



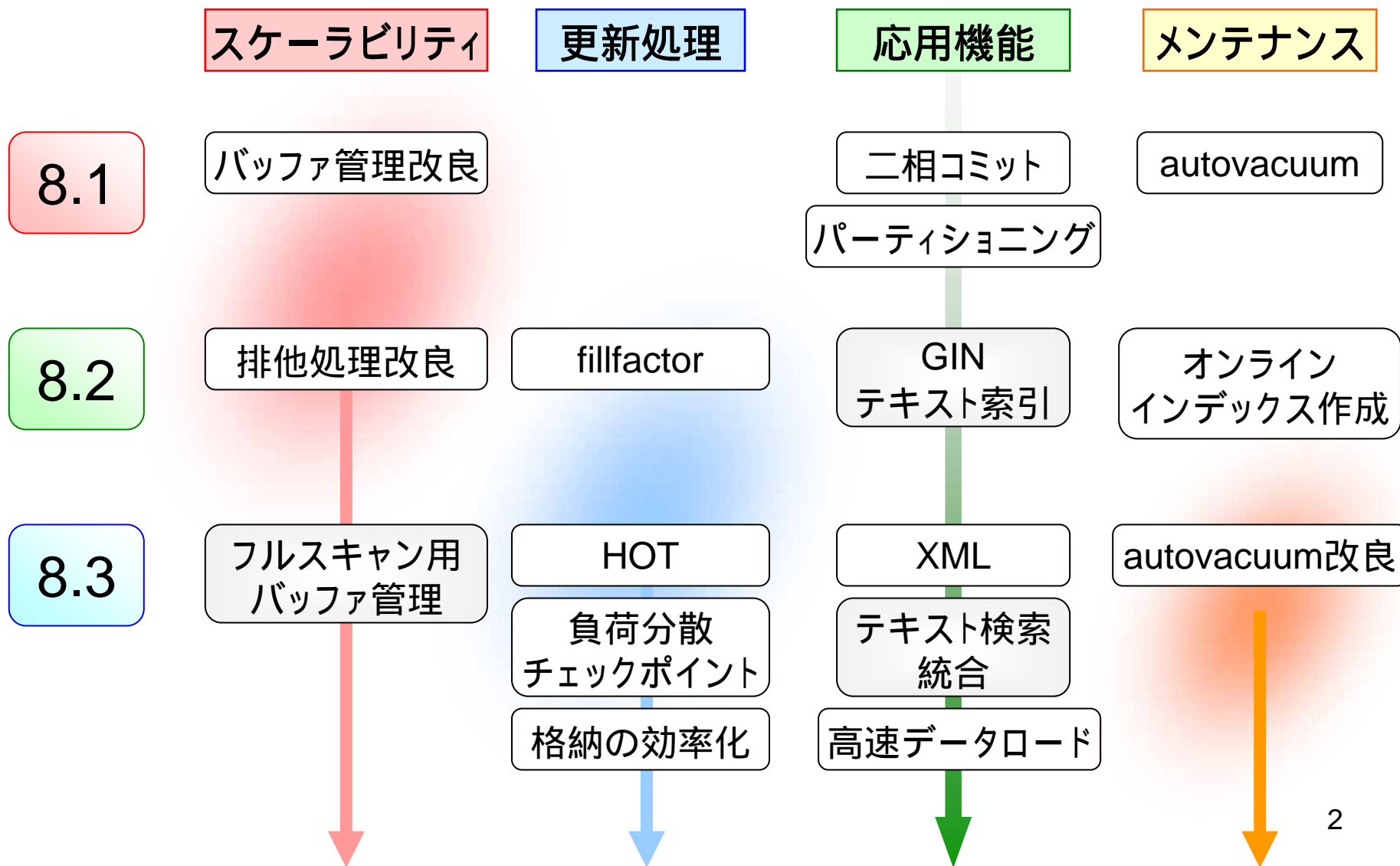
PostgreSQL Updates

~ 8.3は更新も速い~

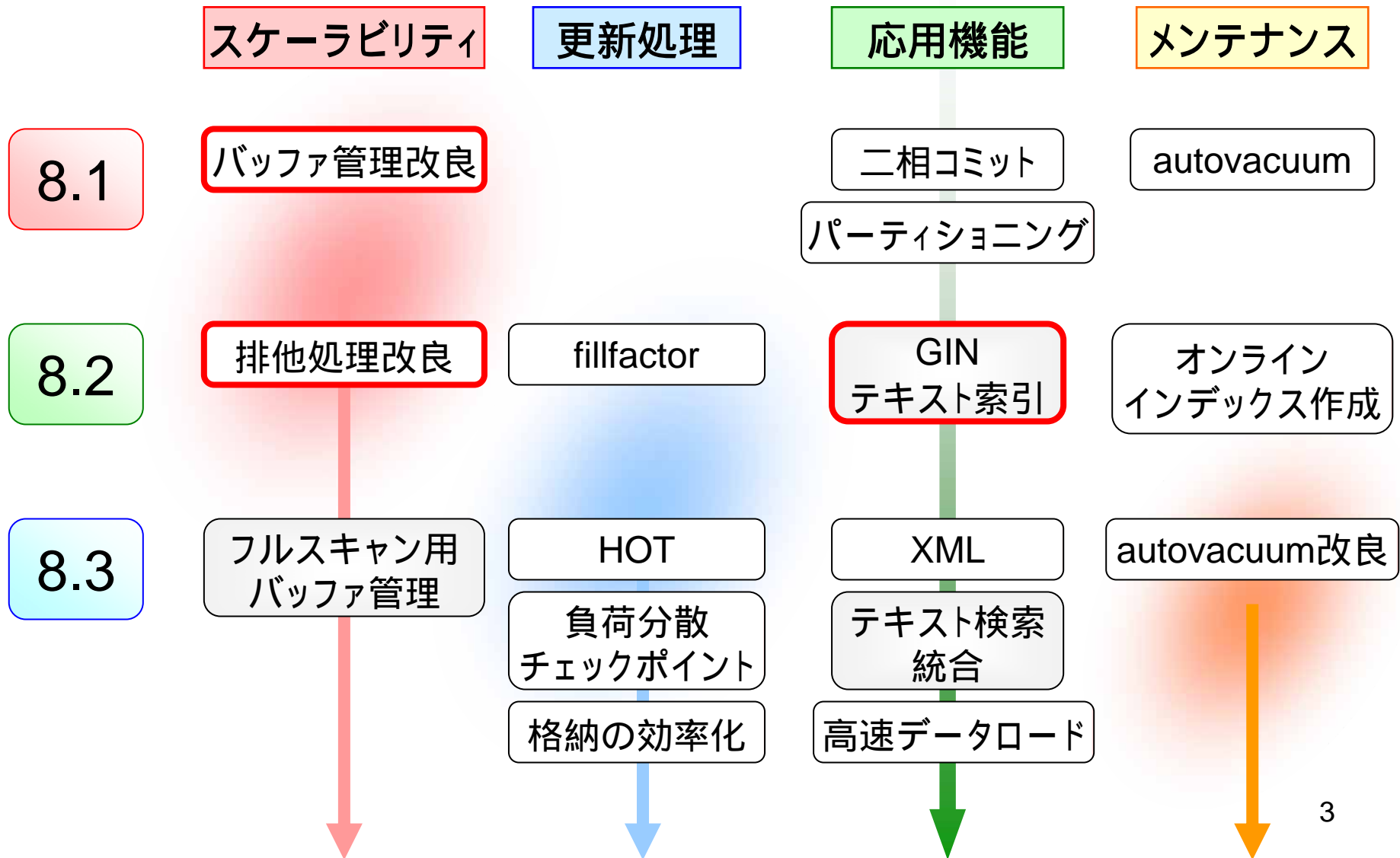
2007.6.23

日本PostgreSQLユーザ会
板垣貴裕

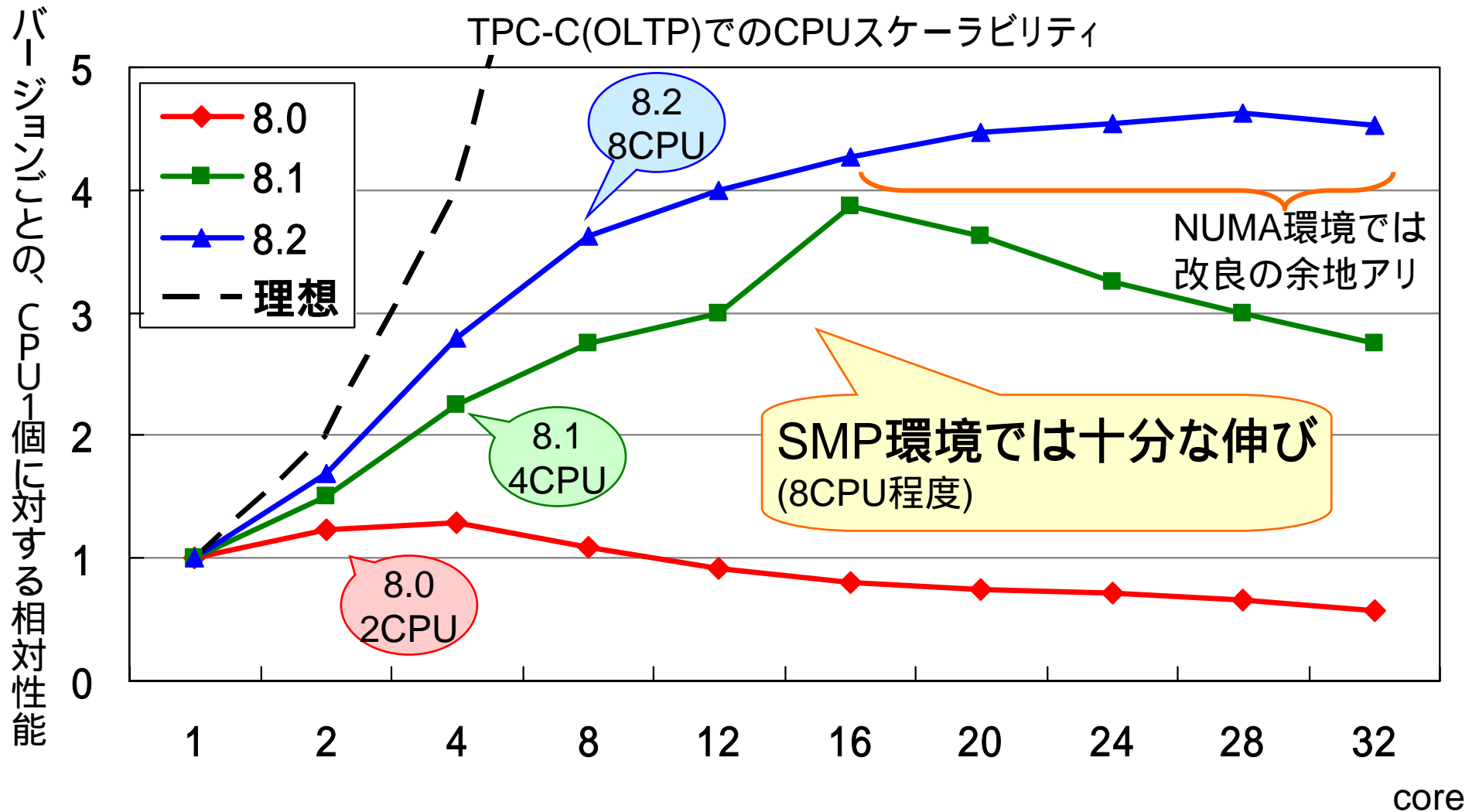
8.1 ~ 8.3 での主要な改善項目



8.1 ~ 8.2 の改善項目おさらい



おさらい: CPUスケールラビリティ



おさらい: 全文テキスト検索

- tsearch2 全文テキスト検索を可能にする拡張

例: `SELECT isbn,title FROM books`

`WHERE fts @@ to_tsquery('word1 & word2');`

- GiST, GIN: 特性の異なる2種類のテキスト検索インデックス
- 共にトランザクション, リカバリをサポート

tsearch2 GiST版 と GIN版 の性能比較

| | 参照 | 更新 | 作成 | サイズ |
|-------------|--------|---------|--------|--------|
| GiST | 112ms | 280ms | 176s | 146MB |
| GIN | 39ms | 3344ms | 532s | 306MB |
| 比率 | × 0.35 | × 11.94 | × 3.02 | × 2.10 |

使い分けを推奨

- 更新が多い GiST
- 参照が多い GIN

- 日本語を扱うには、ひと手間が必要。まだ「公式」サポートは無い

おさらい: VACUUMによるゴミ処理

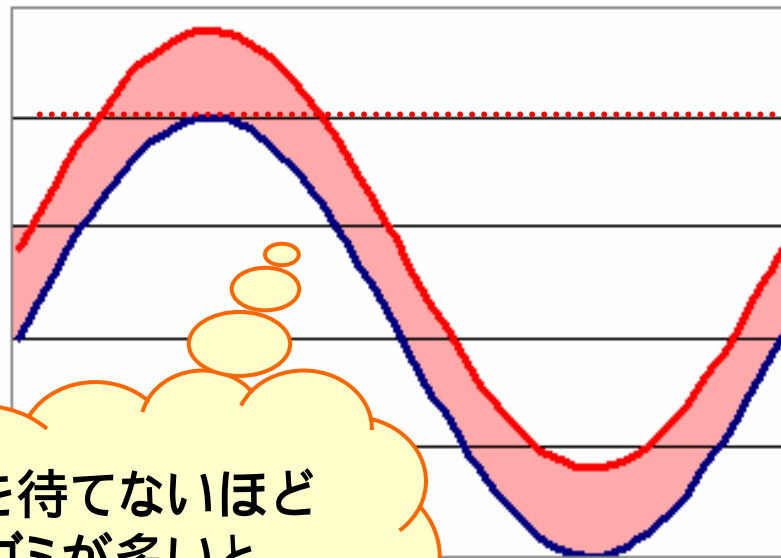
- 「追記型」= UPDATEでゴミが発生する
 - 忙しいときは「ゴミを気にしない」
 - 暇になったら「後からゴミを処理」

ゴミ処理

主処理

サーバの
リソース
使用率↑

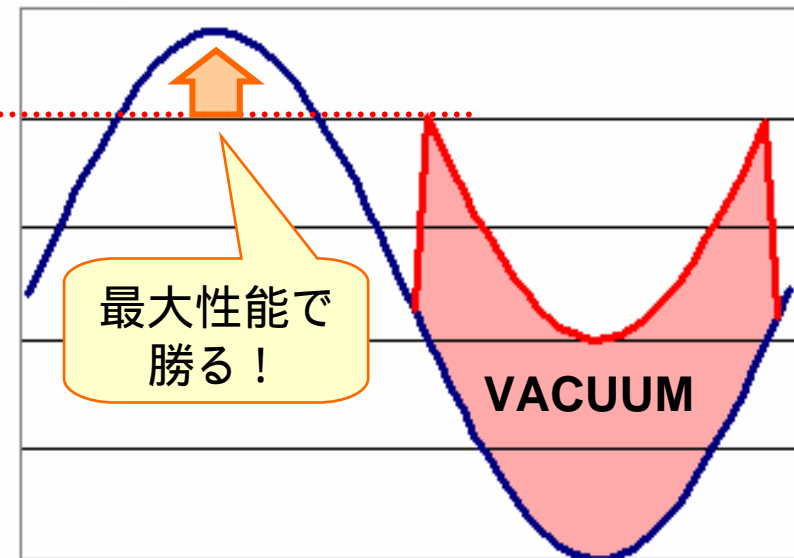
その場でゴミ処理する場合



暇を待てないほど
ゴミが多いと
利点を活かさない!



後からゴミ処理する場合



時間→

PostgreSQL 8.3 の目玉

1. 更新が速い！

- HOTでOLTPも得意に

2. 性能が安定！

- 負荷分散チェックポイント

3. データロードが速い！

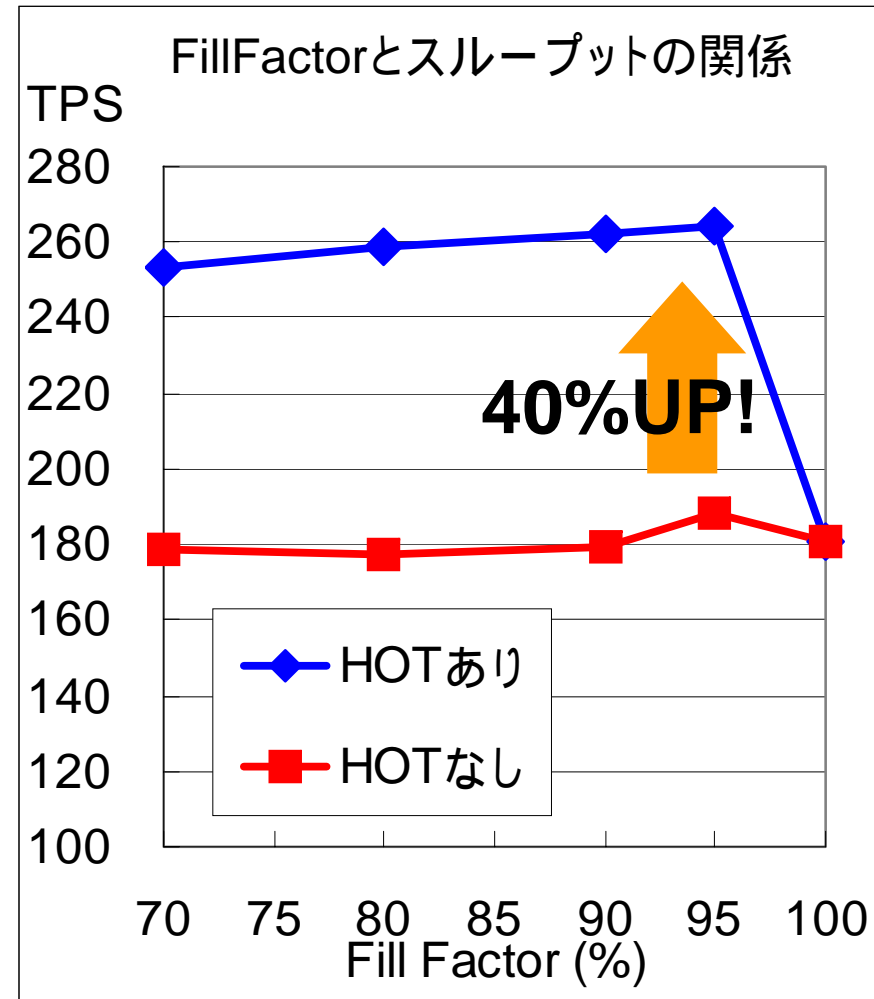
- データロード時にWALをスキップ

4. VACUUMが手間いらず！

- VACUUMは改良型autovacuumにおまかせ

(1) 8.3は更新が速い！

- 新機能「HOT」の効果
 - 更新処理を効率化
 - pgbenchにてスループットが40%向上
- 効果
 - インデックスの更新をスキップ
 - VACUUMを待たずにゴミ処理
 - VACUUMの回数を減らせる
- ポイント: FillFactor
 - 初期状態でページの隙間を残す
 - 100%(デフォルト)よりも90~95%ほどが良い
 - ALTER TABLE *name* SET (fillfactor=95);



pgbench -s400 (5GB)

NTT OSS Center 調べ

HOT インターナル (1)

- 8.2までのUPDATE時の動作

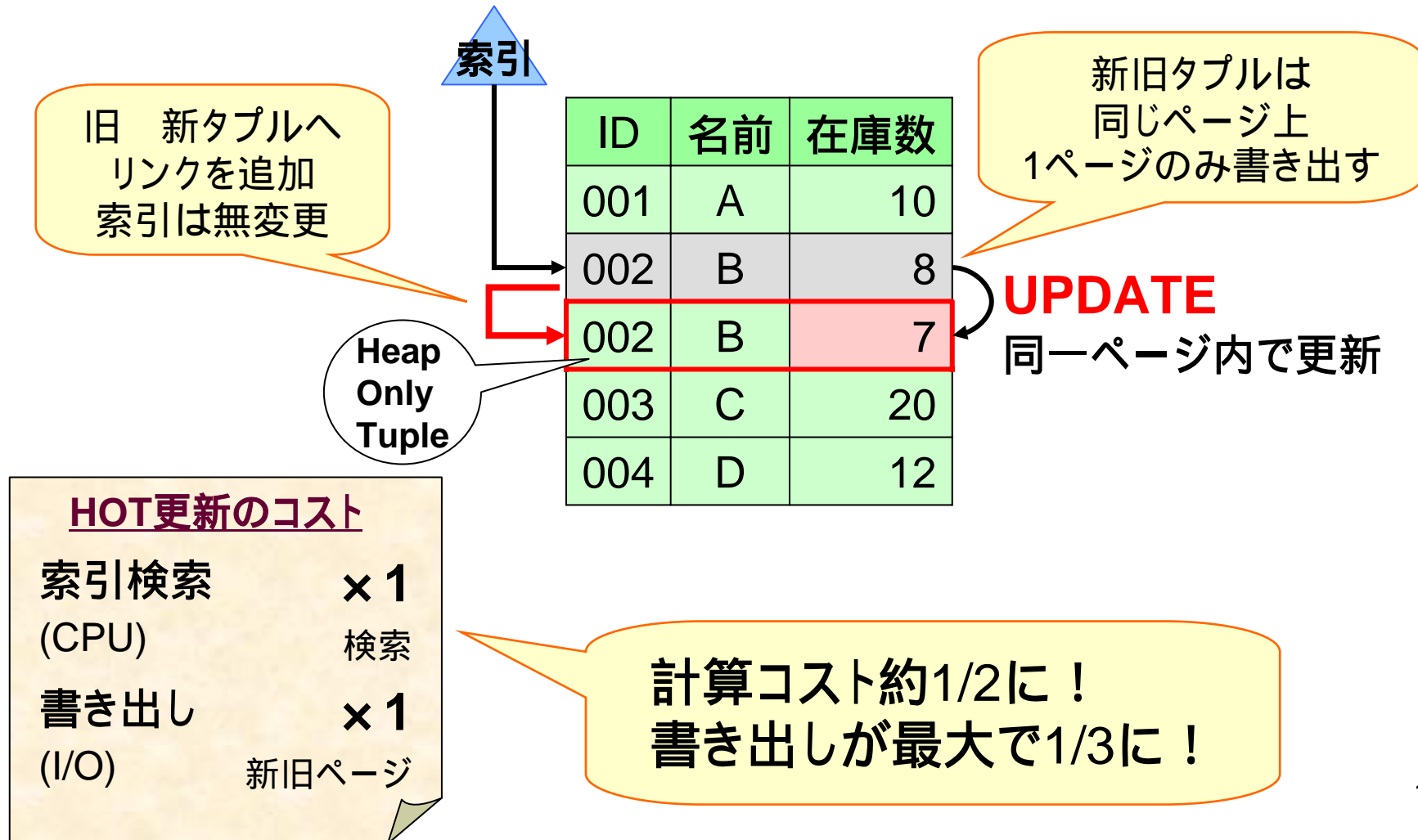
索引

| ID | 名前 | 在庫数 |
|-----|----|-----|
| 001 | A | 10 |
| 002 | B | 8 7 |
| 003 | C | 20 |
| 004 | D | 12 |

索引が張られていない
列「在庫数」のみを
変更する場合でも...

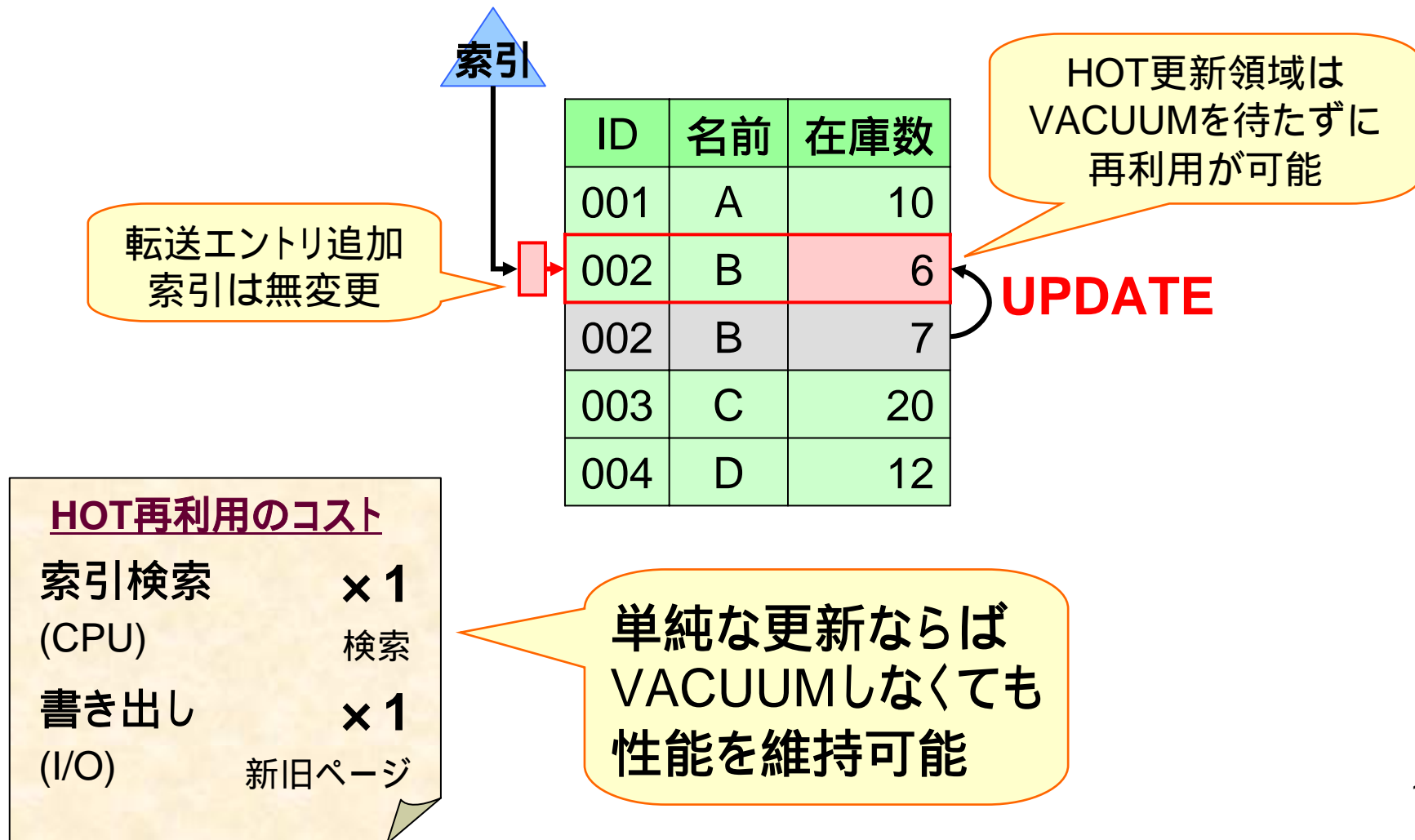
HOT インターナル (3)

• 8.3 HOT UPDATE時での動作



HOT インターナル (4)

- 8.3 HOT UPDATEの後、不要領域の再利用



HOT更新が利用可能な条件

- UPDATEである
 - DELETE+INSERTはダメ
- インデックスが張られた列を更新しない
 - 「単純な更新」限定だが、この使い方は多い
 - 使われていないインデックスは削除する
- 同一ページ上に空き領域がある
 - FillFactorを調整するとベスト
 - 一度VACUUMをかけるだけでもよい

計算コスト、書き出しコストを節約
VACUUMの必要性も削減

PostgreSQL 8.3 の目玉

1. 更新が速い！

- HOTでOLTPも得意に

2. 性能が安定！

- 負荷分散チェックポイント

3. データロードが速い！

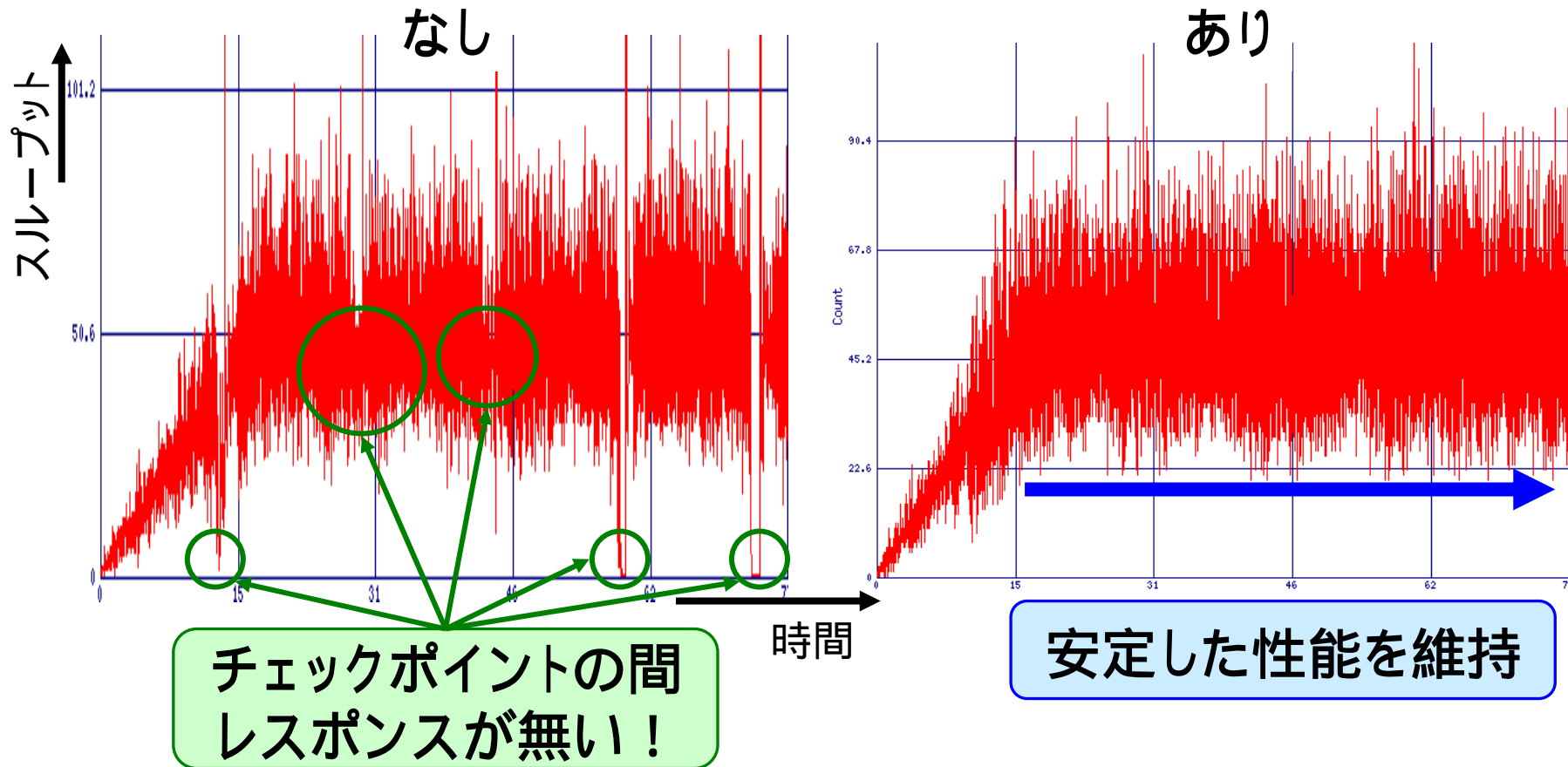
- データロード時にWALをスキップ

4. VACUUMが手間いらず！

- VACUUMは改良型autovacuumにおまかせ

(2) 8.3は性能が安定！

- 負荷分散チェックポイント (LDC)



負荷分散チェックポイント インターナル

- チェックポイント時の動作
 - 全ての変更されたページを書き出す (write)
 - 全ての変更されたファイルを同期する (fsync)

8.2 上記の処理を全速力で行う

8.3 上記の処理を合間にスリープを挟みながら行う

いたって単純

ただし、PostgreSQLに導入された
初の自発的なI/Oスケジューリング
今後のI/O処理の改善に期待

PostgreSQL 8.3 の目玉

1. 更新が速い！

- HOTでOLTPも得意に

2. 性能が安定！

- 負荷分散チェックポイント

3. データロードが速い！

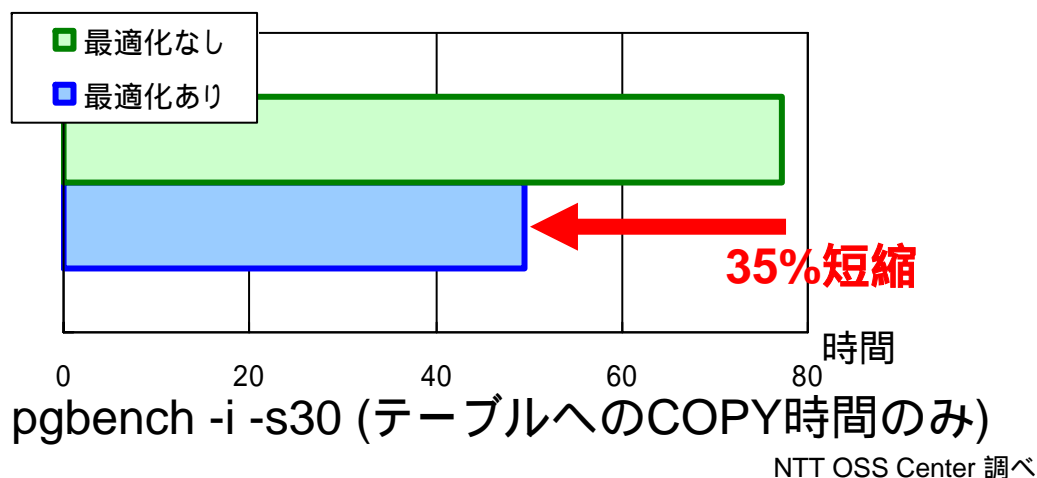
- データロード時にWALをスキップ

4. VACUUMが手間いらず！

- VACUUMは改良型autovacuumにおまかせ

(3) 8.3はデータロードが速い！

- 新規テーブルへのロードが高速化
 - WALをスキップすることでI/O量が約1/2に



高速化するおまじない

```
BEGIN;  
TRUNCATE t;  
COPY t FROM ...;  
COMMIT;
```

- 既存テーブル (インデックス付き) へはまだ苦手
 - 周辺ツールで補う
 - InfoFrame DB Maintenance (NEC)
 - pg_bulkload (オープンソース, pgFoundry)

8.3対応よろしく！

(4) 8.3はVACUUMが手間いらず！

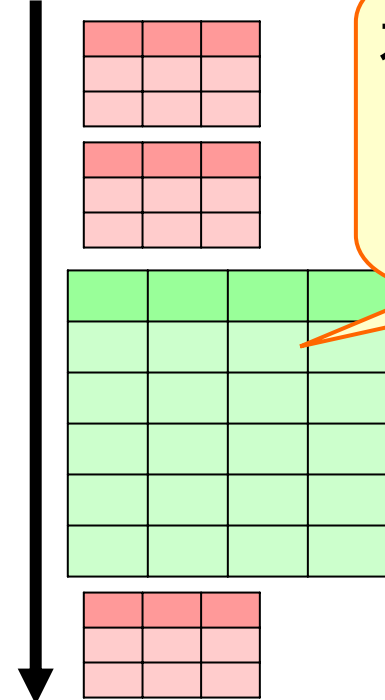
- autovacuum によるVACUUMが柔軟になった
 - 複数プロセスを割り当てられる (autovacuum_max_workers)

8.2

常に1プロセス

VACUUM不足の可能性

プロセス



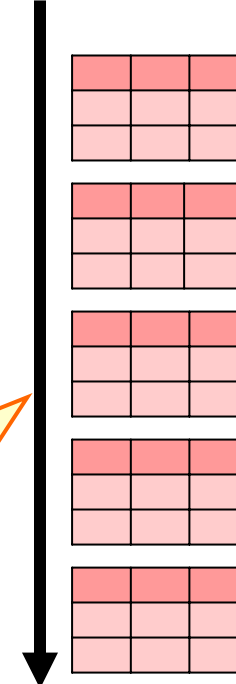
大きなテーブルへの
VACUUM中に
小さなテーブルの
処理が滞る

過不足なく
平行処理できる

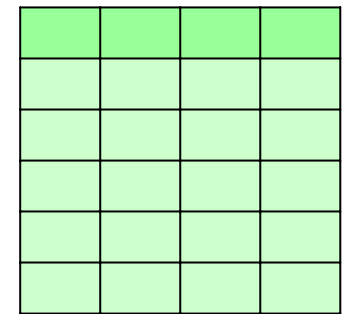
8.3

複数プロセス

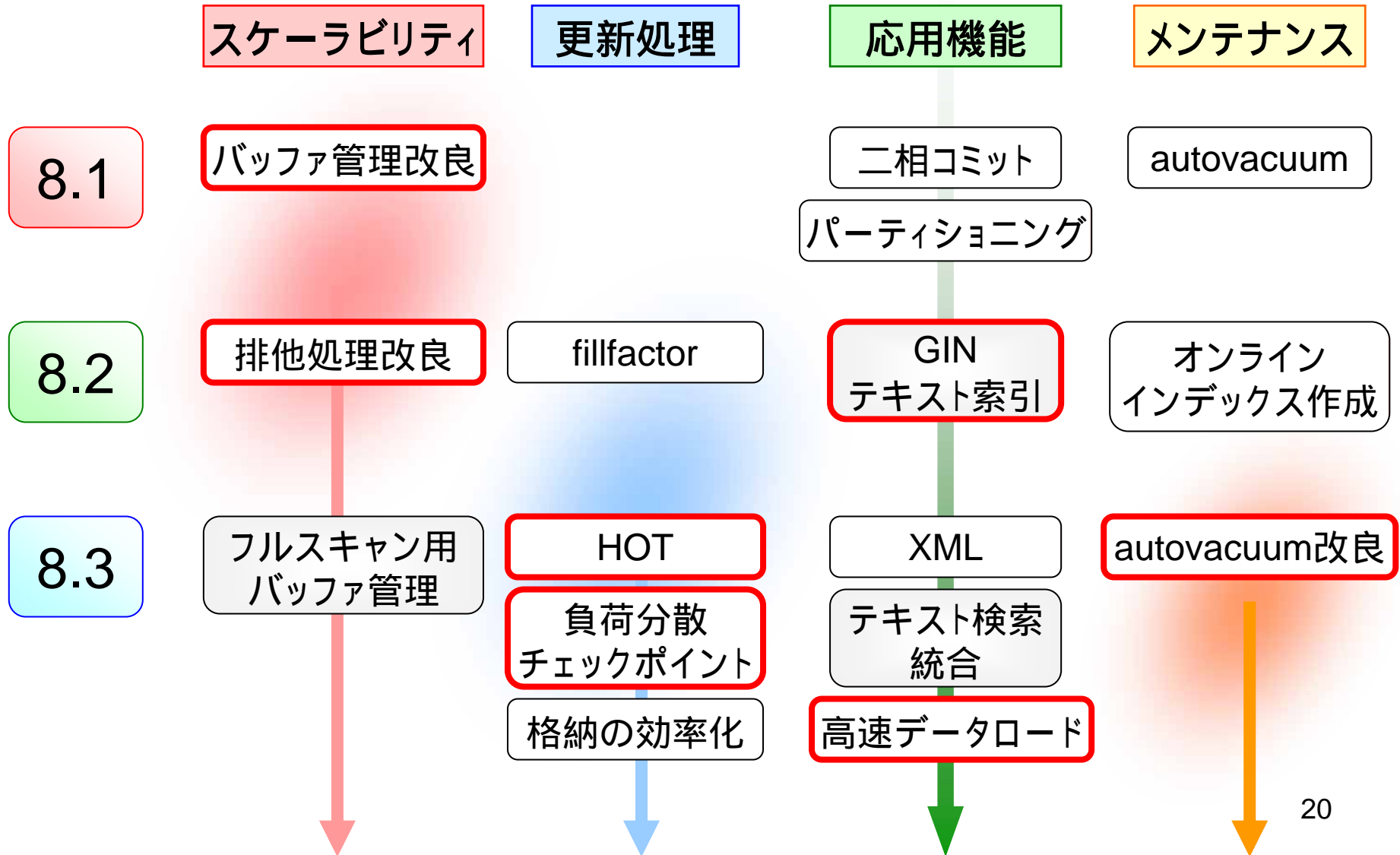
プロセスA



プロセスB



紹介した主要な改善項目



まとめ

PostgreSQL 8.3 には期待大

- もう、更新が遅いとは言わせない！
 - HOTを初めとする各種OLTP向けの改良
- チューニングと運用の手間が軽減
 - VACUUM, Background Writer (チェックポイント用)
- その他 様々な機能追加や改善
 - SQL/XML, 実行プランアドバイザ, 遅延コミット, etc.

PostgreSQL 8.4 の足音が聞こえる...

- HOTのリミッタ解除
 - 実装をシンプルにするために、現状あえて最適化していない箇所も
- autovacuumスケジューラ
 - メンテナンス時間帯を考慮したVACUUMの実施
- ホットスタンバイ型 負荷分散レプリケーション
 - WALをスタンバイ側へ転送 & スタンバイ側で参照処理を許可



PostgreSQLとコミュニティの動向 - PGCon2007を通して -

2007.6.23

日本PostgreSQLユーザー会

笠原 辰仁



目次

- PGCon2007の概要
 - Tutorial
 - 各セッション・BOF・Lightning Talk
- 海外コミュニティの動向
 - ユーザ
 - 周辺ツール
 - 性能面への注目・他
- 国内コミュニティの動向
 - JPUGとイベント
 - ユーザ貢献
 - 注目されている課題
- まとめ

PGCon2007

- PostgreSQLのあれこれを話し合う1年に1回のカンファレンス
 - 今年は120人くらい？が世界中から参加
 - 日本からは、JPUGのメンバー含め9名ほど参加
 - 5/21 ~ 5/24 @ Ottawa University
- 5/21、22
 - Tutorial(1日2本)
- 5/23、24
 - 各種セッション
 - BOF(Birds Of a Feather)
 - Lightning Talk



PGCon2007 集合写真



http://www.postgresql.org/files/community/pgcon07/pgcon_group_photo.html



Tutorial

- Introduction to Hacking PostgreSQL
 - PostgreSQLの内部の俯瞰
 - Hacking&コミュニティへのフィードバックについて
 - レビューアーの不足の話が...
- Performance Whack-a-Mole
 - 性能問題の90%は、10%の要因を直せば改善(多くはアプリ由来)
 - 用途別のボトルネック問題とその解析方法
- Pro PostgreSQL
 - PostgreSQLのDBAとして知っておくべきこと
 - PostgreSQLに限らず、どのDBMSにも通じる話
- PI/Perl - Best of Both Worlds
 - PL/perlの活用方法



各種セッション・BOF・Lightning Talk

- keynote(総括的な話題など)・・・3件
 - Great Steps in PostgreSQL History
 - PostgreSQL Inside
- ツール(周辺ツールなど)・・・5件
 - pgsnmpd
 - PGCluster-II
- 事例(移行事例など)・・・6件
 - Developing New Zealand's Electoral Roll on PostgreSQL
 - Migrating to PostgreSQL in Academia
- コア(本体に関連した話題など)・・・10件
 - Full Text Search in PostgreSQL
 - PostgreSQL performance research

海外コミュニティの動向 - ユーザ -

- PostgreSQLへの移行事例など
 - 大規模環境、24H365D、その他もろもろ・・・
 - 更なる実例が求められている
- なぜPostgreSQLを使うのか？
 - コスト面で有利
 - コミュニティが活発
 - エンタープライズ機能が比較的充実
 - ライセンスがビジネスフレンドリー
 - 開発者がPostgreSQLを扱いなれている！
- 性能面で他のDBMSより優位性がある・・・という話も

海外コミュニティの動向 - 周辺ツール -

- レプリケーションは注目度がかなり多かった分野
 - 昨年と同じく、PostgreSQLを束ねて使う方法が注目されている
- 以下、紹介されたレプリケーションツール
 - PGCluster-II
 - Oracle RAC相当の機能
 - 使えるようになるにはまだ時間がかかりそう？
 - pgpool-II
 - パラレルクエリ
 - Slony-I
 - 非同期のマスタ/スレーブ型
 - 現在マルチマスタ型となる”Slony-M”開発中らしい
 - Sequoia
 - JDBCラッパ
 - Bucardo
 - PL/perl利用
 - 同期型/非同期型を選択可能
 - ODBC Link

海外コミュニティの動向

- 性能面への注目 -

○ 安定性

- チェックポイントやVACUUM中も安定した性能にしたい
 - 現状(8.2系)では、チューニングにより回避可能
 - 8.3でさらに改善される予定！

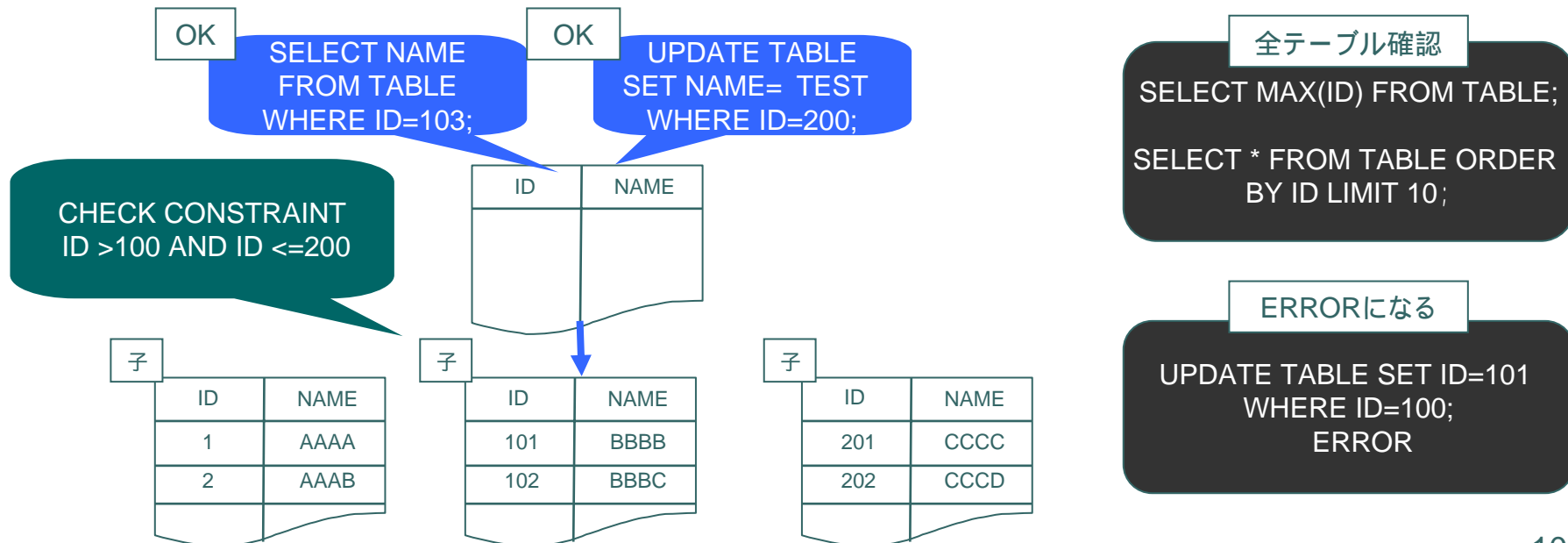
○ チューニング

- チューニングノブが充実してきた
- チューニング技法が今後の課題か？
 - いずれは自律的チューニングを可能にしたい

海外コミュニティの動向

- 大規模環境への注目 -

- 大規模対策として、パーティショニングが注目されている
 - 制約を考慮して、8.1:SELECT、8.2:UPDATEの際に特定のテーブルだけのアクセスをするよう改良
 - しかし、まだ一部の処理は全テーブルを見てしまう or ERROR
 - min/max
 - order by (パーティショニングキーでのソート)
 - パーティショニング跨りの更新



国内コミュニティの動向

- JPUGとイベント -

- JPUG(日本PostgreSQLユーザー会)
 - 世界各国にあるPostgreSQLのユーザーの集まりの中で最も？規模の大きいコミュニティ
 - 会員数は約3000人、日本全国で9つの支部
 - フランスに大きいコミュニティがあるらしい
 - 海外のイベントと比べると、若干企業色が強い？
 - PostgreSQLカンファレンス2007 と PGCon2007を比較しての印象
- 最近のJPUGの活動(2007年1月以降)
 - 海外コミュニティとの交流
 - PostgreSQL技術セミナー2007春
 - コアメンバーのMagnus Hagander氏を招待
 - PostgreSQLカンファレンス2007
 - コアメンバーのJosh Berkus氏を招待
 - PGCon2007へのJPUGメンバー参加
 - 国内でのオフライン活動
 - JPUG 2006年度合宿
 - OSC2007/Spring
 - OSC2007.DB

国内コミュニティの動向 - ユーザ貢献 -

- 事例・評価
 - 1ヶ月25億PVのサイトでPostgreSQLを活用(リクルート)
 - DBT-1、DBT-3によるPostgreSQLの性能評価(IPA)
- ツール開発&OSS化
 - PostgreSQLで不足している機能を補うツール達
 - 運用
 - InfoFrame(NEC) 商用
 - pg_bulkload(NTT)
 - pgsnmpd(JPUGさいとうさん Joshua Tolley)
 - レプリケーション
 - pgpool-II(SRA)
 - PGCluster-II(SRA)
 - セキュリティ
 - ISO版PostgreSQL(NTTデータ)
 - SE-PostgreSQL(海外さん)
 - 8.4から？

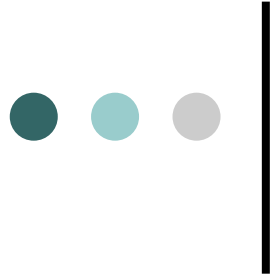
国内コミュニティの動向 - 注目されている課題 -

- 開発者層が薄い
 - 日本 本家へのフィードバックが少ない
 - 開発者支援のための委員会をJPUGで設置
- 機能や仕様が不十分なものがある
 - ログ出力メッセージのエラー体系が不明
 - トラブル時のアクションが分かりづらい
 - 日本語問題
 - 8.3でコアに入る全文検索(tsearch2)がマルチバイト文字には未対応



まとめ

- PostgreSQL本体の開発動向
 - 性能面での不安はかなり払拭されてきた
 - PostgreSQL performance research(PGCon2007)
 - IPA(PostgreSQLカンファレンス2007)
 - 大規模環境、性能安定性、運用面の充実化
 - (機能)負荷分散チェックポイント、VACUUM改良、自動パーティショニング機能
 - (ツール)監視ツール、高速ローダ、レプリケーションツールの多数出現
- コミュニティ
 - 世界レベルでの活動が活発に！
 - コア部分の開発者の増員が課題
 - 実例の更なるアピールを求む



ご清聴ありがとうございました