誕生から15年を迎えたPgpool-IIの過去、現在、そして未来

PgPool Global Development Group
石井 達夫

SRA OSS, INC.

Pgpool-IIの歴史

2003年6月27日: pgpoolの誕生

- コネクションプーリング、フェイルオーバのみ
- サポートするPostgreSQLサーバは2台まで
- Version 2プロトコルのみサポート(まだVersion 3プロトコル=PostgreSQL 7.4はリリースされていなかった)
  - C言語で4,719ステップの規模

ナウマン象(古代の象)

2003年6月27日(金) 22:54:46 JST
[pgsql-jp: 30256] PostgreSQL用コネクションプールサーバ pgpool

石井です.

PHPをはじめ,Perlなど,言語を問わず使える「pgpool」というPostgreSQL用のコネクションプールサーバを作ったので公開します.できたなのでまだアルファ版程度のクォリティですが,よろしかったらお試し下さい.

ftp://ftp.sra.co.jp/pub/cmd/postgres/pgpool/pgpool-0.1.tar.gz

もちろんpgpoolはオープンソースで,ライセンスはPostgreSQLのBSDライセンスと同様のものにしています.

pgpoolを作った動機は,PHPでコネクションプールが使えないことに不満を持ったからです.
一応PHPには「パーシスタントコネクション」というものがあってDBへの接続への接続をキャッシュできますが,少なくともapacheのプロセスの数だけコネクションができるので,DBへ過大な負荷がかかりがちです.
pgpoolを使うとコネクションをキャッシュできるだけでなく,DBへの接続数を適切な数に制限できるので,DBの性能を引き出すことができます.

2004年4月:pgpool 1.0の誕生

- 現在の「ネィティブ・レプリケーションモード」に相当する機能を実装した(まだPostgreSQLにはレプリケーション機能がなかった)
- クエリキャンセル対応
- ラージオブジェクトのレプリケーション対応
- C言語で5,890行
- この頃は、マイナーリリース(x.x)の際にも平気で機能を追加していたりして、かなりいい加減なリリース管理がされていた

現代の象になったがまだまだよちよち歩き

2004年6月: pgpool 2.0へ進化

- 1.0のわずか2ヶ月後にリリース
  - かなり頑張って開発していたようだ
- V3プロトコルにネィティブ対応
- C言語で7,750行
- この後2.5を2005年2月にリリース。ヘルスチェックや、マスタースレーブモードへの対応を追加
  - これでpgpoolとしてのリリースは完了

2006年9月:Pgpool-II 1.0の誕生

- 開発手法の変更
  - 個人プロジェクトから、チーム作業へ
  - IPAの援助で開発
- 機能の大幅追加、現在の姿にほぼ近づく
  - サーバ台数の制限撤廃
  - SQLパーサを搭載して精密な構文解析
  - 管理コマンド(pcp)の実装
  - GUI管理ツール(pgpoolAdmin)の実装
  - パラレルクエリモードの実装
  - C言語で73,511行と、一気に10倍近い規模に増えた(bison, flexコード行数を含む)

Pgpool-II 2.0 – 2.3

- 2007年11月: Pgpool-II 2.0リリース
  - PostgreSQL 8.3パーサ
  - フェイルオーバーコマンドの追加
- 2008年7月: Pgpool-II 2.1リリース
  - recovery_timeoutの追加
- 2009年2月: Pgpool-II 2.2リリース
  - SERIALデータタイプのレプリケーション強化
- 2009年12月: Pgpool-II 2.3リリース
  - 時刻データのレプリケーションが可能に

2010年9月: Pgpool-II 3.0リリース

- PostgreSQL 9.0がリリース、ストリーミングレプリケーションが実装されたのに合わせてストリーミングレプリケーションモードを実装
- ストリーミングレプリケーションに必要な各種機能の実装
  - delay_threshold
  - log_standby_delay
  - white_function_list/black_function_list

2011年9月: Pgpool-II 3.1リリース

- syslogサポート
- アプリケーション名のサポート
- フォローマスターコマンドのサポート
- バックエンドフラグの追加
- ヘルスチェック、ストリーミングレプリケーション遅延チェックのパスワードなどのパラメータ追加

2011年11月pgpool.netへの引っ越し

- それまでホスティングさせてもらっていた pgfoundry の不安定さに手を焼く
- 新しいホストティングサイト pgpool.net を作ることを決意
- pgpool.netをオープン、ソースコード管理も CVS から git に移行した
  - gitへの移行にあたって、フランスのコミュニティのご支援をいただきました

引っ越しはなかなか大変でした

- pgpool.netでは、英語の情報と日本語の情報を同時発信することにした(難しい場合は英語を優先)

その後のPgpool-II

- 2012年 Pgpool-II 3.2リリース
  - watchdogの実装
  - インメモリクエリキャッシュの実装
- 2013年 Pgpool-II 3.3リリース
  - watchdogにheartbeat機能実装
- 2014年 Pgpool-II 3.4リリース
  - ソースツリーの整理
  - メモリ管理、例外処理の導入
- 2015年 Pgpool-II 3.5リリース
  - 拡張問い合わせの高速化
  - watchdogの改良
- 2016年 Pgpool-II 3.6リリース
  - ドキュメントのSGML化
  - watchdogの改良
- 2017年 Pgpool-II 3.7リリース
  - Quorum フェイルオーバの実装
  - AWS Aurora対応

現在の開発体制

- 石井達夫
  - 全体のとりまとめ
  - 日本出身
- Muhammad Usama
  - 3.4の開発からコミッタに就任
  - パキスタン出身
  - 主にwatchdog、認証周りを担当
- 彭 博 (Peng, Bo)
  - 3.3の開発からコミッタに就任
  - 中国出身
  - リリースマネージメント担当、SQLパーサ、PgpoolAdminを担当
- Ahsan Hadi
  - ユーザニーズの取り込み、ベンチマーク
  - パキスタン出身
- 星合拓馬
  - 4.0の開発からコミッタに就任
  - 日本出身
  - ドキュメントなどを担当

現在の開発環境

- Webサイトが活動の中心
  - ソース、PRMのダウンロード
  - Gitリポジトリ
    - 実体は、PostgreSQLがホストしているgitサービス上
  - バグトラッキングシステム
  - メーリングリスト
  - Coverityによるソースコードチェックを利用
- コミッタを中心に、開発を推進
- 日本、パキスタン間で週一の電話会議

リリース/EOLポリシー

- 年に一度メジャーバージョンアップ
  - Pgpool-II 3.7.5
    - メジャーバージョン: 3.7
    - マイナーバージョン: .5
  - 互換性のない機能追加
- 2、3ヶ月に一回バグ修正のためにマイナーバージョンリリース
  - 互換性は維持される
- 各メジャーバージョンは5年間マイナーバージョンアップして保守される
- 現在はCentOS 6, 7用のRPMもリリースしている

Pgpool-IIの現在

- PostgreSQLのクラスタの総合管理ツールに進化
- ストリーミングレプリケーションをより使いやすくすることに注力
  - クエリをプライマリとスタンバイに振り分ける
  - スタンバイに対する検索クエリの負荷分散
  - 自動フェイルオーバ機能の提供
  - スタンバイの昇格管理
- Pgpool-II自体のHA化
  - watchdog

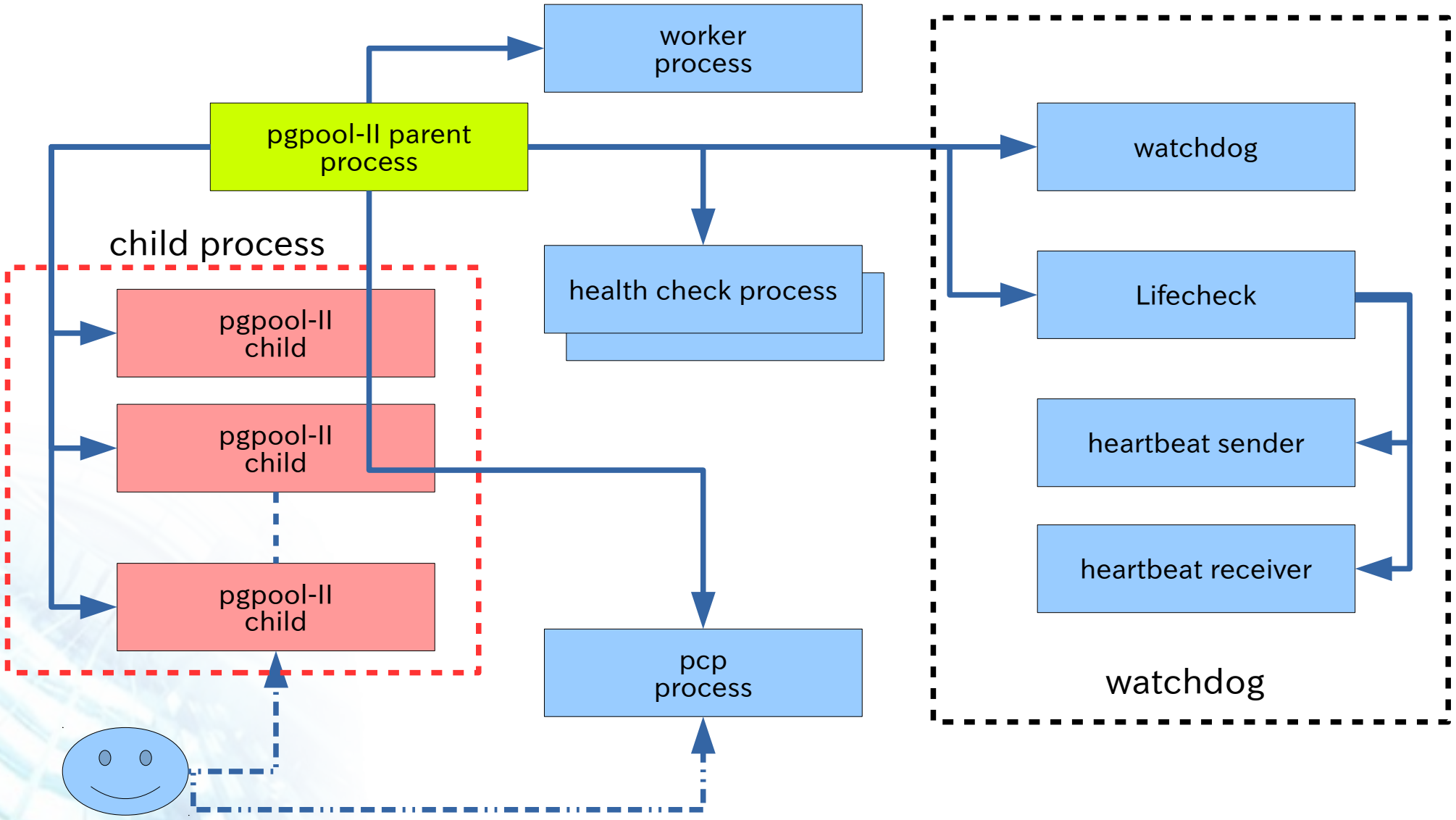
Pgpool-IIの主な機能

性能向上	コネクションプーリング 検索負荷分散 クエリキャッシュ
高可用性	自動フェイルオーバ フェイルオーバスクリプト フォローマスタスクリプト watchdog
クラスタ管理	オンラインリカバリ
クラスタとアプリケーションの親和性	クエリの自動振り分け クエリの振り分けポリシーの設定

Pgpool-IIのアーキテクチャ

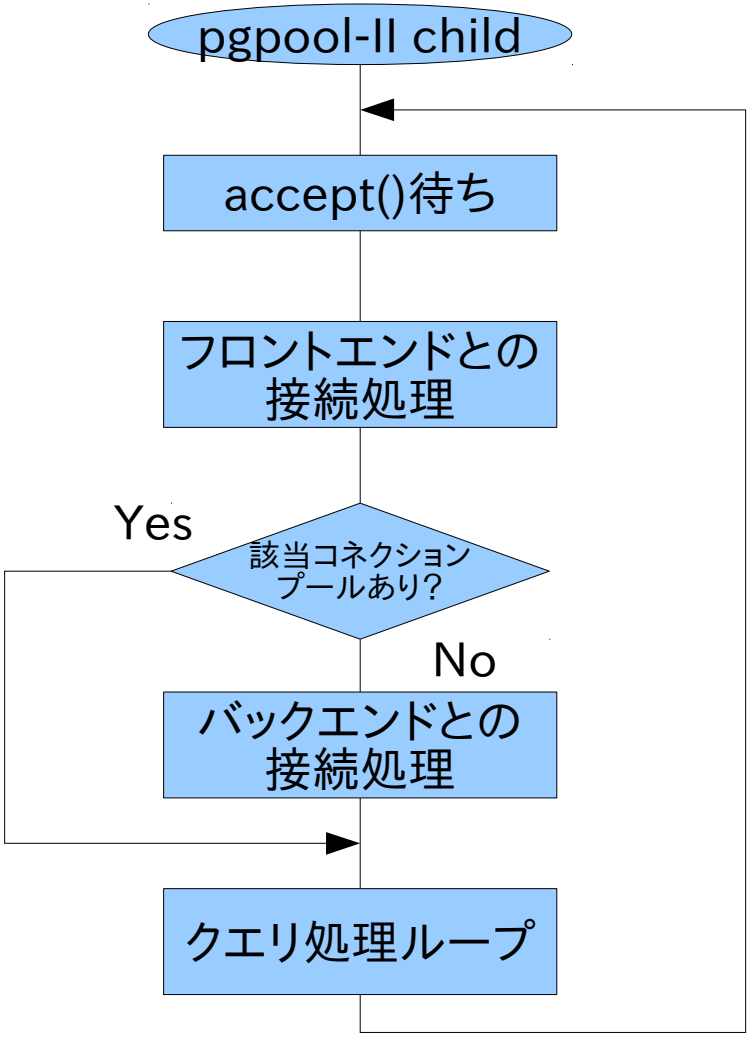
- プロセス構造
- コネクションプーリング
- フェイルオーバ処理
- ヘルスチェック
- 検索負荷分散、watchdogについては後ほどのセッションでお伝えします

Pgpool-IIのプロセス構造

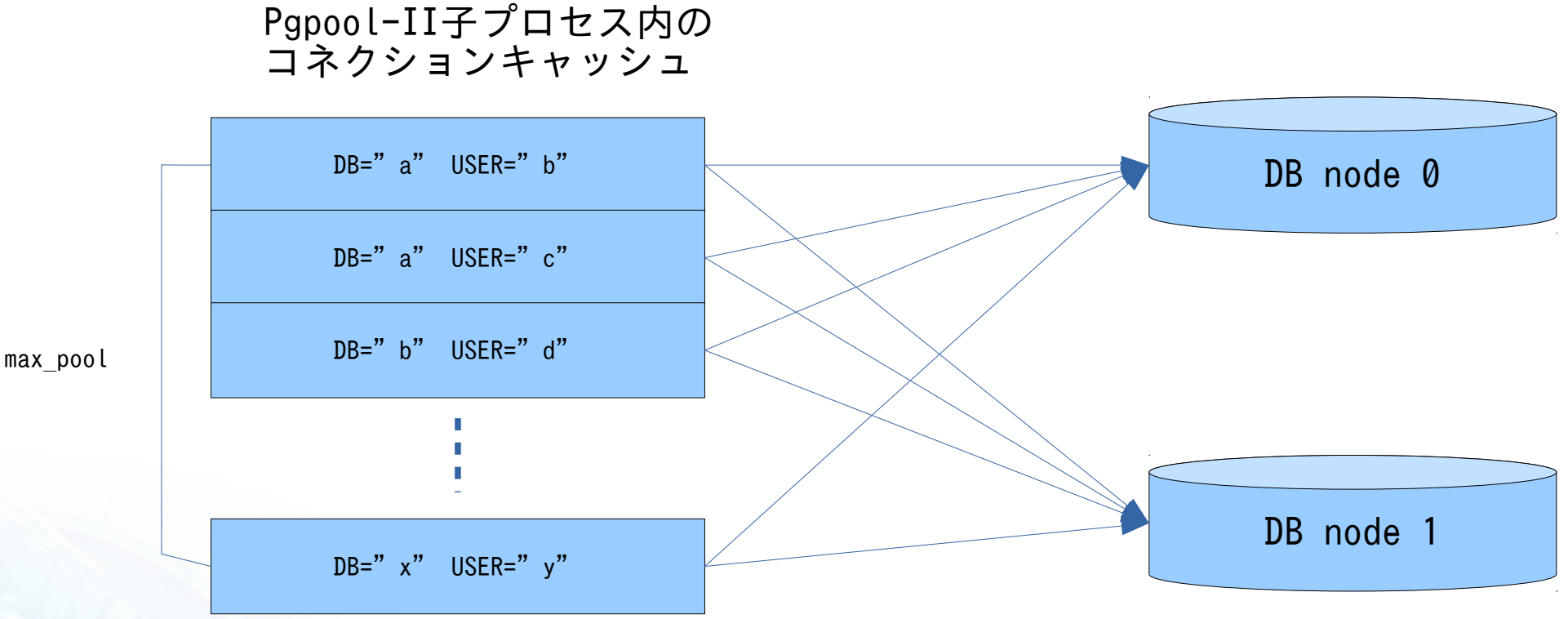


Pgpool-IIの処理概要

- Pgpool-IIでは、親プロセスはlisten()を発行した状態で子プロセスをfork()する
- 子プロセスは、それぞれ個別にaccept()を発行し、フロントエンドからの接続要求があると、OSがどれか一つのpgpool-II子プロセスを選択して処理を渡す
- その子プロセスは、フロントエンドとの接続処理を認証を含めて行い、成功したらすべてのバックエンドと接続を行った後に、フロントエンドからのクエリを受つけてバックエンドに渡す無限ループに入る
- フロントエンドとの接続が切れるか、終了メッセージを受け取ると無限ループを抜けて再びaccept()待ちに入る



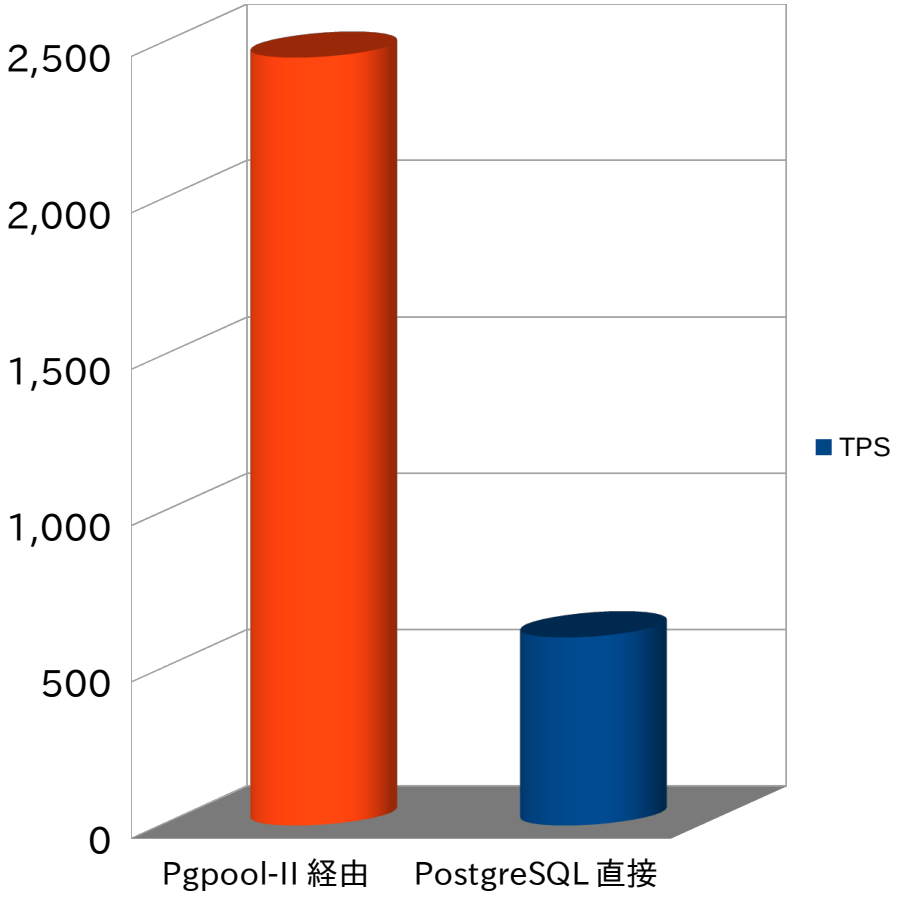
コネクションプーリングの仕組み



max_poolを使い切ると一番古いコネクションが
開放され、そのスロットが再利用される
(LRU管理)

コネクションプーリングの意義

- PostgreSQLは接続に時間がかかる
- DBに接続しっぱなしにして接続時間を節約
- コネクションプーリングの有効時間を設定することも可能
- Javaなどの環境では自前のコネクションプーリングを持っていることがあり、その場合は効果はない

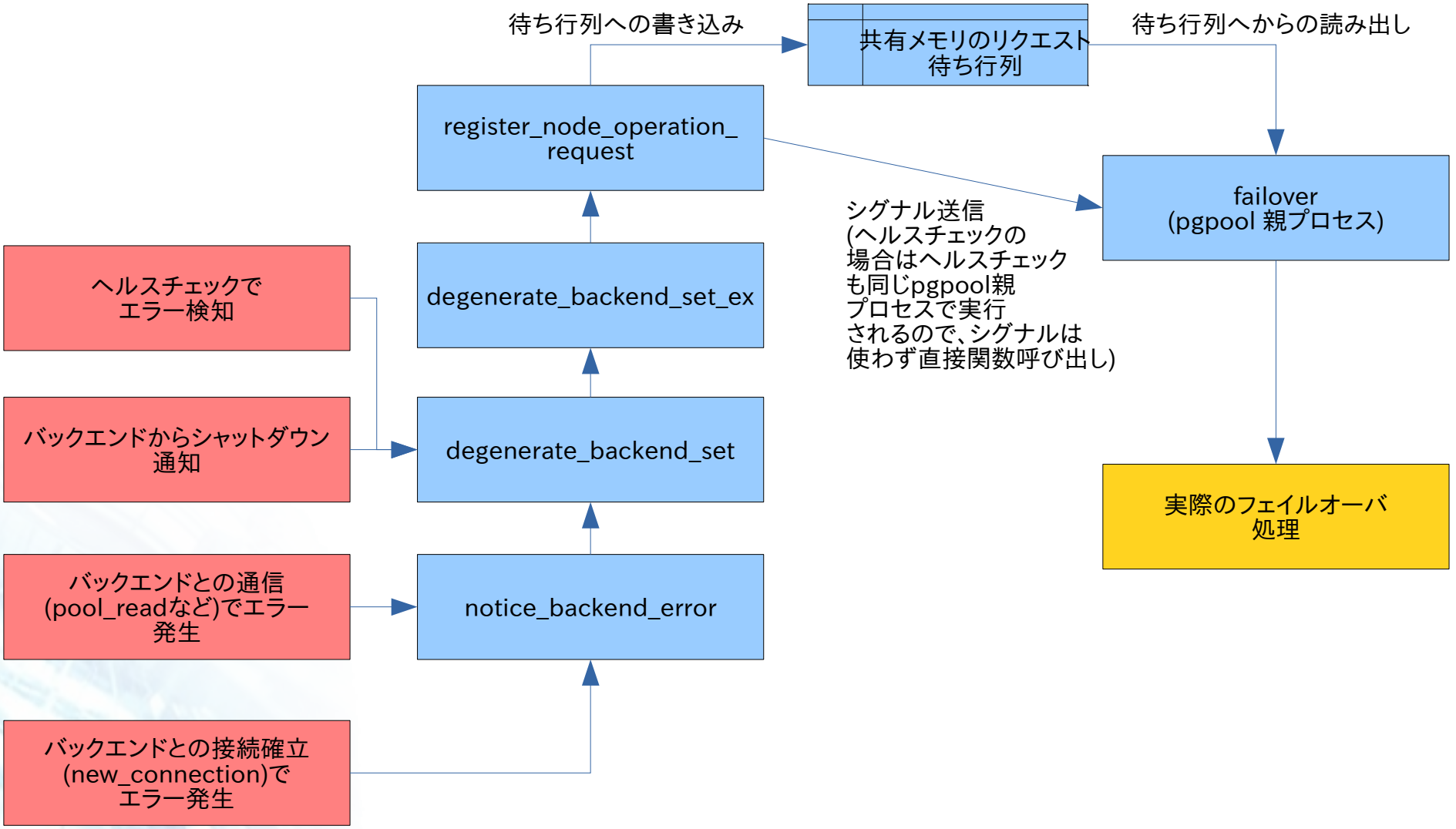


測定条件: PostgreSQL 10.5+Pgpool-II 4.1 beta1
pgbench -S -C -c 1 -T 30

フェイルオーバ処理の概要

- フェイルオーバの引き金はいくつかある
  - ヘルスチェック
  - バックエンドからシャットダウン通知
  - バックエンドとの通信でエラー
  - バックエンドとの接続確立時にエラー
- フェイルオーバ処理は主にpgpoolメインで行われる
  - pgpool子プロセスの再起動
  - 共有メモリ上のステータスの変更
  - フェイルオーバスクリプトやフォローマスターコマンドの起動

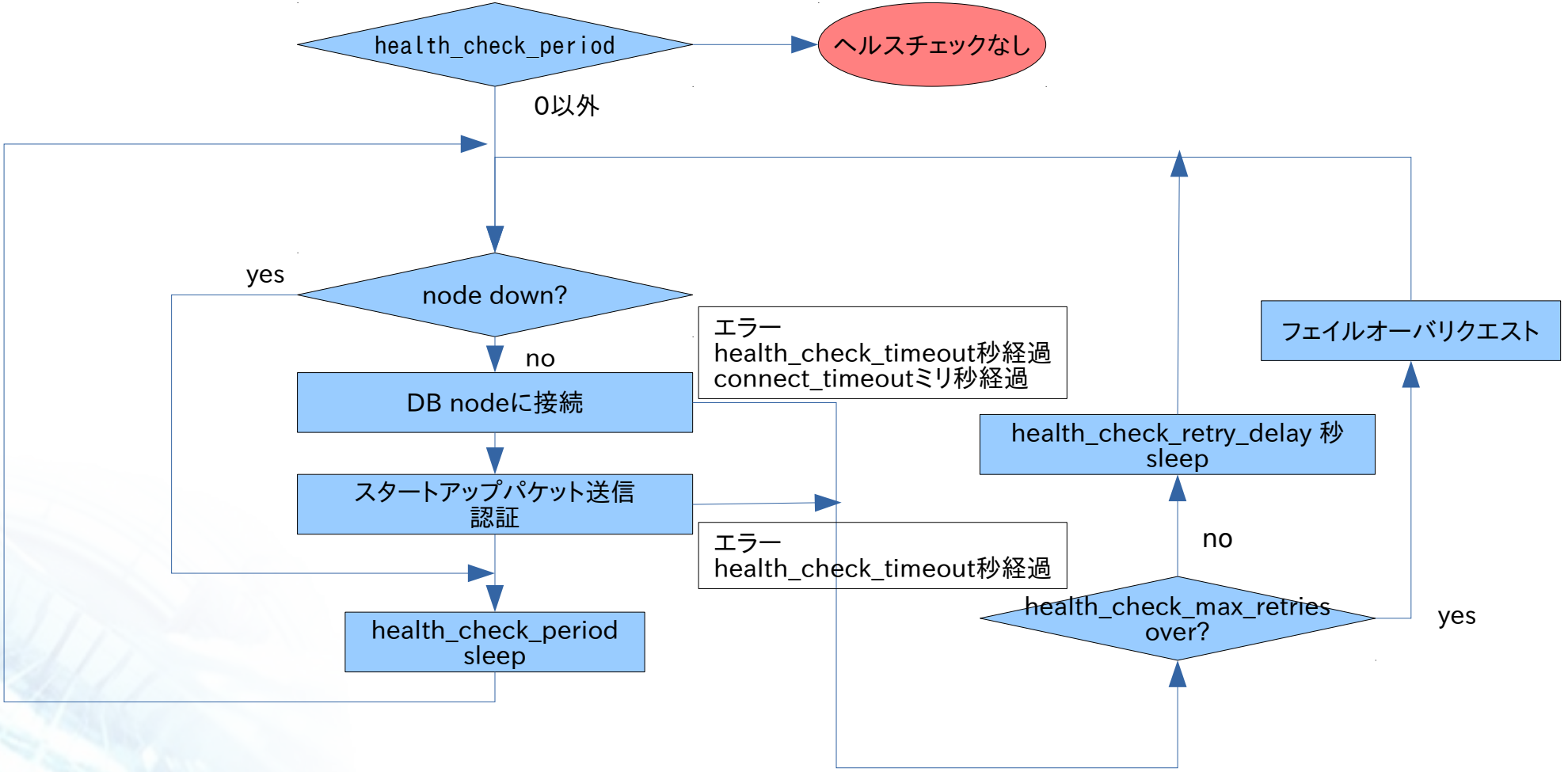
フェイルオーバ発生の契機と通知の流れ



ヘルスチェック

- ヘルスチェックプロセスの中で実施されるDBの監視処理
- すべてのバックエンドに対して、スタートアップパケットを送信し、応答が返ってきたことをもって正常と判断する
- それ以外は異常と見なす
- したがって、バックエンドがダウンした時だけでなく、ネットワークに異常がある場合にも異常と見なされる
- ネットワークに一時的なエラーが発生する可能性がある場合は、ヘルスチェックのリトライを設定するのが効果的
  - Pgpool-II 3.7からは、クォーラムフェイルオーバー機能を使うことで局所的なネットワーク障害による誤ったフェイルオーバーを防ぐことが可能に

ヘルスチェック処理と関係パラメータ



SRA OSS, INC.

2018年Pgpool誕生から15年を迎えて

- Pgpool-IIの次期バージョンは4.0に決定
  - 現在beta1。10月中旬の正式リリースを目指して鋭意開発中!
- 「4.0」にふさわしい充実した新機能の実装を目指しました
  - 検索クエリ負荷分散の設定項目の追加/改良
    - black_query_list
    - disable_load_balance_on_write
    - db_redirect_preference_list
    - app_redirect_preference_list
  - 認証機能の大幅強化
    - SCRAM, Cert認証に対応
    - AES256暗号化パスワード
  - PostgreSQL 11 SQLパーサの移植
  - ソースコードのツールによるフォーマット(インデントなど)
- Pgpool-II 4.0の新機能については、次のセッションでどうぞ!

今後の展望

- PostgreSQLの進化に合わせて機能を追加
  - PostgreSQLにシャーディングなどの機能追加があれば、対応していきたい
- 品質の向上のため、テストの充実
  - regression testの拡充
  - プロトコルレベルの詳細なテスト(pgprotoの利用)
- 内部処理のチューニング、高速化
- 認証機能の充実
  - SCRAM認証、Cert認証の進化
- その他、アイデア募集!

最後に

- 開発に参加してくださる方を募集しています！
- Pgpool-IIの開発に参加するメリット
  - PostgreSQLよりは敷居が低い
    - 規模は1/10位
    - それでいてSQLパーサや例外処理のような重要な部分がポートされているので、PostgreSQLの理解に役立つ
  - ネットワークプログラミングや、マルチプロセスプログラミングの経験蓄積に最適

Thank you!