

## **Large Scale Data Analytics and Visualization**

**Faculty advisor:** Prof. Mariano Consens

The project objectives are to apply data analytics tools and techniques to large datasets to provide useful insights and to support decision making. Several projects are available in different application areas (in particular, social services, smart cities, mining industry). The sources of data include open data as well as data provide by collaborating organizations.

The project activities include the extraction, combination, and cleaning of multiple data sources (leveraging a scalable data management environment such as Hadoop and Spark), followed by the application of data mining and machine learning analysis techniques, and culminated by the preparation of visualizations of the results obtained (using tools such as Tableau, and/or interactive notebook-based visualizations).

Knowledge of SQL, as well as an interest in programming/scripting in notebook environments (e.g., Jupyter, Zeppelin), are required.

Several projects are available.