

# Databases for Large Amounts of Data

Luis Pedro Coelho

§

On the web: <http://luispedro.org>

On twitter: @luispedrocoelho

European Molecular Biology Laboratory

April 4, 2014



Can handle large amounts of data (thousands or millions of records).

## Traditional

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

## Non-traditional

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

## Trendy (NoSQL)

- MongoDB
- CouchDB
- ...

- 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)

# SQL: The language of Databases

**SQL**: Structured Query Language.

The standard way to access databases.

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

```
CREATE TABLE Experiment(  
    ProjectId    INTEGER,  
    ExperimentId INTEGER,  
    NumInvolved  INTEGER,  
    ExperimentDate DATE,  
    Hours        REAL  
);
```

# Create table

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



```
SELECT NumInvolved  
FROM Experiment  
WHERE Hours > 2
```

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

# Select Statement

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

We have a `Experiment.ProjectId` which is foreign key into `Project`!  
We need to join these.

```
SELECT ProjectName
FROM Experiment, Project
WHERE Experiment.ProjectID = Project.ProjectID;
```

We have a `Experiment.ProjectId` which is foreign key into `Project`!  
We need to join these.

```
SELECT UNIQUE ProjectName
FROM Experiment, Project
WHERE Experiment.ProjectID = Project.ProjectID;
```

## Other Commands: INSERT

```
INSERT INTO Experiment  
VALUES(21, 23, 1, DATE('2014-04-04'), 1.4);
```

## You can also access sqlite3 DBs from Python

```
import sqlite3

connection = sqlite3.connect('experiments.db')
cursor = connection.cursor()
for entry in cursor.execute('SELECT * FROM Experiment'):
    print entry
```