Grenoble INP Ensimag
ISI 3A
Gestion de données à grande échelle
NoSQL assignment

Team size: 3 students.
Due December 14, 2016

The objective is to practice data management with a NoSQL system, either MongoDB or Neo4j. Make your choice according to your preferences.

1. Choose an application domain to work on and propose your own subject. The study case has to consider some relationships (1-1, 1-N, N-M) among entities.
2. Create the database corresponding to your case. Insert pertinent data so as to allow executing complex queries.
3. Create appropriate indexes.
4. Choose 5 complex queries (in the corresponding query language) to be evaluated on your data and execute them on your database. At least two queries have to involve indexed data.

For MongoDB users, the case should also

- Involve at least 4 collections and several data types in order to show the flexibility provided by Mongo (e.g, arrays, nested documents, collections with document having different fields).
- Use embedded documents so as references and duplication.
- Include queries concerning several collections. Use lookup operator for some queries but not for all.

For Neo4j users, the case should also

- Involve at least 5 node types. One of them should have at least 2 labels.
- Include several relationship types. Some of them should have properties.
- Define some constraints on the data.
- Include queries concerning several node types and at least one has to be a pattern including 3 nodes types.

This work has to be delivery by "Teide" system (https://intranet.ensimag.fr/teide/). Provide a compressed file that contains the report and two folders:

- Folder 1 contains the data, scripts and instructions to populate the database.
- Folder 2 contains the queries proposed (one file by query)

Report considerations:

- Write a text describing the data stored in the database. This text should not exceed 15 lines. You can include a figure to illustrate.
- Describe the indexes proposed and explain your decision. This text should not exceed 4 lines by index.
- For each query, provide it
  - in natural language. This text should not exceed 3 lines
  - in the language of the system and
  - If there are indexes involved, describe their impact. This text should not exceed 3 lines by index
  - Advantages and disadvantages of the query implementation proposed.
- Text font size: 11 or 12

Note: The compressed file name must follow the next format: "**noSQL_MSBigData_GroupXX**", where XX have to be changed by the corresponding number group assigned by Teide.