

WELCOME

Keep an eye on the big
picture –
RAC monitoring tools

Markus Flechtner

September, 27th 2011

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

1

2011 © Trivadis

Keep an eye on the big picture – RAC monitoring tools
September, 27th 2011

trivadis
makes IT easier. ■ ■ ■

About me ..

- Senior Consultant at Trivadis, Düsseldorf
- Working with Oracle since 1995
 - Development (Forms, Reports, PL/SQL)
 - Support
 - Database Administration
- Working at Trivadis since 2008
- Focus
 - Oracle Real Application Clusters
 - Database Migration
- Course Instructor
 - O-AI – Oracle Architecture and Internals
 - O-RAC – Oracle Real Application Clusters



Background: The project

- Customer running a 26-node-RAC on Linux (64-bit)
 - **largest 11.2.0.2-cluster by node number**
 - **3rd largest cluster by node number**
- 3 databases running on the cluster
 - 4 – 7 instances
 - 11 – 15 TB
- 2 additional databases planned for fall 2011
 - 3 – 5 instances
 - 6 – 14 TB
- **Monitoring such a system is a little bit more difficult than monitoring a 2-node-RAC 😊**



2011 © Trivadis

Keep an eye on the big picture – RAC monitoring tools
September, 27th 2011

AGENDA

1. OS Watcher
2. Cluster Health Monitor
3. RAC Check
4. Grid Control
5. Summary



OSWatcher (1) - Overview

- available since 2006
- available for Linux, Solaris, Tru64, HP-UX, AIX
- set of shell scripts to collect OS and network metrics
- runs in the background on all nodes
- uses OS tools like iostat, vmstat, top, netstat
- keeps data for a given retention period

OSWatcher (2) - Installation

1. Download the tar-file from My Oracle Support (Note 301137.1)
2. Distribute the file among the nodes and un-tar it
3. To start OSWatcher: `./startOSW.sh`
 - Snapshot Interval (seconds, default 60)
 - Archive Interval (hours, default 48)
4. To stop OSWatcher: `./stopOSW.sh`

OSWatcher (3) – OSW as a service (on linux)

- MOS-Note 580513.1 contains an rpm-file osw-service.rpm

```
# rpm -Ivh osw-service.rpm
# /sbin/chkconfig osw on
# /sbin/service osw start
```

- Configuration data is stored in the file /etc/sysconfig/osw

```
# Set OSWHOME to the directory where your OSWatcher tools are
installed, data is stored in ${OSWHOME}/archive
OSWHOME=/home/oracle/doag-regio/osw/
# Set OSWINTERVAL to the number of seconds between collections
OSWINTERVAL=60
# Set OSRETENTION to the number of hours logs are to be
retained
OSWRETENTION=1
# Set OSUSER to the owner of the OSWHOME directory
OSWUSER=oracle
```

OSWatcher (4) – Data Storage

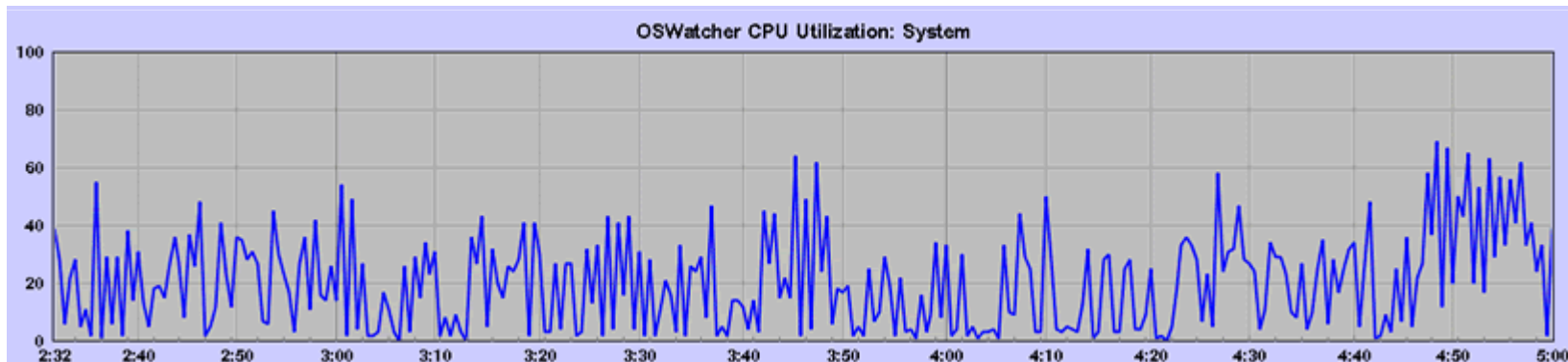
- The collected data is stored in ASCII files

```
oracle@vm104:~/doag-regio/osw/archive/ [grid11202] ls
oswiostat    oswmpstat    oswprvtnet   oswslabinfo  oswvmstat
oswmeminfo   oswnetstat   oswps        oswtop
oracle@vm104:~/doag-regio/osw/archive/ [grid11202] ls oswiostat
vm104.markusflechtner.fma_iostat_11.09.26.2300.dat
vm104.markusflechtner.fma_iostat_11.09.27.0900.dat
vm104.markusflechtner.fma_iostat_11.09.27.1000.dat
oracle@vm104:~/doag-regio/osw/archive/ [grid11202] cat
oswiostat/vm104.markusflechtner.fma_iostat_11.09.27.1000.dat
Linux OSW v3.0.2
zzz ***Tue Sep 27 10:00:17 CEST 2011
`avg-cpu:  %user   %nice %system %iowait  %steal   %idle [...]
           0.67    0.00   1.83    0.33    0.00   97.17
Device:    rrqm/s  wrqm/s   r/s    w/s    rsec/s   [...]
sda        0.00    49.50   0.33  34.88    2.66
sda1       0.00     0.00   0.00   0.00    0.00
[...]
```


OSWatcher (5) – Graphical Output

- The package contains a tool “OSWg” which can be used to generate graphical output from the ASCII files.

```
oracle> export PATH=$ORACLE_HOME/jdk/bin:$PATH
oracle> java -jar oswg.jar -i /home/oracle/doag-
regio/osw/archive
```



OSWatcher (6) - Experiences

- Pros

- Easy to install and run
- Helpful 😊, especially the out-of-box charts
- Oracle Support often requests OSW data (not only) in case of performance problems

- Cons

- Needs a lot of disk space (about 500 GB for all nodes and a storing period of 2 months)

AGENDA

1. OS Watcher
2. Cluster Health Monitor
3. RAC Check
4. Grid Control
5. Summary



Cluster Health Monitor (1) - Overview

- formerly known as “Instantaneous Problem Detection OS Tool (IPD/OS)”
- collects **and analyzes** OS metrics (cluster-wide)
- stores data in a Berkeley database
- installed by default with 11.2.0.2 Grid Infrastructure (Linux, Solaris)
- Download from OTN (version with GUI) for Windows and Linux
 - <http://www.oracle.com/technetwork/database/clustering/downloads/ipd-download-homepage-087212.html>



Cluster Health Monitor (2) - Architecture

- Berkeley Database as a central data storage
- 3 daemons
 - Osysmond = monitoring and collecting OS metrics
 - Ologgerd = receives data, writes data into Berkeley DB
 - Oproxyd = listens for external clients (oclumon, crfgui)

Cluster Health Monitor (3) - Installation

- With 11.2.0.2 CHM is integrated in the Grid Infrastructure (on Linux and Solaris)

```
oracle@vm105:// [grid11202] crsctl status res ora.crf -t -init
ora.crf
      1          ONLINE  ONLINE          vm105
```

- Integration for AIX and Windows is planned for 11.2.0.3
- For the 11.2.0.1 version:

```
Master_node> ./crfinst.pl -i node1,node2 -b <PATH_TO_BDB> -m master_node
on all nodes:
# ./crfinst.pl -f -b <PATH_TO_BDB>
# /etc/init.d/init.crfd enable
```

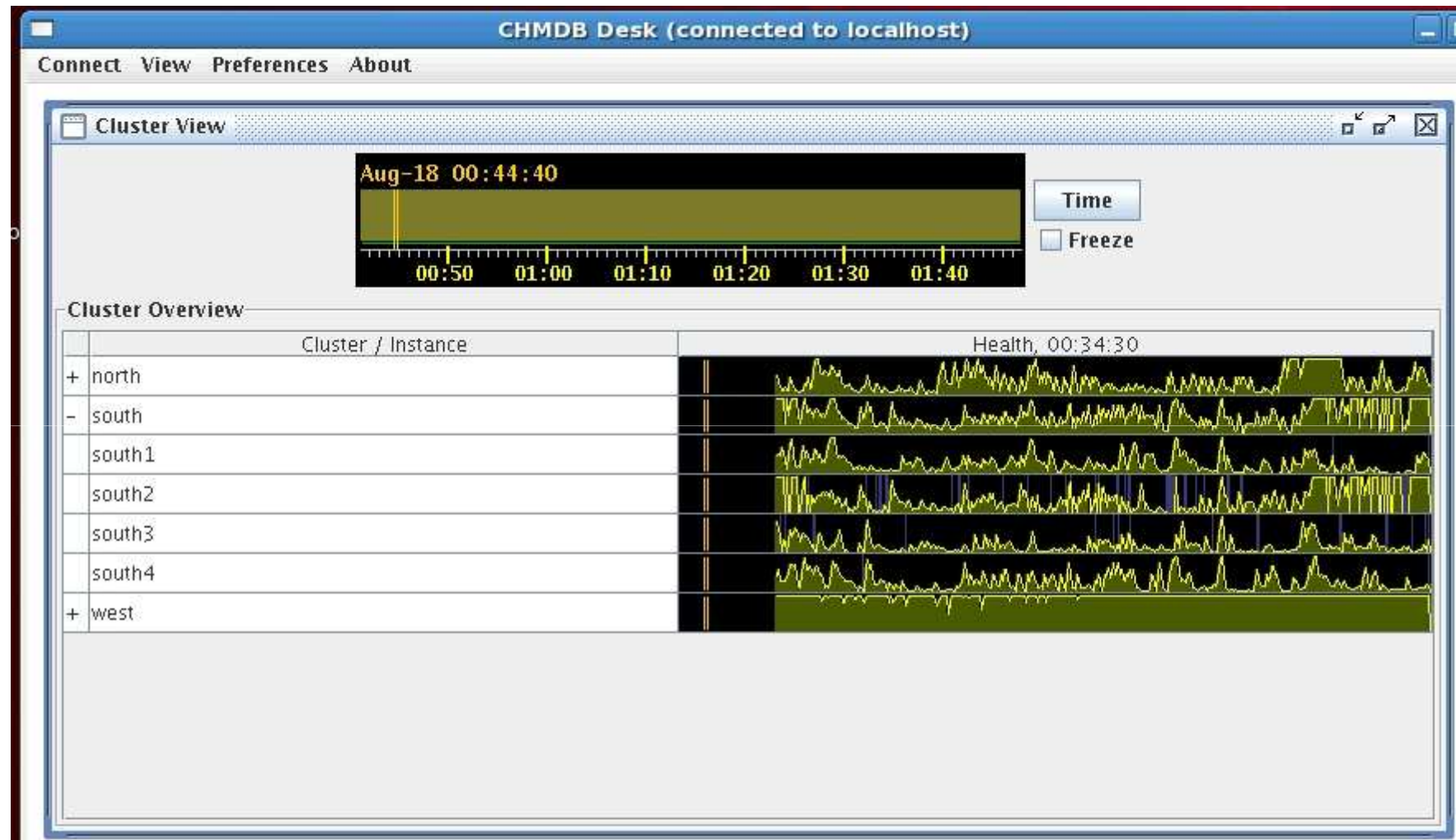
Cluster Health Monitor (4) – CLI (oclumon)

- Commandline-Tool to query the BDB

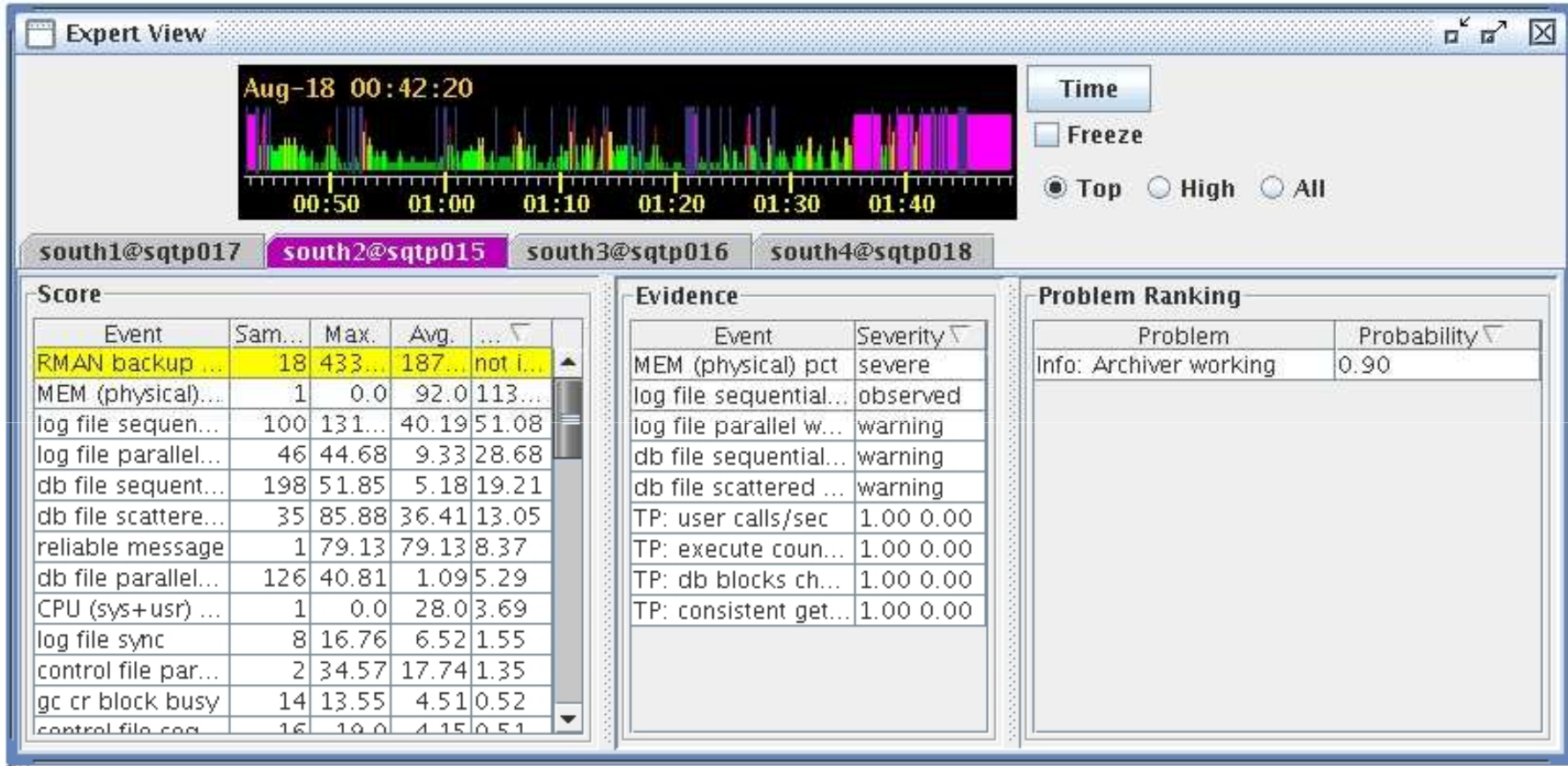
```
oracle@vm105:// [grid11202] oclumon -h
For help from command line : oclumon <verb> -h
For help in interactive mode : <verb> -h
Currently supported verbs are :
showtrail, showobjects, dumpnodeview, manage, version, debug, quit, exit, and
help

oracle@vm105:// [grid11202] oclumon dumpnodeview
dumpnodeview: Node name not given. Querying for the local host
-----
Node: vm105 Clock: '09-27-11 13.56.16' SerialNo:3445
-----
SYSTEM:
#cpus: 2 cpu: 1.10 cpuq: 6 physmemfree: 361884 physmemtotal: 3089940 mcache:
1601244 swapfree: 5144568 swaptotal: 5144568 ior: 38 iow: 79 ios: 17 swpin: 0
swpout: 0 pgin: 38 pgout: 50 netr: 13.93 netw: 25.55 procs: 263 rtprocs: 13
#fds: 5318 #sysfdlimit: 6815744 #disks: 13 #nics: 4 nicErrors: 0
[...]
```

Cluster Health Monitor (5) – GUI (11.2.0.1) – Cluster View



Cluster Health Monitor (6) – GUI – Expert View



Cluster Health Monitor (7) – GUI - „RCA“ & Corrective Actions

The screenshot displays the Cluster Health Monitor GUI with three tabs: south1@sctp017, south2@sctp015, and south3@sctp016. The main window is titled "Dependency" and contains a diagram with a grey box labeled "Root cause" at the top. An arrow points down from this box to a cyan box containing the text "Info: Archiver working" and "P=0.90".

Below the dependency window, a "Corrective Action" panel is visible, also with the same three tabs. It contains the following text:

Corrective Action

The monitor has detected higher than normal latencies.

Corrective actions for "Disk I/O slow" on instance south1 on node sctp017

User may experience performance degradation due to slow disk access (Probability 33%).

More diagnostics: Check disk performance.



Cluster Health Monitor (8) - Experiences

- Will be the standard tool in the future
- Pros
 - It's free
 - Can run in „realtime-mode“ and send alerts
 - Suggests „corrective actions“
 - Integrated in the Grid Infrastructure (11.2.0.2)
- Cons
 - Usability of the GUI (11.2.0.1)
 - So far no GUI for the 11.2.0.2-version of CHM

AGENDA

1. OS Watcher
2. Cluster Health Monitor
3. RAC Check
4. Grid Control
5. Summary



RAC Check (1) - Overview

- new tool (August 2011)
- Available for
 - Linux
 - Solaris
 - AIX
- Supported Oracle Versions: 10gR2 – 11gR2
- Checks
 - RAC best practices
 - Database parameters
 - Infrastructure (Clusterware, ASM) settings
 - OS recommendations (kernel parameters, packages)

RAC Check (2) – Installation and Execution

1. Download RACCheck.zip from My Oracle Support (MOS-Note 1268927.1)
 2. Upload RACCheck.zip to one of the RAC nodes and unzip it
 3. Chmod 755 raccheck
 4. `./raccheck -a`
- root password or sudo-configuration required (or disable root-actions)
 - Takes about 5 minutes per node



RAC Check (3) – Output (excerpt)

[..]

Data collections completed. Checking best practices on vm104.

```
-----  
WARNING => OCR is NOT being backed up daily  
INFO =>   $CRS_HOME/log/hostname/client directory has too many older log files.  
INFO =>   user_dump_dest has trace files older than 30 days for RACDB  
INFO =>   background_dump_dest has files older than 30 days for RACDB  
INFO =>   At some times checkpoints are not being completed for RACDB  
INFO =>   audit_file_dest has audit files older than 30 days for RACDB  
WARNING => kernel.shmmax parameter is NOT configured according to recommendation  
WARNING => One or more ASM disks are found with partnership problem.  
INFO =>   Number of SCAN listeners is NOT equal to the recommended number of 3.  
WARNING => Local listener init parameter is not set to local node VIP. for RACDB  
WARNING => NIC bonding is NOT configured for public network (VIP)  
WARNING => ASM shared_pool_size is < recommended value  
WARNING => OSWatcher is not running as is recommended.  
INFO =>   Jumbo frames (MTU 9000) are not configured for interconnect  
WARNING => NTP is not running with correct setting
```

[..]

RAC Check (4) - Experiences

- Pros
 - Small tool
 - Easy setup and execution
 - Results can be stored in a database
 - Can be configured to run in „silent mode“, without user interaction
- Cons
 - „Best practices“ may not be suitable for a large cluster

AGENDA

1. OS Watcher
2. Cluster Health Monitor
3. RAC Check
4. **Grid Control**
5. Summary



Grid Control (1) - Overview

- Grid Control is running on a 2-node-RAC
- „GC-High Availability“ is established using an active/passive Cluster
- Software is installed on an ACFS-volume
- OMS uses an additional virtual IP and virtual hostname

Grid Control (2) - Output

Databases

View Oracle Load Map Search List

Page Refreshed 19-Aug-2011 01:56:52 PDT

Total Active Load: 97.4 active sessions

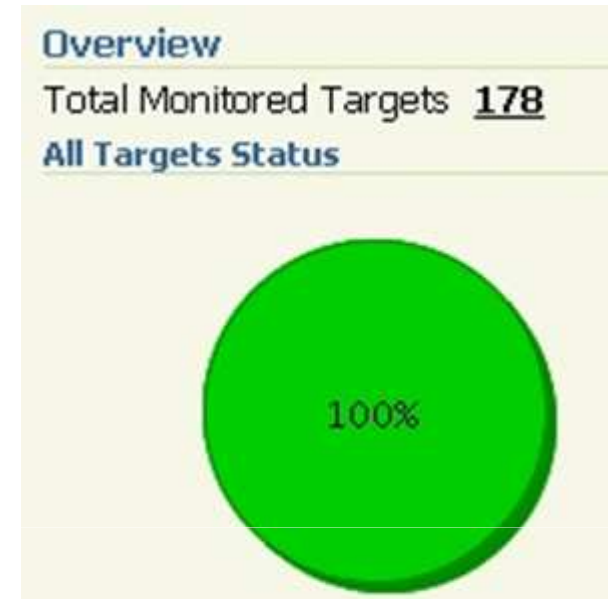
west		north		south		Down/Unavailable No Targets Found
13 (active sessions) west_west1	10 west_west8	7 north_north5 44.6% "User I/O" wait Class 13.4% Top SQL Statements 11.1% Top Segments by "User I/O" and "Cluster"	5 north_north4 57.8% Top SQL Statements 57.7% "User I/O" wait Class 7.1% Top Segments by "User I/O" and "Cluster"	8 south_south2 44.7% "User I/O" wait Class 23.9% Top SQL Statements 12.0% Top Segments by "User I/O" and "Cluster"	8 south_south1 50.6% "User I/O" wait Class 9.6% Top SQL Statements 3.0% Hard Parse	
12 west_west2	9 west_west5	5 north_north2 39.5% "User I/O" wait Class 8.7% Top SQL	3 north_north6 43.1% "User I/O" wait Class 18.4% Top SQL Statements	7 south_south4 56.8% "User I/O" wait Class 18.6% Top SQL Statements		
		3 north_north1 35.1% "User I/O" wait Class 6.4% Buffer Busy - Hot Objects 3.9% Top SQL Statements	1 north_north1 14.6% "User I/O" wait Class	preprd		
			1 north_north1 14.6% "User I/O" wait Class	2 preprd_preprd1 20.0% "User I/O" wait Class	2 preprd_preprd3 20.0% "User I/O" wait Class	0

View Level: Database Instance



Grid Control (3) - Experiences

- Pros
 - Centralized monitoring and administration for all components (hosts, clusterware, ASM, databases)
- Cons
 - The default metrics are not 100% RAC-aware
 - The default metrics are not 100% 11.2-aware
 - Clusterwide events are reported multiple times (e.g. a full ASM diskgroup is reported for all nodes)



AGENDA

1. OS Watcher
2. Cluster Health Monitor
3. RAC Check
4. Grid Control
5. Summary

Summary (1)

- Basically, **large clusters are as easy to manage as small ones!**
- Of course, tools are helpful 😊
- Grid Control is an almost perfect tool, but needs some additional configuration work
- OS Watcher is a very powerful tool to collect metrics, but housekeeping could be easier
- Cluster Health Monitor looks very promising, stronger connection between Grid Control and CHM is expected for the future
- RACCheck is good for checking basic best practices



Summary (2)

- Beside the tools mentioned there are
 - Remote Diagnostic Assistant (RDA)
 - Oracle Configuration Manager (OCM)
 - Diagcollection.pl
 - LTOM (The Lite Onboard Monitor)
 - Cluvfy
 - ..
- Lots of tools ..
- Which one to use when?
- Support always requests the output of the other tool ☹
- Perhaps it's time to consolidate?

ANNEX



References (1)

■ OSWatcher

- MOS-Note 301137.1 - OS Watcher User Guide (incl. link for OSW download)
- MOS-Note 580513.1 – How to Start OSWatcher every system boot

■ Cluster Health Monitor

- <http://www.oracle.com/technetwork/database/clustering/downloads/ipd-download-homepage-087212.html>
- MOS-Note 736752.1 - Introducing Cluster Health Monitor (IPD/OS)
- MOS-Note 1328466.1 – CHM FAQ

References (2)

- **RACCheck**

- MOS-Note 1268927.1 – RAC Check – RAC Configuration Audit Tool
- <http://www.oracle.com/webfolder/technetwork/de/community/dbadmin/tipps/raccheck/index.html>

- **Grid Control in an Active/Passive configuration**

- Grid Control 11.1 Administration Guide, Chapter 18
Section „How to Configure Grid Control OMS in Active/Passive Environment for High Availability Failover Using Virtual Host Names“

THANK YOU

