Databases for Large Amounts of Data

Luís Pedro Coelho

Programming for Scientists

March 31, 2009

University of Pittsburgh

Carnegie Mellon

Databases

For large amounts of data (thousands or millions of records).

Example Databases

Traditional

- MySQL (open source)
- Postgres (open source)
- Oracle
- MS-SQL
- ...

Non-traditional

- SQLite (open source)
- Pytable [HDF5] (open source)
- MS-Access
- ...

Traditional Databases

Two Processes

1. Application
2. Database

Relational Databases

- Based on relations
- Relational database management system (RDMS)

Person

Login	LastName	FirstName
skol	Kovalevskaya	Sofia
mlom	Lomonosov	Mikhail
dmitri	Mendeleev	Dmitri
ivan	Pavlov	Ivan

Project

ProjectId	ProjectName
1214	Antigravity
1709	Teleportation
1737	Time Travel

Experiment

ProjectId	ExperimentId	NumInvolved	ExperimentDate	Hours
1214	1	1	NULL	1.5
1214	2	1	1889-11-01	14.3
1709	1	3	1891-01-22	7.0
1709	2	1	1891-02-23	7.2
1737	1	1	1900-07-05	-1.0
1737	2	2	1900-07-05	-1.5

Involved

ProjectId	ExperimentId	InvolvedId	Login
1214	1	1	mlom
1214	2	1	mlom
1709	1	1	dmitri
1709	1	2	skol
1709	1	3	ivan
1709	2	1	mlom
1737	1	1	skol
1737	2	1	skol
1737	2	2	ivan

~ 2006

(Software Carpentry Website)

Join

Users

Username	Name	Email
lpc	Luís	lpc@cmu
rita	Rita	rita@yahoo
...		

Content

Page	Username	Date
index	lpc	2008-12-12
slides	rita	2009-2-8
project	lpc	2009-3-20
...		

Join

Users

Username	Name	Email
lpc	Luís	lpc@cmu
rita	Rita	rita@yahoo
...		

Content

Page	Username	Date
index	lpc	2008-12-12
slides	rita	2009-2-8
project	lpc	2009-3-20
...		

Page	Username	Email
index	lpc	lpc@cmu.edu
slides	rita	rita@yahoo.com
project	lpc	lpc@cmu.edu
...		

DRY: Don't Repeat Yourself
a.k.a. Single Point of Truth

SQL: The language of Databases

SQL: Structured Query Language.
The standard way to access databases.

(At least for some meanings of the word standard).

SQL: Example

```
CREATE TABLE users (
    username VARCHAR(255) PRIMARY KEY,
    user VARCHAR(255) NOT NULL,
    email VARCHAR(255)
    );
```

Create table

`CREATE` *tablename* (
name type modifiers,
...);

Selects

```
SELECT pages.page, users.username, users.email
FROM pages, users
WHERE pages.username = users.username;
```

Users

Username	**Name**	**Email**
lpc	Luís	lpc@cmu
rita	Rita	rita@yahoo
...		

Content

Page	**Username**	**Date**
index	lpc	2008-12-12
slides	rita	2009-2-8
project	lpc	2009-3-20
...		

Users

Username	Name	Email
lpc	Luís	lpc@cmu
rita	Rita	rita@yahoo
...		

Content

Page	Username	Date
index	lpc	2008-12-12
slides	rita	2009-2-8
project	lpc	2009-3-20
...		

```
SELECT *
FROM users, pages
```

Users

Username	Name	Email
lpc	Luís	lpc@cmu
rita	Rita	rita@yahoo
...		

Content

Page	Username	Date
index	lpc	2008-12-12
slides	rita	2009-2-8
project	lpc	2009-3-20
...		

```
SELECT *
FROM users, pages
```

users.U	users.N	users.E	page.P	page.U	pages.D
lpc	Luís	lpc@cmu	index	lpc	2008-12-12
lpc	Luís	lpc@cmu	slides	rita	2009-2-8
lpc	Luís	lpc@cmu	project	lpc	2009-3-20
rita	Rita	rita@yahoo	index	lpc	2008-12-12
rita	Rita	rita@yahoo	slides	rita	2009-2-8
rita	Rita	rita@yahoo	project	lpc	2009-3-20
...					

Users

Username	Name	Email
lpc	Luís	lpc@cmu
rita	Rita	rita@yahoo
...		

Content

Page	Username	Date
index	lpc	2008-12-12
slides	rita	2009-2-8
project	lpc	2009-3-20
...		

```
SELECT *
FROM users, pages
WHERE users.username = pages.username;
```

users.U	users.N	users.E	page.P	page.U	pages.D
lpc	Luís	lpc@cmu	index	lpc	2008-12-12
lpc	Luís	lpc@cmu	project	lpc	2009-3-20
rita	Rita	rita@yahoo	slides	rita	2009-2-8
...					

In Practice

SQL is a declarative language:
you declare what you want not how to do it.

Select Statement

`SELECT` * or *column names*
`FROM` *dabasases*
`WHERE` *conditions*

Conditions

```
SELECT COUNT(*)
FROM users, pages
WHERE users.username = pages.username AND
    date > DATE('2008-12-31')
    ;
```

What's a NULL?

A NULL is a special object that represents missing.

Foreign Key

We had a *Page.Username* which is foreign key into Users.

```
CREATE TABLE Pages (
    Name VARCHAR(255) NOT NULL,
    Username VARCHAR(255) NOT NULL FOREIGN KEY(Users),
    ...);
```

Other Commands: INSERT

```
INSERT INTO pages
    VALUES('new-page','lpc',DATE('2009-3-31'));
```

Other Commands: UPDATE

```
UPDATE pages
    SET username = 'rita'
    WHERE name = 'index'
```

Other Commands: DELETE

`DELETE FROM` *table*
`WHERE` *condition*;

What We'll Talk About on Thursday

- How to set up a simple database (sqlite)
- Non-traditional databases (pytables)