Introduction

Understanding Databases

Will Doyle

Databases in Data Science

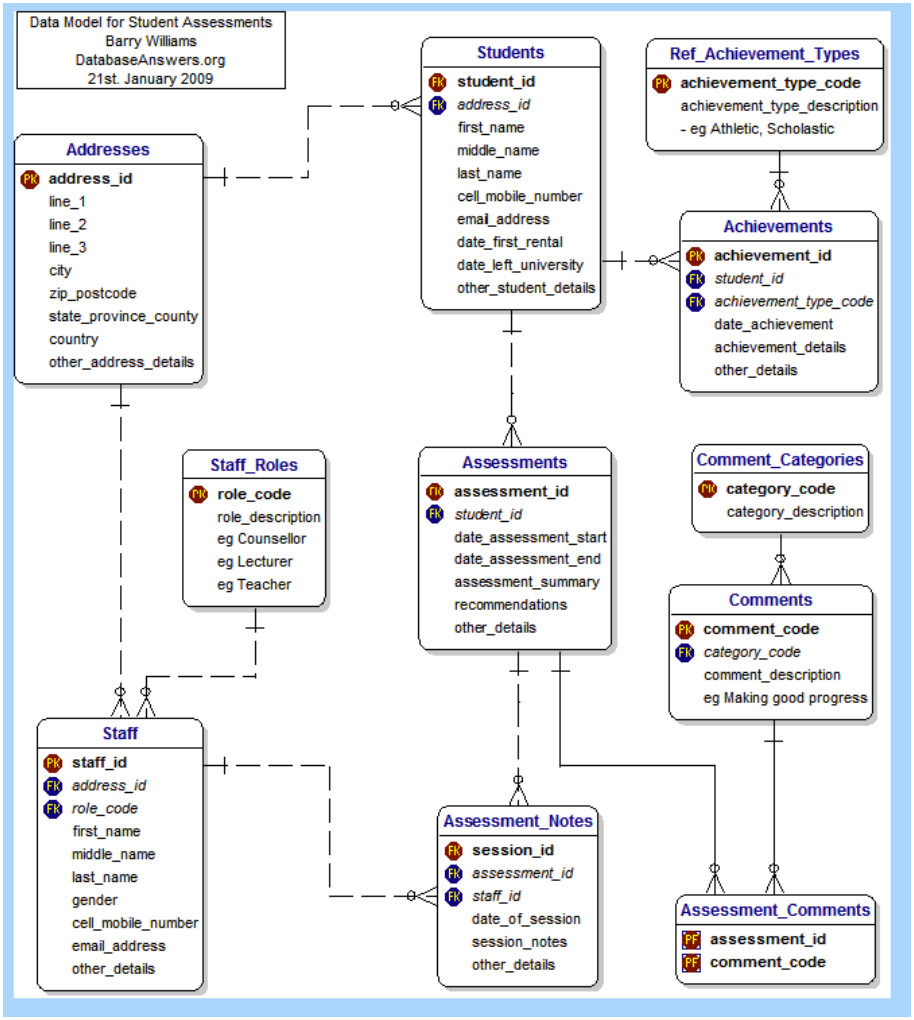
- Databases are the primary means of entering and storing data in any reasonably sized enterprise.
- Design, maintenance, and revision of databases is **well** beyond the scope of this course.
- A few concepts are quite helpful in generating datasets.
- Being able to communicate with database analysts is key.

Database Terms

- Record: set of fields tied to an identifier (think a row in a dataset)
- Table: set of related records
- Field: characteristics of a record
- Query: a request sent to a database
- Identifier: a field in one table that may be used to tie records together across tables

So, a **database** is a set of related **tables,** each of which contains **records** made of multiple **fields** that have **identifiers**

Example Database Schema



Source: http://www.databaseanswers.org/data_models/student_assessments_simple/index.htm

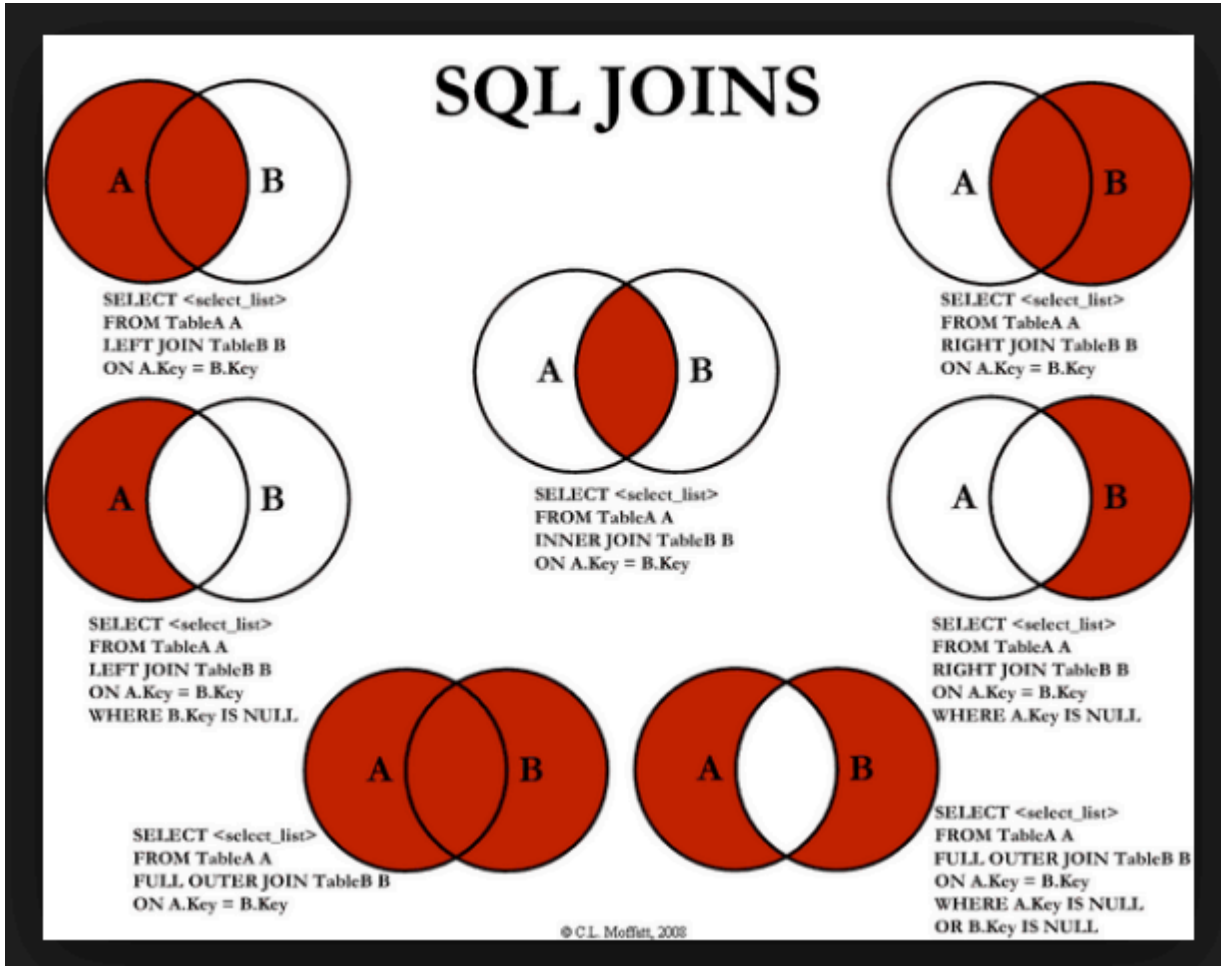
Database Management Systems (DBMS)

- A means by which to organize, maintain, and query tables
- Examples include SQL, MySQL, Oracle, MS Access, and so on
- You do not need to have access to a DBMS to access a database
- The DBMS does imply what kind of operations you may be able to do

Database Verbs

- **Select:** Choose certain fields from a table.
- **From:** Utilize certain tables.
- **Where:** Only select records that meet certain conditions.
- **Join:** Combine records from certain tables.

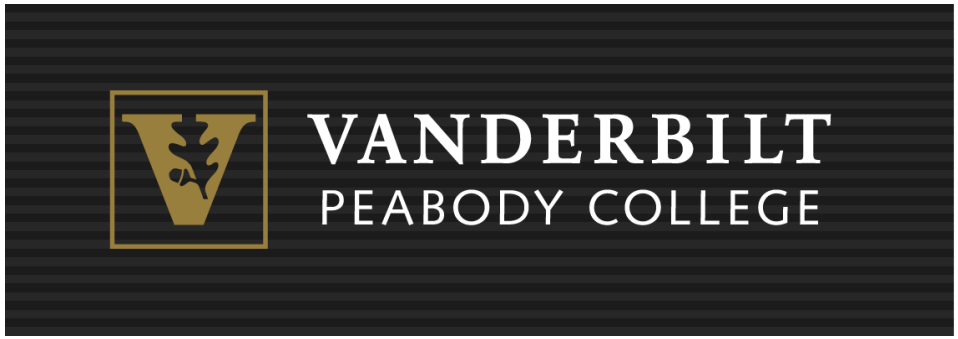
Types of Joins



Source: https://plus.google.com/111053008130113715119/posts/5TdpUKQXxpu

Types of Joins

- 99% of the time, it's best to structure your setup so that you're doing a left join of type 1.
- Other joins get complicated fast.
- If you need to do something more difficult, talk it through with the database analyst.


VANDERBILT
PEABODY COLLEGE