Herrmann & Lenz
Services GmbH

RAC 11.2 mit Data Guard an einem entfernten Standort

Susanne Jahr

DOAG Konferenz + Ausstellung
November 2010

Herrmann & Lenz Services GmbH
Herrmann & Lenz Solutions GmbH

- Erfolgreich seit 1996 am Markt
- Firmensitz: Burscheid (bei Leverkusen)
- Beratung, Schulung und Betrieb/Fernwartung rund um das Thema Oracle Datenbanken
- Schwerpunktthemen: Hochverfügbarkeit, Tuning, Migrationen und Troubleshooting
- Herrmann & Lenz Solutions GmbH
  - Produkt: Monitoring Module
  - Stand auf Ebene 2

Beweggründe / Anforderungen

- Geschäftskritische Webanwendung: Online Bestell-Plattform für Krankenhäuser und Apotheken
- Größte Last täglich zwischen 10:00 und 13:00
- Ausfallsicherheit wichtigster Aspekt
  - Möglichst viele Fehler-Arten sollen kompensiert werden können
  - überwiegt den Aspekt der Skalierung / Performance-Steigerung
- Umzug der Hardware von einem externen Betreiber ins eigene Rechenzentrum

Lösung: RAC mit Data Guard

- Durch RAC abgesicherte Fehler
  - Ausfall von CPU, Memory oder sonstigen Hardware-Komponenten eines Servers
  - Ausfall eines gesamten Servers
- Durch Data Guard abgesicherte Fehler
  - Fehler im Shared Storage des RAC
  - Ausfall des gesamten Clusters
  - Eventuell: Anwender-Fehler (falls Delay eingestellt wird)

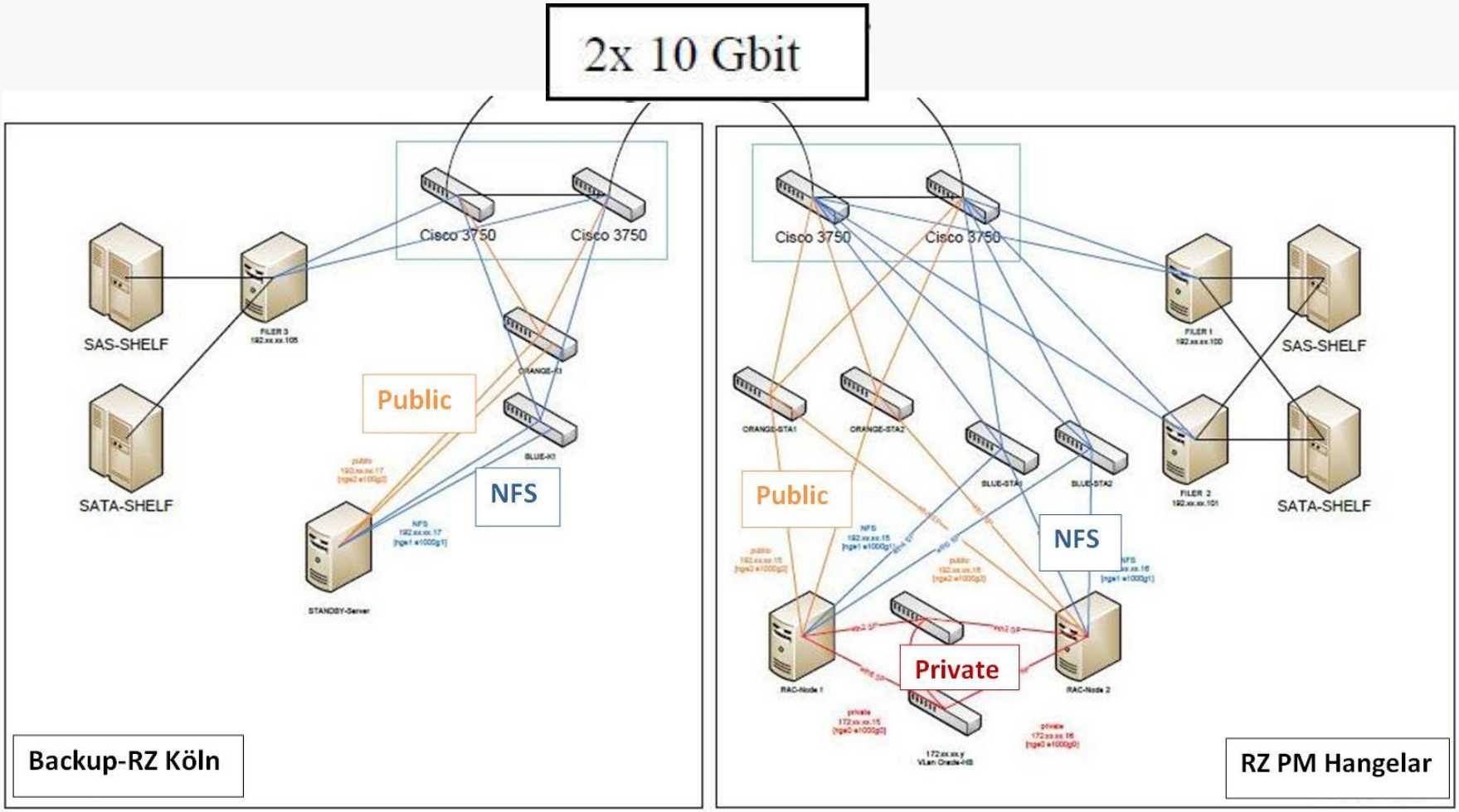
Technische Voraussetzungen (1)

- Betriebssystem
  - Solaris 10 x86-64
- Server
  - SUN Fire X42xx
  - AMD Quadcore, 32 GB RAM
- Shared Storage
  - NetApp FAS2040 File-Server
  - Cluster-Controller
  - Redundante Netzwerk-Anbindung

Technische Voraussetzungen (2)

- Netzwerkverbindung
  - Sowohl lokal als auch zwischen den Rechenzentren: 10GBit über redundant ausgelegte Cisco3750-Switches
  - Physikalische Netzwerkkarten für das private Netzwerk (Cluster Interconnect) sind redundant ausgelegt und über Link Aggregation zu jeweils einer virtuellen Schnittstelle verbunden

Standort-Übersicht



Konfiguration Shared Storage (1)

- Vorbereitung des Shared Storage
  - Cluster-fähige Dateisysteme durch NetApp-Filer bereitgestellt
  - Spezielle NFS-Mount-Optionen beachten!
  - Persistent durch Einträge in die vfstab

Konfiguration Shared Storage (2)

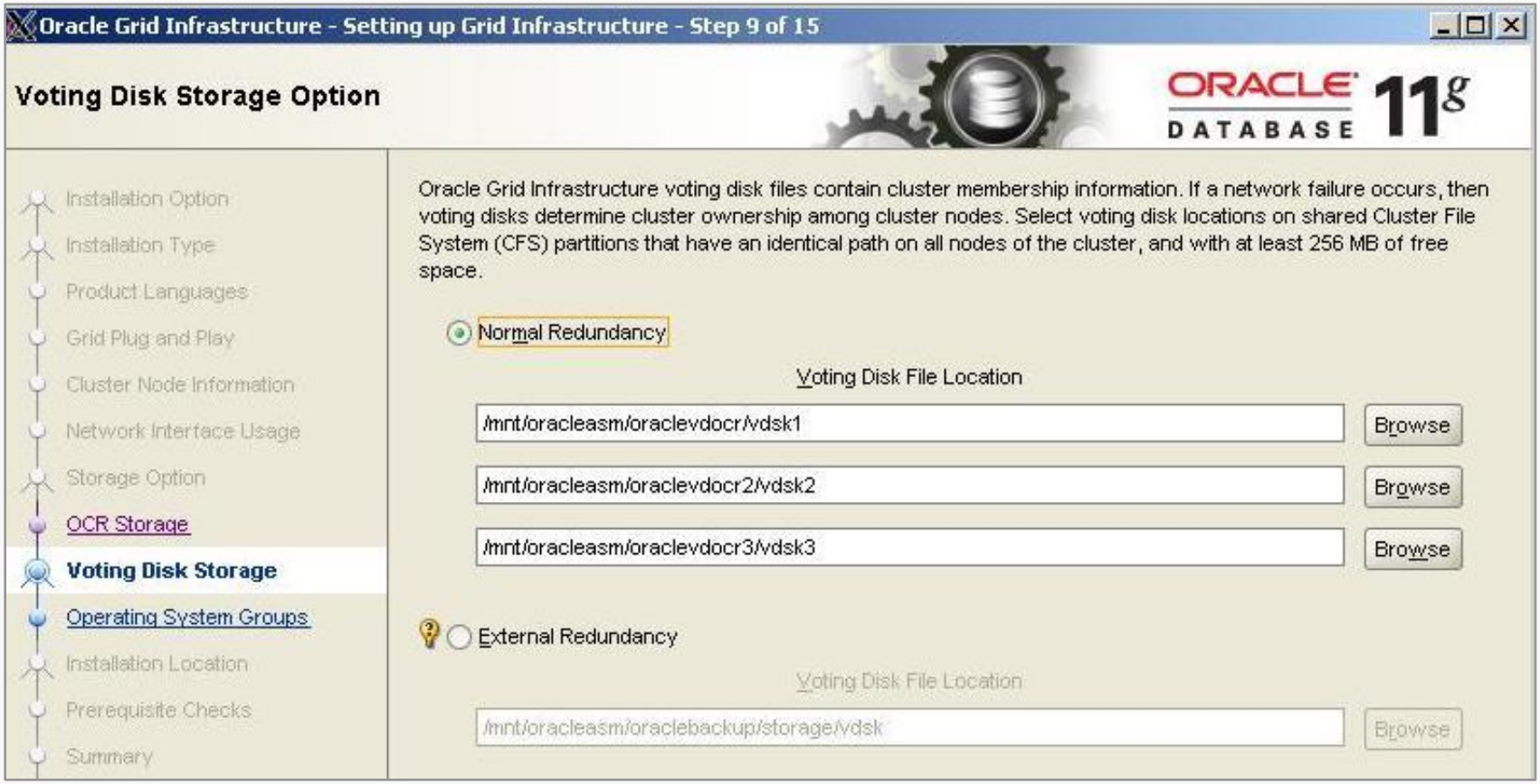
- Beispiel OCR und Voting Disks:

```
<filerSTA1>:/vol/oraclevdocr -
   /mnt/oracleasm/oraclevdocr
   nfs - yes rw,bg,hard,nointr,
   rsize=32768,wsize=32768,
   proto=tcp,vers=3,noac,forcedirectio
```

- Beispiel Datenbankdateien:

```
<filerSTA1>:/vol/oracledata -
  /mnt/oracleasm/oracledata nfs - yes
  rw,bg,hard,nointr,rsize=32768,
  wsize=32768,
  proto=tcp,noac,forcedirectio,vers=3,suid
```

Konfiguration Shared Storage (3)



Konfiguration OS / Oracle-Umgebung

- Kernel-Parameter und Shell-Limits für den oracle-User als Projekt in /etc/project
- Zusätzlich sind einige Parameter-Anpassungen in /etc/system erforderlich!
- ssh-Konfiguration / User Equivalency wird über den Universal Installer konfiguriert (neu in 11.2!)

Konfiguration Netzwerk

- Hinterlegung der Namen und Adressen für
  - Privates Netz
  - Öffentliches Netz und Knoten-VIPs
  - SCAN

  im DNS
- Konfiguration der öffentlichen und privaten physikalischen Netzwerk-Schnittstellen
- Empfehlenswert: redundante Netzwerkkarten verwenden – entweder durch Trunking / Link Aggregation oder durch IPMP (IP Multipathing)

Link Aggregation vs. IPMP (1)

- Durch Link Aggregation werden zwei physikalische Netzwerkkarten virtuell vom Betriebssystem als eine Schnittstelle präsentiert

```
# dladm create-aggr -d nge2 -d e1000g2 1
bash-3.00# dladm show-aggr

key: 1(0x0001) policy: L4 address: :21:28:3d:7b:84(auto)
device    address           speed     duplex  link   state
nge2      0:12:34:5d:6b:78  1000Mbps  full    up     attached
e1000g2   0:87:65:c4:3b:21  1000Mbps  full    up     attached

# ifconfig
aggr1: flags=1000843<UP,BROADCAST,RUNNING,MULTICAST,IPv4>
   mtu 1500 index 2
         inet xxx.xxx.117.15 netmask ffff0000 broadcast
   xxx.xxx.255.255
         ether 0:12:34:5d:6b:78
```

Link Aggregation vs. IPMP (2)

- Bei Verwendung von IPMP: **alle** Adapter, die Mitglieder der IPMP-Gruppe sind, als Netzwerke für den Private Interconnect während der Installation angeben!
- Falls nicht möglich, im Anschluss den betreffenden Adapter manuell hinzufügen (MOS Doc ID # 1069584.1: Solaris IPMP and Trunking for the cluster interconnect in Oracle Grid Infrastructure 11g Rel. 2):

```
$GRID_HOME/bin/oifcfg setif -global
nge2/xxx.xxx.117.0:cluster_interconnect
```

RAC 11.2 – Was hat sich geändert?

- Grid Infrastructure statt Clusterware
  - Beinhaltet OCR und Voting Disk sowie ASM und Listener
  - Keine Unterstützung von Raw Partitions / Block Devices für Cluster-Files in neuen Installationen – im vorliegenden Projekt Speicherung auf NetApp-NFS-Shares
  - Virtueller Name für das gesamte Cluster (SCAN) zum Verbindungsaufbau durch die Clients

Konfiguration SCAN

- Konfiguration des SCAN und der verschiedenen SCAN-Listener können über `srvctl` abgefragt und modifiziert werden:

```
oracle@ RAC-Node2:~> srvctl config scan
SCAN name: myrac-scan.mydomain.com, Network:
  1/192.xxx.xxx.0/255.255.255.0/nge1
SCAN VIP name: scan1, IP:
/myrac-scan.mydomain.com/192.xxx.xxx.20
SCAN VIP name: scan2, IP:
/myrac-scan.mydomain.com /192. xxx.xxx.21
SCAN VIP name: scan3, IP:
/myrac-scan.mydomain.com /192. xxx.xxx.22
```

SCAN in der tnsnames.ora

- Die Clients benutzen den SCAN zur Verbindungsaufnahme mit der Datenbank (keine Adresslisten mehr!):

```
MALLPCL =
   (DESCRIPTION =
       (ADDRESS = (PROTOCOL = TCP)(HOST =
        myrac-scan.mydomain.com)(PORT = 1521))
       (CONNECT_DATA =
            (SERVER = DEDICATED)
            (SERVICE_NAME = MALLPCL)
       )
   )
```

Standby-Systeme (1) - Allgemein

- Drei Datenbanken im Cluster
- Jede Datenbank bekommt eine physikalische Standby-Datenbank im Backup-Rechenzentrum
- db_unique_name der Standby-Datenbanken:

  db_name*stb* (*MALLP* → *MALLPSTB*)

Standby-Systeme (2) - Parameter

- Eintrag jeder Standby-Datenbank in der listener.ora
- Erstellung von Passwortdateien mit identischen SYS-Passworten auf Primary- und Standby-Seite
- Start einer Instanz mit dem korrekten `db_unique_name` und den erforderlichen DataGuard-Parametern
  - `fal_server`: zwei Datenbank-Services, die jeweils auf einer der beiden RAC-Instanzen konfiguriert wurden (`MALLP_SRV1`, `MALLP_SRV2`)

Standby-Systeme (3) - Parameter

- log_archive_1 wird erweitert
  ('LOCATION=/mnt/oracleasm/oraclearc/
  MALLP **VALID_FOR=(ALL_LOGFILES,ALL_ROLES)
  DB_UNIQUE_NAME=MALLPSTB'**)
- log_archive_dest_2 wird als Service zum Log Transport mittels Logwriter konfiguriert
  ('SERVICE=MALLPCL LGWR ASYNC VALID_FOR
  =(ONLINE_LOGFILES,PRIMARY_ROLE)
  DB_UNIQUE_NAME=MALLP')
- log_archive_config: eine DG_CONFIG mit Primär- und Standby-System ('DG_CONFIG=(MALLPSTB,MALLP)'

Standby-Systeme (4) - Anlage

- Erstellung mittels RMAN auf der Standby-Seite
  - Zunächst Anmeldung an die Target-DB (Primary) und die Auxiliary Instance (spätere Standby-DB):

```
rman target sys/xxx@MALLP1_SRV
  auxiliary sys/xxx@MALLPSTB
```

  - Dann Erstellung der Standby-Datenbank

```
RMAN>DUPLICATE TARGET DATABASE FOR
  STANDBY FROM ACTIVE DATABASE;
```

Standby-Systeme (5) – Standby Redolog

- Anlage von Standby-Redolog-Dateien für den Log Transport über den Logwriter
- Anzahl Online-Redolog-Dateien +1 pro Redo-Thread
- Größe Standby-Redolog = Größe Online-Redolog
- Sinnvoll als Vorbereitung für den Failover-Fall:
  - Anlage der Standby-Redologs auch schon in den Primär-Datenbanken
  - Konfiguration beider Seiten sowohl für die Primary- als auch für die Standby-Rolle

Standby-Systeme (6) - Kontrolle

- Kontrolle der Synchronität der Standby-Systeme zu ihren Primär-Datenbanken und der korrekten Konfiguration :
  - `v$archive_dest`
  - `v$dataguard_status`
  - `v$standby_log`
  - `v$log`
  - `v$managed_standby`
  - `v$archived_log`
  - `v$log_history`
  - `v$archive_gap`

Standby-Systeme (7) - Kontrolle

- Nützliches SQL-Skript MOS Doc ID # 241438.1: Script to Collect Data Guard Physical Standby Diagnostic Information

Zusammenfassung

- Einige Neuerungen im RAC 11.2, diese sind nach bisherigen Erfahrungen gut gelungen
- Vereinfachung der Anlage von Standby-Datenbanken (schon in 11.1) durch die DUPLICATE … FROM ACTIVE DATABASE-Option
- Failover-Tests der NetApp-Filer verliefen problemlos, Performance absolut zufriedenstellend
- Es muss also nicht immer ASM sein (jedenfalls in der Enterprise Edition)

Fragen & Kontakt

- susanne.jahr@hl-services.de
- http://www.hl-services.de
- Hier in der Ausstellung Ebene 2 (gelb)