

PostgreSQL - amministrazione

PostgreSQL

Gabriele Bartolini

Flavio Casadei Della Chiesa

Luca Ferrari

Marco Tofanari

Associazione Italiana PostgreSQL Users Group

www.itpug.org

Pisa, 8 Maggio 2009



Licenza Creative Commons

Attribuzione

Non commerciale

Condividi allo stesso modo

2.5 Italia

<http://creativecommons.org/licenses/by-nc-sa/2.5/it/>

Questo documenti è un riadattamento di una presentazione realizzata da [Devise.IT](#) intitolata “Cos'è PostgreSQL”, tenutasi il 23 marzo 2009 all'Università di Pisa (

<http://www.tosslab.it/documentazioneeventi/azstud-2ndquadrant>)



Cos'è PostgreSQL

- Introduzione
- Open-Source
- La storia di PostgreSQL
- Panoramica delle funzionalità più importanti
- I pro e i contro
- Il futuro di PostgreSQL
- La comunità di PostgreSQL e ITPUG



Introduzione

- PostgreSQL è noto come “il più avanzato **sistema di gestione di basi di dati** disponibile in **open-source**”
- Termini:
 - Sistema di gestione di basi di dati (Database Management System – DBMS)
 - Open-Source



Database e DBMS

- Un **database** è *una collezione di dati strutturati*
- Un **DBMS** è *un insieme di componenti software in grado di permettere la creazione e la manipolazione di un database*
- Esistono diversi modelli di database e DBMS:
 - Gerarchico
 - Reticolare
 - **Relazionale** (più diffuso)
 - *Object Relational*
 - Orientato agli oggetti



ORDBMS

- **Object Relational DataBase Management System**
- Un ORDBMS è un DBMS relazionale che:
 - Supporta un modello di database object oriented (implementa oggetti, classi e ereditarietà)
 - L'estensione del modello di dati con tipi di dati e metodi personalizzati
- PostgreSQL è un ORDBMS



Open-Source

- Il **software open-source** (a codice aperto) permette lo studio, il riuso e la modifica del codice secondo determinate **regole**
- Tali regole sono dette **licenze d'uso**
- PostgreSQL è distribuito secondo la licenza BSD, una delle più **permissive**
- L'**utilizzo**, la **modifica**, la **distribuzione** del codice di PostgreSQL non comporta **nessun esborso di denaro** (costi di licenza = 0)
- Totale libertà aziendale (nessuna restrizione)



Licenza BSD

Copyright (c) <year>, <copyright holder>

All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- **Redistributions of source code** must retain the above copyright notice, this list of conditions and the following disclaimer.
- **Redistributions in binary** form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
- Neither the name of the <organization> nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.



Licenze a confronto: PostgreSQL/MySQL/Oracle

- **Costi** per lo studente:
 - **PostgreSQL: no** – Oracle: no – MySQL: no
- **Costi** per lo sviluppatore (librerie client):
 - **PostgreSQL: no** – Oracle: no
 - MySQL: no se FOSS, sì se commerciale
- **Costi** per l'utilizzatore:
 - **PostgreSQL: no** – Oracle: sì
 - MySQL: no se GPL, sì se commerciale
- **Possibilità** di redistribuzione:
 - **PostgreSQL: sì** – MySQL: sì se FOSS



Novità nel mercato del database open-source

- **16 gennaio 2008:** Sun Microsystems acquista MySQL AB per la cifra di 1 miliardo di dollari americani
- **21 aprile 2009:** Oracle Corporation acquista Sun Microsystems per circa 7 miliardi e mezzo di dollari americani



La storia di PostgreSQL

- **Progetto di ricerca** nato presso l'Università di California (Berkeley):
 - 1986 – 1993: University **Postgres** (**Michael Stonebraker**)
 - 1994 – 1995: Postgres95 (Jolly Chen e Andrew Yu)
 - 1996 – oggi: **PostgreSQL**:
 - Fase 1, 1996 – 1998: eliminazione dei crash
 - Fase 2, 1998 – 2001: aderenza allo standard SQL
 - Fase 3, 2001 – 2004: robustezza e stabilità
 - Fase 4, dal 2004 in poi: funzionalità di classe enterprise
- Contributo di centinaia di sviluppatori provenienti da tutto il mondo



Postgres o PostgreSQL?

- Il nome ufficiale è **PostgreSQL**
- Postgres era il nome iniziale del progetto
- Con l'aggiunta del supporto SQL (1996) il progetto cambiò nome in PostgreSQL
- **Postgres** è considerato un sinonimo (più semplice da pronunciare)



Architettura

- Client / Server
- Scritto in linguaggio ANSI C
- Utilizza servizi stabili del sistema operativo
- Altamente portabile:
 - **Linux** (tutte le distribuzioni): Red Hat, Fedora, CentOS, SUSE, Debian, Ubuntu, Gentoo
 - **BSD OS**: FreeBSD, OpenBSD, NetBSD
 - **UNIX** (tutte le principali piattaforme): Solaris, HP-UX, Tru64, AIX
 - **Windows, Mac OS** e altri sistemi operativi



Alcuni casi d'uso

- Applicazioni OLTP
- Sistemi di data warehousing e DSS
- Applicazioni web
- Applicazioni real-time
- Sistemi territoriali (estensione PostGIS)



Alcune fra le funzionalità principali 1/2

- Viste
- Stored procedure (attualmente in 12 linguaggi di programmazione)
- Trigger
- Transazioni ACID (**MVCC** e locking)
- **Estensibilità e funzionalità Object Relational**
- Supporto UNICODE (codifica UTF-8)
- Interfacce di connessione (ODBC e JDBC)
- Schemi



Alcune fra le funzionalità principali 2/2

- Supporto SQL:2003
- Ricerca full text (con dizionari, *lemmi* e sinonimi)
- Tablespace
- Array
- Auto-compressione dei campi testo (> 2000 caratteri)
- Commit asincrona
- Two-phase commit
- Alta disponibilità (**Warm Standby**) e replication
- Localizzazione (NLS)



Sicurezza

- Concetto di “sicurezza di default”
- Modello di sicurezza multi-livello (server, database, tabella, colonna)
- Criptatura completa (SSL)
- Integrazione con l'infrastruttura aziendale di sicurezza (e.g. GSSAPI, LDAP)
- Rapido fix dei *bug* di sicurezza



Replica

- Non è ancora parte integrante di PostgreSQL
- E' al momento delegata a strumenti esterni:
 - **Warm Standby** (modulo *contrib* dalla v. 8.2)
 - Slony
 - Londiste



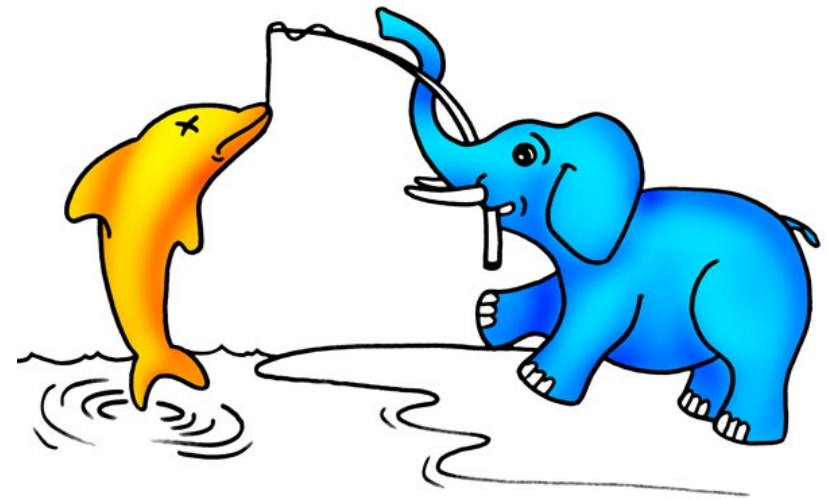
Alcune fra le novità di PostgreSQL 8.4

- Note:
 - Lista non ancora definitiva
 - Rilascio previsto per estate 2009 (?)
- Permessi a livello di colonna
- Common table expression
- Window function
- *Collation* a livello di database



C'è rivalità fra MySQL e PostgreSQL?

- Assolutamente NO
- Target in larga parte diverso
- Caratteristiche MOLTO diverse (talk apposito!)
- Personale MySQL nella comunità di Postgres



Alcuni miti da sfatare

- **“PostgreSQL è lento”:**
 - Giudizio (**fondato**) diffuso nei primi anni 2000
 - Boom di applicazioni web in PHP
 - Paragonato con MySQL (!?!?)
 - Benchmark mono-processo senza tuning
 - **Da PostgreSQL 8.0 questo mito è stato sfatato (vedi benchmark Sun)**
- **“PostgreSQL è difficile da configurare”:**
 - Giudizio fondato fino a PostgreSQL 8.0
 - Configurazione di default adatta per l'uso medio



Curiosità

- Skype utilizza PostgreSQL come database di backend per memorizzare i profili dei propri utenti
- Attualmente il database contiene circa 400 milioni di profili
 - Scalabile fino a 1 miliardo di profili
- Skype sviluppa soluzioni open-source per PostgreSQL (SkyTools) riguardanti :
 - Alta disponibilità (HA)
 - Scalabilità, clustering, partitioning
 - ...



Alcune ragioni per usare PostgreSQL 1/2

- Costo di licenza nullo
- Assenza di vendor lock-in
- Assenza di restrizioni sull'utilizzo
- Funzionalità native mature e stabili:
 - Viste, schemi, stored procedures, trigger, ecc.
- Multi-piattaforma
- Stabilità
- Sicurezza



Alcune ragioni per usare PostgreSQL 2/2

- A partire dalla versione 8:
 - funzionalità *enterprise-class*
 - migliore scalabilità
 - migliori performance
 - adatto per sistemi di supporto alle decisioni
- Vasta e dettagliata documentazione
- Efficiente e tempestivo supporto da parte della **comunità**
- Alta disponibilità, replica e scalabilità
- Estensibilità



Alcuni limiti attuali di PostgreSQL

- Curva di apprendimento ripida (complesso)
- Carenza di tool grafici per l'amministrazione (pgAdminIII e Navicat)
- Carenza di strumenti di design visuale (ERD, UML)
- Marketing delegato esclusivamente alla comunità
- Carenza di supporto professionale 24x7 (ma grossa community)
- Carenza di applicazioni verticali che **ufficialmente** lo supportano
- Carenza di hosting “economico”
- Carenza di sistemi Open-Source per il tuning



Le fasi dello sviluppo di PostgreSQL 1/2

- Tutto nasce dal bisogno di nuove funzionalità
- Bisogni nati per scopi diversi:
 - Commerciali
 - Accademici
 - Interesse personale (in pieno spirito *hacker*)
- Presenza occasionale di **committenti**
- Le nuove funzionalità vengono proposte in due finestre temporali, chiamate **Commit fest**
- Le proposte hanno la forma di vere e proprie **patch** complete di documentazione



Le fasi dello sviluppo di PostgreSQL 2/2

- **Review:** ad ogni patch viene assegnato un esperto che dovrà valutare se applicare la patch al codice
- **Beta:** viene rilasciata una versione per testing diffuso
- Fase di pre-rilascio: la comunità prepara gli annunci nelle principali lingue (italiano incluso)
- Viene rilasciata la versione M.m.0
- Inizia il periodo di manutenzione che può durare diversi anni
- La comunità PostgreSQL supporta le ultime 5 macro release (al momento 8.3, 8.2, 8.1, 8.0, 7.4)



La comunità italiana e europea

- Nata intorno alla lista tecnica psql.it
- Luglio 2007: primo **PostgreSQL Day italiano**, Prato
- Novembre 2007: nasce **ITPUG**, la prima associazione non-profit per la diffusione di PostgreSQL in Italia
- Febbraio 2008: nasce **PostgreSQL Europe**, associazione non-profit per la diffusione di PostgreSQL in Europa
- Ottobre 2008: primo **PostgreSQL Day europeo** / secondo PostgreSQL Day italiano, Prato





ITPUG: cosa è

- Associazione culturale non-profit
- Sviluppatori e utenti PostgreSQL italiani
- Ufficialmente riconosciuta dalle istituzioni
- Fondata su uno statuto redatto in modo collaborativo e democratico
- Aperta a tutti
- Aperta agli individui (non alle aziende)



ITPUG: organizzazione

- Consiglio direttivo:
 - 5 soci
 - Eletti dall'assemblea generale dei soci
 - Prendono decisioni in modo democratico
 - Si ritrovano periodicamente via rete e di persona
- Chapter regionali:
 - Soci di ITPUG che intendono rappresentare la comunità PostgreSQL in una regione, promuovendo attività culturali in loco compatibili con gli obiettivi dell'associazione



ITPUG: alcuni fra i principali obiettivi

- Promuovere PostgreSQL con professionalità verso:
 - Le istituzioni (Pubblica Amministrazione)
 - Le scuole e le università
 - Le aziende
- Organizzare l'evento PostgreSQL Day italiano ogni anno
- Attività di marketing per PostgreSQL
- Partecipare a iniziative/conferenze su software libero e open-source
- Punto di incontro fra la comunità internazionale e quella italiana
- Creare progetti che possano migliorare il progetto, soprattutto per quanto riguarda la localizzazione in italiano



Conclusioni

- PostgreSQL è un potentissimo ORDBMS open-source
- Ha funzionalità molto stabili e standard
- Ha una comunità molto attiva
- Può essere lo strumento adatto in moltissimi casi ma non è assolutamente la “panacea di tutti i mali”
- Vale la pena approfondirlo in quanto le conoscenze acquisite sono trasferibili



Link utili e riferimenti

- <http://www.postgresql.org/> : PostgreSQL
- <http://www.itpug.org/> : Italian PostgreSQL Users Group
- Slide “Un database libero per la Pubblica Amministrazione Italiana”:
<http://www.itpug.org/2008-10-25-itpug-postgresql.pdf>



Link utili e riferimenti

- <http://www.postgresql.org/> : PostgreSQL
- <http://www.itpug.org/> : Italian PostgreSQL Users Group
- Slide “Un database libero per la Pubblica Amministrazione Italiana”:
<http://www.itpug.org/2008-10-25-itpug-postgresql.pdf>
- Slide distribuite in licenza Creative Commons 2.5 BY-NC-SA da 2ndQuadrant Italia (<http://www.2ndQuadrant.it/>) riguardanti la presentazione “Cos'è PostgreSQL”, tenutasi il 23 marzo 2009 all'Università di Pisa (
<http://www.tosslab.it/documentazioneeventi/azstud-2ndquadrant>)

