

高トランザクションシステムへの Postgres Plus適用事例

株式会社富士通ソーシャルサイエンスラボラトリ

2014年11月21日

佐藤 誠

- 富士通ソーシャルサイエンスラボラトリのご紹介
 - 自己紹介
 - 企業情報
 - OSS関連サービス・取り組み
- Postgres Plus適用事例のご紹介
 - Postgres Plus適用概要
 - 適用課題と対応
- 所感
- ノウハウのご紹介

- 富士通ソーシャルサイエンスラボラトリのご紹介
 - 自己紹介
 - 企業情報
 - OSS関連サービス・取り組み
- Postgres Plus適用事例のご紹介
 - Postgres Plus適用概要
 - 適用課題と対応
- 所感
- ノウハウのご紹介

■ 名前

■ 佐藤 誠

■ 略歴

■ 2005/04(入社) - 2009/10

• パッケージ開発

■ 2009/11 - 2012/05

• アプリ開発

■ 2012/06 -

• OSSミドルウェア構築、インフラ構築

■ 得意分野

■ アプリケーション開発、インフラ構築

• Java、C#

• PostgreSQL

■ ちなみに

■ 日本全国での同姓同名ランキング

第20位

■ 社長に多い名前

第1位 

- 社名
 - 株式会社富士通ソーシャルサイエンスラボラトリ（略称：富士通SSL）
 - <http://www.ssl.fujitsu.com/>
 - <http://www.facebook.com/FujitsuSSL>
- 事業所
 - 武蔵小杉本社
 - 関西事業所（大阪）
 - 東海事業所（名古屋）
 - 東海事業所刈谷分室
- 設立
 - 1972/07/12
- 社員数
 - 1,252名（2014年3月末現在、連結ベース）
- 事業内容
 - SI事業
 - ソリューション事業
 - PoweredSolution
- 関係会社
 - 株式会社SSLパワードサービス



FUJITSU 富士通ソーシャルサイエンスラボラトリ

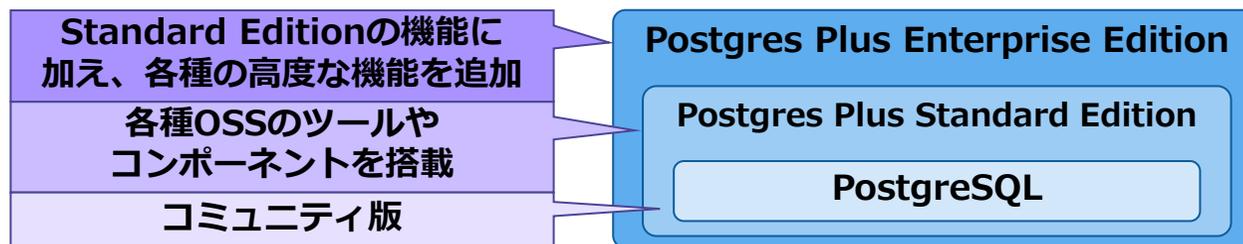
■ PoweredSolution

PoweredSolutionは、お客様のビジネスの成功を強力にプロデュースします。
 当社は、業務系・情報系から、共通インフラまで、幅広い範囲のソリューションを横断的に提供し、
 お客様のビジネスの成功を強力にプロデュースします。



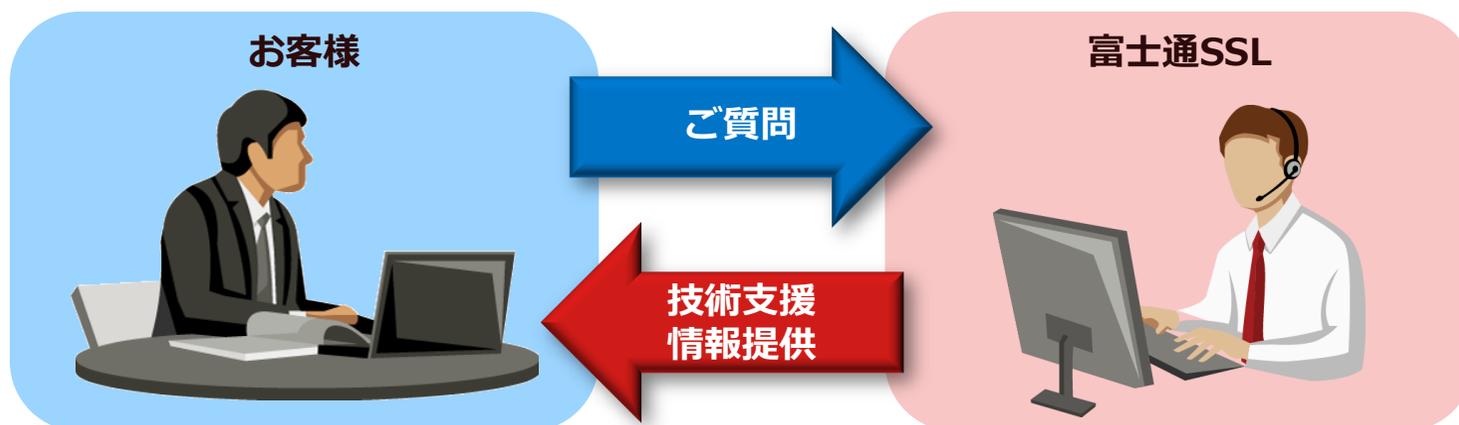
■ Postgres Plus サポート・サービス (<http://www.ssl.fujitsu.com/products/oss/msupport/postgresplus.html>)

- PostgreSQLをベースに開発された「Postgres Plus」のサポートを提供します。
- 一般的な利用において必要となるOSSのツール、コンポーネントなどをまとめた「Standard Edition」と、商用DBとの互換性や性能を重視して機能追加された「Enterprise Edition」をご用意しております。



■ OSSサポートサービス (<http://www.ssl.fujitsu.com/products/oss/support/>)

- お客様に安心してOSSをご利用いただくために、OSSのサポート、セキュリティ情報・アップデート情報の提供、お問い合わせによる技術支援などを行います。





当社は「LPI-Japanビジネスパートナー制度」へ参加し、OSS-DB認定者の技術向上に取り組んでおります。



私たちは、全員取得。
 (株)富士通ソリューションズラボラトリーは、社内でOSS-DBの資格取得を推進しています。

Gold
 Gold認定は、大変なぶん、得るものも多い。
 PostgreSQL 9.4.5
 (株)富士通ソリューションズラボラトリーの内務課への研修受講より、資格が早く取得でき、嬉しく入賞でした。しかしその研修受講にしっかりと取り組んでいます。

Gold
 自身の技術力を高め、さらに向上を目指すために。
 PostgreSQL 9.4.5
 (株)富士通ソリューションズラボラトリーのサポート研修について、研修員としての経験から、今後は研修の参加に合わせて研修-研修の参加を推進できるようにしていきたいです。

Silver
 技術者として、資格取得に励みます。
 PostgreSQL 9.4.5
 開発経験と技術力を伸ばすための一歩であり、資格取得を目指して取り組んでいきたいです。そのための研修受講、受講の資格取得、受講の受講を希望します。

当社は2005年からOSS-DBを含むOSSソフトウェアのサポートサービスをご提供しています。昨年発表したPostgreSQLの普及推進組織「PostgreSQLエンタープライズ・コンソーシアム」にも無償個人企業として参加、その推進活動の一環として、普及推進の一助となるために認定者になるべく努力を怠りません。貴社との連携にもなることを考え、この資格が実際に役立つこと、また全員で一つの目標に向かうことで研修が活性化し、両社関係に繋がることが期待しています。
 (株)富士通ソリューションズラボラトリー 代表取締役 山口 昌幸

DBスペシャリストを認定する資格 / OSS-DB技術者認定試験

OSS-DB
 Open Source Software Database Professional Certification

Silver Gold

株式会社創設者
 OSS-DB推進員
 エルピーアイジャパン
 LPI-Japan

TEL: 03-5689-4482
 Fax: 03-5689-4483
 https://www.oss-db.jp/
 E-mail: info@lpi.jp

OSS-DB 検定

- 富士通ソーシャルサイエンスラボラトリのご紹介
 - 自己紹介
 - 企業情報
 - OSS関連サービス・取り組み
- Postgres Plus適用事例のご紹介
 - Postgres Plus適用概要
 - 適用課題と対応
- 所感
- ノウハウのご紹介

■ 社名

- 株式会社ISAO (ISAO Corporation)
 - <http://www.isao.co.jp/>

■ 所在地

- 東京都台東区浅草橋5-20-8 CSタワー7階

■ 設立

- 2010年2月3日 (創業: 1999年10月1日)

■ 事業内容

- サービス企画/開発/運営事業
- **課金/決済代行事業**
- サーバー構築/運用事業
- カスタマーサポート事業



企業理念

“たのしい！”をうみだしとどける

■システム概要

課金/決済システム

- オンラインの課金決済や会員管理、ポイント管理
- 24時間365日の連続稼働
- 数千万～1億トランザクション/日
- 数百GB以上のデータ量

■作業内容

- OS・各種ミドルウェアのバージョンアップ
- システムの信頼性・安全性を向上

■システム障害

論理障害の発生により、下記が必要となる。

- PostgreSQLのサポート
- 周辺ツールのサポート
- 脆弱性情報の提供
- PostgreSQLの専門知識

■Postgres Plus Advanced Server適用理由

- PostgreSQLを利用
- PostgreSQLの高度なサポート付き
- PostgreSQLの周辺ツールが付属（サポート込み）
- PostgreSQLのプロ
（Core Team、Committers、Contributors）が在籍
- Postgres Plus独自で性能改善している
- セキュリティ情報提供あり
- バグの修正パッチ提供あり（緊急度の高いもの）

■ご紹介できない内容

- 課金/決済システム内容
- システム詳細構成（ハードウェア・ソフトウェア）
- パラメータ設定内容
- 使用技術の詳細
- 性能情報

■製品名

Postgres Plus Advanced Server (PPAS) は、
Postgres Plus Enterprise Edition (PPEE)
に読み替えてください

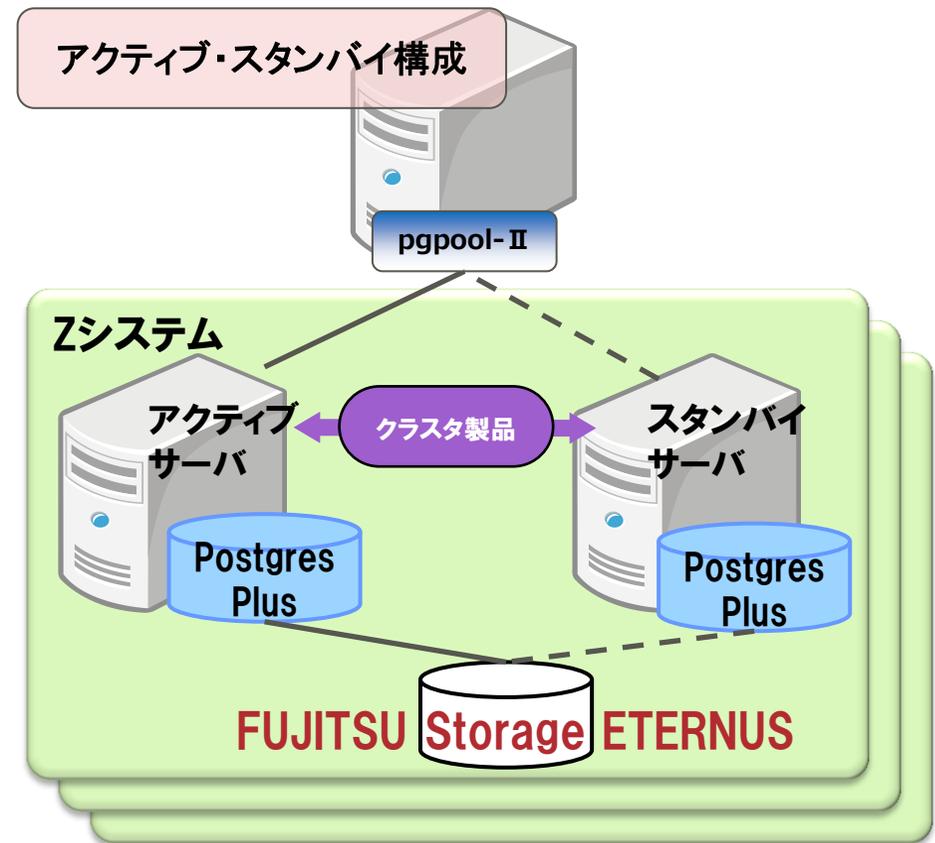
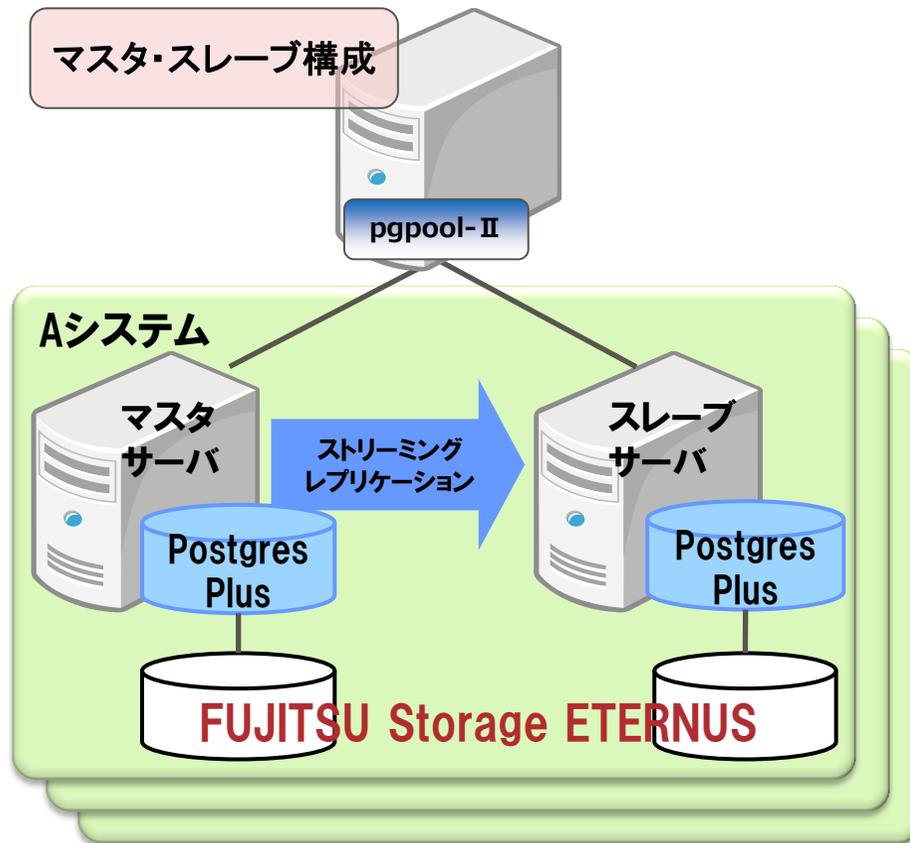
■以下の項目を検討

大項目	中項目	内容
信頼・安全	24時間365日稼働	サービスの止まらないシステム
	障害時の復旧	どのような障害が起きても瞬時に復旧
性能	トランザクション数	Postgres Plus適用判断
	チューニング	Postgres Plusの最適なチューニング
	レプリケーション	Streaming Replicationの性能
	バックアップ・リカバリ	膨大なデータ量のバックアップ・リカバリ
運用	運用手順	運用手順の簡易化
移行	バージョンアップ	サービス停止時間の短縮

安全・信頼（24時間365日稼働）

■サービスレベルに応じたシステム構成

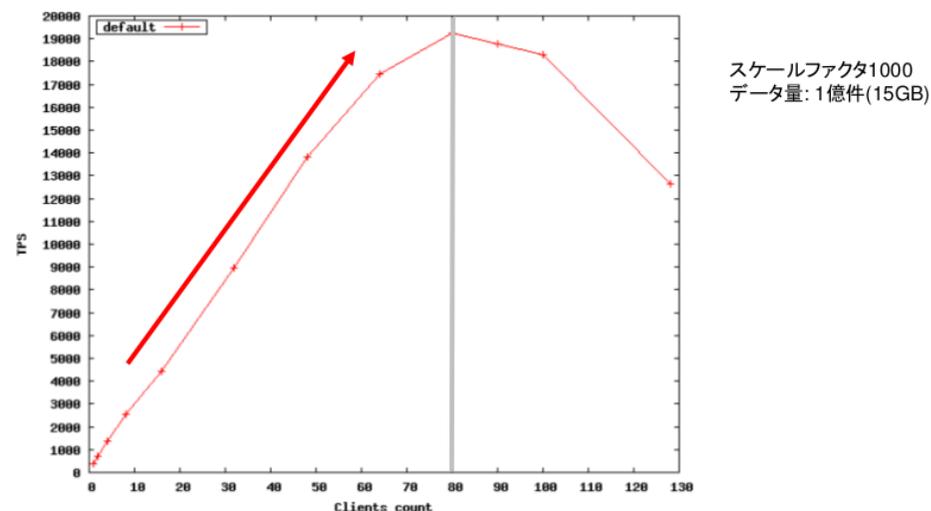
- マスタ・スレーブ構成 : 24時間365日の要求（切り替え時間：数秒）
- アクティブ・スタンバイ構成 : 24時間365日の要求（切り替え時間：数分）



性能（トランザクション数）

■高トランザクション領域でPostgres Plusが使えるのか？

- スケーラビリティの改善（PostgreSQL9.2）
 - ✓ 64コアまでの性能検証
- PostgreSQL エンタープライズ・コンソーシアムでの性能検証
 - ✓ 80コアまでの性能検証
- PostgreSQLの動作実績



<http://creativecommons.org/licenses/by/2.1/jp/>
<https://www.pgecons.org/>

性能（チューニング）

■高トランザクション領域での特殊なチューニングは必要？

- 共有バッファに使用するメモリ
 - ✓ shared_buffers
- ソートやハッシュ処理で使用するメモリ
 - ✓ work_mem
- VACUUM、CREATE INDEX等、保守操作で使用するメモリ
 - ✓ maintenance_work_mem
- チェックポイントを起動するWALセグメントの数
 - ✓ checkpoint_segments

一般的なPostgreSQLのチューニングで性能問題は起こらない。
チューニングノウハウは、すぐに検索可能

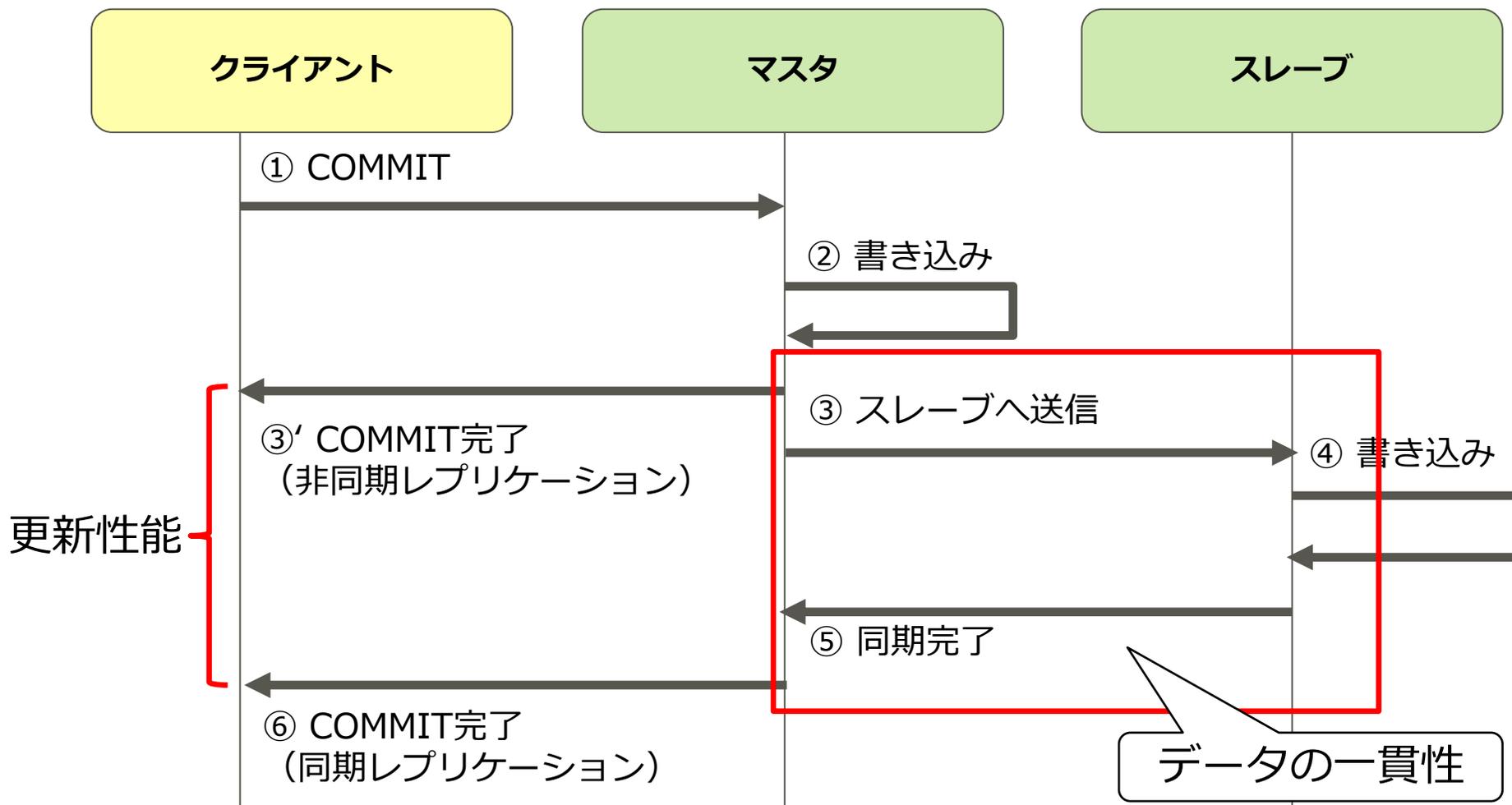
性能（レプリケーション）

■レプリケーションが遅延原因にならないか？

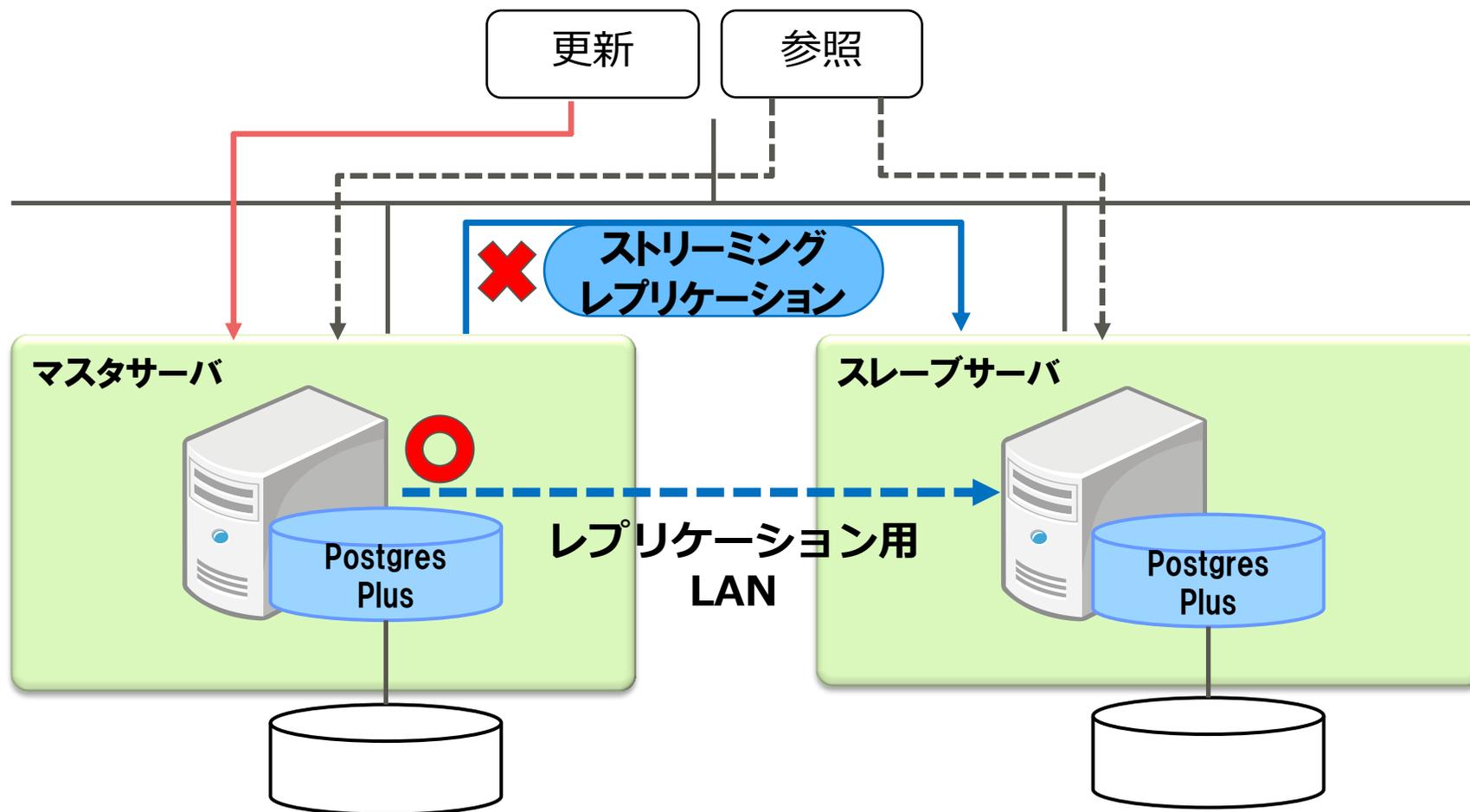
- Streaming Replication（ストリーミングレプリケーション）
 - ✓ 同期レプリケーション
 - ✓ 非同期レプリケーション

課金決済システムのため、マスタ・スレーブ間で
データの一貫性を保てる『同期レプリケーション』を
選定するが更新性能が低下する恐れがある。

■レプリケーションの仕組み



■同期レプリケーションで大丈夫か？



■レプリケーションが遅延原因にならないか？

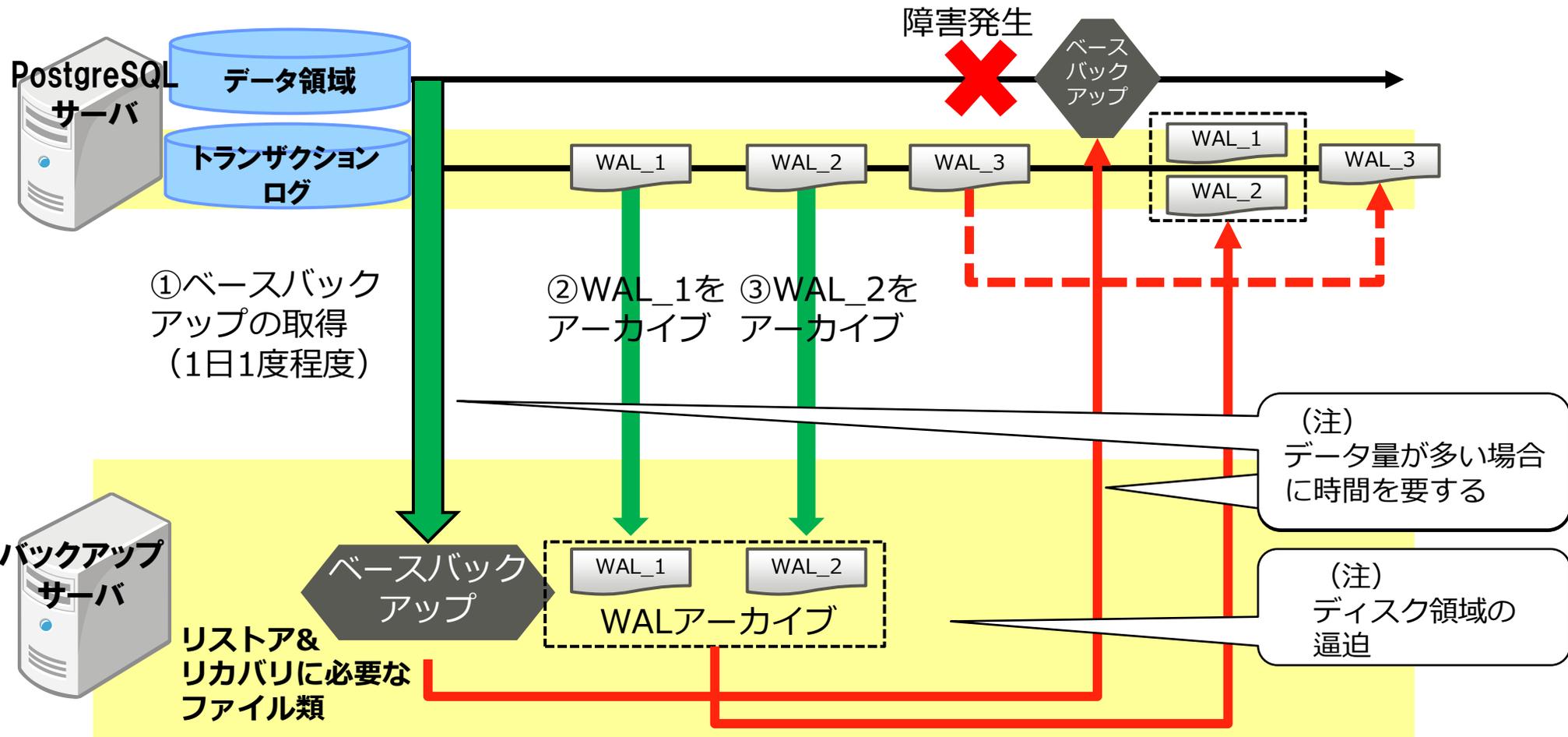
- Streaming Replication（ストリーミングレプリケーション）
 - ✓ 同期レプリケーション

レプリケーション専用LANを用意し、ネットワーク負荷を分散することで、問題なくサービス運用できる。

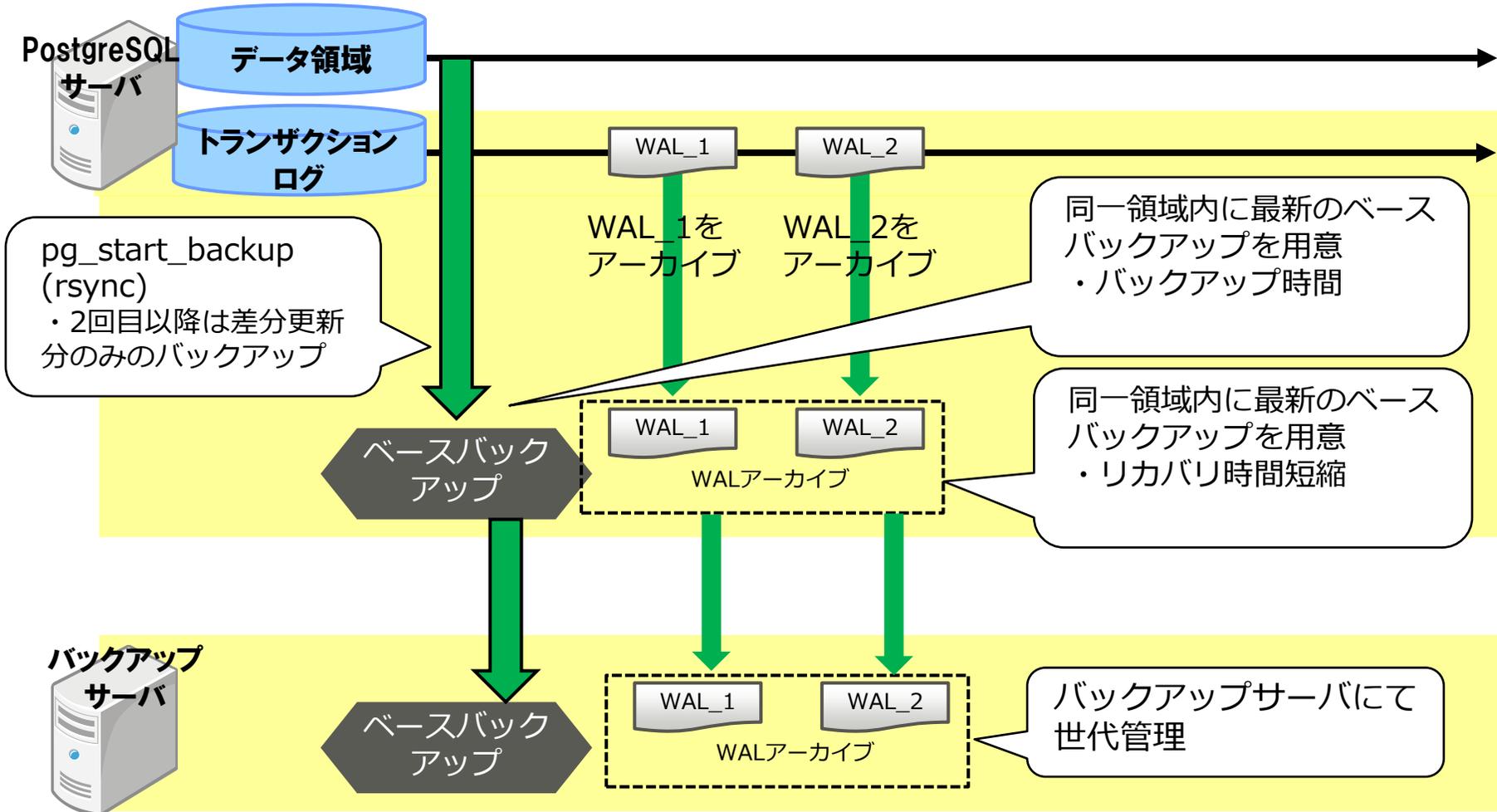
遅延（性能問題）も発生していない。

性能（バックアップ・リカバリ）

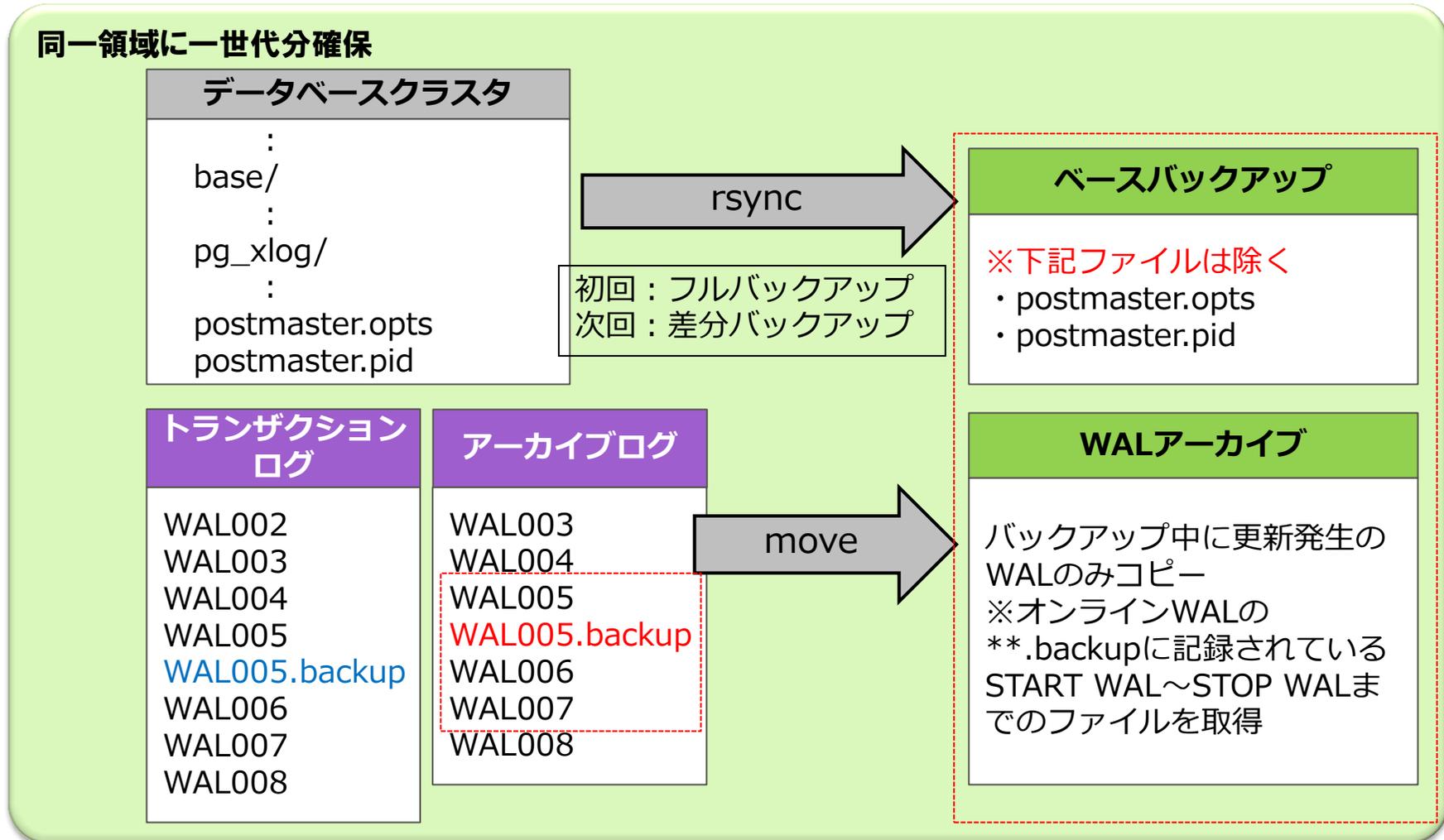
■一般的なバックアップ・リカバリの仕組み



■バックアップ・リカバリ構成 (時間短縮)



■バックアップ・リカバリ方法（時間短縮）



■アーカイブログ領域の逼迫によりPostgreSQLが停止する

- ディスク設計、運用設計が重要
 - ✓PITR（ポイント・イン・タイム・リカバリ）は必須
 - ✓更新データが多いほど、アーカイブログが出力
 - ✓アーカイブログ自動削除機能はない
 - archive_cleanup_command（PostgreSQL 9.0）

弊社では、バックアップのタイミングで削除することが一般的
急なトランザクションに対応するための対策も検討必要である。

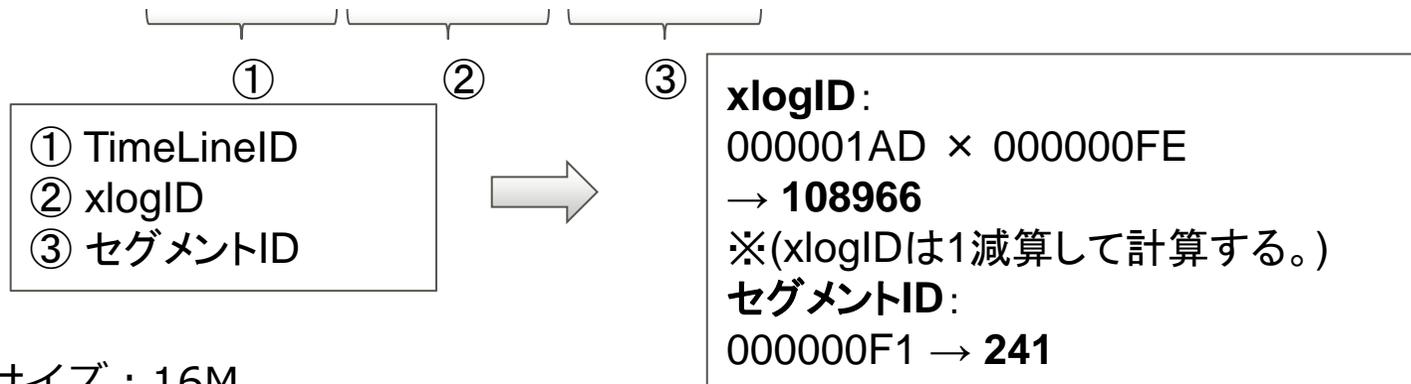
■アーカイブログのディスク設計

- アーカイブログファイル名から必要容量を計算

【アーカイブログ出力ディスク領域の設計 (今回使用した方法)】

- ・WALファイル名から、1日のWAL出力量を求める。

最新WALファイル名 : 00000001 000001AE 000000F1



WALファイルサイズ : 16M

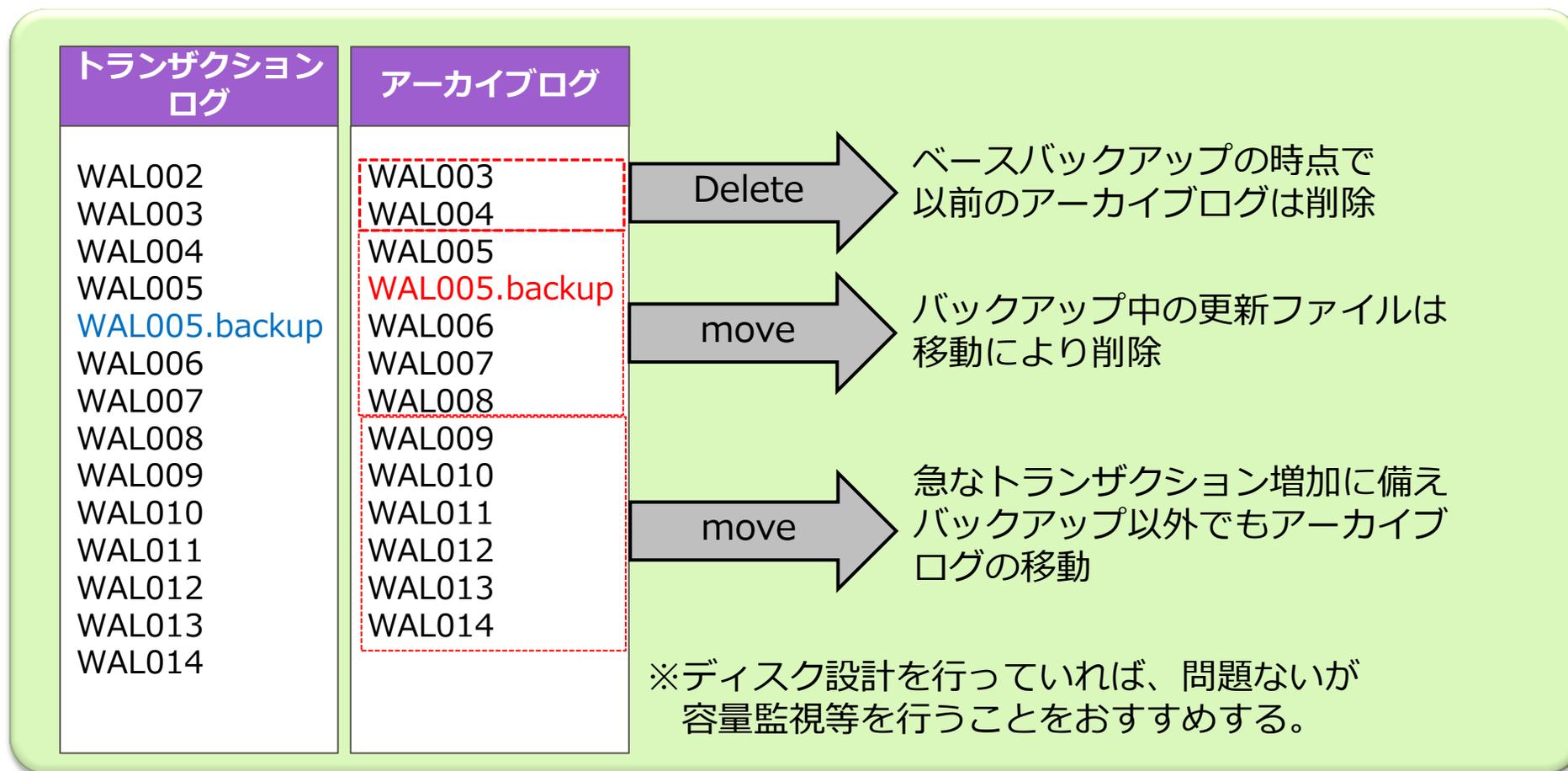
WALファイル数 : xlogID + セグメントID = 109207

サイズ合計 : 109207 × 16M = 1747GB

1日あたり : サイズ合計 ÷ PostgreSQL稼働日数 (365日) = **4.8GB**

■アーカイブログの運用設計

●アーカイブログファイルの削除タイミング



■バックアップがサービス性能に影響を与える？

- バックアップ処理で負荷が上がる
 - ✓rsync利用等を検討し、バックアップサイズの縮小化
- データ量が多いため、バックアップに時間を要する
 - ✓同一領域へ最新データを確保する
(世代管理はバックアップサーバを利用)
- アーカイブログによるディスク領域の逼迫
 - ✓余裕を持ったディスク設計
 - ✓バックアップタイミングでの削除
 - ✓アーカイブログの増加量を把握し、差分移動
 - ✓急なアクセス増加に備え、定常監視

設計・運用設計を行いサービスへの影響を抑えることが可能

運用（手順の簡易化、パターン化）

■システム構成の複雑化に伴い、運用手順も複雑化する？

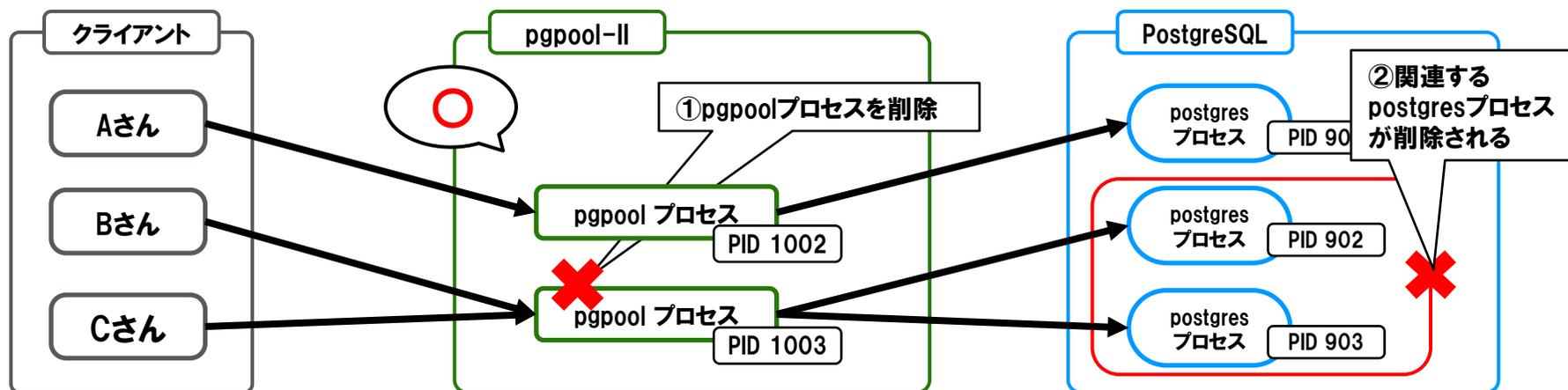
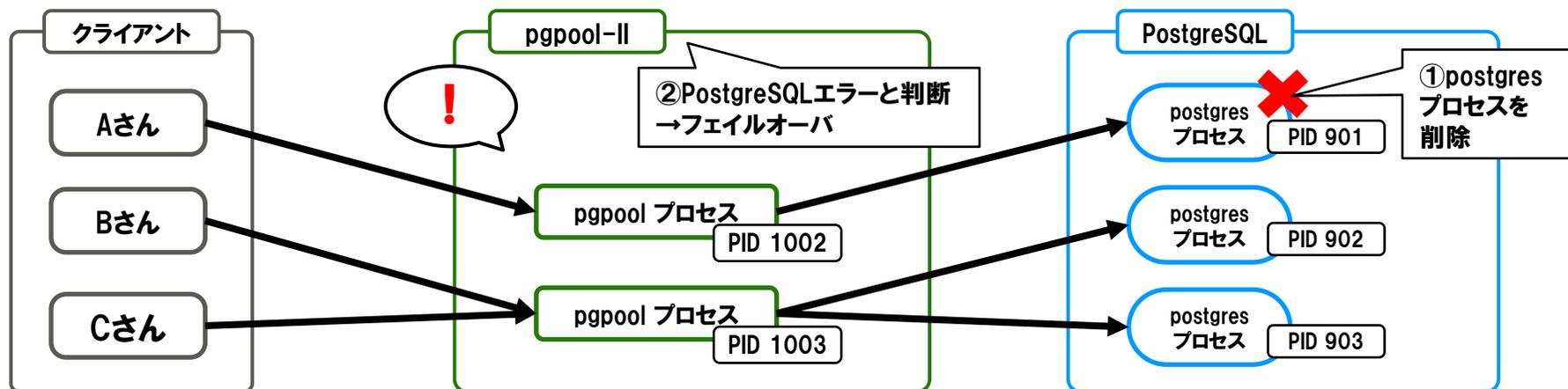
- 運用手順は複雑になる
 - ✓冗長化構成（マスタ・スレーブの管理）
 - ✓障害時の復旧手順
 - ✓運用時の操作（起動・停止順序など）
- 障害パターン数も増加する
 - ✓DBサーバ停止
 - ✓ネットワーク切断
 - ✓APサーバ停止

■DBサーバ停止時の運用

障害パターン	障害パターン概要	イベント概要	自動 / 手動	スクリプト要否
A	PostgreSQLの マスタが停止	pgpool-IIがマスタの障害を検知する	自動	否
		pgpool-IIがフェイルオーバを実行する ・フェイルオーバスクリプト	自動	要
		旧マスタをスレーブとして追加する ・オンラインリカバリスクリプト	自動	要
B	PostgreSQLの スレーブが停止	pgpool-IIがスレーブの障害を検知する	自動	否
		pgpool-IIがスレーブを切り離す ・同期非同期切り替えスクリプト	自動	要
		再度スレーブとして追加する ・オンラインリカバリスクリプト	自動	要

※バックアップ、アーカイブログ削除もスクリプト化

■pgpool-II 利用によるフェイルオーバー発生の注意



■システム構成の複雑化に伴い、運用手順も複雑化する？

- 運用手順は増える
 - ✓障害時の切り替え・復旧を自動化
 - ✓手順書をまとめる
- 運用ミスを防ぐ
 - ✓不要なフェールオーバーを発生させない
 - ✓使用製品の仕組みを理解する

運用手順をまとめ、さらに障害発生原因を把握することができれば運用手順を標準化できる。

しかし、すべての障害パターンを把握するには、運用しながらノウハウを集めるしかない。

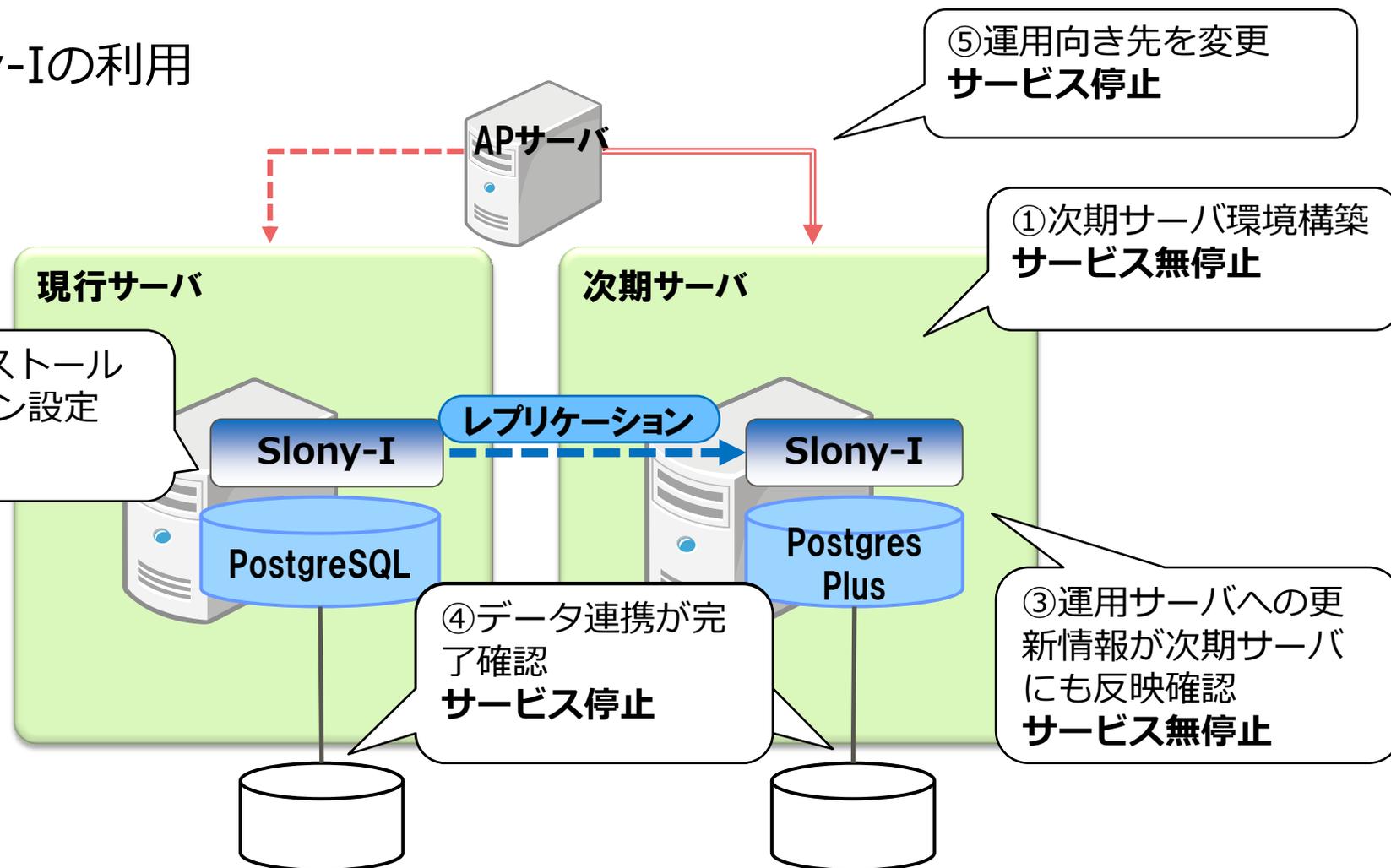
移行手法（サービス停止時間短縮）

■サービス影響を最小限に抑えてバージョンアップするには？

- 移行切り替え時間を考慮
 - ✓データの移行時間
 - ✓ミドルウェアのバージョンアップ時間
- PostgreSQL、Postgres Plusのデータ連携
 - ✓データ移行方法
 - ✓バージョン違いによる差分有無

■システム停止時間を短くするための移行

●Slony-Iの利用



■ システム停止時間を短くするための移行（注意点）

● Slony-Iのバージョンをあわせる

（－）は未検証

	PostgreSQL8.4	PostgreSQL9.1	PostgreSQL9.2	PostgreSQL9.3	PPAS9.2	PPAS9.3
Slony-I 2.0.3	○	－	×	×	×	×
Slony-I 2.1.4	△	－	○	○	○	○
Slony-I 2.2.0	△	－	○	○	○	○
SlonyRep 2.1.2	×	－	－	－	○	×
SlonyRep 2.2.0	×	－	－	－	×	○

△から○へ移行したい場合(参考)

- ① 移行元PostgreSQLが `-enable-thread-safety` オプションでインストールされているか確認
→インストールされていない場合は、別ディレクトリへインストール（例. `/usr/local/pgsql_safe`）
 - ② Slony-Iインストール時に`configure`に上記ディレクトリの`bin`を指定（例. `/usr/local/pgsql_safe/bin`）
 - ③ Slonyモジュールファイルを移行元PostgreSQLのライブラリへ移動
（例. `mv /usr/local/pgsql_safe/lib/slony1_funcs.2.2.1.so /usr/local/pgsql/lib`）
- 上記対策で、Slonyが動作することを確認しました。

- 富士通ソーシャルサイエンスラボラトリのご紹介
 - 自己紹介
 - 企業情報
 - OSS関連サービス・取り組み
- Postgres Plus適用事例のご紹介
 - Postgres Plus適用概要
 - 適用課題と対応
- 所感
- ノウハウのご紹介

高トランザクション領域へ適用

■高トランザクション領域への適用について

- ソフトウェアの問題はない
- 業務内容にあわせた運用設計が重要

■大規模システムへの適用について

- Enterprise領域への適用ノウハウが整備されつつある
- 国内での事例に乏しい

今後、国内事例が発表されていくことで、Postgres Plusの国内適用は加速していくと予想する。

弊社も、PostgreSQLが発展するように発信していく。

- 富士通ソーシャルサイエンスラボラトリのご紹介
 - 自己紹介
 - 企業情報
 - OSS関連サービス・取り組み
- Postgres Plus適用事例のご紹介
 - Postgres Plus適用概要
 - 適用課題と対応
- 所感
- ノウハウのご紹介

ノウハウご紹介

■ xDBレプリケーションのチューニングお作法

サーバリソースを十分に使用できるようにチューニングを行う

【Java】

- xDB起動パラメータに「-Xmx: * * * * MB」を指定（デフォルトから増加させる）

※ヒープ領域を確保

【設定ファイル】

- snapshotParallelLoadCount=8

コア数を指定

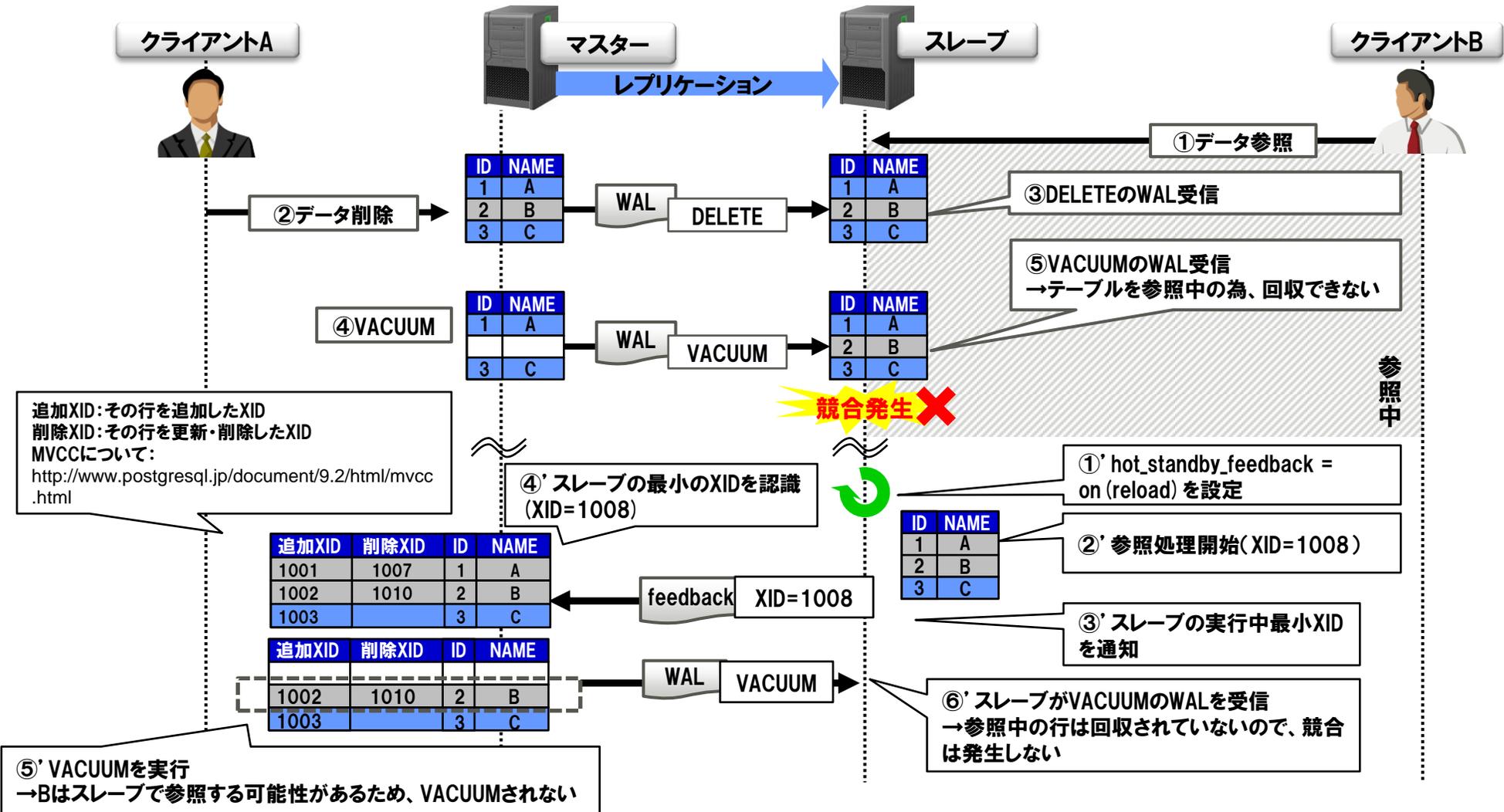
- cpBatchSize=256

サイズは調整してください

cpBatchSize * snapshotParallelLoadCountが-Xmxを超えないようにする。

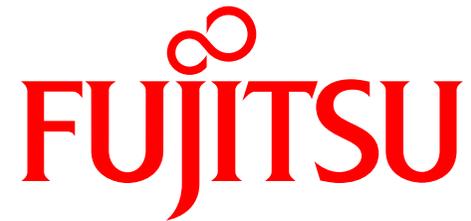
上記の設定で、レプリケーションの速度が飛躍的に伸びることを確認しました。

■ 参照負荷分散によるVacuum処理との競合



■ご清聴ありがとうございました。

※記載の会社名、商品名は、各社の商標または登録商標です。
※記載された情報は、予告なく変更することがあります。
※記載の内容は、2014年6月現在のものです。

The Fujitsu logo consists of a red infinity symbol positioned above the word "FUJITSU". The word "FUJITSU" is rendered in a red, serif typeface with a classic, slightly condensed appearance. The infinity symbol is a simple, clean line drawing.

shaping tomorrow with you