



PGECcons
PostgreSQL Enterprise Consortium

PostgreSQLをエンタープライズシステムで 利用しよう

2014年11月
PostgreSQL エンタープライズ・コンソーシアム
技術部会

<http://www.pgecons.org/>



目次

1. はじめに (本日お伝えしたいこと)
2. PGEConsについて
3. 技術部会の活動内容
4. エンタープライズ向けPostgreSQL技術紹介
5. さいごに



1. はじめに(本日お伝えしたいこと)

- ① 企業の基幹業務にPostgreSQL の導入が加速しており、ミッションクリティカル性の高い領域への普及が促進されています。さらに、それを後押しすべく課題解決に奮闘している企業連携の団体 (PGECons) があること
- ② PGEConsの活動内容と活動成果のご紹介



PostgreSQLのエンタープライズにおける活用

【PGEConsで紹介してきた活用事例】

■ 住友電気工業様

- 2005年より社内標準DBをPostgreSQLに

■ ヤマハモーターソリューション様

- 世界各地の拠点の基幹業務システムに採用

■ 株式会社キャム様

- クラウド・SaaS型 統合基幹業務システム「CAM MACS」

■ 株式会社スポットライト様

- スマートフォンを使った共通来店ポイントサービス「スマポ」

PostgreSQLのエンタープライズにおける活用

【その他の導入事例】

■ 英気象庁

- 年間サポート費用と保守費用の負担が大きいため、商用DBMSへの依存度を低下させることを狙って、PostgreSQLを活用

- ITメディア様の記事より

<http://techtarget.itmedia.co.jp/tt/news/1407/24/news01.html>

■ 仏社会保障システム

- メインフレームからの移行

- Let's postgres の記事より

http://lets.postgresql.jp/documents/case/Bull_Case_in_France/



2. PGEECons (PostgreSQL Enterprise Consortium) について

2. PGEECons について

■ PGEEConsの発足と目的

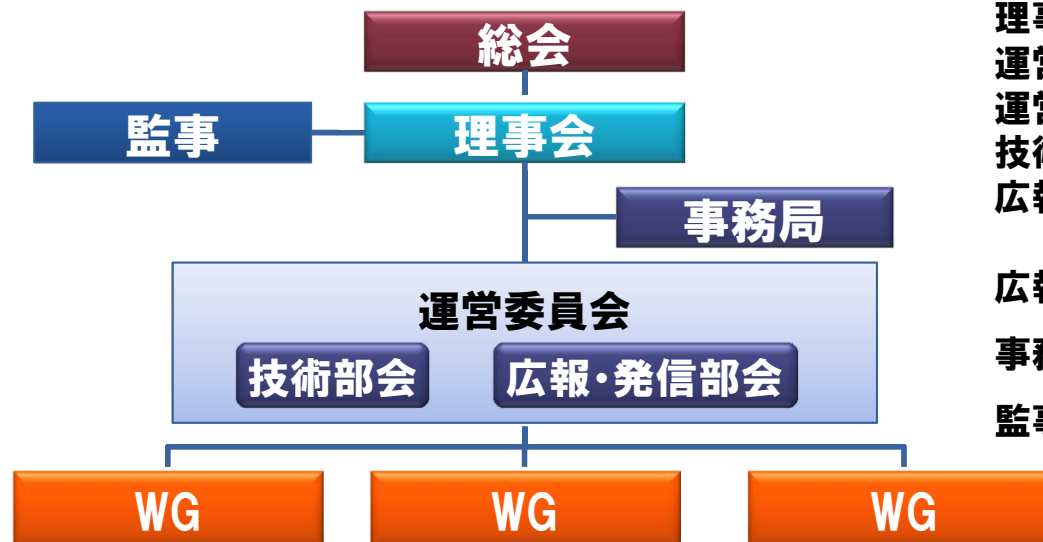
- 2012年4月11日発足
- ミッションクリティカル性の高い**エンタープライズ領域へのPostgreSQLの普及を推進**するため、各種ツールやPostgreSQL本体に関する**利用技術情報**の収集と提供および、その整備などの**活動を企業ベースで展開**する

活動項目	概要
情報発信	会員の導入実績を基に、PostgreSQLおよび周辺ツールに関する情報を集約し、情報発信サイトや 세미나等を通じて提供する
共同検証	エンタープライズ領域への適用に向けて必要となる情報を、実証を通じて充実を図る
開発コミュニティへのフィードバック	よりミッションクリティカル性の高い領域への適用に向けた技術的な課題を集約し、開発コミュニティに要望を発信する
開発プロジェクト支援	会員間での機能拡張に関する連携開発や、必要な周辺ツールの開発プロジェクト支援を進める

平成26年度 体制・会員構成

■ 会員は法人およびそれに準ずる団体で構成

平成26年度体制



理事長	: 日本電信電話株式会社
運営委員長	: 日本電気株式会社
運営副委員長	: 株式会社 日立製作所
技術部会長	: 富士通株式会社
広報・発信部会長	: 日本ヒューレットパッカ ード株式会社
広報・発信副部会長	: 株式会社アシスト
事務局長	: SRA OSS, Inc. 日本支社
監事	: 税理士法人ジャストスタッフ

種別		概要	総会議決権
正会員	理事	理事会に参加、理事長および運営委員長は理事のうちから就任する	あり
	運営委員	運営委員会に参加、部会長およびWG長は運営委員から就任する	
		ワーキンググループ(WG)に参加し、活動に貢献	
一般会員		メーリングリストやWebなどから、活動情報を取得することが可能	無し

平成26年度 参加会員一覧(1 / 2)

(2014年7月2日 現在)

■ 46社(正会員 16社、一般会員30社)が活動

参加法人名 (順不同)			
正会員 (16社)	運営委員	(株)アシスト	日本ヒューレット・パッカード(株)
		SRA OSS, Inc. 日本支社	(株)日立製作所
		NECソリューションイノベータ(株)	(株)日立ソリューションズ
		日本電気(株)	富士通(株)
		日本電信電話(株)	(株)富士通ソーシャルサイエンスラボラトリ
		TIS(株)	特定非営利活動法人 LPI-Japan
	サイオステクノロジー(株)		フューチャーアーキテクト(株)
	NTTソフトウェア(株)		ZEKKO(株)

平成26年度 参加会員一覧(2/2)

(2014年7月2日 現在)

	参加法人名（順不同）		
一般会員 (30社)	(株)アイ・ティ・プロデュース	(株)キャム	日本アイ・ビー・エム(株)
	岩通ソフトシステム(株)	キーウェアソリューションズ(株)	(株)フィックスターズ
	(株)エクサ	ジャパンシステム(株)	(株)マインド
	SFKメディカル(株)	住友電気工業(株)	ミュートック(株)
	(株)エニブラ	住友電工情報システム(株)	(株)メトロシステムズ
	(株)エム・オー・エム・テクノロジー	(株)seiwa	ヤマハモーターソリューション(株)
	エンタープライズDB(株)	(株)中電シーティーアイ	ローリーコンサルティング(株)
	(株)オンザマーク	(株)デジタル・ヒュージ・テクノロジー	(株)アイ・アイ・エム
	関電システムソリューションズ(株)	(株)ニイズ	(株)インフォメーションクリエイティブ
	クオリカ(株)	東芝ソリューション(株)	(株)ギークフィード

WGの活動風景

■ 検討会



■ 実機検証





3. 技術部会の活動内容

技術部会の活動内容と方針

■ 技術部会

- 性能WG (WG1) : 性能向上に向けた機能の活用指針を作成
 - 性能向上を狙ったハードウェア活用、パーティショニングを観点に測定
 - バージョンアップによる差を定点観測
 - 実測結果と共に、性能ボトルネックを解決するノウハウをご紹介
- 移行WG (WG2) : DB移行ガイドの作成
 - DB移行時の各プロセスのガイドの詳細化や広範囲化を中心に活動
 - PostgreSQL活用の課題である、移行費用見積のベースとなる成果物を作成
- 設計運用WG (WG3) : 設計や運用面の技術情報を整備
 - 可用性をテーマにエンタープライズ領域に求められる技術情報の整備・検証の実施
 - PostgreSQLの安定運用に必要な技術ノウハウをご紹介

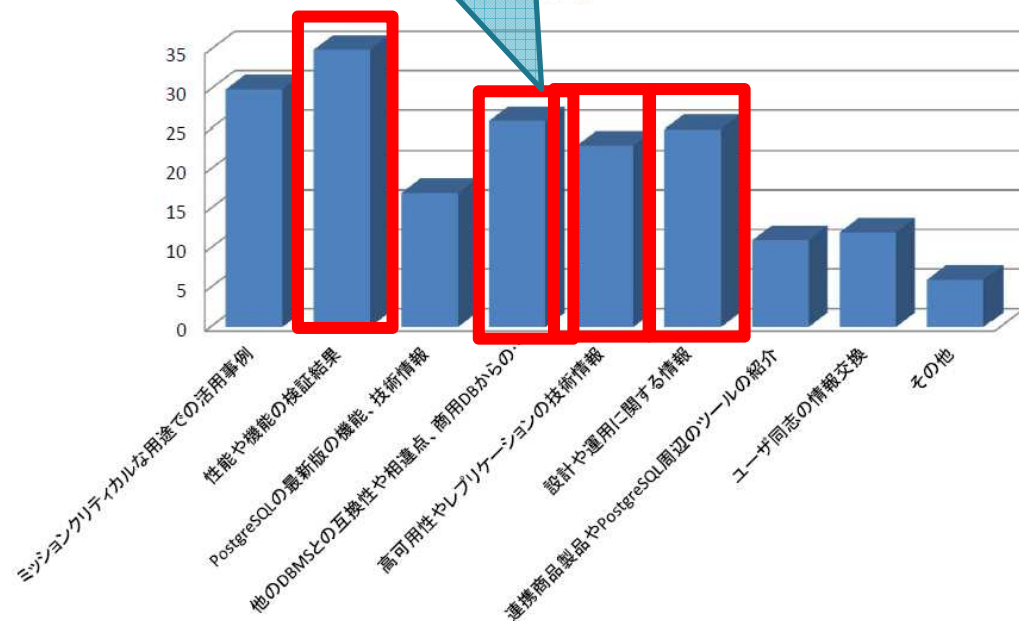
過去セミナーのアンケートよりテーマを選定

PGEConsへの

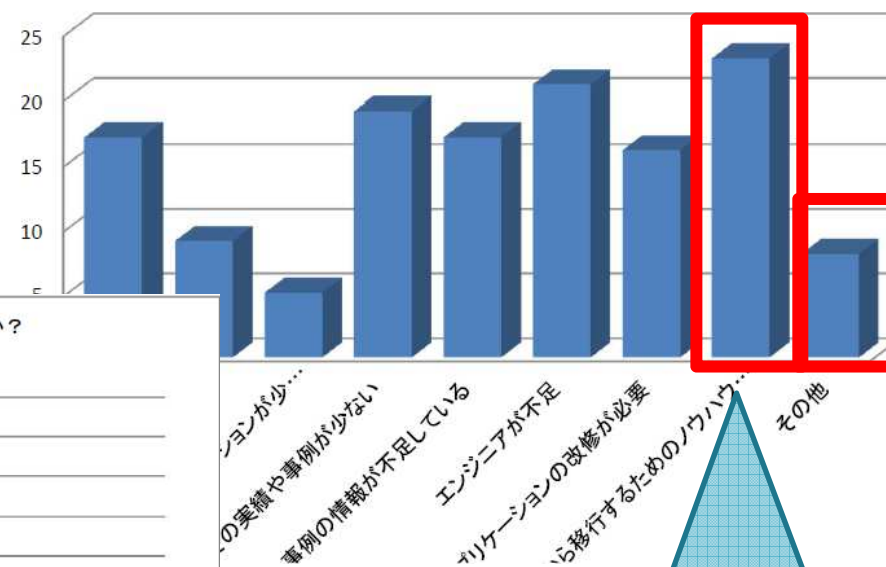
期待

「性能や移行、高可用性、
設計・運用に関する情報
発信」

Q6. PGEConsに何を期待していますか？またどんな情報を発信してほしいですか？
(複数回答可)



Q5. PostgreSQLをよりミッションクリティカル性の高いエンタープライズ領域で採用するための課題は何だとお考えですか？(複数回答可)



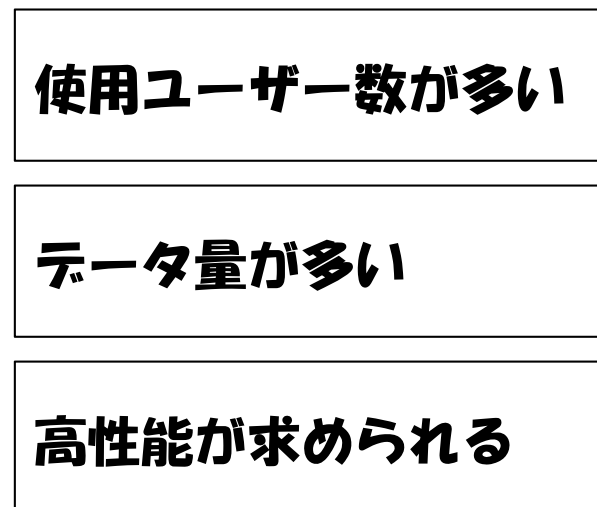
PostgreSQL採用の
課題

「個々のコメントから具体的
テーマを選定」

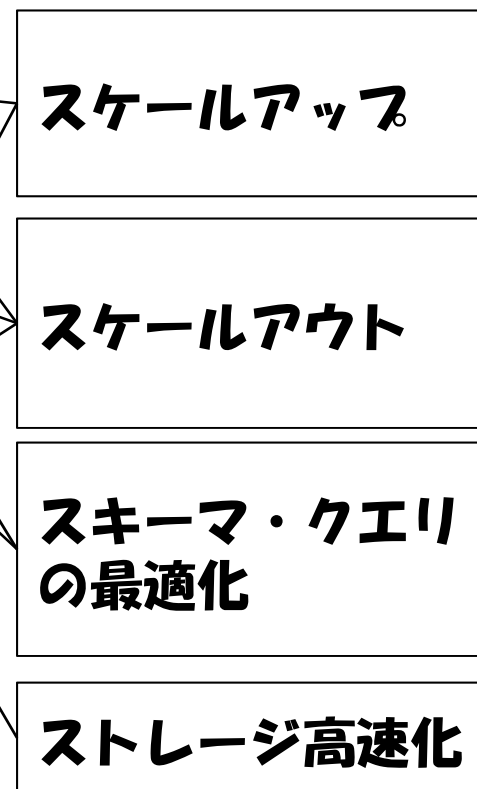
WG1(性能WG)の活動テーマ

大規模DBを見据えたPostgreSQLの性能検証

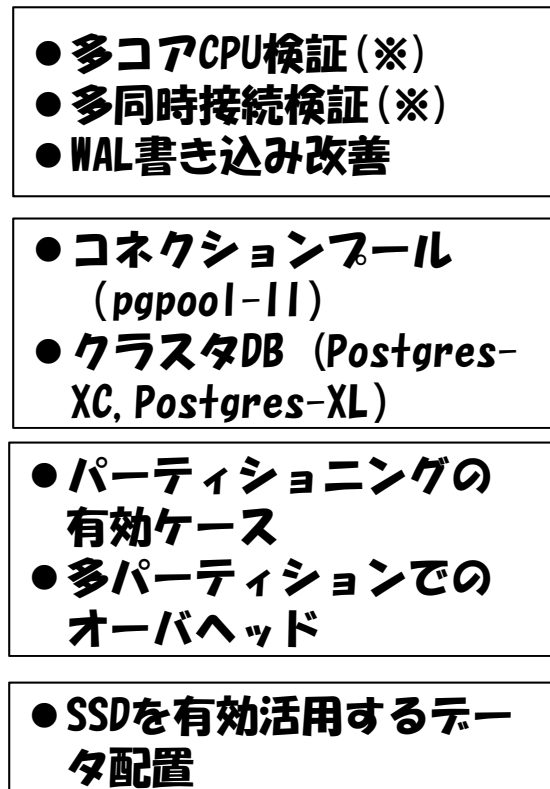
■ 大規模DBの特徴



■ 対応方法



■ 主な検証内容



※「定点観測」と呼び毎年実施

WG1 活動内容

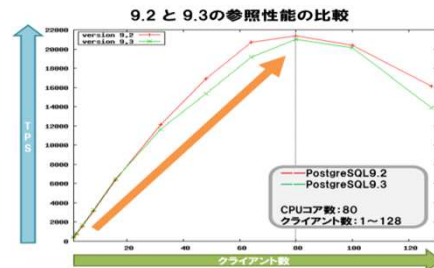
活動体制

- 検証内容検討
テーマの候補を集約して今年度取り組むテーマと担当するメンバを選定。議論はメーリングリストの他毎月1,2回の検討会で実施。
- 性能検証実施
テーマごとに担当社が中心となって検証用シナリオを検討。実機検証はメンバ各社が検証センタの環境を提供し、一定期間に集中して実施。
- 活動成果報告書の作成
報告書は、検証の位置づけや手法の説明、結果の紹介と考察を盛り込んで担当社が起稿し、検討会でのレビューを経て完成。

2013年度活動成果 ～多数のデータと詳細な解説の報告書(67p.)を作成～

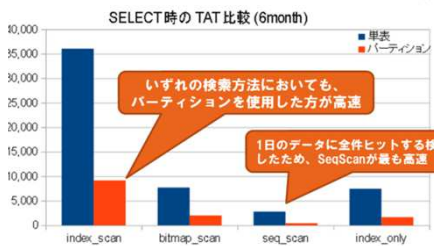
■ 定点観測

CPUコア数1～80に対して、良好なスケール性を確認。



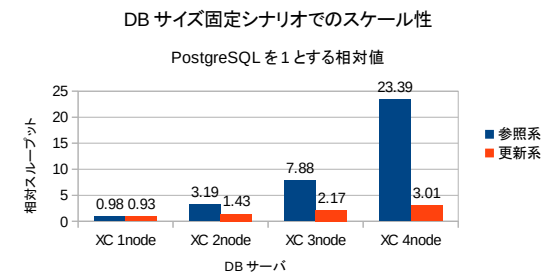
■ パーティショニング検証

約180パーティションの表の検索性能が高速であることを確認。データロードも動的関数の使用により効率化可能。



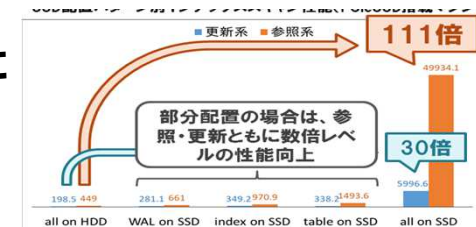
■ Postgres-XC検証

スケールアウトによる性能向上を確認。特に参照系におけるスケールが顕著。



■ SSD検証

更新系処理の高速化に非常に有効であることを確認。全データをSSDに配置すると特に顕著



WG2(移行WG)の活動テーマ

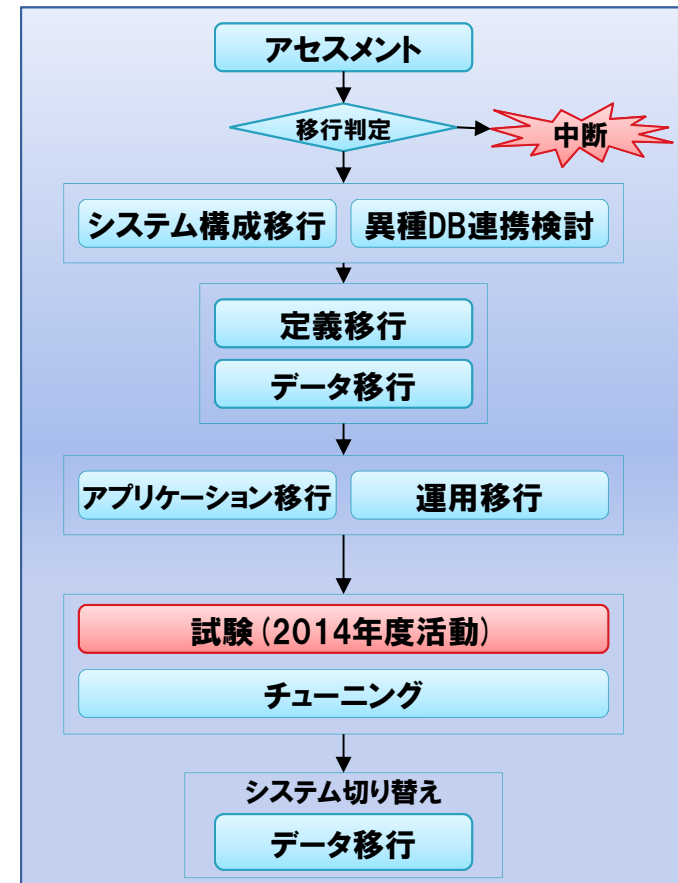
- 今年度は、システム移行後に実施する試験項目および手順の検討を行っている。
- 前提となるシステム移行は前年度までに作成した移行手法を参考に実施。

【現段階で挙げられている試験大項目】

- 移行結果確認試験
 - テーブル定義移行結果確認試験
 - SQL移行結果確認試験
 - データ移行結果確認試験
 - 性能試験
 - ロール/権限移行結果確認試験

WG2 「異種DBMSからPostgreSQLへの移行ガイド」の構成

- 移行作業の全体像を解説
 - DB移行フレームワーク編 (21ページ)
- 移行作業に含まれる作業内容、手順の調査
 - システム構成調査編 (29ページ)
 - 異種DB間連携調査編 (18ページ)
 - スキーマ移行調査編 (25ページ+別表)
 - データ移行・文字コード変換編 (49ページ)
 - ストアドプロシージャ移行調査編 (34ページ)
 - アプリケーション移行調査編 (10ページ)
 - SQL移行調査編 (18ページ+別表)
 - 組み込み関数移行調査編 (15ページ+別表)
 - チューニング編 (30ページ+別表)
 - バージョンアップ編 (39ページ+別表7)
 - **試験編<NEW!>**
- 移行作業を試行する検証
 - データ移行調査および実践編 (60ページ+別表3)
 - アプリケーション移行実践編 (25ページ+別表)



移行プロセス全体像

WG3(設計・運用WG)の活動テーマ

■ 可用性

- PostgreSQLの代表的なシステム構成(シングル・HA・レプリケーション)の特徴・適用領域の整理と検証
- 災害対策を想定したシステム構成の拡張と検証

■ バックアップ

- バックアップ手法の整理と運用例、および検証

■ 監視

- 死活・性能監視に必要な情報と分析手法の整理、対処方法の検証

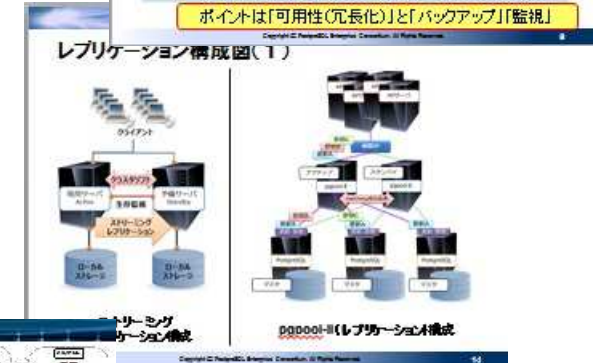
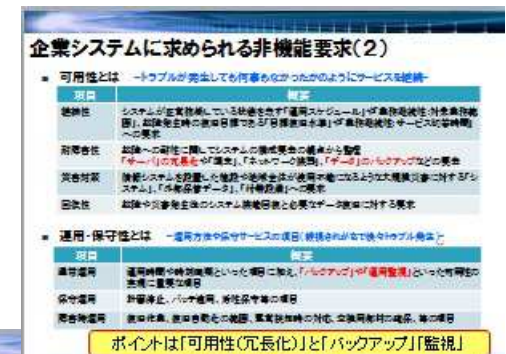
■ セキュリティ・監査

- 利用できる機能の整理と実用レベルのバランス調査と検証

PostgreSQL運用に必要な技術ノウハウ提供を目指す

WG3 活動内容

- **企業システムに求められる「非機能要求」とDBMSに求められる要求(要件)を整理**
- **DB要件を実現するPostgreSQLの代表的なシステム構成の整理と分類**
- **バックアップ／リカバリ、監視に求められる要件の整理**
- **運用技術検証**
 - ・基礎検証、高負荷下での可用性検証
 - ・バックアップ／リカバリ検証
 - ・監視ケーススタディ



商用データベースにも引けを取らないことを確認！



4. エンタープライズ向け PostgreSQL技術紹介

- 2013年度WG3活動より -

2013年度WG3の活動概説

■ 2013年度の活動テーマ

- 可用性
- バックアップ／リカバリ
- 監視

■ なぜ、これらのテーマ？

■ エンタープライズ領域でのPostgreSQL活用にあたり サービス継続性の向上、安定運用が必要

- 世の中には様々な情報はあふれているが、体系立てた調査・検証が必要と判断
- 非機能要求（可用性、性能・拡張性、運用・保守性、以降性、セキュリティ、システム環境・エコロジー）から可用性を柱に関連の強い運用・保守性（バックアップ／リカバリ、監視）をメインテーマに選定

PostgreSQLの可用性向上（概説）

- 活動を通じて、PostgreSQLで取り得る下記構成をシングル構成（1台のPostgreSQLのみ）と比較するかたちで特徴をまとめた
 - HAクラスタ構成（共有ストレージ方式）
 - HAクラスタ構成（シェアードナッシング方式）
 - ストリーミングレプリケーション
 - pgpool-II（レプリケーションモード）
 - Slony-I（トリガーベース）

データ同期性やコスト、オーバヘッド等の観点で
各構成のメリット・デメリットを整理

PostgreSQLの監視（概説）

- 死活監視と性能監視の2つの切り口で、PostgreSQLを安定稼働させるために何を、どのように監視すべきかをまとめた
 - 死活監視：可用性
 - PostgreSQL、サーバの異常発生を検知し、サービス停止時間/切り替え時間を短くする
 - 性能監視：運用保守性
 - 性能問題発生を検知し、サービスへの影響を未然に防ぐ

PostgreSQLの機能だけでなく、
OSの情報収集についても整理

PostgreSQLのバックアップ (1 / 8)

- PostgreSQLのバックアップにはどのようなものがあるの？
- コールドバックアップ、論理バックアップ、オンライン物理バックアップがある
 - それぞれの特徴を整理し、適材適所利用すべき

バックアップ方法	サービス継続	手順	リカバリ・ポイント
コールドバックアップ	× オフライン	○ 単純	△ 取得時点まで
論理バックアップ	○ オンラインで実行可	○ ツールあり	△ 取得時点まで
オンライン物理バックアップ	○ オンラインで実行可	△ ツールはあるが煩雑	○ 任意の時点

PostgreSQLのバックアップ (2/8)

■ コールドバックアップ

- もっとも単純なバックアップ方式。
- PostgreSQLを停止し、データベースクラスタを退避したものをバックアップとする
- リストアもバックアップをデータベースクラスタとして展開して起動するだけとシンプル

□ 注意点

- 戻せる時点は、バックアップ取得時点まで
- テーブル空間 (データベースクラスタ以外の領域にデータベースオブジェクトを配置する) を利用している場合は、テーブル空間に配置されたファイルのバックアップも忘れずに

PostgreSQLのバックアップ (3 / 8)

■ 論理バックアップ (pg_dump/pg_dumpall)

□ pg_dump

- オンラインで論理バックアップを取得する
- ターゲットはデータベース (オプションによりデータのみやスキーマのみも可)
- 出力形式をオプションで選択 (プレーンテキスト、カスタム、tar、ディレクトリ) できる

□ pg_dumpall

- オンラインで論理バックアップを取得する
- ターゲットはデータベースクラスタ (オプションによりデータのみやスキーマのみも可)
- 出力形式はPlainText一択 (選択できない)

PostgreSQLのバックアップ (4/8)

■ 論理バックアップのリストア (pg_restore/psql)

- pg_dump/pg_dumpallで取得したバックアップをリストアする際には、オプション (出力形式) に応じてpg_restoreとpsqlを使い分ける
- pg_restore
 - pg_dumpのカスタム形式、tar形式、ディレクトリ形式で取得したバックアップをリストアする際に利用
 - バックアップファイルの中身を確認する際にも利用できる
- psql
 - PostgreSQLの標準クライアントプログラム
 - pg_dumpのプレーンテキスト形式、pg_dumpallで取得したバックアップをリストアする際に利用
 - -fオプションでバックアップファイルをバッチモードで実行

PostgreSQLのバックアップ (5 / 8)

■ 論理バックアップのまとめ

□ オンライン論理バックアップ取得コマンドのまとめ

バックアップ			リストア
コマンド	ターゲット	オプション	コマンド
pg_dump	データベース	-Fp (PlainText)	psql
		-Fc (Custom)	pg_restore
		-Ft (Tar)	
		-Fd (Directory)	
pg_dumpall	データベース クラスター	なし (PlainText)	psql

PostgreSQLのバックアップ (6/8)

■ オンライン物理バックアップとPITR

- オンライン中に取得したデータベースクラスタ (後述) をベースバックアップとし、必要なWAL (トランザクションログ) を適用することで任意の時点へのリカバリ (Point-In-Time-Recovery) を実現する
- ベースバックアップ取得のために必要な設定
 - wal_levelの調整
 - デフォルトではクラッシュリカバリに必要最低限な情報しかWALに出力されないため、“archive”もしくは“hot_standby”に
 - archive_mode/archive_command
 - WALは循環して再利用されてしまうため、PITRで利用したい場合は適宜別の領域へ退避させる

PostgreSQLのバックアップ (7 / 8)

■ オンライン物理バックアップとPITR

□ ベースバックアップの取得は以下の流れで

- pg_start_backup関数の実行
- データベースクラスタの退避 (tarコマンド等)
- pg_stop_backup関数の実行

(PostgreSQL9.2以降からpg_basebackupコマンドを利用することもOK。ストリーミングレプリケーションのスタンバイでの実行もできる)

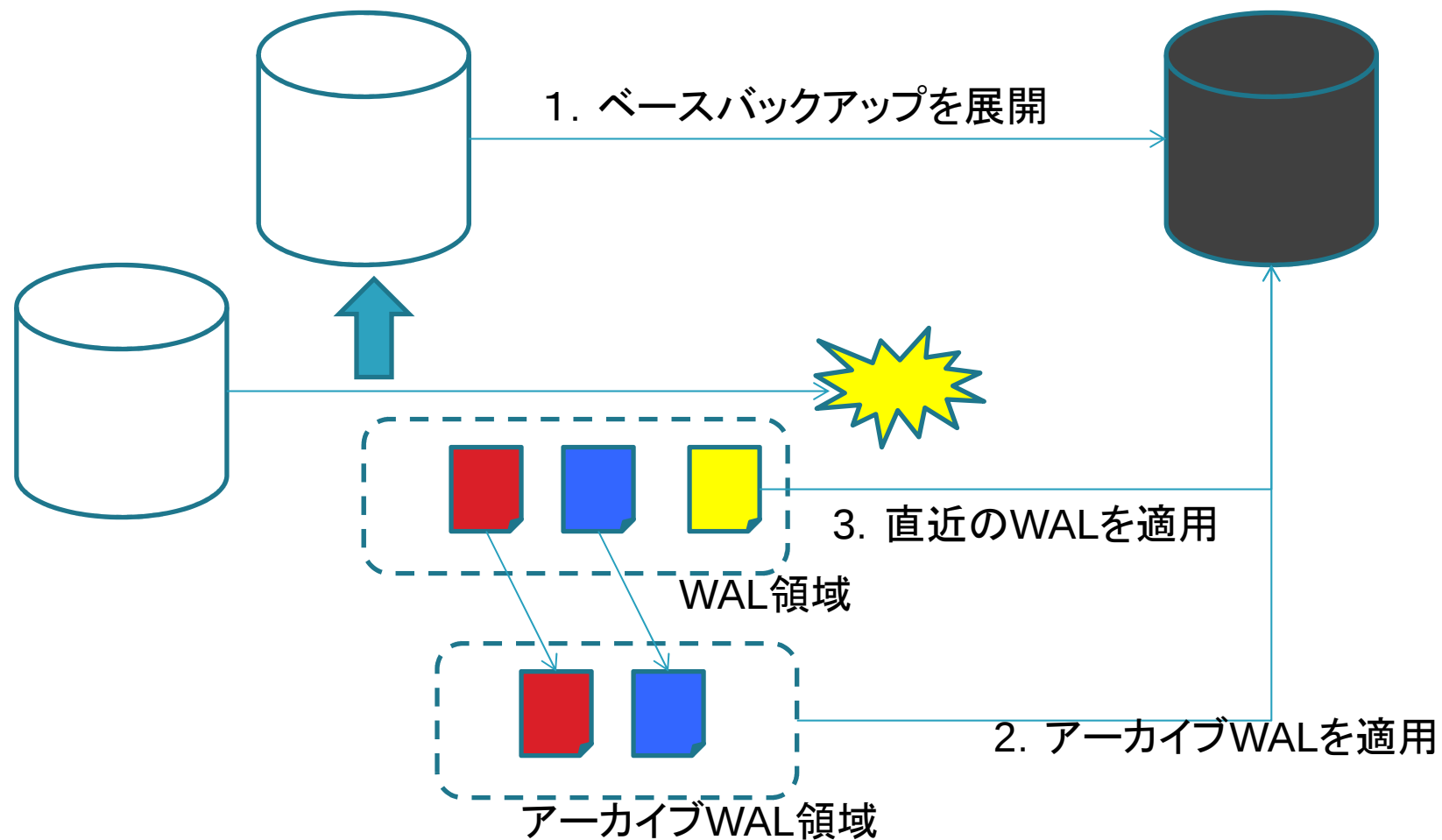
□ PITRは必要なファイルを展開してPostgreSQLを起動

- ベースバックアップ
- アーカイブWALファイル
- (直近までリカバリするなら) 直近のWALファイル
- recovery.conf

- PITRに関する設定を行うファイル
- リカバリしたい時刻 (recovery_target_time) などの指定が可能

PostgreSQLのバックアップ (8/8)

■ オンライン物理バックアップとPITRの流れ



バックアップ検証:A

- マニュアル的には理解できたけど、、、
 - 実際に動いてる環境で問題なく論理バックアップを取得できる？
 - 手順はどんな感じになるのだろうか？
 - 留意点はなにかある？
-
- といった疑問を払しょくすべく、実際にやってみた。

シナリオ (1/2)

- システム構成

- シングル構成

- システム概要

- 参照系システム
 - 更新は土日のバッチ処理のみ

- バックアップ

- 毎週更新処理完了後pg_dumpを利用して全体の論理バックアップを取得

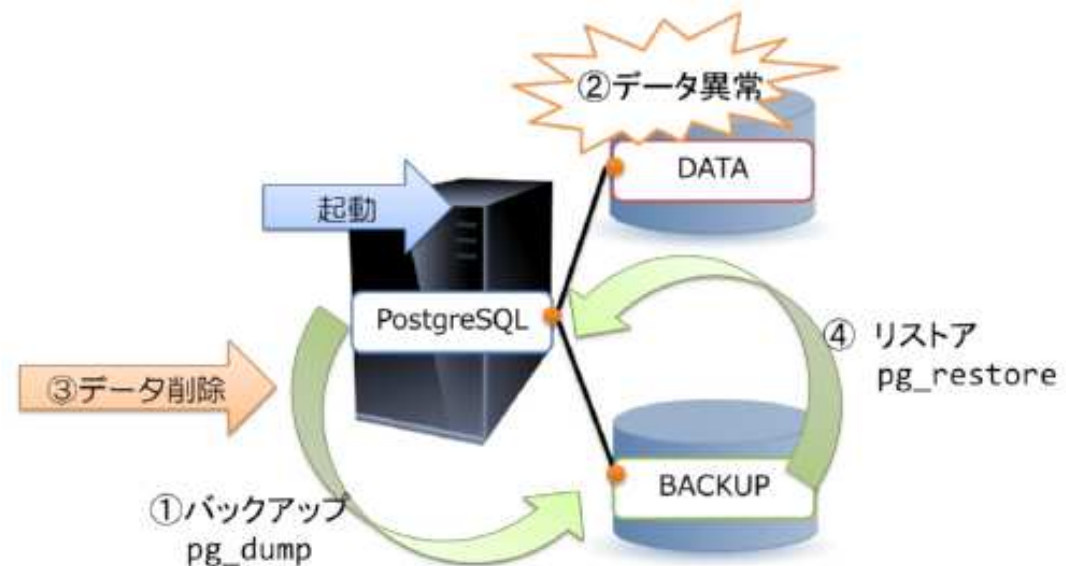
シナリオ (2/2)

■ 検証のイメージ

- ある週の火曜日から、時々エラーが発生するようになった
- データが壊れた可能性が高いが壊れたデータの特定ができてない
- カスタム形式でバックアップを取得している
- 正しかった時点のデータをリストアしたい

検証手順概要

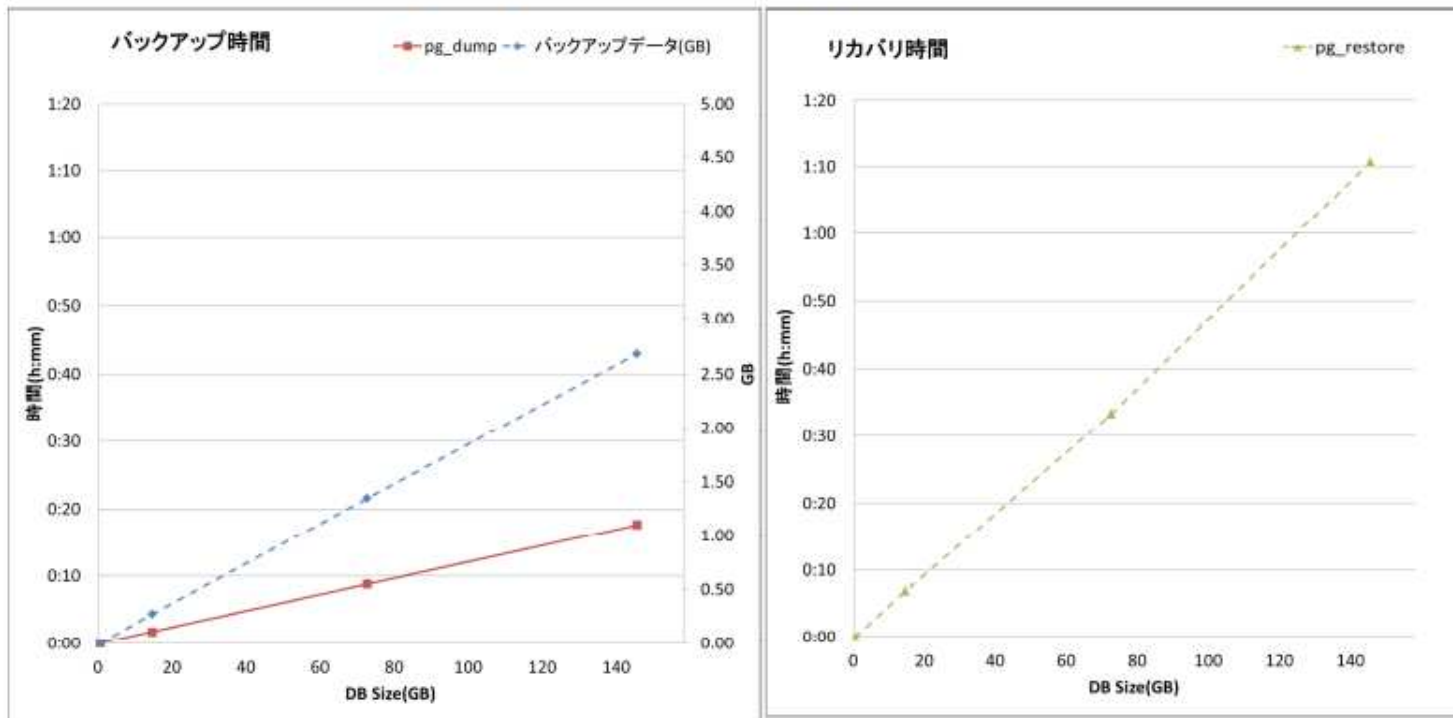
- pgbenchで初期データ投入
- バックアップ取得
 - pg_dump
- データに異常が発生したと仮定し、データ削除
 - dropdbでデータをデータベースごと削除
 - createdbで 再生成
- リストア
 - pg_restore



バックアップ検証:A補足

バックアップとリストア時間の目安について

- データサイズとバックアップ／リストア所要時間の例
 - データ量に比例して線形で増加するため見積もりは簡単
 - バックアップ時間よりもリストア時間の方が長い



バックアップ検証:B

- マニュアル的には理解できたけど、、、
 - 実際に動いてる環境で問題なくオンライン物理バックアップを取得できる？
 - 指定した時刻に戻すことはできる？
 - 手順はどんな感じになるのだろうか？
 - 留意点はなにかある？
-
- といった疑問を払しょくすべく、実際にやってみた。

シナリオ (1/2)

■ システム構成

- シングル構成

■ システム概要

- 更新系システム
- 毎月新しいテーブルを生成しその月の売上を記録
- 1年前のデータをtruncateで削除

■ バックアップ

- PITRによる物理バックアップ
- アーカイブモードはONで運用中

シナリオ (2/2)

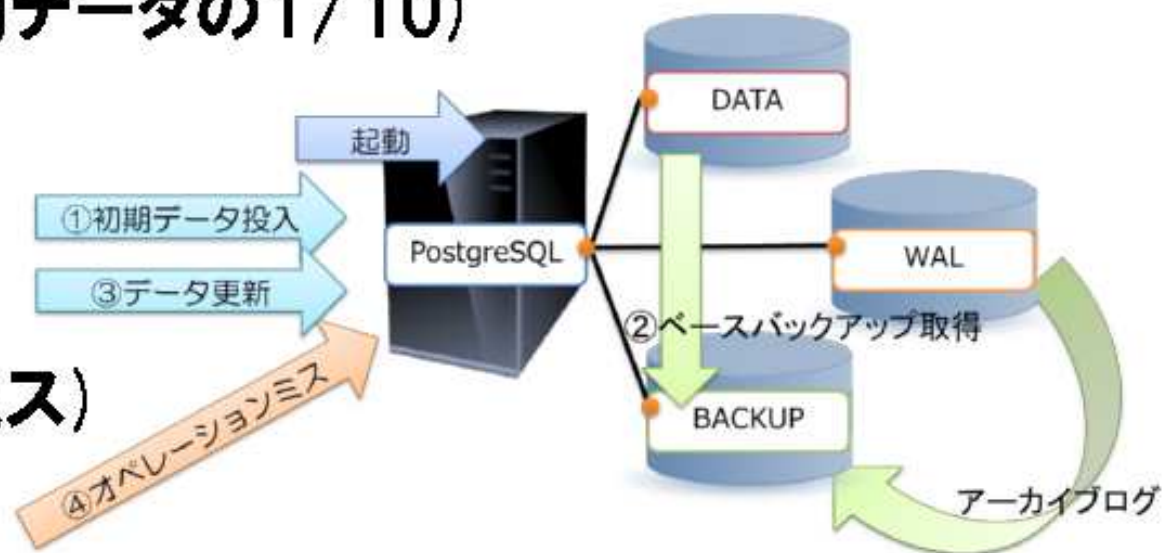
■ 検証のイメージ

- オペレーションミスで今月の売上データをtruncateしてしまった
- ベースバックアップ+アーカイブログの取得により運用中
- truncateした時刻が分かっている、なるべく最新のデータに戻したい

検証手順概要 (1/2)

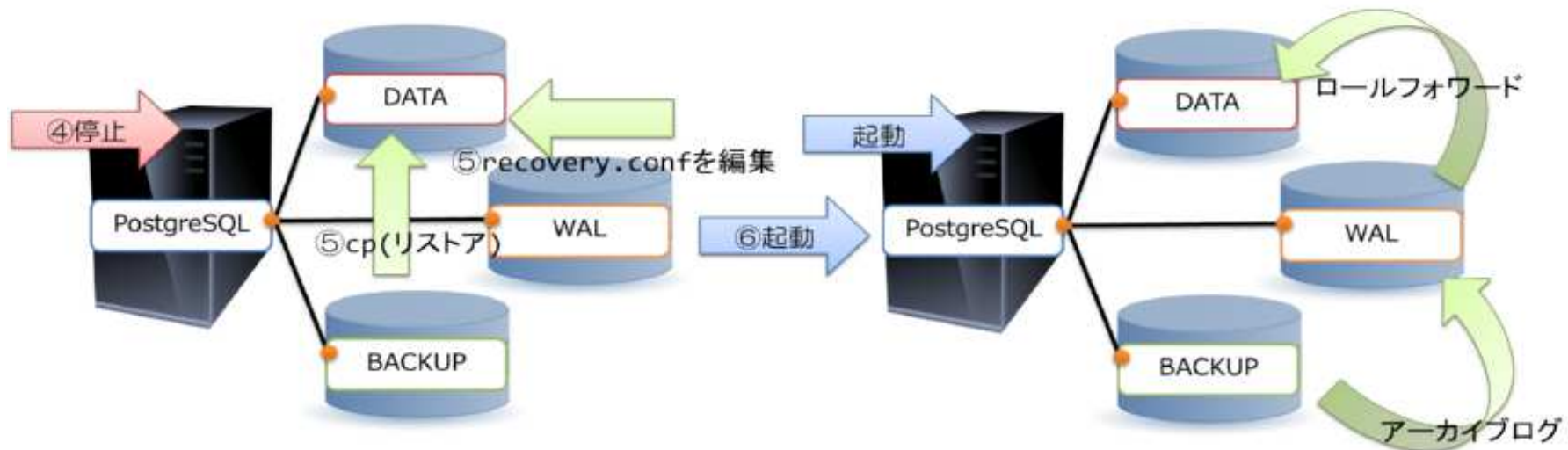
- 初期データ投入
- ベースバックアップ取得
 - `pg_start_backup()`
 - `tar`でデータディレクトリを取得
 - `pg_stop_backup()`
- データ更新 (初期データの1/10)

- `table`をtruncate (オペレーションミス)



検証手順概要 (2/2)

- サーバ停止
- ベースバックアップからリストア
- recovery.confを作成してリカバリ時刻を指定
- サーバスタート

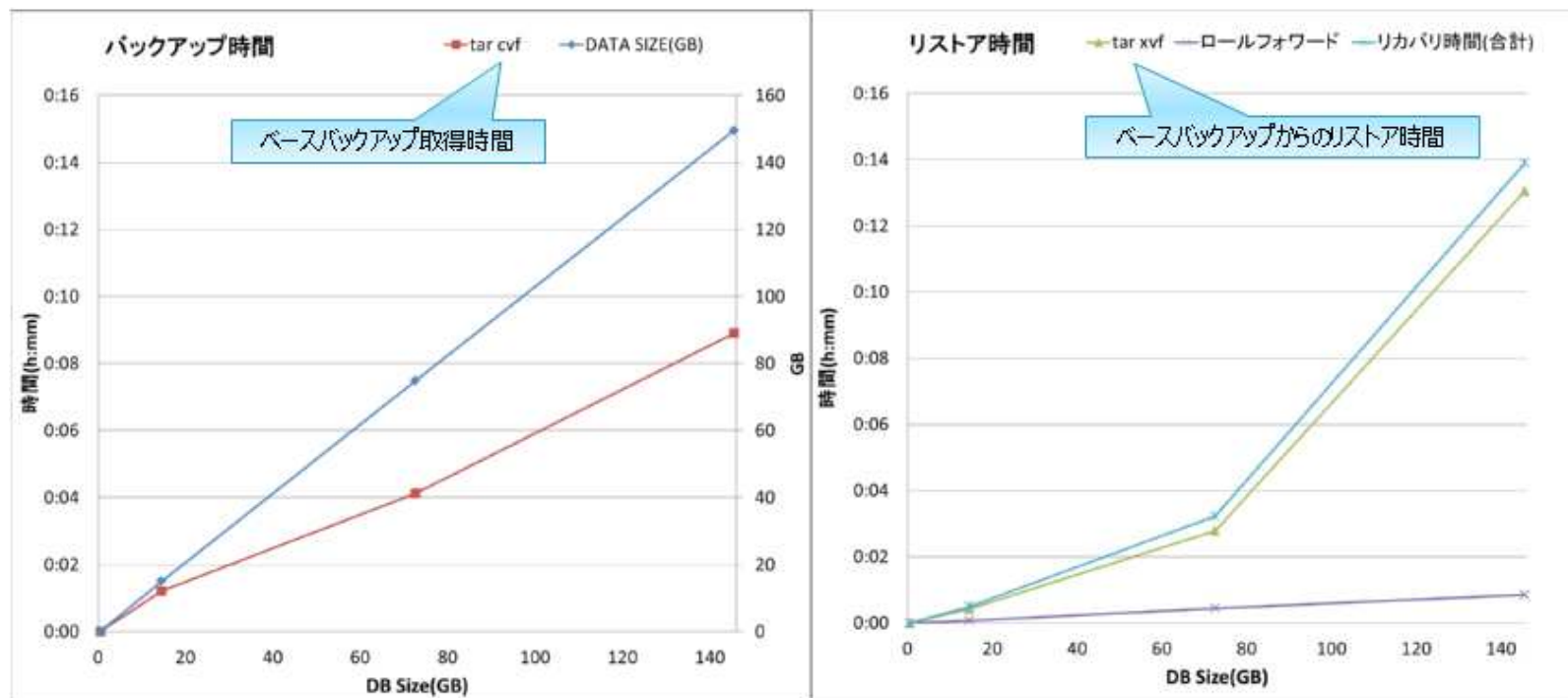


バックアップ検証:B補足

バックアップとリカバリ時間の目安について

■ データサイズとバックアップ・リカバリ所要時間の例

- ロールフォワード時間は更新データ量に比例するため定期的に新しいベースバックアップを取得する必要がある
 - ロールフォワードするログの量はDBサイズの1/10で実施





運用シナリオ検証

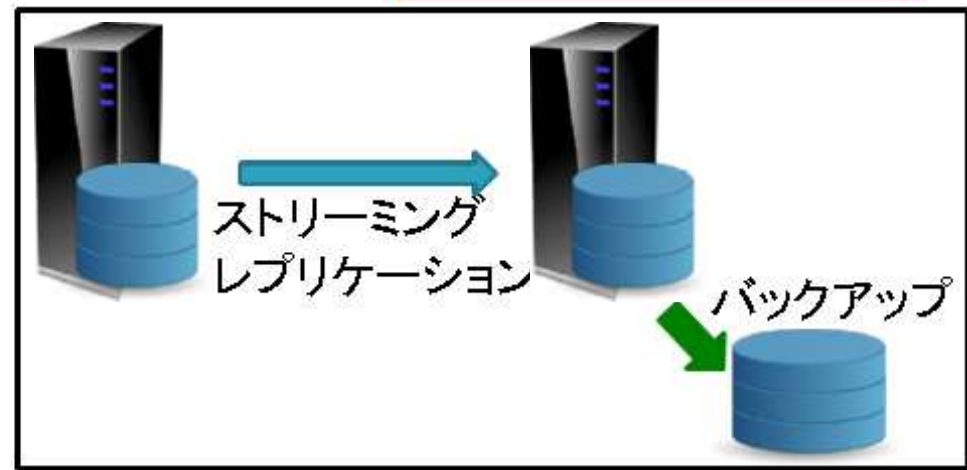
- シングル構成でバックアップ/リカバリがうまくできることは理解できたけど、、、
- 実際の運用環境でもうまくいくのか？
- ということで、より実運用の環境に近い(であろう)構成でやってみた。

運用シナリオ検証

(1) 想定するシステム概要

- 非同期ストリーミングレプリケーション構成。
- 参照更新ともにあり。更新頻度は頻繁で、夜間もあり
- 通常時業務で使用するのはマスタのみ
- 毎日スレーブ側DBからバックアップ取得
- バックアップはオンラインで取得
- ユーザは100人程度
- 蓄積データは大量
- 監視あり
- 障害復旧は手動

同期レプリケーションにも対応できるように



運用シナリオ検証

(1) シナリオ概要

- ○○日△△時××分、オペレーションミスにより、データを誤って削除。
- スレーブサーバの状況を確認したところ、スレーブサーバでもすでに削除。
- スレーブサーバで毎日取得しているバックアップからマスタを○○日△△時××分の直前までリカバリ。
- SR環境再構築

運用シナリオ検証

■ 流れとポイント

- ベースバックアップをスタンバイからマスタへ
- マスタで時刻指定してPITR
- マスタからベースバックアップを取得
- スタンバイをマスタにつないでSR再開

■ ざっくりの流れは上記の通りで問題なくSR再開できる ただし、「マスタからベースバックアップを取得」しないといけない点は要注意！

- マスタで時刻指定してPITRしてるので、旧ベースバックアップを使えない
 - 抜け穴はあるかも…

まとめ

- 2013年度WG3の活動を通じて、サービス継続性、安定運用を意識したPostgreSQLの利用方法を整理・検証できた
 - 性能面については、利用する環境やシステムのワークロードにより変化するため、検証環境等を用意して十分なリハーサルを行うことをお勧めします！
 - バックアップからのリカバリは一部内容や手順が煩雑であるため、有事の際にミスを起こさないようPGEConsが提供する資材もご活用ください！

2014年度WG3活動状況

■ 可用性

- PostgreSQLの代表的なシステム構成(シングル・HA・レプリケーション)の特徴・適用領域の整理と検証
- 災害対策を想定したシステム構成の拡張と検証

■ バックアップ

- バックアップ手法の整理と運用例、および検証

■ 監視

- 死活・性能監視に必要な情報と分析手法の整理、対処方法の検証

■ セキュリティ・監査

- 利用できる機能の整理と実用レベルのバランス調査と検証

昨年度の残課題である
「DR」「監査」をメインテーマに活動中！

2014年度WG3活動状況

■ DR

- とりうるDR構成を検討
- バックアップ/リストアとは違った観点で評価
- オンプレ環境でのDRだけでなく、クラウド環境を活用したDRにも着目

■ セキュリティ・監査

- 商用データベースとの比較をPCIDSSをベースに検討
- 具体的な要件への対応手順を検証

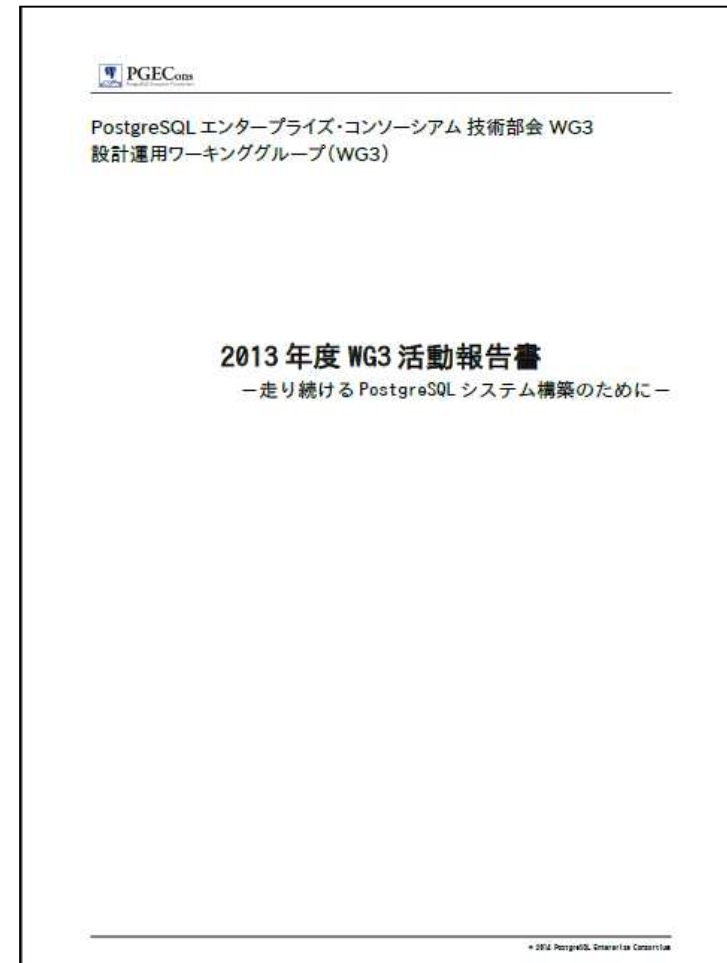
2014年度の成果物にも
ご期待ください！



5. さいごに

成果物の公開

- PGECconsサイトから、無料でダウンロードいただけます
- <http://www.pgecons.org/>



会員募集

- **正会員・一般会員を広く募集いたします**
 - WGと一緒に活動を行っていただける団体様
⇒ 正会員
 - PostgreSQLのエンタープライズ領域に興味を持っている団体様 ⇒ 一般会員

お問い合わせ先:

PostgreSQLエンタープライズ・コンソーシアム事務局

メール : jimukyoku@pgecons.org

Web : <http://www.pgecons.org/>



ご清聴ありがとうございました！

**POSTGRESQL ENTERPRISE
CONSORTIUM**



PGECons
PostgreSQL Enterprise Consortium