OPITZ CONSULTING

Menschen. Innovationen. Lösungen.

Oracle 11gR2 Erfahrungen

Hochverfügbarkeit

Rainier Kaczmarczyk
Solution Architect

OPITZ CONSULTING München GmbH

ORACLE Platinum Partner
Specialized
Oracle Database

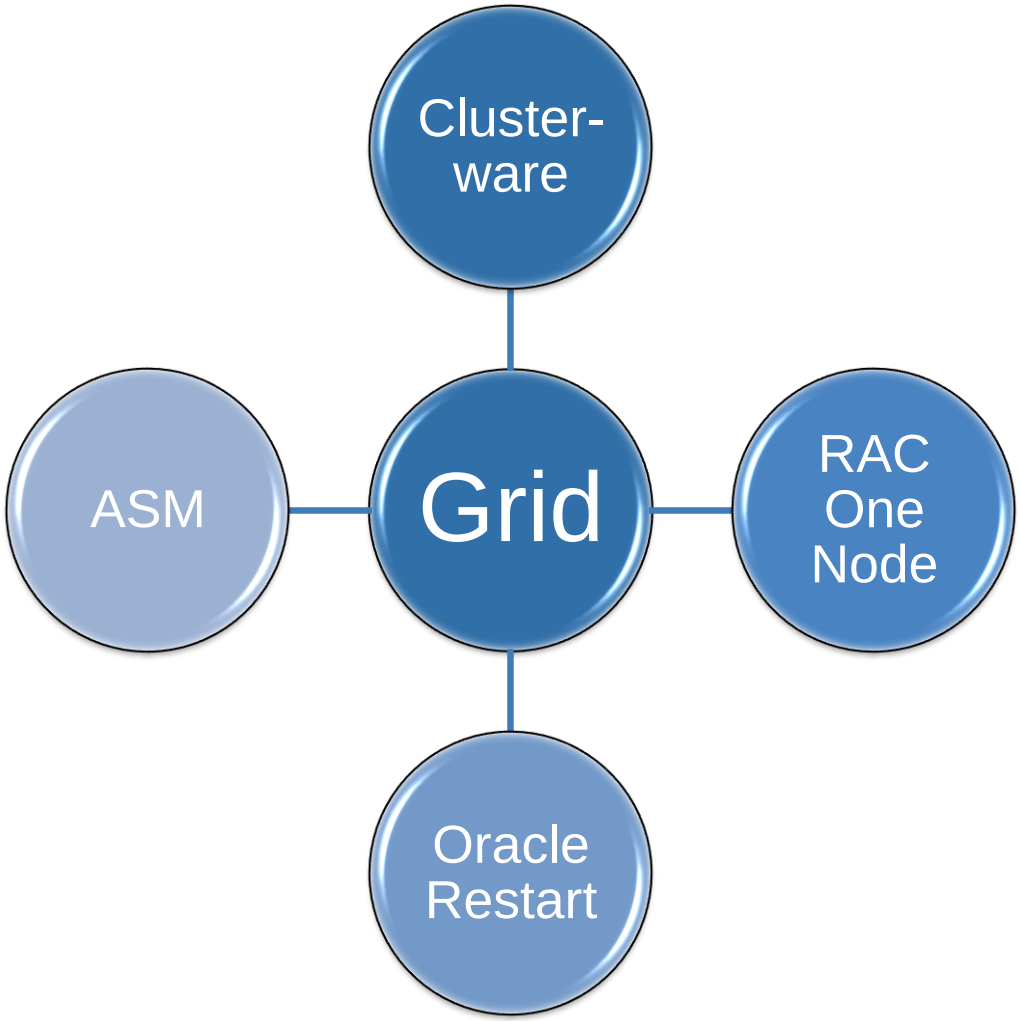
DOAG Regionaltreffen München, 16.05.2011

Agenda

1. Grid Infrastucture
2. Oracle Restart
3. One Node RAC
4. Real Application Cluster
5. ASM Cluster Filesystem
6. Data Guard / Active Data Guard

1 Grid Infrastucture

Grid Infrastructure



Oracle Restart

Oracle Restart: wie war es früher beim Startup?

- **Kein automatischer Start von Oracle Diensten durch Oracle**
  - Dies musste durch das Betriebssystem gemacht werden
  - Eigene Runlevel-Scripts erstellen (/etc/rc3.d/S99oracle)

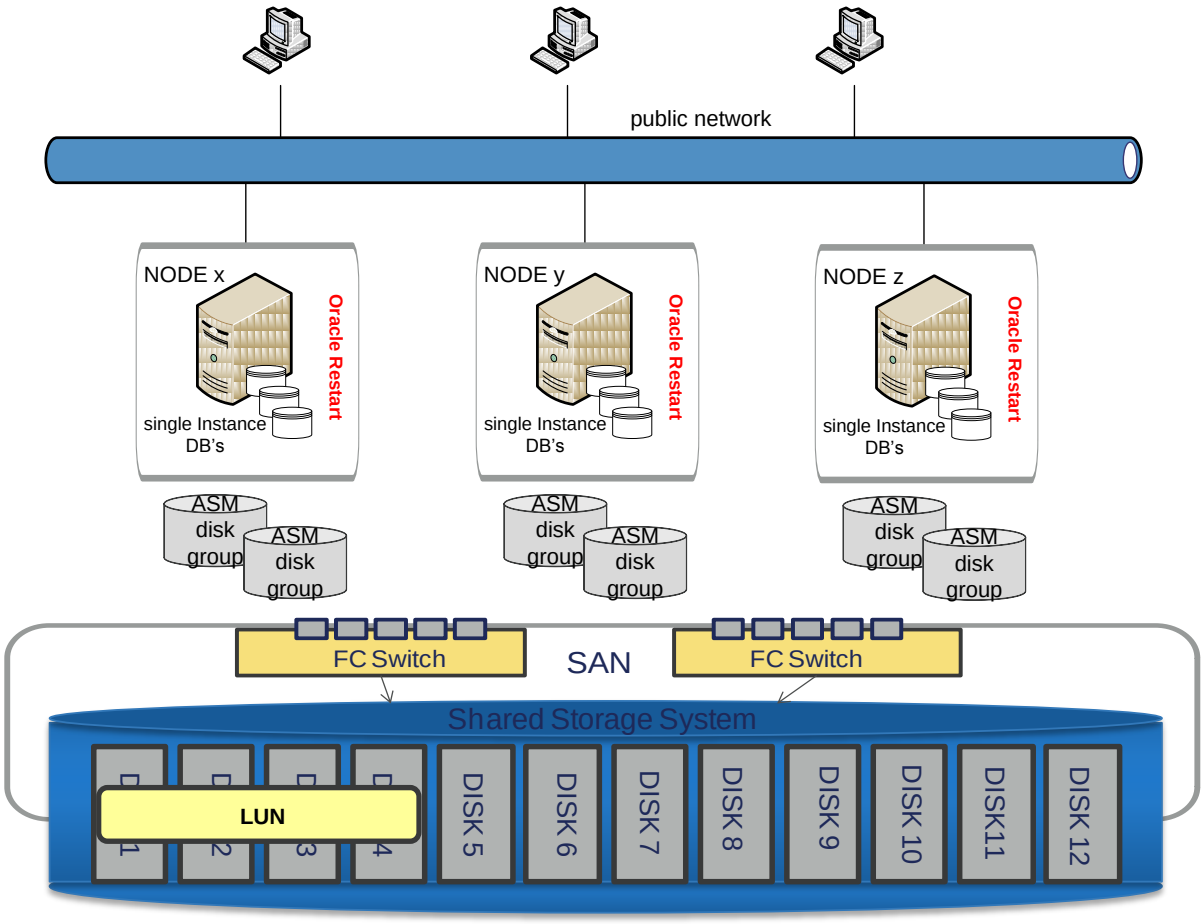
```
case "$1" in
  start)
     ORAENV_ASK=NO
     export ORACLE_SID=DB11201
     . oraenv
     lsnrctl start
     echo "startup" | sqlplus / as sysdba
  ;;
```

- **Keine automatische Überwachung / Restart der Prozesse**
  - Bei der Datenbanksoftware war keine solche Möglichkeit dabei
  - Nur zusätzliche Produkte (z.B. Grid Control) haben die Prozesse überwacht

Was ist Oracle Restart?

- **Framework zum Starten / Stoppen und Überwachen der Oracle-Komponenten**
  - Datenbankinstanz
  - Listener
  - Datenbank-Services
  - ASM
- **Basiert wie bei RAC auf der Clusterware (Grid Infrastructure)**
  - Ein grosser Teil von Clusterware besteht aus der von Digital/HP Tru64 übernommenen Cluster-Software
  - Installation der „Grid“ Software auf einem Standalone-Node
  - Für ASM ist es eine zwingende Voraussetzung

Oracle Restart Übersicht



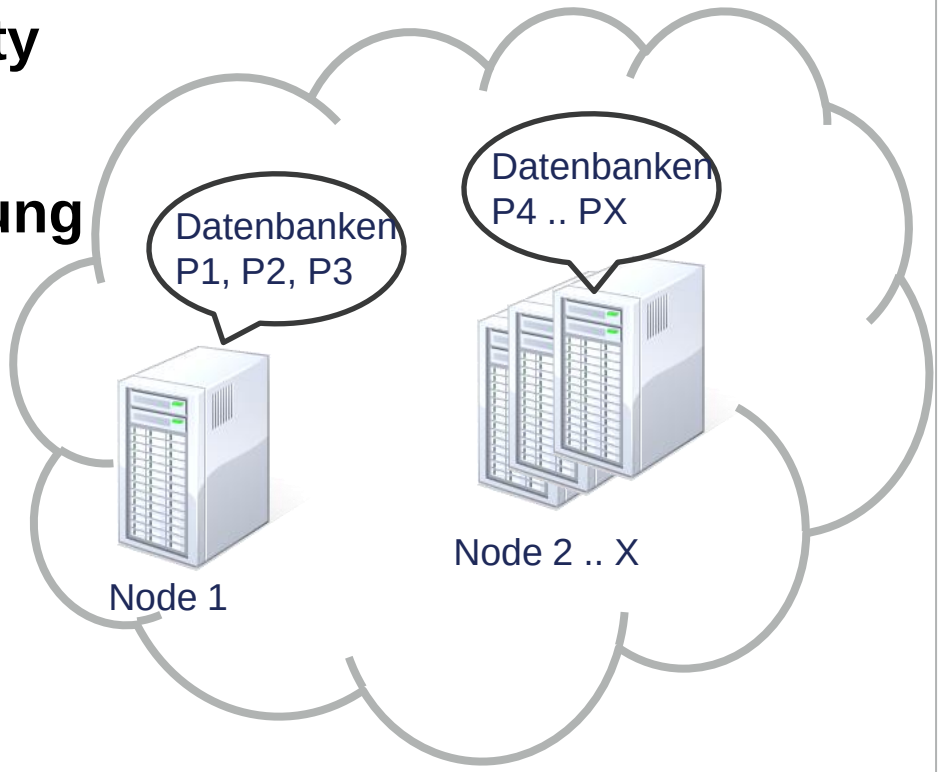
Oracle Restart Erfahrungen

- **Planung**
  - Komplexität der Grid Infrastructure berücksichtigen
  - Den Aufwand für Know How-Aufbau berücksichtigen
- **Implementierung**
  - Implementierung etwas umständlich, aber reibungslos
- **Betrieb**
  - Überwachung auf DB-Prozessebene ➔ Verfügbarkeit wird deutlich erhöht
  - Einige neue Dateien, OS Prozesse, Logs etc. ➔ Know How aufbauen!
  - Alle Komponenten werden automatisch und in der richtige Reihenfolge gestartet ➔ weniger administrativer Aufwand
  - Datenbanken sollen mit **srvctl** Befehlen gestoppt und gestartet werden

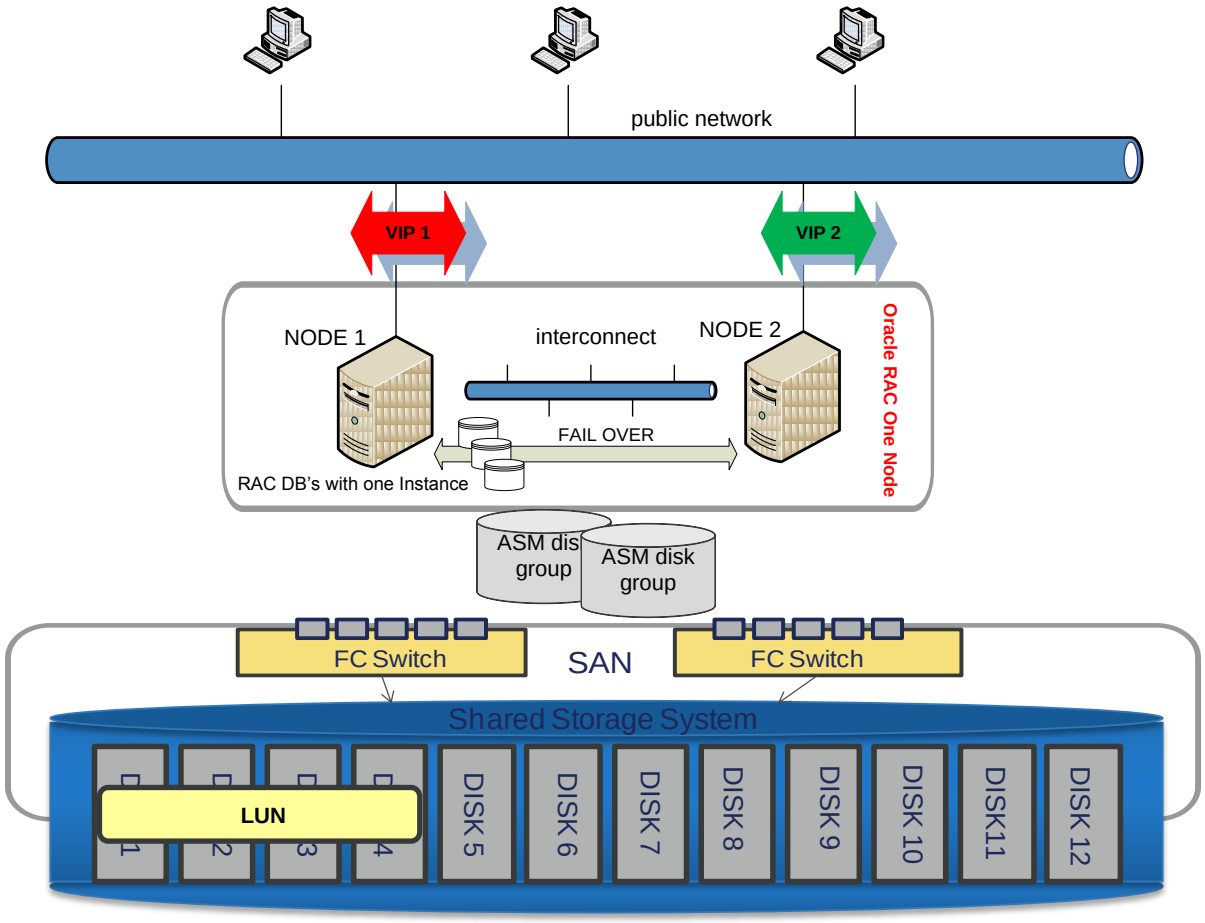
Oracle One Node RAC

Oracle One Node RAC (ONR) – ein paar Fakten

- **Version 11.2.0.1 ist nur für Linux zertifiziert**
- **Ab der Version 11.2.0.2 für alle (11gR2) zertifizierten Plattformen**
- **Nicht supported für 3-rd Party Clusterware (Veritas, IBM …)**
- **Enterprise Edition Lizenzierung**

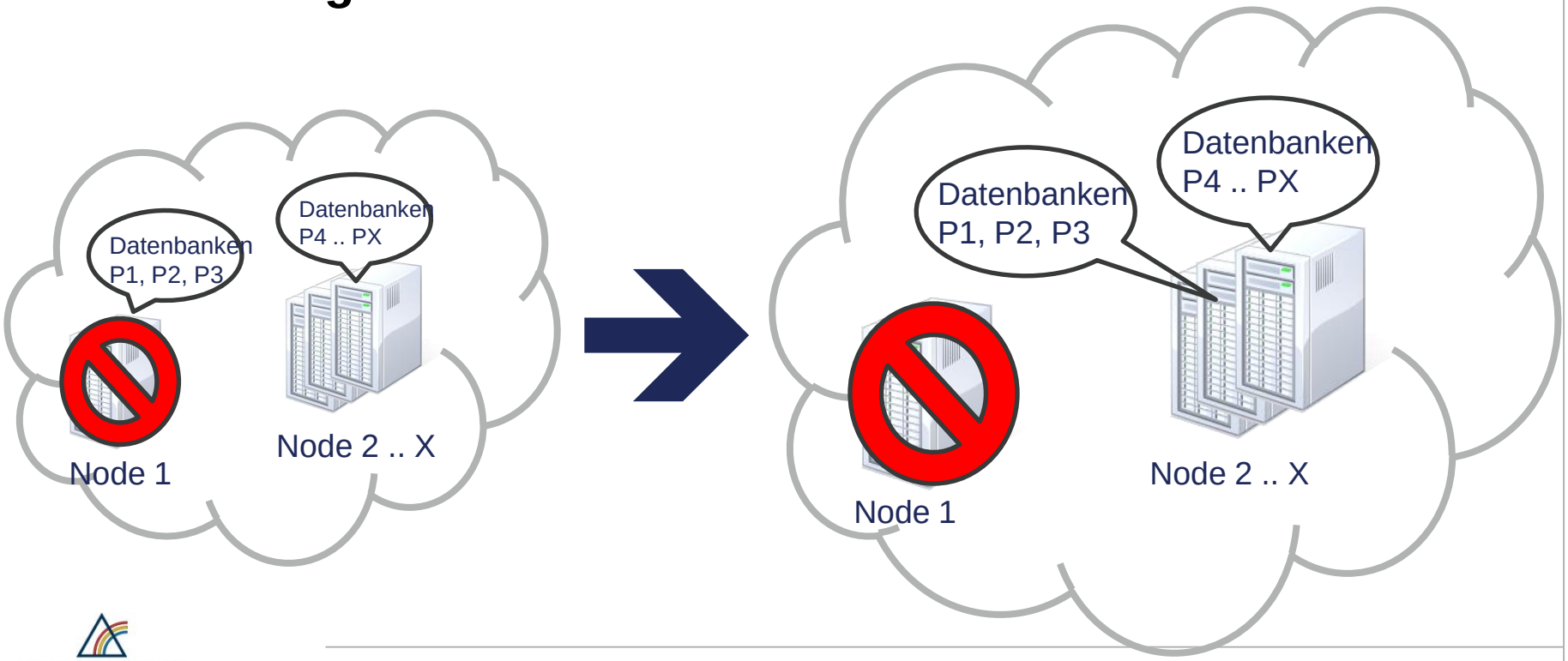


ONR Architektur



Das ONR-Prinzip (1/2)

- **Im Falle eine SW- oder HW-Defekts wird die Instanz auf einem anderen Cluster-Node neu erstellt (!) und gestartet**
- **Anschließend können die Clients mit der neuen Instanz Verbindungen aufbauen**



Das ONR-Prinzip (2/2)

- **Unterschied zu RAC**
  - ONR ist ein RAC, deren Datenbank nur auf einem Knoten läuft
- **Unterschied zu Oracle Restart (OR)**
  - OR startet ausgefallenen Dienste (z. B. Datenbank) neu
  - Fällt ein ganzer Server aus, nutzt dies nichts
  - ONR kann den Dienst auf einem anderen Cluster-Knoten starten (ähnlich Failover Cluster)
- **Unterschied zum Failover Cluster**
  - ONR automatisiert den Failover auf anderen Knoten des Clusters

ONR Erfahrungen

- **Planung**
  - Komplexität der Grid Infrastructure berücksichtigen
  - Kosten und Nutzen Analyse durchführen
- **Implementierung**
  - Ähnliche Voraussetzungen wie bei RAC
  - Bei guter Vorbereitung – reibungslose Installation
- **Betrieb**
  - Gleiche Architektur wie bei RAC ➔ Know How aufbauen!
  - Failover Prozess kann einige Minuten dauern ➔ Kann die Anwendung damit umgehen?
- **Kritische Fragestellung: rechtfertigen die höheren Lizenzkosten und die Komplexität den Einsatz von ONR? Gibt es Alternativen? Failover Cluster mittels Clusterware?**

Real Application Cluster

Clusterware-, ASM- und Datenbank-Software

- **Oracle 10g, 11g Release 1**
  - /u00/app/root/product/crs1020 — Cluster-Software
  - /u00/app/oracle/product/10.2.0 — Datenbank- und ASM-Software
- **Oracle 11g Release 2**
  - /u01/11.2.0/grid — Cluster- und ASM-Software
  - /u00/app/oracle/product/11.2.0/db_1 — Datenbanksoftware
- **ASM und Listener sind vom Datenbank – in das Cluster-Home umgezogen**
- **Die Cluster-Software läuft mehrheitlich im Kontext von root**
  - Es spricht technisch nichts dagegen, diese ausserhalb von ORACLE_BASE zu installieren (/u00/app/root). OFA Prinzip bleibt erhalten
    - Parent-Directory wird auf Owner root gesetzt

Änderungen gegenüber 10g

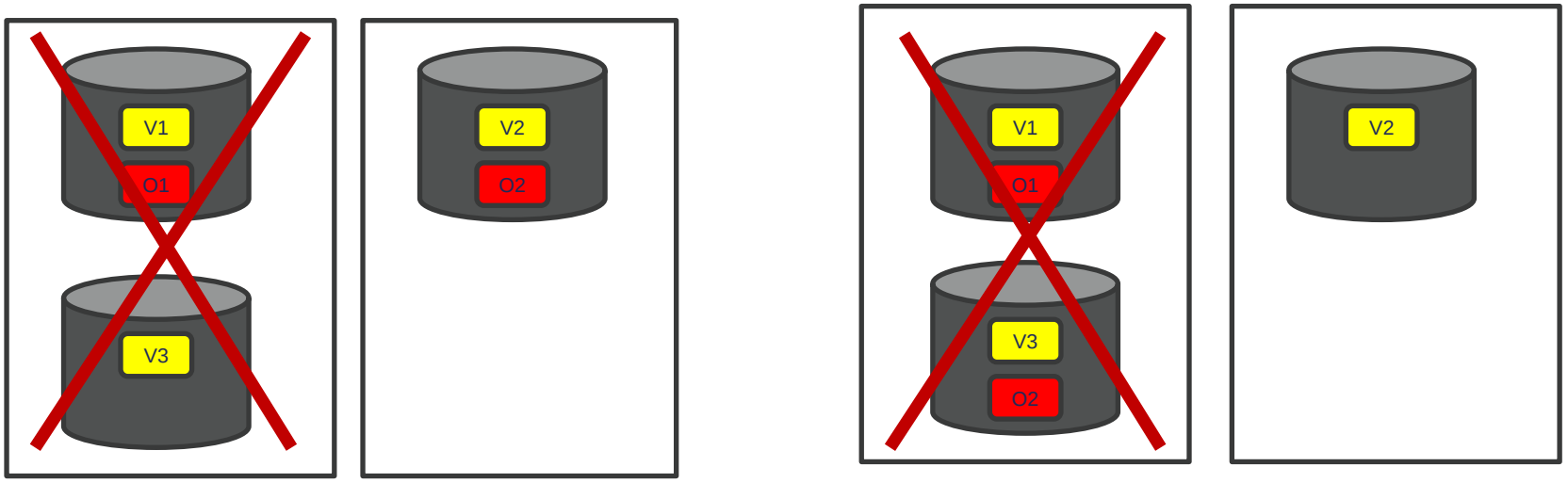
- **Rollentrennung zwischen dem Grid Admin und DBA**
- **ASM wird mit der Clusterware installiert (nicht mehr mit der Datenbank)**
- **Listener läuft im Grid Home**
- **OCR und Voting Disk in ASM**
- **ACFS**
  - Cluster Filesystem, das auf ASM aufsetzt
- **SCAN (Single Client Access Name)**
  - Ein Hostname für den gesamten Cluster
  - Mehrere IP Adressen (typischerweise 3)

Voting Disk

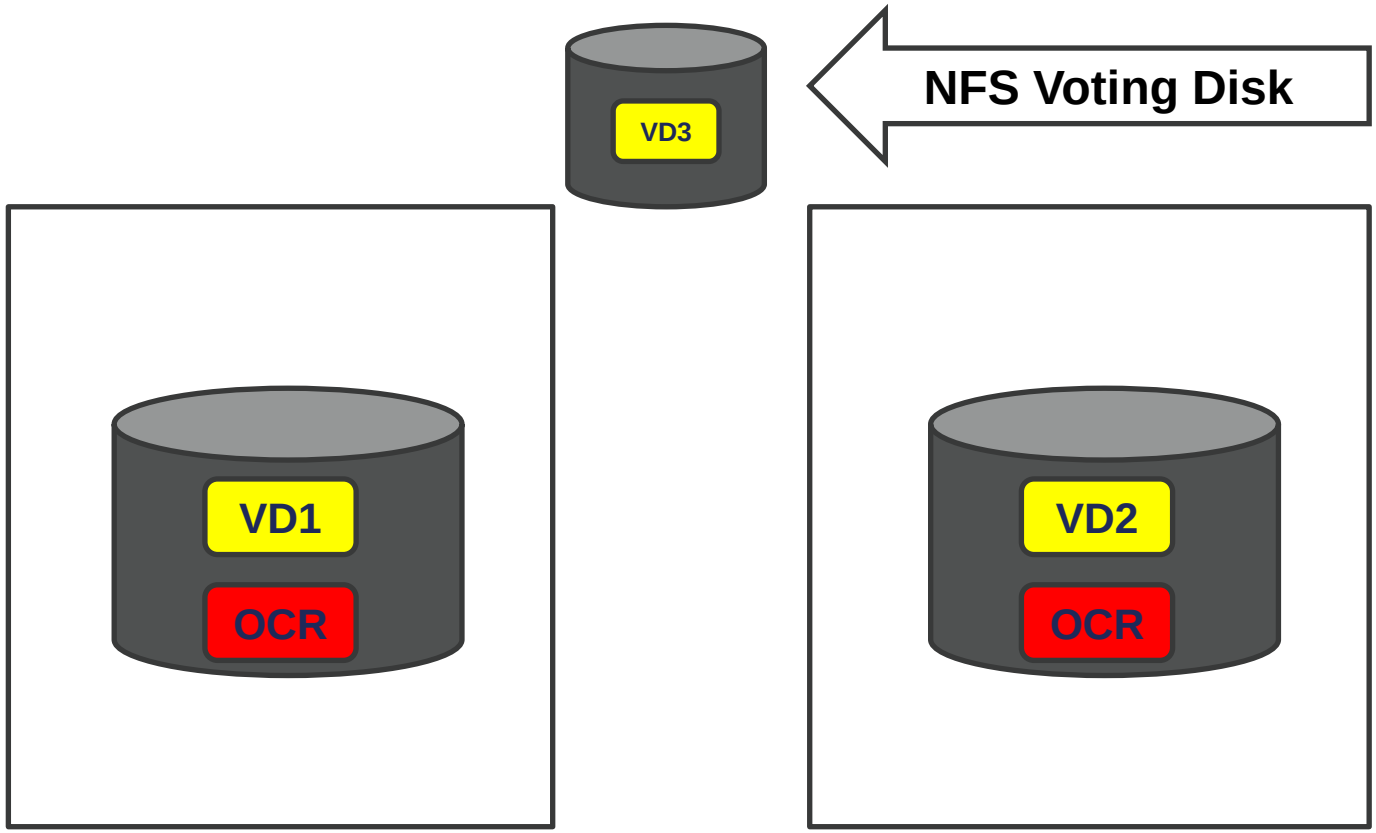
- **Voting Disk wird in einer ASM Diskgroup gespeichert**
- **Normal Redundancy: 3 ASM Disks erforderlich**
- **Clusterware greift direkt auf ASM Disks zu**
  - Beim Start der Clusterware ist ASM noch nicht verfügbar, also auch keine Diskgroups
- **Altes 2 Rechenzentren-Problem bleibt bestehen**
  - Eine Seite hat zwei Voting Disks = Quorum
  - NFS Voting Disk einrichten
  - ASM Disk auf NFS Share
  - Markieren als „Quorum“
    - Stellt sicher, daß ausschließlich die Voting Disk auf dem NFS Share gespeichert wird und nicht etwa z.B. die OCR
- **Voting Disk Backup in OCR**

OCR

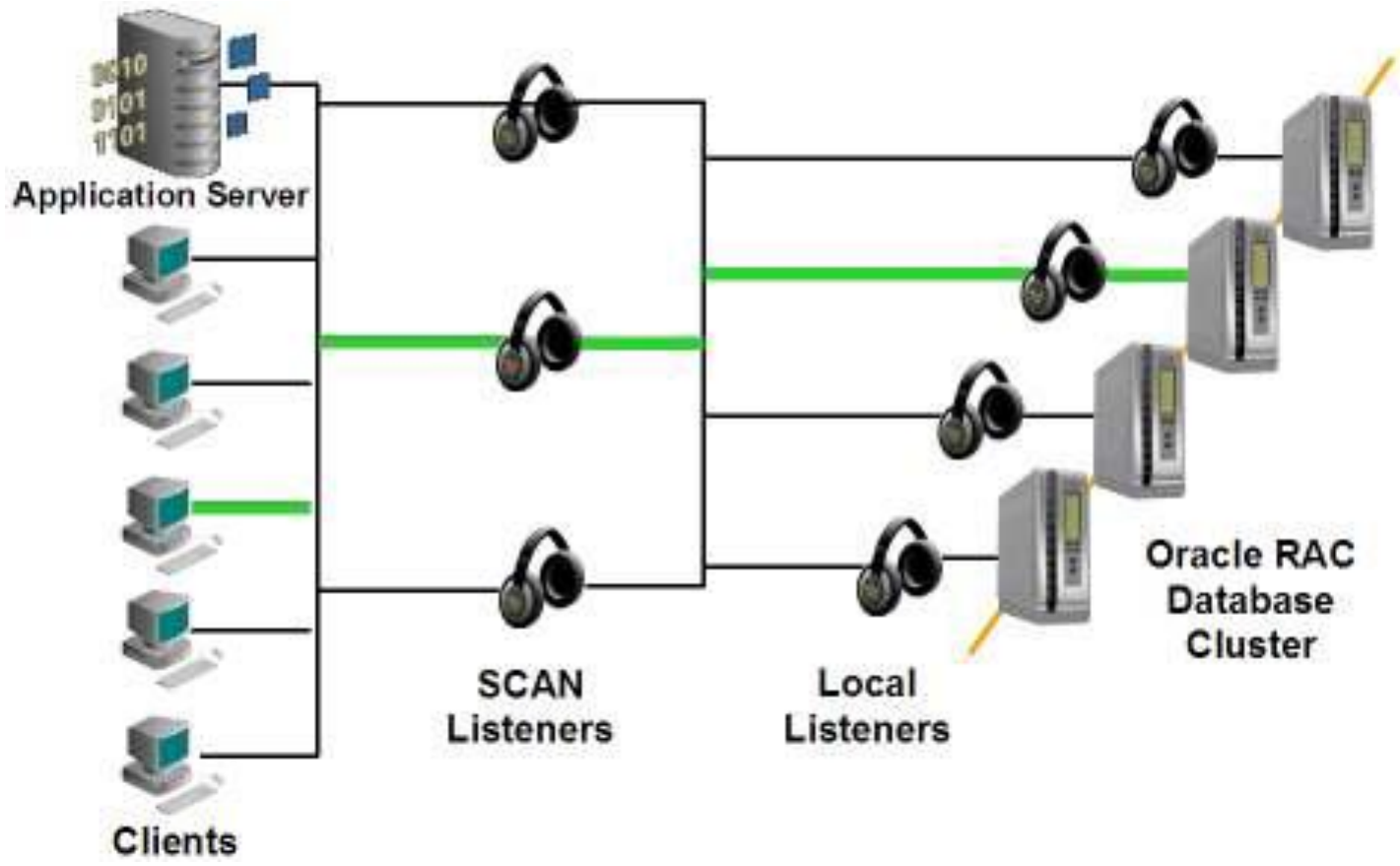
- **OCR Speicherung in ASM Diskgroup**
- **Normal Redundancy: 2 Disks**
- **Installation: eine Diskgroup wird erstellt**
  - OCR und Voting Disks
  - 3 ASM Disks (genauer: Failgroups)
  - 2 Rechenzentren Problem



OCR und Voting Disks aufräumen



Single Client Access Name (SCAN)

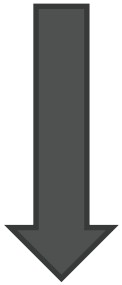


Quelle: http://www.oracle.com/technetwork/database/clustering/overview/scan-129069.pdf

SCAN

- **11gR2 Client wird vorausgesetzt**

```
DB10G_VIA_SCAN =
      (DESCRIPTION =
         (ADDRESS_LIST =
           (ADDRESS = (PROTOCOL = TCP)(HOST = my-scan-host)(PORT = 1521)))
            (CONNECT_DATA =
           (SERVER = DEDICATED)
           (SERVICE_NAME = DB10G_SCAN)))
```



- **Workaround bei älteren Clients**

```
DB10G_VIA_SCAN =
      (DESCRIPTION =
         (ADDRESS_LIST =
           (ADDRESS = (PROTOCOL = TCP)(HOST = 10.1.10.200)(PORT = 1521))
           (ADDRESS = (PROTOCOL = TCP)(HOST = 10.1.10.201)(PORT = 1521))
           (ADDRESS = (PROTOCOL = TCP)(HOST = 10.1.10.202)(PORT = 1521)))
        (LOAD_BALANCE = yes)
        (CONNECT_DATA =
          (SERVER = DEDICATED)
          (SERVICE_NAME = DB10G_SCAN)))
```

Rollentrennung Grid Admin und DBA

- **Nach Best Practice Empfehlung sollten die Grid und Datenbank Benutzer getrennt werden:**
  - /usr/sbin/useradd -u 501 -g oinstall -G asmadmin,asmdba,asmoper **grid**
  - chown -R grid:oinstall /u01/11.2.0/grid
  - /usr/sbin/useradd -u 502 -g oinstall -G dba,asmdba **oracle**
  - chown -R oracle:oinstall /u01/app/oracle/product/11.2.0/db_1
- **Nach der Grid und DB Installation können keine Remote-Verbindungen zur Oracle DB hergestellt werden:**
  - ORA-12537 "TNS: connection closed"
- **MOS ID 1069517.1: ORA-12537 if Listener (including SCAN Listener) and Database are Owned by Different OS User**

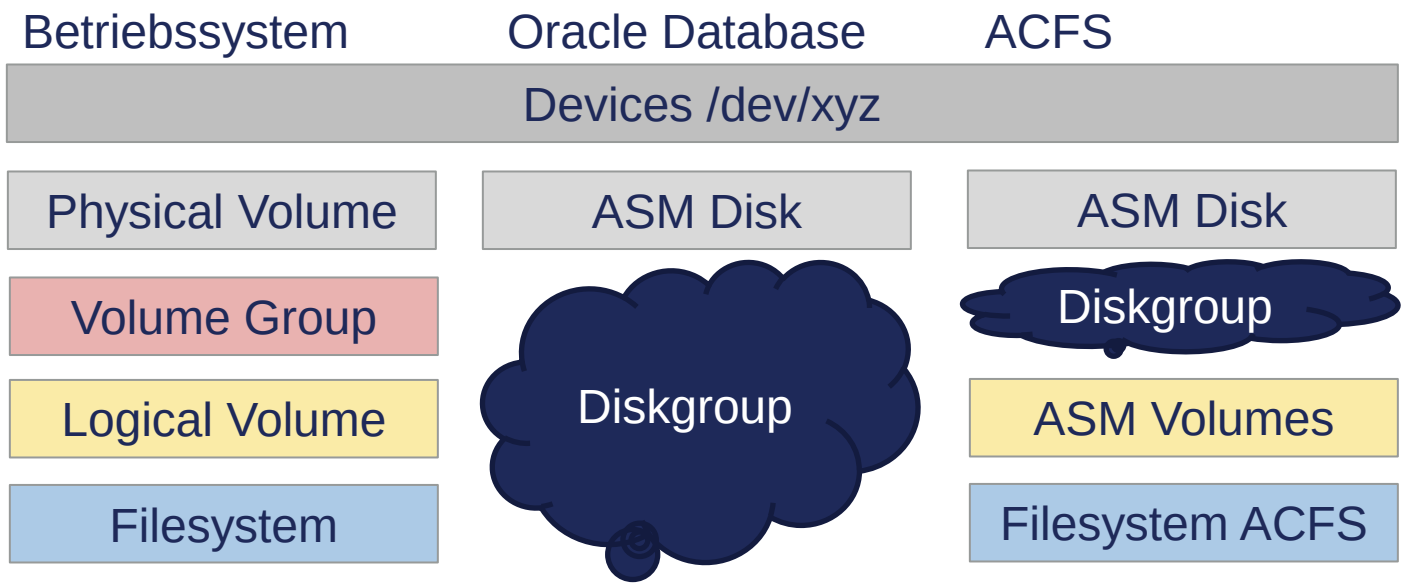
RAC 11gR2 Erfahrungen

- **Planung**
  - Neue Architektur, viele neue Begriffe (mittlerweile zahlreiche Dokumentationen)
  - Deutlich mehr Aufwand bei der Planung, dafür aber weniger Aufwand bei der Implementierung
  - Mehr IP‘s und DNS Names müssen eingeplant werden
- **Implementierung**
  - Bei guter Vorbereitung – i.d.R. reibungslose Installation
  - Verbesserte und intuitiver Installer – Installationsvoraussetzungen können mit der Hilfe vom Installer korrigiert werden
- **Betrieb**
  - Stabile und robuste Systeme, altbewährte Technik
  - Bei Timeouts deutlich weniger Reboots von Knoten
  - Neue Verzeichnisse, OS Prozesse, Logs – Knowhow muss zwingend aufgebaut werden

ASM Cluster Filesystem

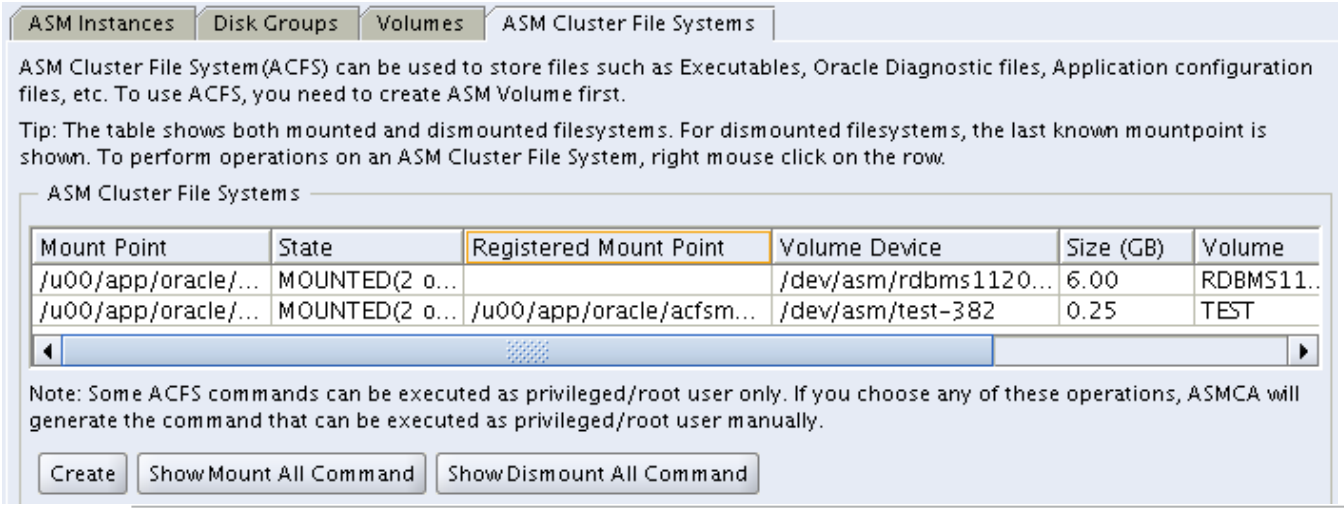
ACFS

- **Clusterfilesystem auf ASM Diskgroups**
  - Mirroring über ASM Diskgroup Redundancy
  - Nicht für Datenbankfiles (direkt in ASM speichern)



ACFS (Fortsetzung)

- **Bestandteil von OCFS (Oracle Cloud File System)**
  - Die kostenpflichtigen Datenbank Editionen Standard Edition One, Standard Edition und Enterprise Edition beinhalten eine restricted use Lizenz für OCFS, beschränkt auf das Speichern von Oracle Software Binaries, Metadaten und Diagnose Files. Wenn User-Files in OCFS gespeichert werden sollen, ist OCFS zu lizenzieren.
- **asmca**
  - GUI zur Administration

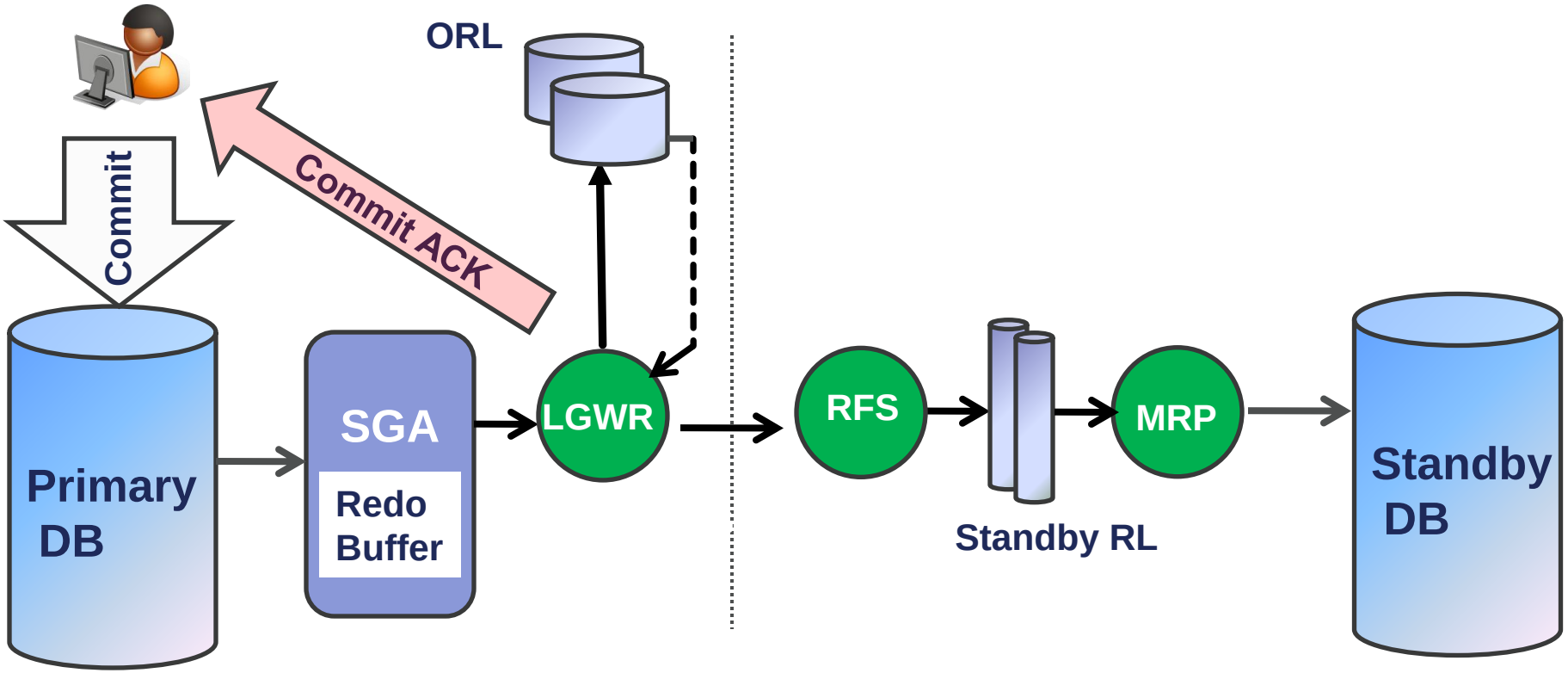


ACFS Erfahrungen

- **Planung**
  - Abhängigkeiten gegenüber Backup Media Manager muss geprüft werden
    - Diverse Hersteller unterstützen ACFS als FS-Typ nicht => kein Backup
  - Überlegung was in ACFS abgelegt wird. Ist es als echtes Cluster File System ernst zu nehmen
- **Implementierung**
  - Wenn ASM und LVM KnowHow vorhanden, sehr einfach und schnell einzubinden
- **Betrieb**
  - Sicherung des Filesystems mittels Legato Networker problematisch
  - Verkleinern von Filesystemen laut Doku möglich, in der Praxis – nicht
  - Aktuell wenig im Einsatz

Data Guard / Active Data Guard

Physical Standby 11G



Duplicate from Active Database

- **Passwortfile auf die Gegenseite kopieren**

```
alter system archive log current;
$ORACLE_HOME/bin/rman
connect target sys/***@prim
connect auxiliary sys/***@sby
startup clone nomount;
DUPLICATE TARGET DATABASE FOR STANDBY FROM ACTIVE DATABASE;
```

oder

```
DUPLICATE TARGET DATABASE
FOR STANDBY
FROM ACTIVE DATABASE
DORECOVER
SPFILE
SET "db_unique_name"="foou" COMMENT ''Is a duplicate',
SET LOG_ARCHIVE_DEST_2="service=inst3 ASYNC REGISTER
VALID_FOR=(online_logfile,primary_role)„
SET FAL_SERVER="inst1" COMMENT "Is primary„
NOFILENAMECHECK;
```

Active Data Guard

- **Active Data Guard: kostenpflichtige Option**
- **Auch „Real Time Query“ genannt**
- **Selects (Reports) bei laufendem Redo apply möglich**
- **Aktivierung**
  - ALTER DATABASE OPEN;
  - ALTER DATABASE RECOVER MANAGED STANDBY DATABASE DISCONNECT;

Active Data Guard (Fortsetzung)

- **ALTER SESSION SET STANDBY_MAX_DATA_DELAY=**
  - Festlegen, wie zeitnah man gerne an der Primary dran wäre
  - Wenn Delay nicht gehalten werden kann:
    ORA-03172: STANDBY_MAX_DATA_DELAY of 0 seconds exceeded
  - ALTER SESSION SYNC WITH PRIMARY;
  - Für STANDBY_MAX_DATA_DELAY=0 muß LogXPTMode=sync sein.
- **Switchover**
  - Neue Standby ist nach Switchover auch im real-time Query Modus
- **Automatic Block Recovery**
  - Korrupte Blöcke in der Primary werden automatisch aus der Standby repariert

Active Data Guard: Achtung Lizenzverletzung!

- **Active Data Guard: kostenpflichtige Option**
- **Es reicht ein „alter database open“**
  - Broker startet den Redo Apply
- **Achtung für Windows**
  - Früher:
    Windows Dienst startet Standby Datenbank im „open“ Mode
    Broker schließt die Datenbank wieder und startet Redo Apply
  - Jetzt:
    Windows Dienst startet Standby Datenbank im „open“ Mode
    Broker startet Redo apply
  - Ergebnis:

```
SQL> select open_mode from v$database;
OPEN_MODE
--------------------
READ ONLY WITH APPLY
```

Bewertung HA Lösungen

Methode	Alle Knoten lizenz-pflichtig	Lizenzs-pflichtige Option	Verfügbarkeit	Komplexität im Betrieb	Out-of-the-Box Lösung	OC Empfehlung
Oracle Restart	Ja	Nein	Niedrig	Niedrig		
Oracle One Node RAC	Ja	Ja	Hoch	Mittel		Je nach Anforderung
Failover Single DB mittels Clusterware	Nein, 10 Tage Regel	Nein	Hoch	Mittel	Nein, manuelle Konfiguration notwendig	
HACMP , Linux HB etc.	Nein, 10 Tage Regel	Nicht aus Oracle Sicht	Mittel, der Zustand der DB wird nicht berücksichtigt	Mittel	Jain, manuelle Konfiguration notwendig	Ja, falls keine andere HA Lösungen vorhanden
RAC	Ja	EE – ja SE – nein	Sehr Hoch	Hoch		
DG /ADG	Ja	Nein/Ja	Sehr Hoch	Hoch		

Fragen und Antworten

Kontaktdaten

Rainier Kaczmarczyk, Solution Architect

OPITZ CONSULTING München GmbH
rainier.kaczmarczyk@opitz-consulting.com
Telefon +49 89 680098 – 5409
Mobil +49 173 5773806