

PostgreSQL para um bilhão de usuários

Fernando Ike de Oliveira

B2BR - Grupo TBA

Setembro de 2008 / PGCon-BR 2008



- 2004 libera um conjunto de ferramentas para replicação, balanceamento de carga e alta-disponibilidade para PostgreSQL.
- Essas ferramentas são conhecidas como PL/Proxy, PgBouncer, Skytools.
- A licença é BSD.
- Instalação à partir do código-fonte ou pacotes *.deb *.rpm.
- Oficialmente o Debian, Fedora tem pacotes binários.

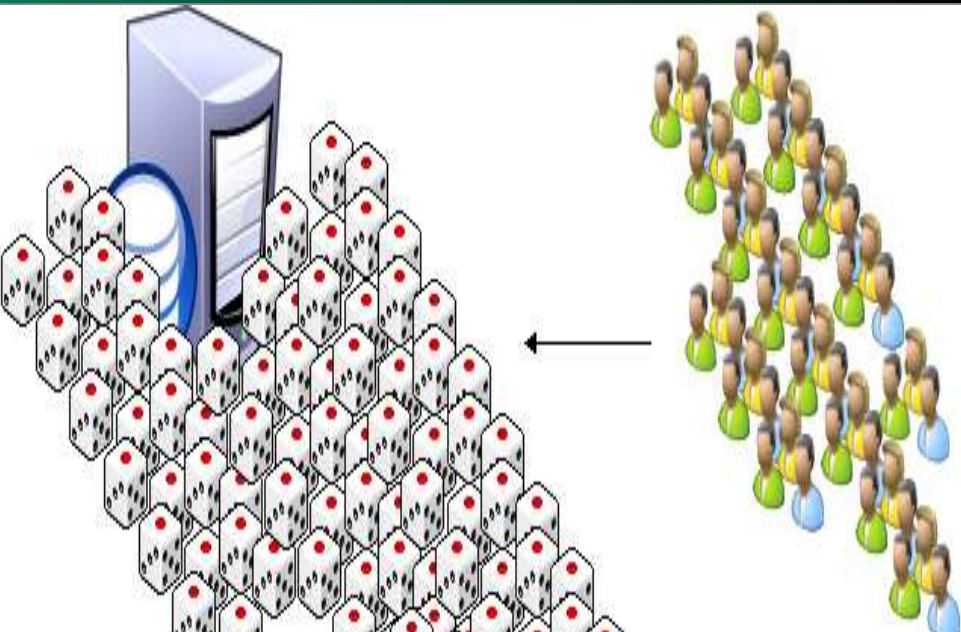
- PL/Proxy é uma linguagem usada para chamadas remotas e particionamento de banco de dados usando hash dos dados.
- PL/Proxy permite criar funções de proxy usando hash para especificar destino (base de dados alvo).
- PL/Proxy é comparável à um roteador de rede que encaminha a expressão SQL para o instância correta.

Exemplo

```
"SELECT (hashtext('pgcon1'))%3" = 0
"SELECT (hashtext('pgcon2'))%3" = -1
"SELECT (hashtext('pgcon3'))%3" = -2
"SELECT (hashtext('pgcon4'))%3" = 0
"SELECT (hashtext('pgcon5'))%3" = 0
"SELECT (hashtext('pgcon6'))%3" = 2
"SELECT (hashtext('pgcon7'))%3" = -1
"SELECT (hashtext('pgcon8'))%3" = 0
"SELECT (hashtext('pgcon9'))%3" = 0
"SELECT (hashtext('pgcon10'))%3" = 0
```

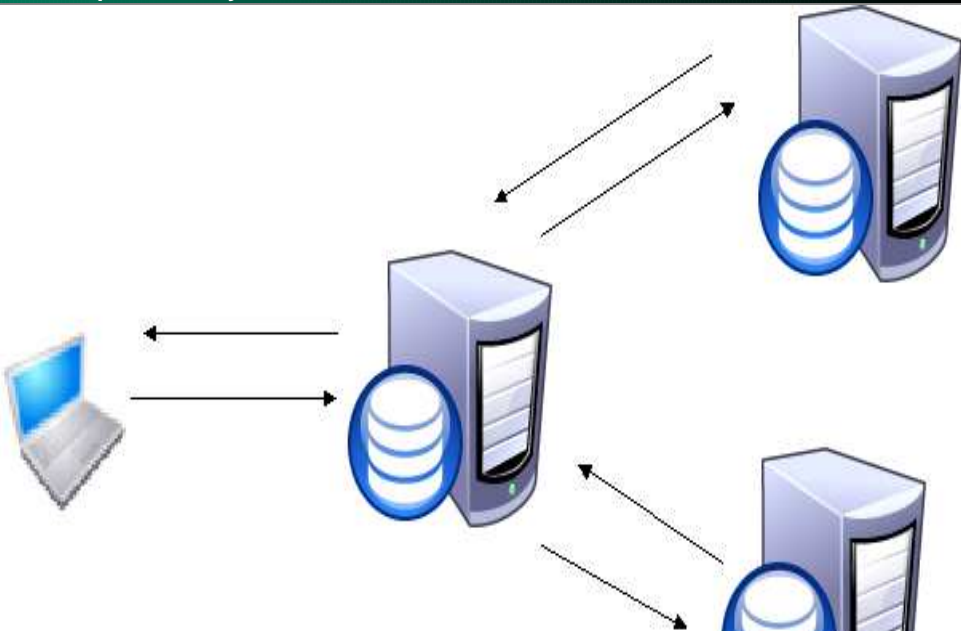
- Como um barramento, para todos as servidores PostgreSQL
- Tabelas
- Particionamento usando funções de Proxy

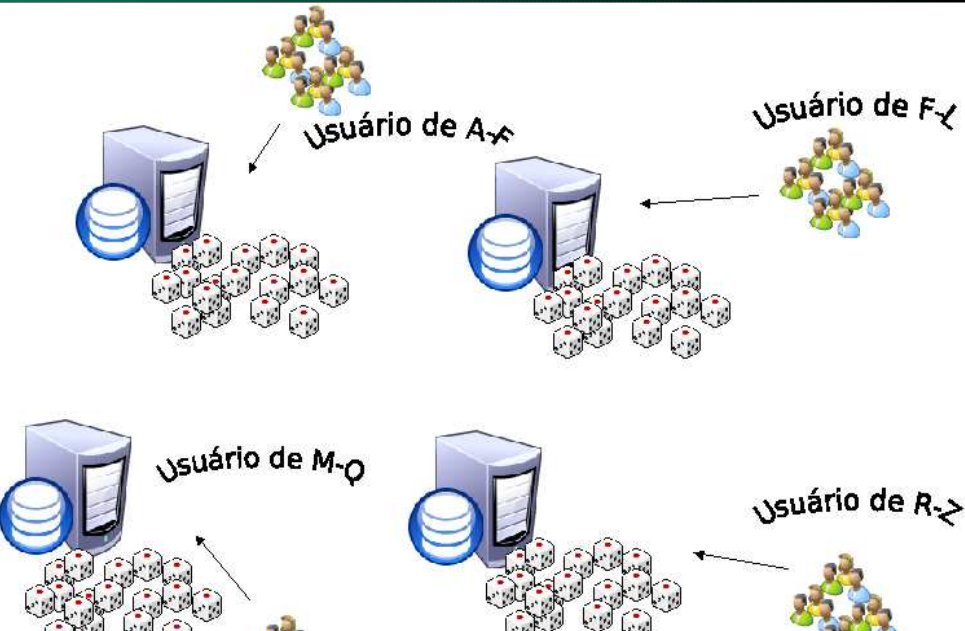
PL/Proxy Problema



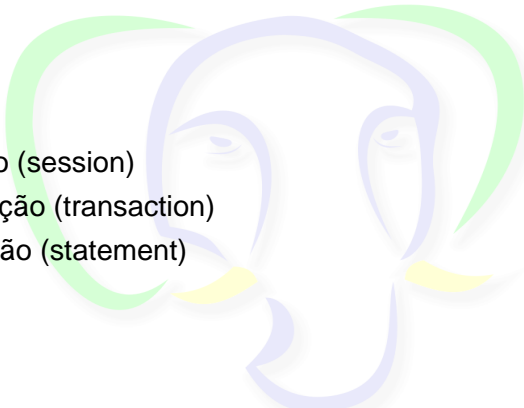
PL/Proxy

Princípio do Proxy





- Baixo consumo de recurso (2k por conexão)
- Suporta reconfiguração sem reiniciar o serviço
- Suporta reinício/atualização sem derrubar a conexão cliente
- Suporta o protocolo v3 ou superior, somente => 7.4
- O Parse SQL não é muito rápido e consome pouco tempo de cpu
- Tem uma interface console de gerenciamento
- Tem estrutura sua (própria) estrutura de autenticação mas similar ao arquivo similar ao arquivo de senha pg_pwd do PostgreSQL.
- Permite autenticação do tipo: trust, texto plano, crypt, md5.

- 
- Sessão (session)
 - Transação (transaction)
 - Instrução (statement)

- Quando um cliente conecta, o PgBouncer retira uma conexão do pool e entrega para a aplicação
- Ao o cliente desconectar, o PgBouncer re-aloca a conexão para o pool. geralmente essa configuração é recomendada para aplicações legadas (Por não usarem de forma eficiente o pool)

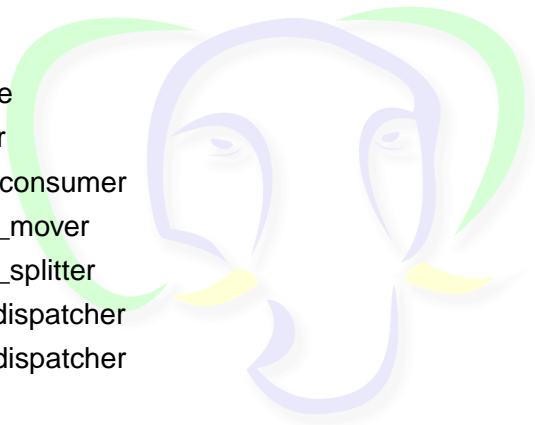
- Servidor mantém para o cliente a conexão somente durante a transação. Quando PgBouncer notifica que transação acabou, o servidor devolve a conexão para o pool.
- Essa opção não deve ser usada com servidor de aplicações que gerenciam pool (Jboss, por exemplo)

- Este é o modo usado usado com o PL/Proxy. Por ser mais agressivo, ele retornar o conexão para o pool depois que a consulta termina.
- Transações muito longas com múltiplos Statements são desabilitados neste modo.

- Skytools são um conjunto de scripts para gerenciar cluster de servidores PostgreSQL.
- Desenvolvido em C e Python
- Permitem usar para replicação assíncrona
- Permitem replicar dados particionados nos servidores "slave(s)"
- Possível extender usando as API do PgQ

- Centralização de log gerenciamento de exceções.
- Gerenciamento de conexões ao banco de dados
- Gerenciamento de configuração
- Gerenciamento de scripts...

- londiste
- walmgr
- serial_consumer
- queue_mover
- queue_splitter
- table_dispatcher
- cube_dispatcher



- ... um sistema de replicação
- ... Master/Slave(s) como tipo de replicação
- ... de replicação é assíncrona
- ... baseada fortemente nas idéias do no Slony-1.
- ... um replicador baseado em gatilhos

- ... uma ferramenta para gerenciar replicação por WAL
- ... similar ao pg_standby (contrib)
- ... possível gerenciar replicação baseada em PITR (Warm Standby).
- ... escrito em python
- ... uma ferramenta de replicação que usa túnel SSH.

Ambos

- Usa o PgQ como transporte dos dados
- Move/copia os dados para os servidores Slave
- Move/copia os dados em lote(batch)
- O processamento é/são nos slave(s)

queue_mover

- Usado para mover/copia os dados para OLTP, Web

queue_splitter

- Usado para BI

Ambos

- Ferramentas para replicar dados em tabelas particionadas.
- Usados para preparar os dados para outros servidores

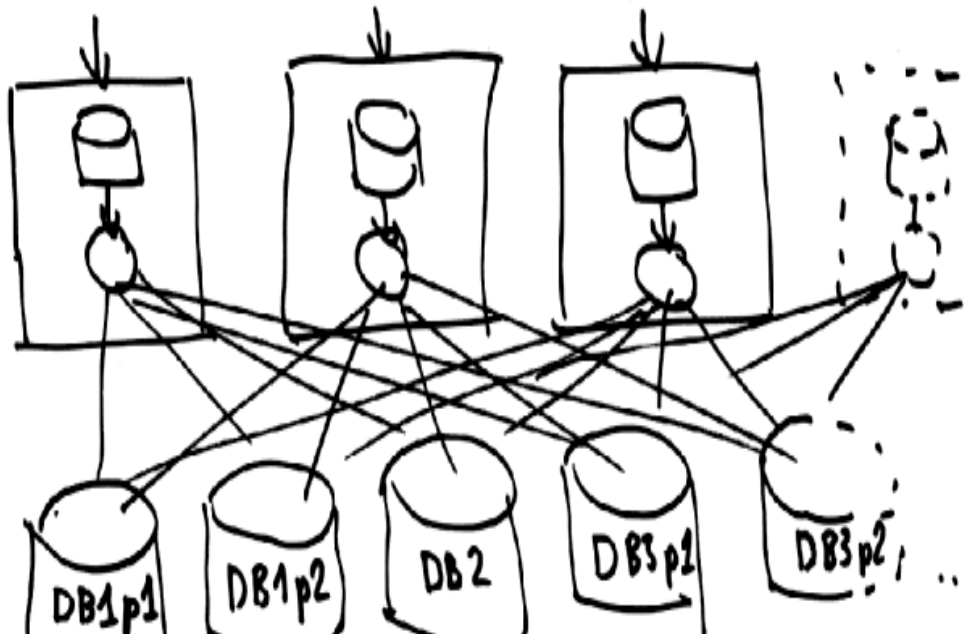
cube_dispatcher

- Prepara os dados para banco de dados do tipo BI, OLAP, Cubo...
- Não tem suporte para operações de remoção de registro
- Caso haja duas versões do mesmo registro, ele irá enviar somente a última versão.

table_dispatcher

- script para configuraro particionamento de uma ou mais tabela.
- possibilita particionar uma tabela por mês, por exemplo.

Exemplo 1



Exemplo 2



Links e Listas de discussão:

- <https://developer.skype.com/SkypeGarage/DbProjects/SkypePostgresqlWhitepaper>
- <https://developer.skype.com/SkypeGarage/DbProjects/PlProxy>
- <http://pgfoundry.org/mailman/listinfo/plproxy-users>
- <https://developer.skype.com/SkypeGarage/DbProjects/PgBouncer>
- <http://pgfoundry.org/mailman/listinfo/pgbouncer-general>
- <https://developer.skype.com/SkypeGarage/DbProjects/SkyTools>
- <http://pgfoundry.org/mailman/listinfo/skytools-users>
- <http://joacosme.wordpress.com/2008/07/03/comecando-com-o-plproxy/>

Contatos:

- `fernando.ike@b2br.com.br`
- `fernando.ike@gmail.com`
- `http://www.midstorm.org/~fike/weblog`

PGCon Brasil 2008

- `http://pgcon.postgresql.org.br`



Imagens

- Joao Comes `http://joacosme.wordpress.com`