
帝塚山学院大学
 TEZUKAYAMA GAKUIN UNIVERSITY

データベース概論

NonSQL, GoogleのDB

ビッグデータ


中野秀男

情報メディア学科

1 データベース概論 NonSQL,Google 2019/1/21


今日の話

- ▶ 授業アンケートに回答
- ▶ 復習
- ▶ マルチメディアのDBを簡単に
- ▶ 質問やコメントの回答
- ▶ NonSQL
- ▶ グーグルのDB
- ▶ ビッグデータ(時間があれば)
 - ▶ ビッグデータ
 - ▶ ビッグデータの例
 - ▶ ビッグデータの利用方法

2 データベース概論 NonSQL,Google 2019/1/21
 
帝塚山学院大学
 TEZUKAYAMA GAKUIN UNIVERSITY


今までの整理(1)

- ▶ データベースは集める、整理する、検索する
- ▶ データ: 実体、データの性質: 属性、そしてデータの集合
 - ▶ データはレコードで、その属性がフィールド、全体でテーブル
- ▶ ExcelやAccessでデータ整理はできる
 - ▶ Accessはデータベース用アプリ
- ▶ 複数の利用者が同時に利用、権限制御などはSQLで統一
- ▶ 高速に検索や処理のために便利なB-treeのデータ構造
- ▶ レベルに応じて
 - ▶ Excel, Access, MySQL, PostgreSQL, SQLserver, Oracle

3 データベース概論 NonSQL,Google 2019/1/21
 
帝塚山学院大学
 TEZUKAYAMA GAKUIN UNIVERSITY


今までの整理から(2)

- ▶ データの集合: 表, テーブル
 - ▶ いろいろなデータの集合があるので、
 - ▶ それぞれに主キーを決めて、
 - ▶ 相互の関係は外部キーで関係をつける
 - ▶ マスターテーブル(台帳): 基本になるテーブル
 - ▶ データテーブル: 日々更新されるデータ
- ▶ SQL
 - ▶ 複数の利用者が同時に利用、権限制御などでSQLで統一
 - ▶ ネットを通して利用できるように: サーバとクライアント(端末)
 - ▶ いろいろな言語から使えるように
 - ▶ PHP, Perl, Python, Ruby, C, C++, C#, Java
- ▶ NonSQL: SQLでは扱うのが難しい半構造データ
 - ▶ 例えば: 名前を名前だけとか姓と名とか、複数あるメアドとか

4 データベース概論 NonSQL,Google 2019/1/21  帝塚山学院大学
TEZUKAYAMA GAKUIN UNIVERSITY


マルチメディア

- ▶ 一般にマルチメディアは
 - ▶ シンプルテキスト: plain text, txt
 - ▶ 書式付きテキスト: rich text, rtf
 - ▶ 静止画像、動画像、音声、音楽
- ▶ 非テキスト型メディア
 - ▶ 静止画像、動画像、音声、音楽
- ▶ 非テキスト型メディアのデータベースの特徴
 - ▶ データ量が膨大
 - ▶ コンピュータが扱いにくいデータ
- ▶ メタデータ
 - ▶ マルチメディアの中身を表すデータ

5 データベース概論 NonSQL,Google 2019/1/21  帝塚山学院大学
TEZUKAYAMA GAKUIN UNIVERSITY


マルチメディアDBシステムの困難性

- ▶ 画像データベースの処理の流れ
 - ▶ 1. ユーザが「どんな画像が欲しいか」をDBに伝える
 - ▶ 2. DBシステムがユーザの検索要求を解析
 - ▶ 3. DBシステムがユーザの要求に近いメタデータを探し選択
 - ▶ 4. DBシステムが検索結果(候補画像)をユーザに提示
- ▶ 要求の難しさ
 - ▶ 要求があいまい
- ▶ 画像や音や映像にハッシュタグのようなメタデータを

6 データベース概論 NonSQL,Google 2019/1/21  帝塚山学院大学
TEZUKAYAMA GAKUIN UNIVERSITY


質問やコメント(1)

- ▶ マルチメディアはよくわからない
- ▶ マルチメディアDBがよくわからない
- ▶ あまりメタデータは聞かない
- ▶ メタはどういう意味か
- ▶ メタデータとデータの違い
- ▶ インスタやツイッターでハッシュタグが役に立つ
- ▶ 「Googleが描いた馬」の説明がわかりやすかった
- ▶ Googleの描画も進歩している
- ▶ 映画の容量は大きい
- ▶ 検索サイトの画像のところはマルチメディアDBか
- ▶ 動画を入れるのに容量が小さくできる日は

▶ 7 データベース概論 NonSQL,Google 2019/1/21  帝塚山学院大学
TEZUKAYAMA GAKUIN UNIVERSITY


質問やコメント(2)

- ▶ Googleドライブについて学んで見たい
- ▶ カベルスキーのソフトの強みは
- ▶ 電子マネーが広まらないのは規格がバラバラ
- ▶ ドメイン名のルートは「.」
- ▶ ロボホンにクリスマス限定のものは
- ▶ ゼミでクリスマスイベントはあるか
- ▶ SPIは今からでも間に合うか
 - ▶ 新聞を読めと言われても

▶ 8 データベース概論 NonSQL,Google 2019/1/21  帝塚山学院大学
TEZUKAYAMA GAKUIN UNIVERSITY

半構造

- ▶ SQLの限界
 - ▶ あいまいな記述ができない
 - ▶ 例:メールアドレスが二つある
 - ▶ 例:携帯電話を二台持っているので電話番号が二つある
 - ▶ リンクしとけば簡単に検索できるのに、そんな機能がない
- ▶ 半構造
 - ▶ 自己記述にしておく:データにラベルをつけておく
 - ▶ リンクも記述する

▶ 9 データベース概論 NonSQL,Google 2019/1/21  帝塚山学院大学
TEZUKAYAMA GAKUIN UNIVERSITY

Googleのデータベース(1)

- ▶ 世界中のウェブを自動巡回プログラム(ロボット)で収集
- ▶ 安価なコンピュータや補助記憶でDBを構成
- ▶ 列(フィールド)の単位で整理せずに、行の単位で整理
- ▶ 複数の行(文書)を複数のコンピュータに保存
- ▶ 検索要求(クエリ)はありそうな複数のコンピュータに要求
- ▶ 多くの同時にくる大量のクエリに平均的に対応
- ▶ シャーディング: データの断片化
- ▶ レプリケーション: 複製
- ▶ ページランク

▶ 10

データベース概論 NonSQL Google 2019/1/21



Googleのデータベース(2)

- ▶ Googleのデータベース説明の図が入ります。

▶ 11

データベース概論 NonSQL Google 2019/1/21



Googleのデータベース(3)

- ▶ Googleのデータベース説明の図が入ります。


▶ 12

データベース概論 NonSQL Google 2019/1/21




ビッグデータ (Wikipedia)

- ▶ 通常のデータベースでは扱えない量
- ▶ 構造化データや非構造化データ
- ▶ ペタ、エクサ、ゼタ、更に上
 - ▶ K, M, G, T, P, E, Z
- ▶ 格納や活用(検索/共有/分析/可視化)
- ▶ いろいろな分野で先端技術を使う

13 データベース概論 NonSQL,Google 2019/1/21  帝塚山学院大学
TEZUKAYAMA GAKUIN UNIVERSITY


ビッグデータの例

- ▶ 政府や自治体を持つ住民や組織情報
- ▶ 家電やシステムの操作ログ
- ▶ ライフログ
- ▶ センサーネット情報
- ▶ 顧客データ
 - ▶ 政府、自治体、企業、NPO
- ▶ ソーシャルメディアデータ
- ▶ マルチメディアデータ
- ▶ ウェブサイトデータ

14 データベース概論 NonSQL,Google 2019/1/21  帝塚山学院大学
TEZUKAYAMA GAKUIN UNIVERSITY


ログ

- ▶ 操作ログ
 - ▶ 家電メーカーの思惑
 - ▶ B-CAS問題
 - ▶ フォレンジック(PC操作の記録)
- ▶ ライフログ
 - ▶ 歩数等の活動の記録
 - ▶ 読んだもの等の知識の記録
 - ▶ 写真等の記憶の記録
 - ▶ 体重や脳波等の身体の記録
 - ▶ レシート等の身体の記録
 - ▶ 位置情報等の環境の記録
 - ▶ 気分等の精神の記録

15 データベース概論 NonSQL,Google 2019/1/21  帝塚山学院大学
TEZUKAYAMA GAKUIN UNIVERSITY


ビッグデータのステップ

- ▶ 多様な複数のデータを
- ▶ 総合的に収集・分析することで
- ▶ 可視化する/新たな知見を得る
- ▶ プラットフォームの構築(高速、大容量)
- ▶ サードパーティの参加(複数で多く)

▶ 16 データベース概論 NonSQL,Google 2019/1/21  帝塚山学院大学
TEZUKAYAMA GAKUIN UNIVERSITY


時代の背景

- ▶ コンテンツ:
 - ▶ ソーシャルメディア:
 - ▶ Twitter, Facebook, Instagram, Tik Tok
 - ▶ 動画配信: Youtube
- ▶ プラットフォーム
 - ▶ クラウド、Android OS
- ▶ ネットワーク
 - ▶ ワイヤレス通信、ブロードバンド
- ▶ デバイス
 - ▶ 各種センサーの普及, IoTの時代
 - ▶ ICT端末の多様化:携帯,PC,TV, スマホ

▶ 17 データベース概論 NonSQL,Google 2019/1/21  帝塚山学院大学
TEZUKAYAMA GAKUIN UNIVERSITY

仮説を検証

- ▶ 仮説を立て
- ▶ 複数の大量のデータを集め
 - ▶ 大容量/分散ストレージ
- ▶ データマイニング等の手法で推論(?)
 - ▶ 高速/並列計算、アルゴリズム
- ▶ 例
 - ▶ 紙オムツを買った人は缶ビールも買う
 - ▶ 冬にスキー用品を買った人は、夏にテニス用品を

▶ 18 データベース概論 NonSQL,Google 2019/1/21  帝塚山学院大学
TEZUKAYAMA GAKUIN UNIVERSITY
