OSC2023 Online/Fall

PostgreSQL
http://www.postgresql.jp

超初級PostgreSQL入門

自己紹介

喜田紘介 @kkkida_twtr

日本PostgreSQLユーザ会 理事長
エクスチュア株式会社 Data Specialist

DB技術が進歩を続けるなかで、RDBの価値？需要あるの？という話も聞いたり。
そんな中でも、幸いなことにPostgreSQL人気の高まりは日々感じています。

PostgreSQLが好きな人も、DBが好きな人も、これから勉強する人にとっても、親しみやすいコミュニティであり、その環境やつながりを思い思いに使っていただいて、皆さまにとって何かしらを得られる場を作っていきたいと思っています。

本日のアジェンダ

データベースとは？なぜ必要？

PostgreSQL入門

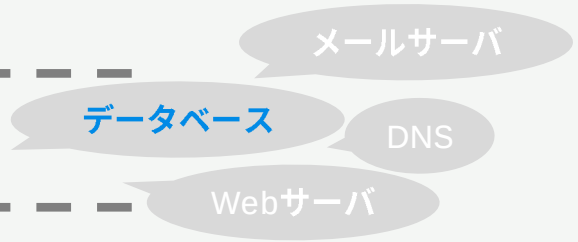
SQL開発とデータベース管理

データベースとは？なぜ必要？

データベースとは？データの一元管理とは？

データを一元管理するためのミドルウェア

アプリケーション	やりたいことを実現するための、目的に特化したプログラム
ミドルウェア	**アプリケーション動作に必要な各固有の機能を提供する、特定の役割を持ったプログラムの総称**
OS	上位レイヤにハードウェアを制御するための命令を提供する汎用的なソフトウェア
ハードウェア	CPU、メモリ、ディスクなどの物理機器



データの一元管理とは

データは現代社会を維持する根幹

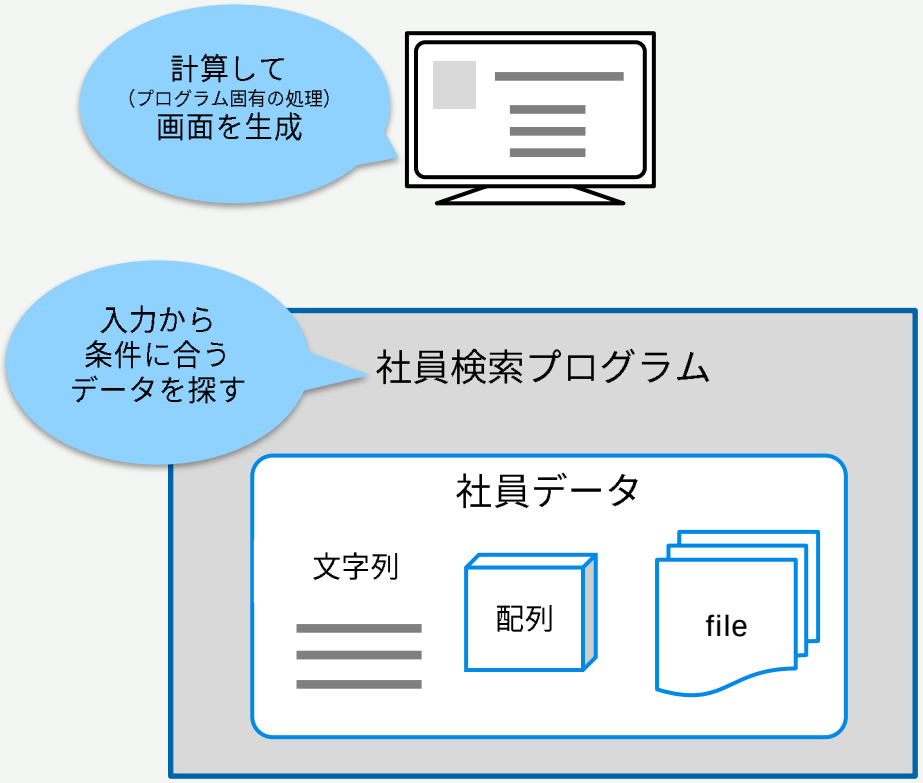
- ✓ 間違ってはいけない
- ✓ 失われてはいけない
- ✓ 適正な手段を経て参照・修正が可能

アプリケーションを都度開発するのではなく、
データの管理に特化したレイヤを設ける

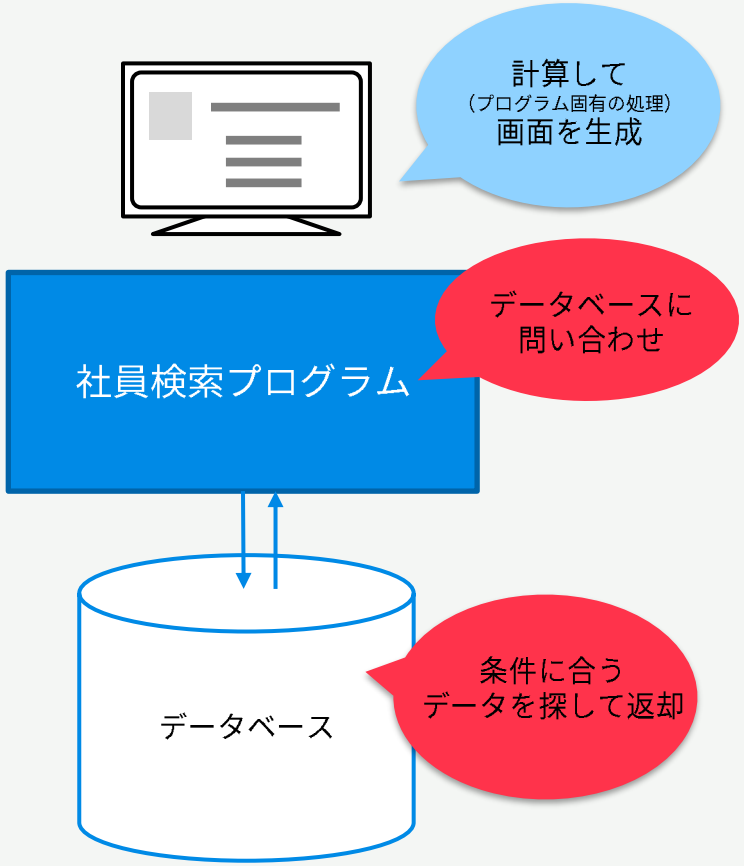
- ✓ 1箇所に統合することで正しさを保証
- ✓ 1箇所に統合することで故障から保護
- ✓ いつでも、だれでも参照・修正（正しい権限を持つ限り）

プログラムとデータの分離

プログラム内部にデータを持っている例

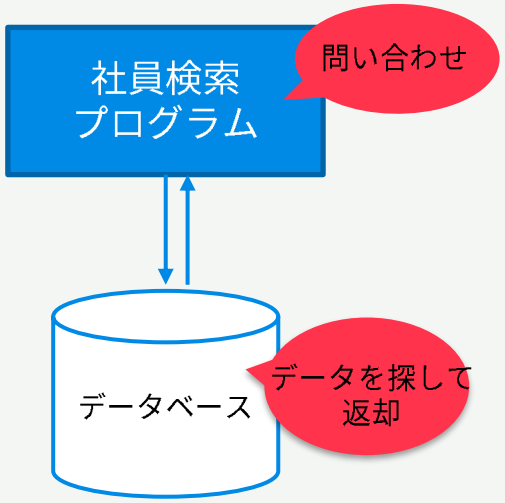


データベースにアクセスするプログラム



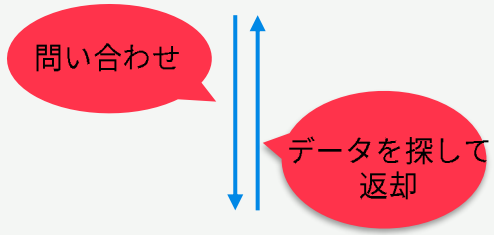
データベースへの問い合わせ

データベースへのアクセスはSQLという標準化された問い合わせ言語を用いる




```
SELECT 社員番号、名前、部署
FROM メンバー表
WHERE 部署 = 営業
```

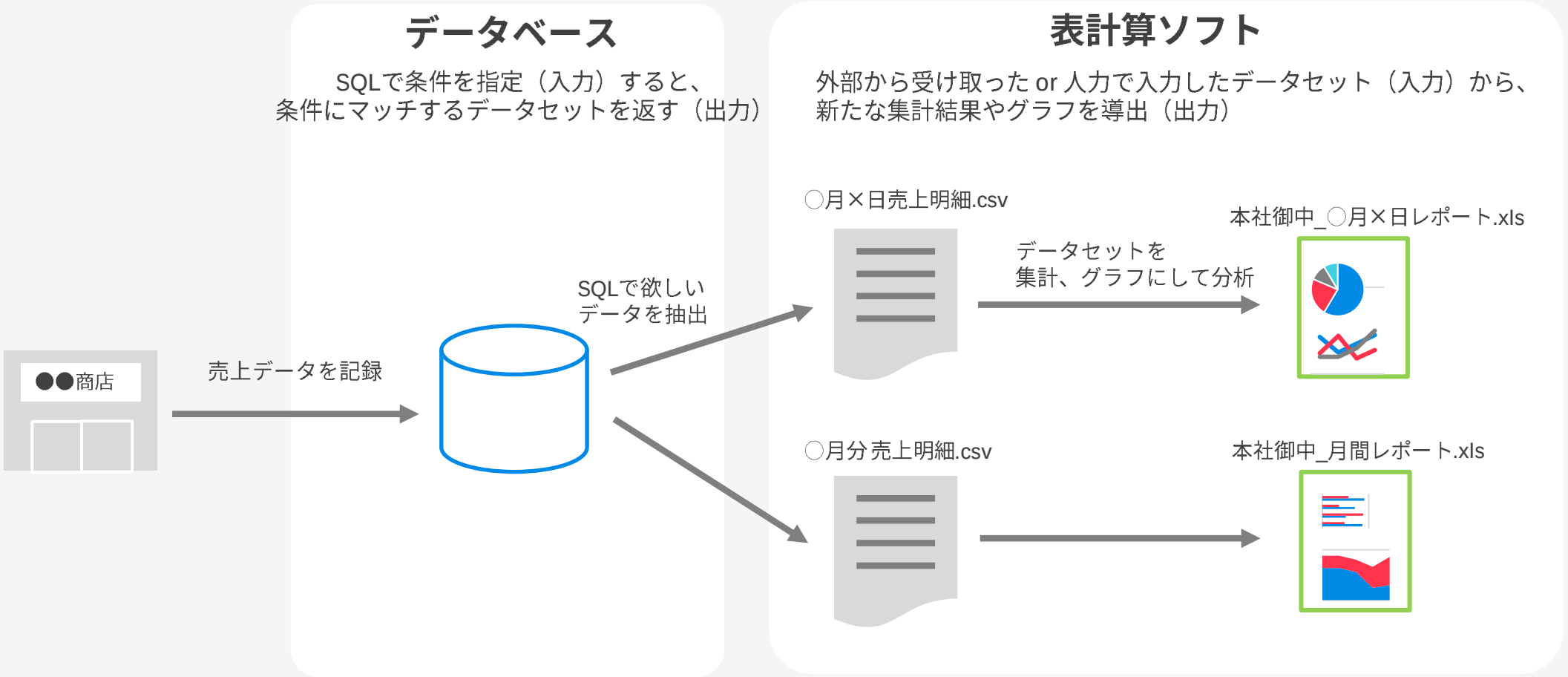
社員番号	名前	部署
102	佐藤	営業
104	高橋	営業



社員番号	名前	部署	メールアドレス
101	田中	技術	tanaka@mail.com
102	佐藤	営業	sato@mail.com
103	加藤	技術	kato@mail.com
104	高橋	営業	takahashi@mail.com
105	稲葉	本部	inaba@mail.com

データベース vs 表計算ソフト

「表形式でデータを保管するもの」と理解されがちだが、利用目的が異なる



PostgreSQL入門

PostgreSQLとは

多機能、高性能、かつオープンソースのリレーショナルデータベース管理システム（RDBMS）

- INGRES(‘70),POSTGRES(’80)由来の歴史
  （カリフォルニア大学バークレイ）
- BSDライセンス（に似た）利用しやすいライセンス
- 特定のオーナー企業がない

バージョン	6.0	7.0	8.0	9.0	10 以降
	SQLに対応	WALに対応	Windows対応 VACUUMの弱点改善	レプリケーション構成 パラレルクエリ	論理レプリケーション パーティショニング 開発者体験

基本機能の実装 → 基本機能の改良・発展 → 高度な要件に応じた機能を充実

PostgreSQLの主要機能

3rd Party Tools

- クラウド統合
- SQL Hint
- 拡張インデックス
- マルチマスタ レプリケーション
- データ連携 移行ツール
- 監査ログ取得

PostgreSQL 近年の強化ポイント

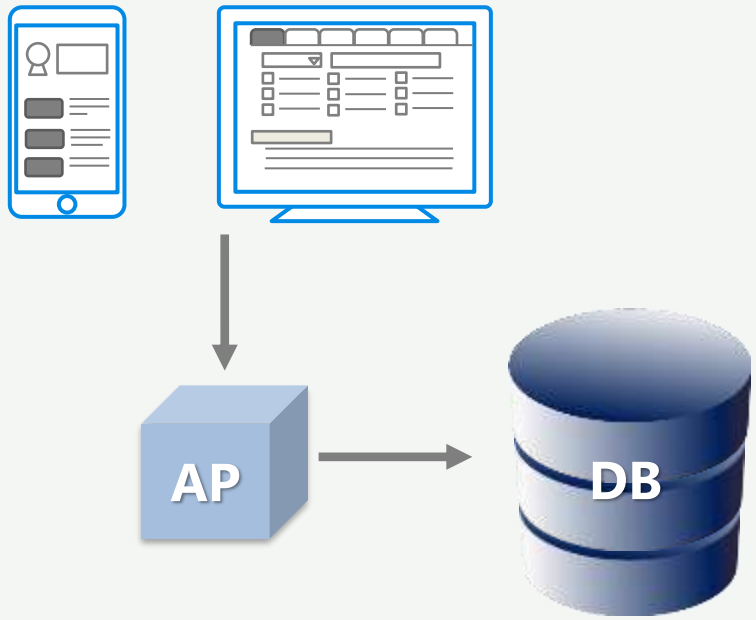
- 内部ロック改善
- 待機イベント
- 自動Prewarm
- 参照可能 レプリケーション
- 外部データラッパ
- 行単位アクセス制御
- テーブル パーティショニング
- パラレルクエリ
- JITコンパイリング
- 論理レプリケーション
- 豊富な計画タイプ
- PL/pgSQL

RDBMSの基本

- MVCC
- 行レベルロック
- 無停止バックアップ
- 論理バックアップ
- ユーザアクセス制御
- システムカタログ
- COMMIT ROLLBACK
- 変更履歴（WAL）
- PITR
- 制約
- 標準SQL
- インデックス

PostgreSQLへの期待

リクエストに対して正しい結果を高速に返す



READ 求める結果を高速に 方法はお任せ

- ✓ WHERE条件にマッチ
- ✓ 聞いた時点のデータ
- ✓ 非手続き型

WRITE データを間違わない 失わない

- ✓ 正しいデータの維持
- ✓ データ保全

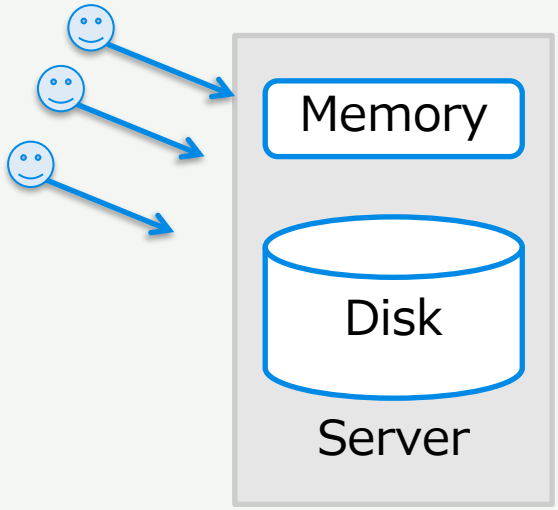
OTHER いつでも使える 安心・安全

- ✓ 可用性
- ✓ セキュリティ
- ✓ 各種調査用の情報

高度な要件への対応

- パラレルクエリ
- 豊富な計画タイプ
- テーブル パーティショニング
- 内部ロック改善
- 論理レプリケーション
- PL/pgSQL
- 待機イベント
- 参照可能 レプリケーション
- 行単位アクセス制御

ハードウェアの限界を超えて必要な機能を実現する

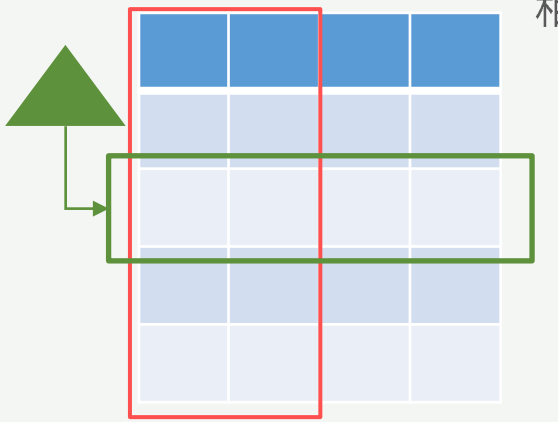


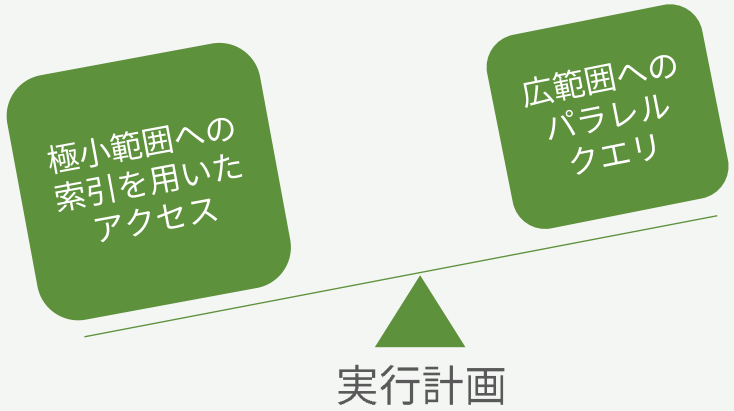
- 性能の観点では、メモリのみで処理を続けることが理想
  ただしデータは永続化したい
- 変更履歴をシーケンシャルI/Oでディスクに保存する事で、
  性能影響を抑えて永続化

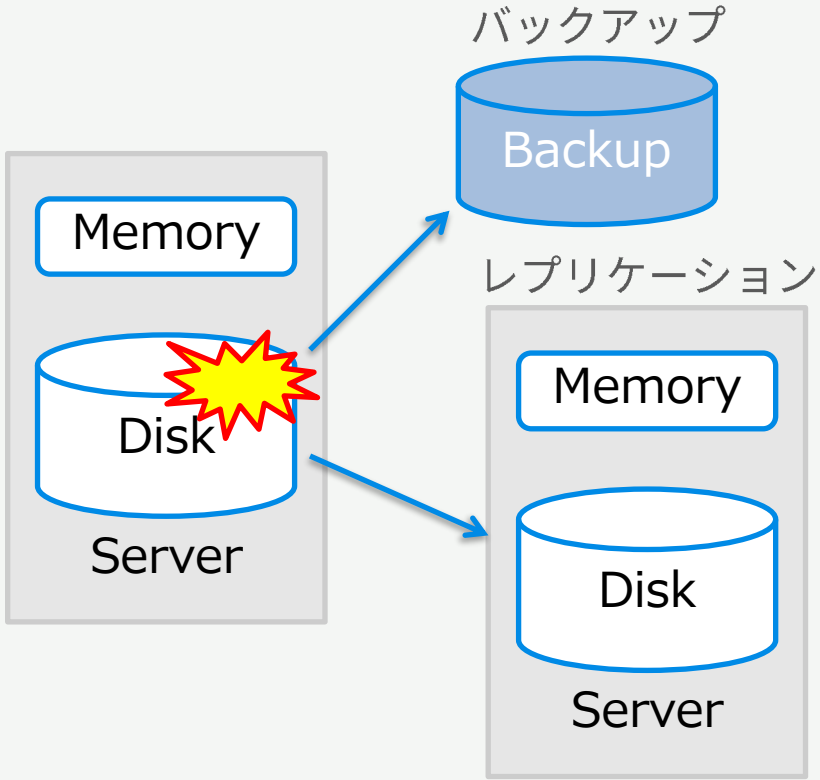
シーケンシャル I/Oの負荷 << ランダム I/Oの負荷

SELECT **名前、TEL**
FROM **メンバー表**
WHERE **社員番号**

- 集合全体が何行であっても
  そこから取り出す結果が何行であっても
  相応に適したパフォーマンスが得られる







共有バッファ上の更新＋WALによる変更履歴の永続化

①クライアントから更新処理を発行

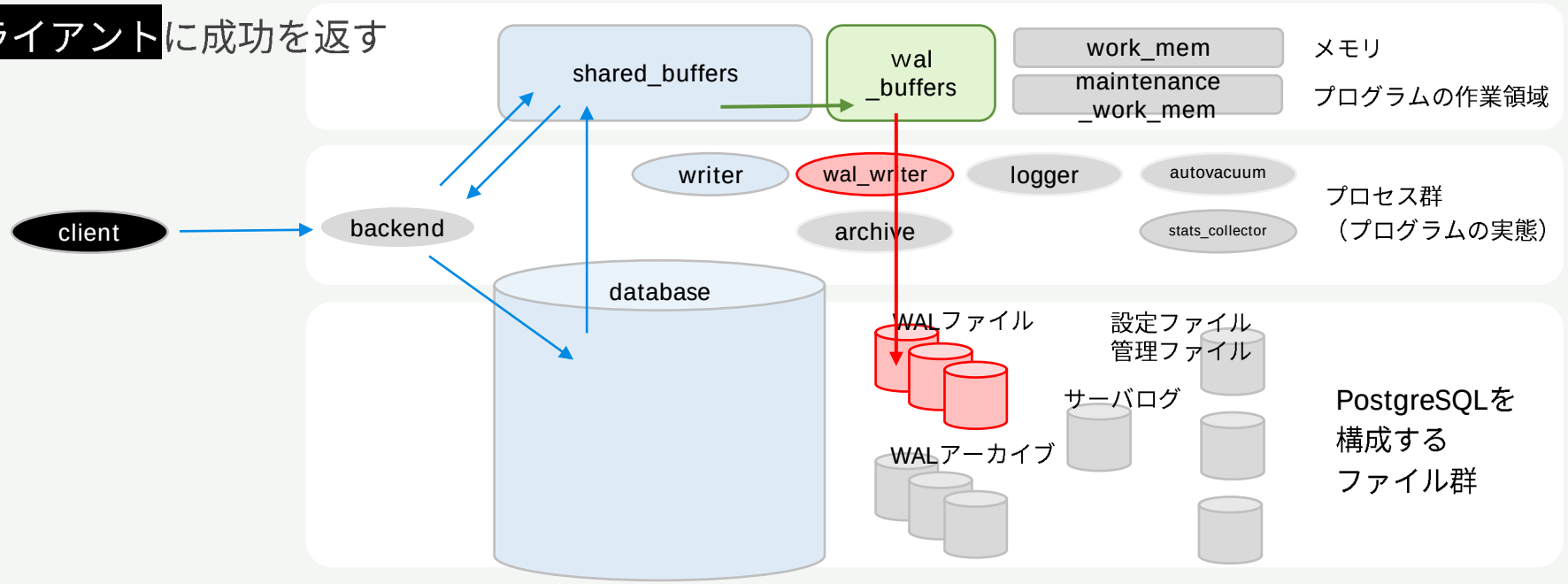
②backendプロセスが必要なデータを共有バッファから探す

③バッファ上に無い場合は、ディスクの該当ブロックをバッファに載せる

④変更内容をWALバッファ上のWALレコードとして作成

⑤共有バッファ上のデータを更新

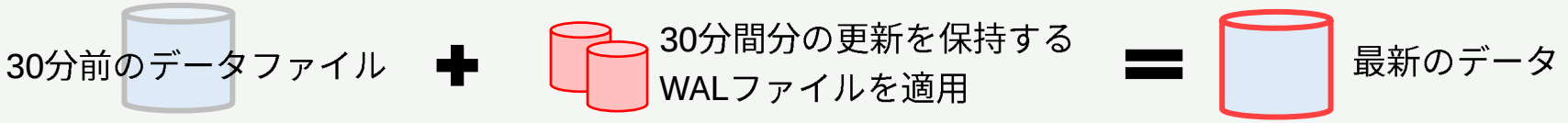
⑥クライアントが変更を確定(COMMIT)すると、WALレコードをWALファイルに永続化し、
WAL書き込みが成功したらクライアントに成功を返す



ディスクに永続化されたデータは復旧のために利用可能

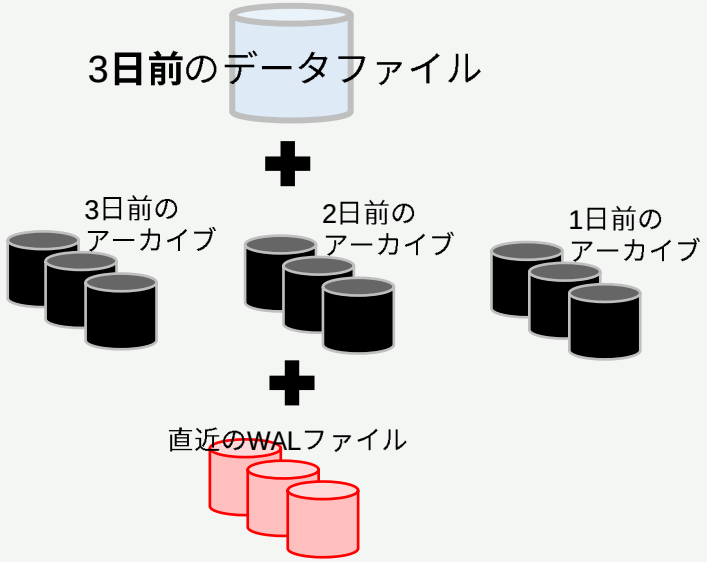
共有バッファ上のデータは、データファイルに定期的に書き込まれる（例：30分間隔）

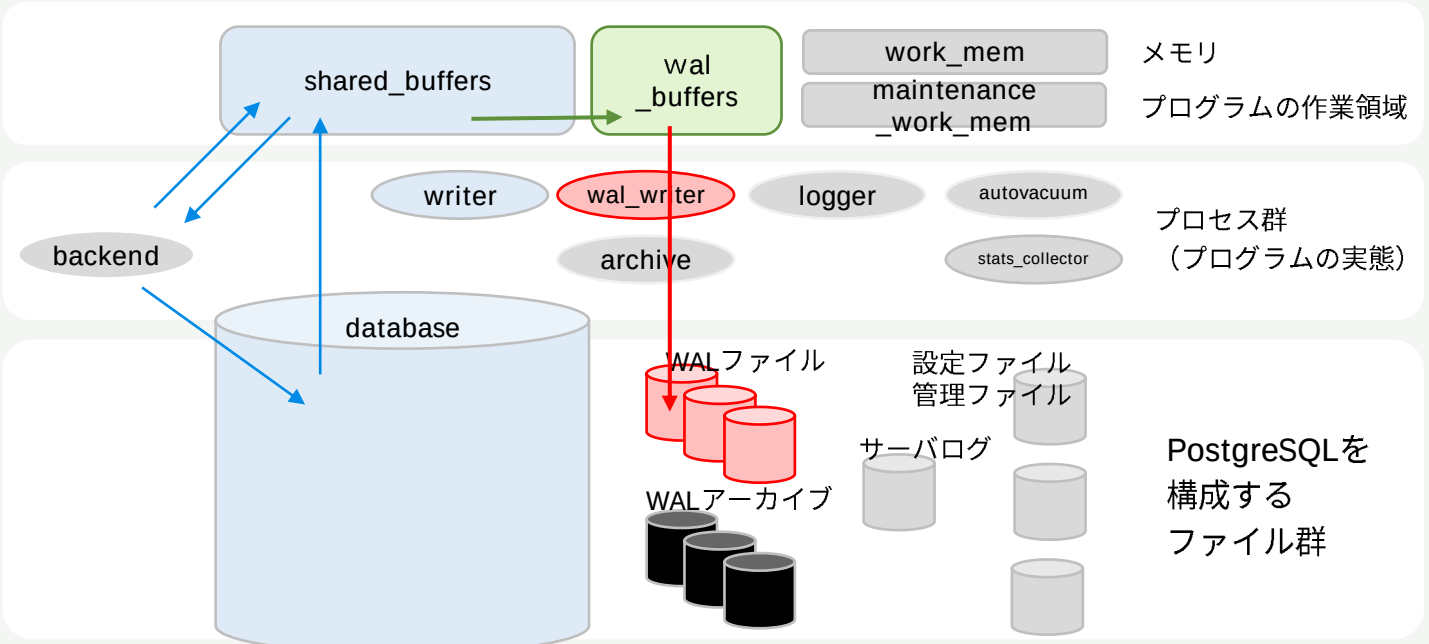
データファイルにまだないデータはWALファイルに永続化されている



ただし、WALファイルは上限サイズを超えないよう循環されるため、無限にさかのぼって保持することは不可能

WALファイルをユーザ管理のもと別領域に退避するアーカイブ機能を利用





SQL開発とデータベース管理

アプリ開発者とデータベース管理者

データベース初学者にとって、今取り組んでいる業務/課題は
アプリ開発観点？データベース管理者の観点？を理解することは非常に重要

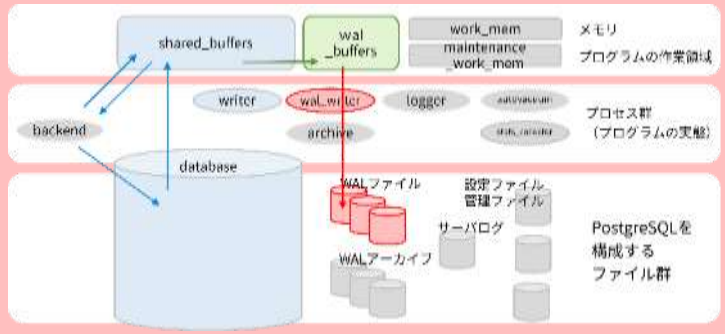
アプリ開発者

データモデリング
SQL開発
パフォーマンス調査

SELECT 名前、TEL
FROM メンバー表
WHERE 社員番号

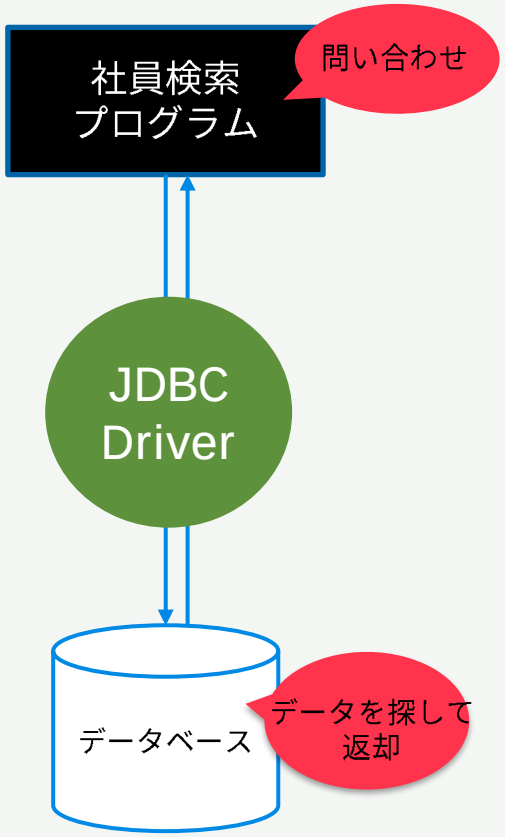
DBA、インフラ管理者

PostgreSQLの設計・実装
メンテナンス操作



アプリケーションデベロッパー：SQLや業務アプリケーションのコーディング

業務アプリ開発の仕事の中で、Java等アプリケーションと一緒にSQLを書く



PostgreSQL用の接続定義

PostgreSQL開発コミュニティで
公開しているJDBCドライバを利用

PostgreSQLへの問い合わせ

問い合わせそのものは
Javaプログラム中にSQL文を直接記述

結果の受け取り、整形、エラー制御

Javaプログラムとして受け取った
データをどう使うかを記載

```java
import java.sql.Connection;
import java.sql.DriverManager;
import java.sql.ResultSet;
import java.sql.SQLException;
import java.sql.Statement;

public class PostgreSQLAccess {

  private static final String DB_URL = "jdbc:postgresql://localhost:5432/mydb";
  private static final String USER = "myuser";
  private static final String PASS = "mypassword";

  public static void main(String[] args) {
    try {
      // PostgreSQL JDBCドライバをロード
      Class.forName("org.postgresql.Driver");

      // データベースへの接続
      Connection connection = DriverManager.getConnection(DB_URL, USER, PASS);

      // SQLクエリの実行
      Statement statement = connection.createStatement();
      ResultSet resultSet = statement.executeQuery("SELECT * FROM mytable");

      while (resultSet.next()) {
        // レコードのデータを取得
        int id = resultSet.getInt("id");
        String name = resultSet.getString("name");
        System.out.printf("ID: %d, Name: %s%n", id, name);
      }

      // リソースを閉じる
      resultSet.close();
      statement.close();
      connection.close();

    } catch (ClassNotFoundException e) {
      System.out.println("PostgreSQL JDBC Driver is not found. Include it in your library path.");
      e.printStackTrace();
    } catch (SQLException e) {
      System.out.println("Connection failed. Check output console.");
      e.printStackTrace();
    }
  }
}
```

上流工程：データモデリング

現実世界の事実をコンピュータで扱うために「モデル化」する

関係モデル

- ✓ 事実のモデル化のための1つの方法
- ✓ システム開発の容易さから最も普及

商品一覧

種類	金額	産地
りんご	200円	青森産
みかん	50円	愛媛産
バナナ	100円	沖縄産

在庫

種類	個数
りんご	50個
みかん	300個
バナナ	20房

売上実績

日時	もの	個数
9:00	りんご	1個
9:01	りんご	2個
10:00	みかん	10個

果物屋の販売業務をモデル化
・この形で表現しきれないケースを考えてください
・表現しきれない=モデル化の誤り
・どのような形なら解決できますか？

現実世界の事実をコンピュータで扱うために「モデル化」する

関係モデル

- ✓ 事実のモデル化のための1つの方法
- ✓ システム開発の容易さから最も普及

商品一覧

種類	金額	産地
りんご	200円	青森産
みかん	50円	愛媛産
バナナ	100円	沖縄産
りんご	150円	山梨産

在庫

種類	個数
りんご	50個
みかん	300個
バナナ	20房
りんご	30個

売上実績

日時	もの	個数
9:00	りんご	1個
9:01	りんご	2個
10:00	みかん	10個
10:05	りんご	2個

別の産地のりんごが商品に加わったとき、在庫や売り上げ実績が破綻

果物屋の販売業務をモデル化
- この形で表現しきれないケースを考えてください
- 表現しきれない=モデル化の誤り
- どのような形なら解決できますか？

商品一覧

商品ID	種類	金額	産地
001	りんご	200円	青森産
002	みかん	50円	愛媛産
003	バナナ	100円	沖縄産
004	りんご	150円	山梨産

在庫

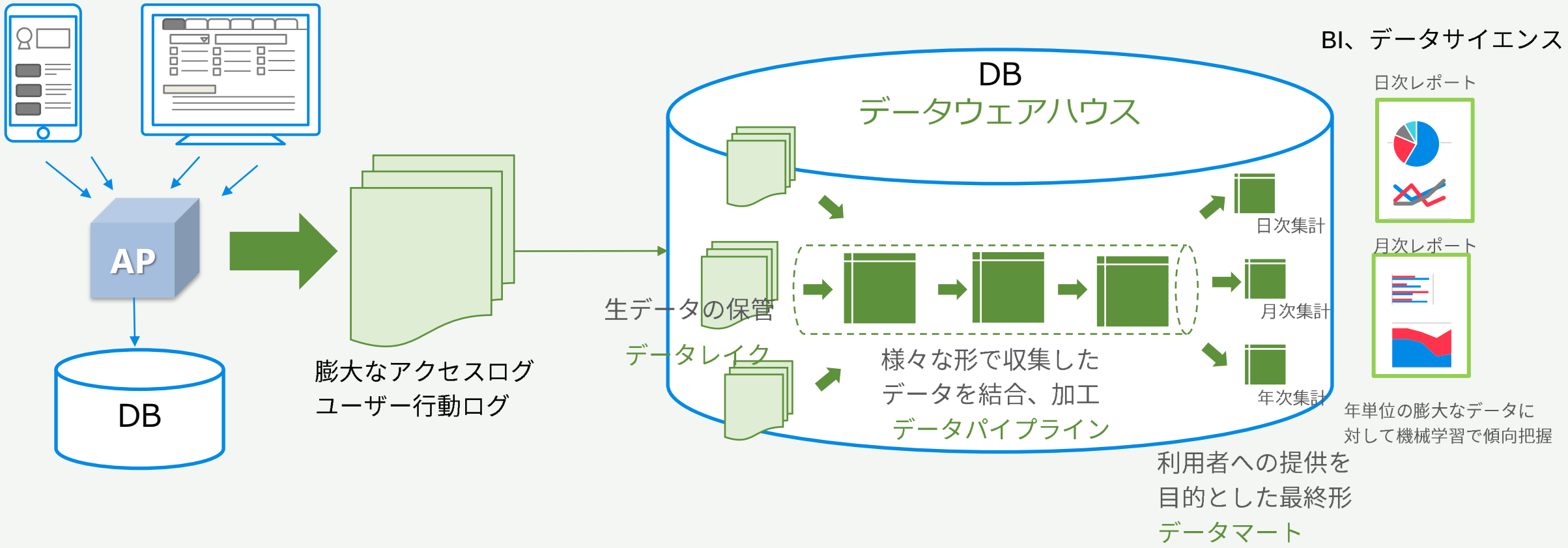
商品ID	個数
001	50個
002	300個
003	20房
004	30個

売上実績

日時	商品ID	個数
9:00	001	1個
9:01	004	2個
10:00	002	10個

データエンジニア

データドリブン時代の大量データ処理を担う専門領域



データベース管理

アプリケーションに対して常に安定的にデータベース機能を提供するために

DBA、インフラ管理者

- 構築・メンテナンス
- 監視
- 監査
- 運用処理開発、手順整備
- バックアップ計画
- コスト・ライセンス管理

おわりに

PostgreSQLやデータベース、SQLを学習することは

✓あらゆるアプリケーション開発

✓企業のDXやデータドリブン経営といったトレンドの領域

に無くてはならない必須知識（かつ、学習コストは低め）

はじめてのデータベース学習にPostgreSQLを用いる理由

- ✓ 無償で配布されているまたは各社クラウドサービス上にラインナップされ、どんな学習環境でも用意できる
- ✓ SQL標準に準拠し、幅広いSQL構文・機能を学習できる
- ✓ 日本語マニュアルや技術記事が豊富にヒットし困りごとを解決しやすい

PostgreSQL**の最新情報にアクセス**

Slack：postgresql-jp
https://tinyurl.com/pgsql-jp-slackin

PostgreSQLカンファレンス2023
https://www.postgresql.jp/jpug-pgcon2023