



## BIG DATA

INF/01 - 6 CFU - 2° Semester

### Teaching Staff

#### ALFREDO PULVIRENTI

**Email:** apulvirenti@dmi.unict.it

**Office:** Stanza 35, Terzo Blocco Dipartimento di Matematica e Informatica.

**Phone:** 095-7383087

**Office Hours:** Martedì 10-11.

---

### LEARNING OBJECTIVES

General teaching training objectives in terms of expected learning outcomes.

**Knowledge and understanding:** The course aims to give the knowledge and basic and advanced skills to the analysis of large amounts of data.

**Applying knowledge and understanding:** the student will acquire knowledge about the models and algorithms for analyzing data such as: mining high support, recommendation systems, search for similarities high dimension, map-reduce and spark, complex networks analysis, text mining and the document tagging systems.

**Making judgments:** Through concrete examples and case studies, the student will be able to independently develop solutions to specific problems related to big data.

**Communication skills:** the student will acquire the necessary communication skills and expressive appropriateness in the use of technical language in the general area of big data.

**Learning skills:** The course aims to provide students with the necessary theoretical and practical methods to deal independently and solve new problems that may arise during a work activity. For this purpose, different topics will be covered in class by involving students in the search for possible solutions to real problems, using benchmarks available in the literature.

---

### DETAILED COURSE CONTENT

High Support Data Mining. Recommendation Systems. Map-Reduce. Beyond or map-reduce Similarity search of higher dimensions: shingling, Min-Hashing, LSH, Min-LSH. Dimensionality reduction: SVD, CUR, Application to LSI Johnson-Lindenstrauss theorem. Link Analysis: PageRank, link spam, Hub-Authorities,

Applications on Map-Reduce. Web Advertising: online Algorithms, Adword and its implementations. Graph mining: subgraph matching, motif finding, community detection, Network alignment and network analysis. Text mining: TF.IDF, Bag-Of-Word, Entity annotation.

---

## **TEXTBOOK INFORMATION**

Mining of Massive Datasets

[Jure Leskovec](#), [Anand Rajaraman](#), [Jeff Ullman](#)

<http://www.mmids.org>

---