PostgreSQL

ITWS 2110 & CSCI 2960
Web Systems Development
Fall 2015

Ahmed Eleish

Agenda

- Intro
- Postgres
- Object-relational DBMS
- Features of PostgreSQL
  - Inheritance
  - Datatypes
  - GiST
  - Compliance
  - MVCC
  - Full text search
  - Extensibility
  - Community
- PHP and PostgreSQL

Database

- An organized collection of data
- Has a database model
- Is managed by a Database Management System

Database Management System

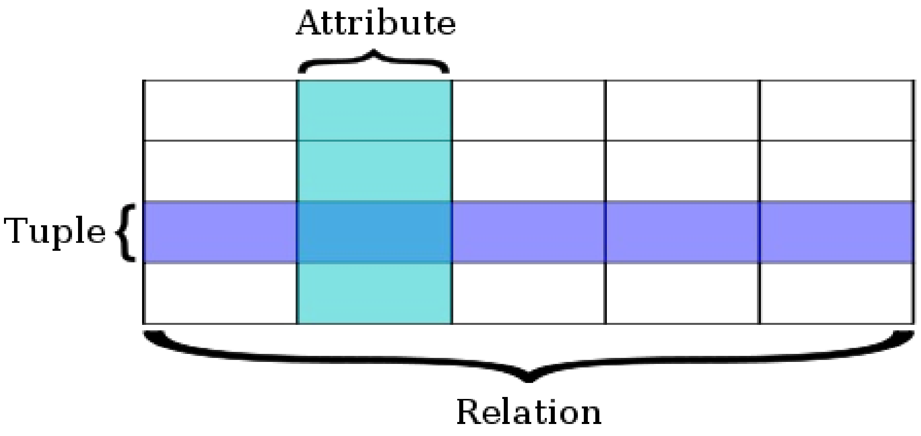
- Application that manages the creation, querying, update, and administration of databases

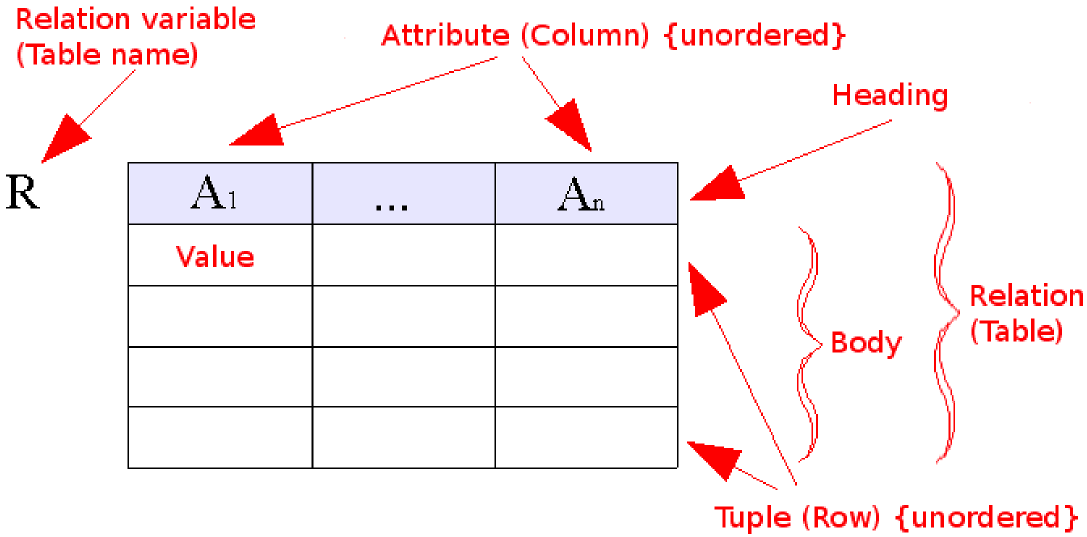
- Supports particular logical and physical data models

- Allows users to create/manipulate data/information while taking care of the underlying data structures

Relational Database Management Systems

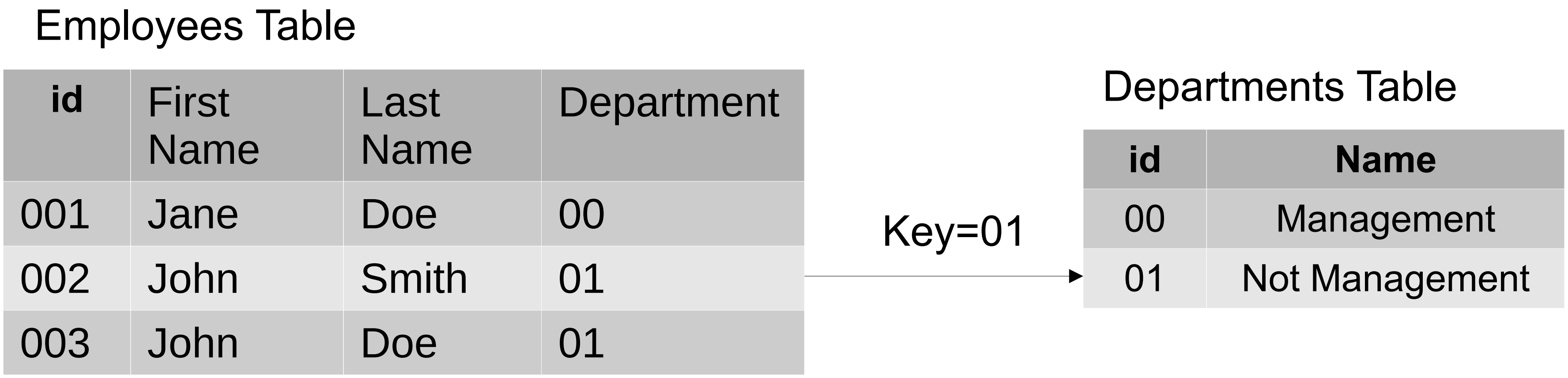
- Popular since the 1980s
- MySQL, PostgreSQL, Oracle, MS SQL Server, IBM DB2
- Implement the **Relational Model** proposed by E.F. Codd in 1969





Relational Database Management Systems

- Tables, keys, and constraints
- Compact and clean
- Mathematical base allows deep analysis of data



Employees Table

id	First Name	Last Name	Department
001	Jane	Doe	00
002	John	Smith	01
003	John	Doe	01

Key=01

Departments Table

id	Name
00	Management
01	Not Management

Structured Query Language (SQL)

- Programming language for RDBMS
- Originally based on relational algebra & relational calculus
- Data Definition Language (DDL):
  - CREATE, ALTER, DROP
- Data Manipulation Language (DML):
  - SELECT..FROM, INSERT..INTO, UPDATE..SET, DELETE..FROM
- Date Control Language (DCL):
  - GRANT, REVOKE

ACID-Compliance

- Atomicity
  - “All or nothing”: if part of the transaction fails the entire transaction fails
- Consistency
  - Write operations will bring the DB to a valid state, according to constraints, cascades, and triggers
- Isolation
  - Serializability, result of concurrent execution of transactions as if executed serially
- Durability
  - Committed transactions will not be lost due to hardware/software failure

PostgreSQL

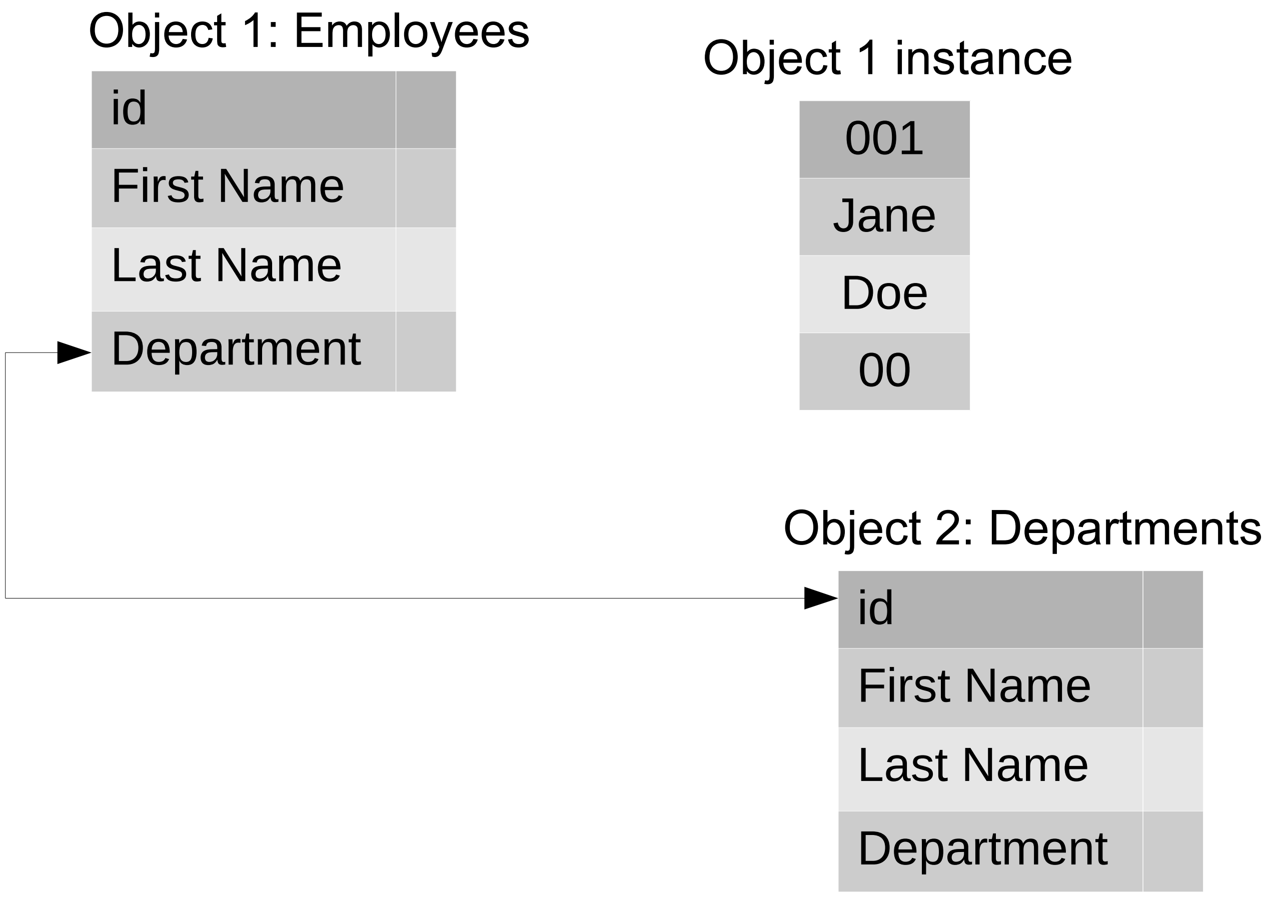
- Object-relational database system
- Runs on all major OS
- ACID-compliant
- Programming interfaces for Java, .NET, Perl, Python, Ruby, C/C++, and its own PL/pgSQL
- Enterprise class, highly scalable

PostgreSQL

- Evolved from project Ingres at UC, Berkeley
- Originally titled Postgres
- Renamed PostgreSQL in '96
- Developed by PostgreSQL Global Development Group
- **Free and open-source**

Object-relational DBMS

- Database model is hybrid of relational model and **object-oriented model**



Object 1: Employees

id	
First Name	
Last Name	
Department	

Object 1 instance

001
Jane
Doe
00

Object 2: Departments

id	
First Name	
Last Name	
Department	

Why Object-relational

- Relational Database
  - Mathematical relation allows set theory techniques in data analysis
  - Impedance mismatch between application and DB

- Object-oriented Database
    - Allows sets, lists, user-defined datatypes and nested objects
    - No mathematical base for deep analysis

- ORDBMS bridge the gap
    - Complex data, nested objects, type inheritance
    - Mathematical base for analysis

PostgreSQL: Type Inheritance

- Tables can inherit from parent tables
- Data in child tables will appear as if they exist in parent tables
- Can be used for data partitioning
  - e.g. Separate tables for monthly sales data
    - When querying date range only relevant tables are scanned
    - Smaller tables with smaller indexes mean faster searches
    - Easier maintenance
- Has shortcomings regarding unique constraints and foreign keys

PostgreSQL: Supported Datatypes

- Native datatypes:
  - Boolean, Character, Binary, Date/Time, bigint, integer, double precision, XML, JSON, IPv4, IPv6 addresses, MAC addresses

- User-defined datatypes:
  - Complex objects, domains, ranges, geometric types such as points, lineStrings, Polygons,…
  - Access methods for these types need to be defined

Generalized Search Tree (GiST)

- Disk-based indexing system, generalized form of the B+ tree
- Allows creation of custom datatypes with indexed access methods
- Can be used to implement B+ Trees, R-Trees, among others
- Includes support for variable-length keys and composite keys
- GiST is extensible

Generalized Search Tree (GiST)

- Can support nearest neighbor search (KNN-GIST)

- Can be used to find similar words or close objects or locations within geospatial data
  - This makes it ideal for GIS databases

PostGIS

- Adds support for geographic objects to PostgreSQL
- Follows the Simple Features spec. from the Open Geographic Consortium
- Features
  - Geometry types for Points, LineStrings, Polygons, MultiPoints,…
  - Spatial predicates
  - Geospatial operators for measurements and set operations

PostgreSQL: Compliance

- Implements majority of SQL:2011
  - Makes migration of data to other systems easy
- Fully ACID-compliant
  - High reliability and emphasis on data integrity
  - Implements MVCC to manage concurrency

Multiversion Concurrency Control

- Every transaction gets a “snapshot” of the database
- Uses timestamps and incrementing transaction IDs
- A transaction sees only transactions completed before it started

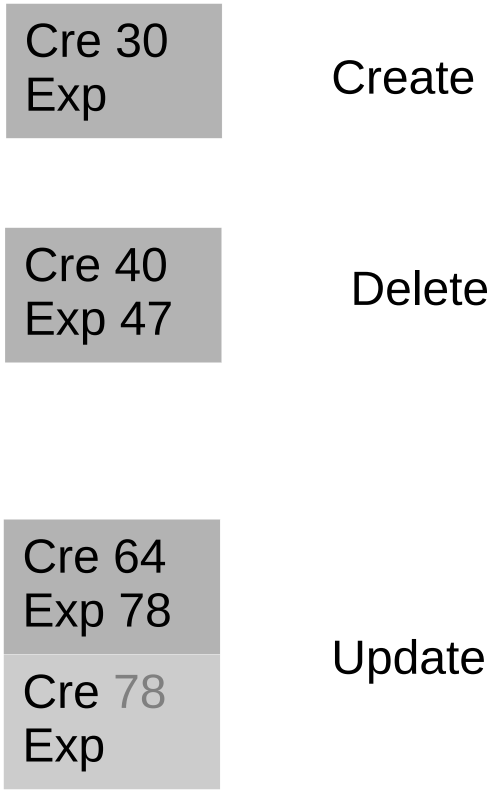
Multiversion Concurrency Control

- On query start PostgreSQL records the transaction counter
- Visible rows must have a creation transaction id that:
  - Is a committed transaction
  - is less than the transaction counter recorded at query start

    and
  - was not in-process at query start
- Visible rows must also have an expire transaction id that:
  - is blank or aborted or
  - is greater than the id stored at query start

    or
  - was in-process at query start

Multiversion Concurrency Control



Cre 30
Exp

Create

Cre 40
Exp 47

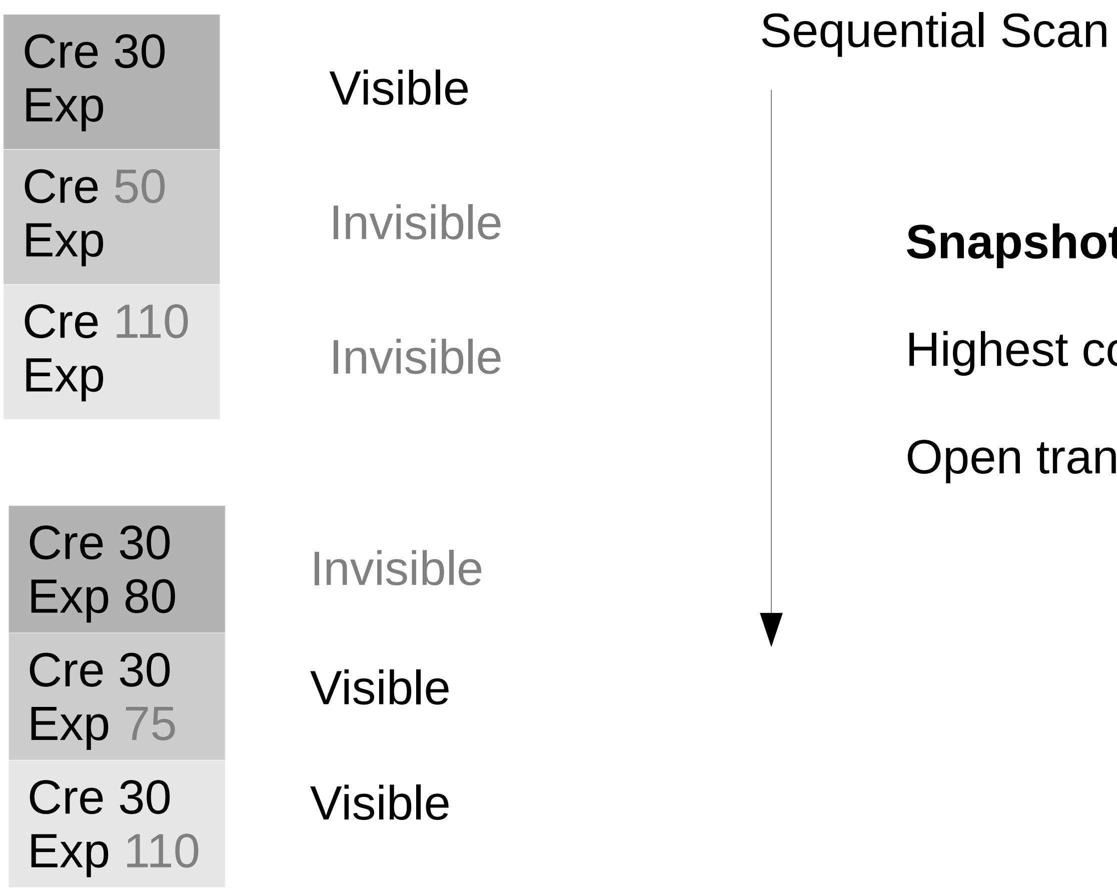
Delete

Cre 64
Exp 78

Cre 78
Exp

Update

Multiversion Concurrency Control



Sequential Scan

Cre 30
Exp

Visible

Cre 50
Exp

Invisible

Cre 110
Exp

Invisible

Cre 30
Exp 80

Invisible

Cre 30
Exp 75

Visible

Cre 30
Exp 110

Visible

Snapshot

Highest committed transaction: 100

Open transactions: 25, 50, 75

PostgreSQL: Full Text Search

- Problems with regular text search in DB systems:
  - No linguistic support
  - Search results are not ranked
  - Slow without index support
- In PostgreSQL:
  - Documents are preprocessed by:
    - Parsing documents into tokens
    - Converting tokens into lexemes using *dictionaries*
    - Storing preprocessed documents

PostgreSQL: Full Text Search

- Dictionaries:
  - Programs that convert tokens into lexemes
  - Normalize words so that derived forms of the same word will match
  - PostgreSQL provides predefined dictionaries for many languages
  - Templates are available to create new dictionaries

PostgreSQL: Extensibility

- Open-source with permissive free-software license

- Programming interfaces for C/C++, Java, .NET, Perl, Python, ODBC,…

- Stored procedures can be written in over a dozen languages and loaded as libraries

- Many third-party tools for designing and managing the system

PostgreSQL: Extensibility

- GiST enables the creation of new indexing methods allowing users to specify..
  - What to store
  - How to store it

    and
  - Define new ways of search it

PostgreSQL: Community

- PostgreSQL Global Development Group
  - Diverse group of companies and individual contributors
- Active community including mailing lists and IRC
- Commercial support options also available

When to use PostgreSQL

- Where standards and compliance to SQL are expected
- If data integrity and reliability are the top priority
- When complex, custom procedures are needed
- Where complex database designs are required
- If an enterprise class DBMS is needed with low costs

When not to use PostgreSQL

- In systems with a simple structure

- Where *read* performance is the top requirement

- Where database and system administration experience may be lacking

PHP and PostgreSQL

- PHP includes an API for connecting to, and manipulating data in a PostgreSQL database
- PostgreSQL 6.5 or later
- Connecting to DB
  - pg_connect, pg_close
- Manipulating data
  - pg_query, pg_update
- Prepared Statements
  - pg_prepare, pg_execute

PHP and PostgreSQL

- PHP also includes PDO_PGSQL which implements PDO to enable access to PostgresQL
- Functions
  - PDO_PGSQL DSN – Connecting to database
  - PDO::pgsqlCopyToArray – Copying data from db to PHP