

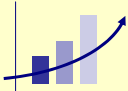


Oracle Datenbanken und RAID-Technologie

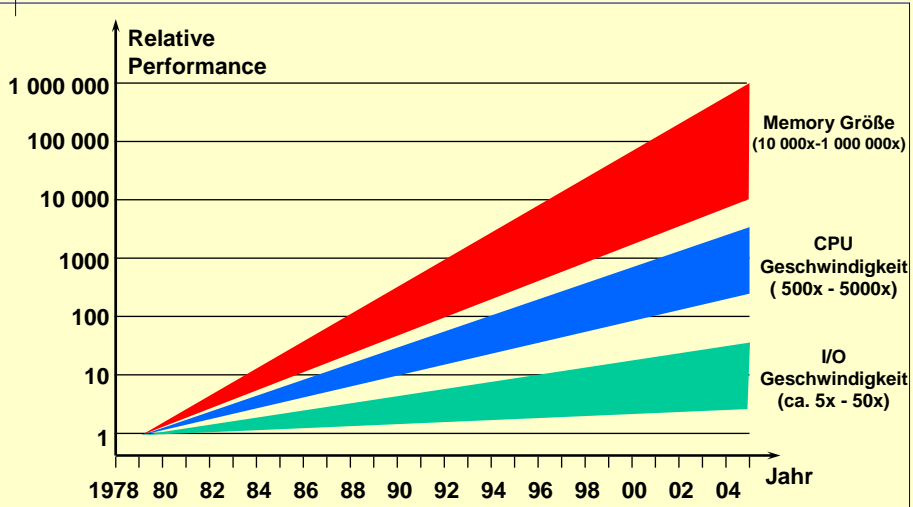
28. DECUS München e.V.
Symposium 2004
Maritim Hotel Bonn
1C03

Hermann Brunner, Angerwiese 15, 85567 Grafing
www.brunner-consulting.de

brunner consulting Oracle Datenbanken und RAID-Technologie 1

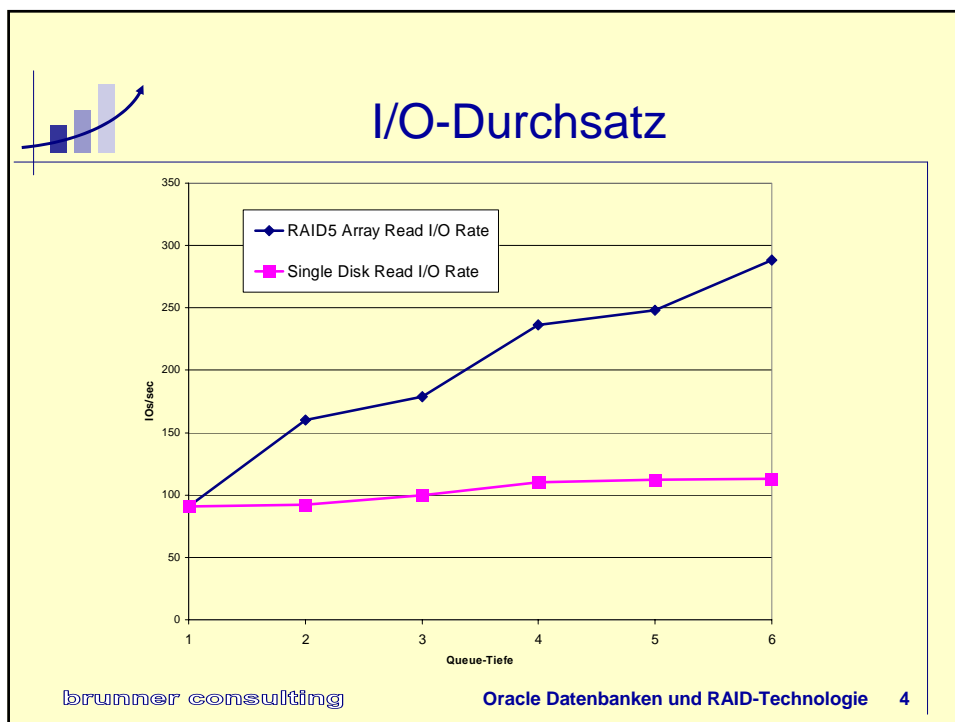
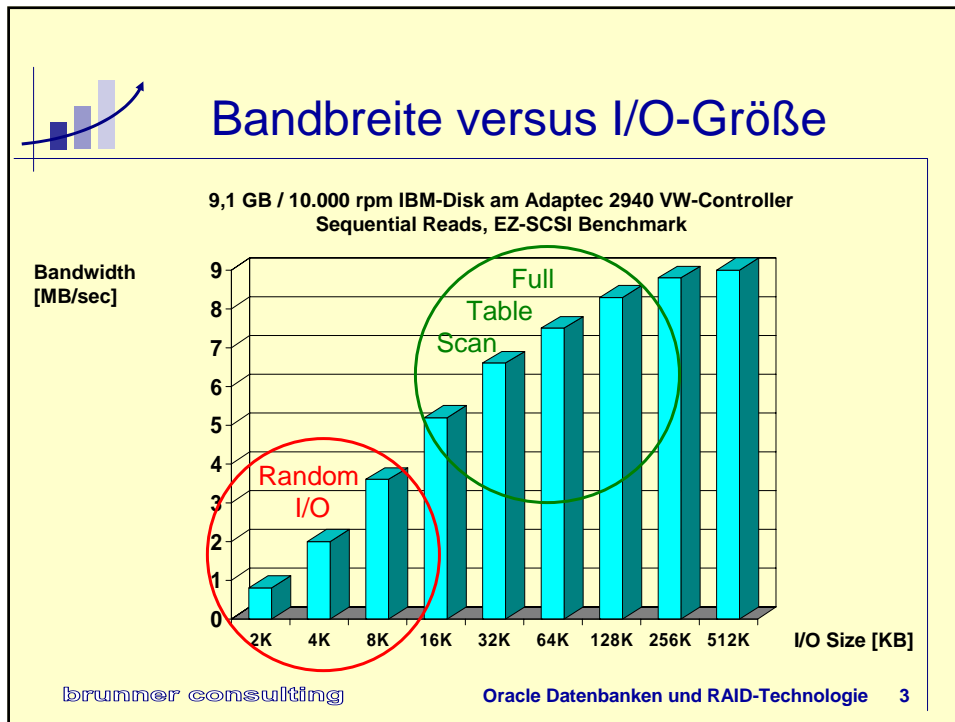


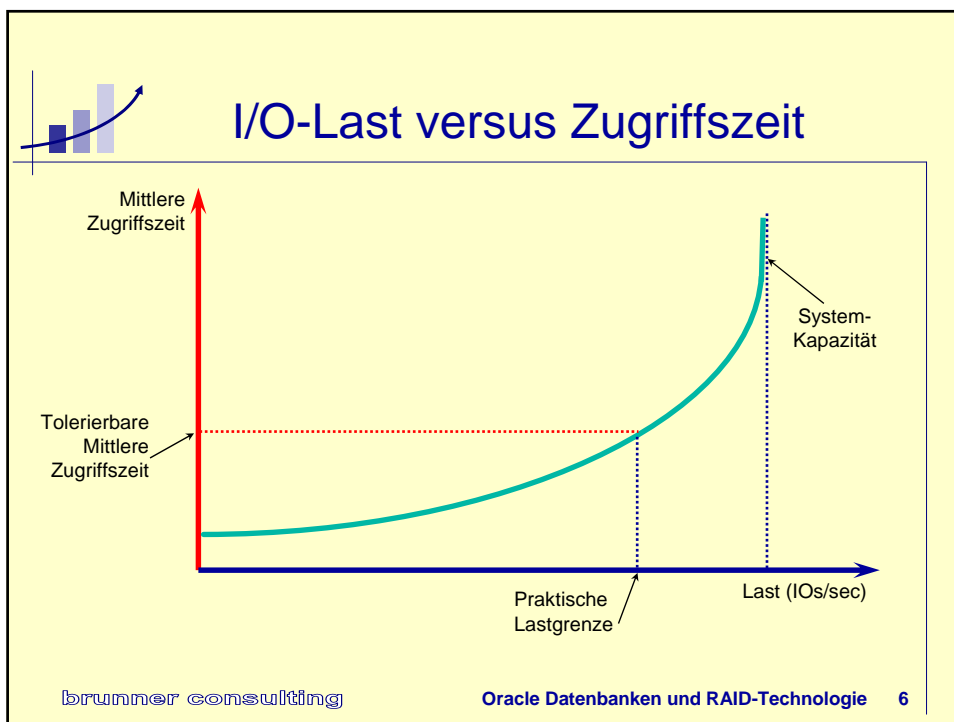
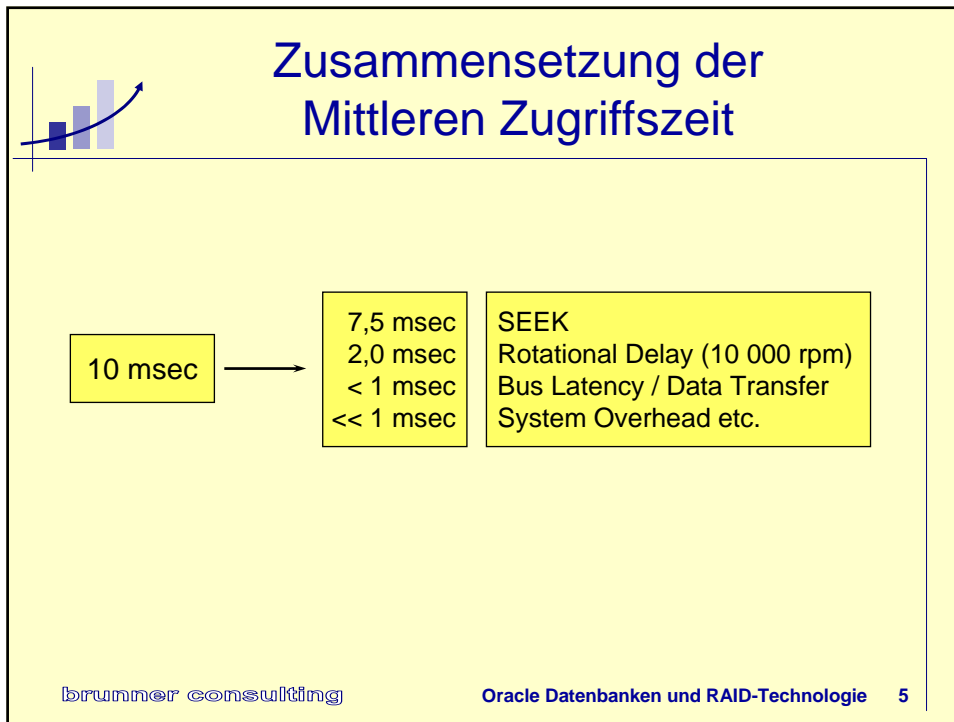
Wozu überhaupt über I/O Performance nachdenken?



| Jahr | Memory Größe (10 000x-1 000 000x) | CPU Geschwindigkeit (500x - 5000x) | I/O Geschwindigkeit (ca. 5x - 50x) |
|------|-----------------------------------|------------------------------------|------------------------------------|
| 1978 | 1 | 1 | 1 |
| 1980 | 10 | 10 | 10 |
| 1982 | 100 | 100 | 100 |
| 1984 | 1 000 | 1 000 | 1 000 |
| 1986 | 10 000 | 10 000 | 10 000 |
| 1988 | 100 000 | 100 000 | 100 000 |
| 1990 | 1 000 000 | 1 000 000 | 1 000 000 |
| 1992 | 10 000 000 | 10 000 000 | 10 000 000 |
| 1994 | 100 000 000 | 100 000 000 | 100 000 000 |
| 1996 | 1 000 000 000 | 1 000 000 000 | 1 000 000 000 |
| 1998 | 10 000 000 000 | 10 000 000 000 | 10 000 000 000 |
| 2000 | 100 000 000 000 | 100 000 000 000 | 100 000 000 000 |
| 2002 | 1 000 000 000 000 | 1 000 000 000 000 | 1 000 000 000 000 |
| 2004 | 10 000 000 000 000 | 10 000 000 000 000 | 10 000 000 000 000 |

brunner consulting Oracle Datenbanken und RAID-Technologie 2





Impact of Read and Write Caches

| | XACTION I/O (Small) | | Large File I/O | |
|----------------|---------------------|-----------|----------------|-----------|
| | Read | Write | Read | Write |
| JBOD | OK | Sehr gut | OK | Gut |
| RAID 0 (Strip) | Sehr gut | Exzellent | Gut | Sehr gut |
| RAID 1 (Shad) | Gut | Sehr gut | OK | Gut |
| RAID 0+1 | Exzellent | Exzellent | Gut | Sehr gut |
| RAID 3 | Schlecht | Gut | Sehr gut | Exzellent |
| RAID 4 | Sehr gut | Gut | Sehr gut | Gut |
| RAID 5 | Sehr gut | Sehr gut | OK | Gut |
| RAID 6 | Exzellent | Sehr gut | Sehr gut | Gut |

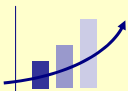
brunner consulting
Oracle Datenbanken und RAID-Technologie 7

Schlußfolgerungen

Typischerweise wird die Zugriffszeit lange, bevor das Speichersystem die maximale Belastung erreicht hat, inakzeptabel.

Normalerweise verdoppelt sich die Mittlere Zugriffszeit bereits bei ca. 50% der maximalen Last.

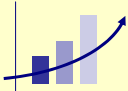
brunner consulting
Oracle Datenbanken und RAID-Technologie 8



Oracle

brunner consulting

Oracle Datenbanken und RAID-Technologie 9



Oracle-Instance

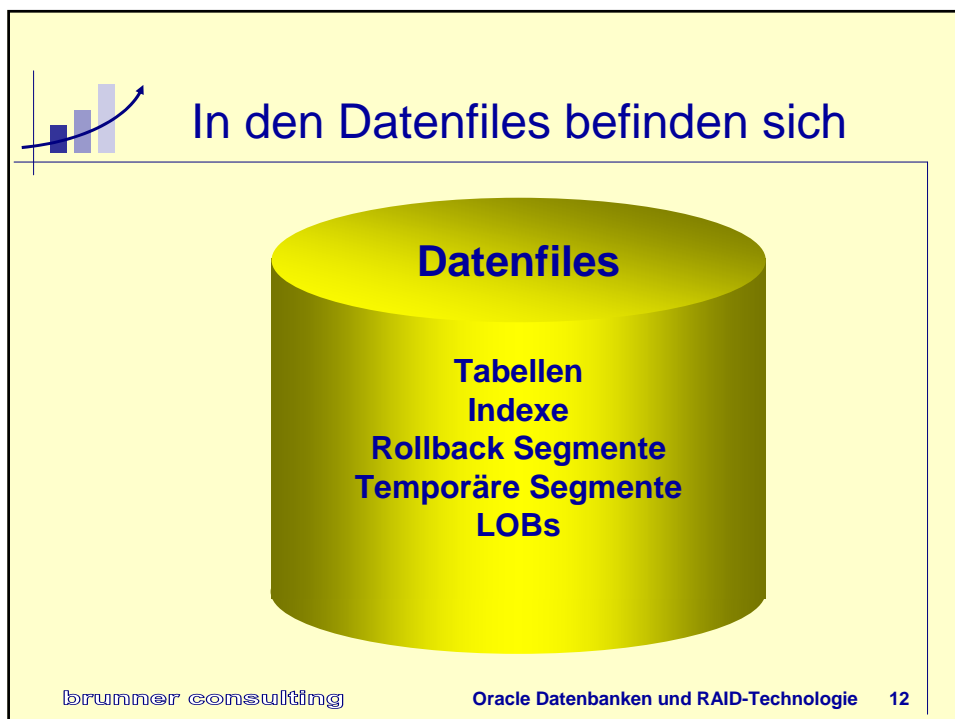
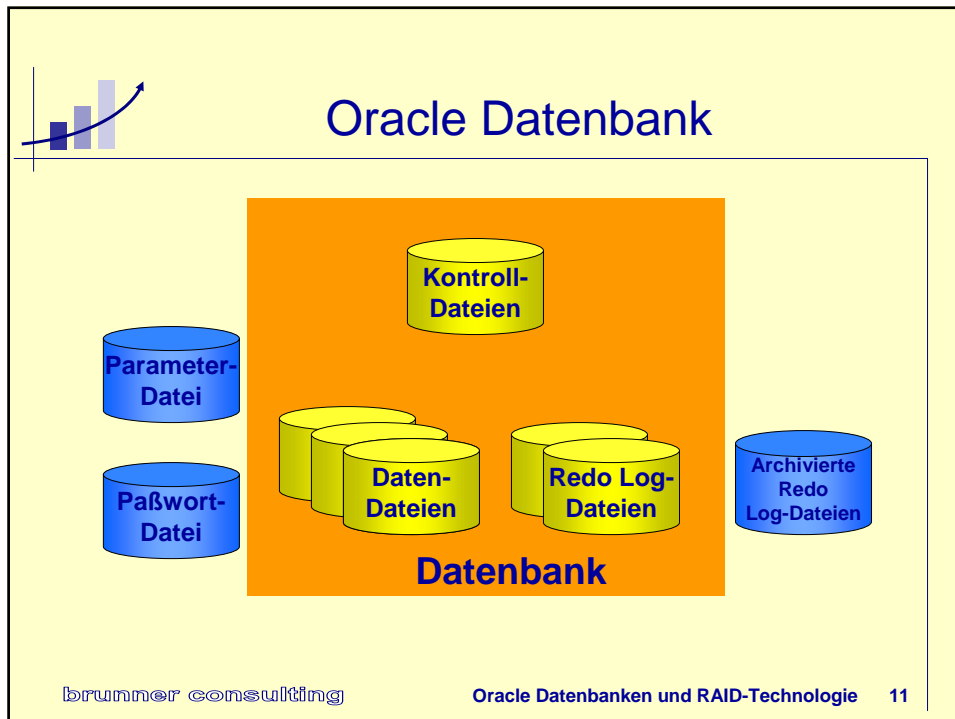
Instance

| | | |
|-----------------|---------------------------------------|-----------------|
| DB Buffer Cache | LRU-Liste | REDO Log Buffer |
| | Dirty-Liste | |
| | Shared Pool | |
| | LIBR Cache Data Dictionary Cache | |

DBWR SMON PMON CKPT LGWR ARCH

brunner consulting

Oracle Datenbanken und RAID-Technologie 10





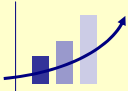
Inhalt der Kontrolldatei

Kontrolldatei

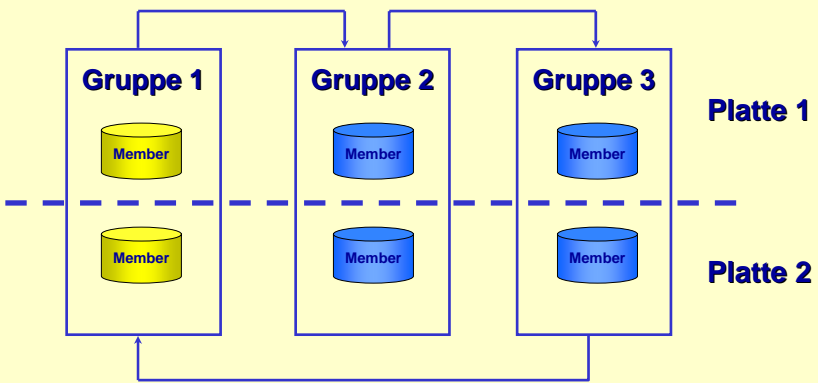
- Datenbankname
- Speicherorte der Datendateien
- Speicherorte der Redo Log-Dateien
- Tablespace-Namen
- Aktuelle Log Sequence-Nummer
- Checkpoint-Informationen
- Log-History
- Backup-Informationen

**Kann mit ORACLE-Mitteln gespiegelt werden
→ erhöht Ausfallsicherheit**

brunner consulting Oracle Datenbanken und RAID-Technologie 13



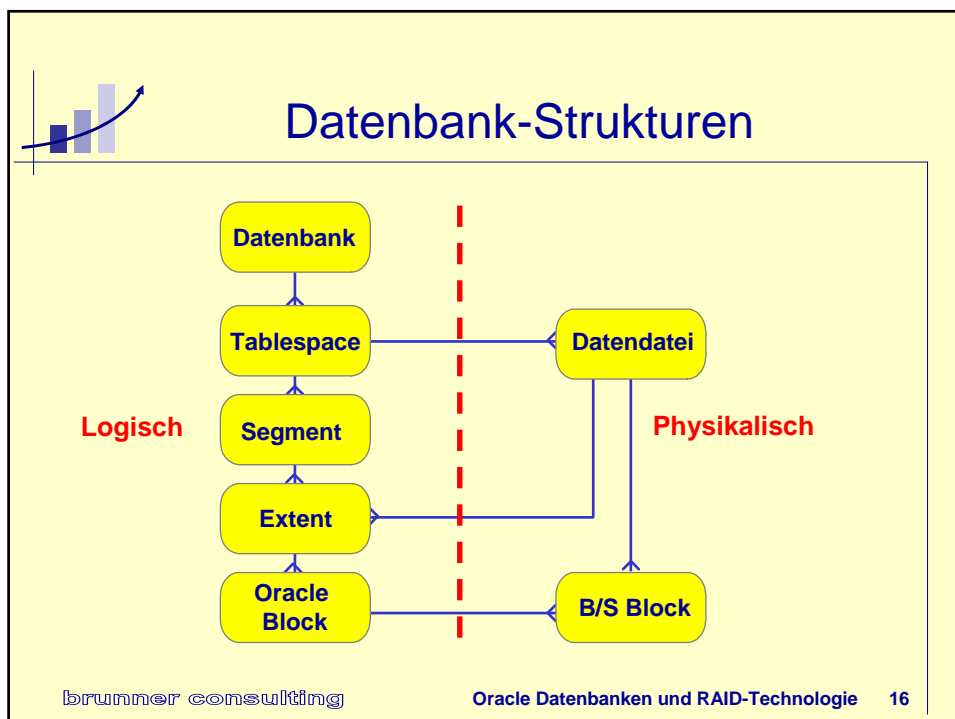
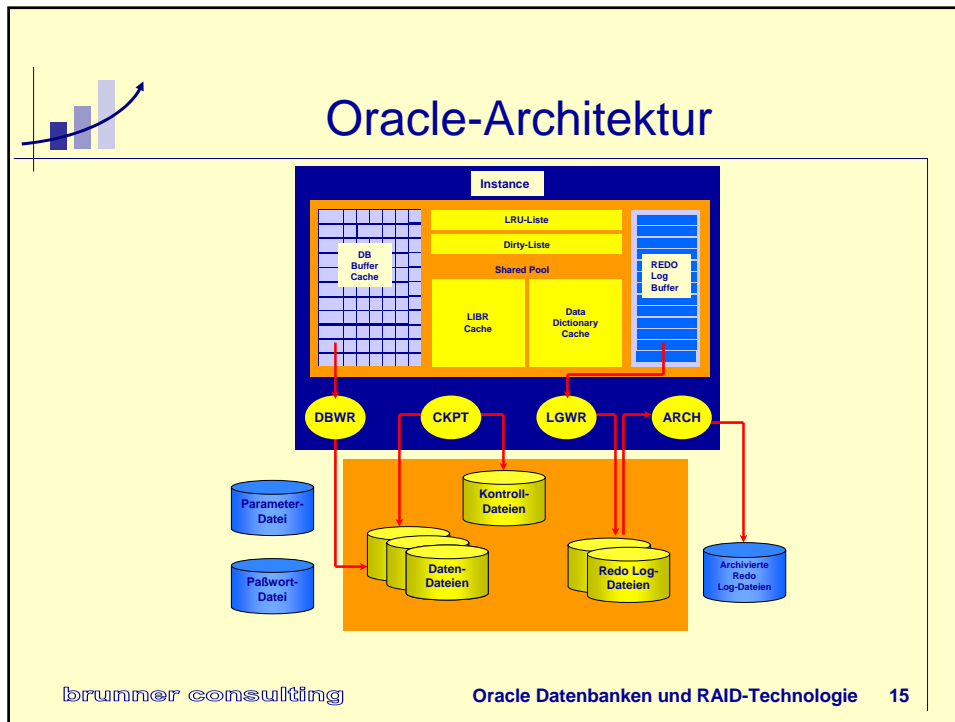
Redo Log-Gruppen und Member

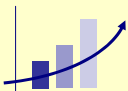


The diagram illustrates three Redo Log Groups (Gruppe 1, Gruppe 2, Gruppe 3) distributed across two disks (Platte 1 and Platte 2). Each group contains two members. Gruppe 1 has two active members (yellow cylinders) on Platte 1 and Platte 2. Gruppe 2 has one active member (yellow cylinder) on Platte 1 and one inactive member (blue cylinder) on Platte 2. Gruppe 3 has one active member (yellow cylinder) on Platte 1 and one inactive member (blue cylinder) on Platte 2. Arrows indicate the flow of data between the groups and across the disks. A legend at the bottom indicates that a yellow cylinder represents an active member and a blue cylinder represents an inactive member.

■ = Aktiv ■ = Inaktiv

brunner consulting Oracle Datenbanken und RAID-Technologie 14

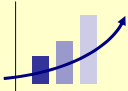




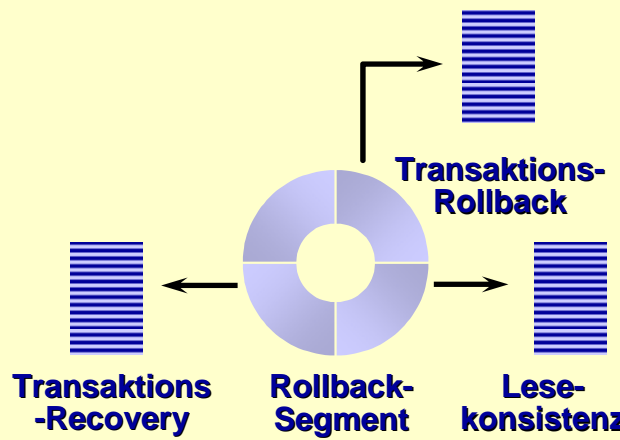
Tablespaces anlegen

```
CREATE TABLESPACE apps_data
DATAFILE '/DISK1/apps01.dbf' SIZE 1000M,
         '/DISK2/apps02.dbf' SIZE 1000M,
         '/DISK3/apps03.dbf' SIZE 1000M,
MINIMUM EXTENT 500K
DEFAULT STORAGE (INITIAL 500K NEXT 500K
MINEXTENTS 3 MAXEXTENTS 300 PCTINCREASE 0);
```

brunner consulting Oracle Datenbanken und RAID-Technologie 17

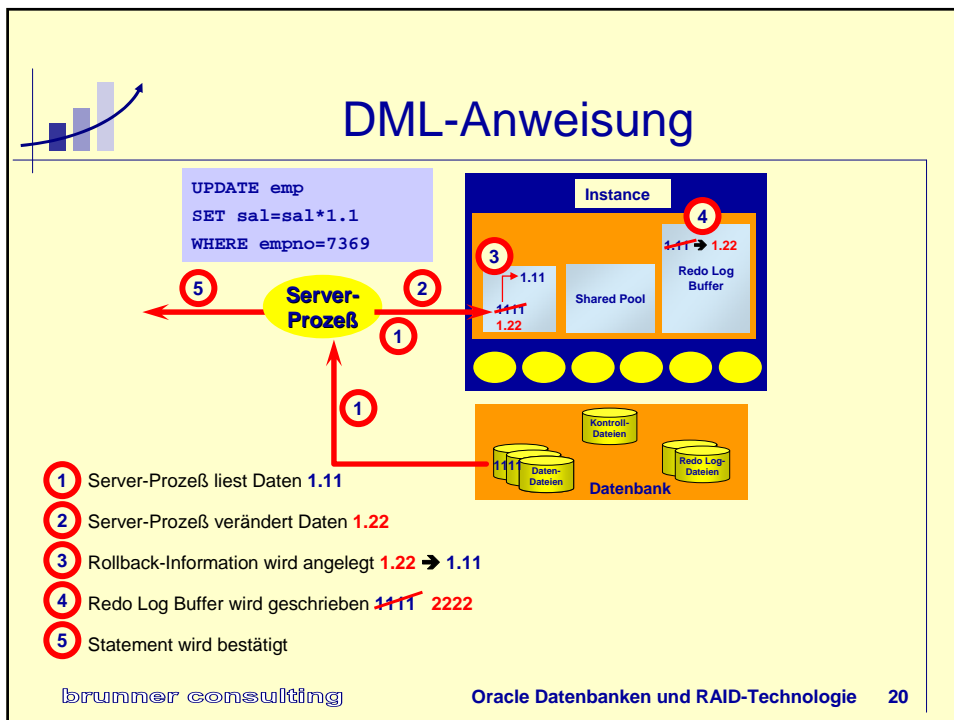
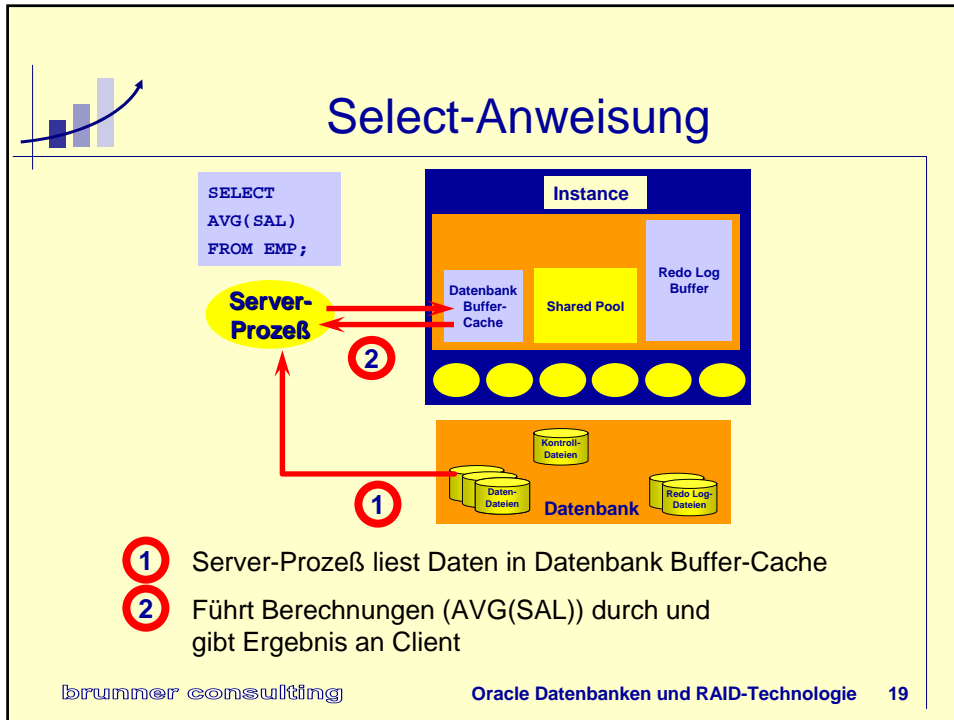


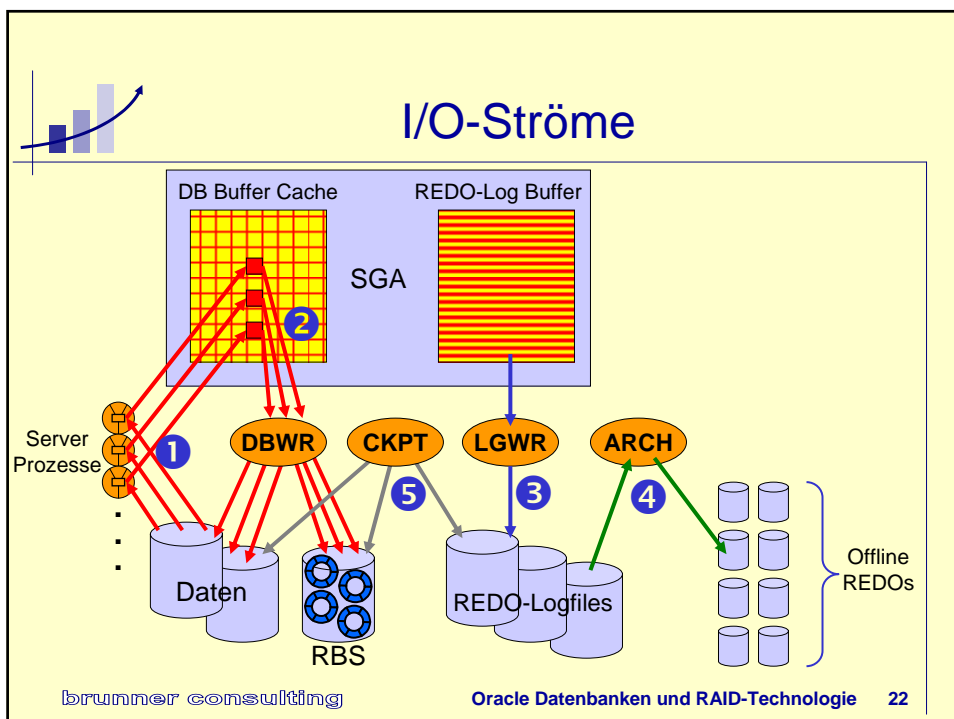
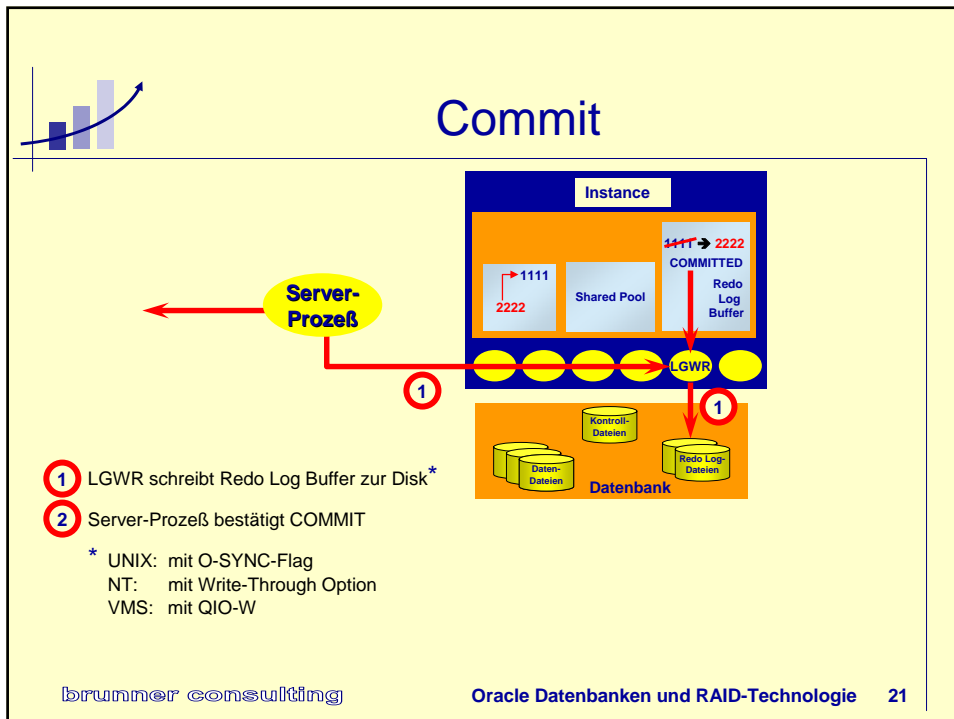
Rollback-Segmente: Aufgabe

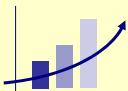


The diagram illustrates the tasks of a Rollback Segment. A central circle labeled "Rollback-Segment" is divided into four quadrants. Three arrows point from this central circle to three separate stacks of horizontal lines representing data blocks. The arrows are labeled "Transaktions-Recovery" (pointing left), "Lese-konsistenz" (pointing right), and "Transaktions-Rollback" (pointing up and right).

brunner consulting Oracle Datenbanken und RAID-Technologie 18



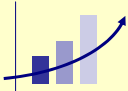




Es entstehen folgende I/O-Ströme

- 1 Viele parallele Server-Prozesse
Lesen aus den Datafiles
(**Zeitkritisch** → geringere Response Time gefragt)
- 2 DBWR - schreibt nahezu kontinuierlich in Datafiles
(Daten + RBS-Segmente → **nicht zeitkritisch**)
- 3 LGWR - schreibt spätestens bei jedem Commit in die REDO-
Logfile → **SEHR zeitkritisch**
- 4 ARCH - liest REDO-Log, schreibt archivierte Offline-REDOs →
nicht zeitkritisch, aber minimaler Durchsatz muß erfüllt sein
- 5 CKPT - schreibt gelegentlich zu CTL-Files (→ **nicht zeitkritisch**)

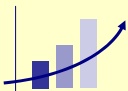
brunner consulting Oracle Datenbanken und RAID-Technologie 23



Database Writer

- Schreibt ASYNCHRON, wenn
 - ⇒ Anzahl Dirty Buffers zu hoch
 - ⇒ Zu wenig Platz in Buffer-Cache
 - ⇒ Zeitintervall abgelaufen (3 sec)
 - ⇒ Logwriter löst CHECKPOINT aus
(z.B. wegen Logfile-Switch)

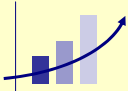
brunner consulting Oracle Datenbanken und RAID-Technologie 24



Optimale Konfiguration (1)

- ① Datafiles
 - „normale Disks“ → O.K.
 - Mirrorsets → sicher, teuer
 - RAID 5 → sicher, nicht ganz so teuer
 - Stripe Sets → optimale Performance, erspart Striping auf Tablespace-Ebene

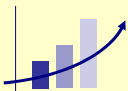
brunner consulting Oracle Datenbanken und RAID-Technologie 25



Optimale Konfiguration (2)

- ② Redo-Logs
 - „normale Disks“ → O.K., wenn mit Oracle-Mitteln gepiegelt
 - Write Back-Cache → Erhöht COMMIT-Geschwindigkeit, aber nur sicher mit **sehr guter** Hardware
 - Mirror-/Shadow-Sets → für höchste Ausfallsicherheit

brunner consulting Oracle Datenbanken und RAID-Technologie 26



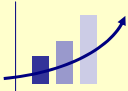
Optimale Konfiguration (3)

3 Alle anderen Files
(PWD-File, Ctl-File, INITxxxx.ORA)

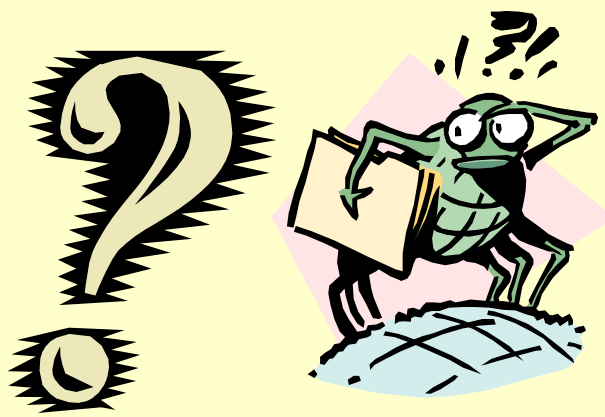
Prinzip:
Am wenigsten Ärger gibt es mit Spiegelung.

(Da diese zentralen Dateien enorm wichtig für das Überleben der DB sind, sollte es auch keine Kostenfrage sein...)

brunner consulting Oracle Datenbanken und RAID-Technologie 27



Fragen?



brunner consulting Oracle Datenbanken und RAID-Technologie 28