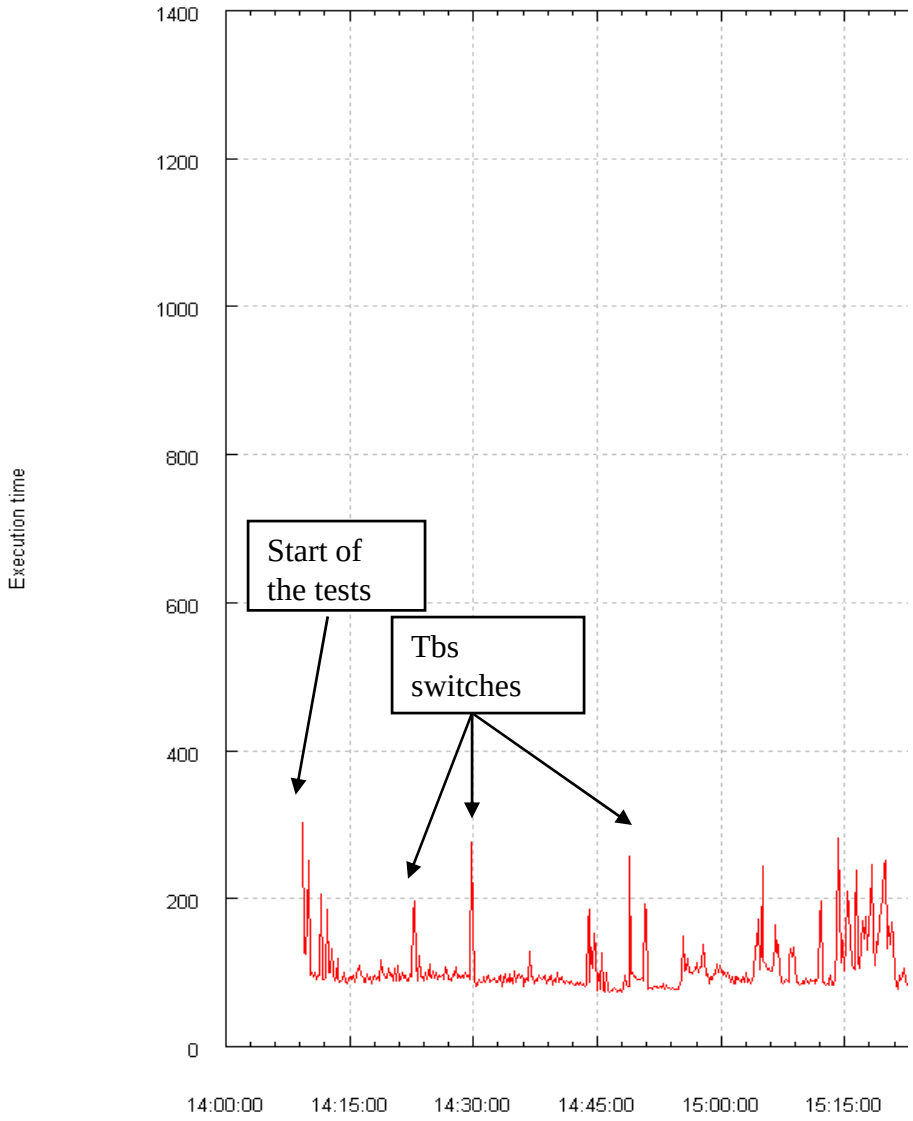
- 1000 changes/s as a bulk insert by each session
- Single Swingbench instance runs 15 sessions towards a dedicated schema
- 10 swingbench instances in total needed to generate top load of 150 000 changes/s

Tests



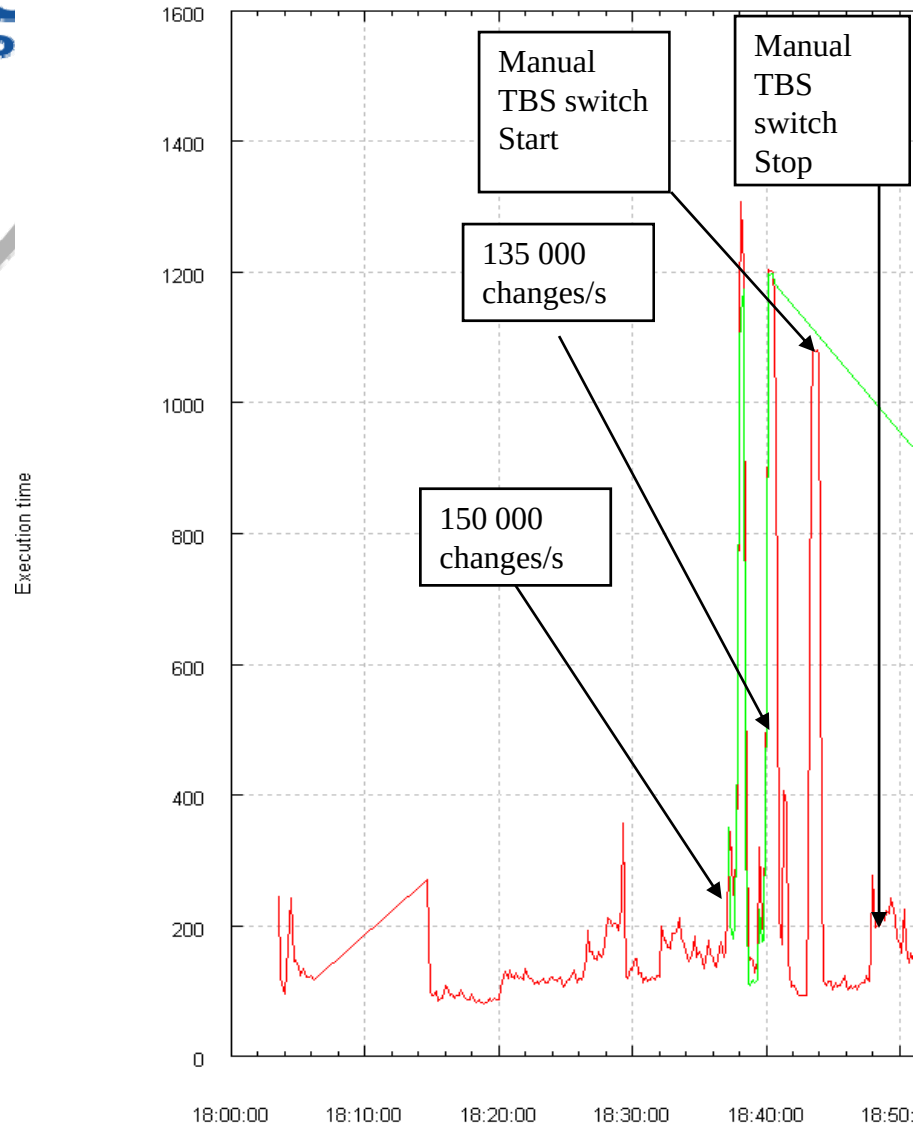
- 4-Node RAC with 4 Cells storage
- 10 GB SGA
- 20 GB Tablespaces, 5 MB Uniform size
- Last stable point 145000 changes/s

Fast file creation (1)



- Feature of Exadata
- _cell_fcre = true
- 17 seconds for 20 GB Tablespace
- Little spike in execution times
- Much below 1000 ms
- threshold

Fast file creation (2)



- _cell_fcre = false
- ~ 2 minutes to create 20GB tablespace
- Much bigger spike
- Higher then 1000ms threshold

Conclusions

- 4-Node RAC setup with Exadata storage:
  - could sustain up to 145 000 changes/s
  - bottleneck on concurrent change of control files
  - Much faster file creations lead to less spikes in execution times
- Overall better performance with Exadata storage features on.
- Next steps:
  - possibility get hardware onsite
  - or get Exadata software installed on our storage

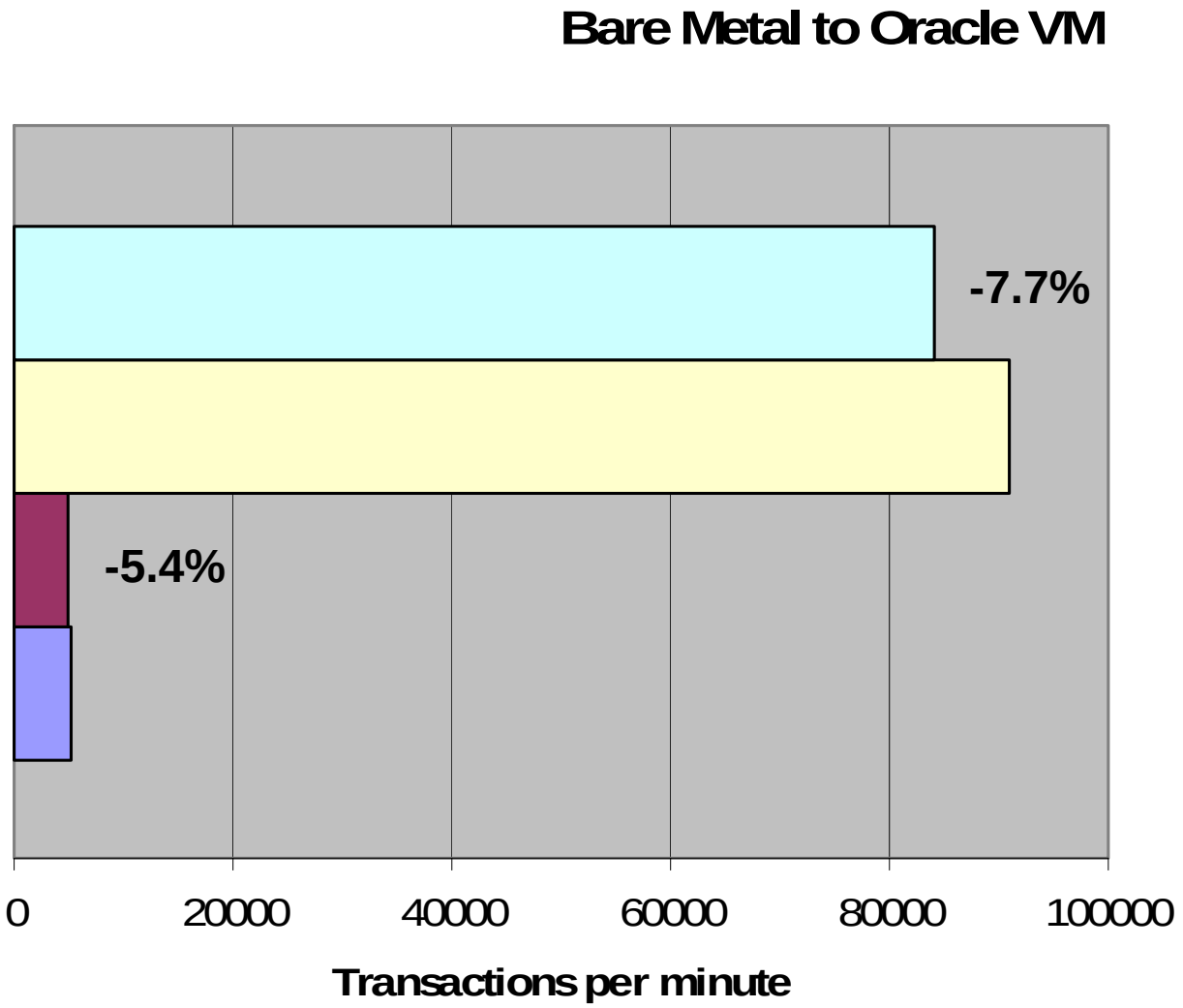
Virtualization

Oracle Database Virtualization

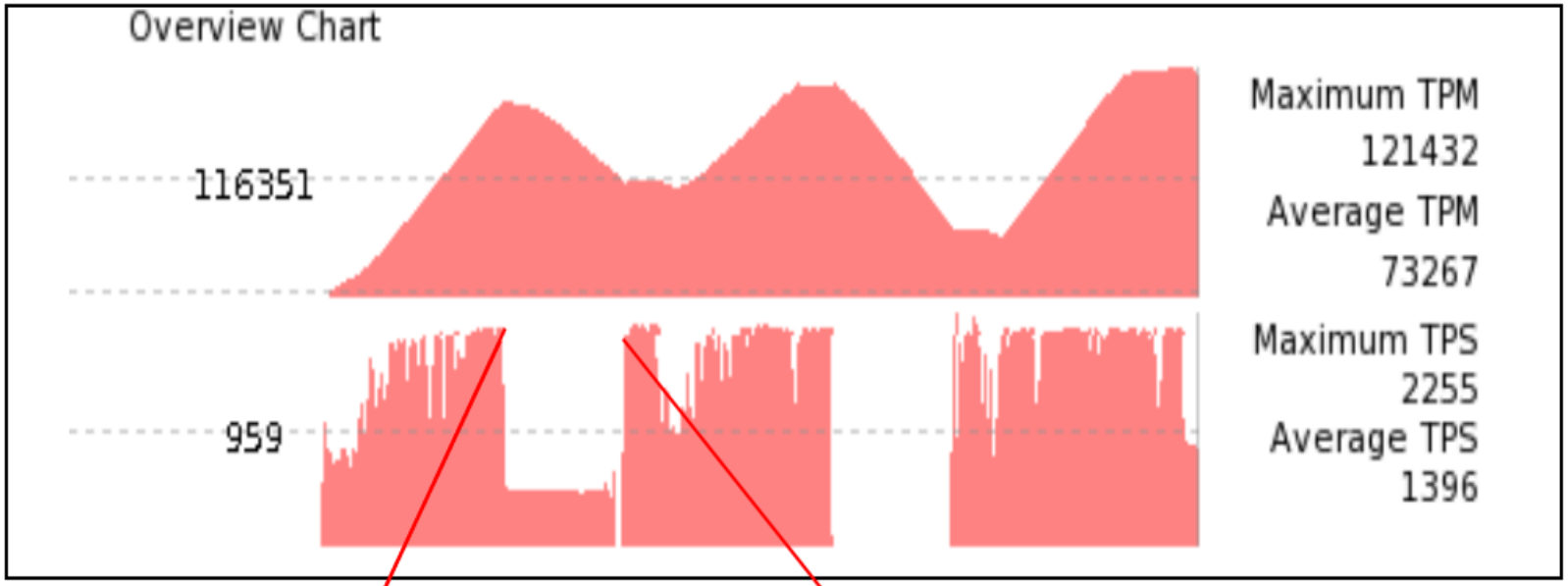
- Main focus:
  - Measure the overhead
  - Test of Oracle RAC on Oracle VM and OEL5
  - Test of Oracle RAC on OEL5 and pure XEN
  - Test of Hardware Virtualization vs. para-virtualization
- Work done by Andrei Dumitru and Anton Topurov
- Results show better performance and ease of use for Oracle VM solution
- Live migration has almost no downtime

Bare Metal
vs
Oracle VM

Bare Metal to Oracle VM



Live Migration



Node 1

```
#xm list
Name          ID   Mem VCPUs    State   Time(s)
Domain-0       0   834    8       r-----  1773.7
virt04         8  4096    8       -b----   517.4
```

```
# xm migrate virt04 node2 --live
```

```
# xm list
Name          ID   Mem VCPUs    State   Time(s)
Domain-0       0   834    8       r-----  1785.7
migrating-virt04 8 4096    8       r-----   538.3
```

```
# xm list
Name          ID   Mem VCPUs    State   Time(s)
Domain-0       0   834    8       r-----  1851.5
```

Node 2

```
# xm list
Name          ID   Mem VCPUs    State   Time(s)
Domain-0       0   834    8       r-----  2410.8
```

```
# xm list
Name          ID   Mem VCPUs    State   Time(s)
Domain-0       0   834    8       r-----  2444.8
virt04        11  4096    0       -bp---     0.0
```

```
# xm list
Name          ID   Mem VCPUs    State   Time(s)
Domain-0       0   834    8       r-----  2481.1
virt04        11  4096    8       -b----     6.4
```

Virtualisation - Conclusions

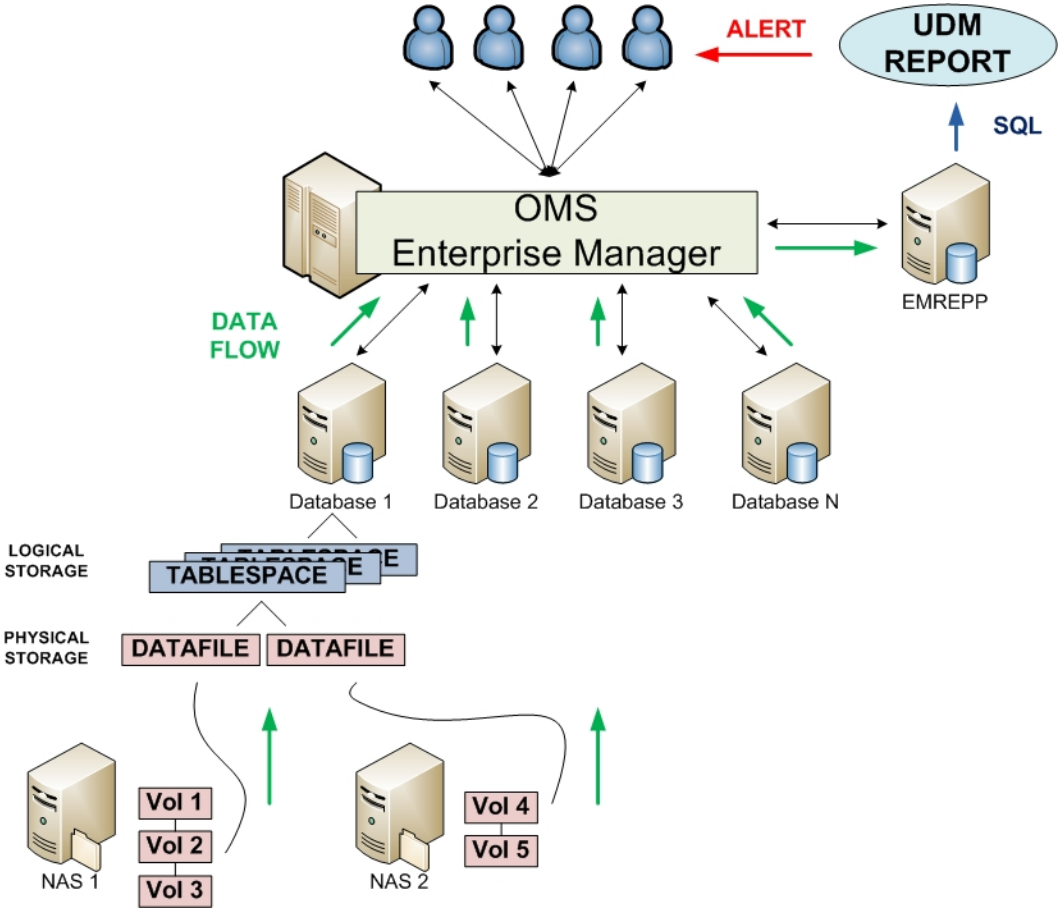
- Both Xen and Oracle VM are quite stable
- Oracle VM is tuned for database usage
- Oracle supports only Oracle VM based setup
- RAC is now certified, but with restrictions
- Live migration is a big plus for deployment (patching, hardware intervention, move to new HW…)
- Next EM is virtualization aware
- Deploying devdb11 as a pilot virtualized database

Enhanced storage monitoring

- ‘Plug-ins’ to Grid Control give space and performance information at storage level.
- BUT, no link between datafile growth and volume extension.
- Tighter integration between database and storage layers is required.

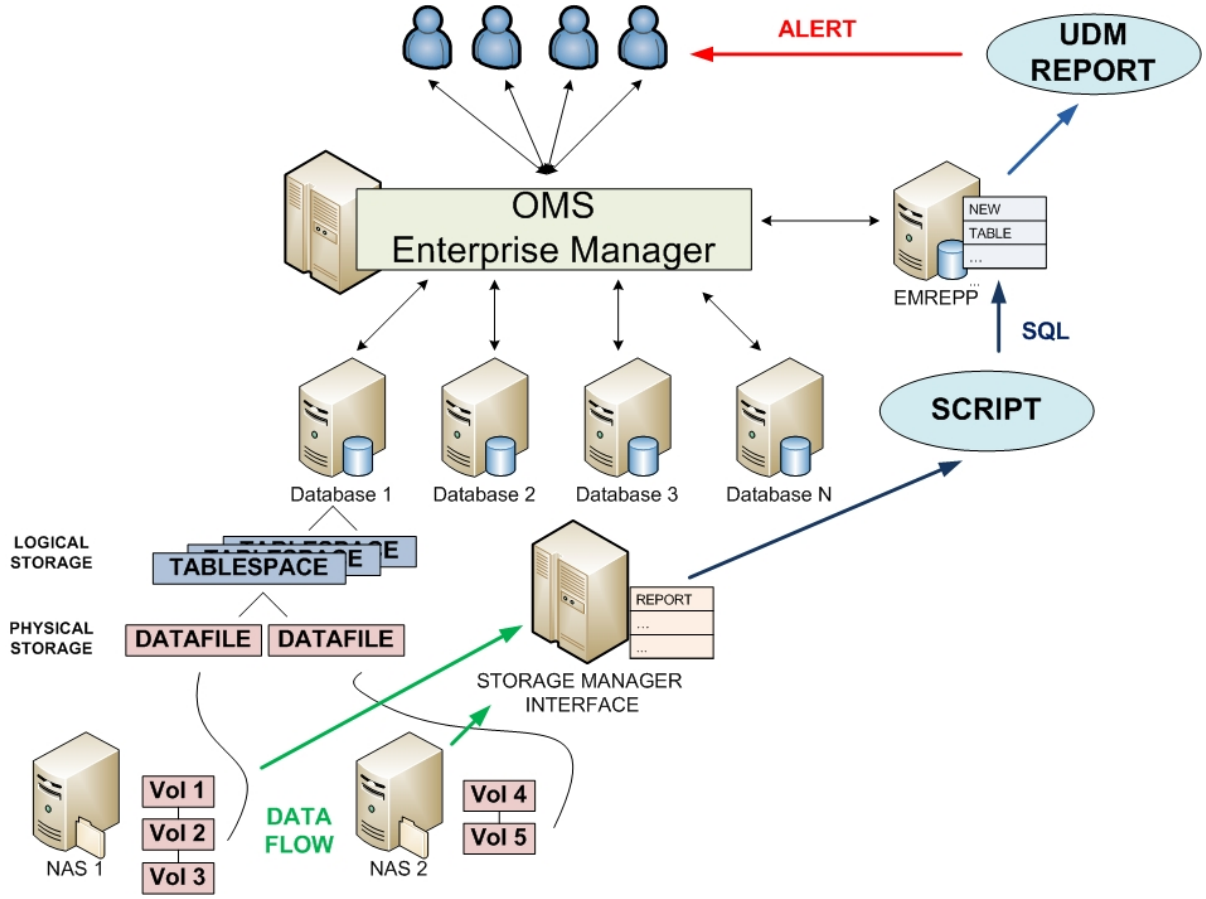
Enhanced storage monitoring

- Out of the box monitoring



Enhanced storage monitoring

- Extended Storage monitoring



Enhanced storage monitoring

Advantages

- Identify datafiles that cannot extend because underlying volume cannot grow
- Identify underlying volumes that are not mounted on our servers

Future

- Oracle Weblogic
- Enterprise Manager and security
- Newer version of Enterprise Manager
- Upgrade repository database to 11.1, benefit of Active Dataguard (already using DataGuard)
- Virtualisation