Relational Databases and SQL

Professor Widom’s Instructional Odyssey

www.professorwidom.org

Stanford University

Google

ib

Instabase

Google Cloud Platform

acm Association for Computing Machinery

Very Large Data Bases Endowment Inc.

amazon web services™

Relational Database Management Systems

- Around for more than 40 years
- $50+ billion industry
- No sign of slowing down
- Why so successful?
  - Simple data model
  - High-level expressive query language
  - Reliable systems
  - Scalable systems

Popular RDBMSs

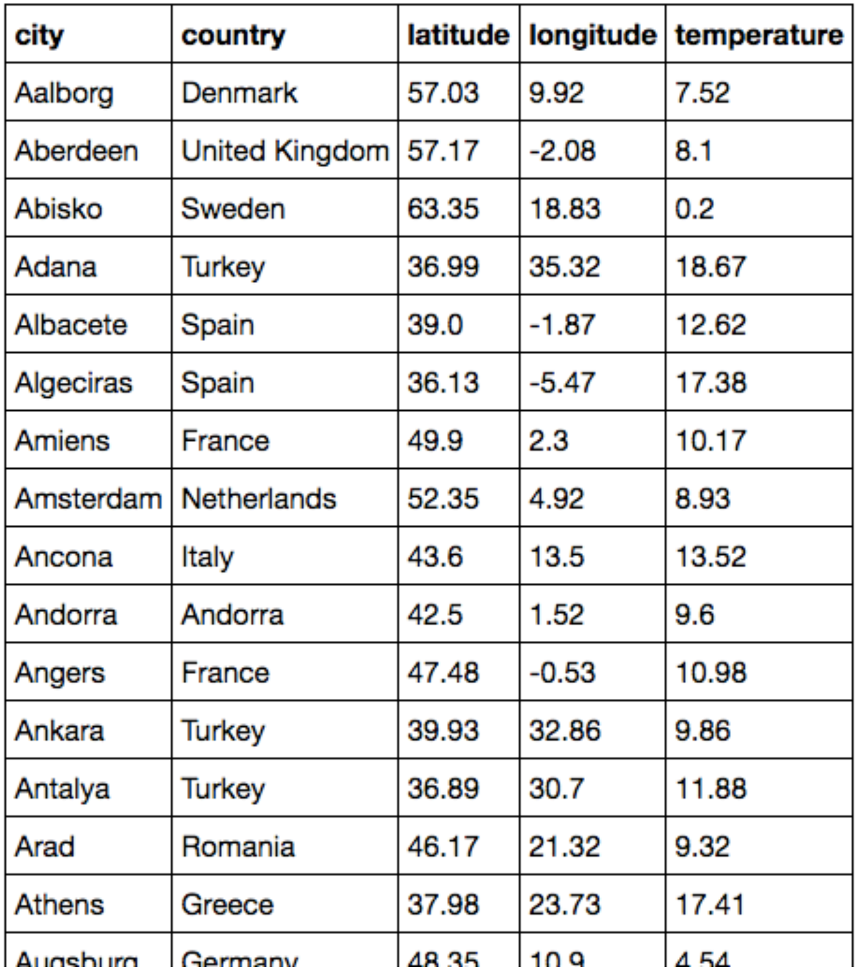
- Commercial proprietary systems
  - Oracle
  - Microsoft SQL Server
  - IBM DB2
  - Others …
- Open-source systems
  - MySQL
  - SQLite
  - PostgreSQL
  - Others …

Some general-purpose languages and platforms now support a subset of SQL (e.g., Excel, Spark)

Basic Concepts

- Relation (table)
- Attribute (column)
- Tuple (row)

Cities

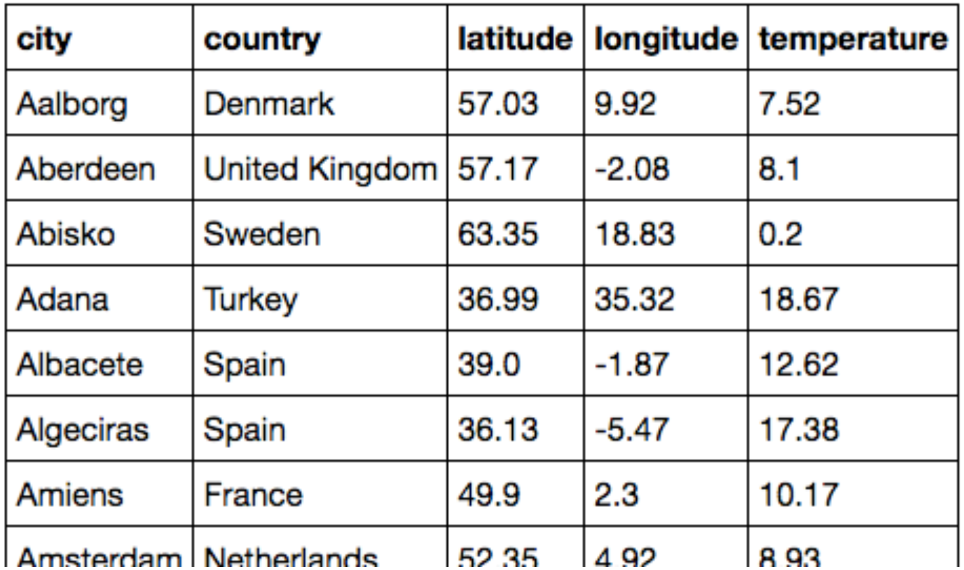


city	country	latitude	longitude	temperature
Aalborg	Denmark	57.03	9.92	7.52
Aberdeen	United Kingdom	57.17	-2.08	8.1
Abisko	Sweden	63.35	18.83	0.2
Adana	Turkey	36.99	35.32	18.67
Albacete	Spain	39.0	-1.87	12.62
Algeciras	Spain	36.13	-5.47	17.38
Amiens	France	49.9	2.3	10.17
Amsterdam	Netherlands	52.35	4.92	8.93
Ancona	Italy	43.6	13.5	13.52
Andorra	Andorra	42.5	1.52	9.6
Angers	France	47.48	-0.53	10.98
Ankara	Turkey	39.93	32.86	9.86
Antalya	Turkey	36.89	30.7	11.88
Arad	Romania	46.17	21.32	9.32
Athens	Greece	37.98	23.73	17.41
Augsburg	Germany	48.35	10.9	4.54

Differences Between Table and Spreadsheet

- Name is significant
- Order is not significant – can change on re-open
- Regular structure, more “row-oriented”
- Data elements always values, not formulas

Cities



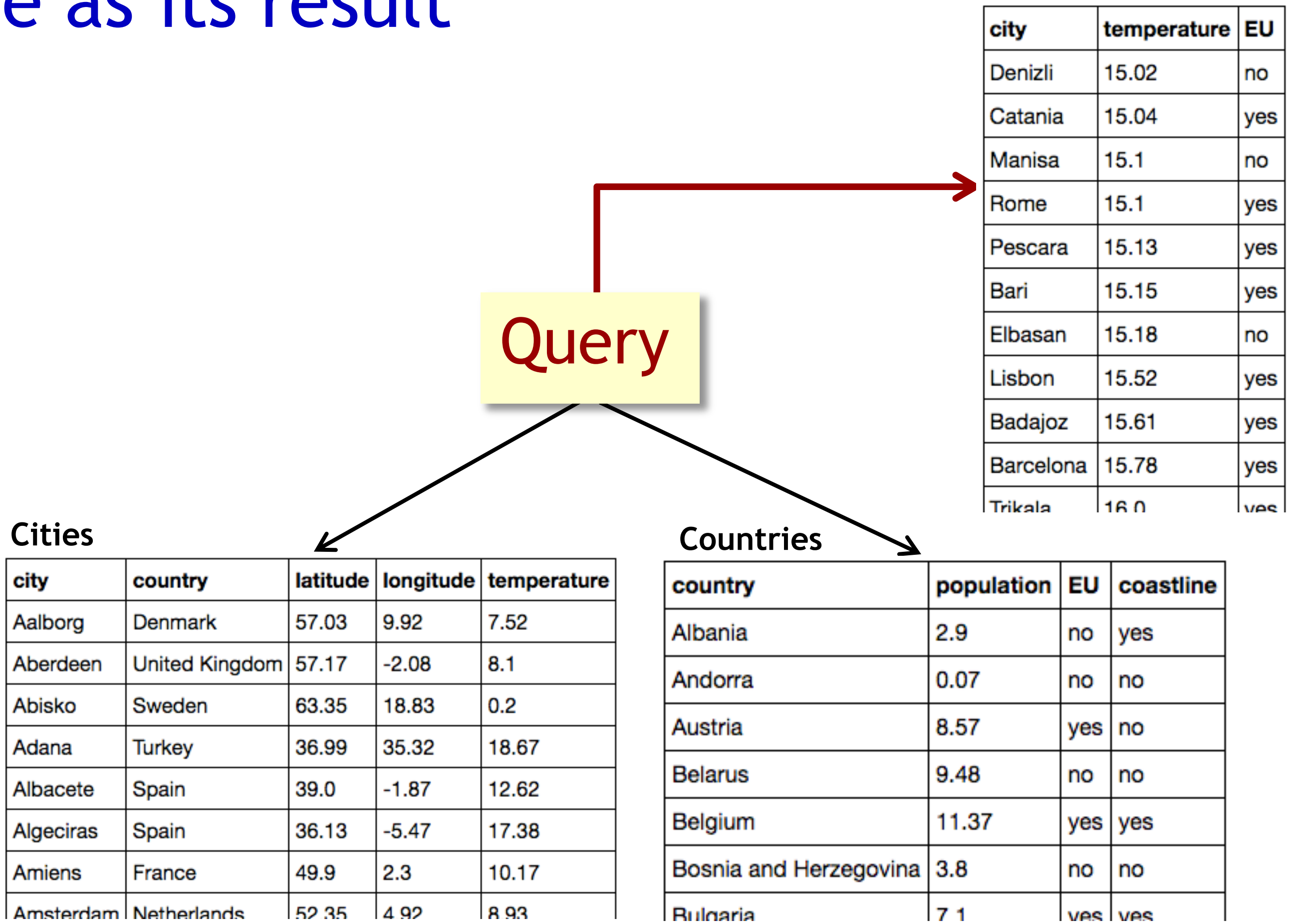
city	country	latitude	longitude	temperature
Aalborg	Denmark	57.03	9.92	7.52
Aberdeen	United Kingdom	57.17	-2.08	8.1
Abisko	Sweden	63.35	18.83	0.2
Adana	Turkey	36.99	35.32	18.67
Albacete	Spain	39.0	-1.87	12.62
Algeciras	Spain	36.13	-5.47	17.38
Amiens	France	49.9	2.3	10.17
Amsterdam	Netherlands	52.35	4.92	8.93

Creating and Loading Data

System-dependent, but can nearly always start with CSV file or similar

Querying

Query executed over one or more tables, returns table as its result



Query

city	temperature	EU
Denizli	15.02	no
Catania	15.04	yes
Manisa	15.1	no
Rome	15.1	yes
Pescara	15.13	yes
Bari	15.15	yes
Elbasan	15.18	no
Lisbon	15.52	yes
Badajoz	15.61	yes
Barcelona	15.78	yes
Trikala	16.0	yes

Cities

city	country	latitude	longitude	temperature
Aalborg	Denmark	57.03	9.92	7.52
Aberdeen	United Kingdom	57.17	-2.08	8.1
Abisko	Sweden	63.35	18.83	0.2
Adana	Turkey	36.99	35.32	18.67
Albacete	Spain	39.0	-1.87	12.62
Algeciras	Spain	36.13	-5.47	17.38
Amiens	France	49.9	2.3	10.17
Amsterdam	Netherlands	52.35	4.92	8.93

Countries

country	population	EU	coastline
Albania	2.9	no	yes
Andorra	0.07	no	no
Austria	8.57	yes	no
Belarus	9.48	no	no
Belgium	11.37	yes	yes
Bosnia and Herzegovina	3.8	no	no
Bulgaria	7.1	yes	yes

Querying

Query executed over one or more tables, returns table as its result

- Find all cities with temperature between 15 and 25, return city and country
- Find average city temperature for each country
- Find all cities in countries that are in the EU but don’t have coastline, return city and country
- Find all pairs of cities that are close together, i.e., longitude and latitude are less than 0.5 apart
- Find the westernmost city

The SQL Language

- Also more than 40 years old

  One of oldest languages still in use (others?)

- Supported by all RDBMSs, standardized across products

  More or less …

- Interactive or embedded in programs

- Also can be used to modify the database

Databases

Europe Temperatures

Cities (city, country, latitude, longitude, temperature)

Countries (country, population, EU, coastline)

2010 World Cup

Teams (team, ranking, games, wins, draws, losses, goalsFor, goalsAgainst, yellowCards, redCards)

Players (surname, team, position, minutes, shots, passes, tackles, saves)

Titanic

Titanic (last, first, gender, age, class, fare, embarked, survived)

Jupyter Notebooks

(formerly iPython notebooks)

- Modeled after “laboratory notebooks”
- In one notebook can combine text boxes with boxes containing executable code in a wide variety of languages
- Can run/re-run boxes (cells) individually, or run/re-run entire notebook

Rapid adoption in many sectors

Jupyter Notebooks

- Can download to your computer (recommend *Anaconda*) but no one-button download yet
- We will use notebooks in the cloud, via *Google Colab*
- Either way, notebooks run in a web browser

To execute a code cell, click inside the box then click ▶.

Or use *shift*, *control*, or *command* with *enter* or *return*

Agenda: Basic SQL

(Creating and populating tables)

1. Basic SELECT statement
2. Ordering
3. Joins
4. Basic aggregation
5. Limit clause

Plenty of your turn!

For help while working with SQL:

Tutorials and help pages
(website Course Materials)

➢ Web search

Agenda: Advanced SQL

1. Duplicates
2. Table variables
3. Subqueries of all types
4. Advanced aggregation
5. Data modification

Plenty of your turn!

Some Features Not Covered

- Set operators: Union, Intersect, Except
- Keys
  - Designated column that must have unique value in each row
  - Or designated set of columns
- Null values
  - Special value usually denoting unknown or undefined
  - Not included in aggregations, =, <, etc.
  - Example: … where temp <= 10 or temp > 10
- Outer joins

Relational Databases and SQL

Professor Widom's Instructional Odyssey

www.professorwidom.org

Stanford University

Google

Association for Computing Machinery

Very Large Data Bases Endowment Inc.

INSTABASE

Google Cloud Platform

amazon web services™