

PostgreSQL
replizieren

Jetzt erst Recht!



Wednesday, October 13, 2010

Michael Renner

Wednesday, October 13, 2010

Mein Name ist...

Vergangenheit Sysadmin, Scrummaster, Projectmanagement

Interesse als DBA & Daten

Mitglied der DE & EU PGUG

Werde euch Neuerungen zum Thema Replikation in PG 9.0 erzählen

Schnelle Fragerunde

Wednesday, October 13, 2010

Wer verwendet schon PostgreSQL?

Mit Replikation?

Möchte jemand ein MySQL Setup ablösen?

Weil wir schon bei Fragen sind...: Gleich stellen oder notieren

Postgre... was?

Wednesday, October 13, 2010

Robuste SQL Datenbank, seit 1995, (Select, Insert, Update, Views, Triggers, etc.)

Standardsorientiert

Community-Driven, keine einzelne Entität dahinter

Unter einer BSD Lizenz

Replikation heisst...

Wednesday, October 13, 2010

(Daten)Replikation Definition:

Mehrfache Speicherung von selben Daten, verschiedene Standorte, inklusive Synchronisation

Im Vortrag:

Klassische Replikation, asynchron Master -> Slave

Master RW, Slave RO



*You have to know the past
to understand the present*

~ Dr. Carl Sagan

Wednesday, October 13, 2010

Replikation bei PG kein neues Thema
Varianten gibts seit über 6 Jahren

Da Replikation bei DBs umfangreiches Gebiet,
hilft einen Überblick der Möglichkeiten zu haben

Eine der ältesten...

Trigger & Statements

Wednesday, October 13, 2010

Einer der ältesten Varianten

Einzelne SQL-Abfragen analysiert und verteilt

Statements werden repliziert, entweder Trigger oder Middleware

Bekannte Vertreter:

Slony (2004)

PG-Pool (mindestens 2004)

Sehr flexibel

Erhöhter Wartungsaufwand

Schema-Änderungen deutlich komplizierter

Log shipping

Wednesday, October 13, 2010

Seit 8.2 (2006) Wenn man von Logs spricht...

Write Ahead Log od. WAL

Postgres sammelt Änderungen an Tables od. Indizes

Werden für Crash Recovery oder Live-Backup benötigt

Perfekte Quelle f. Replikation

Allerdings: Kein Read Only Slave möglich

16MiB Segmente, nur abgeschlossene werden gesendet -> hohe Latenz

Forks & Patches

Wednesday, October 13, 2010

Unterschiedliche Ansätze in Forks
zB von EnterpriseDB od. CommandPrompt

Wartung durch Third Parties, keine zentrale Koordination

Teststellungen für die native Integration

Is This Good for the PROJECT?

Wednesday, October 13, 2010

Wenn man Situation betrachtet: Keine Replikation sicherlich falsch

Allerdings: Wählbarer Schmerzgrad

standby Slaves: Geringer Adminaufwand

lesbare Slaves: hoher Adminaufwand

Situation UNZUFRIEDENSTELLEND

Kurs wechsel



Wednesday, October 13, 2010

Vergangenheit: Replikation für Core Team keine Priorität, sagt Replikation third party

Fehlende native Replikation großes Hemmniss für neue Benutzer

Vor zwei Jahren Treffen bei Konferenz, Entschluss Thema selbst anzugehen

8.4?

Leider nein!

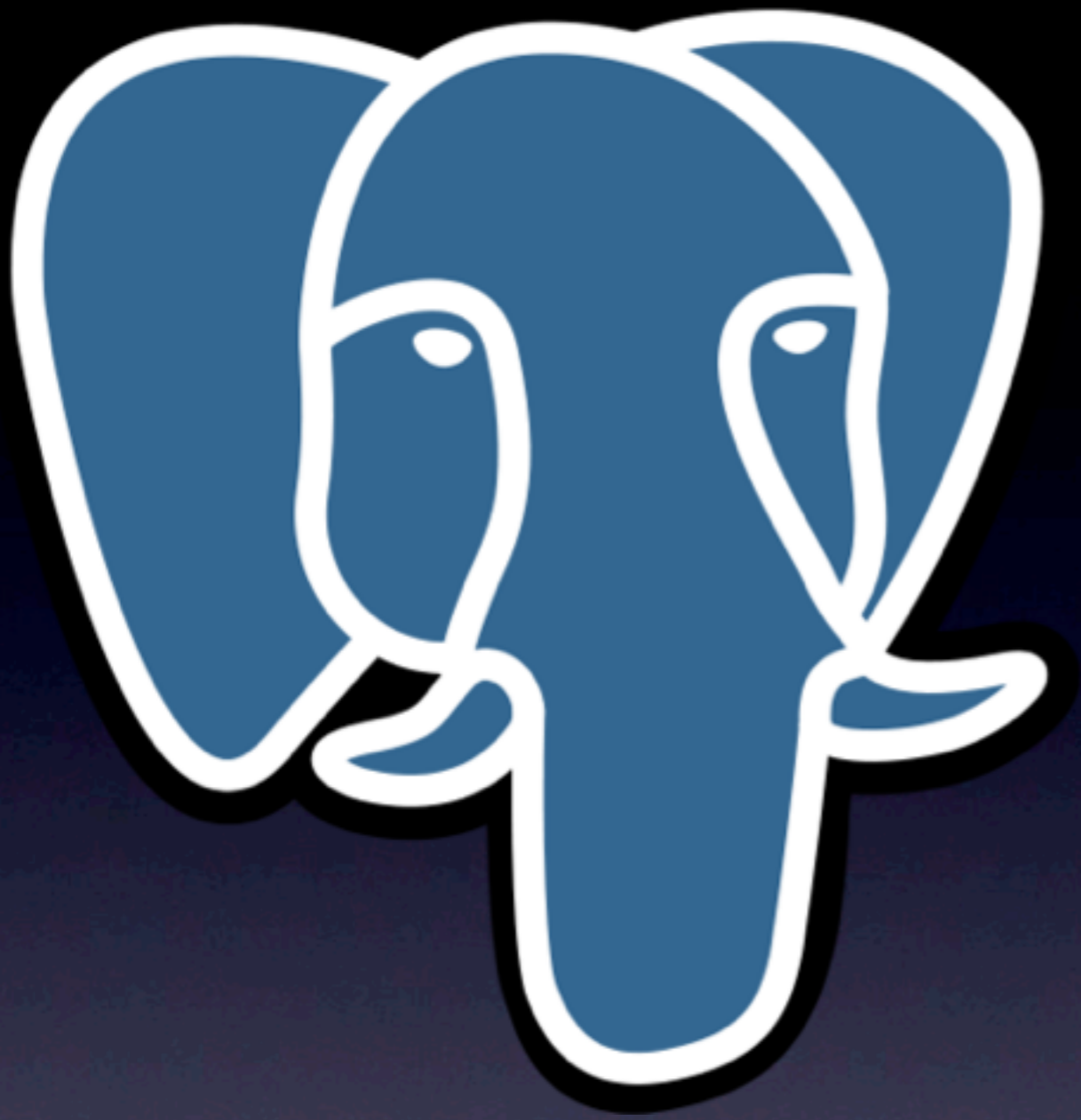
Wednesday, October 13, 2010

Erste Drafts f. 8.4 angekündigt

Zu umfangreiche Änderungen, besonders Hot Standby
Änderungen an WAL Format Neuland f. ganzes Team

Deswegen auf nächstes Release verschoben

Deswegen 9.0 und nicht 8.5



Es repliziert!

Wednesday, October 13, 2010

Und jetzt sind wir bei 9.0 und...

PostgreSQL repliziert!

Und das möchte ich euch gleich zeigen

DEMO

Wednesday, October 13, 2010

Wenns Fragen gibt, einfach losschiessen

Cheat sheet!

Nach Demo: Replikation ermöglicht durch zwei wesentliche Neuerungen

Streaming Replication

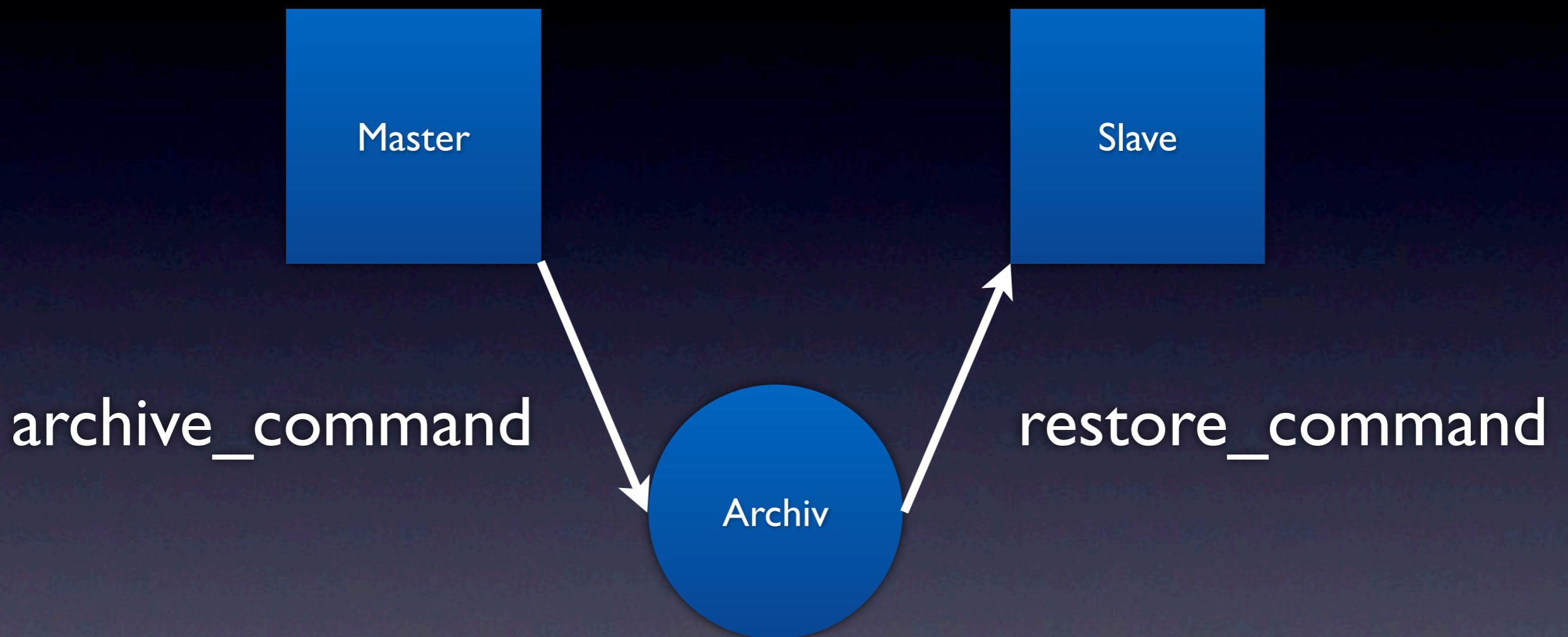
Wednesday, October 13, 2010

SR: by Nippon Telegraph and Telephone Corporation

Ermöglicht direkte Verbindung zwischen Master und Slave Server

Vorher sah das ungefähr so aus:

Log Shipping...



Wednesday, October 13, 2010

Wenn man Logshipping...

Zwei Server

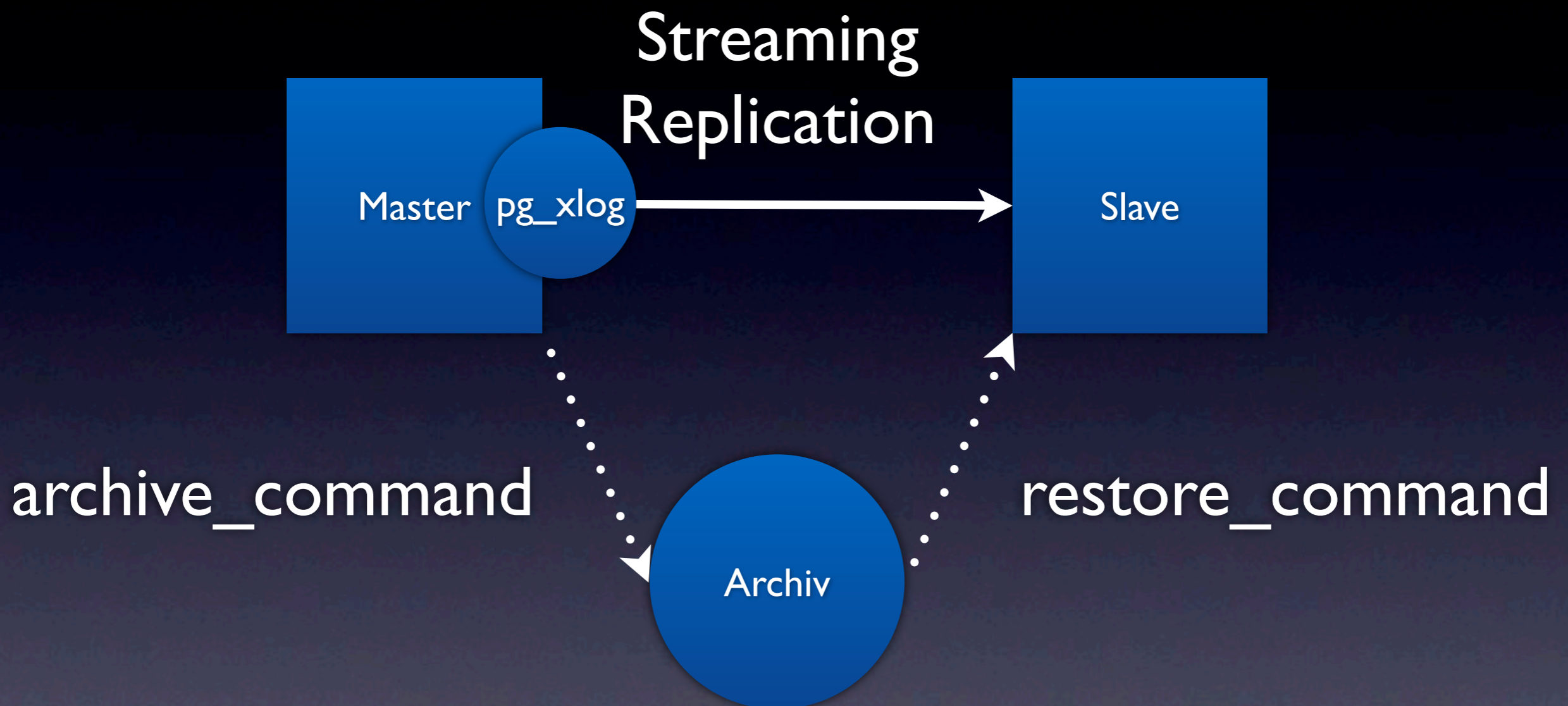
Master schreibt regelmässig WAL (od. Write Ahead Log) Files

Für jedes fertige File: archive_command ins Archiv

Slave: Regelmässig restore_command

Wenn neues File da, änderungen übernommen

...aufgebohrt



Wednesday, October 13, 2010

Jetzt direkte TCP-Verbindung zwischen Master & Slave

Alte Log Shipping Infrastruktur noch da

Notwendig wenn ein WAL-File aus dem Logverzeichnis gelöscht wurde

Im Normalbetrieb nicht notwendig

Relevant wenn Slave offline oder sehr lagged

ohne Archiv: haendisches Synchronisieren der Datenbanken

Hot Standby

Wednesday, October 13, 2010

HS: Simon Riggs, 2nd quadrant

Ermöglicht lesende Statements auf Slave

Readonly verlässlich: Garantiert keine Datenänderungen auf Slaves

Einiges an Refaktorisierungsarbeit notwendig gewesen

Größere WAL-Files weil mehr Information, muss am Master konfiguriert werden

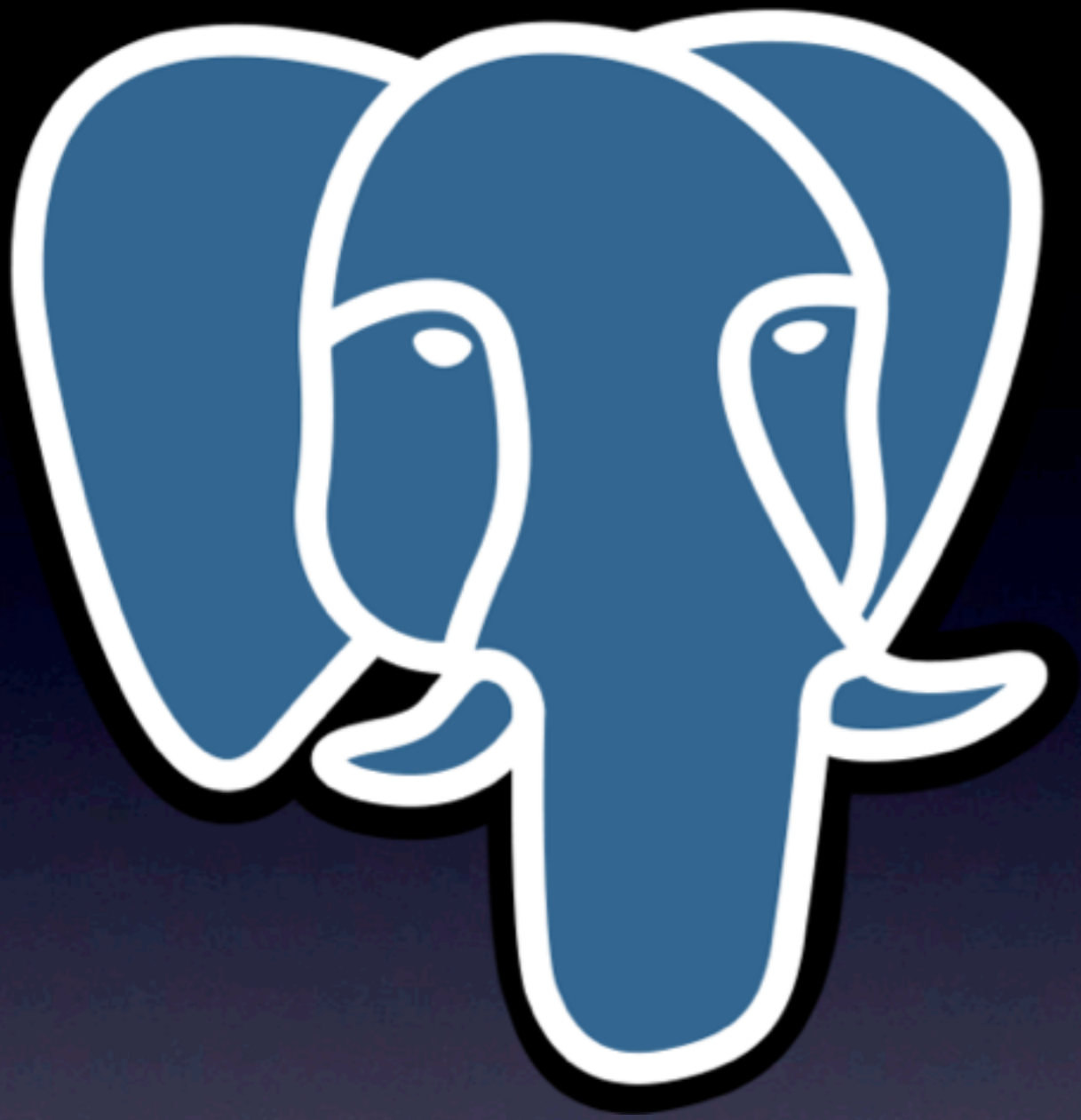
SETUP

Wednesday, October 13, 2010

Um zu zeigen unter der Haube aussieht

Einmal Slave Server löschen und neu konfigurieren

Wenn Fragen...



Das Kleingedruckte

Wednesday, October 13, 2010

Keine Lösung kann alle Use Cases abdecken

Hier einige Dinge die berücksichtigt werden sollten

Query Conflicts



Wednesday, October 13, 2010

Query Conflicts passieren

wenn Änderungen am Master mit Transaktionen am Slave kollidieren

Passiert, weil Master nichts über Zustand des Slaves weiß
allerdings viele Nebeneffekte und Interaktionen gibt

Häufigster Fall Cleanups durch Master

Ein Beispiel

Master

TX I



Slave

Wednesday, October 13, 2010

Abstrakter Fall, Master schreibt 3 Rows in der ersten TX

Slave repliziert

Slave Abfrage wird gestartet, läuft lang

Master löscht Wert (Markiert ihn als wiederverwendbar)

Slave repliziert, kein Problem, Daten noch da

Master überschreibt alten Wert

Slave kann die Änderungen wegen TX nicht übernehmen

Replikation Warten oder Slave Query killen

Master

TX 1



Slave



Wednesday, October 13, 2010

Abstrakter Fall, Master schreibt 3 Rows in der ersten TX

Slave repliziert

Slave Abfrage wird gestartet, läuft lang

Master löscht Wert (Markiert ihn als wiederverwendbar)

Slave repliziert, kein Problem, Daten noch da

Master überschreibt alten Wert

Slave kann die Änderungen wegen TX nicht übernehmen

Replikation Warten oder Slave Query killen

Master

TX 1



Slave



Slave TX



Wednesday, October 13, 2010

Abstrakter Fall, Master schreibt 3 Rows in der ersten TX

Slave repliziert

Slave Abfrage wird gestartet, läuft lang

Master löscht Wert (Markiert ihn als wiederverwendbar)

Slave repliziert, kein Problem, Daten noch da

Master überschreibt alten Wert

Slave kann die Änderungen wegen TX nicht übernehmen

Replikation Warten oder Slave Query killen

Master

TX 1



TX 2



Slave



Slave TX



Wednesday, October 13, 2010

Abstrakter Fall, Master schreibt 3 Rows in der ersten TX

Slave repliziert

Slave Abfrage wird gestartet, läuft lang

Master löscht Wert (Markiert ihn als wiederverwendbar)

Slave repliziert, kein Problem, Daten noch da

Master überschreibt alten Wert

Slave kann die Änderungen wegen TX nicht übernehmen

Replikation Warten oder Slave Query killen

Master

TX 1



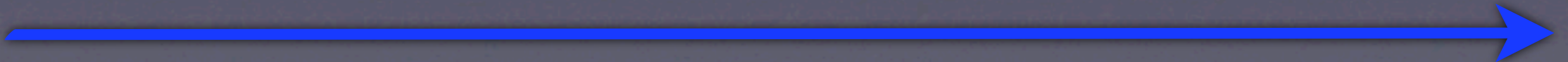
TX 2



Slave



Slave TX



Wednesday, October 13, 2010

Abstrakter Fall, Master schreibt 3 Rows in der ersten TX

Slave repliziert

Slave Abfrage wird gestartet, läuft lang

Master löscht Wert (Markiert ihn als wiederverwendbar)

Slave repliziert, kein Problem, Daten noch da

Master überschreibt alten Wert

Slave kann die Änderungen wegen TX nicht übernehmen

Replikation Warten oder Slave Query killen

Master

TX 1



TX 2



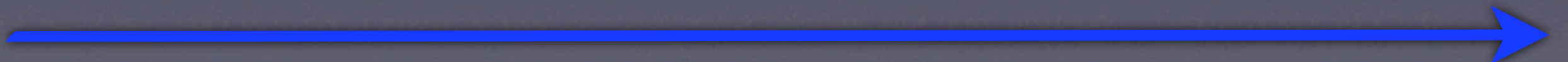
TX 3



Slave



Slave TX



Wednesday, October 13, 2010

Abstrakter Fall, Master schreibt 3 Rows in der ersten TX

Slave repliziert

Slave Abfrage wird gestartet, läuft lang

Master löscht Wert (Markiert ihn als wiederverwendbar)

Slave repliziert, kein Problem, Daten noch da

Master überschreibt alten Wert

Slave kann die Änderungen wegen TX nicht übernehmen

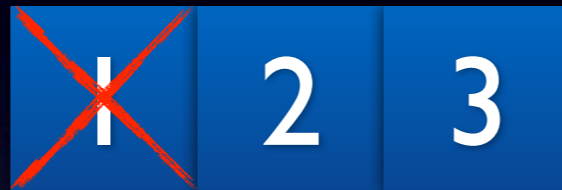
Replikation Warten oder Slave Query killen

Master

TX 1



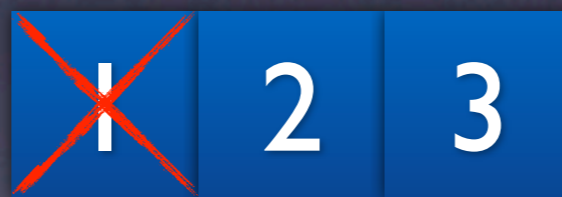
TX 2



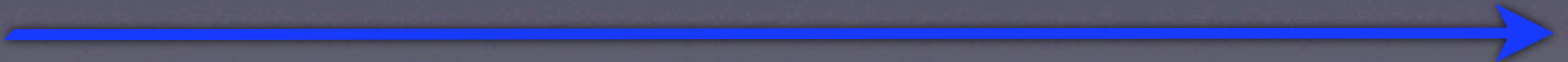
TX 3



Slave



Slave TX



Wednesday, October 13, 2010

Abstrakter Fall, Master schreibt 3 Rows in der ersten TX

Slave repliziert

Slave Abfrage wird gestartet, läuft lang

Master löscht Wert (Markiert ihn als wiederverwendbar)

Slave repliziert, kein Problem, Daten noch da

Master überschreibt alten Wert

Slave kann die Änderungen wegen TX nicht übernehmen

Replikation Warten oder Slave Query killen

Weiteres Konfliktpotential

Wednesday, October 13, 2010

Nicht einziger Fall wo passieren kann, u.a.:

Table-Locks (Alter Table etc)

DROP DATABASE

DROP TABLESPACE (mit temptables am slave)

max_standby_streaming_delay

Wednesday, October 13, 2010

Um Abbruch vs. Warten zu klären: max_standby_streaming_delay

Wieviele Sekunden Slave wartet

Differenz zwischen Empfangszeitpunkt und aktueller Uhrzeit

Abschliessend: Langlaufende TX können und werden gecancelt werden – Applikation muss darauf vorbereitet sein

Statementfreie Zone

Wednesday, October 13, 2010

WAL-Replikation -> Master und Slave immer ident

Keine Serverspezifischen Indizes

Keine selektive Replikation für DBs/Tables

Keine Baumreplikation (Master -> Slave -> Slave)

Primär Loadbalancing f. überschaubare Datenbanken, Backup, Failover

Failover

Wednesday, October 13, 2010

Keine eigene Funktionalität für HA o.ae., siehe Linux-HA

Infrastruktur gibts allerdings:

Um Slave zu Master zu machen: triggerfile anlegen (Datei, die überwacht wird)

Sobald ein Server Master wird, kein Slavebetrieb mehr möglich

Wenn Notwendig, neues Backup von "Master" ziehen

Don't be afraid

Wednesday, October 13, 2010

All das soll nicht verschrecken

Status Quo: Robuster Startpunkt f. weitere Entwicklung

9.0 Verwendbar für Produktionsbetrieb

Kein Letztstand, Gotchas/Todos/etc. werden gefixt!



...but is it done?

Wednesday, October 13, 2010

9.0 ist fertig

9.0 wurde am 20. September veröffentlicht

9.0.1 am 4. Oktober

Primär Security fixes für stored Procedures, keine dramatischen Probleme
Keine Probleme bei Replikation & Hot Standby

Die Zukunft

Wednesday, October 13, 2010

Am Horizont sind schon

Sync Replikation

Setup w/o manual copy

Easier archiving

wiki.postgresql.org

“Streaming Replication”

Wednesday, October 13, 2010

[Setup-Instruktionen im Wiki, nach... suchen](#)

[Detailinformationen in der ausführlichen Doku](#)

postgresql.org/community

Wednesday, October 13, 2010

PostgreSQL hat eine gut informierte und hilfsbereite Community

auch im deutschsprachigen Bereich

Nähere Infos unter...

Mailing Lists, IRC Kanäle sowie das Wiki

Probiert PostgreSQL!



Wednesday, October 13, 2010

Wenn ihr ein neues DB-Projekt startet, probiert PostgreSQL

Wenn ihr mit eurer aktuellen SQL-Datenbank unzufrieden seid, probiert PostgreSQL

Fragen?
Beschwerden?
Meine DB ist...?



Wednesday, October 13, 2010

Das war's von meiner Seite...

Fragen?
Beschwerden?
Meine Datenbank ist besser als deine!

Danke!

Michael Renner

Wednesday, October 13, 2010

Danke für's zuhören

Bei Fragen/Diskussionen: Jetzt hier, oder später via Kontaktmöglichkeiten, siehe Benutzerseite Wiki

Bildnachweis:

Seite 12: Selena Marie (<http://www.flickr.com/photos/selenamarie/2382172535/>)

Seite 22: "Reservoir Dogs": Adam Mateja (<http://admat.deviantart.com/gallery/#/d1h8noz>)

Seite 32: "Hypnotoad": copyright FOX, its related entities and the Curiosity Company

All other: istockphoto.com