

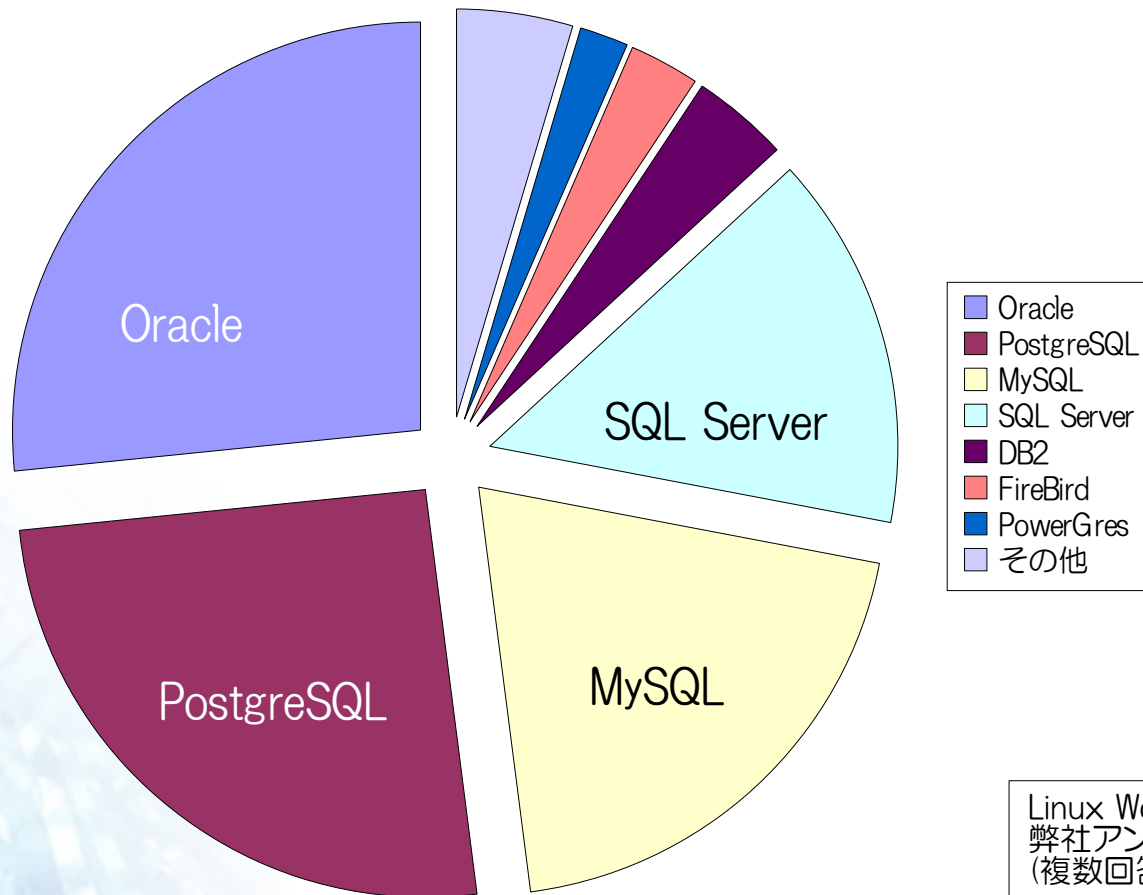
最強のOSS DB PostgreSQLの最新技術情報

SRA OSS, Inc. 日本支社
石井 達夫

2008/9/13 OSC島根

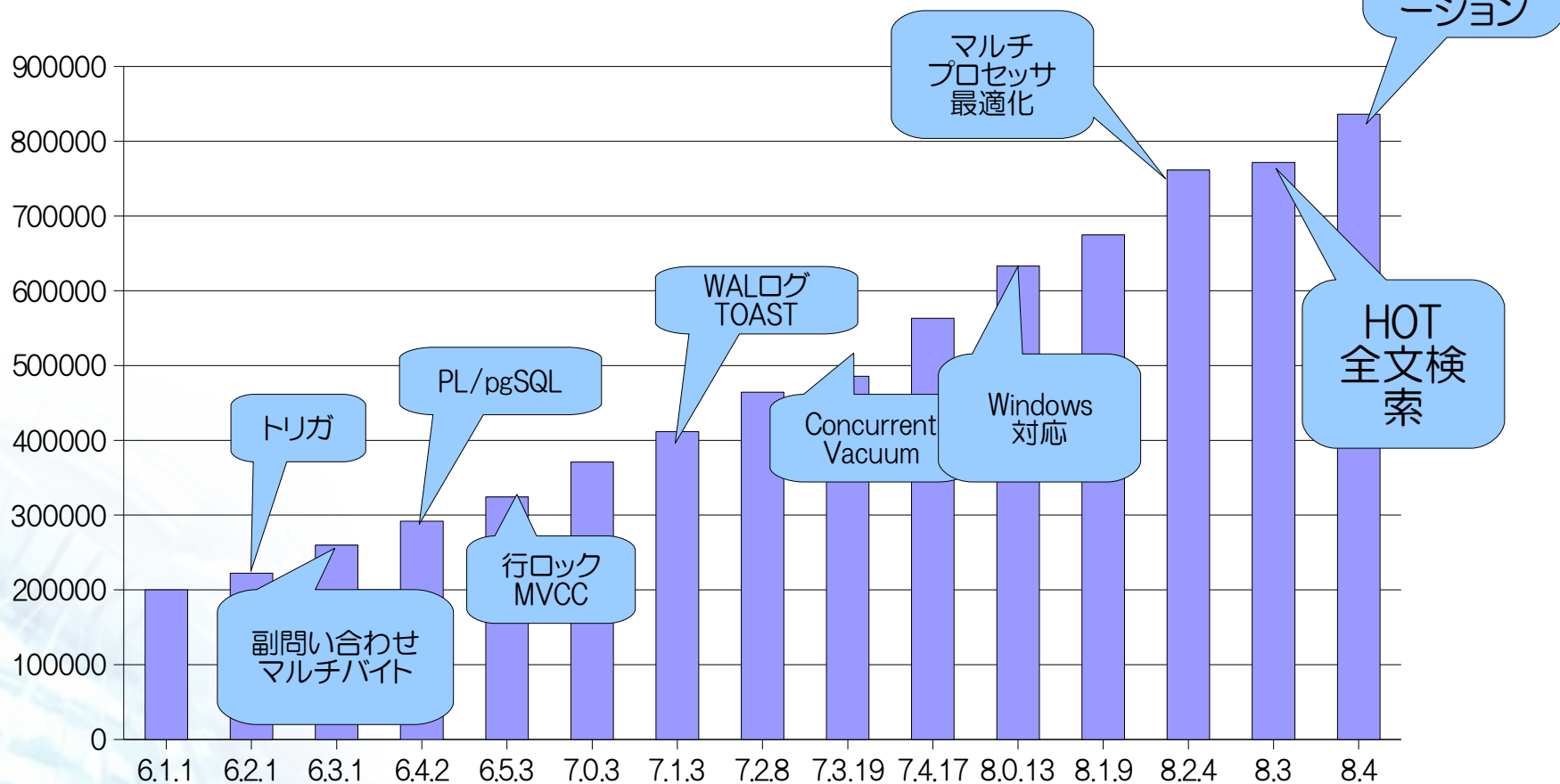
最強のOSS DB PostgreSQL

PostgreSQLは 日本におけるDB製品では 第2位のシェア



Linux World 2007における
弊社アンケート結果による
(複数回答可能)

順調に成長を続けるPostgreSQL



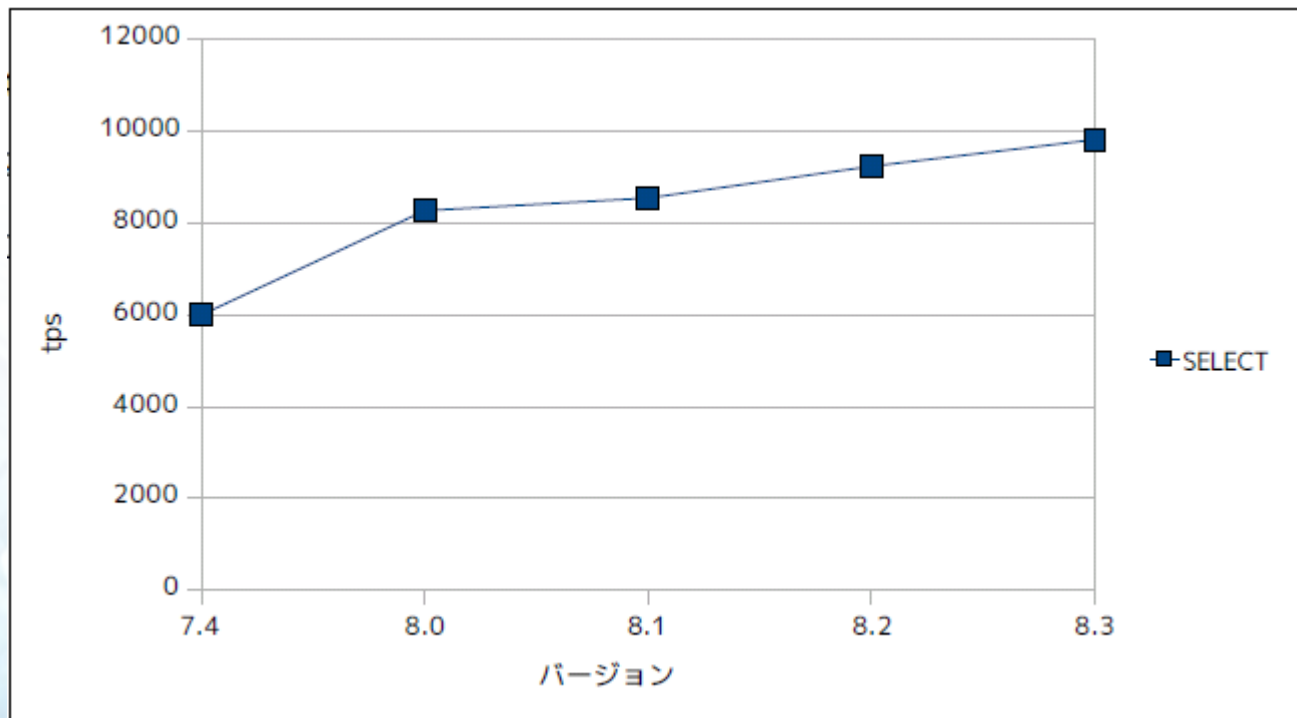
この5年間でPostgreSQLの性能は どれだけ向上したか？

- DBマガジンの記事より
- 最新版8.3の更新性能の
向上に注目！



検索性能の向上

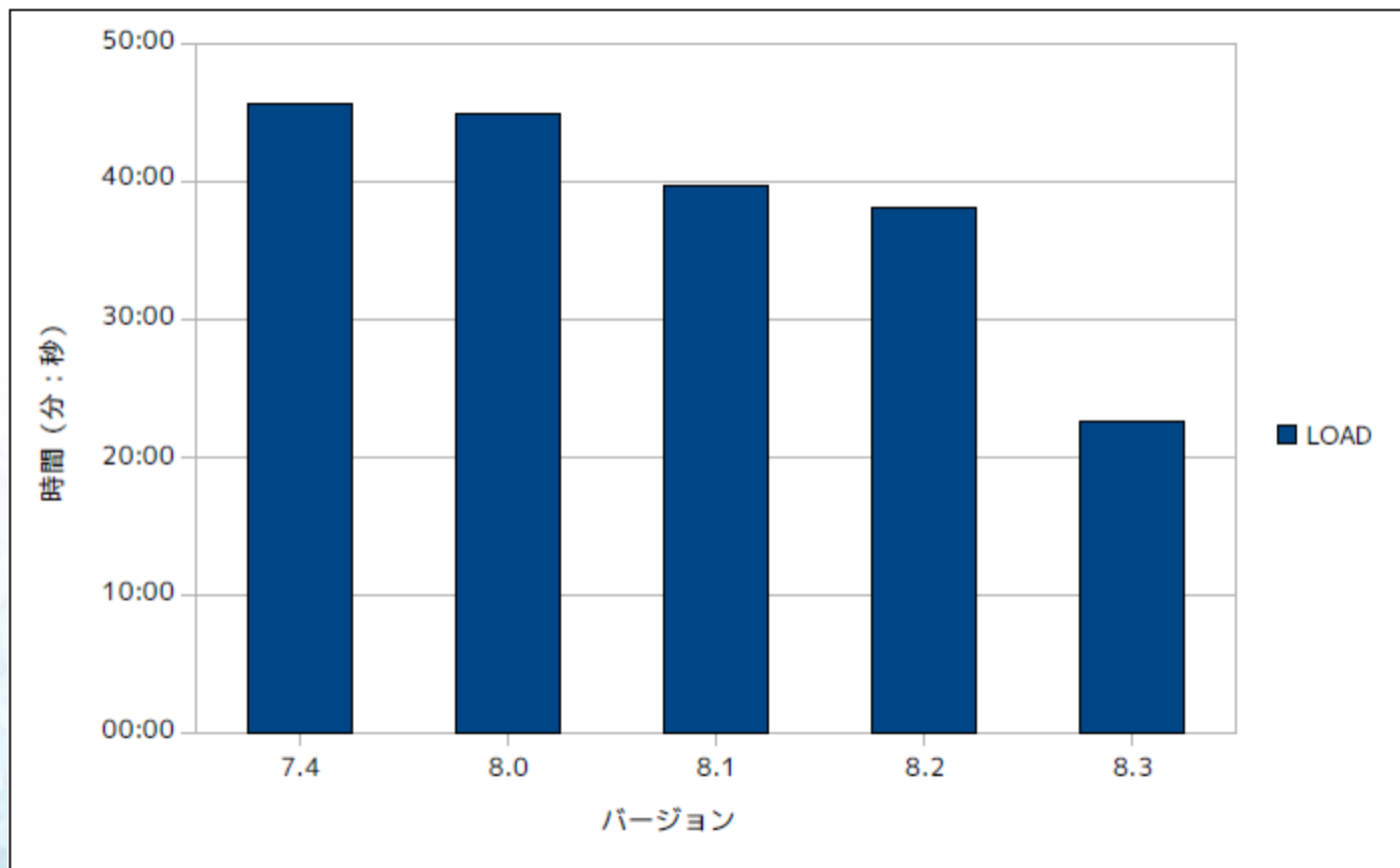
7.4(2003/11)から8.3(2008/2)の間に**164%**の性能向上



一括ロード性能の向上

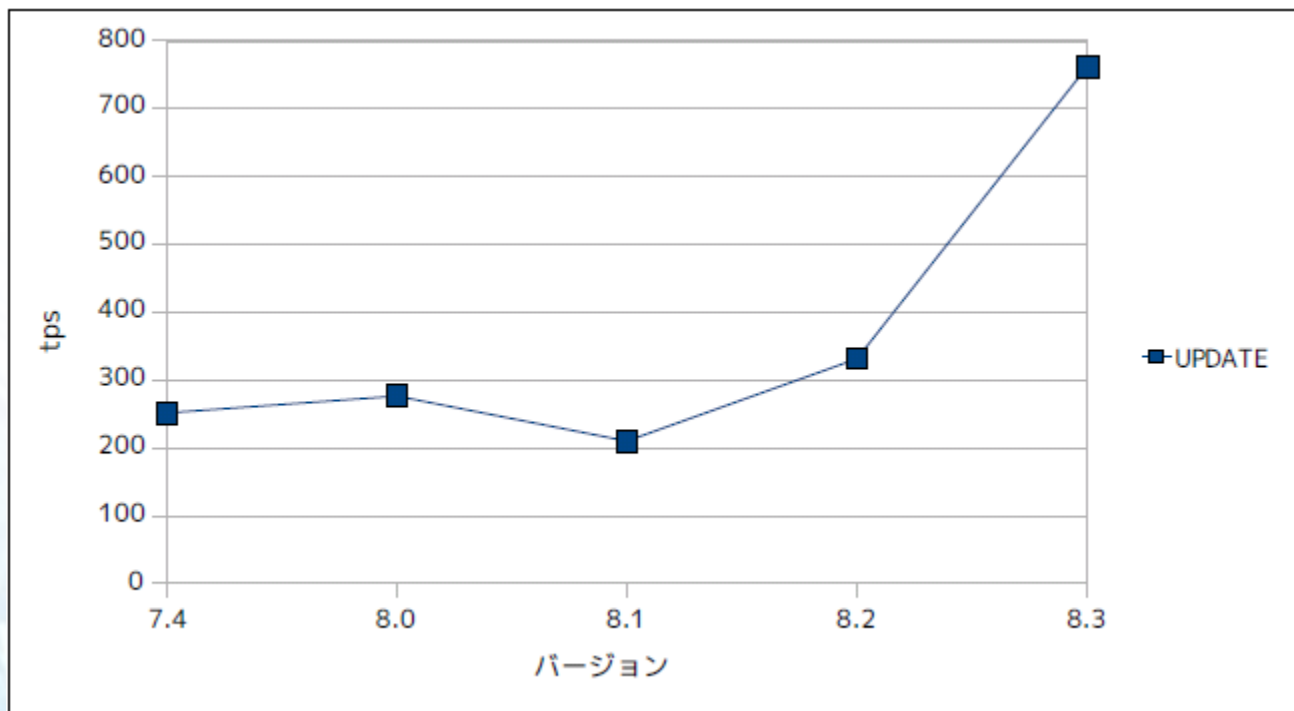
7.4(2003/11)から8.3(2008/2)の間に**200%**の性能向上

1機中(15GB)のロード



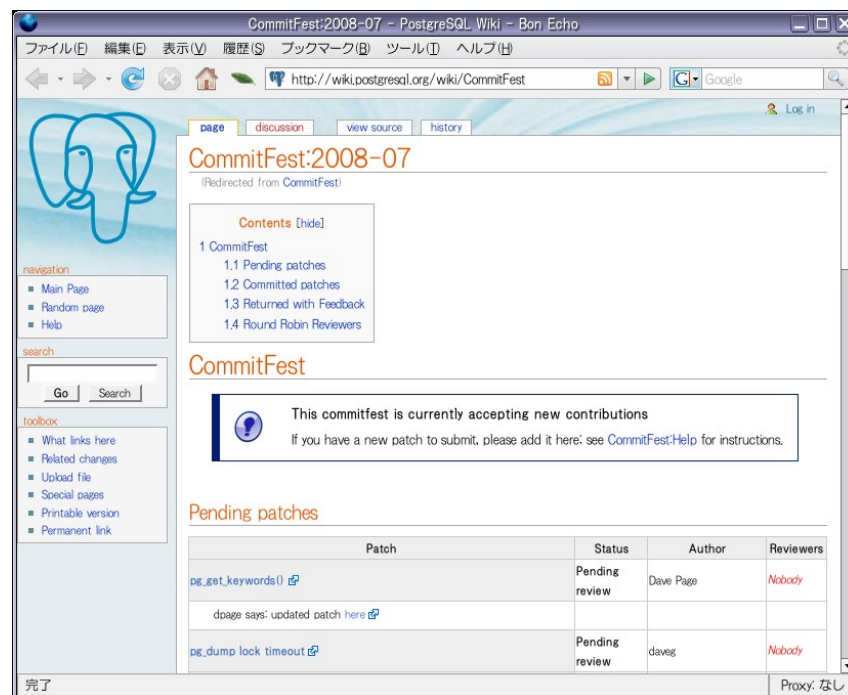
更新性能の向上

7.4(2003/11)から8.3(2008/2)の間に**300%**の性能向上



PostgreSQL 8.4の動向

- 8.4からは”commit festa”方式で開発
 - 一ヶ月おきに「集中コミット月間」を設け、パッチをレビュー、コミットする
- 現在のリリース目標は2009年の第一四半期

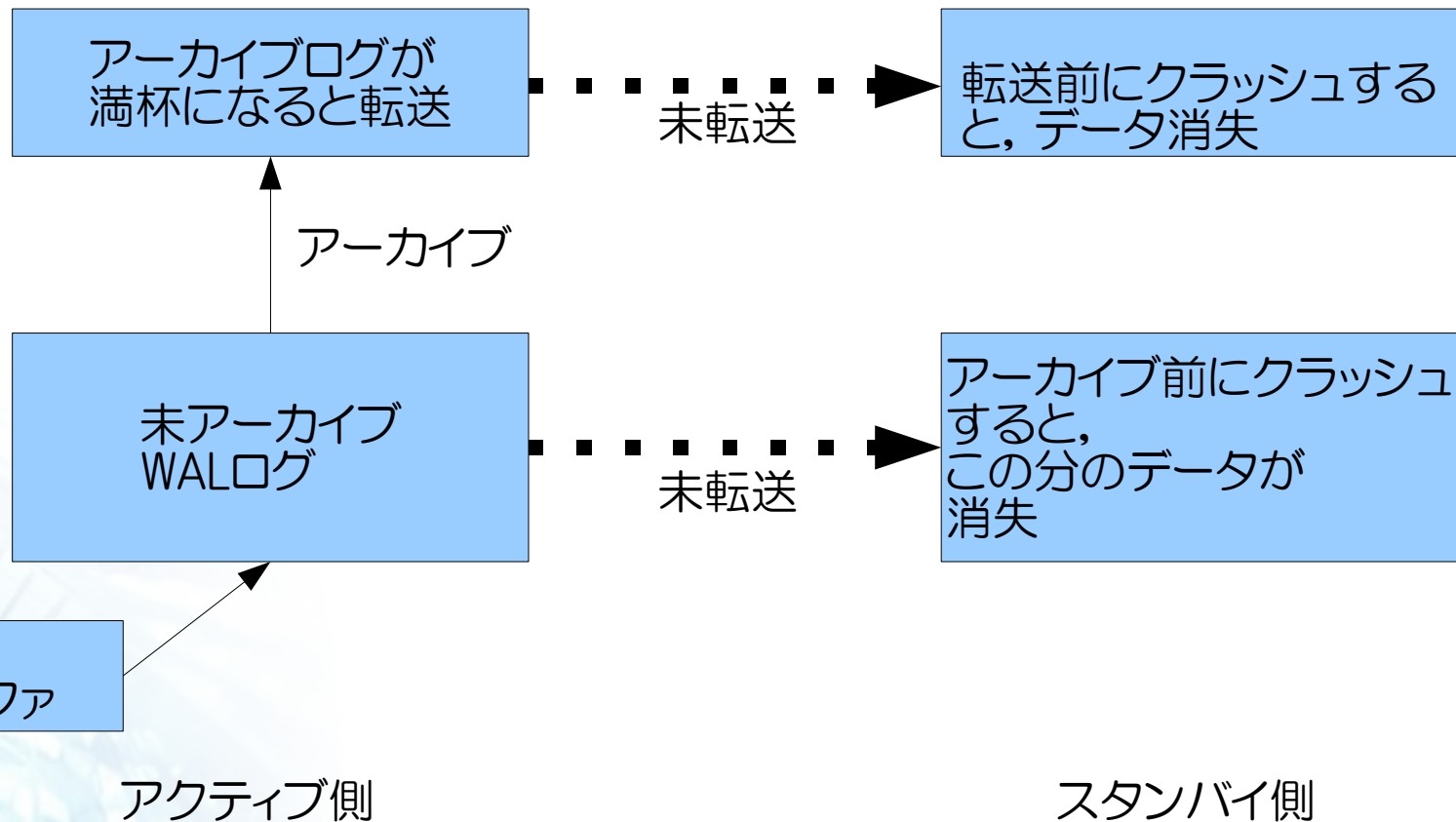


8.4注目の新機能(1)

・Synchronous Log Shipping Replication

- WALログをスレーブに転送することによってレプリケーションを実現
- warm standbyと違って同期式レプリケーションなので、コミットしたデータの消失がない

warm standbyにおける データ消失



データ消失を最小限にするためには？

- アーカイブログセグメントが満杯(16MB)になるまではアーカイブログが転送されないので、最悪16MB分のトランザクションデータが消失
- archive_timeoutを設定することによって、データ消失を最小限にすることが可能
- しかし、DBがアイドルでも、必ず16MBデータが転送されるので非効率
- しかも、archive_timeoutは無限に小さくできないのでデータ消失の可能性がわずかだが必ず残る

Synchronous Log Shipping Replication による解決

- アーカイブログセグメント単位ではなく, ログレコード単位に毎回転送
- ログレコードの転送が完了するまではコミットしない
- コミットしたデータはstandby側にも転送されていることが保証されているので, データの消失がない

Synchronous Log Shipping Replication の評価 (NTT-OSSによる)

- 特長
 - データ消失なし
 - 透過的なレプリケーション(レプリケーションすることによる機能上の制限なし)
 - 性能への影響はわずか(7%)
- 制限事項
 - スタンバイ側のDBは利用できないので、レプリケーションノードを増やしてスケールアップを狙うことはできない
 - Slony-Iやpgpoolを必要とするユースケースを完全に置き換えるものではない
 - ディザスタリカバリ目的でスタンバイ側を遠隔地に置くのはOK?

8.4注目の新機能(2)

- 再帰SQLのサポート
- リスト, 木構造などのデータ構造から再帰的にデータを取得できる
 - 今はアプリで処理するか, PL/pgSQLなどの関数を使わなければならなかった
- 住友電工とSRA OSSの協力により開発中
 - ユーザ企業が積極的にPostgreSQLの開発に関与する新しい形態(エコシステム)
- 実装はSQL標準の”WITH RECURSIVE”句を採用

単純な再帰SQL

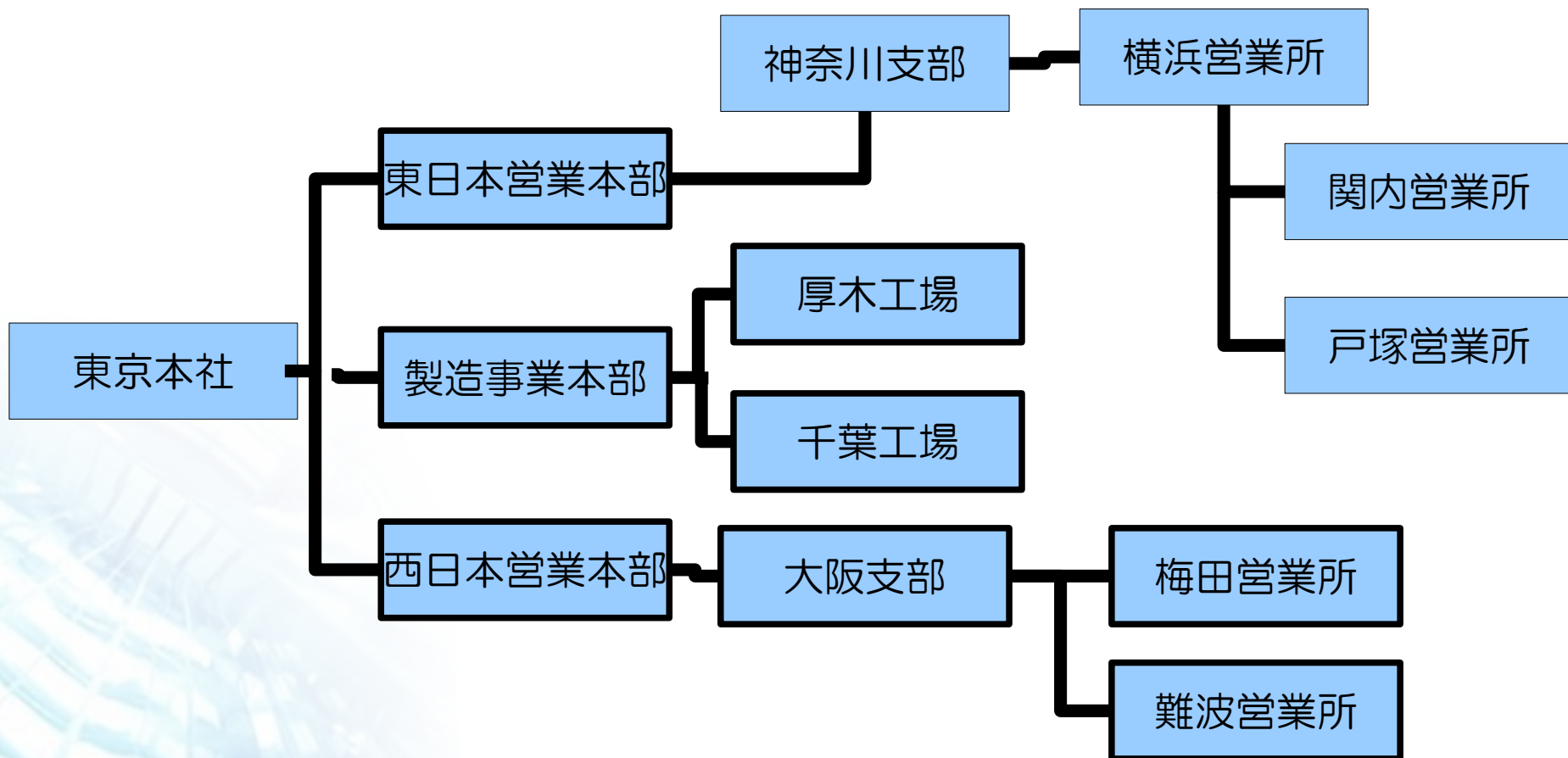
```
WITH RECURSIVE t(n) AS (  
    VALUES (1) -- 非再帰項
```

```
UNION ALL
```

```
    SELECT n+1 FROM t WHERE n < 100 -- 再帰項  
)  
SELECT sum(n) FROM t;
```

```
sum  
-----  
5050
```

サンプルデータ



RDBでの表現

事業所名	管轄事業所名	人員数
東京本社		100
東日本営業本部	東京本社	50
神奈川支部	東日本営業本部	50
横浜営業所	神奈川支部	20
関内営業所	横浜営業所	15
戸塚営業所	横浜営業所	10
製造事業本部	東京本社	50
厚木工場	製造事業本部	200
千葉工場	製造事業本部	200
西日本営業本部	東京本社	50
大阪支部	西日本営業本部	50
梅田営業所	大阪支部	20
難波営業所	大阪支部	20

東日本営業本部以下の人員数を求める

再帰SQLの定義

```
WITH RECURSIVE 東日本営業人員 (事業所名, 人員数) AS
  (SELECT 事業所名, 人員数 FROM 人員構成表
   WHERE 事業所名 = '東日本営業本部'

   UNION ALL

   SELECT PEN.事業所名, PEN.人員数 FROM 人員構成表 AS PEN,
    東日本営業人員 AS EBP
   WHERE PEN.管轄事業所名 = EBP.事業所名)
```

```
SELECT SUM(人員数) FROM 東日本営業人員;
```

sum

145
(1 row)

SELECT本体

組織階層の「深さ」を表示

```
WITH RECURSIVE 東日本営業人員(事業所名, レベル, 人員数) AS
(SELECT 事業所名, 1, 人員数 FROM 人員構成表
WHERE 事業所名 = '東日本営業本部'
```

```
UNION ALL
```

```
SELECT PEN.事業所名, EBP.レベル+1, PEN.人員数 FROM 人員構成表 AS PEN,
東日本営業人員 AS EBP
WHERE PEN.管轄事業所名 = EBP.事業所名)
```

```
SELECT * FROM 東日本営業人員;
```

事業所名	レベル	人員数
東日本営業本部	1	50
神奈川支部	2	50
横浜営業所	3	20
関内営業所	4	15
戸塚営業所	4	10

(5 rows)

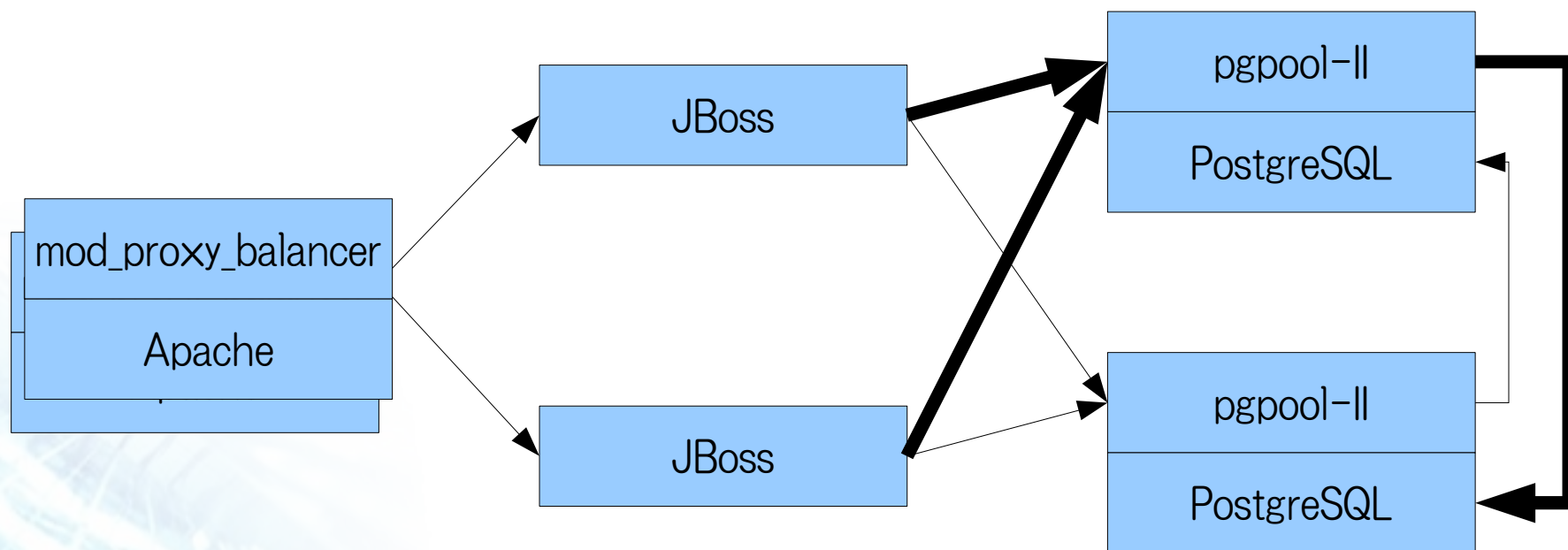
pgpool-II 2.1リリース！

- PostgreSQL専用のオープンソースクラスタソフト
- PostgreSQL 7.4から8.3まで対応
- レプリケーション, 負荷分散, パラレルクエリなどの機能
- 管理用GUIツールあり
- 18万以上のダウンロード実績
- 最新版では, 主に, オンラインリカバリと, 拡張プロトコル回りの安定性の改善

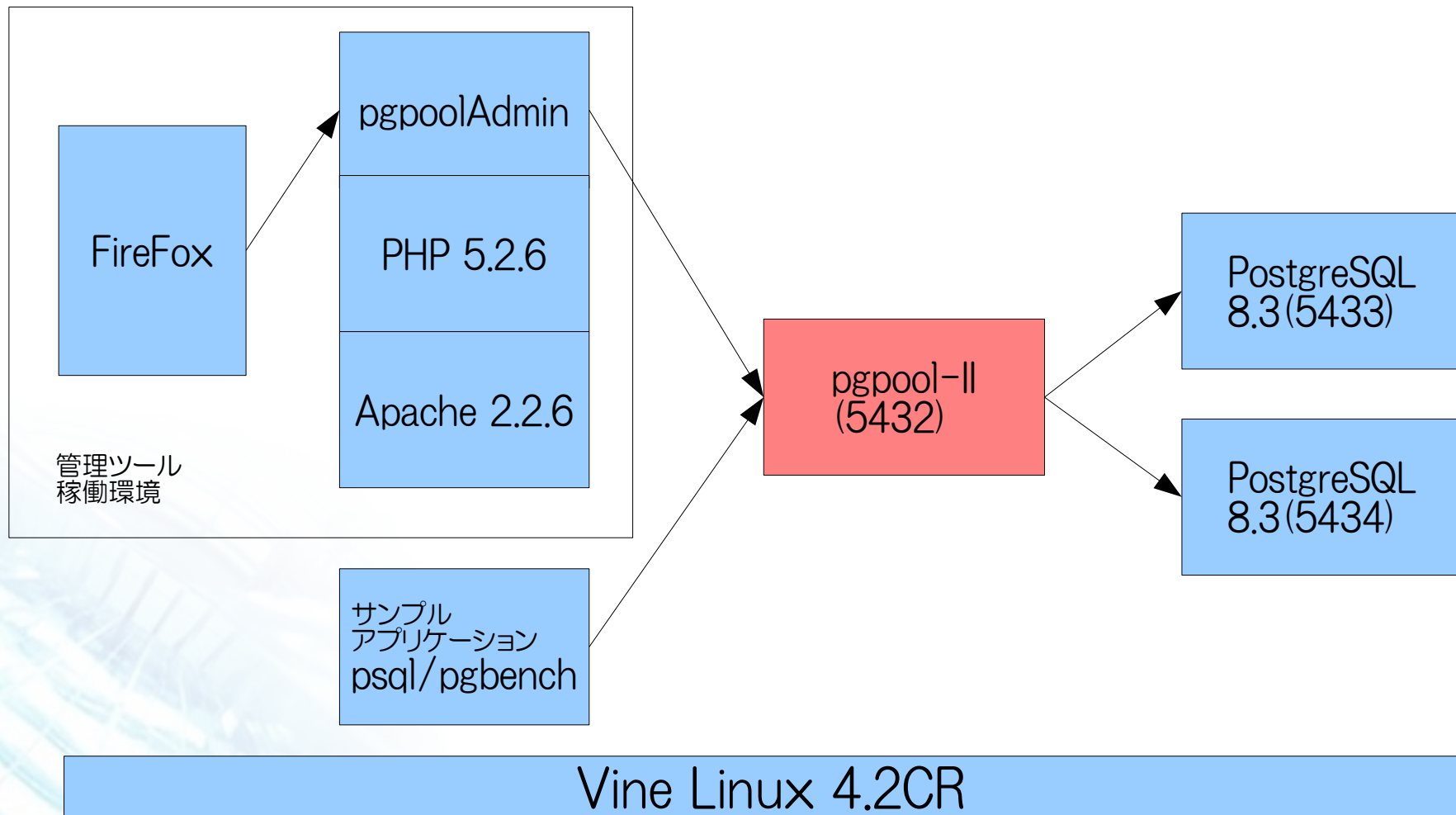
なぜpgpool?

- pgpoolを導入するメリット
 - レプリケーションにより、データのリアルタイムバックを実現
 - 同期型レプリケーションなので、複数サーバ間でデータのずれが起きない
 - 負荷分散機能により、検索系の性能アップが可能
 - PostgreSQLアプリケーションの修正は最小限
 - DML, ラージオブジェクトもレプリケーション可能
 - 運用を止めずにDBの切り離し, 復帰, 追加が可能
- PostgreSQLの可用性, 性能を向上可能

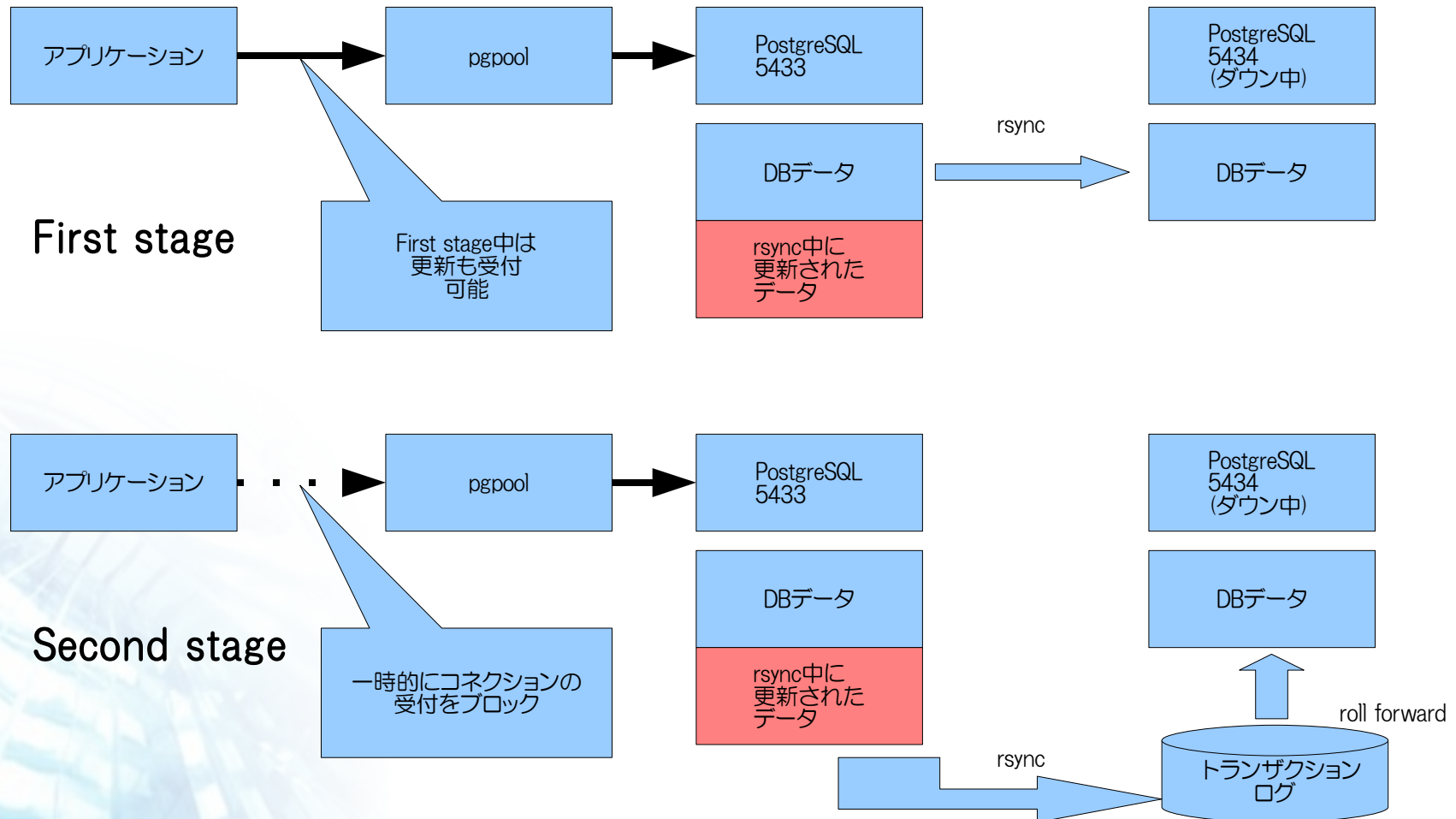
某地方自治体における pgpool-IIの導入事例



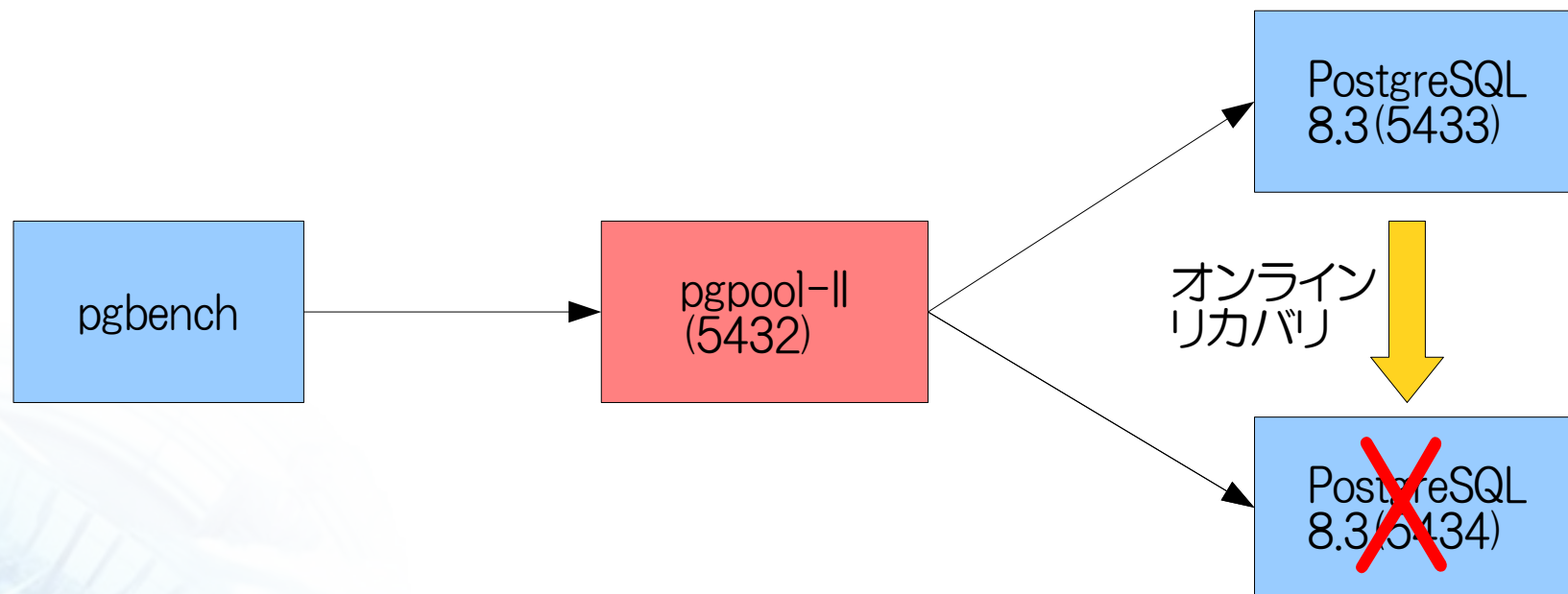
pgpool-II デモシステムの構成



オンラインリカバリの仕組み



デモシナリオ



pgpool-IIの関連情報

- pgpool-II開発サイト
 - <http://pgfoundry.org/projects/pgpool/>
- 「PostgreSQLには絶対pgpool-II」(Think IT)
 - <http://www.thinkit.co.jp/article/98/1/>

ご清聴ありがとうございました！