

Administracao de Banco de Dados SGBD

PostgreSQL: uma alternativa para sistemas gerenciadores de banco de dados de código aberto

Uma das questões importantes no desenvolvimento de qualquer aplicação para computadores que faça uso de um banco para armazenamento de dados é sobre a escolha de um sistema gerenciador de banco de dados, ou SGBDs. Embora possamos administrar e armazenar os dados de formas alternativas, o SGBD é a escolha mais eficiente e também mais compatível com as tecnologias utilizadas no mercado.

Dentre os sistemas mais notórios, o PostgreSQL é um sistema gerenciador de banco de dados objeto relacional que se destaca pelo código aberto e por ser um dos mais avançados em termos de recursos neste segmento.

Este artigo tem como propósito apresentar o PostgreSQL, ressaltando as suas vantagens e apontando as desvantagens perante outros sistemas similares disponíveis. A partir do histórico do sistema, serão discutidos os recursos oferecidos, as decisões de projeto tomadas e a situação do PostgreSQL na atualidade.

1. História do PostgreSQL

O PostgreSQL é um projeto que foi inicialmente desenvolvido sob a licença BSD (Berkeley Software Distribution, de código aberto) na Universidade de Berkeley, Califórnia, e liderado pelo professor Michael Stonebraker, que lançou a sua primeira versão em 1989. O principal objetivo da criação do PostgreSQL foi criar um SGBD que compreendesse diferentes tipos de dados, ou objetos, e fosse capaz de descrever relações entre eles.

Nesta versão, o sistema suportava a linguagem QUEL, que foi substituída pela linguagem SQL cinco anos depois, pelos alunos Andrew Yu e Jolly Chen, fato que deu origem ao sufixo no nome atual do sistema. Com o seu lançamento na internet, o desenvolvimento do PostgreSQL tomou uma nova direção.

Atualmente, o desenvolvimento segue o modelo Bazar [1], no qual o código é desenvolvido de forma totalmente aberta e pública. Seus desenvolvedores são, em sua maioria, voluntários espalhados pelo mundo que se comunicam pela Internet. A validação ocorre através de um grupo seletivo, designado especialmente para este fim. O PostgreSQL é financiado por grandes empresas, como a Fujitsu, Hub.Org, NTT Group, Red Hat, Skype e SRA, além de receber doações.

Entre seus usuários estão multinacionais, órgãos governamentais de diferentes países e grandes universidades. O PostgreSQL recebeu ainda diversos prêmios, dos quais se destacam: melhor sistema de banco de dados open source pela Linux New Media Award, e ganhador do prêmio Linux Journal Editors' Choice como melhor SGBD open source por cinco vezes.

2. Recursos do PostgreSQL

O PostgreSQL se destaca como um dos sistemas de código aberto que oferece grande número de recursos frente a sistemas similares, como o MySQL, sendo inclusive comparado a sistemas comerciais de grande porte, como o Oracle.

Ele possui muitas características de sistemas modernos, como: consultas complexas,

triggers (disparadores), transações e controle de concorrência.

Consultas complexas diz respeito a consultas a dados do banco que envolvem diversas condições, junções de tabelas e funções sobre estes dados. Disparadores, são funções criadas que são ligadas a eventos - um disparador pode ser configurado, por exemplo, para ser executado antes ou depois de qualquer inserção, atualização ou exclusão de dado do banco. O conceito de transações está intimamente ligado à segurança de completude das operações. Finalmente, o controle concorrência visa controlar tentativas de acessos simultâneos ao banco de dados.

Além disso, é possível estendermos todas as suas funcionalidades através da adição de: novos tipos

de dados, funções, operadores, funções agregadas, métodos indexados e linguagens procedurais.

Associada a todas essas características, a principal vantagem que torna o sistema ilimitado em seu poder de extensão é a possibilidade de ser modificado e distribuído livre de qualquer cobrança, para fins comerciais, privados ou acadêmicos.

3. Popularidade

Quando comparado ao seu concorrente open source MySQL, é notável a superioridade deste em relação à popularidade de uso. A razão disso pode ser explicada pelo senso comum de que o MySQL era mais simples que o PostgreSQL. Pelo fato de o PostgreSQL ser mais robusto e ter mais funções, exigia um conhecimento técnico mais profundo. No entanto, mudanças significativas foram implementadas visando melhorar a usabilidade e essa diferença em relação ao MySQL já não é mais tão notável.

Outro fato relevante que dificultou a popularização do PostgreSQL é a disponibilização tardia do SGBD para o ambiente Windows, o que veio a ocorrer apenas na sua oitava versão.

O PostgreSQL é um Sistema Gerenciador de Base de Dados Relacional (SGBDR) de alta performance, de fácil administração e utilização em projetos. Esse SGBDR permite a utilização da linguagem SQL, triggers (disparadores) e tantos outros recursos presentes nos mais famosos sistemas SGBDR do mercado, como Oracle, InterBase, SQL Server, MySQL, etc., além de permitir "Embedded SQL" com pré-compilação (com as Linguagens C e C++).

Além disso, o PostgreSQL possui interface com diversos ambientes e linguagens de programação, como C, C++, MS Visual Basic, Perl e Java. Todos esses recursos proporcionam ao programador e ao administrador de banco de dados realizar suas tarefas e atender suas perspectivas mais específicas.

Uma de suas principais características que motivou este artigo é o fato de PostgreSQL ser um software livre, ou seja, seu uso, modificação ou distribuição são livres. Desde que PostgreSQL se tornou software livre, em meados de 1996 quando seu código-fonte foi aberto, várias contribuições têm sido agregadas a fim de torná-lo uma ótima opção extremamente competitiva em nível de funcionalidade com os mais modernos e conhecidos SGBDRs do mercado.

Sistema de Gerenciamento de Banco de Dados

Um Sistema de Gerenciamento de Banco de Dados (SGBD) — do inglês Data Base Management System (DBMS) — é o conjunto de softwares responsáveis pelo gerenciamento de um banco de dados. Seu principal objetivo é retirar da aplicação cliente a responsabilidade de gerenciar o acesso, a persistência, a manipulação e a organização dos dados. O SGBD disponibiliza uma interface para que seus clientes possam incluir, alterar ou consultar dados previamente armazenados. Em bancos de dados relacionais a interface é constituída pelas APIs (Application Programming Interface) ou drivers do SGBD, que executam comandos na linguagem SQL (Structured Query Language).

Todas as organizações, por menor que sejam, possuem quantidades cada vez maiores de dados e informações a armazenar. Todavia, a manipulação dessas informações tornou-se impossível de ser realizada manualmente (via papéis, principalmente), pois sua utilização, além de demorada (devido a catalogação dos dados), é passível de erros, principalmente ocasionados pelo desgaste do operador em conseguir resgatar informações requisitadas.

Nesse sentido, torna-se mais fácil encontrar a informação numa base de dados que recorre a uma das tecnologias de informação de maior sucesso e confiança. Ou seja, as bases de dados estendem a função do papel ao guardar a informação em computadores. Qualquer empresa que pretenda garantir um controle efetivo sobre todo o seu negócio, tem obrigatoriamente de recorrer a sistemas de gestão de bases de dados. Existem muitos tipos de ferramentas, completas e com funcionalidades acrescidas, que elevam outros níveis, a capacidade operacional de gerar informação de valor para a organização.

Um sistema de gerenciamento de banco de dados não é nada mais do que um conjunto de programas que permite armazenar, modificar e extrair informações de um banco de dados. Há muitos tipos diferentes de SGBD. Desde pequenos sistemas que funcionam em computadores pessoais a siste-

mas enormes que estão associados a mainframes. Um SGDB implica a criação e manutenção das bases de dados, elimina a necessidade de especificação de definição de dados, age como interface entre os programas de aplicação e os ficheiros de dados físicos, e separa as visões lógica e de concepção dos dados. Assim sendo, são basicamente três os componentes de um SGBD:

- Linguagem de definição de dados (especifica conteúdos, estrutura a base de dados e define os elementos de dados);
- Linguagem de manipulação de dados (para poder alterar os dados na base);
- Dicionário de dados (guarda definições de elementos de dados e respetivas características — descreve os dados, quem os acede, etc.) (Gouveia; 2009).

Qualidade de Dados

Um banco de dados é meio caminho andado para que a empresa tenha a informação que precisa. Para isso, outras medidas devem ser tomadas para ter certeza de que os dados sejam confiáveis. Alguns dos erros são causados por dados incoerentes produzidos por múltiplos sistemas. Se o banco de dados for projetado adequadamente, a ocorrência de dados incoerentes será pequena. Porém a maioria dos problemas de qualidade com nomes digitados incorretamente, números trocados ou códigos faltantes, ocorre durante a entrada de dados, esses erros ficam mais comuns quando as empresas transferem parte dos seus dados para a Internet, e permite que clientes e fornecedores insiram seus dados no site e isso efetue alterações no sistema interno. Os problemas com qualidade de dados não são só empresariais, eles também representam sérios problemas às pessoas, afetando sua condição financeira e até mesmo seu emprego.

Descrição

Um modelo de SGBD define como os dados serão armazenados no banco de dados. Os quatro modelos mais conhecidos são:

- hierárquico;
- em rede;
- relacional;
- orientado a objetos

Existem também outros modelos, variando com o autor:

- o modelo de dados objeto-relacional é praticamente uma mistura do modelo relacional com o orientado a objetos.
- o modelo relacional estendido, é uma adição de características do modelo orientado a objetos ao relacional
- o semiestruturado é dedicado a documentos em formatos semiestruturados, normalmente em XML;
- estruturas de dados otimizadas, que possam manipular uma grande quantidade de informação;
- uma linguagem que possibilite a criação, atualização e consulta dos dados armazenados. Normalmente esta linguagem é dividida em partes:
 - Linguagem de definição de dados ou LDD (ou DDL, do inglês), com comandos como CREATE, DROP e ALTER TABLE;
 - Linguagem de manipulação de dados, ou LMD (ou DML, do inglês), com comandos como UPDATE, SELECT, INSERT e DELETE;
 - Linguagem de controle de dados, ou LCD, com comandos para controle de acesso dos usuários do sistema, como GRANT e REVOKE, em SQL.

- um mecanismo transacional que garanta a consistência, entre as operações, dos dados armazenados. Também é possível definir uma linguagem adicional para restrições, como a OCL. As principais linguagens para manipular bancos de dados são: SQL, em seus vários padrões, como SQL2 e SQL3; e OQL.

PostgreSQL

Características

Hoje, o PostgreSQL é um dos SGBDs (Sistema Gerenciador de Bancos de Dados) de código aberto mais avançados, contando com recursos como:

- Consultas complexas
- Chaves estrangeiras
- Integridade transacional
- Controle de concorrência multi-versão
- Suporte ao modelo híbrido objeto-relacional
- Ele tem facilidade de Acesso
- Gatilhos
- Visões
- Linguagem Procedural em várias linguagens (PL/pgSQL, PL/Python, PL/Java, PL/Perl) para Procedimentos armazenados
- Indexação por texto
- Estrutura para guardar dados Georeferenciados PostGIS

Histórico

O PostgreSQL é um dos resultados de uma ampla evolução que se iniciou com o projeto Ingres, desenvolvido na Universidade de Berkeley, Califórnia. O líder do projeto, Michael Stonebraker, um dos pioneiros dos bancos de dados relacionais, deixou a universidade em 1982 para comercializar o Ingres, porém retornou a ela logo em seguida.

Após seu retorno a Berkeley, em 1985, Stonebraker começou um projeto pós-Ingres com o objetivo de resolver problemas com o modelo de banco de dados relacional. O principal problema era a incapacidade do modelo relacional compreender “tipos” (atualmente, chamados de objetos), ou seja, combinações de dados simples que formam uma única unidade.

O projeto resultante, chamado Postgres, era orientado a introduzir a menor quantidade possível de funcionalidades para completar o suporte a tipos. Estas funcionalidades incluíam a habilidade de definir tipos, mas também a habilidade de descrever relações - as quais até este momento eram amplamente utilizadas, mas completamente mantidas pelo usuário. No Postgres, o banco de dados “compreendia” as relações e podia obter informações de tabelas relacionadas utilizando regras.

Iniciando em 1986, a equipe divulgou uma série de documentos descrevendo a base do sistema e em 1988 o projeto possuía um protótipo funcional. A versão 1 foi liberada para um grupo pequeno de usuários em junho de 1989, seguida pela versão 2 com um sistema de regras reescrito em junho de 1990. Para a versão 3, liberada em 1991, o sistema de regras foi reescrito novamente, mas também foram adicionados suporte para múltiplos gerenciadores de armazenamento e um melhorado motor de consultas. Já em 1993, Postgres havia crescido imensamente em popularidade e possuía uma grande demanda por suporte e por novas funcionalidades. Após a liberação da versão 4, a qual era uma simples versão de limpeza, o projeto foi oficialmente abandonado pela Universidade de Berkeley.

Entretanto, devido ao fato do seu código fonte estar sob uma licença BSD, o seu desenvolvimento foi continuado. Em 1994, dois estudantes, Andrew Yu e Jolly Chen, adicionaram um interpretador SQL para substituir a linguagem QUEL (desenvolvida para o Ingres) e o projeto foi renomeado para **Postgres95**. Com a divulgação de seu código pela Internet, Postgres95 iniciou uma nova vida como software open source.

Em agosto de 1996, Marc Fournier, Bruce Momjian e Vadim B. Mikheev lançaram a primeira versão externa da Universidade de Berkeley e deram início à tarefa de estabilizar o código herdado. Também em 1996, o projeto foi renomeado para **PostgreSQL** a fim de refletir a nova linguagem de consulta ao banco de dados: SQL. A primeira versão de PostgreSQL, a 6.0, foi liberada em janeiro de 1997. Desde então, um grupo de desenvolvedores e de voluntários de todo o mundo, coordenados pela Internet, têm mantido o software e desenvolvido novas funcionalidades.

As principais características acrescentadas nas versões 6.x são o en:MVCC (Multiversion Concurrency Control – Controle de Concorrência Multiversões), melhorias no SQL e novos tipos de dados nativos (novos tipos de datas e hora e tipos geométricos).

Em maio de 2000 foi liberada a versão 7.0. As versões 7.x trouxeram as seguintes novas funcionalidades: Write-Ahead Log (WAL), esquemas SQL, outer joins, suporte a IPv6, indexação por texto, suporte melhorado a SSL e informações estatísticas do banco de dados.

A versão 8.0 foi lançada em janeiro de 2005 e entre outras novidades, foi a primeira a ter suporte nativo para Microsoft Windows (tradicionalmente, o PostgreSQL só rodava de forma nativa em sistemas Unix e, em sistemas Windows - através da biblioteca Cygwin). Dentre as muitas novidades da versão 8.x, pode-se destacar o suporte a tablespaces, savepoints, point-in-time recovery, roles e Two-Phase Commit (2PC). Em setembro de 2010 foi lançada a versão mais recente: 9.0.

Desenvolvimento do Projeto

O PostgreSQL é um projeto open source coordenado pelo PostgreSQL Global Development Group. Embora as atividades do grupo sejam patrocinadas por diversas organizações de todo o mundo, seu modelo de desenvolvimento é o modelo Bazar (originalmente apresentado em A Cathedral e o Bazar de Eric S. Raymond).

Portanto, o desenvolvimento do PostgreSQL é feito por um grupo de desenvolvedores, em sua maioria voluntários, espalhados por todo o mundo e que se comunicam via Internet. Logo, trata-se, de um projeto direcionado pela comunidade de desenvolvedores e de usuários, a qual qualquer pessoa pode se juntar, bastando se inscrever em listas de discussão e participar delas.

Voluntários interessados em contribuir com o projeto também podem consultar as sugestões de tarefas de desenvolvimento de novas funções e de correções de erros que são publicadas na lista TO-DO ou apresentar suas próprias sugestões. O código desenvolvido é submetido à equipe do projeto que pode aceitá-lo e incluí-lo nas novas versões ou recusá-lo. Voluntários também podem colaborar gerando documentação ou realizando traduções!

As ferramentas utilizadas para o apoio ao desenvolvimento são o sistema de gestão de fontes CVS (Concurrent Version System), listas de discussão, servidor de news e salas de bate-papo (IRC).

Estruturas de Decisão

O Projeto PostgreSQL é desenvolvido e direcionado pela sua comunidade de desenvolvedores e de usuários. Para coordenar o projeto há uma equipe central (core team) composta por sete desenvolvedores e um grupo de committers CVS. O código fornecido por voluntários é avaliado e aceito ou rejeitado pelos committers.

Este modelo de desenvolvimento de software, baseado no modelo Bazar originalmente apresentado em A Cathedral e o Bazar de Eric S. Raymond, possibilita o desenvolvimento de software com qualidade devido, principalmente, a permitir:

- Tratar usuários como parceiros e/ou desenvolvedores. Eles contribuem diretamente com o desenvolvimento do software apresentando os problemas enfrentados, suas necessidades, suas sugestões

de solução e, até mesmo, seu próprio código fonte de solução. Assim, usuários auxiliam de forma pró-ativa nas melhorias e na depuração do software.

- Reutilizar código fonte.
- Lançar rapidamente e frequentemente novas versões. Com uma base ampla de usuários testadores do software, os eventuais problemas são rapidamente identificados e sugestões de solução também aparecem com rapidez.

Estado Atual

Além de doações, o projeto PostgreSQL se sustenta pelo patrocínio de diversas empresas, entre as quais se destacam: Fujitsu, Hub.Org, NTT Group, Red Hat, Skype e SRA.

O software tem adquirido prestígio na comunidade Linux, tendo recebido diversas vezes o prêmio Linux Journal Editor's Choice de melhor sistema de gerenciamento de banco de dados (SGBD).

Uma diversidade de módulos adicionais está disponível através do pgfoundry.

A aceitação do PostgreSQL tem se ampliado para além da comunidade de código aberto. Há entre os seus usuários grandes empresas internacionais, órgãos governamentais de vários países e universidades de prestígio mundial. Existe uma lista dos principais usuários no Brasil e no mundo e, também, há alguns estudos de caso de aplicações que utilizam o PostgreSQL.

Radiografia

Nesta seção são apresentadas medidas e estimativas referentes ao PostgreSQL 8.3.0 obtidos pelo SLOCCount de en:David A. Wheeler. Esta versão de PostgreSQL apresenta mais de 570 mil linhas de código, o que segundo o modelo COCOMO exigiria um esforço estimado de pouco mais de 157 pessoas-ano para o seu desenvolvimento.

Estima-se que o tempo necessário para uma empresa desenvolver um software deste porte é de 3 anos e meio (aproximadamente 43 meses), o que corresponderia a cerca de 43 programadores trabalhando em paralelo. Considerando que o salário médio de um programador nos EUA é de cerca de US\$ 56.286,00 por ano, e multiplicando por 2,4 para considerar custos adicionais além do salário dos programadores, o custo estimado para o desenvolvimento de um software do porte do PostgreSQL 8.3.0 é de US\$ 21.220.945,00.

Estado atual do PostgreSQL	
Página web	http://www.postgresql.org/
Início do projeto	1985
Licença	Licença BSD
Versão atual	9.6.3
Data da versão atual	2017-05-08
Linhas de código fonte	571.821
Esforço estimado de desenvolvimento (pessoa-ano / pessoa-mês)	157,09 / 1.855,10
Estimativa de tempo de desenvolvimento (anos)	3,66
Estimativa de nº de desenvolvedores em paralelo	42,39
Estimativa de custo (US\$)	21.220.945,00

A principal linguagem de programação utilizada no desenvolvimento do PostgreSQL 8.3.0 é o ANSI C. Entretanto, também são utilizadas minoritariamente outras linguagens conforme mostra a tabela.

Linguagens de Programação utilizadas no PostgreSQL 8.3.0		
Linguagem	Linhas de código	%
ANSI C	541.312	94,66
yacc	15.527	2,72
lex	5.580	0,98
sh	5.412	0,95
perl	3.913	0,68
asm	65	0,01
python	12	0,00

[illegible]