

Oracle RAC und NFS

Johannes Ahrends



- Oracle Spezialist seit 1992
 - 1992: Presales bei Oracle in Düsseldorf
 - 1999: Projektleiter bei Herrmann & Lenz Services GmbH
 - 2005: Technischer Direktor ADM Presales bei Quest Software GmbH
 - 2011: Geschäftsführer CarajanDB GmbH
- 2011 → Ernennung zum Oracle ACE
- Autor der Bücher:
 - Oracle9i für den DBA, Oracle10g für den DBA, Oracle 11g Release 2 für den DBA
- DOAG Themenverantwortlicher Datenbankadministration, Standard Edition
- Hobbies:
 - Drachen steigen lassen (Kiting) draußen wie drinnen (Indoorkiting)
 - Motorradfahren (nur draußen)

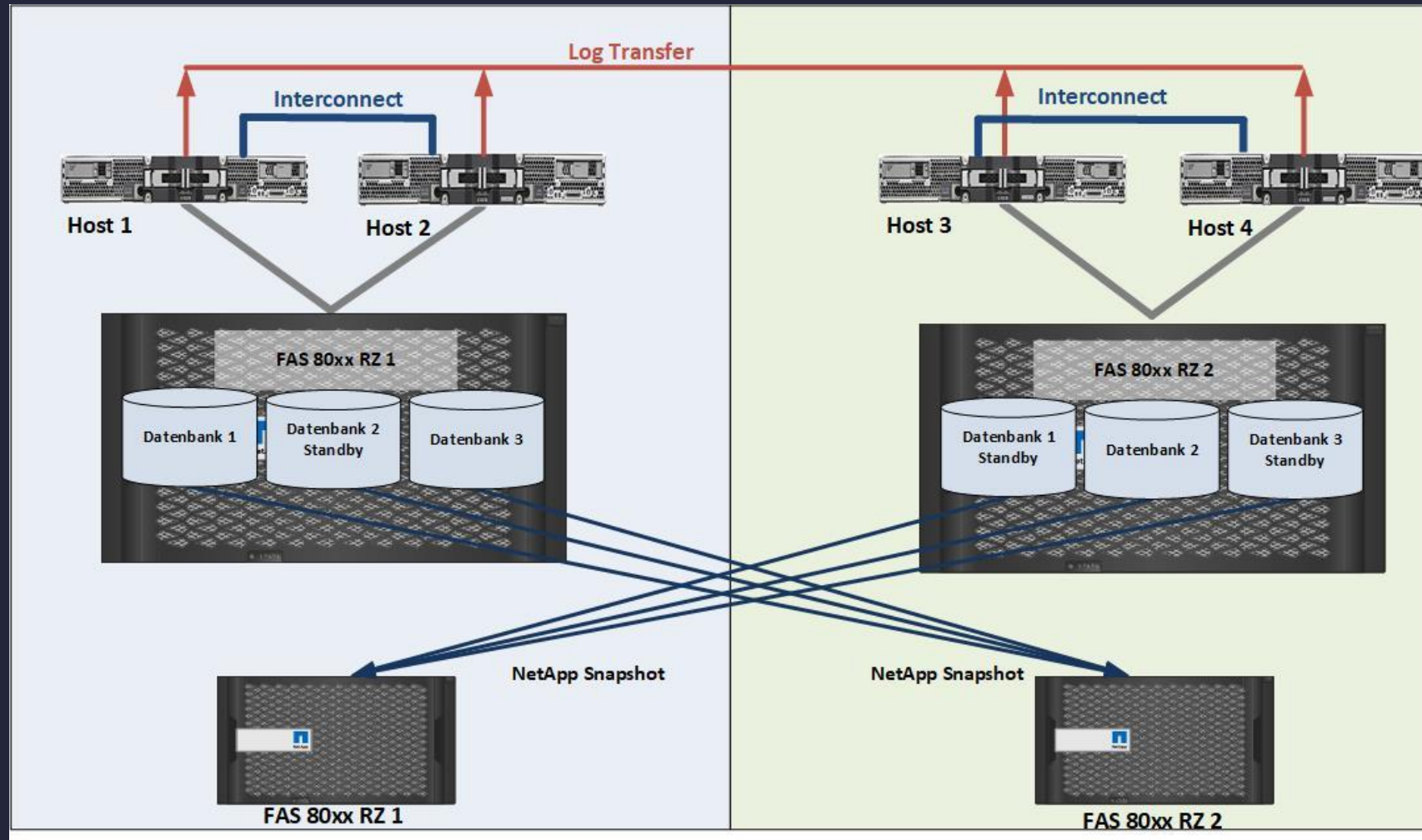


Das Projekt

- Oracle 12c (12.1.0.2)
- 2 Knoten RAC
- NetApp Storage mit Direct NFS
- Data Guard
- 2 Datenbanken
- Projektstart Oktober 2015

- Migration von 8 Oracle 11g Datenbank 6 Knoten Stretched RAC nach
 - Oracle 12c (12.1.0.2)
 - 2 Knoten RAC
 - Data Guard
 - Multitenant
 - 4 Stages (Maintenance, Test, Vorproduktion, Produktion)
 - FlexPod Architektur mit NetApp Storage
- Start des Projektes: August 2015

Schematische Darstellung



Gründe für NFS

- Unterstützung durch NetApp
 - Umfangreiche Dokumentation
- Backup durch Snapshots
- Snapshot Cloning
- Automatische Storageerweiterung und Reduzierung
- Umfangreiches Knowhow vorhanden

Direct NFS

Was ist Direct NFS?

- Einführung mit Datenbankversion 11g
 - Oracle Direct NFS kurz DNFS
 - Integrierter NFS-Client, unabhängig vom OS-basierten Client
 - Oracle-Instanz erhält direkten Zugriff auf NFS-Storage
- Oracle-eigene Optimierungen
 - Multipathing über bis zu vier parallele Pfade (unabhängig von Switch, NIC-bonding)
 - Instanz konfiguriert die NFS-Kommunikation automatisch – leichtere RAC Konfiguration
 - NFS-Storage auch unter Windows nutzbar
- Unterstützung
 - 11g NFS Server mit NFSv3
 - 12c NFS Server mit NFSv3 bis 4.1

- Aktivierung Direct NFS Client - Linux
 - Anmelden als Oracle Software installation owner

```
> cd $ORACLE_HOME/rdbms/lib  
> make -f ins_rdbms.mk dnfs_on
```

- Alternativ manuell bzw. vor Version 11.2

```
> cd $ORACLE_HOME/lib  
> mv libodm11.so libodm11.so.stub  
> ln -s libnfsodm11.so libodm11.so
```

- Bei Oracle RAC nicht notwendig

- Aktivierung Direct NFS Client – Windows
 - Standard ODM Library, oraodm11.dll durch die ODM NFS Library, oranfsodm11.dll im bin Verzeichnis ersetzt werden.

```
> copy oraodm11.dll oraodm11.dll.stub  
> copy /Y oranfsodm11.dll oraodm11.dll
```

- Datenbank neu starten – Linux + Windows
 - Beim Starten der Datenbank sollte folgende Zeile im Alert Log zu finden sein:

```
Oracle instance running with ODM: Oracle Direct NFS ODM Library Version 3.0
```

- Konfiguration fstab bzw. mtab - Linux
 - vi /etc/fstab – default Optionen sollten genügen
 - bei Problemen: NFS Buffer Size Parameter „rsize“ und „wsize“ setzen

```
nas1:/volume1/oranfs      /mnt/oranfs
nfs      rw,bg,hard,rsize=32768,wsize=32768,nfsvers=3,nointr,timeo=600,actimeo=0,tcp,addr=x.x.x.x      0 0
```

- Konfiguration (alternativ / zusätzlich) oranfstab - Linux
 - Zusätzlich kann die oranfstab erstellt werden um Oracle-eigene Optionen zu nutzen
 - Direct NFS Client bestimmt die Einstellung des Mount Points nacheinander
 1. \$ORACLE_HOME/dbs/oranfstab
 2. /etc/oranfstab
 3. /etc/mtab

- orafstab - Linux
 - /etc/orafstab

```
server: nas1                # Alias frei wählbar
path: 172.16.0.2            # IP NFS-Server
local: 172.16.30.40        # Optional IP Datenbank-Server
path:                      # Optional bis zu 4 path/local Paare für Multipathing
local:                     # Optional weiterer Pfad
export: /volume1/orafns/oradata mount: /mnt/orafns/oradata
                             # export: NFS-Server Pfad, mount: Lokaler Pfad
export: mount:              # Optional weitere export/mount Pfade - unbegrenzt
mnt_timeout: 30            # Optional Timeout bei fehlgeschlagenem Mounten
dontroute                  # Optional Messages should not be routed by OS
nfs_version: nfsv4.1       # Optional ab 12c + NFSv4 (default: NFSv3)
management                 # Optional ab 12c use management interface for SNMP
community                  # Optional ab 12c community string for SNMP

server: nas2                # weitere Server durch Leerzeile getrennt
path: 172.16.0.3
export: /volume/path/nfs mount: /local/path
```

- Oranfstab - Voraussetzung unter Windows
 - oranfstab mit Editor unter ORACLE_HOME\dbs\oranfstab anlegen

```
server: nas1                # Alias frei wählbar
path: 172.16.0.2            # IP NFS-Server
local: 172.16.30.40        # Optional IP Datenbank-Server
path:                       # Optional bis zu 4 path/local Paare für Multipathing
local:                     # Optional Multipathing für Lastverteilung und Failover
export: /volume1/oranfs/oradata mount: D:\oranfs\oradata
                           # export: NFS-Server Pfad, mount: Localer Pfad
export: mount:              # Optional weitere export/mount Pfade - unbegrenzt
mnt_timeout: 30            # Optional Timeout bei fehlgeschlagenem Mounten
uid: 65534                 # Optional User ID für NFS Server Zugriff
gid: 65534                 # Optional Group ID für NFS Server Zugriff
nfs_version: nfsv4.1       # Optional ab 12c + NFSv4 (default: NFSv3)

server: nas2                # weitere Server durch Leerzeile getrennt
path: 172.16.0.3
export: /volume/path/nfs mount: D:\local\path
```

- NFS-Server mit mindestens NFS-Version 3
 - Speziell bei Linux-basierten Servern kann es ein Problem mit der Port-Reservierung geben, da die Oracle-Instanz nicht als root-User läuft. Um eine Abweisung durch den NFS-Server zu verhindern muss ggf. auf diesem die Export-Option **insecure** in der `/etc/exports` gesetzt werden. Hiermit werden Verbindungen über nicht-privilegierte Ports über 1024 zugelassen.

```
nas1> cat /etc/exports
```

```
/volume1/oranfs
```

```
*.carajandb.intra(rw,async,no_wdelay,insecure,root_squash,insecure_locks,sec=sys,anonuid=1024,anongid=100)
```

- Berechtigungen für den Export-Ordner setzen - Linux

```
nas1> cd /volume1/oranfs  
  
> chown oracle:dba oradata  
> chmod 755 oradata
```

- Berechtigungen für den Export-Ordner setzen – Windows
 - Der Direct NFS Client nutzt standardmäßig uid:65534 und gid:65534, was nobody und nogroup auf dem NFS-Server entsprechen sollte. Der Zugriff darüber sollte gewährleistet sein kann aber im Zweifel explizit in der oranstab gesetzt werden. Für die richtigen Permissions legen wir auf dem NFS-Server einen **User oracle** und die **Group ORA_DBA** an und setzen die Berechtigungen entsprechend:

```
nas1> cd /volume1/oranfs  
  
> chown oracle:ORA_DBA oradata  
> chmod 755 oradata
```

- Abfrage ob der NFS Server verbunden ist

```
select * from v$dnfs_servers;
```

ID	SVRNAME	DIRNAME	MNTPORT	NFSPORT	WTMAX	RTMAX
---	-----	-----	-----	-----	-----	-----
1	nas1	/volume1/oranfs/oradata	892	2049	131072	131072

- Ist die Konfiguration noch nicht richtig, liefert die Abfrage keine Zeilen zurück. Zudem muss der Pfad beim ersten mal angesprochen werden bevor die View ein Ergebnis liefert.
- Zum Test ein File auf dem Mount anlegen und obige Abfrage nochmal absetzen:

```
SQL> create tablespace test_tbs datafile '/mnt/oranfs/oradata/test_tbs.dbf' size 10M;
```

```
SQL> create tablespace test_tbs datafile 'D:\oranfs\oradata\test_tbs.dbf' size 10M;
```

- Abfrage Zugriff über Direct NFS oder NFS-Client OS

```
select * from v$dnfs_files;
```

FILENAME	FILESIZE	PNUM	SVR_ID
-----	-----	----	-----
d:/oranfs/oradata/test_tbs.dbf	10493952	10	1

- Liefert die Abfrage keine Zeilen zurück wird unter Linux weiter der NFS-Client des OS genutzt.
- Weitere Views:

```
select * from v$dnfs_channels;  
select * from v$dnfs_stats;
```

- „Best Practices to configure a dNFS client“
- „Best Practices for Oracle Databases on NetApp Storage“
- „Oracle Databases on ONTAP“

dNFS Mountoptionen

- \$ORACLE_HOME/dbs/oranfstab
- /etc/oranfstab
- /etc/mtab

- OCR Devices → Kein Direct NFS
- ORACLE_HOME
- ADR: NFS Optionen noac und/oder actimeo=0 löschen
 - Hostcaching erlaubt → besserer Durchsatz für Traces, Logs, etc.
 - „Do not place ADR data on a file system with noac or actimeo=0 because performance problems are likely. Separate ADR data into a different mount point if necessary.”

- Auszug aus „Oracle Databases on ONTAP“

Oracle Objekte	NFS Optionen
Controlfiles, Datafiles, Redolog-Files	<code>rw,bg,hard,vers=3,proto=tcp,timeo=600,rsiz=65536,wsiz=65536, nointr,actimeo=0</code>
CRS	<code>rw,bg,hard,vers=3,proto=tcp,timeo=600,rsiz=65536,wsiz=65536, nointr,noac,actimeo=0</code>
ORACLE_HOME (Dedicated)	<code>rw,bg,hard,vers=3,proto=tcp,timeo=600,rsiz=65536,wsiz=65536</code>
ADR_HOME	<code>rw,bg,hard,vers=3,proto=tcp,timeo=600,rsiz=65536,wsiz=65536, actimeo=0</code>

- Grid: \$GRID_BASE/diag
- Oracle DB: \$ORACLE_BASE/diag
- ADR von Grid kann nicht geändert werden
 - Eigener Mountpoint /u01/app/gridbase
 - Vorsicht: Mountpoint /u01/app existiert für Oracle und Grid Software
- ADR der Datenbank kann geändert werden
 - Eigener Mount Pfad (/u01/oradb):

```
sqlplus / as sysdba  
SQL> ALTER SYSTEM SET diagnostic_dest='/u01/oradb';
```



Upgrade 12.2

Upgrade SLES 12 + Oracle 12.2

- Upgrade SuSE SLES 11 nach SLES 12
- Bedingt Neuinstallation des Betriebssystems
- Upgrade Oracle 12.2
 - OCR Devices nur noch auf ASM unterstützt!
 - ASM on NFS („Guidelines for Configuring Oracle ASM Disk Groups on NFS“)

- Rolling Upgrade
 - Delete Node 1 (runInstaller oder Deinstall)
 - Grid Infrastructure und Oracle DB Software
 - Betriebssysteminstallation Knoten 1
 - Add Node 1 (addnode.sh)
 - Grid Infrastructure und Oracle DB Software
 - dito Knoten 2
 - Upgrade GI Node 1
 - Upgrade GI Node 2
- Vorteil: RAC immer verfügbar
- Nachteil: sehr aufwändig, mehrfaches Schwenken der Services

- Data Guard First
 - Betriebssysteminstallation Knoten 1 und Knoten 2
 - Mounten aller Filesysteme
 - Löschen der GI Software (/u01/app/grid) auf beiden Knoten
 - Installation Grid Infrastructure 12.2
 - addnode.sh Oracle DB Software
 - Starten der Datenbanken
- Nachteil: Cluster nicht verfügbar → Kein DR während des Upgrades
- Vorteil: Schnell, keine „Altlasten“
 - Aktuell: Downtime ca. 2 Stunden

- NFS Devices

- /u02/grid/ocrA (5 GB)
- /u02/grid/ocrB (5 GB)
- /u02/grid/ocrC (5 GB)
- /u02/grid/GIMR (70 GB)

- GI Installationsproblem → Filesysteme müssen leer sein

- NetApp erstellt .snapshot Verzeichnis unter jedem Mountpoint

```
mkdir /u02/grid/ocrA/OCR  
mkdir /u02/grid/GIMR/MGMTDB
```

- 12.2 GI Installation Responsefile

```
oracle.install.asm.storageOption=ASM_ON_NAS  
oracle.install.asmOnNAS.ocrLocation=/u02/grid/ocrA/OCR  
oracle.install.asmOnNAS.configureGIMRDataDG=true  
oracle.install.asmOnNAS.gimrLocation=/u02/grid/GIMR/MGMTDB
```

- OCR „Devices“ anlegen

```
dd if=/dev/zero of=/u02/grid/ocrA/ocrA bs=1024k count=2000
dd if=/dev/zero of=/u02/grid/ocrB/ocrB bs=1024k count=2000
dd if=/dev/zero of=/u02/grid/ocrC/ocrC bs=1024k count=2000
chmod 660 /u02/grid/ocrA/ocrA /u02/grid/ocrB/ocrB /u02/grid/ocrC/ocrC
chgrp asmadmin /u02/grid/ocrA/ocrA /u02/grid/ocrB/ocrB /u02/grid/ocrC/ocrC
```

- ASM Diskgruppe erstellen

```
sqlplus / as sysasm
CREATE DISKGROUP OCR NORMAL REDUNDANCY
    FAILGROUP OCRA DISK '/u02/grid/ocrA/ocrA'
    FAILGROUP OCRB DISK '/u02/grid/ocrB/ocrB'
    QUORUM FAILGROUP ocrC DISK '/u02/grid/ocrC/ocrC'
ATTRIBUTE 'compatible.asm'='12.1', 'compatible.rdbms'=',12.1';
```

- Fehlermeldung beim Hinzufügen eines neuen Knotens

```
cluvfy stage -pre nodeadd -n server3
PRCR-1154 : Failed to create file output stream with file name:
/u01/app/gridbase/crsdata/server1/cvu/cvutrace.log/u01/app/gridbase/crsdata/server1/cvu/cvutrace.log.0 (Permission
denied)

ERRORMSG(vrzlora103): PRVG-0809 : Mount options for file system "x.x.x.x:/ora_ocrC" mounted on path
"/u02/grid/ocrC/ocrC" at node "server3" did not meet the requirements for this platform
[Expected =
"rw,!ro,!hard|soft,rsiz>=32768,wsiz>=32768,tcp|proto=tcp,v3|nfsv3|vers=3|nfsvers=3|v4|nfsv4|vers=4|nfsvers=4|vers=4
.0|vers=4.1|nfsvers=4.1,timeo>=600,acregmin=0&acregmax=0&acdirmin=0&acdirmax=0|actimeo=0" ;
Found =
"rw,relatime,vers=3,rsiz=65536,wsiz=65536,namlen=255,acregmin=0,acregmax=0,acdirmin=0,acdirmax=0,hard,proto=tcp,tim
eo=600,retrans=2,sec=sys,mountaddr=x.x.x.x,mountvers=3,mountport=635,mountproto=tcp,local_lock=none,addr=x.x.x.x"]
```

- MOS: Creating File Devices On NAS/NFS FileSystems For ASM Diskgroups. (Doc ID 1620238.1)
 - “NOTE: If Quorum is hard mounted then when NFS is not available then we may hit bug 19373893, so For quorum disks soft mount is recommended.”

- NFS Optionen 12.2

Oracle Objekte	NFS Optionen
/u01/app	rw,bg,hard,vers=3,proto=tcp,timeo=600,rsiz=65536,wsiz=65536
/u01/app/gridbase /u01/oradb	rw,bg,hard,vers=3,proto=tcp,timeo=600,rsiz=65536,wsiz=65536, actimeo=0
/u02/grid/ocrA /u02/grid/ocrB /u02/grid/GIMR	rw,bg,hard,vers=3,proto=tcp,timeo=600,rsiz=65536,wsiz=65536, nointr,noac,actimeo=0
/u02/grid/ocrC	rw,bg,soft,vers=3,proto=tcp,timeo=600,rsiz=65536,wsiz=65536, nointr,noac,actimeo=0
/u03/oradata/<DBNAME>/redoA /u03/oradata/<DBNAME>/redoB /u03/oradata/<DBNAME>/CDB /u03/oradata/<DBNAME>/pdb /u03/oradata/<DBNAME>/temp	rw,bg,hard,vers=3,proto=tcp,timeo=600,rsiz=65536,wsiz=65536, nointr,actimeo=0

SLES 12 Kernel Bug

- Wenn Anzahl NFS Mounts > 50 → Mounts fehlen teilweise
- Bug im Kernel behoben mit Kernel 4.4.140 (!)

- Transparent
- Einfache Installation
- Vollständig Supportet
- Nutzung von Storage Snapshots
- Einfaches Clonen von Datenbanken (Copy On Write)
- Bei Data Guard können Dateien einfach kopiert werden

Aus aktuellem Anlass

- OJVM in der Datenbank
- Oracle Security Alert for CVE-2018-3110

CVE#	Component	Package and/or Privilege Required	Protocol	Remote Exploit without Auth.?	CVSS VERSION 3.0 RISK (see Risk Matrix Definitions)									Supported Versions Affected	Notes
					Base Score	Attack Vector	Attack Complex	Privs Req'd	User Interact	Scope	Confidentiality	Integrity	Availability		
CVE-2018-3110	Java VM	Create Session	Oracle Net	No	9.9	Network	Low	Low	None	Changed	High	High	High	11.2.0.4, 12.1.0.2, 12.2.0.1, 18	

MOS Note: 2217053.1

RAC Rolling Install Process for the "Oracle JavaVM Component Database PSU" (OJVM PSU) Patches

*"Beginning with the Jan2017 OJVM PSU patchset for 11.2.0.4 and for 12.1.0.2, this document defines a few specific situations where the OJVM PSU patchset can be postinstalled into each database **while the database remains in unrestricted "startup" mode**. This will allow a "Conditional Rolling Install" ability for the OJVM PSU patchsets for 11.2.0.4 and for 12.1.0.2"*

- Oracle 12.2. mit Release Update 180717 (Juli 2018)
- Bug 26430323 : ORA-07445[KKQGBPCHECKVALIDITYDP] WHEN SELECTING FROM DUAL
- “As you are running your 12.2.0.1 version on Windows platform, one off patches are not available for Windows. On Windows bug fixes are released in quarterly patch bundles. The bug will be fixed in Windows although cannot give you a timeframe when this will happen.”

- 20.11. 14 Uhr Die Zukunft des DBAs: über oder unter den Wolken?
- 20.11. 18 Uhr Open DB Mic Session
- 21.11. 14 Uhr Das Battle – Oracle vs. PostgreSQL
- 21.11. 16 Uhr Multitenant in der Standard Edition
- 22.11. 09 Uhr Oracle 18c XE
- 23.11. DOAG Schulungstag „Oracle XE, die kostenlose Oracle Datenbank“
- 21.11 10 Uhr Oracle Monitoring – Reicht der OEM (Fabian Greis)
- 21.11 15 Uhr Nur 3 Oracle Homes patchen bei 200 produktiven Datenbanken (Andre Lünsmann, Barmenia Versicherung)

- Experten mit über 25 Jahren Oracle Erfahrung
- Firmensitz in Erftstadt bei Köln
- Spezialisten für
 - Oracle Datenbank Administration
 - Hochverfügbarkeit (RAC, Data Guard, Failsafe, etc.)
 - Einsatz der Oracle Standard Edition
 - Oracle Migrationen (HW, Unicode, Standard Edition)
 - Replikation
 - Performance Tuning
 - Datenbank Cloning (Delphix, Actifio, CloneDB)
- Fernwartung
- Schulung und Workshops (Oracle, Toad)



- E-Mail: johannes.ahrends@carajandb.com
- Homepage: www.carajandb.com
- Adresse:
 - CarajanDB GmbH
Siemensstraße 25
50374 Erftstadt
- Telefon:
 - +49 (22 35) 1 70 91 84
 - +49 (1 70) 4 05 69 36
- Twitter: carajandb
- Facebook: johannes.ahrends
- Blogs:
 - blog.carajandb.com
 - www.toadworld.com