

Tudo o que você queria saber sobre PostgreSQL mas tinha vergonha de perguntar

Euler Taveira de Oliveira

Timbira - A empresa brasileira de PostgreSQL

04 de novembro de 2011

timbira

Introdução

Perguntas

Conclusão

Sobre esta apresentação

- ▶ esta apresentação está disponível em:
<http://www.timbira.com.br/material>
- ▶ esta apresentação está sob licença *Creative Commons Atribuição 3.0 Brasil*:
<http://creativecommons.org/licenses/by/3.0/br>

O que é?

- ▶ perguntas mais frequentes
- ▶ curiosidades
- ▶ conceitos de bancos de dados
- ▶ como fazer
- ▶ história

O que não é?

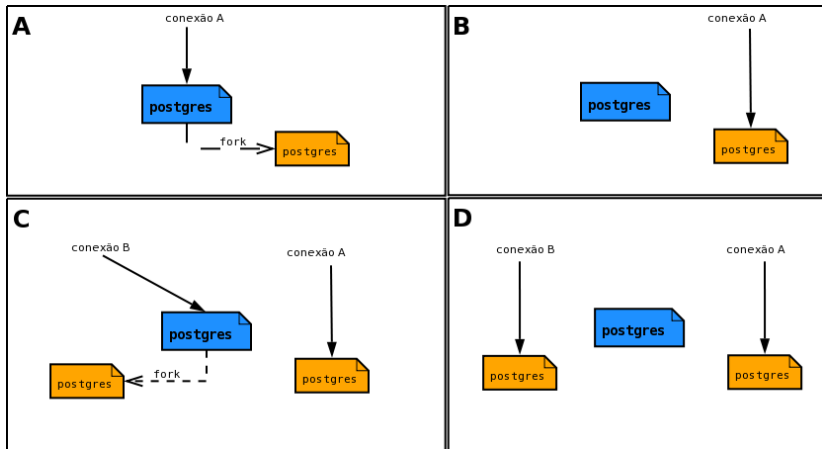
- ▶ tópicos avançados
- ▶ modelagem
- ▶ consultas SQL
- ▶ comparação com outros SGBDs
- ▶ sistema operacional
- ▶ hardware

Introdução

Perguntas

Conclusão

Como é o processo de conexão no PostgreSQL?



Por que o PostgreSQL não inicia?

- ▶ no arquivo de log ou na saída de erro padrão (stderr):

Mensagem de erro

FATAL: não pôde criar segmento de memória compartilhada:

Argumento inválido

DETALHE: Falhou ao executar chamada de sistema

shmget(key=9100001, size=**2217328640**, 03600).

Solução

```
# sysctl kernel.shmmax kernel.shmall
```

```
kernel.shmmax = 33554432
```

```
kernel.shmall = 2097152
```

```
# sysctl -w kernel.shmmax=2217328640
```

```
kernel.shmmax = 2217328640
```

```
# sysctl -w kernel.shmall=2217328640
```

```
kernel.shmall = 2217328640
```

Por que o PostgreSQL não conecta?

- ▶ no arquivo de log ou na saída de erro padrão (stderr):

Mensagem de erro

FATAL: Peer authentication failed for user "**foo**"

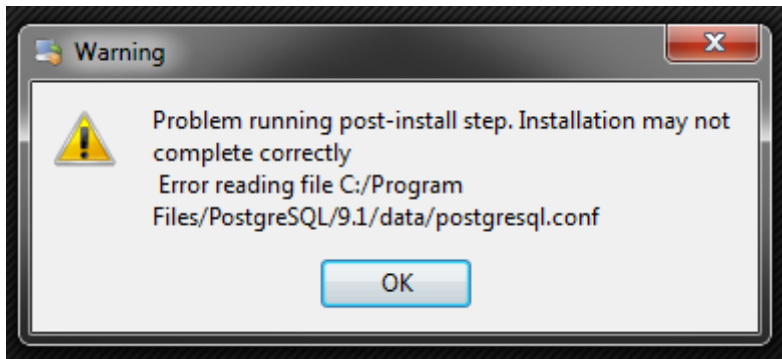
...

FATAL: nenhuma entrada no pg_hba.conf para máquina "**192.10.30.1**", usuário "**foo**", banco de dados "**pgbr2011**"

Solução

- ▶ editar \$PGDATA/pg_hba.conf
- ▶ primeira regra que combina é utilizada
- ▶ "peer authentication" = usuário do SO + conexão local

Por que o PostgreSQL não instala (no Windows)?



Por que o PostgreSQL não instala (no Windows)?

- ▶ difícil testar em todos os sabores deste sistema operacional
- ▶ atualmente 4 desenvolvedores e 2 QAs para instalador (EnterpriseDB)
 - ▶ **Windows:** XP, Vista, 7, 8(?), 2003, 2003R2, 2008, 2008R2
 - ▶ **Arquitetura:** 32 bits, 64 bits
 - ▶ **Modelo:** Basic, Home, Professional, Business, Enterprise, Ultimate, Embedded (desktop); Standard, Enterprise, Datacenter (server)
 - ▶ **Rede:** domínio, standalone
 - ▶ **Conta:** admin local, usuário comum, admin UAC, admin domínio
 - ▶ **Correção:** muitos *service packs*
 - ▶ **Antivírus:** vários
 - ▶ **Group Policy:** inúmeras possibilidades

Por que o PostgreSQL não instala (no Windows)?

- ▶ é uma versão do SO suportada?
- ▶ tente sem (desabilitar geralmente não adianta) antivírus e firewall
- ▶ usuários
 - ▶ **super-usuário**: usuário como todas as permissões no PostgreSQL
 - ▶ **serviço**: usuário do SO que executa o PostgreSQL
 - ▶ **net user postgres /delete**: remove o usuário do serviço
- ▶ verificar o erro em %TEMP%/bitrock_installer*.log
- ▶ relatar **detalhadamente** na lista de discussão

Relatando um erro de instalação no Windows

- ▶ a mensagem exata vista quando instalador falhou
- ▶ a versão exata do PostgreSQL
- ▶ instalador 32 bits ou 64 bits?
- ▶ sistema operacional e versão (i.e., “Windows XP Professional with Service Pack 3”)
- ▶ sistema operacional 32 bits ou 64 bits?
- ▶ como executou o instalador? linha de comando? usuário?
- ▶ que antivírus e firewall tem instalado?
- ▶ o que você já tentou fazer para solucionar o problema?
- ▶ há uma versão anterior do PostgreSQL instalada? Qual?
- ▶ informações adicionais

O que é memória compartilhada?

- ▶ é um método de comunicação entre processos
- ▶ permite o uso de informação por diferentes processos na mesma máquina

O que são shared buffers?

- ▶ é a cache do PostgreSQL
- ▶ a descarga da cache é feita em cada CHECKPOINT

O que é WAL?

- ▶ “Write Ahead Logging” (Registro de Escrita Prévia)
- ▶ método padrão para garantir integridade (ACID)
- ▶ mudança nos arquivos de dados só ocorrem após serem registradas
- ▶ não precisa descarregar mudanças nos arquivos a cada COMMIT
- ▶ em caso de queda do servidor, podemos refazer as transações a partir dos logs

O que é checkpoint?

- ▶ é o ponto no log de transação que todos os arquivos de dados são descarregados no disco
- ▶ ele garante que se houver uma queda, a recuperação começará a partir do último CHECKPOINT
- ▶ o PostgreSQL realiza CHECKPOINTS: por intervalo de tempo ou após preencher x arquivos de log de transação (o que ocorrer primeiro)
- ▶ é uma operação dispendiosa para servidores que possuem uma escrita intensa
- ▶ quanto maior a quantidade de buffers compartilhados, maior pode ser o tempo de CHECKPOINT

Como o PostgreSQL armazena os dados?

- ▶ \$PGDATA: vários subdiretórios
- ▶ \$PGDATA/base: cada diretório é um banco de dados
- ▶ cada tabela, índice e sequência é um ou mais arquivos (1 GB cada)
- ▶ colunas grandes: tabela TOAST
- ▶ \$PGDATA/pg_tblspc: link simbólico para diretório da tablespace
- ▶ \$PGDATA/base/pgsql_tmp: tabelas temporárias

Como o PostgreSQL armazena os dados?

```
$ ls
base      pg_hba.conf      pg_multixact      pg_stat_tmp      pg_twophase      postgresql.conf
global    pg_ident.conf    pg_notify         pg_subtrans      PG_VERSION       postmaster.opts
pg_clog   pg_log           pg_serial         pg_tblspc        pg_xlog          postmaster.pid
$ ls global/
11669      11802           11808            11813            11819            11825           11917
11669_fsm  11804           11809            11815            11820            11826           pg_control
11669_vm   11805           11810            11816            11821            11911           pg_filenode.map
11671      11806           11810_fsm        11817            11821_fsm        11913           pg_internal.init
11672      11806_fsm       11810_vm         11817_fsm        11821_vm         11915
11800      11806_vm        11812            11817_vm         11823            11916
$ ls base/
1 11910 11918 16389
$ ls pg_xlog/
00000001000000000000000001 archive_status
$ ls pg_stat_tmp/
pgstat.stat
$
```

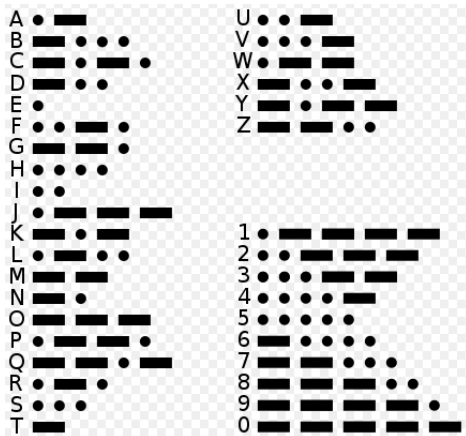
O que é MVCC?

- ▶ **Multiversion Concurrency Control**
- ▶ gerenciar acesso concorrente aos dados
- ▶ cada transação “vê” um instantâneo dos dados (a algum tempo atrás)
- ▶ mesmo se aquele dado estiver sido removido ou modificado
- ▶ esse mecanismo protege a sessão de dados inconsistentes (alterados em transações concorrentes)
- ▶ isolamento de transação para cada sessão do banco de dados
- ▶ minimiza os travamentos do banco de dados causados por acesso concorrente
- ▶ provê performance razoável em ambientes multi-usuários
- ▶ vantagem: leitura não bloqueia escrita e vice-versa
- ▶ outros SGBDs suportam MVCC (ex. Oracle, MS SQL Server, DB2 e MySQL+InnoDB)

O que é locale?

- ▶ configuração regional
- ▶ alfabeto, ordenação, formatação de número, formatação de moeda, zona horária
- ▶ LC_COLLATE: ordenação de *strings*
- ▶ LC_CTYPE: classificação de caracteres
- ▶ LC_MESSAGES: língua das mensagens (i18n)
- ▶ LC_MONETARY: formato de moeda
- ▶ LC_NUMERIC: formato de números
- ▶ LC_TIME: formato de data e hora
- ▶ especificar durante a criação do agrupamento (initdb)
- ▶ altera comportamento de: ORDER BY, upper(), lower(), to_char()

O que é encoding?



código de Morse

timbira

O que é encoding?

מבחינתנו
לפי, יש
רצוננו, עשנו
למחמם, יהיה
מבחינתנו
שם, יש
נענו, עשנו
הענו, יהיה
לפי, עשנו
עשנו, עשנו
נענו, עשנו

hebraico

timbira

O que é encoding?

- ▶ é o processo na qual a informação da fonte é convertida em símbolos para ser apresentada/transmitida
- ▶ no Brasil as codificações mais utilizadas são: UTF8 e Latin1 (ISO-8859-1)
- ▶ UTF8 possui a vantagem de agregar inúmeras línguas
- ▶ UTF8 agrega mais línguas do que o ISO-8859-*

O que é collation?

- ▶ apresentar informação em uma determinada ordem (LC_COLLATE)
- ▶ comportamento de classificação (LC_CTYPE)
- ▶ PostgreSQL suporta collation por coluna
- ▶ collation está “amarrada” a codificação de caracteres (encoding)
- ▶ no Brasil utilizamos: pt_BR ou pt_BR.utf8
- ▶ CREATE COLLATION

O que é autovacuum? E como ele funciona?

- ▶ execução automática dos comandos ANALYZE e VACUUM
- ▶ opcional mas altamente recomendado (≥ 8.3)
- ▶ coletor de estatísticas habilitado
- ▶ processos
 - ▶ *launcher*: 1
 - ▶ *worker*: 0 ou mais
- ▶ *número de tuplas alteradas (e novas) > fator de escala * número de tuplas + limite base*
- ▶ transações antigas em um banco de dados também disparam o autovacuum
- ▶ pode herdar parâmetros do VACUUM
- ▶ parâmetros podem ser definidos por tabela

Como funciona a replicação?

- ▶ recuperação de registros do log de transação no servidor secundário
- ▶ entrega: arquivo ou fluxo (“stream”)
- ▶ no primário: processo *walsender*
- ▶ no secundário: processo *walreceiver*
- ▶ privilégio REPLICATION (≥ 9.1)
- ▶ configuração: `recovery.conf` e `postgresql.conf`
- ▶ monitoramento: `pg_current_xlog_location` (primário) e `pg_last_xlog_{receive, replay}_location` (secundário)

Por que a consulta está lenta?

- ▶ estatísticas desatualizadas (ANALYZE)
- ▶ uso de índices
- ▶ estimativa ruim
- ▶ "deficiência" do planejador
- ▶ alto número de conexões (concorrência)

Solução

- ▶ habilitar autovacuum / aumentar frequência do ANALYZE
- ▶ criar novos índices
- ▶ diminuir número de conexões (pool)
- ▶ procurar o melhor plano (desabilitar parâmetros enable_*)

Quais os parâmetros do postgresql.conf adequados?

- ▶ **não** existe fórmula mágica
- ▶ os parâmetros adequados dependem da
 - ▶ carga de trabalho
 - ▶ tipo de transações
 - ▶ sistema operacional
 - ▶ hardware
- ▶ conhecer
 - ▶ conceitos
 - ▶ arquitetura do PostgreSQL
 - ▶ $\pm 20\%$ dos parâmetros (performance)
 - ▶ aplicações
 - ▶ hardware
 - ▶ sistema operacional

Como faço consulta em 2 bancos de dados distintos?

- ▶ dblink (contrib) – utiliza *foreign data wrapper*
- ▶ executar comandos SQL em outros bancos de dados PostgreSQL
- ▶ pgsql_fdw (?) irá substituir dblink?

É possível fazer consulta em outros SGBDs?

- ▶ dbilink (DBI, Perl)
- ▶ SQL/MED
 - ▶ oracle_fdw
 - ▶ mysql_fdw
 - ▶ odbc_fdw
 - ▶ couchdb_fdw
 - ▶ redis_fdw
 - ▶ file_fdw (contrib)
 - ▶ twitter_fdw
 - ▶ ldap_fdw
 - ▶ ...

http://wiki.postgresql.org/wiki/Foreign_data_wrappers

Como instalar um contrib?

- ▶ apt-get install postgresql-contrib-x.x ou
- ▶ compilar código e instalar contrib
- ▶ versão ≤ 9.0
 - ▶ executar arquivo
`$$SHAREDIR/postgresql/contrib/minhacontrib.sql`
- ▶ versão 9.1
 - ▶ CREATE EXTENSION minhacontrib

Como funciona o particionamento?

- ▶ **não** há sintaxe para criação/manipulação de partições
- ▶ utiliza herança de tabelas e uma técnica de otimização de consulta (`constraint_exclusion`)
- ▶ com `constraint_exclusion` habilitado, o planejador examina as restrições das partições e só inclui no plano aquelas partições que satisfazem a condição
- ▶ tabela pai geralmente está vazia
- ▶ tabela filho contém exatamente a mesma definição da tabela pai (mas possui a restrição)
- ▶ criar gatilho `BEFORE INSERT` para redirecionar dados a partição correta
- ▶ `ANALYZE` tabela-pai não atualiza estatísticas nas tabelas filho
- ▶ `UPDATE?` `DELETE?`

O PostgreSQL suporta hints em consultas?

- ▶ NO hints, please!
- ▶ planejador possui opções para “forçar” a escolha de outro plano
- ▶ planejador possui parâmetros para ajustar custos de operações (busca sequencial/por índice, cpu)
- ▶ PostgreSQL têm interesse em melhorar o otimizador
- ▶ contribua consultas/planos não ótimos (*pgsql-performance*)

Como faço a consulta ignorar acentos, maiúsculas e minúsculas?

- ▶ **não** há uma função ou parâmetro mágico
- ▶ `unaccent()`: remove acentos
- ▶ `lower()`, `upper()`: normalização
- ▶ índices sobre expressão
- ▶ TODO: suportar um collate que faça isso!

Origem do nome PostgreSQL

- ▶ Postgres: nome original do projeto em Berkeley
- ▶ Post (após) + INGRES = Postgres
- ▶ Postgres + SQL
- ▶ SQL: nova linguagem implementada

nova versão

- ▶ altera **x** e **y** ou somente **y** (x.y.z)
- ▶ novas funcionalidades
- ▶ pode alterar funcionalidades existentes
- ▶ atualização somente com cópia de segurança e restauração

versão corretiva

- ▶ altera **z** (x.y.z)
- ▶ **não** altera funcionalidades
- ▶ corrige bugs
- ▶ atualização **não** necessita de cópia de segurança e restauração

- ▶ BSD (qualquer propósito)
 - ▶ venda
 - ▶ versão personalizada com código fechado
- ▶ diferente da licença GPL

- ▶ o PostgreSQL segue a risca o padrão SQL
- ▶ isso **não** quer dizer que não haja suporte a coisas que não estão no padrão

PGDG e o futuro. Pode ser comprado?

- ▶ definitivamente **não**
- ▶ **não** há uma empresa que detém os direitos do código fonte
- ▶ **não** há uma empresa que empregue todos os desenvolvedores
- ▶ funcionalidades são introduzidas se e somente se houver um consenso da comunidade



timbira

Como perguntar na lista?

- ▶ detalhes do problema (versão, SO, parâmetros, consulta, plano, ...)
- ▶ evite *top-posting*
- ▶ evite *off-topic*
- ▶ seja claro
- ▶ conheça as regras da lista

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

Introdução

Perguntas

Conclusão

Outras inúmeras perguntas...

- ▶ A sua pergunta na lista pgbr-`{geral, dev}`
- ▶ A sua pergunta na lista pgsq-`{general, performance, hackers}`
- ▶ histórico das listas
- ▶ blogs
 - ▶ <http://planeta.postgresql.org.br>
 - ▶ <http://planet.postgresql.org>
- ▶ wiki
 - ▶ <http://wiki.postgresql.org>

- ▶ Mesa redonda: "O mercado de serviços em PostgreSQL no Brasil" (*Euler Taveira*)
- ▶ Estripando o Elefante - dividindo seus problemas em problemas menores (*Dickson Guedes*)
- ▶ Fazendo uma manada de elefantes passar por baixo da porta (*Fabio Telles*)
- ▶ Tudo o que você queria saber sobre PostgreSQL mas tinha vergonha de perguntar (*Euler Taveira*)

?

Euler Taveira de Oliveira
euler@timbira.com.br
<http://www.timbira.com.br>