

# Homework 0: Environment Setup

*COMS W4111: Introduction to Databases*  
*Sections 002, V02*  
*Fall 2022*

## Introduction and Overview

*Homework 0* provides step-by-step instructions on how to install and set up necessary tools this course will require. We start the semester with a HW 0 environment setup prior to HW 1. This allows us to resolve configuration and set up issues before students begin working on graded assignments.

Tests and screenshots are required to show that the environment was set up successfully. Submission for this assignment is the accompanying W4111\_HW0.ipynb file with completed tests and screenshots converted to PDF format.

This assignment is for both Programming and Non-Programming sections. All steps will be required for both tracks unless specifically marked for Programming only.

A points earned score between 0 and 100 determines the final grade for the semester. We will explain the grading, HW and exams in the first lecture. HW 0 is not worth any points, i.e. the points earned for submission is 0. Failing to submit HW 0 **on time without using late days will result in a 5 points deduction from HW 1**, however.

**Due date: September 18, 11:59 PM EDT on GradeScope**

Please note: You may NOT use late days for the submission of this assignment. Check Courseworks for GradeScope access.

The following tools will need to be installed and or setup:

1. Anaconda/Python 3/Jupyter Notebook
2. PyCharm (Programming Track only, but strongly recommended for all students).
3. MySQL Server Community Edition.
4. DataGrip

In the zip file accompanying this assignment you will find:

- HW0 Test folder
- W4111\_HW0.ipynb (what you will submit as a pdf)
- DDL.sql
- smallRelationsInsertFile.sql

- lahman-mysql-dump.sql

## Anaconda/Python 3/Jupyter Notebook

Follow the online instructions to download the Anaconda Individual Edition [here](#). It should default to installing Python 3.9. Choose the option to “install only for me.”

Once installation is complete:

1. Launch/start Anaconda Navigator
2. Select and Launch Jupyter Notebook

A file system/folder navigator page will appear in the browser. Navigate to the folder/directory containing the HW 0 files. You created the folder/directory either by cloning the class repository and making a copy of the relevant folder, or by downloading a zip file and unzipping.

Click on the link W4111\_HW0.ipynb. This will launch the Jupyter Notebook in a new browser window/tab. Switch to the notebook in the browser. Follow the instructions for the sections:

- Add Student Information
- Testing Anaconda

## PyCharm (Programming Track only, recommended for all)

Next, download the professional version of PyCharm [here](#). You need to [apply](#) for an education license. Please install PyCharm Professional. You can install the trial version and apply the free license later.

PyCharm requires configuration for each project. In the test for this section you will practice configuring a Python Environment.

Start PyCharm, and in PyCharm, open the HW0 Test subfolder of the homework folder, it will open up a new project window. Click on the HW0PyCharmTest.py file. The Anaconda installation should make Python 3.8 the default interpreter for your projects. If you get a message about missing an interpreter or cannot find the run option, please follow [online instructions](#) to configure a Python interpreter for the project.

Follow the instructions and replace your uni in the Python file where specified. Right click on the HW0PyCharmTest.py file, either on the tab or anywhere in the file's body. Click “Run HW0PyCharmTest” to execute. Take a screenshot of your window and embed it into the test ipynb file provided where specified. There is a Markdown cell. You just need to replace the name of the file in Image("./PyCharmScreenshot.png").

If your file will not run because it is missing a configuration or interpreter follow the instructions below, otherwise move onto the next section.

To configure the interpreter, click “Add Configuration” in the top right corner then the “+” button. Scroll down and select Python. You can name this configuration how you choose, I suggest naming it the Python version you use. Select the appropriate Python version in the Python Interpreter field then click OK.

## MySQL Server Community Edition

You must install MySQL (Server) Community. There is a [download link](#) and you can select your operating system. The installation is relatively easy. There are

- [Instruction for Windows](#). Use the MySQL Installer method and choose the Developer Default.
- [Instructions for Mac](#).
- [Linux](#).

At some point, you will be promoted for/have to set login/authentication options.

- Write down and remember the *root* user ID and *password*. You are installing on a local machine with no sensitive data. A simple password is OK. I recommend the password *dbuserbdbuser*. Using a common password solves problems due to your forgetting your password.
- **Choose the Legacy Authentication method.**

Installing MySQL registers MySQL Server as a service. It should start automatically. If you are ever unsure if MySQL Server is running, there are online OS specific instructions for determining status, starting and stopping the server.

## DataGrip

DataGrip will be the GUI (Graphical User Interface) and tool that allows you to visualize and edit data table definitions and data on your MySQL server. You need to [apply](#) for an education license from JetBrains if you didn't already do so in the PyCharm setup.

Download [DataGrip](#) and setup. You will need the education license so you can use DataGrip beyond your free trial period. You can start with the trial version of professional and apply the credit later.

You will need to set up a connection in DataGrip to your local MySQL Sever. There are [online instructions](#) for configuring a connection. In addition to the general instructions, there are [MySQL specific](#) instructions.

Once DataGrip is installed, you need to install two databases. You do this by running 3 SQL scripts that are in the HW 0 folder:

- DDL.sql
- smallRelationsInsertFile.sql
- lahmans-mysql-dump.sql

You will run the SQL scripts in that order. The DataGrip documentation [provides instructions](#) on how to run an SQL script from a file (hard drive). Follow the instructions for each of the SQL files.

When your Lahman database has been successfully loaded, copy, paste, then run the following query in the console:

```
SELECT * FROM lahmansbaseballdb.People LIMIT 100;
```

Don't worry about what this query means for now, we will cover this in class.

You should get an output similar to the screenshot below. Take a screenshot of your window and insert your screenshot in the appropriate section of the .ipynb provided. You need to save the screenshot to a file in the HW 0 directory and replace the file name in the DataGrip section on the notebook.

Example screenshot:

The screenshot displays a database console window with a dark theme. The left sidebar shows a tree view of the database schema, including tables like 'AllstarFull', 'Appearances', 'AwardsManagers', 'AwardsPlayers', 'AwardsShareManagers', 'AwardsSharePlayers', 'batting', 'BattingPost', 'CollegePlaying', 'Fielding', 'FieldingOF', 'FieldingOFsplit', 'FieldingPost', 'HallOfFame', 'HomeGames', 'Managers', 'ManagersHalf', 'Parks', 'People', 'Pitching', and 'PitchingPost'. The main area shows a SQL query: `SELECT * FROM lahman2017raw.People LIMIT 100;`. Below the query, the results are displayed in a table with columns: `playerID`, `birthYear`, `birthMonth`, `birthDay`, and `bir`. The results show 100 rows of player data. The bottom panel shows the execution log with two entries: `[2021-01-09 13:38:36] 100 rows retrieved starting from 1 in 516 ms (execution: 149 ms, fetching: 367 ms)` and `[2021-01-09 14:14:48] 100 rows retrieved starting from 1 in 168 ms (execution: 124 ms, fetching: 44 ms)`. The status bar at the bottom indicates 'Connected (42 minutes ago)' and '1:1 (45 chars) CRLF UTF-8 4 spaces'.

playerID	birthYear	birthMonth	birthDay	bir
26 abernte02	1933	3	6	USA
27 abernwo01	1915	2	1	USA
28 aberscl01	1921	8	28	USA
29 ablesha01	1883	10	4	USA
30 abnersh01	1966	6	17	USA
31 abramca01	1924	3	2	USA
32 abramge01	1899	11	9	USA
33 abregjo01	1962	7	4	USA
34 abreubo01	1974	3	11	Venez
35 abreujo01	1913	5	24	USA
36 abreujo02	1987	1	29	Cuba

You will also need to run some code cells that connect your jupyter notebook to your Lahman database.