

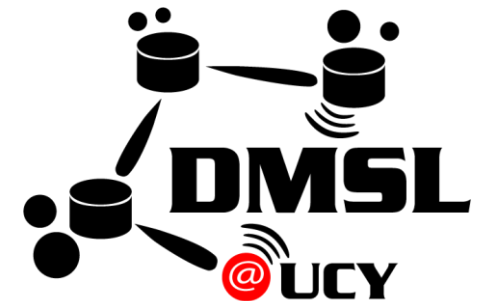
EPL646 – Advanced Topics in Databases  
**ChromaDB**  
&  
**Overview of Transaction  
Management**

Christoforos Panayiotou

<http://www.cs.ucy.ac.cy/~dzeina/courses/epl646/labs/lab.html>



University  
of Cyprus



# Chroma DB installation

- On your machine
  - You will need a full installation of Python
    - Cannot use the portable installation shown in Lab 2
  - Go to <https://python.org>
    - Select the appropriate version for your system, download and install it
  - You will need the Microsoft C++ Build Tools
    - You can download them from: <https://visualstudio.microsoft.com/visual-cpp-build-tools/>
    - It will need around 7GB of space!
    - Check this link for instructions: <https://learn.microsoft.com/en-us/answers/questions/136595/error-microsoft-visual-c-14-0-or-greater-is-require>

# Chroma DB installation

- In our linux labs
  - Create a new virtual environment (**venv**) for python
    - Use the command: `python3 -m venv epl646env`
    - `epl646env` is the directory in which the virtual environment will be created
  - Start the **venv**
    - Use the command: `source epl646env/bin/activate`
    - To close the **venv** use the command: `deactivate`
  - Install Pandas, Dask and Parquet
    - Use the command: `pip install pandas dask pyarrow`
  - Install chromaDB
    - Use the command: `pip install chromadb`

# Fixing the inbstallation

- Our labs have an incompatible version of **SQLite** for **Chroma DB** so we need to fix this
  - Install the `pysqlite3-binary` with the command: `pip install pysqlite3-binary`
  - Then in any python program you want to use **Chroma DB** add the following lines at the top:

```
__import__ ('pysqlite3')
import sys
sys.modules['sqlite3'] = sys.modules.pop('pysqlite3')
```

# Practice

- Run the code found in slide 14 of lecture 5b
- Do the same for the code in slide 17 of lecture 5b

- You must change the line:

```
client =  
chromadb.Client(Settings(chroma_db_impl="duckdb+parquet", persist_directory=".chroma/my-db"))
```

- This is deprecated and no longer works!

- Instead use:

```
client =  
chromadb.PersistentClient(path=".chroma_db")
```

- Also use function `get_or_create_collection()` instead of `create_collection()`

- Do not forget to add the fix for SQLite!

## Exercise 16.1

**(Exercise 16.1)** Give brief answers to the following questions:

- a. What is a transaction? In what ways is it different from an ordinary program (in a language such as C)?
- b. Define these terms: atomicity, consistency, isolation, durability, schedule, blind write, dirty read, unrepeatable read, serializable schedule, recoverable schedule.

## Exercise 16.2

- **(Exercise 16.2)** Consider the following actions taken by transaction  $T_1$  on database objects  $X$  and  $Y$  :

**$R(X), W(X), R(Y), W(Y)$**

- a. Give an example of another transaction  $T_2$  that, if run concurrently to transaction  $T_1$  with-out some form of concurrency control, could interfere with  $T_1$ .

## Exercise 16.3

**(Exercise 16.3)** Consider a database with objects  $X$  and  $Y$  and assume that there are two transactions  $T_1$  and  $T_2$ . Transaction  $T_1$  reads objects  $X$  and  $Y$  and then writes object  $X$ . Transaction  $T_2$  reads objects  $X$  and  $Y$  and then writes objects  $X$  and  $Y$ .

- Give an example schedule with actions of transactions  $T_1$  and  $T_2$  on objects  $X$  and  $Y$  that results in a write-read conflict.
- Give an example schedule with actions of transactions  $T_1$  and  $T_2$  on objects  $X$  and  $Y$  that results in a read-write conflict.
- Give an example schedule with actions of transactions  $T_1$  and  $T_2$  on objects  $X$  and  $Y$  that results in a write-write conflict.



# Questions?

<http://www.cs.ucy.ac.cy/~dzeina/courses/epl646/labs/lab.html>

