PostgreSQLレプリケーション
～pgpool／Slony-Iの運用性とその評価～

SRA OSS, Inc. 日本支社
http://www.sraoss.co.jp/
佐藤 友章
sato@sraoss.co.jp

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

アジェンダ

- はじめに
- レプリケーションとは？
- pgpool／Slony-Iの紹介
- pgpool／Slony-Iの運用性とその評価
- まとめ

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

はじめに

- IPAによる「2006年度OSS活用基盤整備事業」の一環としてOSSの性能・信頼性評価を実施
  - pgpool、Slony-Iの運用性に関する評価もその1つ
  - その他にPostgreSQL、MySQL、JBoss、Geronimoについても評価を実施
- 評価結果はOSS iPediaにて公開中
  - http://ossipedia.ipa.go.jp/
  - 「性能評価」ページの「性能評価情報を探す」から検索

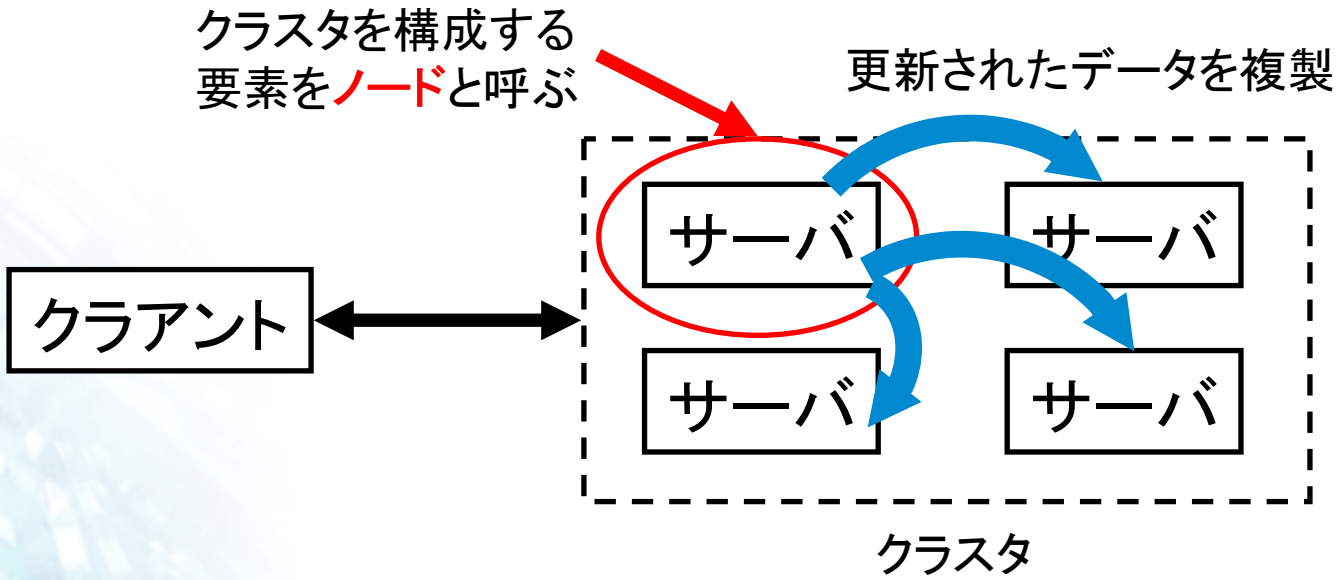
Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

レプリケーションとは？

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

レプリケーションとは？

- レプリケーション（replication）：複製すること
- 複数台のシステムに同じデータを複製する仕組み



Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

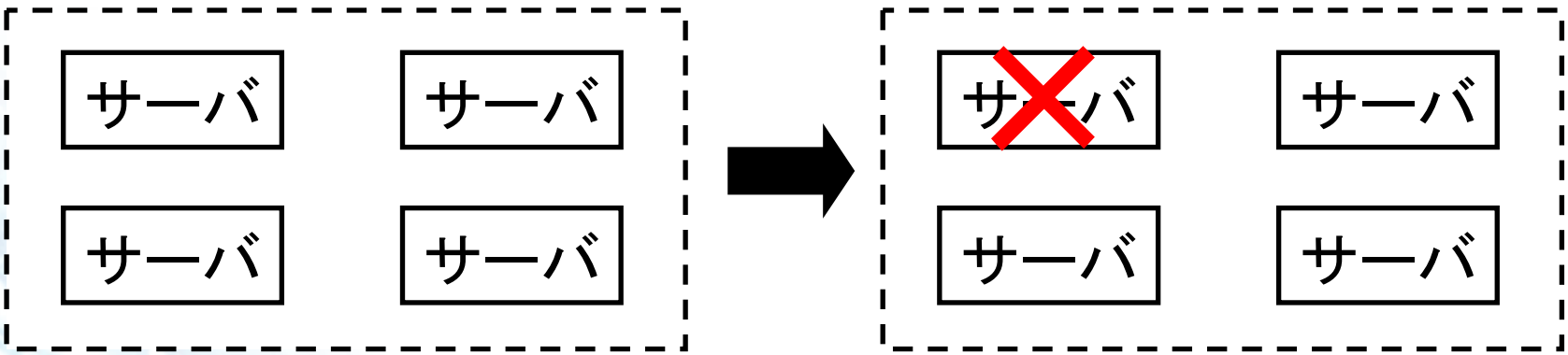
レプリケーションの利点・欠点

- 利点
  - 冗長化
  - 高速化
- 欠点
  - 非常に複雑なため、導入するにはある程度知識が必要
  - 異常が発生した場合の切り分けが大変
  - システム構成の金額がはねあがる

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

冗長化とは？

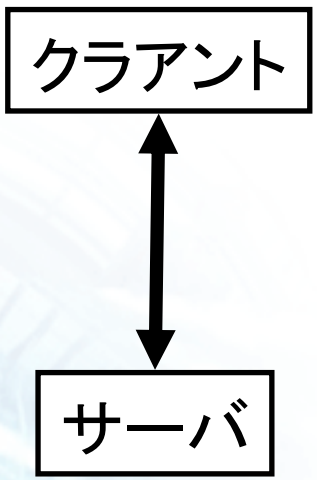
- 1台異常が発生してもサービスを提供し続けることができる



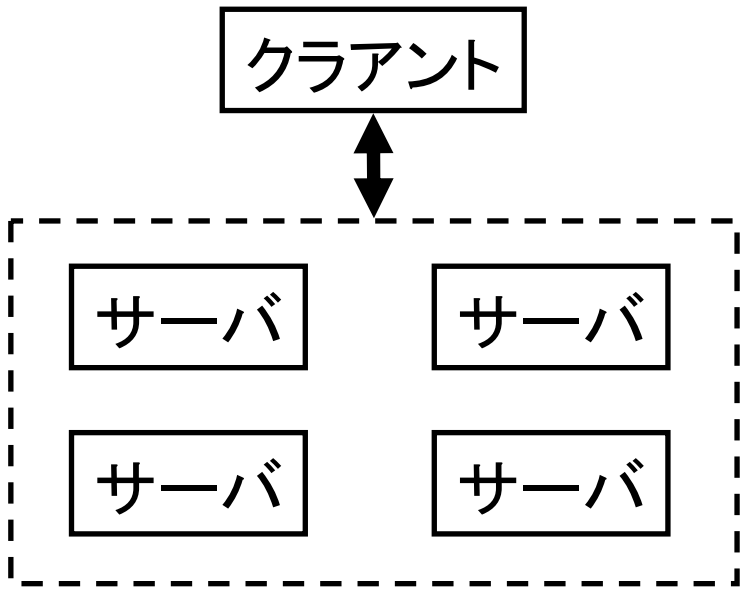
Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

高速化とは？

- 1つの仕事を複数のシステムが行う
  - レプリケーションをすることで検索系クエリの実行を分散



負荷が集中



負荷を分散

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

PostgreSQLのレプリケーション

- PostgreSQL自体にレプリケーションの機能はない
- レプリケーションするためのミドルウェアがいくつかある
  - pgpool
  - Slony-I
  - PGCluster
  - …

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

pgpool／Slony-Iの概要

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

pgpoolとは？

- 2003年開発スタート
  - 石井達夫氏が個人で開発
  - 最初はコネクションプーリング機能のみ(pgpoolという名前の由来はここからきている)
- ライセンス
  - BSDライセンス
  - ソースコードが公開され、自由に改造してもよい
- 現在の開発体制
  - pgpool Global Development Groupという開発団体に移行

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

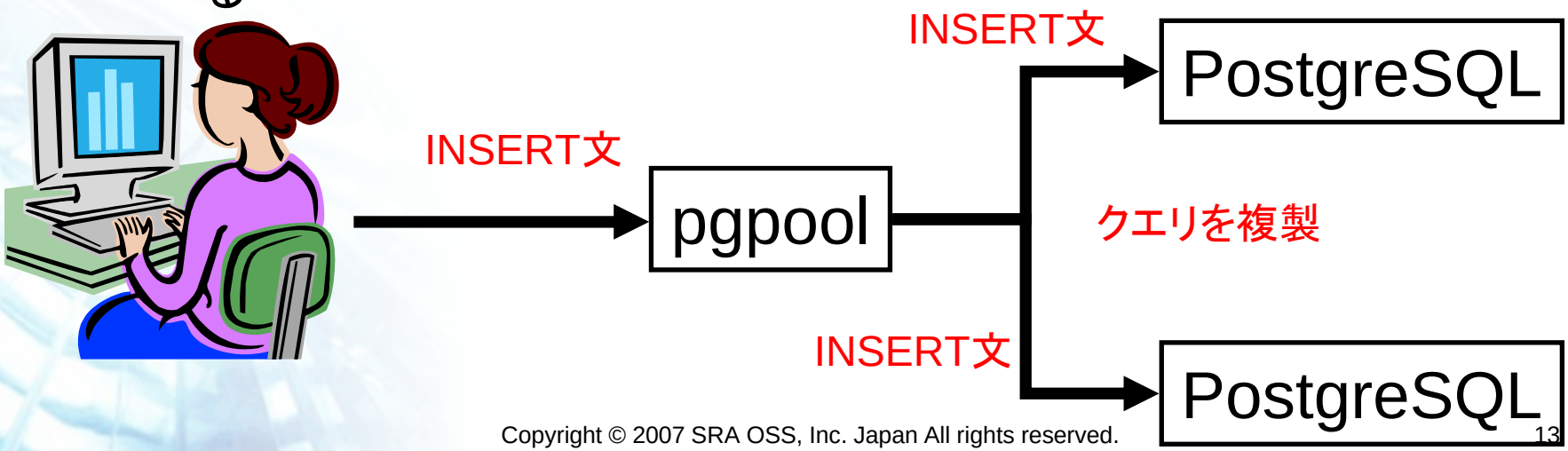
pgpoolの機能

- コネクションプーリング
  - 接続時のオーバーヘッドを軽減
- 同期レプリケーション
  - 2台までの制限あり
- ロードバランス
  - レプリケーション設定時にのみ検索を振り分けることが可能

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

pgpoolのレプリケーション

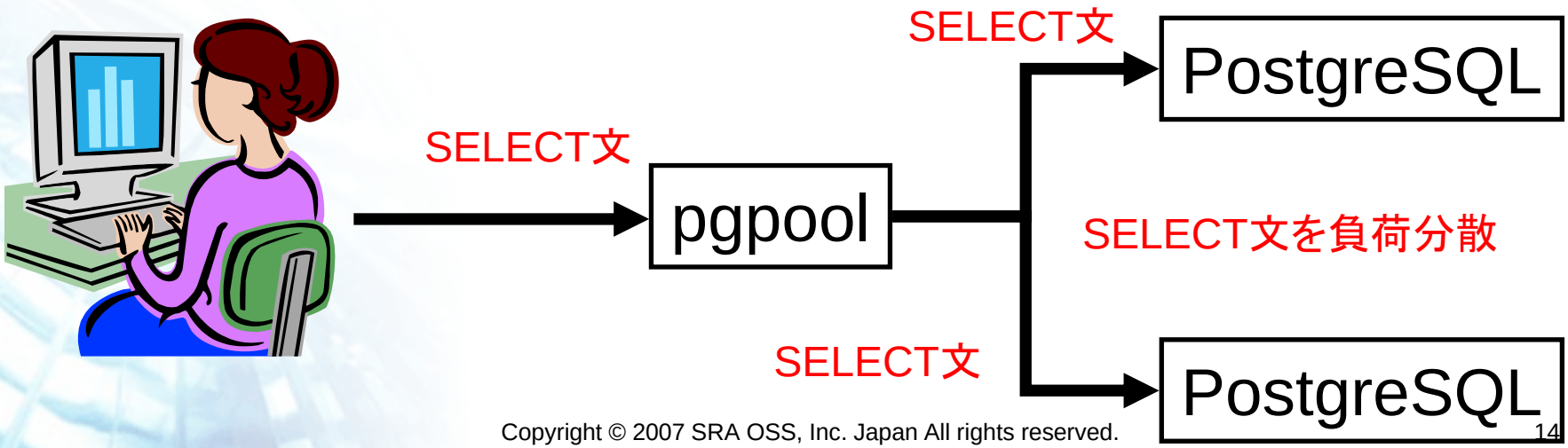
- 2台に更新系クエリを送信することで、データの同期を取る
  - 異常が発生した場合にも運用が可能(縮退運転)
  - 常に同期を取るため、通常より更新系クエリは時間がかかる



Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

pgpoolのレプリケーション

- ロードバランス
  - 同じデータを2台で持つので、検索についてはどちらかに問い合わせすればよいので、負荷分散が可能
  - ロードバランス先の割合を指定可能



Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

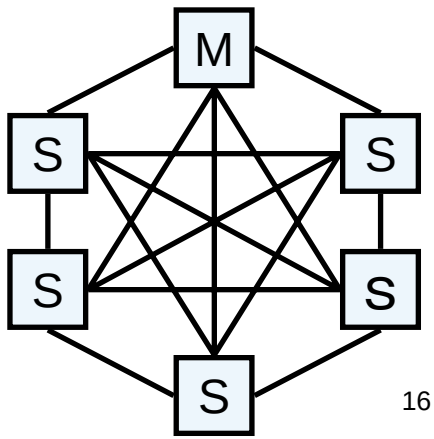
Slony-Iとは？

- Slony Development Groupによる開発
  - PostgreSQL開発コアメンバーのJan Wieck氏が中心
- レプリケーションに特化
  - コネクションプーリングやロードバランス機能はない
  - pgpoolと組み合わせて使用することが可能
- BSDライセンス

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

Slony-Iの機能

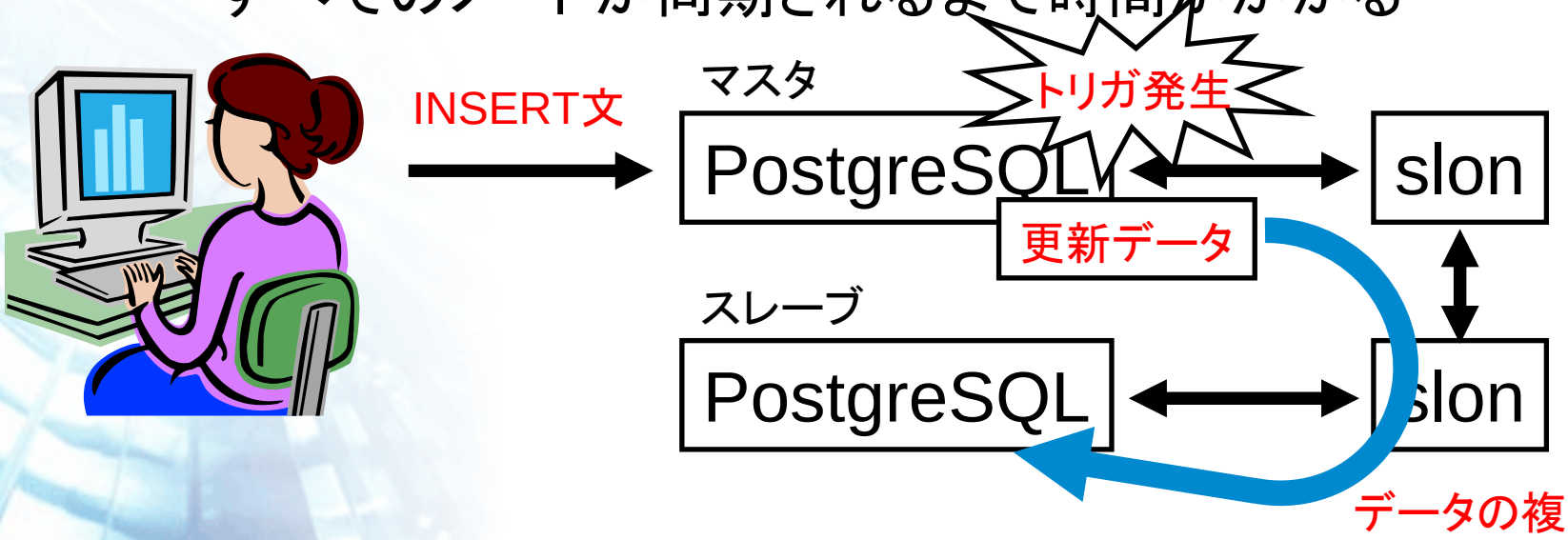
- 非同期・マスタスレーブ構成
  - レプリケーションには遅延が発生
  - 更新を受け付けられるのはマスタのみ
- ノードのカスケード接続
  - マスタ・スレーブ間だけでなく、スレーブ・スレーブ間のレプリケーションが可能
  - 複数台のノードによる複雑な構成が可能
- スレーブからマスタへの切り替え



Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

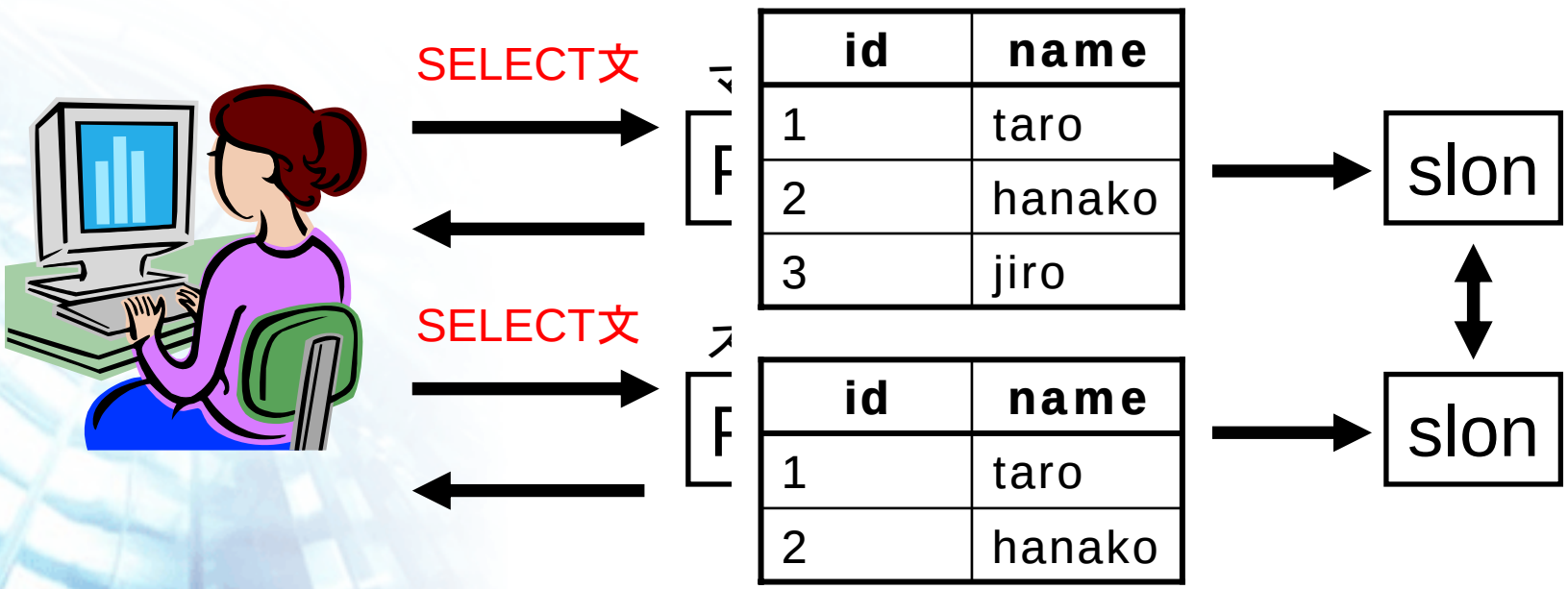
Slony-Iのレプリケーション

- 更新系クエリを実行、トリガが更新データを記録
- slonデーモンが非同期に更新データを複製
  - 更新系クエリでもあまり時間がかからない
    - 遠隔地へのバックアップが可能
  - すべてのノードが同期されるまで時間がかかる



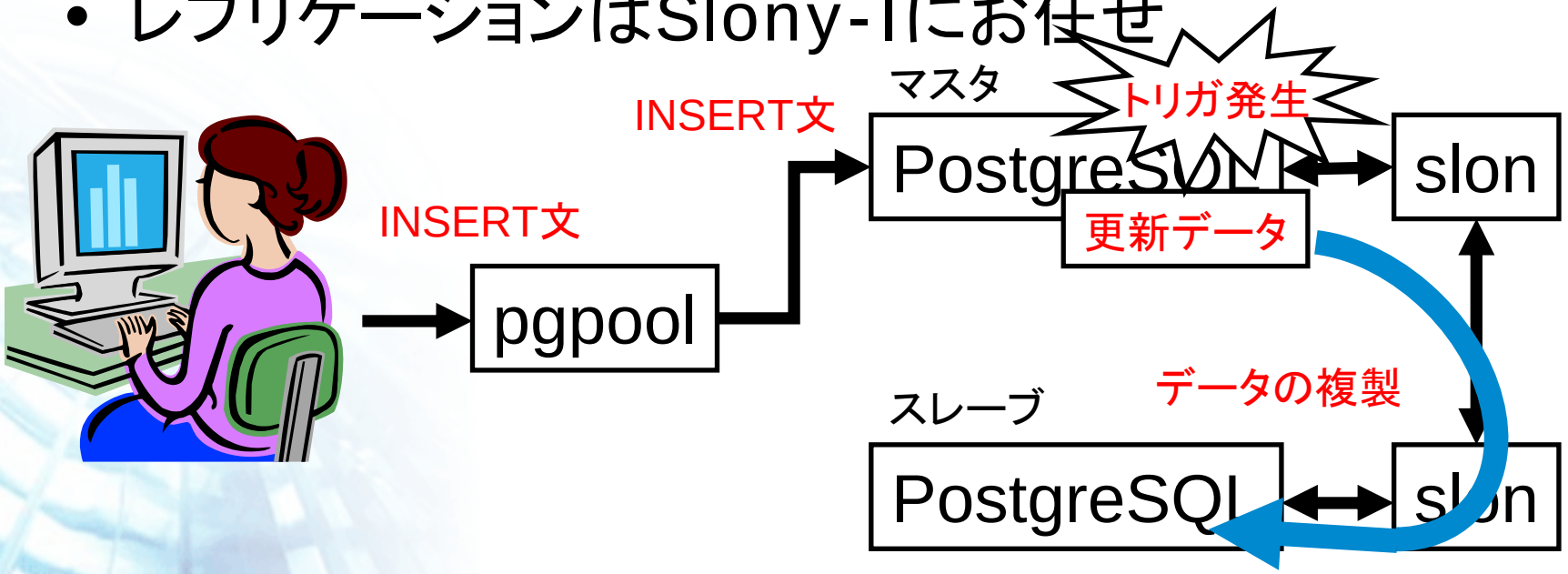
Slony-Iのレプリケーション

- ロードバランス
  - ロードバランス機能がないので、クライアントで検索を振り分ける必要がある
  - 非同期なのでノードによって検索結果が異なることがある



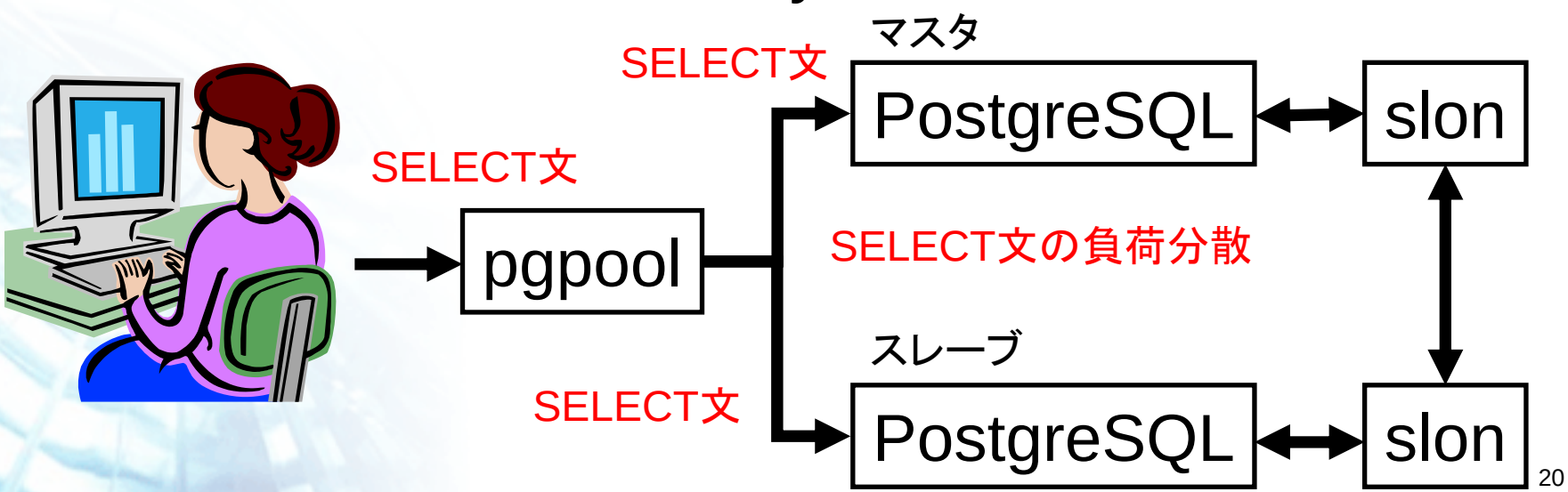
pgpoolとSlony-Iの組み合わせ

- pgpoolはマスタスレーブモードで動作
  - コネクションプール、ロードバランス機能のみを提供
  - INSERT文はマスタのみに、SELECT文の実行を負荷分散
- レプリケーションはSlony-Iにお任せ



pgpoolとSlony-Iの組み合わせ

- pgpoolはマスタスレーブモードで動作
  - コネクションプール、ロードバランス機能のみを提供
  - INSERT文はマスタのみに、SELECT文の実行を負荷分散
- レプリケーションはSlony-Iにお任せ



pgpool／Slony-Iの運用性とその評価

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

運用性評価の概要

- 評価対象
  - pgpool
  - Slony-I
- 評価の観点
  - インストール・設定の容易性
  - 機能の有効性
  - PostgreSQLからの移行性
  - 耐障害性

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

インストール・設定の容易性

- pgpool、Slony-Iともに

```
./configure && make && make install
```

で基本的にインストール可能

- 設定は…

pgpool	• 設定ファイルpgpool.confのみ • 3.2からpgpool_hba.conf（クライアント認証）も設定可能 • PostgreSQLの書式と同じ
Slony-I	• slonikコマンドプロセッサを使って設定するが分かりにくい • altperlやpgAdmin IIIを使えば楽になるかも

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

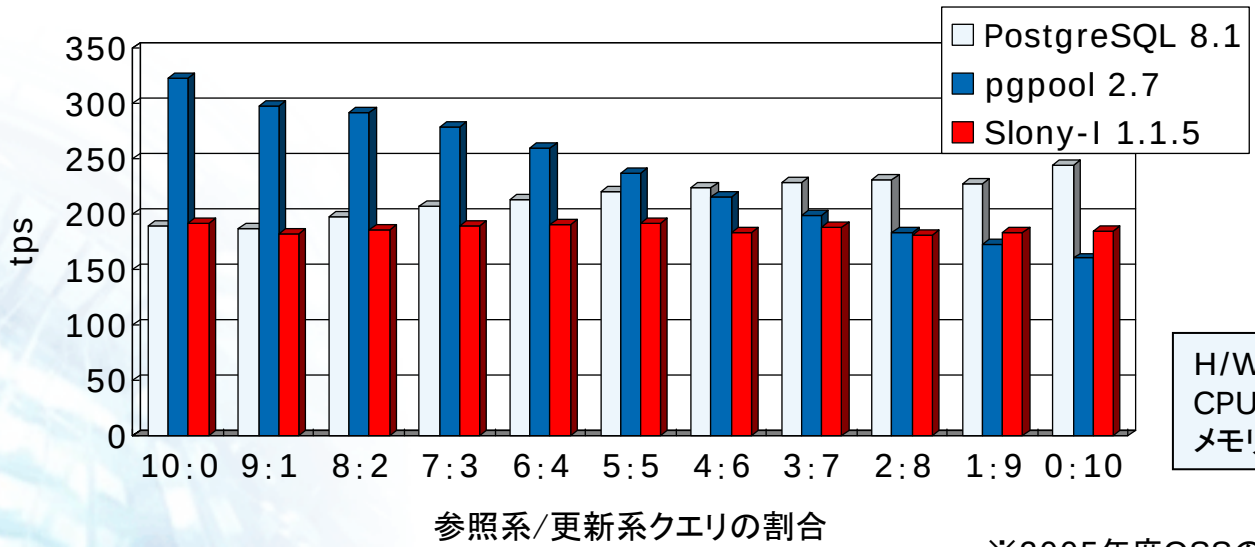
インストール・設定の容易性
Slonikコマンドプロセッサを使った設定例

```
cluster name = slony_test;
node 1 admin conninfo = 'dbname=test_db host=db1 user=postgres';
node 2 admin conninfo = 'dbname=test_db host=db2 user=postgres';
init cluster ( id=1, comment = 'Master');
table add key (node id=1, fully qualified name='public.history');
create set (id=1, origin=1, comment='All tables');
set add table (set id=1, origin=1, id=1,
       fully qualified name = 'public.accounts', comment='accounts');
set add table (set id=1, origin=1, id=2,
       fully qualified name = 'public.branches', comment='branches');
set add table (set id=1, origin=1, id=3,
       fully qualified name = 'public.tellers', comment='tellers');
set add table (set id=1, origin=1, id=4,
       fully qualified name = 'public.history', comment='history', key=serial);
store node (id=2, comment = 'Slave');
store path (server = 1, client = 2, conninfo='test_db host=db1 user=postgres');
store path (server = 2, client = 1, conninfo='test_db host=db2 user=postgres');
store listen (origin=1, provider = 1, receiver =2);
store listen (origin=2, provider = 2, receiver =1);
```

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

機能の有効性

- pgpool、Slony-Iともにマニュアルに書いてある機能は問題ない
- では、どのくらい有効なのかと言えば…



H/W:HP ProLiant DL380 G3
CPU:Intel Xeon 2.8GHz × 2
メモリ:2.5GB

※2005年度OSSの性能・信頼性評価の結果より

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

PostgreSQLからの移行性
pgpoolの場合

- 実行するタイミングで結果が変わるSQLはレプリケーションできない
  - 例：random()やcurrent_timestampなど
  - クライアントで値を求めてSQLに埋め込めば対応可能
- serial型やシーケンスはテーブルのロックが必要
  - pgpool.confでinsert_lockを有効にすればINSERT文を実行する際に自動的にテーブルをロック
- DDLはそのまま移行できる

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

PostgreSQLからの移行性
Slony-Iの場合

- DDLはレプリケーションできない
  - 例：CREATE TABLE、ALTER TABLE文など
  - slonikコマンドプロセッサでEXECUTE SCRIPTを実行すれば対応可能
- ラージオブジェクトはレプリケーションできない
  - システムカタログにはトリガを定義できないため
- random()やcurrent_timestampなどはそのまま移行可能

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

耐障害性

- pgpool
  - ノードの障害時には縮退運転でサービス継続可能
  - 障害復旧時にはサービスの停止が必要
- Slony-I
  - スレーブに障害が発生してもサービスを継続できる
  - オンラインリカバリが可能
  - マスタに障害が発生すると更新系クエリの実行が不可

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

まとめ

- 目的に応じて使い分けことが重要
- pgpool:
  - 検索性能を向上させたい
  - 簡単に導入したい(ノードは2台まで)
- Slony-I:
  - 遠隔地へのバックアップを行いたい
  - オンラインリカバリを行いたい
- 詳細はOSS iPediaを参照
  - http://ossipedia.ipa.go.jp/

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

pgpoolの参考URL・ML

- 開発サイト
  - http://pgfoundry.org/projects/pgpool/’
- コミュニティサイト
  - http://pgpool.sraoss.jp/
- メーリングリスト
  - pgpool-general-jp@sraoss.jp（一般向け、日本語）
  - pgpool-general@pgfoundry.org（一般向け、英語）
  - pgpool-hackers@pgfoundry.org（開発向け、英語）

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

Slony-Iの参考URL・ML

- 開発サイト
  - http://slony.info/
- メーリングリスト
  - slony1-general@pgfoundry.org（一般向け、英語）
  - slony1-hackers@pgfoundry.org（開発向け、英語）

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.

ありがとうございました

SRA OSS, Inc. 日本支社
http://www.sraoss.co.jp/
佐藤 友章
sato@sraoss.co.jp

Copyright © 2007 SRA OSS, Inc. Japan All rights reserved.