



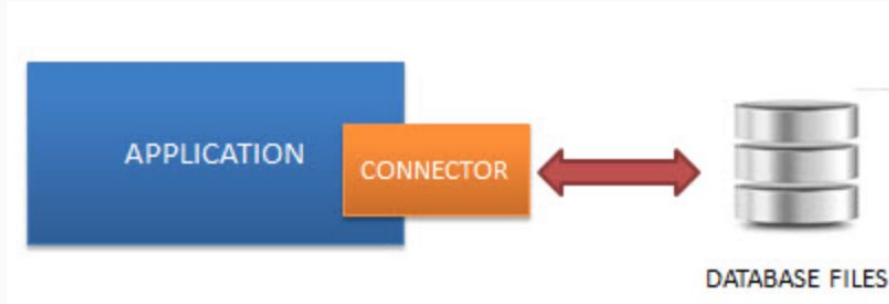
NYU

Center for  
Data Science

# Lab 1: RDBMS

# What is SQLite?

- SQLite is a “lightweight” version of SQL
- SQLite is serverless: DB is stored in a single file
- Transactions in SQLite are fully ACID-compliant
  - Atomic, Consistent, Isolated, and Durable.



Reference: <http://www.sqlitetutorial.net>

# What is the difference and how is it better?

- SQLite does not need to be "configured" as a server like MySQL.
- Serverless- no separate server process is needed
- Transactional SQL database engine
- Stable Cross-Platform Database File
  - can be written/ copied on different machines with different architecture

# Basic Commands

Attach Databases

Create Table, Insert Into the table

Retrieve Information (SELECT)

Filter data (WHERE, LIMIT, BETWEEN, IN)

Group By Data with HAVING clause

Sort Data ASC/DESC

Joins (LEFT, OUTER, INNER, CROSS)

Alter Table, Update Table

Drop Table, Delete Table

Aggregations

Transactions

## Order of Execution

1. From (Join + On)
2. Where
3. Group By
4. Having
5. Select
6. Order By
7. Limit



# Python and SQLite

To perform any operation on a SQLite database via Python's sqlite3 module

A connection needs to be opened to a SQLite database file

Then you can **execute()** queries through the connection object

```
import sqlite3
conn = sqlite3.connect('music-small.db')

t = (1990,) # note that this is a tuple!

for row in conn.execute("""SELECT count(*) FROM artist
                           WHERE artist_active_year_begin=?""", t):
    print(row)
```

# Do's and Don'ts

**Don't use this as this is insecure.**

```
value = '1998'  
conn.execute("SELECT * FROM tracks WHERE year = '%s' " % value)
```

**Always do this:**

```
value = (1998,)   
conn.execute("SELECT * FROM tracks WHERE year =?", value)
```

# Lab 1 materials

- Just like lab 0: accept the github classroom invitation
- A data file (music-small.db) is available on Brightspace
  - Content → Relational databases → music-small.db
  - Download this to your machine in the same folder as your code
  - **Do not add it to git!**
- Two parts:
  - part 1: write queries to answer different questions
  - part 2: optimize the DB for an existing query

# Things to keep in mind

- Use sqlitebrowser to explore the data
- Keep a “clean” backup copy of the database in case you change things
- Commit your partial results as you go.
- Don't forget to update RESULTS.md
- Don't forget to push to github!