Large Scale and Multi-Structured Data Bases
University of Pisa
Academic Year 2022-2023
Workgroup Project Discussion Guidelines

The project discussion will be carried out in remote mode using the dedicate channel in MS teams.

The discussion will last maximum 40 minutes.

The fist 10 minutes will be dedicated to the live presentation of application. After a brief introduction of the application (running it live), the most relevant functionalities must be shown. Analytics, statistics, and suggestions must be highlighted. Examples involving all the considered databases must be shown.

Afterwards, a presentation with slides (10/15 minutes) should support the discussion.

We expect:

- The complete UML analysis class diagram for the complete data model (no need to show the use case class diagram, it will be checked in the documentation by the instructors and the main use cases should be discussed while running the application). 1-2 Slides
- A complete description of the dataset adopted and stored in the databases. 1-2 Slides
- A discussion on non-functional requirements must be provided in a couple of slides, including especially how the CAP theorem issues have been handled in the application. 1-2 Slides
- A discussion on the design of the data bases adopted for the application (collections organization, architecture of the graphs, structure of the namespaces and keys). Moreover, for each DB, the most relevant queries, those associated to analytics, statistics, and suggestions, must be shown in the slides. 1-2 Slides for each considered DB.
- Details on how data consistency has been managed and a discussion on a possible data sharding. 1-2 slides
- A discussion on the Software and Hardware architecture. 1 slide
- Some final considerations. 1-2 slides.

As in the final documentation, which must be submitted before the project discussion by the fixed deadline, all the choices must be justified.

During the presentation, instructors will ask for questions and clarifications.

After the 40 minutes presentation and discussions, the group will be asked to show the details of the code and of the databases.