

An Introduction to Databases

Christopher G. Wilson, Ph.D.

Loma Linda University
Dept. of Pediatrics and Center for Perinatal Biology

May 21, 2015

Outline

1. Review
2. R vs Python
3. Databases
4. Structured Query Language
5. NOSQL and NewSQL
6. SourceTree Demo

R is ...

- ▶ A comprehensive and expandable statistical environment and programming language
- ▶ Has powerful graphics
- ▶ Most widely used language in bioinformatics
- ▶ Is standard for data mining and biostatistical analysis
- ▶ Did everyone get through Holly's tutorial??

R versus Python: head-to-head

http://blog.datacamp.com/r-or-python-for-data-analysis/?imm_mid=0d2357&cmp=em-data-na-na-newsltr_20150520

What is a database?

- ▶ A *database* is an organized collection of related records (files, data, etc.)
- ▶ Records previously stored in separate files can be organized in a common pool of data
- ▶ A *Database Management System* (DBMS) is a software application that allows you to create, define, query, update, and administer a database
- ▶ Commonly used DBMSs include: *MySQL, PostgreSQL, Oracle, Sybase, IBM DB2, etc.*

A brief history of DBMSs

- ▶ **1960s** The introduction of the term database coincided with the availability of direct-access storage (disks and drums) from the mid-1960s onwards. The Oxford English dictionary cites a 1962 report by *System Development Corporation* as the first to use the term "data-base" in a specific technical sense.
- ▶ **1970s** Edgar Codd (IBM) was primarily involved hard disk development. He was unhappy with available DBMSs—notably the lack of a "search" facility—he wrote a groundbreaking paper "A Relational Model of Data for Large Shared Data Banks" that was the foundation for Structured Query Language (SQL)

A brief history of DBMSs

- ▶ **1980s** The age of desktop computing—new computers empowered their users with spreadsheets like *Lotus 1-2-3* and database software like *dBASE*, which was lightweight and easy to understand
- ▶ **1990s to present** Object-oriented programming alters the landscape for database interfaces. Programmers and designers began to treat the data in their databases as objects. This allows for relations between data to be relations to objects and their attributes and not to individual fields.

SQL

- ▶ *Structured Query Language (SQL)* is a special-purpose programming language designed for managing data in relational database management systems (RDBMS)
- ▶ Originally based upon relational algebra and tuple relational calculus, SQL consists of a data definition language and a data manipulation language
- ▶ The scope of SQL includes data insert, query, update and delete, schema creation and modification, and data access control
- ▶ SQL became an American National Standards Institute (ANSI) in 1986, and of the International Organization for Standardization (ISO) in 1987

SQL (cont.)

- ▶ Despite the existence of such standards, though, most SQL code is not completely portable among different database systems without changes
- ▶ *Structured Query Language (SQL)* is a special-purpose programming language designed for managing data in relational database management systems (RDBMS)
- ▶ Originally based upon relational algebra and tuple relational calculus, SQL consists of a data definition language and a data manipulation language
- ▶ The scope of SQL includes data insert, query, update and delete, schema creation and modification, and data access control

Post-relational Databases

- ▶ XML databases are a type of structured document-oriented database that allows querying based on XML document attributes
- ▶ XML databases are mostly used in enterprise database management, where XML is being used as the machine-to-machine data interoperability standard.

NewSQL

- ▶ *NewSQL* is a class of modern relational databases that aims to provide the same scalable performance of NoSQL systems for online transaction processing (read-write) workloads while still using SQL
- ▶ These databases include *ScaleBase*, *Clustrix*, *EnterpriseDB*, *MemSQL*, *NuoDB*, *VoltDB*, etc.

Some YouTube links

[https : //www.youtube.com/watch?v = FR4QleZaPeM](https://www.youtube.com/watch?v=FR4QleZaPeM)

[https : //www.youtube.com/watch?v = 4Z9KEBexzcM](https://www.youtube.com/watch?v=4Z9KEBexzcM)

[https : //www.youtube.com/watch?v = ql_g07C_Q5I](https://www.youtube.com/watch?v=ql_g07C_Q5I)

Today

- ▶ Jon will demo *SourceTree* for us
- ▶ Time to start on the next project. . .