

PostgreSQL

freeBSD



Was ist PostgreSQL?

PostgreSQL ist ein objektrelationales Datenbankmanagementsystem (ORDBMS), das als Open-Source-Programm frei verfügbar ist und ohne Lizenzierung heruntergeladen und benutzt werden darf. Ursprünglich wurde POSTGRES als Projekt an der University of California at Berkeley Computer Science Department entwickelt. Seither arbeiten viele Entwickler an PostgreSQL.

Es ist das fortschrittlichste Datenbanksystem im Open-Source-Bereich. PostgreSQL unterstützt die SQL92, SQL99 und SQL2003 Standards und es bietet außerdem eine Reihe eigener Erweiterungen an.

Es sind in PostgreSQL die Speicherung nicht atomarer Daten, Vererbung und Objektidentitäten implementiert wie es objektrelationale Datenbanken bieten. Es ist Benutzern erlaubt, das System um selbstdefinierte Datentypen, Operatoren und Funktionen zu erweitern. Neben der Unterstützung der referentiellen Integrität und einem fortschrittlichen Transaktionsmanagement bietet PostgreSQL die Definition von Triggern und Regeln, mit denen Zugriffe auf Datenbankobjekte gesteuert werden können.

Lizenz

PostgreSQL unterliegt wie FreeBSD der bekannten BSD-Lizenz, die bewusst klar und einfach gehalten ist.

Was bietet PostgreSQL

- ▶ PostgreSQL unterstützt den ANSI-SQL-Standard SQL92. Von den fünf Kategorien des SQL99-Standards, Framework, Foundation, Call Level Interface, Persistent Stored Modules und Host Language Bindings sind die Kategorien 1, 2 und 5 implementiert. Darüberhinaus bietet PostgreSQL eine Menge eigener Erweiterungen an, beispielsweise geometrische Datentypen.
- ▶ *Der Sprachumfang:* Mengenoperationen auf einer oder mehreren Tabellen sind mit PostgreSQL ebenso machbar wie Subselects, Outer Joins (auch der Full Outer Join wird unterstützt) oder Views.
- ▶ *Referentielle Integrität:* Mit der Definition von Primär- und Fremdschlüsseln können sich Tabellen referenzieren. So werden die Beziehungen zwischen Tabellen auf die Datenbank abgebildet und das System kontrolliert selbständig die Korrektheit dieser Beziehungen.
- ▶ *Transaktionen:* PostgreSQL unterstützt das ACID-Prinzip. Durch die Verwendung der fortschrittlichen MVCC-Technologie (**M**ulti **V**ersion **C**oncurrency **C**ontrol). In PostgreSQL, müssen Lesezugriffe niemals auf Schreibzugriffe warten und umgekehrt. Desweiteren werden noch Zwei-Phasen-Commitment und Savepoints angeboten.
- ▶ *Trigger:* In PostgreSQL sind „Before“- oder „After“-Trigger implementiert
- ▶ *Programmierung:* Zur Programmierung von Client-Programmen bietet PostgreSQL Schnittstellen zu ODBC und JDBC sowie zu zahlreichen Programmiersprachen.
- ▶ *Stored Procedures:* Das sind Funktionen, die in den unterschiedlichsten Programmiersprachen aber auch in SQL oder der PostgreSQL eigenen prozeduralen Programmiersprache PL/pgSQL geschrieben werden können. Sie werden kompiliert und im Datenbankserver abgelegt, weshalb sie sehr schnell sind.
- ▶ *Objektrelationale Eigenschaften:* PostgreSQL erlaubt die Speicherung nicht-atomarer Datentypen in Arrays oder in geometrischen Datentypen. Beides sind Erweiterungen von PostgreSQL, die nicht im SQL-Standard enthalten sind. Alle Datenbankobjekte, ob Tabelle, View, Sequenz oder einzelner Datensatz haben in PostgreSQL eine eigene Objektidentität. Außerdem wird die Vererbung von Tabelleneigenschaften auf abgeleitete Kindtabellen unterstützt.
- ▶ *Erweiterbarkeit:* Das Typsystem ist um eigene Datentypen, Operatoren und Funktionen erweiterbar, so dass das System individuell anpassbar ist. Funktionen können in SQL oder verschiedenen prozeduralen Sprachen geschrieben werden. Mit PL/pgSQL steht eine eigene prozedurale Sprache zur Verfügung.
- ▶ *Large Objects:* Binärdateien, zum Beispiel Sounddateien oder Bilder können in Tabellen gespeichert werden und entweder ganz oder in Teilen wieder ausgelesen werden. Zur Speicherung dieser Large Objects kann man zwischen zwei Datentypen wählen.
- ▶ *Zeichensätze:* PostgreSQL 'versteht' viele internationale Zeichensätze. Seit der Version 7.3 ist die Multibyte-Codierung standardmäßig aktiviert und ermöglicht den Umgang mit EUC (Extended Unix Code) oder Unicode.
- ▶ *SSL:* Die Datenkommunikation zwischen Client und Server kann mit SSL-Verbindungen verschlüsselt übertragen werden.

- ▶ *Kerberos Authentifizierung:* Kerberos ist ein Protokoll zur Authentifizierung in einem Netzwerk. Es wurde speziell zur sicheren Client-Server-Kommunikation entwickelt und kann direkt in PostgreSQL einkompiliert werden. Eine freie Implementation dieses Protokolls steht beim Massachusetts Institute of Technology zur Verfügung.
- ▶ *Im Betrieb:* PostgreSQL nutzt das Multi-Processing, unterstützt mittels „Slony“ Replikation, läßt dem Anwender die freie Wahl des Tablespaces und bietet Point-in-Time Recovery.
- ▶ *Clustering und Loadbalancing:* Die separat erhältliche Software „PG-Cluster“ bietet Loadbalancing und Clustering eines PostgreSQL-Servers.

Technische Daten

Parameter	Wert
Maximale Größe einer Datenbank	Unbegrenzt, es existieren Implementationen mit einer Datenbankgröße von 1 TeraByte
Maximale Größe einer Tabelle	16 TeraByte (ab Version 8: 32 TeraByte)
Maximale Größe einer Zeile	1,6 TeraByte
Maximale Größe eines Felds	1 GigaByte
Maximale Zahl von Zeilen in einer Datenbank	unbegrenzt
Maximale Zahl von Spalten	250 bis 1600, abhängig vom Spaltentyp
Maximale Zahl von Indeces einer Tabelle	unbegrenzt

„unbegrenzt“ bedeutet, dass das Dateisystem hier die Obergrenze festlegt.

FreeBSD und PostgreSQL

FreeBSD bietet dank seiner sehr guten Stabilität eine ideale Basis für PostgreSQL. Weitere Merkmale wie schnelles und sicheres Dateisystem, schneller TCP/IP-Stack und Security sind Säulen für eine optimale Betriebssystemplattform, die einen guten Datenbankserver auszeichnet.

FreeBSD hat im Hinblick auf diese Features einiges zu bieten:

- ▶ *Sehr gute Stabilität des Betriebssystems:* Für Datenbankserver ist es sehr wichtig, auf einem stabilen und ausgereiften Betriebssystem installiert zu sein. FreeBSD zeigt hier kaum Schwächen
- ▶ *Schnelles Dateisystem:* UFS-2 mit Softupdates stellt eine ideale Ausgangsbasis dar. Softupdates sind eine effiziente Methode zum Schreiben der Metadaten, die das Filesystem stets in einem konsistenten Zustand halten. Dies dient natürlich auch der Datensicherheit.

Auch große Tabellen im Dateisystem stellen bei UFS-2 mit seiner maximalen Dateigröße von 32 PiB (PebiByte = 2⁵⁰ Bytes) bzw. ma-

ximalen Volumengröße von 1YiB (YobiByte = 2^{80} Byte) kein Problem dar.

- ▶ **Schneller TCP/IP-Stack:** FreeBSD hat den in der Fachwelt anerkanntermaßen schnellsten und stabilsten TCP/IP-Stack. Nur so ist es möglich, Anfragen von Clients aus dem Netzwerk fast ohne Verzögerung an den Datenbankdienst zu leiten bzw. das Abfrageergebnis an den Client zurück zu senden.
- ▶ **Security:** Ein Datenbankserver muss gegenüber einer „feindlichen“ Umwelt geschützt sein. Dafür bietet gerade FreeBSD einige wichtige Features, die die meisten anderen Betriebssysteme vermissen lassen:
 - ▶ **Paketfilter:** mit IPFW und PF stehen zwei leistungsfähige Paketfilter zur Verfügung, wobei PF von OpenBSD portiert wurde. Diese Filter erlauben es dem Administrator, Anfragen bzw. Angriffe von unerwünschten Clients gezielt zu unterdrücken und somit die Datenbank(en) vor unerwünschten Subjekten zu schützen.
 - ▶ **Jail:** Dieses Feature geht einen Schritt weiter. Es stellt eine Art virtuelles FreeBSD dar. Prozesse, die in einer Jail gestartet werden, haben keinerlei Möglichkeit, das Wirtssystem zu infiltrieren.
- ▶ **Summa Summarum:** FreeBSD ist das ideale Betriebssystem für Datenbank-Server auf PostgreSQL-Basis. Beide Systeme ergänzen sich nahezu ideal.

Software für PostgreSQL auf FreeBSD-Basis

- ▶ Systemsoftware

PGCluster (Loadbalancing und Clustering) - Features: synchrone Arbeitsweise; Replikation auf Tabellenebene; Multi Master, zwei oder mehrere Datenbankserver können simultanen Anfragen von Clients erhalten; Replikation basiert auf Anfragen, diese werden innerhalb einer Gruppe von Datenbankservern von Server zu Server gesendet; replizieren von Sequenzen und Large Objects; Daten vom Typ serial werden synchronisiert.

Slony (Replikation von Datenbanken) - Features: asynchrone Arbeitsweise; Single Master to multiple Slaves; Slaves können kaskadiert werden, benutzt Trigger, Replikation auf Tabellenebene, keine exklusiven Locks notwendig, kann Schemaänderungen replizieren, alle Tabellen müssen einen Primärschlüssel haben, alle Knoten im Netzwerk müssen immer erreichbar sein, kann keine Large Objects replizieren, kein automatischer Failover

- ▶ Administration

pgAdmin: Umfangreiches grafisches Tool zur Administration, das dem Anwender fast keine Wünsche offen läßt. Es besticht durch einfache Bedienung und übersichtliche Darstellung der Datenbankstruktur. Eine installierte GTK2-Umgebung ist Voraussetzung.

phpPgAdmin: Eine auf PHP basierende Web-Anwendung, die un-

abhängig vom auf dem Client installierten Betriebssystem benutzt werden kann.

- ▶ Client-Software für PostgreSQL

ERP/CRM: Compiere, SQL-Ledger

Connectivity: ODBC/JDBC-Treiber, Ruby-, Python und Perl-Bindings, C/C++-Libraries, Qt-Plugin

Stored-Languages: pl/Ruby, pl/Python, pl/Pearl, pl/Java

Groupware: OpenGroupware, diverse POP3/IMAP-Server, AntiSpam-Verwaltung

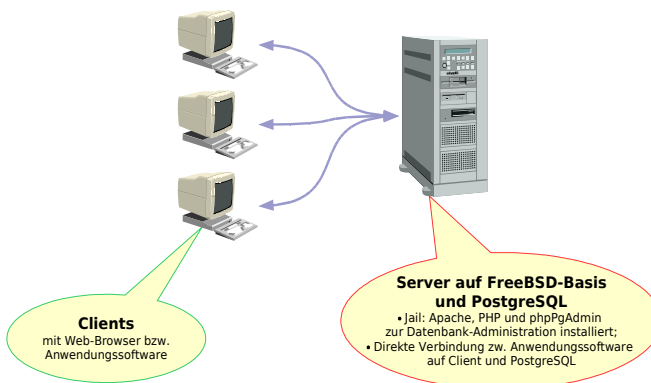
CMS/Shop-Systeme: LivingShop, XIST4C

Bilddatenbank: gPictureBase

Anwendungsbeispiele

- ▶ Kleiner bis Mittlerer Datenbankserver

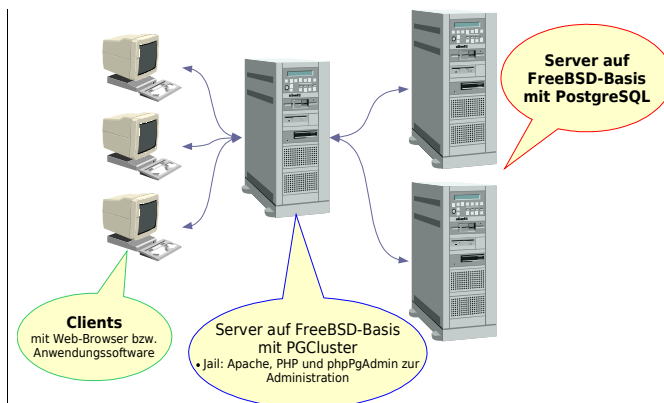
Ein kleiner bis mittlerer Datenbank-Server stellt geringere Anforderungen an die Hardware. Daher ist es wichtig bestimmte Teile der Installation in eine Jail zu verlagern, um den PostgreSQL-Prozess zu schützen.



Die Administrations-Software wird in einer Jail installiert. So kann der Administrator über das Web-Interface phpPgAdmin die Datenbank verwalten. Die Client-Software hat dagegen direkten Zugriff auf PostgreSQL.

- ▶ Großer Datenbankserver

In diesem Fallbeispiel wird auf zwei Servern PostgreSQL installiert und auf einem dritten Rechner PGCluster, das für Replikation und Loadbalancing sorgt. Die Administrations-Software wird in einer Jail installiert. So kann der Administrator über das Web-Interface phpPgAdmin die Datenbank verwalten. Die Clients kommunizieren hier aber mit dem PGCluster-Server.



Weiterführende Information

- ▶ Literatur

PostgreSQL (2nd Edition)

Korry Douglas, Sams Publishing, ISBN: 0672327562

Beginning Databases with PostgreSQL:

From Novice to Professional, Second Edition

Neil Matthew, Apress, ISBN: 1590594789

PostgreSQL Ge-packt

Peter Eisentraut, mitp-Verlag, ISBN 3826614933

PostgreSQL

Paul Weinstabl, C&L-Verlag, ISBN 3936546223

PostgreSQL. Das offizielle Handbuch.

Peter Eisentraut, mitp-Verlag, ISBN: 3826613376

PostgreSQL Developer's Handbook

Ewald Geschwinde, Sams Publishing, ISBN 0672322609

Beginning PHP and PostgreSQL 8

W. Jason Gilmore, Apress, ISBN 1590595475

PHP and PostgreSQL Advanced Web Programming

Ewald Geschwinde, Sams Publishing, ISBN 0672323826

Literatur zum Thema FreeBSD bitte auch den Flyer „BSD-Literatur“ beachten!

- ▶ Internet

PostgreSQL-Homepage: www.postgresql.org

deutsche PostgreSQL-Homepage: www.postgres.de

PGCluster: <http://pgcluster.projects.postgresql.org/index.html>

FreeBSD: www.freebsd.org