Data Management: Projects

Tutor

Riccardo Valentini

Academic Year 2021/2022
About myself..

Lectures:
• Today (18/05/2022): projects and graph databases (Neo4j);
• Next Thursday (26/05/2022): data warehousing (Google Data Studio);

Office Hours:
• On Wednesday from 10:00 AM to 12:00 PM, Room B221 (DIAG) or ONLINE (further details in next slide!)

Contact me:
On email: valentini@diag.uniroma1.it
Subject: [DM]: «anything you want»

Riccardo Valentini
Bachelor and Master Degree in Engineering in Computer Science (2020);

PhD Student at DIAG.
Office Hours rules

- **Appointments** (ONLY for office hours, **not project presentations**) will be scheduled in a dynamic way, through Google Calendar.

- You will be enabled to book a 1 or 2 slots (of 15 minutes each) from Google Calendar. Please, book 2 slots ONLY if you really have enough things to be discussed to fill the whole time, otherwise, **you could deny other colleagues the possibility of useful office hours**.

- **DO NOT** book **more than 2 slots**, otherwise I will remove **ALL** the slots you booked.

- You will always allow choosing online or in presence office hours. There may be a case in which I will manually set it to online (of course, I will warn you and provide you with the corresponding google meet address for the online meeting).

- **MAY 2022 (ONLINE ONLY, ONLY 1 SLOT per student / group):**
  https://calendar.google.com/calendar/u/0/appointments/schedules/AcZssZ27VLQx_0OsAK1kLWTBQlek2sbMuVqBsQ55O3BC7AXMWHT9r9355dZLoTJMxXFyPbhkxTk0tvD5i

- **JUNE-JULY 2022:**
  https://calendar.google.com/calendar/u/0/appointments/schedules/AcZssZ0FEpOw-cJwlc5v-UJzGD2wr32_RGi4RUyL5mCwIQ4mTCfVD_S6bqF6nL20hBefrSAV-kTNU1NP

*Please refer to those two links in case of changes in the schedules.*
Office Hours procedure

Riccardo Valentini
Office Hours
Appuntamenti di 15

Seleziona un orario per l'appuntamento

Giugno 2022

<table>
<thead>
<tr>
<th>D</th>
<th>L</th>
<th>M</th>
<th>M</th>
<th>G</th>
<th>V</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td>31</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
</tr>
</tbody>
</table>

MER
8

10:00AM
10:20AM
10:40AM
11:00AM
11:20AM
11:40AM

Sapienza email address.
Project general instructions

- Projects can be done either **individually** or in **groups of 2 students**;
- **2 options** for the project:
  1. **Option 1**: it is required to prepare a **report** *(at least 12, at most 20 pages)* about a certain topic;
  2. **Option 2**: it is required the usage of one (or more!) **data management/analytics tool** which you can select.

*Further details about both options will be given in the next slide.*

- Both options require a 10-15 minutes **slide presentation** about the assessed challenges/topic;
- Both options require an **approval before starting** to work on the project;
- Both options require **the material to be sent to ME** ([valentini@diag.uniroma1.it](mailto:valentini@diag.uniroma1.it)) and to fix an appointment (online or in presence) with either me or Prof. Lenzerini;

In order to have the **exam registered**, every student enrolled in the A.Y. 2021/22 must present a project in dedicated sessions with the professor or me:

This means you can either pass the written test and then present a project or present a project and then passing the written test.
Project topics

For the updated list of projects (with their category, option 1 or 2), please refer to the web pages of the course: https://www.diag.uniroma1.it/~lenzerin/index.html/?q=node/53
Project steps

To submit a project:

1. Select something which interested you among the topics discussed;

2. (OPTION 2 only) Select a relevant dataset, which could be suitable w.r.t. the chosen tool (at this point, some interesting dataset could be the occasion to think about your tool choice in case you want to change it/them);

3. Think about the structure of your project (discuss with your partner if you have one!) and prepare a one page pdf document indicating:
   a) The names and matricolae of the team members (OPTION 1 and 2);
   b) The topic you select to write the report (OPTION 1 only);
   c) The tool/tools you want to use (OPTION 2 only);
   d) The dataset you want to use (OPTION 2 only);
   e) The description (5-10 lines) of what you desire to develop (OPTION 2 only).

4. One of the team members has to send me (valentini@diag.uniroma1.it) an email with the subject «[DM]: project proposal», attaching the pdf described at point 3;

5. After my approval, you can start working on your project, feeling free to contact me to express doubts, simple questions or literally anything..

6. Once you feel you completed the project... (next slide!)
**Complete a project**

Once you feel you have completed the project, you can write an email with the subject «[DM]: project complete» explicitly telling me if you want an online or in presence appointment and attaching:

- **Both options**: a slide presentation containing the summary of your work and how you proceeded (step by step) to reach your results (length 10-15 minutes);
- **OPTION 1 only**: the report you wrote (from 12 to 20 pages);

I will reply with an appointment to present your work to either Prof. Lenzerini or me. At presentation time, you are required to do the following:

- The slide presentation you sent us before (length 10-15 minutes, few changes are acceptable, but no revolutions w.r.t. to the one attached to the mail);
- **OPTION 2 only**: A short demo (length 3-5 minutes) showing the relevant features of your project;

Contextually, (BOTH OPTIONS) some questions could be asked to the group members to assess each one’s contribution to the team.

**Very important!** DO NOT forget to have your project approved BEFORE presenting your work since the approval step (points 4 and 5 in the previous slide) is crucial if I suggest some corrections to your proposal.

At the end of your presentation, we will tell you if your project work is completed (in this case, you will receive a final official email recognizing your completed task) or if you should complete/change some aspects of your project and come back to present it again!
General rules

You are allowed to:

✓ use more than one tool, e.g. making comparisons between SQL and NoSQL in performing some tasks. But try to auto-assess the effort of the project you are planning to do, or it can be too much!

✓ use different tools w.r.t. the ones I will show you during the lectures (as long as they are of the same type), explicitly writing it in your proposal;

✓ use your personal dataset, but try to ensure that it matches the suitable characteristics w.r.t. what you want to present;

✓ divide the effort between the two group members, being transparent about what each of you has done, e.g., each member shows a subset of the total results (but both have to be aware of how to use the tools you selected).

You are NOT allowed to:

✗ present a project BEFORE my approval; of course, you can contact me for your doubts/questions, but you can show your work only after my OK.

✗ use private datasets without the consensus of the owner of the data, e.g., the dataset of a company you worked for.

✗ have a team of 2 people, but with only 1 working. At presentation and demo time, I will check if both the group members have a clear view of the project itself and the tool(s) used;

✗ have a team of more than 2 people.
**Tips**

- Try to select **challenging topics, tools, and datasets**: let this project be an opportunity to grow professionally and learn something practical and valuable for your career!
- Be **proactive**: if something is wrong or you feel stuck on a problem, or whatever, please write me an email ([valentini@diag.uniroma1.it](mailto:valentini@diag.uniroma1.it)) or (even better!) come to my office hours.
- Be **creative**: challenging problems require « thinking *outside the box.* »
- Choose wisely your **teammate**: working in pairs is another occasion to grow up professionally and build solid relationships!
- Focus on the **content of your projects** and not on the tool itself: you do not have to sell me some product, but explain to me what you have done (which is what I am interested in!), so please let your presentation be a clear description of your work.
- Consider your project also an opportunity to improve your **presentation skills**, both in terms of language and in terms of communication to the audience.