



CS196

Data Science

Brief Foray into SQL

Week 9 of Data Hackerspace

What is SQL?

PostgreSQL



MySQL®



Microsoft®
SQL Server®

- “Structured Query Language”
- Query language for relational databases
 - Strict schemas
 - “Tables” to hold structured data
- Declarative in nature
 - Instruct *what* to do, but let the system decide *how* to do it
 - Allows for multiple implementations (MySQL, PostgreSQL, etc.)
 - Allows for database level optimizations

What does SQL look like in practice?

- Databases contain Tables which contain Columns

- Table Creation

```
CREATE TABLE Student (  
    StudentID int,  
    LastName varchar(255),  
    FirstName varchar(255),  
    School varchar(255)  
);
```

- Table Insertion

```
INSERT INTO People (1, "Congdon", "Benjamin", "UIUC");
```

- Table Queries

```
SELECT * FROM People WHERE FirstName="Ben";
```

Why do we use SQL?

- What happens when we can't fit all of our data in memory?
 - Solution for "Medium Data" problems
- Structured data is (generally) cleaner data
- Note: "Big Data" outscales traditional SQL servers. Other data processing platforms like Hadoop, Spark, etc. are popular for these types of applications

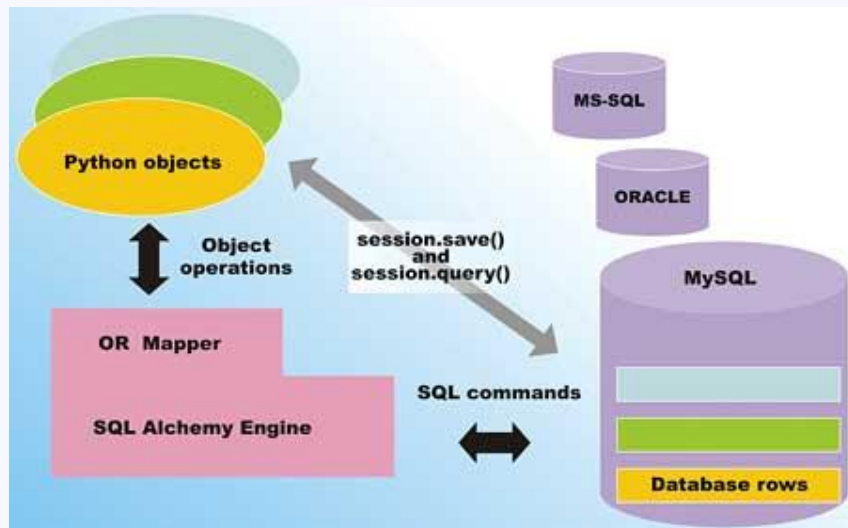
How do we use SQL?

- Language Specific Drivers
 - Usually database specific
 - MySQL ⇨ PyMySQL
 - PostgreSQL ⇨ Psycopg2
- Raw SQL Queries
 - Powerful/Flexible
 - Failure Prone (Application not schema aware)

How do we use SQL?

SQLAlchemy

- Object Relational Mappers (ORMs)
 - Abstraction layer between database and application
 - Automatically creates tables, constructs and executes queries, etc.



conda install -c anaconda sqlalchemy



