


帝塚山学院大学
 TEZUKAYAMA GAKUIN UNIVERSITY

データベース概論
NoSQL, GoogleのDB
まとめ


中野秀男
 帝塚山学院大学非常勤講師
 大阪市立大学名誉教授、堺市情報セキュリティアドバイザー

1 データベース概論 NoSQL,Google 2020/1/27

今日の話

- ▶ 質問やコメントの回答
- ▶ 半構造(復習)
- ▶ NonSQL
- ▶ ゲーグルのDB
- ▶ まとめ


2 データベース概論 NoSQL,Google 2020/1/27


帝塚山学院大学
 TEZUKAYAMA GAKUIN UNIVERSITY

質問やコメント(1)


- ▶ ExcelとAccessのどちらを使うか
- ▶ 非テキスト型メディアの検索を高性能にするには何が必要
- ▶ DBシステムは画像検索でも使われている
- ▶ マルチメディアのDBは身近にあるもの全てか
- ▶ Googleの検索に自分の名前を
 - ▶ エゴサーチ
 - ▶ 権利は
- ▶ キーワード検索はすごい
 - ▶ 複数のキーワードが入力できる
- ▶ GoogleとYahooのデータベースはどちらが大きいか

3 データベース概論 NoSQL,Google 2020/1/27


帝塚山学院大学
 TEZUKAYAMA GAKUIN UNIVERSITY


質問やコメント(2)

- ▶ 個人を特定するのは簡単か
- ▶ C-learningが新しくなる
- ▶ 陸,海,空,宇宙,サイバーの次は
- ▶ ヨドバシカメラの梅田がマルチメディア

4 データベース概論 NonSQL,Google 2020/1/27  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

5 データベース概論 NonSQL,Google 2020/1/27  帝塚山学院大学
TEZUKAYAMA GAKUIN UNIVERSITY


NoSQL(1)

- ▶ NoSQL: Not only SQL
- ▶ Wikipedia
 - ▶ NoSQL(一般に "Not only SQL" と解釈される)とは、関係データベース管理システム (RDBMS) 以外のデータベース管理システムを指す **おおまかな分類語** である。関係データベースを杓子定規に適用してきた長い歴史を打破し、それ以外の構造のデータベースの利用・発展を促進させようとする **運動の標語** としての意味合いを持つ。
- ▶ RDBMSの歴史
 - ▶ 1977年 Oracle Database(Oracle)
 - ▶ 1983年 DB2(IBM)
 - ▶ 1989年 SQL Server(Microsoft)
 - ▶ 1989年 Postgres(OSS)
 - ▶ 1995年 MySQL(OSS)

6 データベース概論 NonSQL,Google 2020/1/27  帝塚山学院大学
TEZUKAYAMA GAKUIN UNIVERSITY


NoSQL(2)

- ▶ 時代の推移
 - ▶ ITシステムの成長(CPU、メモリ、ストレージ、ネットワーク)
 - ▶ ムーアの法則(18ヶ月で2倍、3年で4倍)
 - ▶ ギルダーの法則/カオの法則(通信網の帯域幅は6ヶ月で2倍)
 - ▶ 大量のデータとボトルネック
 - ▶ CPU - メモリ - ストレージ - ネットワーク(ネットワークがボトルネック)
- ▶ 解決策は分散ファイルシステム/分散データベース
 - ▶ Google File System(GFS)
 - ▶ Hadoop Distributed File System

7 データベース概論 NonSQL,Google 2020/1/27  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

8 データベース概論 NonSQL,Google 2020/1/27  帝塚山学院大学
TEZUKAYAMA GAKUIN UNIVERSITY


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。

9 データベース概論 NonSQL,Google 2020/1/27  帝塚山学院大学
TEZUKAYAMA GAKUIN UNIVERSITY


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。

10 データベース概論 NonSQL,Google 2020/1/27  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

11 データベース概論 NonSQL,Google 2020/1/27  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

12 データベース概論 NonSQL,Google 2020/1/27  帝塚山学院大学
TEZUKAYAMA GAKUIN UNIVERSITY
