

DOAG Regionaltreffen München/Südbayern

Oracle Private Database Cloud (DBaaS) - Example Implementation

Robert Bialek

Principal Consultant, MU-IMS

December 10, 2012

BASEL BERN LAUSANNE ZÜRICH DÜSSELDORF FRANKFURT A.M. FREIBURG I.BR. HAMBURG MÜNCHEN STUTTGART WIEN

1

2012 © Trivadis

Oracle Private Database Cloud (DBaaS) - Example Implementation
December 10, 2012

trivadis
makes IT easier. ■ ■ ■

Who Am I

- Principal consultant and trainer at Trivadis GmbH in Munich
 - robert.bialek@trivadis.com
- Focus: Oracle database high availability
 - Architecture design
 - Review, troubleshooting, coaching
 - Backup and recovery
 - Performance tuning
 - Linux administration/troubleshooting
- Trainer for the following Trivadis courses:
 - Oracle Grid Infrastructure (O-GRINF)
 - Oracle Real Application Clusters (O-RAC)
 - Oracle Data Guard (O-DG)

ORACLE
Certified Master

trivadis
makes IT easier. ■ ■ ■

Disclosure

- Project implemented at a big **global** acting company
- Due to the internal policy, Trivadis is **not allowed** to mention the name of the company in this presentation

AGENDA

1. Introduction
2. Operating System
3. Grid Infrastructure
4. Database Management

Introduction (1)

- In 2011 Trivadis took over at the customer the **24/7 operation** of all Oracle databases
 - Including middleware (OAS, OID)
 - And operating system (SUN Solaris, Linux)
- One of the biggest challenges was to replace the **old database infrastructure** with a **new Oracle Private Database Cloud (DBaaS)**
- The **old** infrastructure
 - Hardware: Sun Fire E6900
 - 8 domains, Sun Solaris 10
 - Veritas failover cluster
 - Veritas volume manager
- **Key project drivers**
 - Performance/scalability/stability/support issues with the old infrastructure
 - More flexibility and process optimization necessary



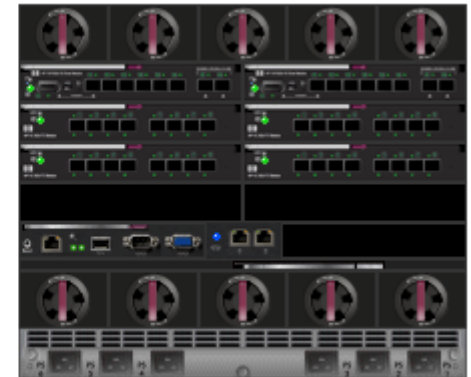
Introduction (2)

- Considering the project requirements, the decision was made to use the following infrastructure
- **The new hardware:**
 - 4 x HP BladeSystem c7000 Enclosures
 - 40 ProLiant BL460c Gen[6/7/8]
 - **Two** spare blades pro enclosure
 - Prod: 256GB RAM, Test: 196 GB
 - All hardware components are redundant
- Each server has it's **own** server profile (VC defined NIC **MAC** and HBA **WWPN**)
- Ethernet and FC connections can be changed for each server separately

Front View



Rear View



Introduction (3)

- **The new software stack:**
 - Oracle Enterprise Linux
 - Oracle Grid Infrastructure
 - Real Application Clusters (stretched configuration) – Data Guard (optional)
- The new Oracle database environment is best suited for the upcoming **Oracle 12c** version (PDBs, etc.)
 - Each application is using it's own database service (self-contained)
- Change at **EVERY** hardware and software layer
 - EMC DMX vs. Fujitsu Ethernus DX8000
 - SPARC vs. x86-64
 - SAN Solaris vs. Oracle Enterprise Linux
 - Veritas cluster/volume manager vs. Oracle Grid Infrastructure
 - Single instance vs. Real Application Clusters (optionally plus Data Guard)
 - Monitoring, etc...

Introduction (4)

- Project timeline



Pilot Project

- New or very performance demanding applications
- New tools/backup framework
- New monitoring framework
- Ongoing optimizations

Migrations Optimizations

- In many cases database consolidation
- Migrations with Trivadis framework
- Ongoing optimizations

Hardware Change Data Center Migration Optimizations

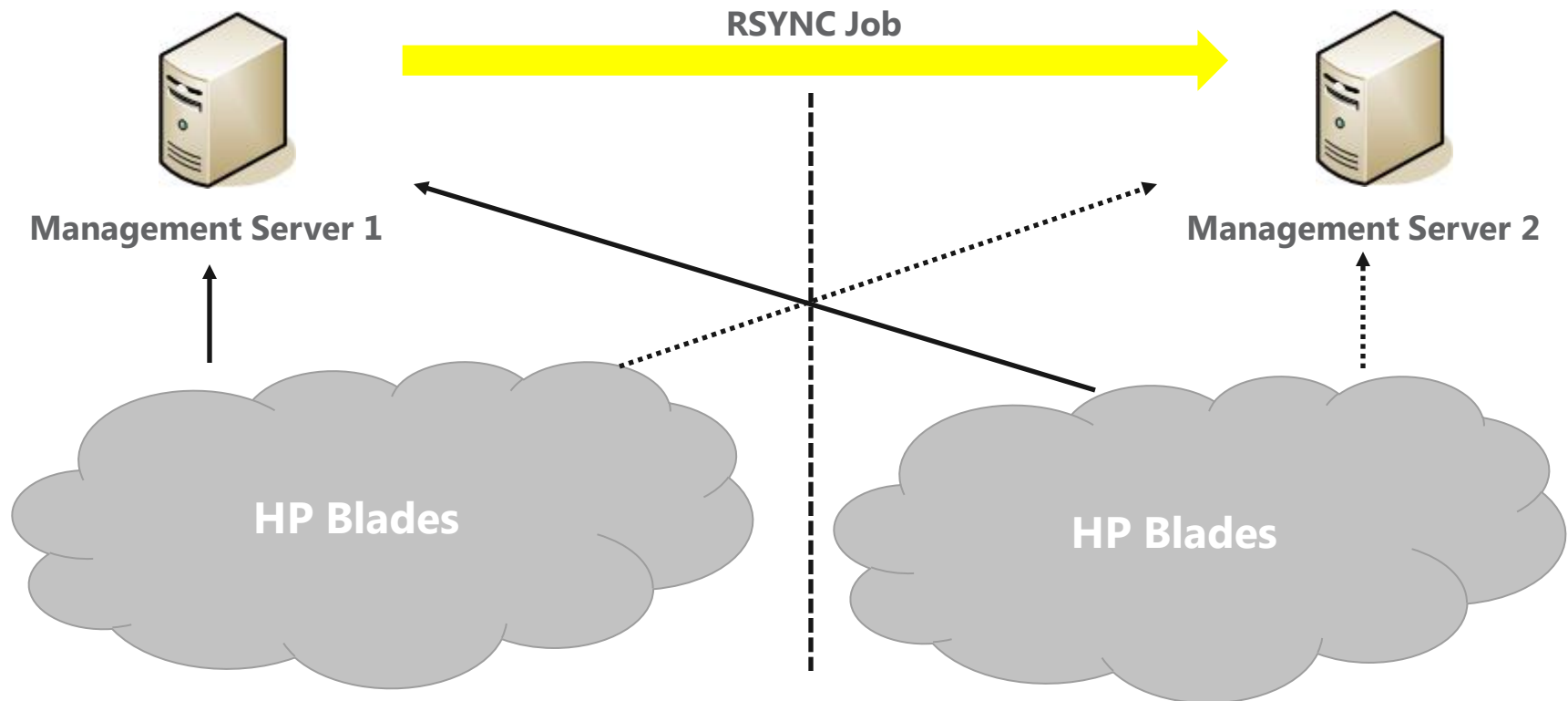
- Migration to the new HP BladeSystem c7000 Enclosures
- Migration from HP BL460c Gen6 to Gen 8
- Security: new clusters, local firewalls, admin. accounts, etc.
- Migration to the new data center
- Even more flexibility - Server Pools, PM databases, new 12c features,...

AGENDA

1. Introduction
2. Operating System
3. Grid Infrastructure
4. Database Management

Operating System – Provisioning (1)

- Two Linux installation servers – **custom** implementation
 - DHCP, TFTP and local YUM repositories
 - Management server 1 → **active**, management server 2 → **passive**



Operating System – Provisioning (2)

- To provision a new server
 - Server “One Time Boot” → PXE NIC boot



HP BL460c

- Installation duration ~ **10-15 min.**
 - Ready to use, patched Oracle Enterprise Linux (OEL) 5.7 x86-64
 - **Minimal** installation (no GUI), **many services** have been deactivated
 - **RH** compatible **kernel** (UEK will be considered in the future)
 - With all users and settings necessary for the Grid Infrastructure/database installation

Management Server

Kernel
Initial RAM Disk

Base YUM Repo

YUM Repo with
OS patches

Konfig. Files

Third Party
Software

Operating System – Shared Storage and I/O Multipath

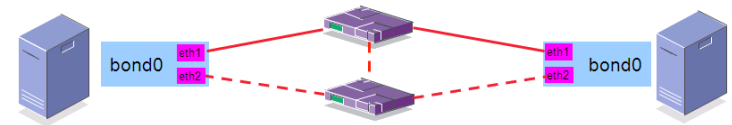


- Multipath software: Linux Device Mapper
 - With OEL 5.7 and RH kernel downgrade for some RPMs is necessary

```
rpm -Uvh --force --nodeps kpartx-0.4.7-46.el5.x86_64.rpm \  
> device-mapper-multipath-0.4.7-46.el5.x86_64.rpm \  
> mkinitrd-5.1.19.6-71.el5.x86_64.rpm  
  
/etc/yum.conf:  
exclude=device-mapper-multipath* kpartx* mkinitrd*
```

- **Shared Storage: ~ 90 TB**
 - Each ASM disk used for databases → **200GB**
 - LUNs for cluster voting files → **2GB** (6 LUNs / cluster)
 - **4 I/O paths** to each LUN
- Example production cluster with 8 nodes
 - **440** multipath device maps with **4 paths** to each LUN → **1760 block devices**
 - **Quality checks** framework (e.g. LUN aliases) strictly necessary

Operating System – Network Configuration



- Linux Bonding for public and private network

```
grid@server1:~/ [+ASM1] oifcfg getif
bond0 10.250.54.0 global public
bond1 192.168.0.0 global cluster_interconnect
```

- Bonding mode: **active-backup**

- Active-Active not possible → links do not terminate on the same switch
- For “manual” load balancing we use the **primary** parameter

```
grid@server1:~/ [+ASM1] cat /etc/sysconfig/network-scripts/ifcfg-bond0
DEVICE=bond0
...
BONDING_OPTS="miimon=100 mode=active-backup primary=eth1"
```

```
grid@server2:~/ [+ASM1] cat /etc/sysconfig/network-scripts/ifcfg-bond0
DEVICE=bond0
...
BONDING_OPTS="miimon=100 mode=active-backup primary=eth0"
```

Operating System – Central Syslog Server

- Both install servers are used also as a central syslog servers
 - syslog-ng installation on both management servers



Management Server 1



Management Server 2



```
Sep 19 11:35:53 server1 kernel: kernel: sd 0:0:19:40: timing out command, waited 300s
Sep 19 11:35:53 server1 multipathd: /sbin/mpath_prio_alua exited with 1
Sep 19 11:35:53 server1 multipathd: error calling out /sbin/mpath_prio_alua /dev/sdsx
```



```
*.info;mail.none;authpriv.none;cron.none /var/log/messages
*.info;mail.none;authpriv.none;cron.none @mgmt1
*.info;mail.none;authpriv.none;cron.none @mgmt2
```

Operating System – Miscellaneous

- **Huge Pages** have been set up on all nodes
 - Default: 50 % RAM (not compatible with AMM)
- Kernel change from **RH** compatible to **UEK**
 - Tests are ongoing (active in one cluster since 10.2012)
 - Change planned for 2013
- Upgrade to OEL 6.X currently not possible (Oracle 10.2 RDBMS support)
- **Ksplice** will not be used
 - With RAC a server reboot is generally not a problem
 - After applying OS patches a server reboot is even **desired**

Operating System – Add-On Tools

- **OSWatcher** has been installed on every database server
 - Snapshot every 30 sec. Archive retention → 5 days

```
[root@server1 ~]# chkconfig --list oswbb
oswbb          0:off  1:off  2:on   3:on   4:on   5:on   6:off

/home/oracle/tools/oswbb/<SERVER>:
drwxr-xr-x 2 nfsnobody nfsnobody 16384 Sep 23 11:00 oswiostat
drwxr-xr-x 2 nfsnobody nfsnobody 16384 Sep 23 11:00 oswmeminfo
drwxr-xr-x 2 nfsnobody nfsnobody 16384 Sep 23 11:00 oswmpstat
drwxr-xr-x 2 nfsnobody nfsnobody 16384 Sep 23 11:00 oswnetstat
...
```

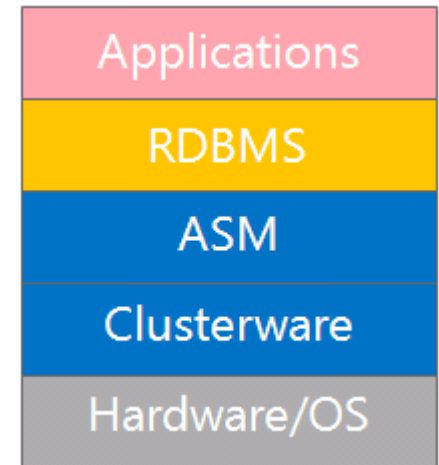
- Daily job to collect **storage data** from every server
 - multipath topology
 - /dev/disk/by-path/*
- **Daily quality** checks
 - LUN aliases correct/consistent across all cluster nodes?

AGENDA

1. Introduction
2. Operating System
3. Grid Infrastructure
4. Database Management

Grid Infrastructure – Overview

- Grid Infrastructure installation type:
 - Software only
- Grid Infrastructure has been installed using **role separation**
 - **grid** user – cluster administration
 - **oracle** user – database administration
- Current topology:
 - **One** production cluster: **8 nodes** [414 cluster resources]
 - **Two** test clusters: **10 nodes** [638 cluster resources] and **2 nodes** [127]
 - DBA test environment: **4 nodes**
 - Dedicated **security** production (**4 nodes**) and test (**4 nodes**) clusters with ASO/Kerberos will be installed in Q1 2013



Grid Infrastructure – SCAN

- **Cluster Name = Single Client Access Name**
 - One **SPOC** for **all** Oracle clients (OCI,JDBC,...)
 - SCAN configured with DNS

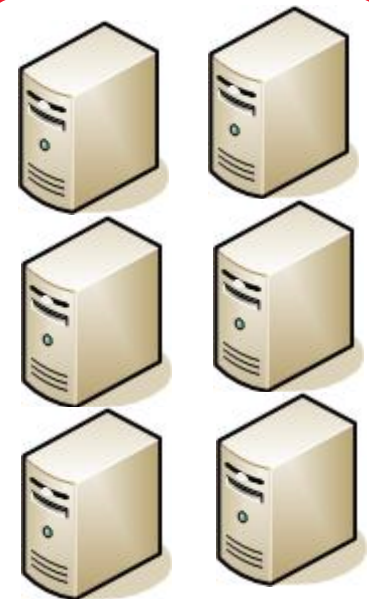


HOST=**CLUSTER1**



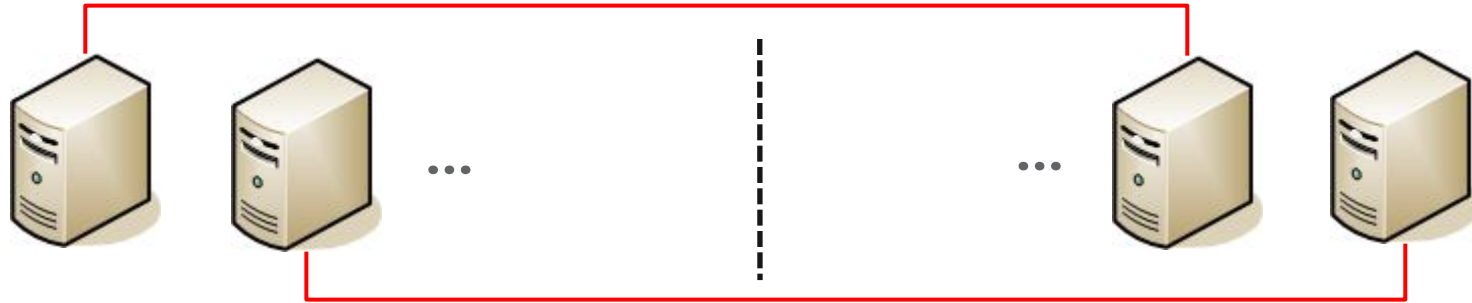
```
jdbc:oracle:thin:@cluster1:15400/app_t  
sqlplus scott@cluster1:15400/app_t
```

CLUSTER1



Grid Infrastructure – Logical Cluster Partitioning

- All clusters have been logically partitioned



- Example:
 - SERVER1 (RZ1) – SERVER5 (RZ2)
 - SERVER2 (RZ1) – SERVER6 (RZ2)
 - SERVER3 (RZ1) – SERVER7 (RZ2)
- **Server Pools** (cardinality approach) are considered to be used later with PM databases

Grid Infrastructure – ASM Disk Groups

- All ASM disk group names (normal redundancy) are **globally** unique
 - Easy LUN (database) migration between clusters
 - For non business critical databases → storage consolidation

+U2010 - Data Files, Control/Redo Files

+U2020 - FRA, Control/Redo Files

Grid Infrastructure – Monitoring/Add-On Tools

- Wrapper scripts for nagios

```
nagios ALL=(grid) NOPASSWD: /usr/local/wrapper/check_voting_files.sh
nagios ALL=(grid) NOPASSWD: /usr/local/wrapper/check_cluster.sh
nagios ALL=(grid) NOPASSWD: /usr/local/wrapper/check_resources.sh
nagios ALL=(grid) NOPASSWD: /usr/local/wrapper/check_ocr.sh
nagios ALL=(grid) NOPASSWD: /usr/local/wrapper/check_asmdg.sh
nagios ALL=(grid) NOPASSWD: /usr/local/wrapper/list_cluster_nodes.sh
```

- **Daily quality** checks

Name	Failgroup	Failgroup_Type	Path
U5310_0021	RZ1	REGULAR	/dev/mapper/u5310_01_rz1p1
U5310_0000	RZ2	REGULAR	/dev/mapper/u5310_01_rz2p1

AGENDA

1. Introduction
2. Operating System
3. Grid Infrastructure
4. Database Management

Database Management (1)

- Many highly **consolidated** RAC databases – **2 instances/database**
 - Some with more than 50 applications
 - Only **two versions**: 11.2.0.3 and 10.2.0.5
 - Fix for Oracle Security Alert for CVE-2012-1675 has been implemented
 - Also with complex add-on features (n-way multimaster Streams replication)
- Each application uses its **own service** name in cluster

`<APPLICATION_NAME>_<P|T>` #Oracle Net alias in the same format in OID

- Default service configuration: **active/passive**

app1_t
app2_t



...



app3_t
app4_t
app5_t



...

T2501

T2502

Database Management (2)

- In the future more flexibility with **server pools** and **policy managed databases**
 - **Active/active** service configuration (stretched cluster!) for selected databases/applications
 - More than 2 instances for **selected** databases/applications
 - **12c**: database consolidation with PDBs
-
- But, for a **real cloud**, a flexible database infrastructure is **not enough**
 - What we need are **appropriate** software frameworks

Database Management – Tools (1)

- User **Self-Service Portal**
 - Database **service** configurator

Start **DB-Konfigurationen** SO-Admin Dienstleister-Admin Anwendungs-Admin

Menü
Neue Konfiguration
 Meine Konfigurationen

1. DB-Auswahl > **< 2. Applikation >** < 3. Hardware > < 4. DB-Klassifizierung > < 5. >

< >

* Applikationsname

* Applikationsgruppe

* Applikationsbeschreibung

Datenbankversion

Datenbank-Zeichensatz

Applikations-Server

Datenbank-Landschaft

	Umgebung	Applikationsname	Termin	Termin Ende
<input checked="" type="checkbox"/>	P		<input type="text" value="01.10.2012"/>	<input type="text"/>
<input checked="" type="checkbox"/>	T		<input type="text" value="01.10.2012"/>	<input type="text"/>
<input type="checkbox"/>	E		<input type="text"/>	<input type="text"/>
<input type="checkbox"/>	S		<input type="text"/>	<input type="text"/>

Konfiguration-ID 858
 Status Temporaer

Kosten pro Jahr
 Kosten pro Monat

Konfiguration

CPU 2.00
 Storage 1000

Serviceklasse

Datenklassifizierung Intern
 Datenintegrität Wichtig
 Kontinuität Kritisch

Database Management – Tools (2)

- **Trivadis Enterprise Service Center (ESC)**
 - Used by application owners, management and DBAs

Trivadis Enterprise Service Center

Overview DBA Service Manager Application Owner Management Administration Pa

Service Manager > SLA Report

SLA Report Parameter zurück

Datum von: 01.09.2012

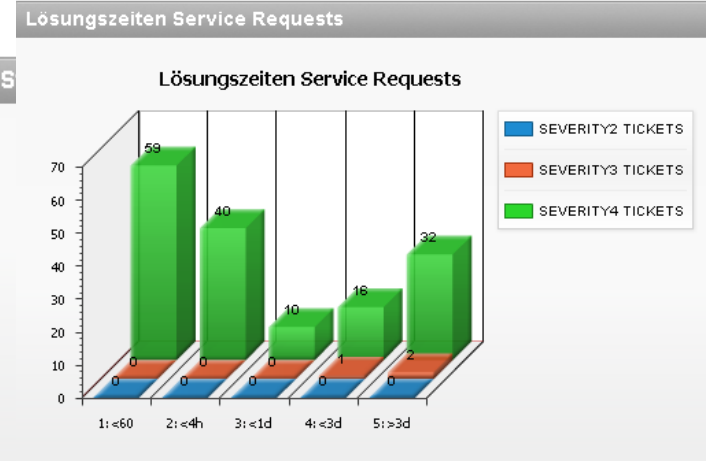
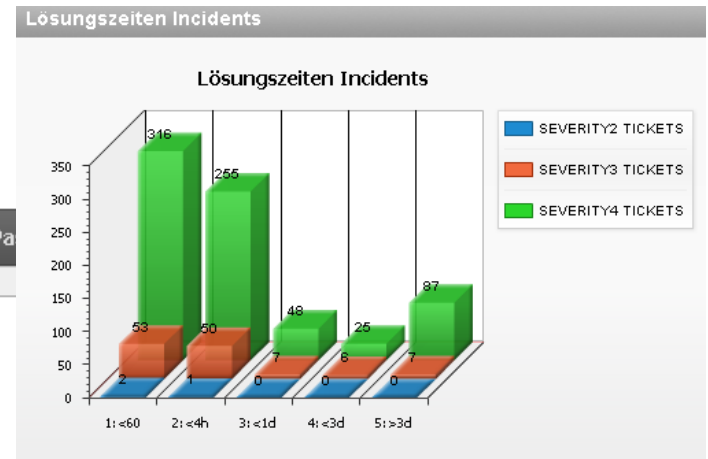
Datum bis: 30.09.2012

Refresh SLA Report

Severity Overview Incidents

Severity Overview Incidents

Severity	Count
Severity 2	1
Severity 3	121
Severity 4	712



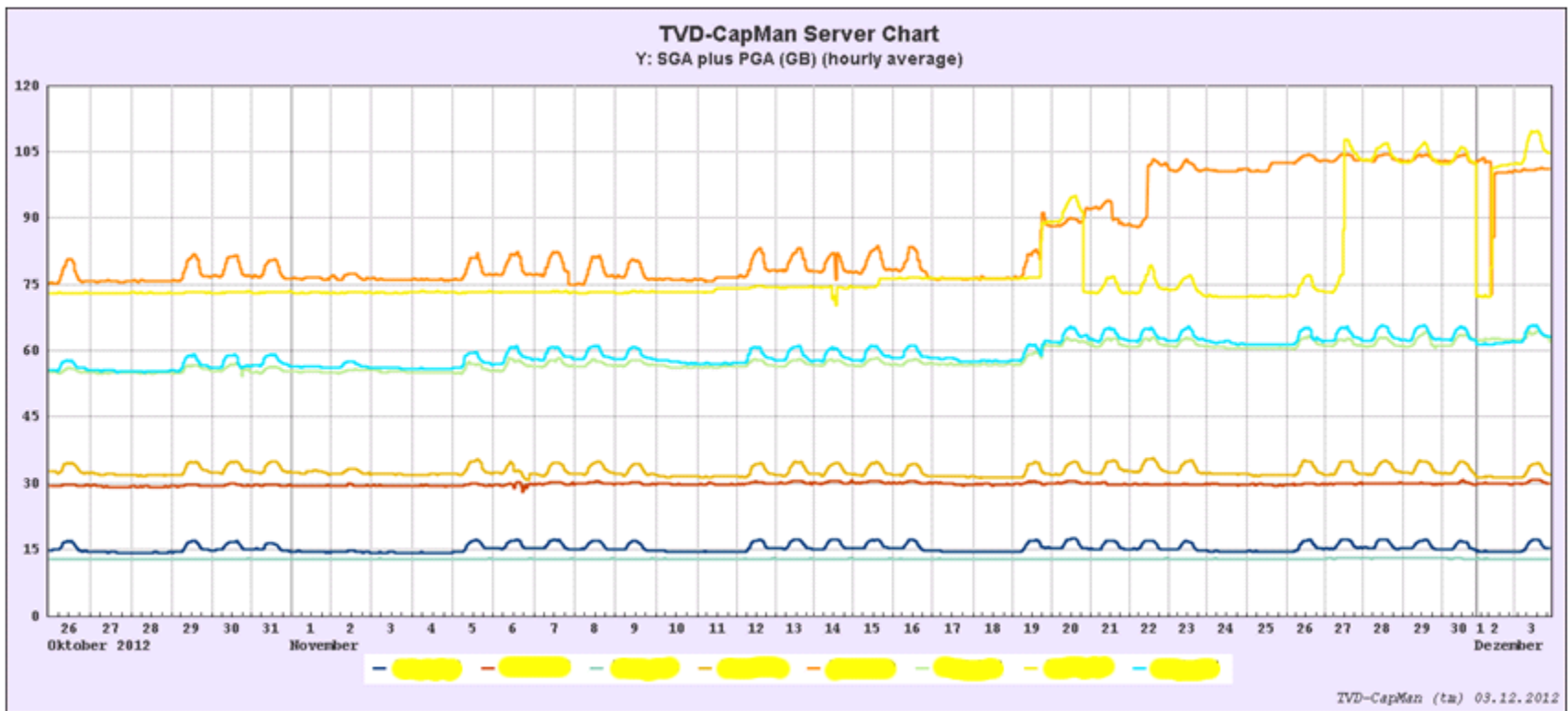
Database Management – Tools (3)

■ Trivadis Enterprise Service Center (ESC)

The screenshot displays the Trivadis Enterprise Service Center (ESC) interface for user management. The top navigation bar includes 'Overview', 'DBA', 'Service Manager', and 'Application'. The current page is 'Application Owner > Self Service'. The main content area is titled 'User Management für MAXIMO V4' and features a sidebar with icons for 'User anlegen', 'User löschen', 'lock/unlock', 'Passwort ändern', and 'Userliste'. The main panel is titled 'Datenbank auswählen' and shows 'DB: E132' and 'Vorlage: MAXIMO-User'. A 'Übernehmen' button is present. Below this, the 'User anlegen für Applikation: MAXIMO V4 in DB: E132' form is shown, with fields for '*User:', '*Passwort Type:' (Local/External), '*Type:' (Connect), 'Passwort:', 'Standard-Tablespace:', 'Other-Tablespaces:', 'Temp-Tablespace:', 'Privilegin und Rollen:', 'Standard Rolle:', 'Profile:', and 'Proxy User:'. An 'Ausführen' button is at the bottom.

Database Management – Tools (4)

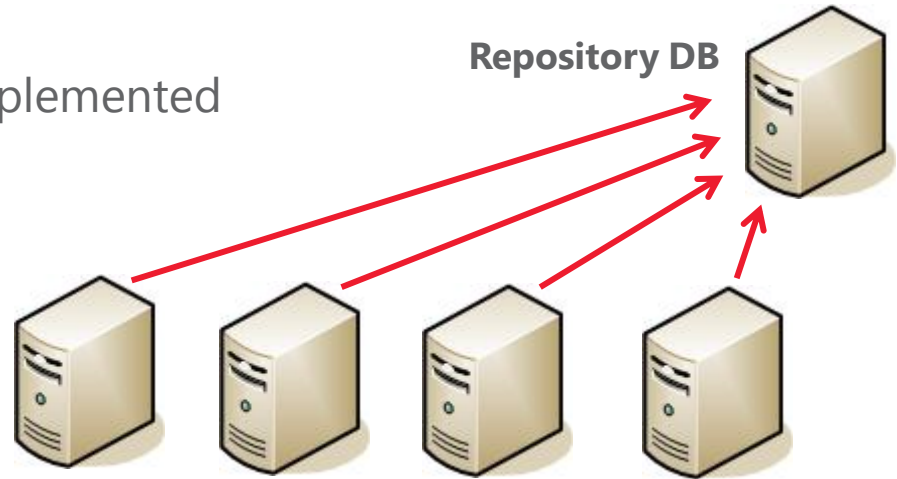
- **TVD-CapMan** - capacity management for Oracle databases
 - Example SGA and PGA memory usage per server



Database Management – Tools (5)

- Backup framework based on **TVD-Backup** with a new **add-on scheduler**

- Currently only CLI, GUI will be implemented



- Backup agent

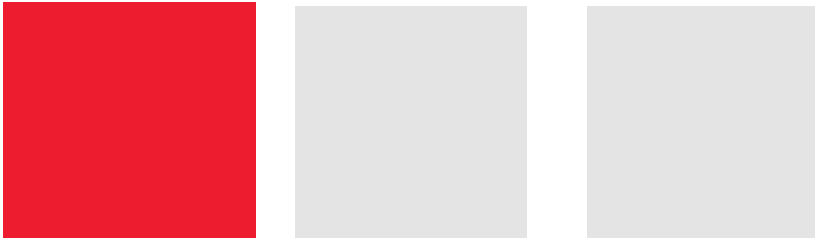
```
*/5 * * * * . /u00/app/oracle/local/dba/bin/basenv.bash >/dev/null 2>&1  
; cd /u00/app/oracle/local/custom/backup; ./backup_agent -p 6 -d -l log
```

- Checks all databases instances on a server
- Sends the information to a repository database
- PL/SQL package used to analyze the information
- Agent receives the information which backups should be started

Database Management – Tools (6)

- **Many** other tools/frameworks to ease the daily administration
 - **TVD-BasEnv**
 - **TVD-ConfMan**
 - **Migration Tool** to efficiently migrate databases/applications from other platforms to Linux
 - etc.

Core Messages



- Very interesting project with a lot of opportunities
- Very good feedback from the end users
- For a database cloud you need add-on software frameworks - the infrastructure is not enough

THANK YOU.

BASEL BERN LAUSANNE ZÜRICH DÜSSELDORF FRANKFURT A.M. FREIBURG I.BR. HAMBURG MÜNCHEN STUTTGART WIEN