PostgreSQL でスケールアウト構成を構築しよう

db tech show tech 2013
2013/11/15

SRA OSS, Inc. 日本支社
技術開発部　長田 悠吾

SRA OSS,INC.

自己紹介

- 長田 悠吾（ながた ゆうご）
- 所属
  - SRA OSS, Inc. 日本支社 技術開発部
- 業務
  - PostgreSQL 関連の技術調査
  - pgpool-II の開発

    ･･･など

- SRA OSS, Inc. 日本支社
  - PostgreSQL を中心としたOSSのサポート/コンサルティング
  - OSS 関連プロダクトの販売
  - 技術者トレーニングサービス

SRA OSS,INC.

本日の内容

- PostgreSQL のスケールアウト構成

  「サーバを複数台使って高い性能を得る」

PostgreSQL 標準のレプリケーション機能

＋

pgpool-II による負荷分散 & 高可用化

実績もあり、構成しやすい組み合わせ

その手法についてご紹介いたします！

SRA OSS, INC.

アジェンダ

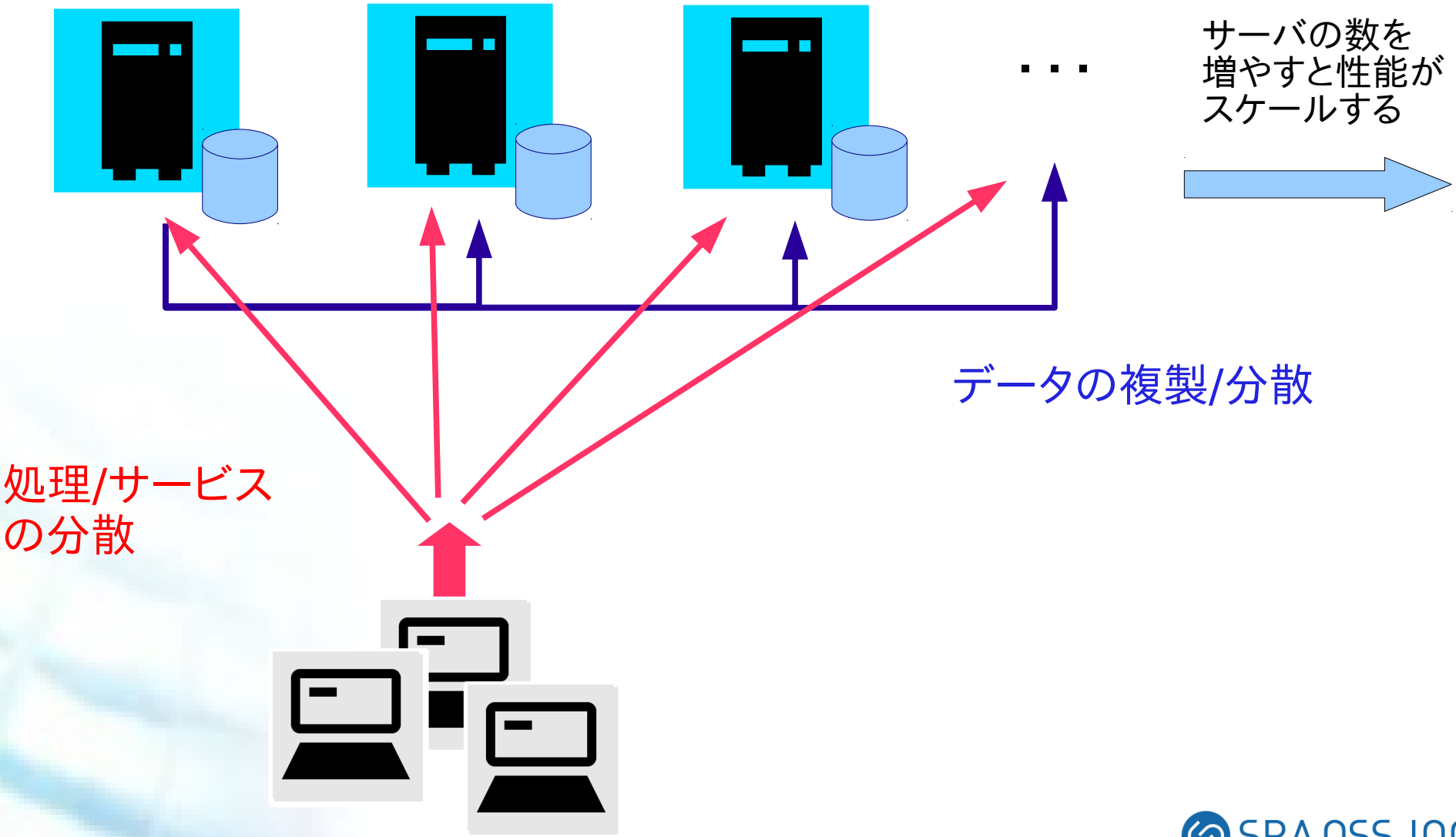
- データベースのスケールアウト構成
- PostgreSQL におけるレプリケーション
- pgpool-II
  - 負荷分散
  - 高可用化
- PostgreSQL と pgpool-II によるシステム構成
- スケールアウト性能
- デモ
- まとめ

データベースの
スケールアウト構成

SRA OSS, INC.

データベースのスケールアウト構成

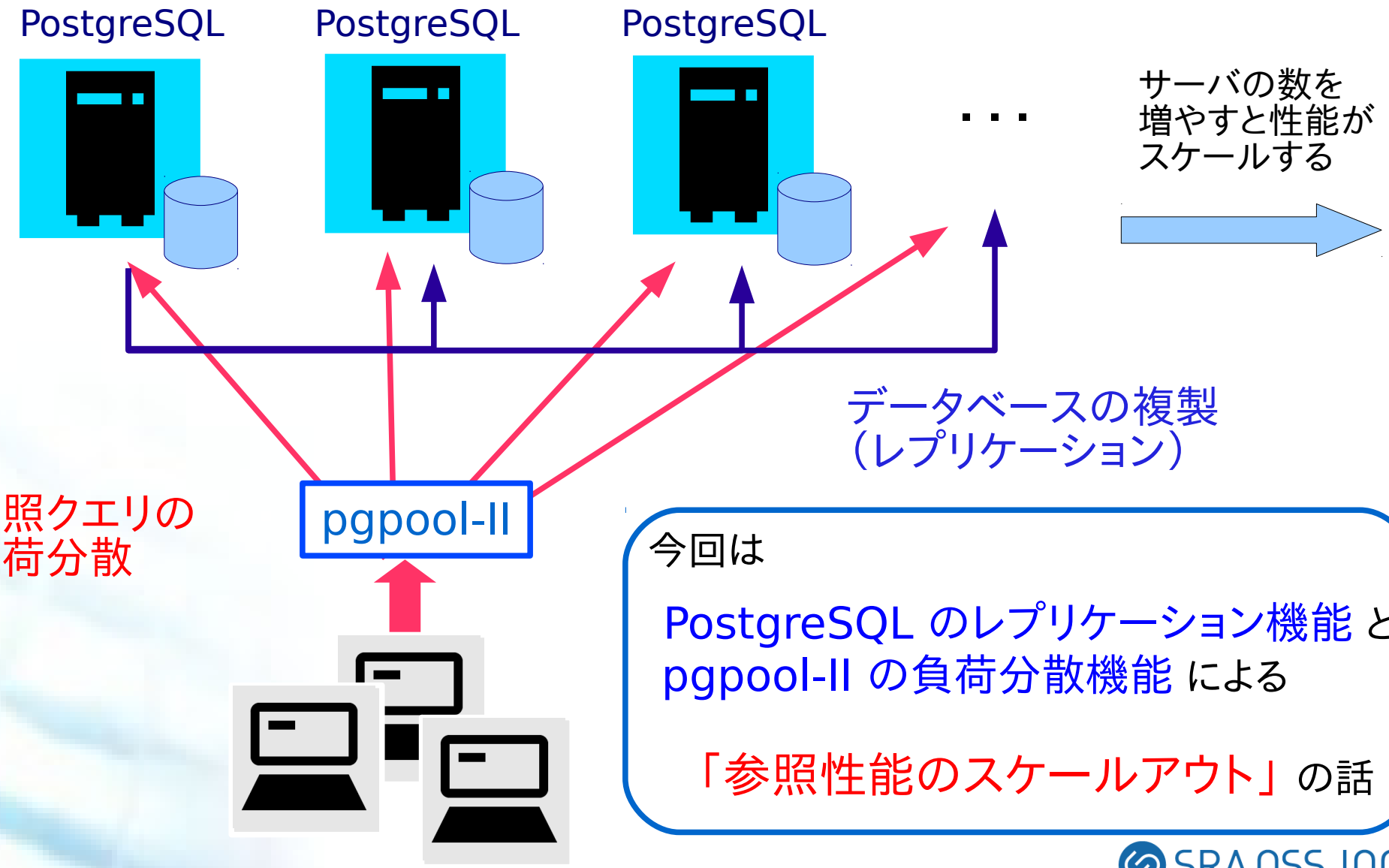
- 複数のデータベースサーバに処理を分散させる



SRA OSS, INC.

データベースのスケールアウト構成

- 複数のデータベースサーバに処理を分散させる



今回は

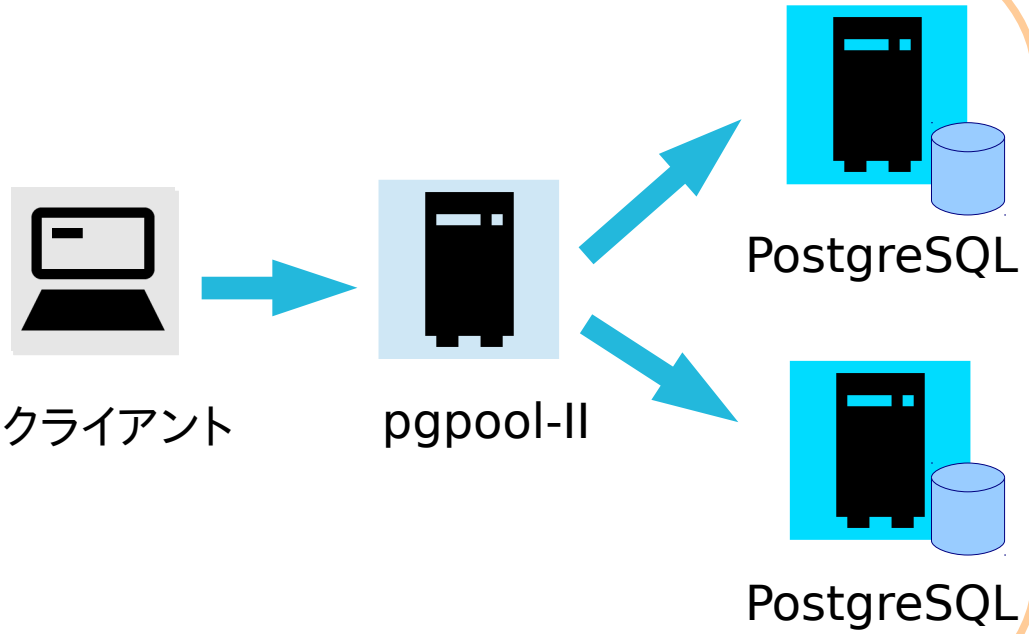
PostgreSQL のレプリケーション機能 と
pgpool-II の負荷分散機能 による

「参照性能のスケールアウト」の話

SRA OSS, INC.

pgpool-II とは?

- アプリケーションと PostgreSQL の間に介在し、クラスタリング機能を提供するミドルウェア
- オープンソースソフトウェア（BSDライセンス）
- 多彩な機能
  - コネクションプーリング
  - 参照負荷分散
  - オンメモリクエリキャッシュ
  - 自動フェールオーバー
  - オンラインリカバリ
  - レプリケーション



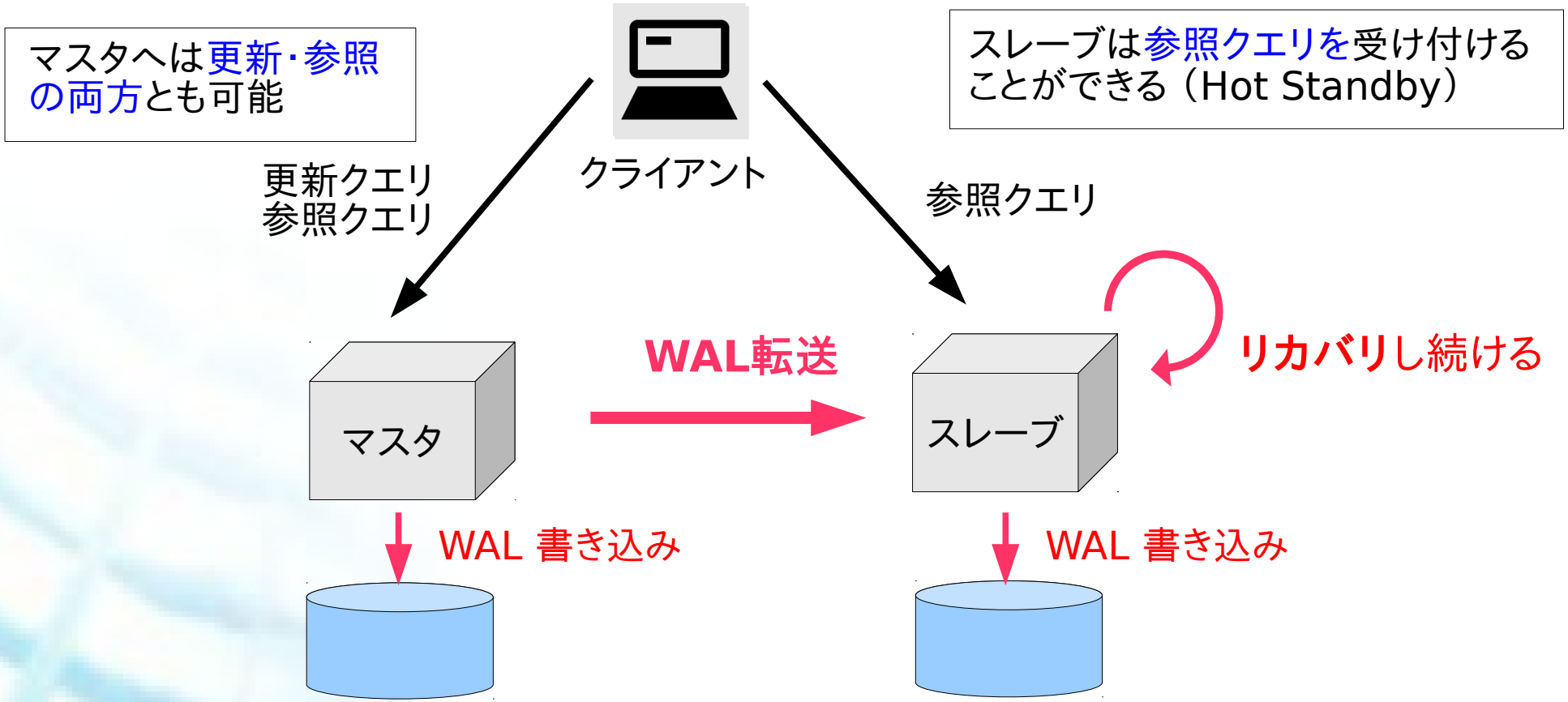
SRA OSS, INC.

PostgreSQL における レプリケーション機能

SRA OSS, INC.

PostgreSQL のレプリケーション機能

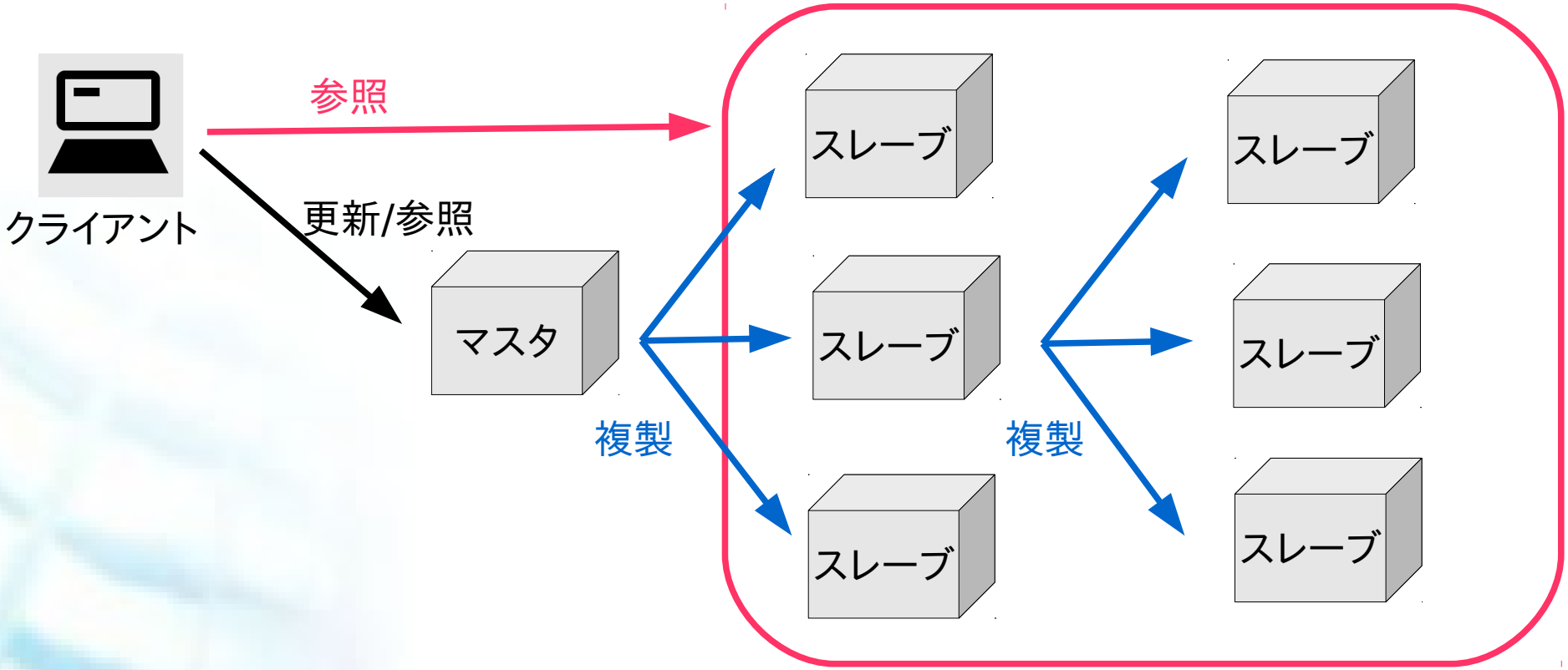
- ストリーミングレプリケーション（PostgreSQL 9.0 ～）
  - マスタからスレーブにトランザクションログ（WAL）を転送することによりデータの複製を実現
  - 転送とリカバリの遅延のため、マスタとスレーブが常に同じ内容とは限らない



SRA OSS, INC.

PostgreSQL のレプリケーション機能

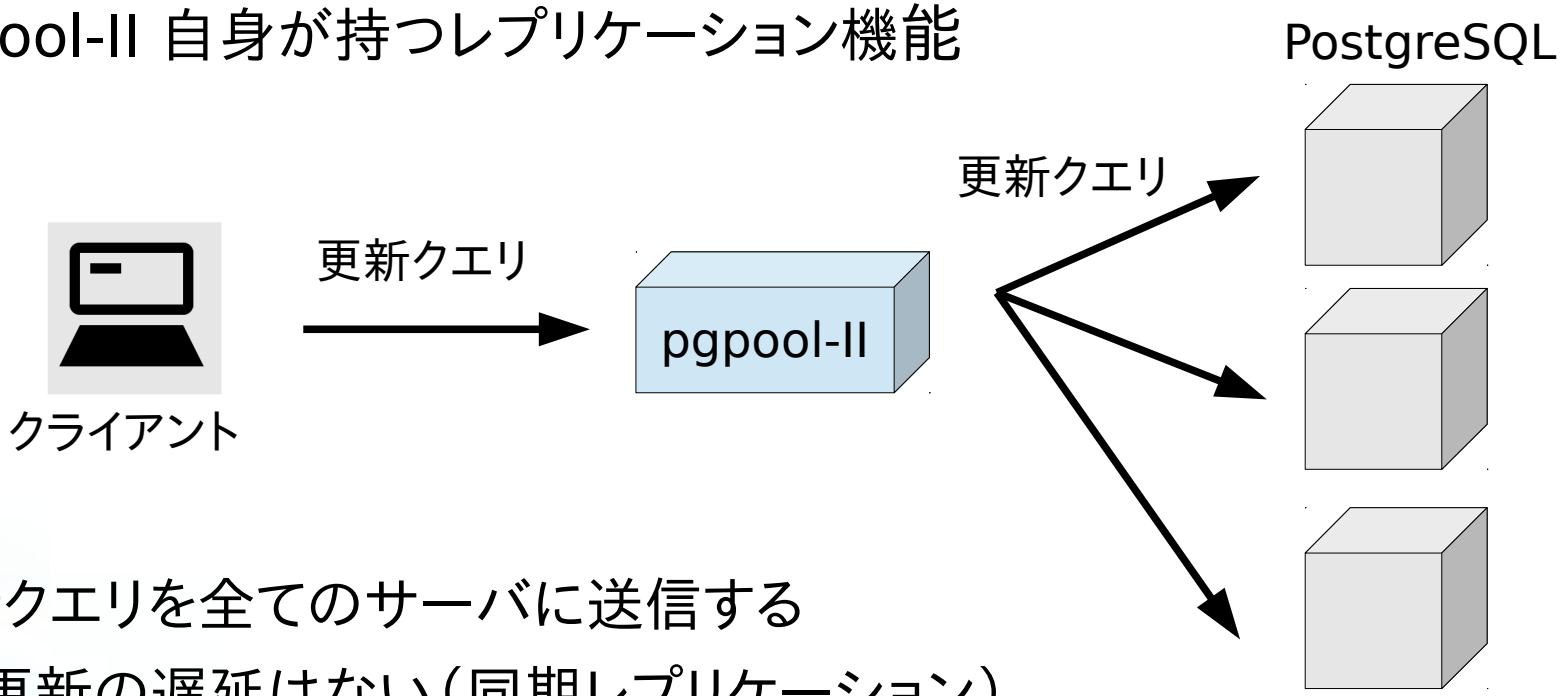
- 1つのマスタから複数のスレーブに複製可能
  - スレーブからさらに別のスレーブへのレプリケーションが可能
    - カスケードレプリケーション（PostgreSQL 9.2 ～）
  - スレーブは参照クエリを受け付けることができる
    - これを利用して参照性能のスケールアウト!



SRA OSS, INC.

（参考）pgpool-II によるレプリケーション

- ネイティブレプリケーションモード
  - PostgreSQL のストリーミングレプリケーション機能を用いない pgpool-II 自身が持つレプリケーション機能



  - 更新クエリを全てのサーバに送信する
    - 更新の遅延はない（同期レプリケーション）
    - 更新性能が 50% に落ちる

- 更新の遅延が問題になる場合以外は、ストリーミングレプリケーションを用いるのがおすすめ

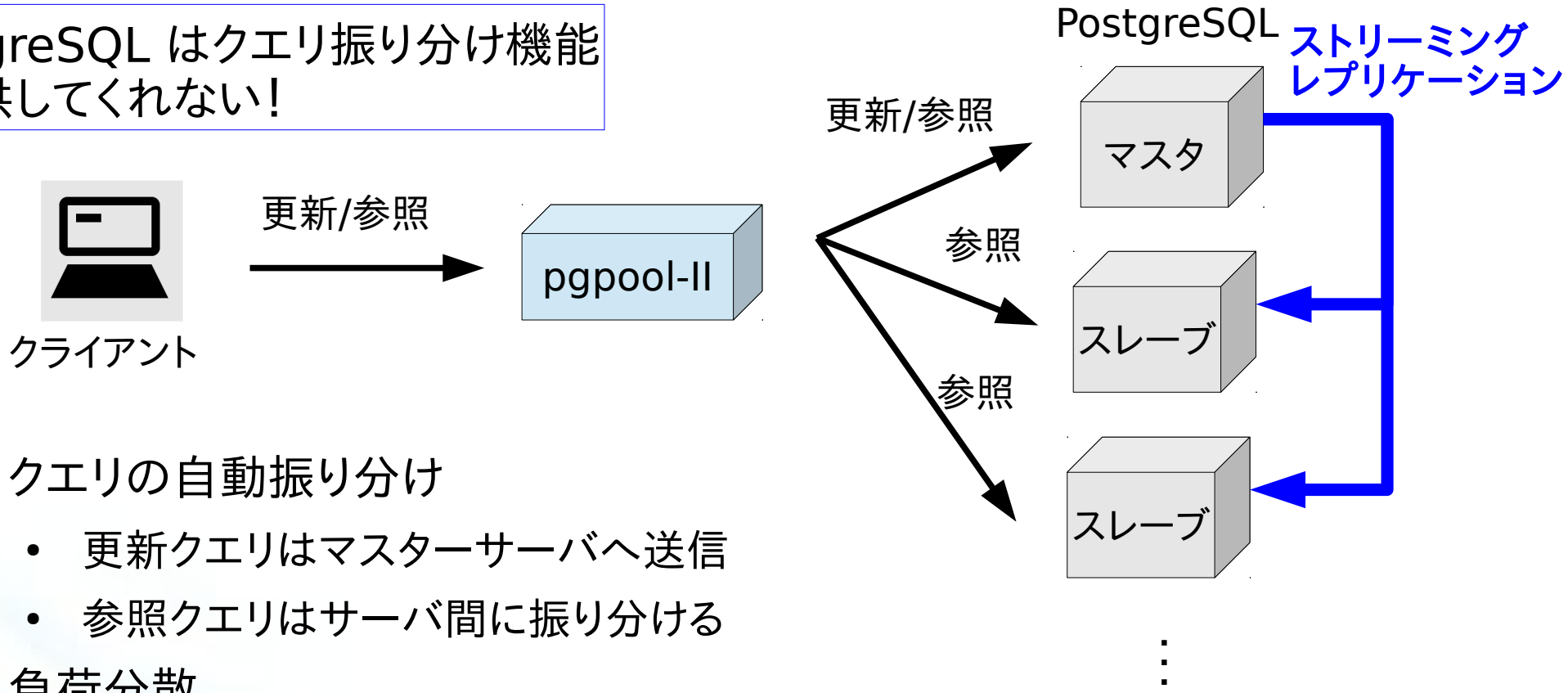
SRA OSS, INC.

pgpool-II による
負荷分散 & 高可用化

SRA OSS, INC.

負荷分散

PostgreSQL はクエリ振り分け機能を提供してくれない！

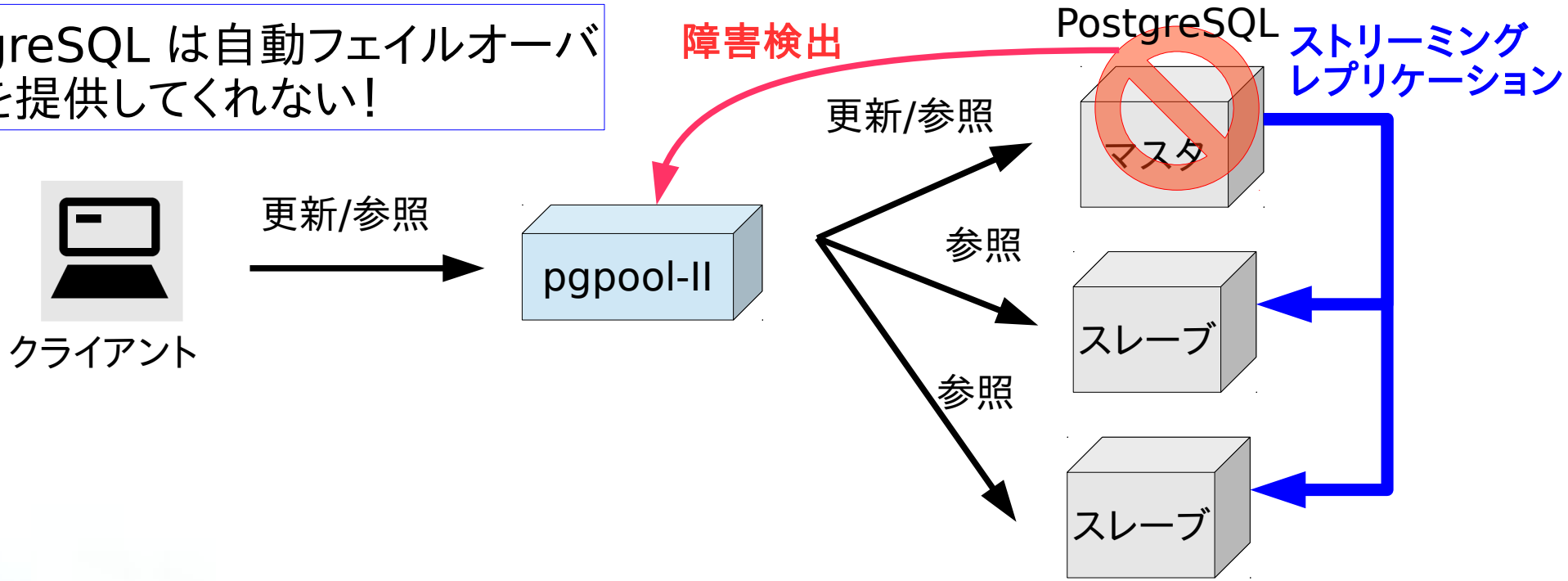


- クエリの自動振り分け
  - 更新クエリはマスターサーバへ送信
  - 参照クエリはサーバ間に振り分ける
- 負荷分散
  - 参照クエリを配分する負荷の重み付けが可能
  - スレーブサーバを増やすことにより参照性能をスケールアウトすることが可能
- レプリケーション遅延の監視
  - 遅延が閾値を越えたら負荷分散の対象から外すことが可能

SRA OSS, INC.

自動フェイルオーバ

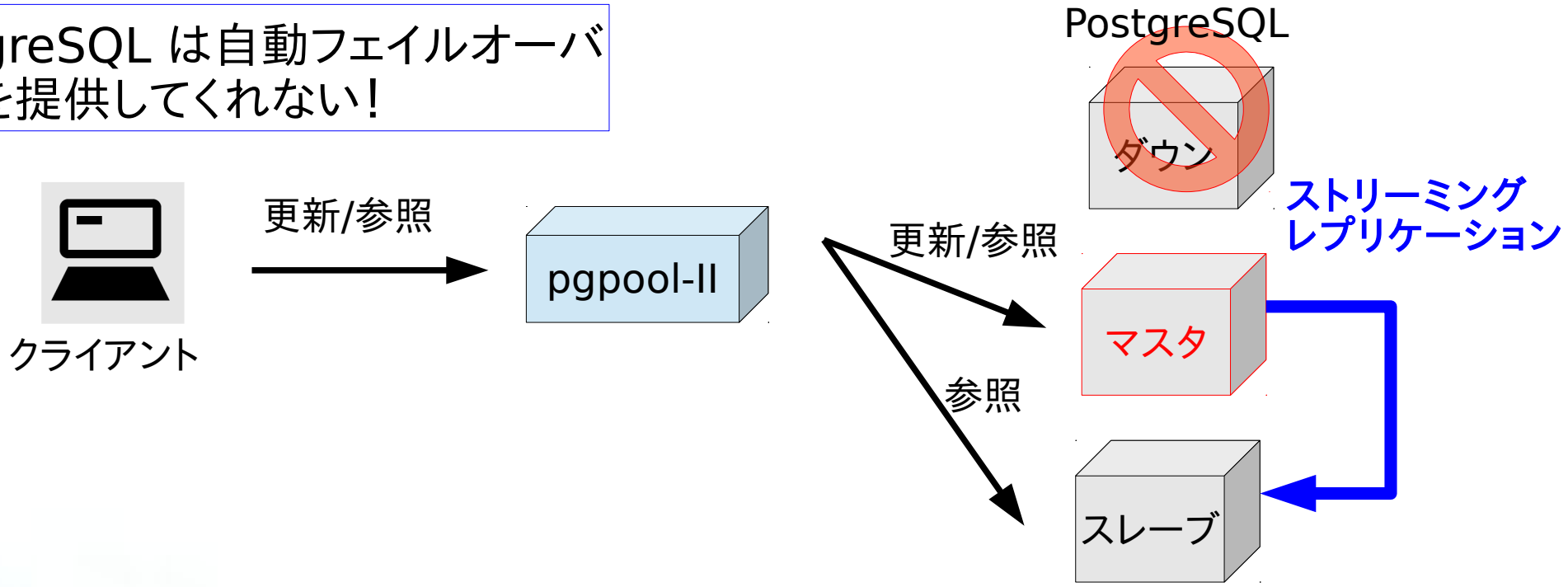
PostgreSQL は自動フェイルオーバ機能を提供してくれない!



- pgpool-II は PostgreSQL を定期的に監視
- ダウンしたノードは自動的に切り離される
  - ノードがダウンした場合の後処理はシェルスクリプトで記述する
    - マスタがダウンした場合には、1つのスレーブをマスターに昇格させ、他のスレーブの「複製元」を新しいマスタに変更する・・・など

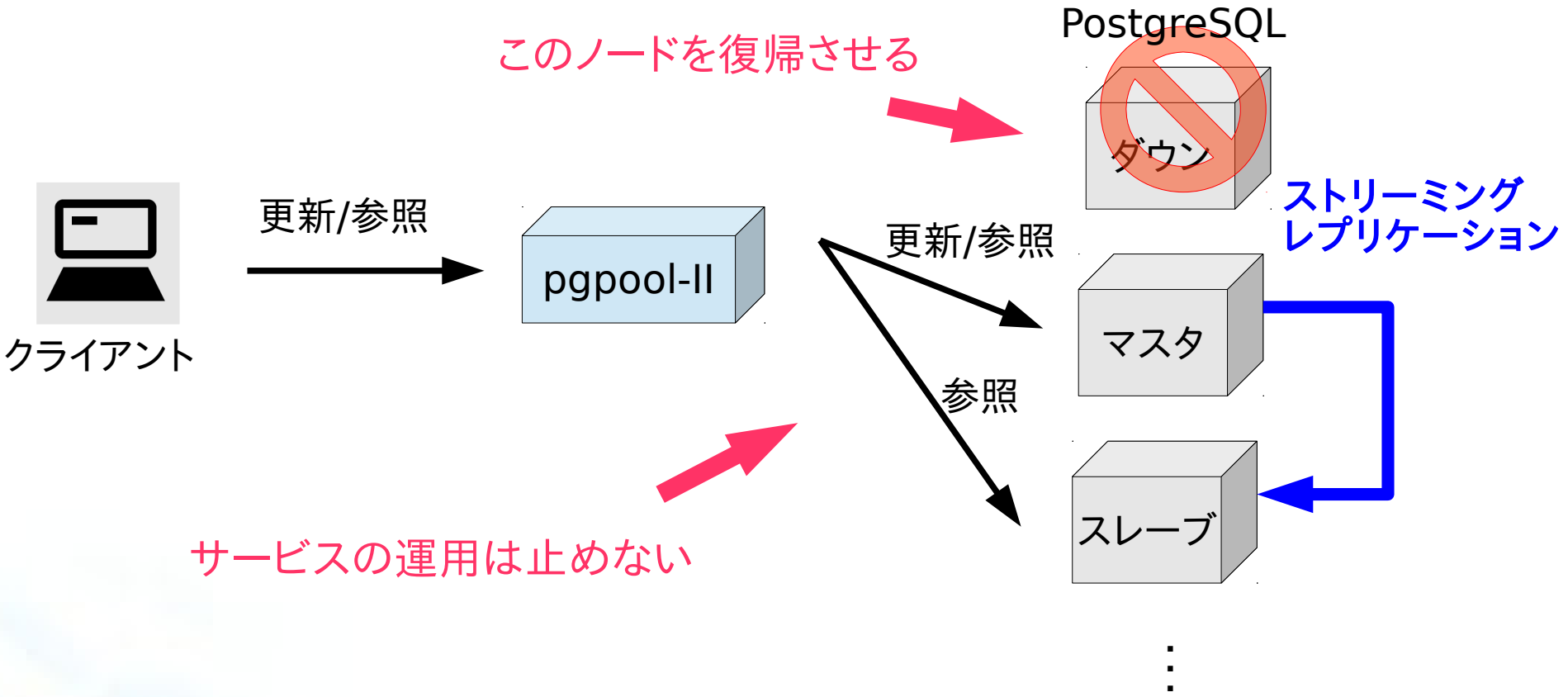
自動フェイルオーバ

PostgreSQL は自動フェイルオーバ機能を提供してくれない!



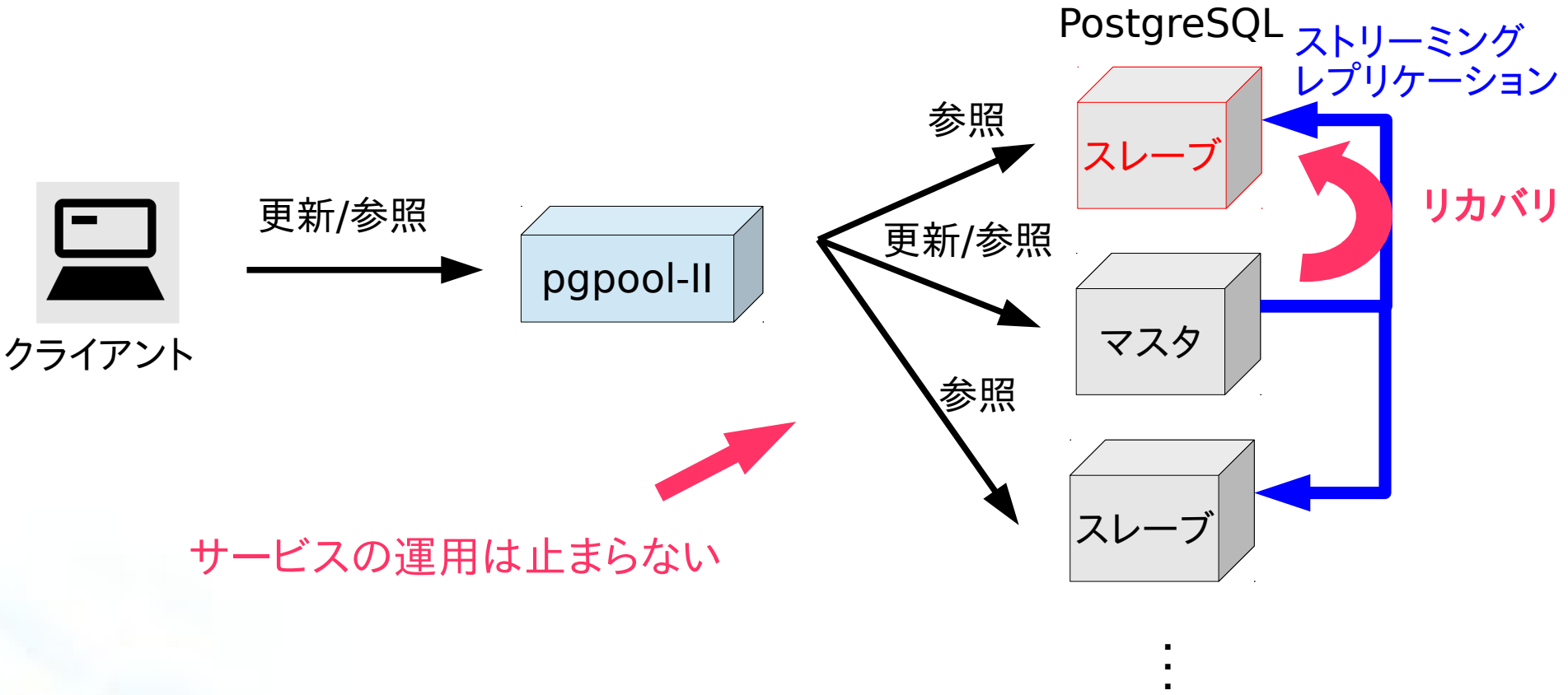
- pgpool-II は PostgreSQL を定期的に監視
- ダウンしたノードは自動的に切り離される
  - ノードがダウンした場合の後処理はシェルスクリプトで記述する
    - マスタがダウンした場合には、1つのスレーブをマスターに昇格させ、他のスレーブの「複製元」を新しいマスタに変更する・・・など

オンラインリカバリ



- システムの運用を止めずにノードを復帰させる機能
  - マスタのベースバックアップを取得し、これを元にリカバリを行う

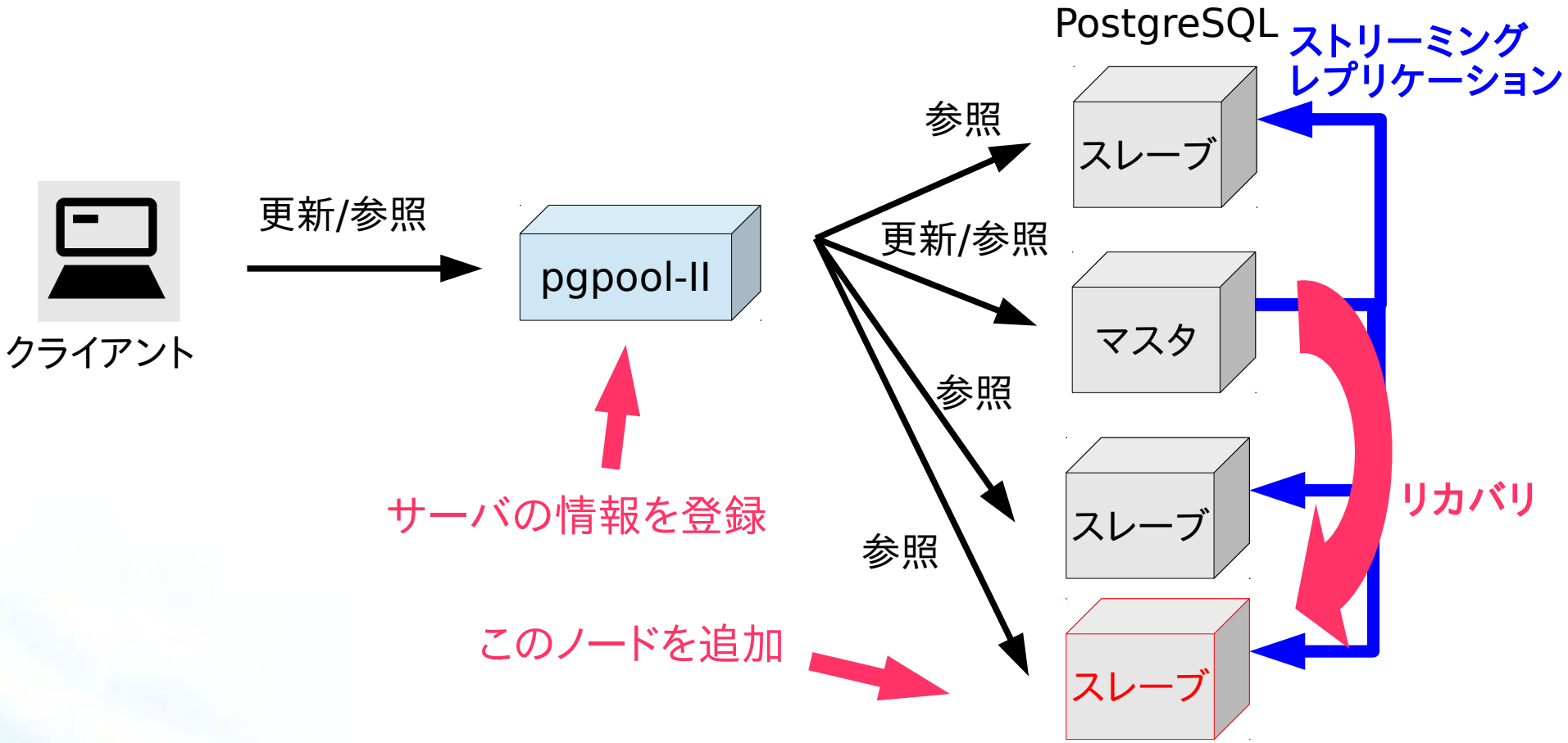
オンラインリカバリ



- システムの運用を止めずにノードを復帰させる機能
  - マスタのベースバックアップを取得し、これを元にリカバリを行う

SRA OSS, INC.

新しいスレーブの追加



- 同じ方法で新しいスレーブサーバの追加が可能
  - pgpool-II に追加するサーバの情報（ホスト名、ポート番号など）を登録
  - その後に、オンラインリカバリを行う

  → 参照性能のスケールアウトが容易に！

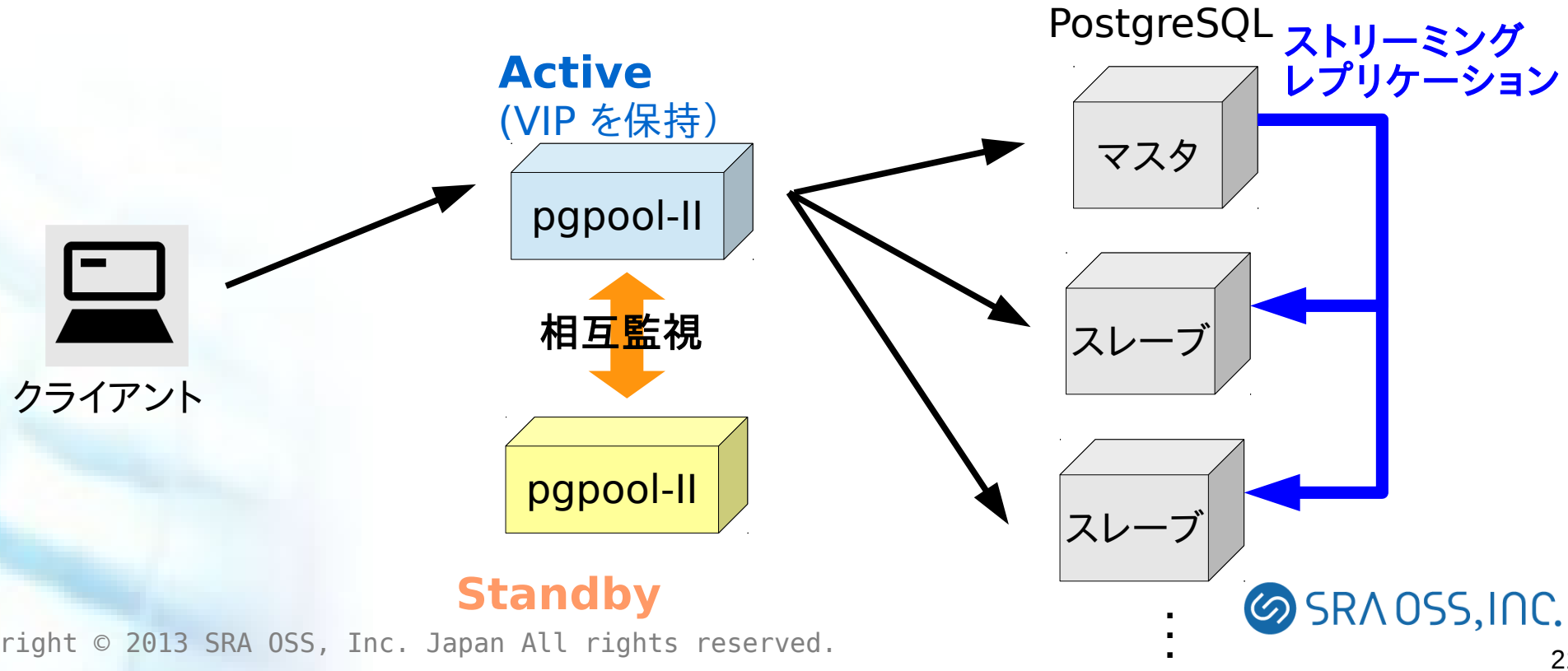
SRA OSS, INC.

PostgreSQL と pgpool-II
システム構成

SRA OSS, INC.

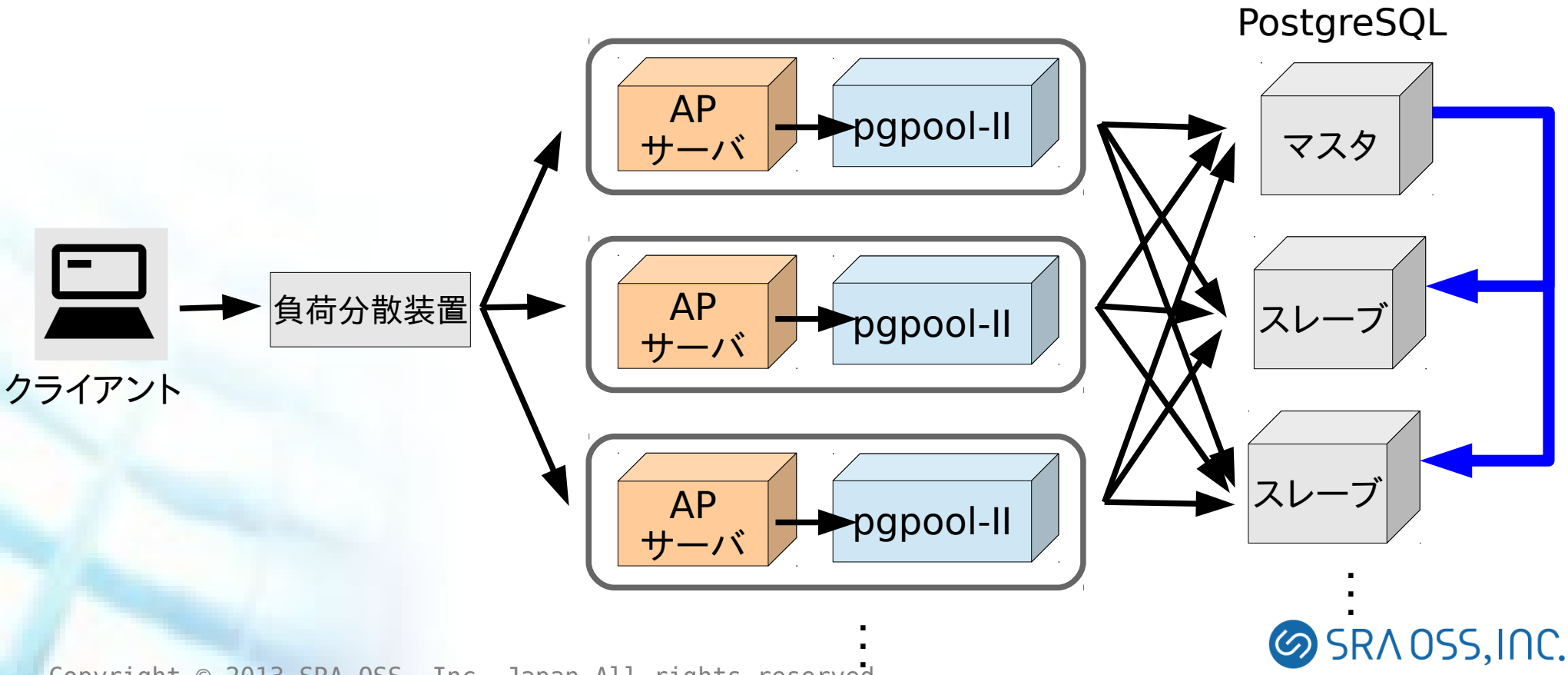
pgpool-II の active/standby 構成

- Watchdog
  - pgpool-II 組み込みの高可用性機能
  - pgpool-II を Active/Standby 構成にすることで SPoF（単一障害点）を回避
    - クライアントは Active pgpool-II に仮想IP(VIP)でアクセスする
    - Active pgpool-II がダウンしたら、standby pgpoo-II が VIP を引き継ぐ



マルチマスタ的構成

- APサーバと pgpool-II を1台のサーバに同居させた構成
  - pgpoo-II が冗長化されている
  - APサーバ/pgpool-II のペアを増やすことで、APサーバの性能をスケールアウト可能
  - PostgreSQL を増やすことでDB参照性能をスケールアウト可能

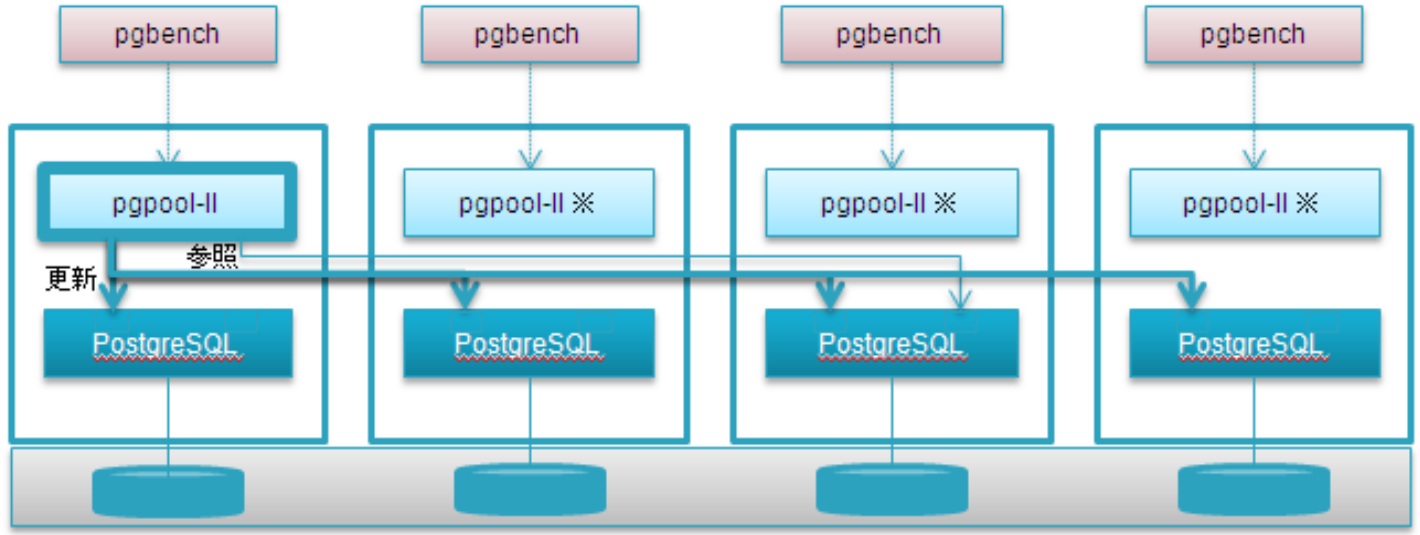


SRA OSS, INC.

スケールアウト性能

スケールアウト性能

- 参照性能は本当にスケールアウトするか？
- pgpool-II(3.2.1) & PostgreSQL(9.2.1)で、ノード数を増やすと全体の処理能力が向上するかを確認
  - 1～4台の PostgreSQL で検証
  - マルチマスタ的構成と似た構成
  - APサーバに相当する位置に、ベンチマークツール（pgbench）が配置されている。

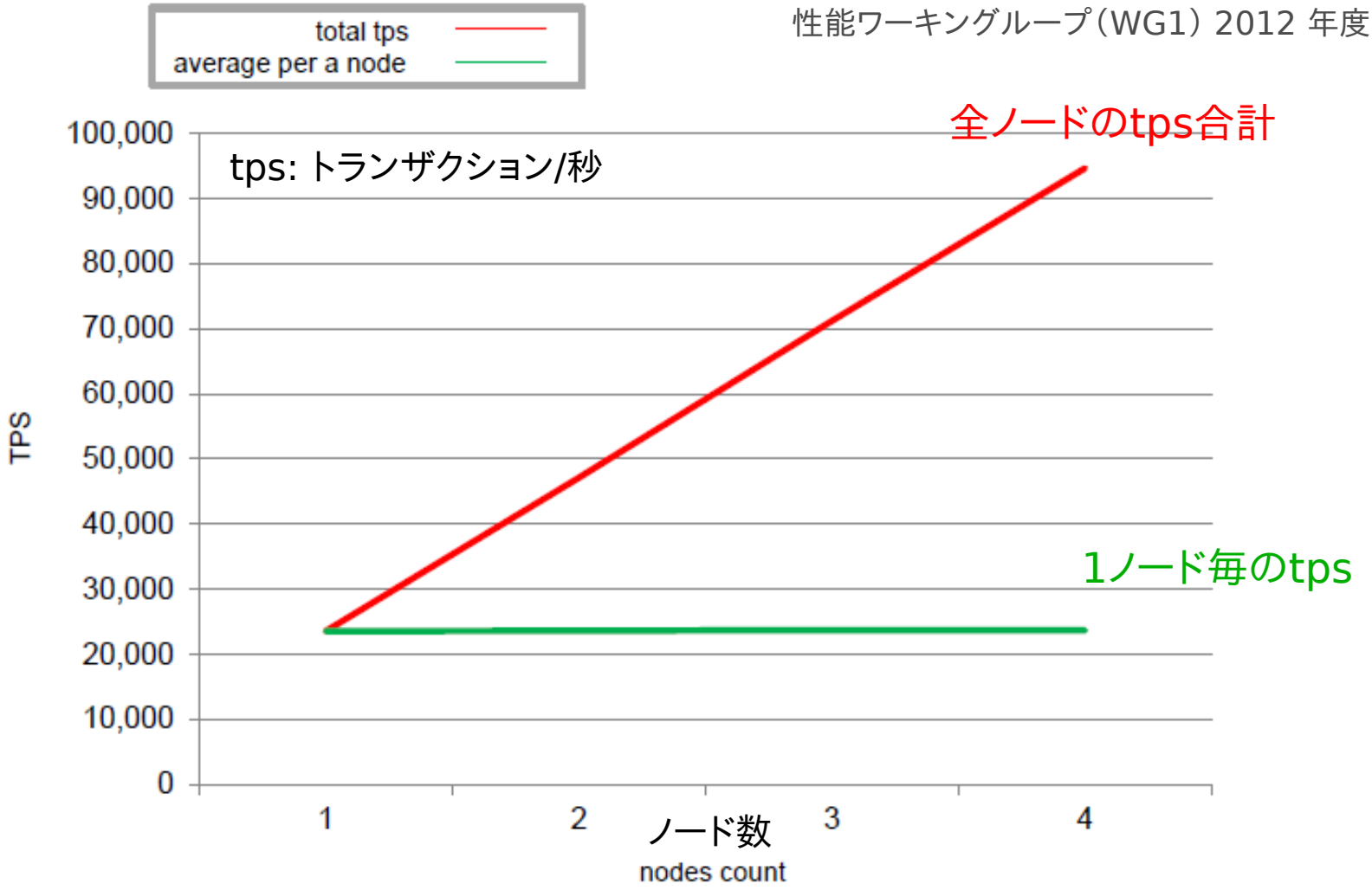


※PostgreSQL エンタープライズ・コンソーシアム
性能ワーキンググループ（WG1）2012 年度成果物より引用

SRA OSS, INC.

スケールアウト性能（結果）

※PostgreSQL エンタープライズ・コンソーシアム
性能ワーキンググループ（WG1）2012 年度成果物より引用



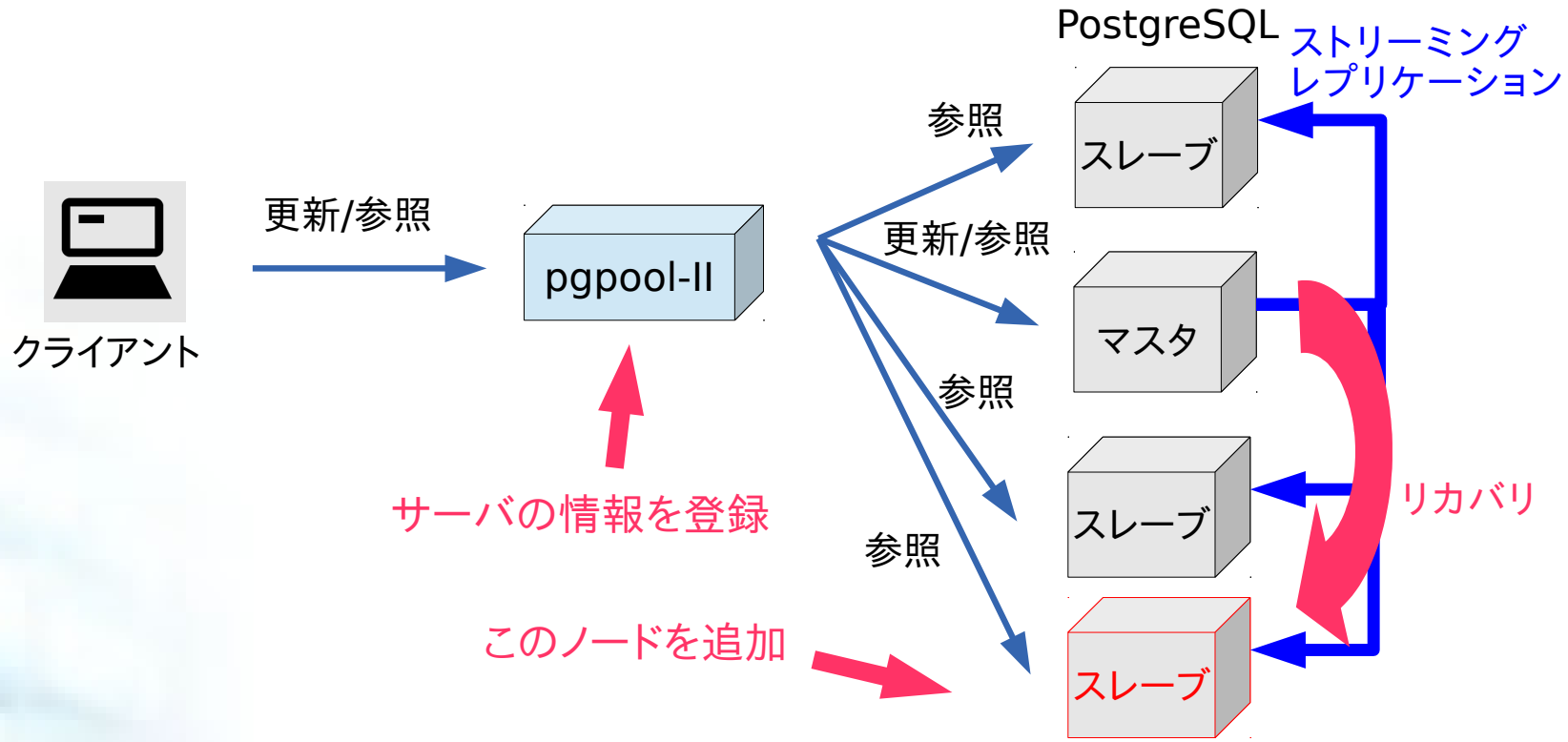
ノード数が増えるほど、合計の tps が増える（スケールメリットあり）

SRA OSS, INC.

デモ

デモの内容

- 運用中のシステムに新しいスレーブを追加
  - ベンチマーク処理を中断せずに新しいノードが追加可能
  - 負荷分散により、自動的に新しいスレーブに参照クエリが振り分けられるようになる



SRA OSS, INC.

まとめ

まとめ

- PostgreSQL 標準のレプリケーション機能
  - ストリーミングレプリケーション
- pgpool-II
  - 負荷分散
  - クエリの自動振り分け
  - 自動フェイルオーバ
  - オンラインリカバリ
  - 新規ノードの追加
- これらの組み合わせによる、PostgreSQL 参照性能スケールアウト構成の紹介
  - 参照性能のスケールアウト
  - 新規ノード追加のデモ

SRA OSS, INC.

参考URL

- PostgreSQL ドキュメント
  - http://www.postgresql.jp/document/9.3/html/
- pgpool-II オフィシャルサイト
  - http://www.pgpool.net/

PostgreSQL エンタープライズコンソーシアム

- 略称:PGECons
- 2012 年度実施報告書に PostgreSQL のスケールアウト検証の結果あり
- http://www.pgecons.org/

Let's Postgres

- PostgreSQL 情報のポータルサイト
- pgpool-II 関連の記事あり
- http://lets.postgresql.jp/

SRA OSS, INC.

ご清聴ありがとうございました。

SRA OSS, INC.