Grundlagen der PostgreSQL Administration

Timo Pagel

Agenda

- Installation/Konfiguration
- Backup/Restore
- Hochverfügbarkeit
- Überwachung
- Ausführungspläne

Agenda

- Installation/Konfiguration
- Backup/Restore
- Hochverfügbarkeit
- Überwachung
- Ausführungspläne

Eigenschaften PostgreSQL 1/2

- „Object Related Data Base Management System“
- ACID Compliant
  - Atomarität, Konsistenz, Isolation, Dauerhaftigkeit
- SQL Standardkonform ANSI-SQL:2008
- Stabilität und Codequalität haben höchste Priorität
- BSD-artige Lizenz

Eigenschaften PostgreSQL 2/2

- Write Ahead Log (pg_xlog) zur Wiederherstellung und zur Replikation sowie für Hot Backup (später mehr)
- Query Planner/Optimizer sind ausgereift (später mehr)
- Multi-Version Concurrency Control (MVCC)
  - Snapshot für Transaktionen

Konfiguration - Installation

- Quellen
  - Selbst kompilieren
  - Pakete
    - Ubuntu
    - Debian
    - RedHat

Konfiguration – Ordnerstruktur 1/2

- /etc/postgresql/9.1/main
  - Basiskonfiguration
    - Authentifizierung in pg_hba.conf
    - Server-IP-Einstellungen in postgresql.conf
    - … alles was konfigurierbar ist

Konfiguration – Ordnerstruktur 2/2

- /var/lib/postgresql/9.1/main
  - Sich verändernde Dateien
    - Tablespace
    - WAL (später mehr)
    - … alles was sich verändert
- /var/log/postgresql/postgresql-9.1-main.log
  - Log-Datei mit Status-Informationen

Tools

- psql
  - Kommandozeilenclient
    - ~/.psqlrc
  - Backslash-Befehle
    - Hilfe: \?
- PgAdmin
  - GUI

Agenda

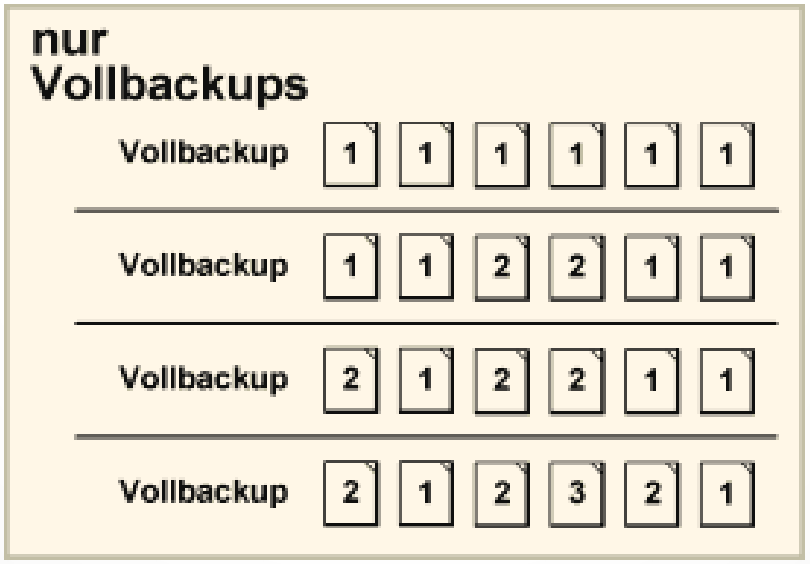
- Installation/Konfiguration
- Backup/Restore
- Hochverfügbarkeit
- Überwachung
- Ausführungspläne

Backup – Arten 1/4

- Vollbackup
  - Sicherung des gesamten Datenbestandes
    - Basis für Inkrementell und Differenziell
- Inkrementelles Backup
  - Benötigt Vollbackup
  - Schrittweise Sicherung aller Änderungen
- Differenzielles Backup
  - Speichert alle Änderungen zum letzten Vollbackup
  - Vorteil: Einspielen schneller als bei inkrementellem Backup

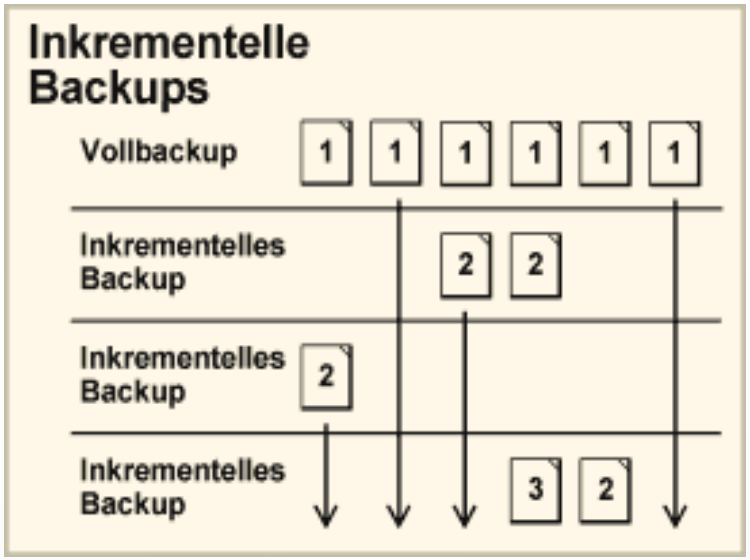
Backup – Arten 2/4

- Vollbackup
  - Sämtliche Daten zu einem bestimmten Sicherungspunkt werden archiviert



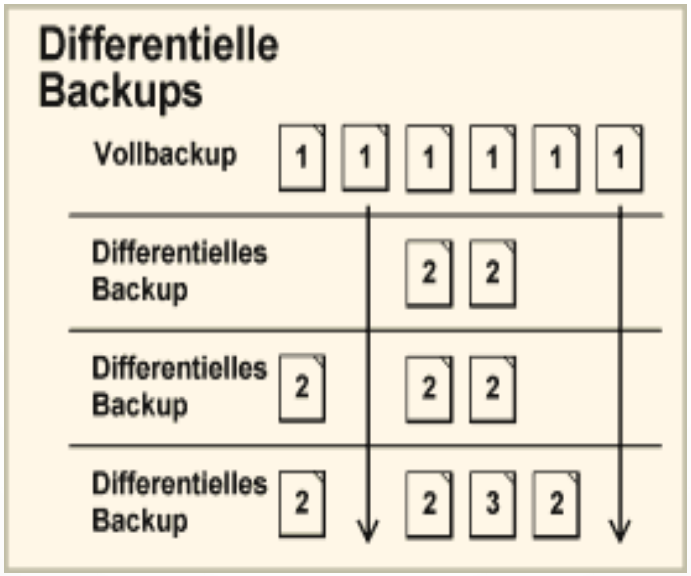
Backup – Arten 3/4

- Inkrementelles Backup
  - Nur die Daten sichern, die sich seit dem letzten Vollbackup geändert haben



Backup – Arten 4/4

- Differentielles Backup
  - Es werden alle Änderungen zum letzten Vollbackup gesichert



Backup – Betriebssystem Tools

- Archivierung:
  - tar
  - zip
- Komprimierung:
  - gzip

Backup – Offline Archivierung 1/2

- Ergebnis: 1:1 Kopie (Binärdateien)
  - Größe: Kompakt
- Geschwindigkeit/Performance: Schnell/Hoch
- Inhalt:
  - Binärdateien → Geringe Portabilität
  - Optional: Log und Konfigurationsdateien

Backup – Offline Archivierung 2/2

- Erreichbarkeit: Server ist während des Backups nicht verfügbar
- Unterstützte Backuparten:
  - Vollsicherung
  - Inkrementelle Sicherung
  - Differentielle Sicherung

Restore – zur Offline Archivierung

1) DB stoppen
2) Vorhandene Verzeichnisse löschen
3) Archiv einspielen
4) Server starten

Backup – Online Archivierung 1/2

- Voraussetzung WAL (siehe später)
  - Start: SELECT pg_start_backup('label');
  - Stop: SELECT pg_stop_backup();
- Ergebnis: 1:1 Kopie (Binärdateien)
  - Größe: Kompakt
- Geschwindigkeit/Performance: Schnell/Hoch

Backup – Online Archivierung 2/2

- Inhalt:
  - Binärdateien → Geringe Portabilität
  - Optional: Log und Konfigurationsdateien
- Erreichbarkeit: Server ist während des Backups verfügbar für Lesezugriffe
- Unterstützte Backuparten:
  - Vollsicherung
  - Inkrementelle Sicherung
  - Differentielle Sicherung

Restore – zur Online Archivierung

1) DB stoppen
2) Vorhandene Verzeichnisse löschen
3) Archiv einspielen
4) Server starten

Backup – Online

- Ergebnis: Dump
  - Größe: Größer als beim Offline Backup
- Geschwindigkeit/Performance: Langsam/Niedrig
- Inhalt:
  - Struktur und Inhalt als SQL-Text → hohe Portabilität
- Erreichbarkeit:
  - Server ist durchgehend erreichbar
    → Lock-Mechanismus erforderlich

Online Backup - Tools

- Dump
  - pg_dump
    - Erzeugt Backup einer Datenbank als Textdatei
  - pg_dumpall für globals
    - Erzeugt Backup aller Datenbanken inkl. Rollen und Tablespaces als Textdatei

Restore - Online Backup

- pg_restore
- psql –set ON_ERROR_STOP=on dbname < dump.txt.sql
  - Gleichwertig, da im textuellen Backup alle notwendigen SQL-Statements enthalten sind

Write Ahead Logging - Grundlagen

- „Fortlaufende“ Datei auf der Festplatte mit allen Änderungen
- Ermöglicht Vorwärts-Wiederherstellung
  - Auch „“REDO“
- 16 MB Segmente
- Timeout konfigurierbar

WAL - Einsatzmöglichkeiten

- Inkrementelles Backup (gleich)
- Hot Standby (später mehr)
- Master/Slave-Systeme
- Intern beim Online Backup

WAL - Backup

1) Vollbackup erstellen
2) Konfiguration ändern
   - archive_mode = on
   - archive_timeout = 60
   - archive_command = 'test ! -f /data/wal_archive/%f && cp %p /data/wal_archive/%f'
3) Kopieren der WAL auf einen Sicherungsserver
   - Konfiguration: archive_command = 'rsync --delay-updates --whole-file -ar -e ssh %p postgres@<Server-IP>:/data/wal_archive/%f </dev/null'

WAL – Restore 1/2

1) DB stoppen
2) Vorhandene Verzeichnisse löschen
3) Vollbackup-Archiv einspielen
4) Wenn pg_xlog-Verzeichnis gesichert wurde
   → Inhalt löschen
5) Kopieren aller WAL-Segmente aus Backupordner in pg_xlog

WAL – Restore 2/2

6) Ggf. neue Benutzer-Verbindungen unterbinden

7) /var/lib/postgresql/9.1/main/recovery.conf

- restore_command = 'cp /mnt/irgendwo/%f %p'

8) Server starten

- DB bleibt im Wiederherstellungsmodus
- Am Ende recovery.conf → recovery.done

9) DBMS hat den gewünschten Zustand wiederhergestellt

- ggf. Benutzerverbindungen wieder aktivieren

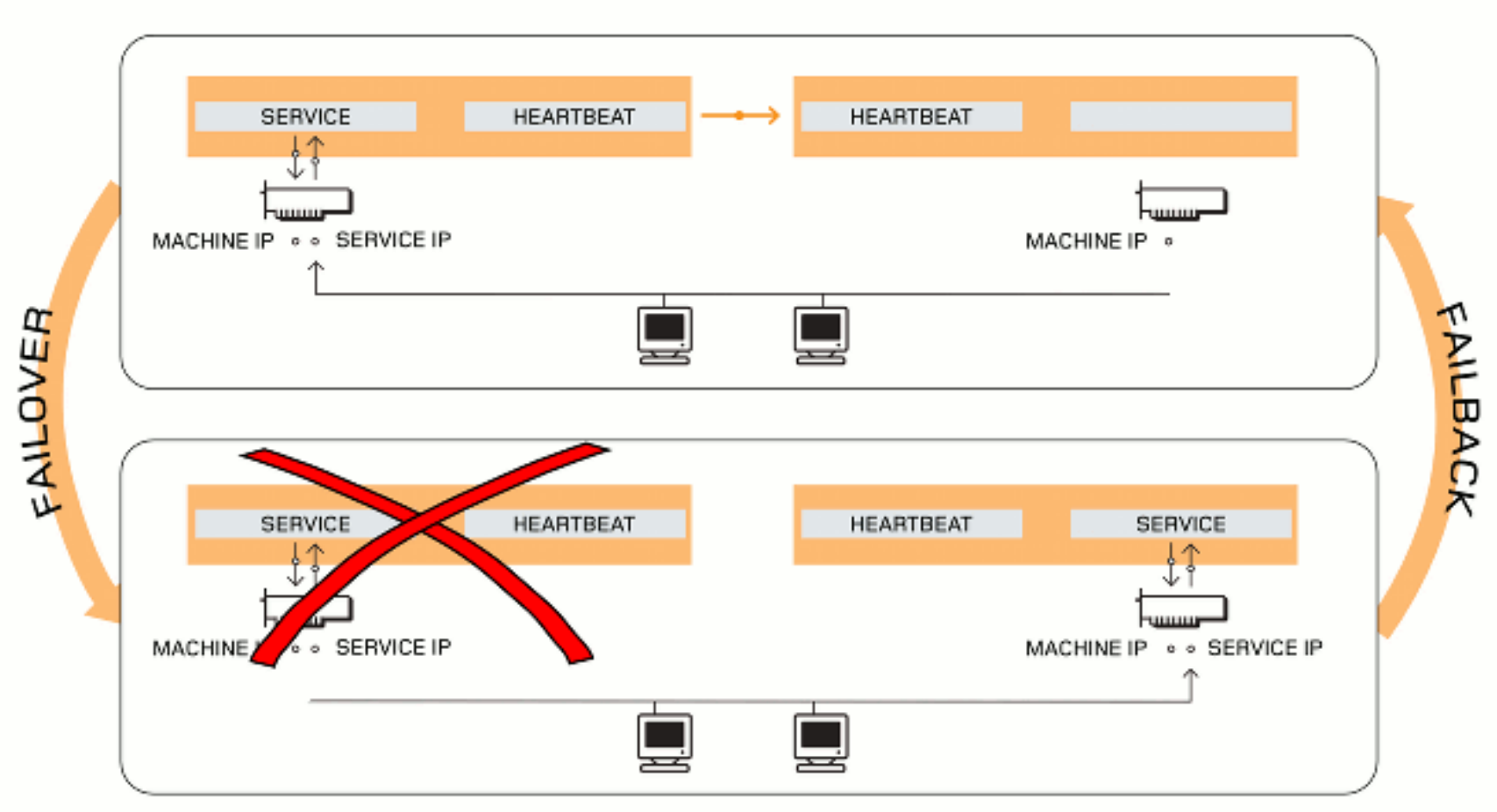
WAL – Point in Time Recovery Klienten-Konfiguration

- recovery.conf
  - restore_command = 'cp /mnt/irgendwo/%f %p'
  - recovery_target_time = '2010-08-25 21:52+02'
  - recovery_target_inclusive = false

Agenda

- Installation/Konfiguration
- Backup/Restore
- Hochverfügbarkeit
- Überwachung
- Ausführungspläne

Hochverfügbarkeit mit Heartbeat



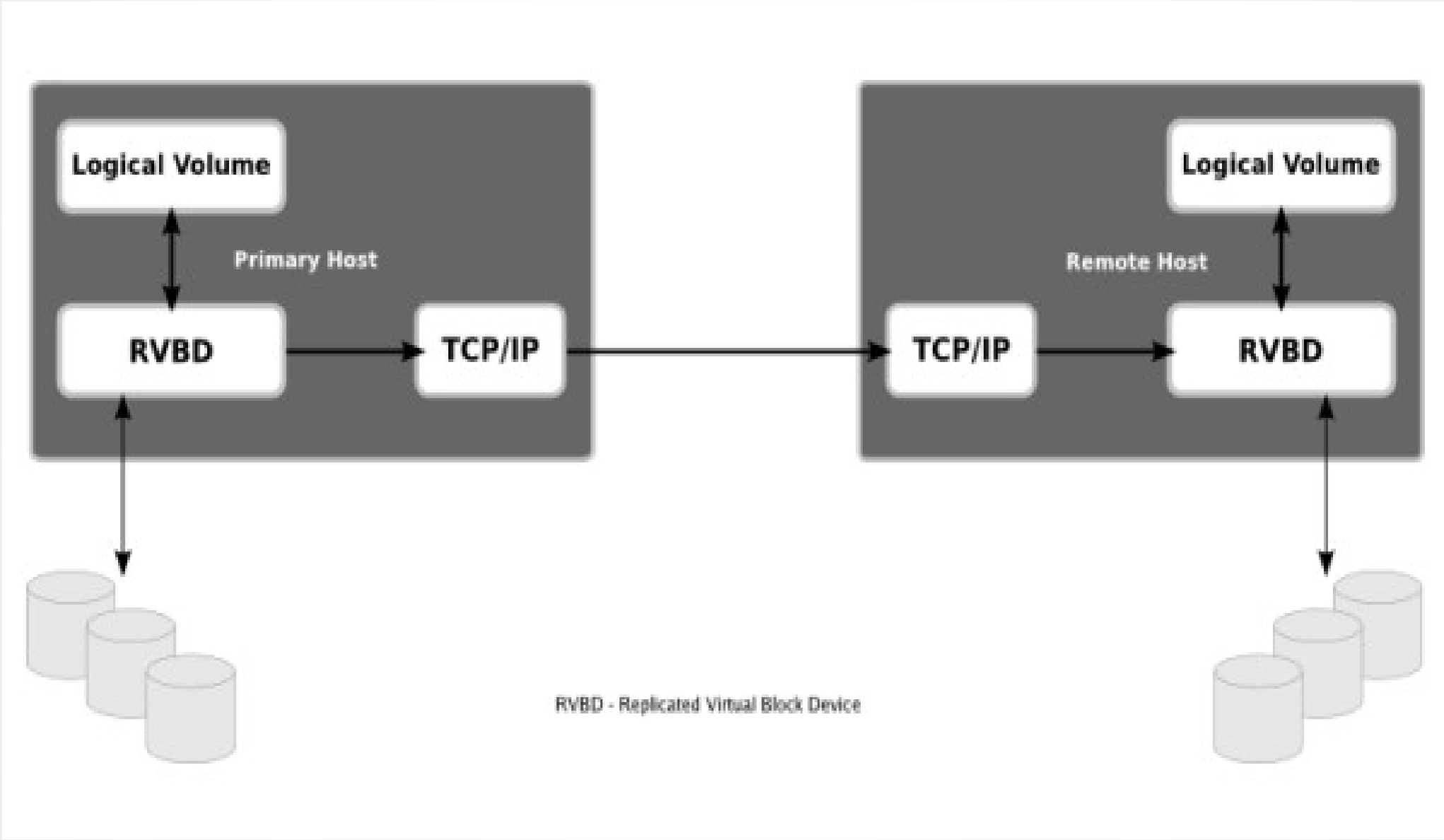
Hochverfügbarkeit - Überblick

- **Shared Disk Failover**
- **File System (Block-Device) Replication**
- **Warm Standby Using Point-In-Time Recovery (PITR)**
- Master-Slave Replication
  - Slave sendet Schreibanfragen an den Master
  - Slave kann Leseanfragen beantworten
- Synchronous Multimaster Replication
  - Schreibperformance schlecht

File System (Block-Device) Replication - Überblick

- Software
  - Kommerziell: Distributed Replicated Block Device
  - OpenSource: openfiler
- Master für Schreibzugriffe
- Slave(s) für Lesezugriffe

File System (Block-Device) Replication - Prinzip



Warm Standby Using Point-In-Time Recovery (PITR)

- Asynchron (ggf. wurden die letzten Änderungen nicht auf den Standby-Server übernommen)
- Nachteil: Es wird nur ein Server aktiv genutzt
- recovery.conf auf Standby Seite
  - standby_mode = 'on'
  - primary_conninfo = 'host=192.168.1.50 port=5432 user=foo password=foopass'
  - restore_command = 'cp /path/to/archive/%f %p'
  - archive_cleanup_command = 'pg_archivecleanup /path/to/archive %r'

Agenda

- Installation/Konfiguration
- Backup/Restore
- Hochverfügbarkeit
- Überwachung
- Ausführungspläne

Monitoring - „Überwachungsserver“

- Systemüberwachung (gleich mehr)
  - Plattenauslastung
  - CPU Verbrauch
  - RAM Verbrauch
  - ...
- Datenbanküberwachung
  - Nagios check_postgres
    - Locks (später mehr)
    - Connections
    - …

Monitoring - Tools

- Systemtools
  - Prozessausgabe (ps)

```
~$ ps -ef  | grep postgres
postgres  1438     1  0 16:09 ?        00:00:00 /usr/lib/postgresql/9.1/bin/postgres [...]
postgres  1443  1438  0 16:09 ?        00:00:00 postgres: writer process
postgres  1444  1438  0 16:09 ?        00:00:00 postgres: wal writer process
postgres  1445  1438  0 16:09 ?        00:00:00 postgres: autovacuum launcher process
postgres  1446  1438  0 16:09 ?        00:00:00 postgres: stats collector process
```

  - Prozessübersicht (top)

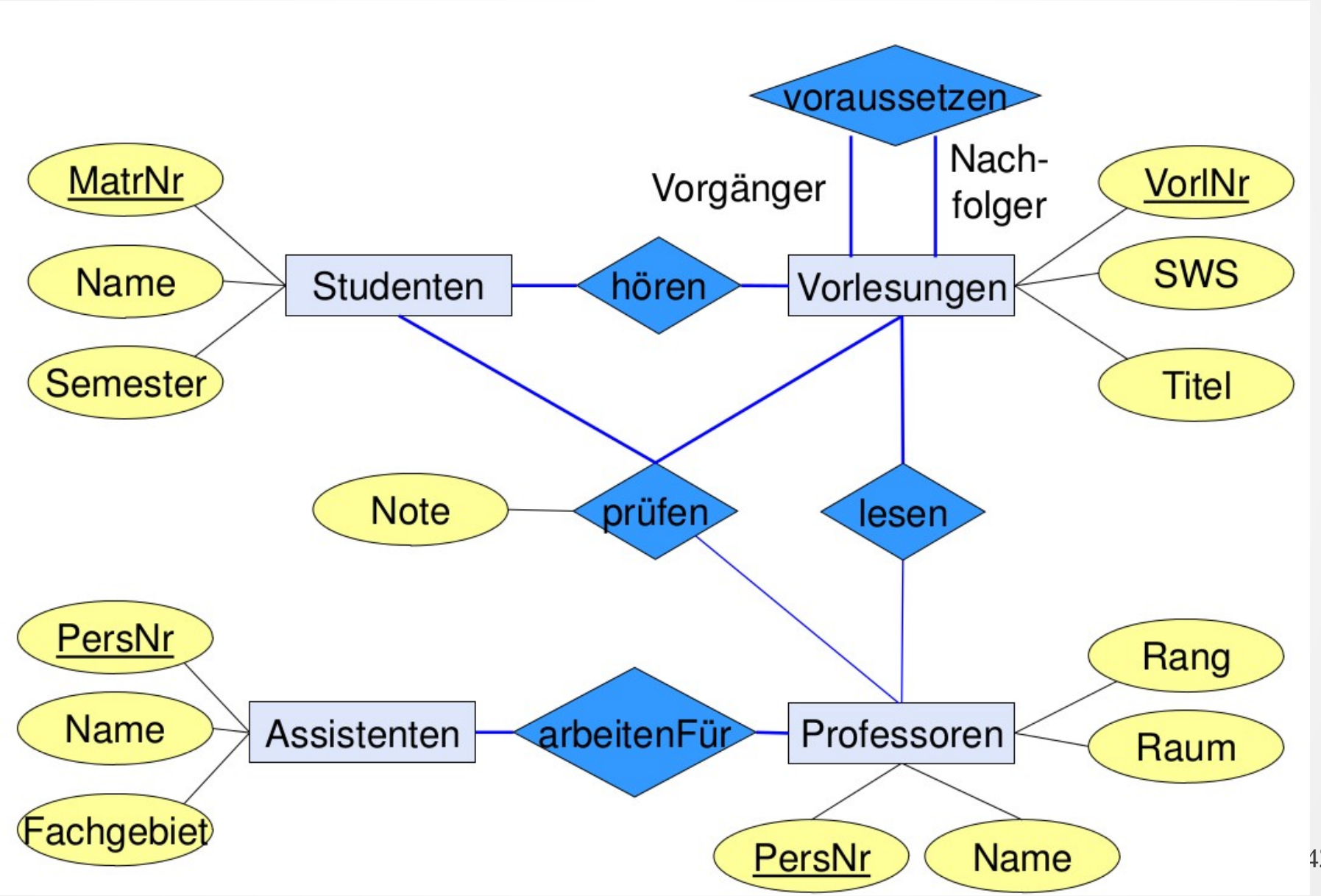
Monitoring - Statische Views

- pg_stat_activity
  - Darstellung aller Serverprozesse
- pg_statio_all_tables
  - Zugriffsstatistik
- pg_locks
  - Lock-Übersicht
- ...

Agenda

- Installation/Konfiguration
- Backup/Restore
- Hochverfügbarkeit
- Überwachung
- Ausführungspläne

Ausführungspläne
Überblick Schema



Ausführungspläne

- Befehl „Explain <Query>“ zeigt Struktur des Ausführungsplan
- Zugriffsarten auf Tabellen
  - seq scan, index scan oder bitmap index scan
- Operationen
  - Joins
  - Sortierungen
  - Aggregierungen

Ausführungspläne

- EXPLAIN [ANALYZE] [VERBOSE] statement
  - Ohne Analyse/Verbose:
    - Zeigt Algorithmus-Ausführungsschätzung an
  - ANALYSE: Zeigt tatsächliche Schleifendurchgänge und konkrete Ressourcenbelegung
  - VERBOSE: Zeigt die gesamte Ausführung

Ausführungspläne – Studenten

- Query:

  EXPLAIN

  select Name, MatrNr from studenten;

- Ausgabe:

```
                          QUERY PLAN
-------------------------------------------------------------
 Seq Scan on studenten  (cost=0.00..16.80 rows=680 width=86)
(1 row)
```

- Seq Scan
  - Zugriffsart
- cost
  - Startkosten (z.B. Sortierung)..Totalkosten
  - Maßeinheit: Plattenzugriffe
- rows
  - Geschätzte Ergebnisdatensätze
- width
  - Durchschnittliche Größe eines Datensatzes in Byte

- Ausgabe:

```
                          QUERY PLAN
------------------------------------------------------------
 Seq Scan on studenten  (cost=0.00..16.80 rows=680 width=86)
(1 row)
```

```
                                  QUERY PLAN
------------------------------------------------------------------------------
 Nested Loop  (cost=18.54..55.41 rows=8 width=86)
   ->  Hash Join  (cost=18.54..52.95 rows=8 width=4)
         Hash Cond: ("prüfen".vorlnr = vorlesungen.vorlnr)
         ->  Seq Scan on "prüfen"  (cost=0.00..27.70 rows=1770 width=8)
         ->  Hash  (cost=18.50..18.50 rows=3 width=4)
               ->  Seq Scan on vorlesungen  (cost=0.00..18.50 rows=3 width=4)
                     Filter: ((titel)::text = 'Informatik II'::text)
   ->  Index Scan using studenten_pkey on studenten  (cost=0.00..0.29 rows=1 width=86)
         Index Cond: (matrnr = "prüfen".matrnr)
(9 rows)
```

EXPLAIN

SELECT Name, MatrNr

FROM (Vorlesungen NATURAL JOIN prüfen NATURAL JOIN Studenten)

WHERE Titel = 'Informatik II';

Fragen?

- Installation/Konfiguration
- Backup/Restore
- Hochverfügbarkeit
- Überwachung
- Ausführungspläne

Dokumentation

- http://www.postgresql.org/docs
- http://www.net.co.at/doc/howto/docs/shell_script_entwicklung/docs/backup.html
- http://tuning.postgresql.de/postgresql_explain

Backup - Nagios