

PostgreSQL

GUIA DO PROGRAMADOR

André Milani

Novatec



CAPÍTULO 1

Bem-vindo ao PostgreSQL

Caro leitor, seja bem-vindo ao mundo do PostgreSQL, um excelente banco de dados com todas as características e propriedades necessárias para atender aos mais exigentes padrões de aplicações do mundo da informática.

O objetivo deste livro é capacitar o leitor para que possa utilizar e administrar o servidor de banco de dados PostgreSQL da melhor forma possível, sabendo utilizá-lo da forma correta de acordo com cada situação, aumentando a segurança das aplicações, otimizando recursos e reduzindo custos com sua integração a outras aplicações.

Neste capítulo inicial será apresentada uma introdução ao PostgreSQL, passando por sua história, licença de uso, características, comparações com outros bancos de dados, casos de uso, até o que se espera desta ferramenta futuramente.

1.1 O que é o PostgreSQL?

O PostgreSQL é um Sistema Gerenciador de Banco de Dados (SGBD) Relacional, utilizado para armazenar informações de soluções de informática em todas as áreas de negócios existentes, bem como administrar o acesso a estas informações.

As soluções de informática estão presentes em diversas áreas de negócios, desde a própria tecnologia da informação, bem como em áreas como setores aéreo, bancário, governamental, automobilístico etc. Toda e qualquer aplicação que armazene dados para uso ou acesso posterior está integrada com um banco de dados de alguma forma, seja armazenamento na memória, arquivos de textos, ou armazenamento em tabelas. Esta é uma das propostas do PostgreSQL.

Apenas o armazenamento de informações não torna o mecanismo utilizado em um Sistema Gerenciador de Banco de Dados (SGBD). Se todas as informações de uma solução forem armazenadas em um único arquivo de texto, qualquer pessoa que tenha acesso de leitura a este arquivo poderá consultá-lo por inteiro, e não apenas a informação que desejar, violando, assim, as leis de sigilo envolvidas nas aplicações. O mesmo ocorre se a pessoa tiver acesso de escrita no arquivo de texto em questão, pois poderá alterar qualquer informação, mesmo a que não lhe pertence. Um SGBD deve controlar além do armazenamento dos dados, assim como seu acesso – quem pode ler cada dado e quem pode alterar cada informação.

Utilizando os conceitos mais básicos, o PostgreSQL é uma ferramenta encarregada de armazenar dados e gerenciar o acesso de cada informação de acordo com regras previamente definidas.

1.1.1 História do PostgreSQL

A ferramenta conhecida atualmente como PostgreSQL teve origem em um projeto chamado de POSTGRES na Universidade Berkeley, na Califórnia (EUA), em 1986. Uma equipe orientada pelo professor Michael Stonebraker foi designada para criar o modelo e as regras de um novo sistema de armazenamento de dados, com o apoio de diversos órgãos, entre eles o Army Research Office (ARO) e o National Science Foundation (NSF).

A primeira versão de demonstração do projeto ficou pronta em 1987, e logo em seguida foi apresentada em diversos meios. Em 1989, a primeira versão estável foi lançada, com sucessivos lançamentos anuais de novas versões com diversas correções de bugs. Em meados de 1991, seu código foi adquirido pela empresa Illustra Information Technologies, a qual se fundiu com a Informix, hoje pertencente à IBM, e utilizado como gerenciador de banco de dados em um importante projeto científico. Vale a pena comentar que a empresa Informix citada era de propriedade do professor Michael Stonebraker, orientador do projeto POSTGRES, e que a mesma foi comprada por US\$ 1 bilhão pela IBM em 2001.

A primeira grande mudança no projeto POSTGRES ocorreu em 1994. Devido à crescente popularidade que a ferramenta estava adquirindo, o projeto foi encerrado, dando origem ao programa Postgres95, que trouxe uma grande vantagem em sua primeira versão: a incorporação da linguagem SQL, pelos desenvolvedores Andrew Yu e Jolly Chen, substituindo a linguagem PostQUEL anteriormente utilizada. Na época também o programa foi totalmente compatibilizado com o padrão ANSI C, tornando-o portátil para mais de uma plataforma, entre outras diversas melhorias, tornando-o um ícone entre os bancos de dados.

No ano de 1996, novas melhorias surgiram, e o nome Postgres95 já estava desatualizado. Novamente a denominação da ferramenta foi mudada, desta vez para o nome como a conhecemos hoje: PostgreSQL. Atualmente, o SGBD encontra-se em sua versão 8.2 estável, contendo todas as principais características que um SGBD pode disponibilizar.

1.1.2 Licença de uso

Ao contrário da maioria dos softwares livres existentes no mercado, o PostgreSQL não utiliza a licença GNU para regularizar a sua utilização, mas, sim, a licença BSD (Berkeley Software Distribution).

Originada juntamente com o sistema operacional FreeBSD, a licença BSD obteve reconhecimento e vários outros softwares atualmente também a utilizam graças à sua liberalidade perante outras licenças de software.

Como é possível notar, tanto o PostgreSQL quanto a licença BSD tiveram origem no mesmo local, a Universidade de Berkeley, na Califórnia. Este é o primeiro fator que faz com que o PostgreSQL utilize esta licença, pois os interesses iniciais da ferramenta e da licença tinham algo em comum. Mesmo com o código adquirido mais tarde, sua licença BSD foi mantida e é utilizada até hoje, sendo atualizada e revisada periodicamente.

A licença BSD possui inúmeras vantagens sobre a licença GNU, ou melhor, inúmeras restrições a menos do que as impostas pela outra licença. Isto torna o código muito mais acessível para diversos tipos de utilizações, incluindo a livre utilização da ferramenta até mesmo para fins comerciais.

Por várias vezes já foi indagado o porquê de a licença BSD não ser migrada para GNU no caso do PostgreSQL, e a resposta mais ouvida foi a que a empresa mantenedora do PostgreSQL gosta e prefere manter as coisas como estão.

1.2 Utilizações recomendadas

O PostgreSQL encontra-se em uma versão perfeitamente estável e confiável, com os principais recursos existentes nos bancos de dados pagos disponíveis no mercado. Suas capacidades são para suprir as necessidades de pequenas, médias e grandes aplicações.

O PostgreSQL não tem limite de tamanho para seus bancos de dados, sendo a única limitação para tal critério o hardware disponível pelo computador em que o PostgreSQL está armazenando suas informações. Sua limitação dá-se em nível de

tabela, com um limite máximo de 32TB por tabela. Além do mais, é possível ter registros (linhas) com até 1.6TB, campos com até 1GB, tabelas com até 1.600 campos e índices ilimitados para aceleração da busca dos resultados.

A estabilidade do PostgreSQL também é um de seus recursos mais interessantes, pois foi projetado para executar no método 24/7 (24 horas por dia, sete dias na semana). Em outras palavras, executar indefinidamente. Assim, este banco de dados tem sido bastante utilizado também na internet, seja em sites, portais, lojas virtuais ou soluções em informática.

1.3 Compatibilidades

A seguir, algumas informações gerais sobre a compatibilidade do PostgreSQL no que se diz respeito a sistemas operacionais que o suportam, linguagens de programação, plataformas de desenvolvimento e versão de SQL utilizada.

Para os programadores e desenvolvedores há bibliotecas e drivers de conexão para o PostgreSQL para as principais plataformas e linguagens utilizadas, podendo-se citar as seguintes: C/C++, Java/JSP, PHP, ASP, .NET, Perl, Python, Ruby, Tcl e driver ODBC, entre outros.

Sobre ambientes de instalação, o PostgreSQL é uma ferramenta extremamente portátil, disponibilizando instalações para diversos sistemas operacionais, como, por exemplo:

- Linux (Fedora Core, Debian, SuSE, RedHat).
- Unix (BSD, Solaris, HP-UX, AIX).
- Max OS X Server.
- Windows (2000, 2003, XP).

Em se tratando de compatibilidade com a linguagem SQL, o PostgreSQL já conta com vários recursos implementados da versão ANSI SQL 2003, sendo o primeiro banco de dados a implementar algumas das mais recentes definições na história do SQL como Gin e DTrace.

1.4 Características

A seguir, algumas das principais características e recursos existentes no PostgreSQL.

1.4.1 SGBD relacional com suporte a ACID (transações)

O PostgreSQL é um SGBD relacional completo, com suporte a operações ACID (Atomicidade, Consistência, Isolamento e Durabilidade). Cada uma destas propriedades garante uma parte da qualidade dos serviços disponibilizados pelos bancos de dados.

No que se diz respeito à integridade referencial, o PostgreSQL possui e valida este tipo de operação em seus relacionamentos armazenados em tabelas. Tanto os conceitos das propriedades ACID quanto de Integridade Referencial estão disponíveis para leitura no Apêndice A.

1.4.2 Replicação

Assim como os demais bancos de dados, o PostgreSQL oferece os recursos necessários para realizar a replicação entre servidores. A vantagem sobre outros bancos de dados é que sua licença abre este recurso para uso gratuito até para aplicações comerciais, diferentemente de outras licenças de software livre utilizadas por outros bancos de dados.

1.4.3 Cluster (alta disponibilidade)

Visando expandir sua capacidade para mais de um servidor (hardware) decorrente das limitações de processamento, é possível configurar o PostgreSQL para que atue como um cluster de informações.

A utilização de cluster envolve o uso de dois ou mais computadores, interligados e sincronizados entre si, para que ambos possam atender às demandas vindas dos usuários da aplicação ou banco de dados em questão, na teoria dobrando a capacidade de utilização.

1.4.4 Multithreads

O PostgreSQL gerencia várias conexões com o banco de dados de uma única vez, por meio do recurso de multithread oferecido pelos sistemas operacionais. Desta forma, mais de uma pessoa pode acessar a mesma informação sem ocasionar atrasos ou filas de acesso.

Algumas operações forçam o uso de filas de acesso aos dados, principalmente aquelas em que mais de um usuário está tentando realizar um acesso de gravação nos mesmos dados. O PostgreSQL administra estas situações para que os dados não sejam corrompidos.

1.4.5 Segurança SSL e criptografia

O suporte nativo à SSL já está embutido no PostgreSQL, possibilitando criar conexões seguras a partir destes canais, tanto para trafegar informações de login quanto aquelas consideradas sigilosas.

Ainda, o PostgreSQL fornece extensibilidade para utilização de algoritmos de criptografia como o SHA1 e o MD5 (já nativo em suas últimas versões).

1.4.6 SQL

Baseado nos padrões estabelecidos pelo ANSI SQL, o PostgreSQL adota este critério na implementação de suas funcionalidades.

As mais recentes melhorias do PostgreSQL basearam-se na versão ANSI SQL 2003, e novos recursos continuam sendo desenvolvidos, sendo grande parte de suas funcionalidades em conformidade com as versões 92 e 99 do ANSI SQL. Vale a pena destacar a impossibilidade de algum banco de dados conseguir implementar imediatamente uma nova versão ANSI SQL tão logo a mesma seja lançada. Uma outra importante observação é que o PostgreSQL foi pioneiro na implantação de alguns recursos (Gin e DTrace), na frente de todos os demais bancos de dados.

1.4.7 Incorporável em aplicações gratuitamente

Por utilizar a licença de uso BSD, o PostgreSQL pode livremente ser incorporado por aplicações pessoais e/ou comerciais, sem nenhum custo para o desenvolvedor ou fornecedor do software em questão.

Muitas micro e pequenas empresas, assim como algumas médias e grandes organizações, têm migrado suas plataformas de desenvolvimento para o software livre, até mesmo as corporações do ramo de tecnologia da informação. Assim, é possível conciliar o alto poder de processamento do PostgreSQL e seu custo inexistente de licenças para desenvolver softwares aos seus clientes.

1.4.8 Capacidade de armazenamento

O PostgreSQL suporta de forma eficiente e confiável grandes tamanhos de informações em suas tabelas, algumas, inclusive, maiores do que a capacidade de tamanho de arquivo fornecida por certos sistemas operacionais.

Um breve resumo de sua capacidade de armazenamento pode ser conferido a seguir:

- Tamanho máximo de um banco de dados – ilimitado.
- Tamanho máximo de uma tabela – 32 TB.
- Tamanho máximo de uma linha – 1,6 TB.
- Tamanho máximo de um campo – 1 GB.
- Tamanho máximo de linhas por tabela – ilimitado.
- Tamanho máximo de colunas por tabela – de 250 a 1.600, dependendo dos tipos de dados utilizados.
- Tamanho máximo de índices por tabela – ilimitado.

1.5 Novidades na versão 8.2

A cada nova versão, o PostgreSQL conta com a implementação de novas melhorias e funcionalidades, melhorando ainda mais a sua qualidade. Dentre as principais novidades lançadas na versão 8.2, destacamos:

- Maior controle na criação e exclusão de objetos.
- Sintaxe SQL com multivalores para o parâmetro `VALUES`, otimizando as operações.
- Novas otimizações para várias operações SQL.
- Melhoramentos nos vetores (arrays) de dados, incluindo o valor `null` para estes.
- Aumento da performance do algoritmo de ordenação de dados.
- Criação de índices sem o bloqueio de operações.
- Novas funcionalidades de monitoramento, log e tuning de SQL.
- Maior flexibilidade na definição de herança no relacionamento entre tabelas.

O comando `COPY TO` pode ser realizado a partir de um comando `SELECT`.

1.6 Comparação com outros bancos de dados

A comparação entre as diferentes alternativas de bancos de dados é importante para saber quais recursos irão oferecer melhores resultados para cada projeto, ou para a empresa, no caso de corporações que optam por um único servidor de banco de dados, de forma a escolher a que melhor atenda à maioria dos projetos - ou os mais importantes.

Não faz parte do escopo deste livro detalhar as etapas da escolha de um banco de dados, pois assume-se que, para o leitor, esta decisão já tenha sido tomada pelo PostgreSQL. Contudo, cabe a este tópico indicar as principais vantagens que o PostgreSQL pode oferecer, bem como suas limitações e sugestões de soluções para as mesmas.

Os primeiros bancos a serem comparados com o PostgreSQL são as outras alternativas em software livre para bancos de dados. Dentre as opções mais populares e utilizadas, podemos citar o MySQL e o FirebirdSQL. Trata-se de três opções que podem ser excelentes para o seu projeto. Existe uma preferência para PostgreSQL e MySQL quando há uma necessidade maior de estabilidade (anos de mercado), pois ambos estão entre os mais utilizados, sendo ainda que estes dois bancos de dados oferecem mais funcionalidades decorrente do seu tempo de existência e desenvolvimento.

Para algumas operações, o MySQL pode apresentar um resultado melhor. Na maioria das operações para o usuário final de uma aplicação, esta diferença de tempo é praticamente imperceptível. Nas operações de carregamento do banco ou processamento em lote de informações, quando se trata de milhões de registros, que esta diferença pode se tornar visível. Neste caso, o melhor seria comparar as operações em questão utilizando os dois bancos de dados para avaliar os resultados, ressaltando que tal diferença só aparecerá para operações que envolvam uma enorme quantidade de informações.

Outro fator importante na comparação entre os bancos de dados livres são suas licenças. Neste ponto é que o PostgreSQL se destaca, pois ele apresenta a licença mais flexível, permitindo que o mesmo seja integrado em aplicações comerciais sem restrição, sem a necessidade de manter o código de sua aplicação aberto ou outros pré-requisitos. Apesar de muitas pessoas comentaram que o MySQL também é assim, quem conhece ambas as licenças sabe que o MySQL não pode ser utilizado em aplicações comerciais sem que algumas condições sejam atendidas.

Na comparação do PostgreSQL com outros bancos de dados pagos, podemos citar o Microsoft SQL-Server e o Oracle. Já ultrapassando a questão do preço, que não existe para o PostgreSQL, abordaremos a das funcionalidades. Os três bancos de dados possuem os principais recursos como replicação, uso de transações ACID, segurança e ferramentas para administração. Existe uma tendência de uso para o PostgreSQL e Microsoft SQL-Server em termos de praticidade de uso (a porcentagem de mercado que utiliza Oracle nos servidores de hospedagem mais utilizados no Brasil é praticamente nula, porém sua popularidade aumenta no mercado privado de TI).

Existe ainda um preconceito contra o uso de software livre em algumas empresas, principalmente pelo fato de que os programas pagos possuem contratos com estas empresas, podendo a mesma culpar o fornecedor e ser ressarcida de alguma forma, caso algum problema venha a ocorrer na ferramenta em si. Ao mesmo tempo, as ferramentas livres estão tão estáveis quanto as alternativas pagas, até pelo fato de seu código ser revisado por dezenas de colaboradores ao redor do mundo, e não restrito a uma equipe de poucos funcionários. Por este motivo, o software livre está cada vez mais tomando o espaço das ferramentas pagas. Vale a pena lembrar que o papel do DBA (Administrador de Banco de Dados) é fundamental para a boa execução de qualquer servidor de banco de dados.

✦ **Importante:** Ao se compararem dois ou mais bancos de dados, é importante levar em conta os aspectos técnicos que cada aplicação ou servidor necessita. Não são todos os casos em que as aplicações utilizarão tabelas de tamanho superior a centenas de terabytes, ou que utilizarão milhões de colunas em uma única tabela. Portanto, é necessário avaliar todas as vantagens e desvantagens das alternativas existentes no mercado antes de decidir qual a melhor para o seu caso.

1.7 Casos de uso

Dentre os principais casos de uso divulgados pelo site do PostgreSQL, pode-se citar o da BASF (Divisão de Produtos para Agricultura na América do Norte). A empresa em questão atinge vendas no valor de US\$ 29 bilhões anuais, com aproximadamente 92.000 empregados ao redor do mundo.

O objetivo da BASF era vender seus produtos na internet, transformando-se na líder em seu segmento. Para isto, criou um projeto de vendas on-line que utiliza o PostgreSQL como banco de dados. O projeto tem a capacidade de atender até 25.000 usuários simultâneos, com tempo de resposta para os usuários finais entre 3 a 5 segundos (levando-se em conta ainda outras tecnologias que utilizam tempo como o servidor web).

Neste sentido, foi criada toda a estrutura com hardwares e softwares configurados visando à otimização de recursos, e o PostgreSQL escolhido pela BASF em decorrência sua licença livre, em sua tecnologia superior, manutenção e suporte. Dentre os requisitos técnicos do projeto, estava o uso de banco de dados relacional com suporte a transações ACID, stored procedures, triggers (gatilhos) e replicação, entre outros. O projeto foi concluído com todas as necessidades atendidas. Vale a pena citar que entre as demais opções estudadas pela BASF encontravam-se Microsoft SQL-Server, Oracle, MySQL, Firebird, IBM Universal Database e SAP DB.

Existem vários casos de uso descritos na página do PostgreSQL. Para conhecer outros grandes usuários desta ferramenta, acesse:

- **Estudos de caso**

<http://www.postgresql.org/about/casestudies>

- **Usuários**

<http://www.postgresql.org/about/users>

1.8 O que esperar das próximas versões do PostgreSQL

Basicamente, os principais objetivos das próximas versões do PostgreSQL estão em cada vez estar mais padronizado de acordo com as mais recentes regras divulgadas pelo ANSI SQL, implementando novas funcionalidades, além de melhorar a performance a cada nova versão, visando aproveitar ao máximo o que cada hardware e sistema operacional podem oferecer.

1.9 Sites relacionados

- **Site oficial**

<http://www.postgresql.org>

- **Site de ajuda (foruns, lista de discussão, canais IRC)**

<http://www.postgresql.org/support>

- **Site do PostgreSQL Brasil (Comunidade Brasileira)**

<http://www.postgresql.org.br>

- **Site com informações da Licença de uso BSD**

<http://www.opensource.org/licenses/bsd-license.php>