CLOUD COMPUTING AND BIG DATA
INF/01 - 6 CFU - 2° semestre

Docente titolare dell’insegnamento
GIUSEPPE PAPPALARDO
- Email: pappalardo@dmi.unict.it
- Edificio / Indirizzo: Dipartimento di Matematica e Informatica, Blocco 3, Ufficio n. 48, Viale A. Doria 6, Catania
- Telefono: 0957383000
- Orario ricevimento: http://www.dmi.unict.it/pappalardo/ricevim.php

OBIETTIVI FORMATIVI

1. **Knowledge and understanding (Conoscenza e capacità di comprensione).** Students will acquire a precise knowledge and understanding of fundamental concepts in the field of cloud computing, chiefly through a guided exploration of the main technological solutions available from the public Cloud, focusing on resources and services oriented to data storage, analysis, visualization and machine learning.

2. **Applying knowledge and understanding (Capacità di applicare conoscenza e comprensione).** Based on the operating knowledge acquired, students will develop an effective "toolset" of practical, application-oriented skills in leveraging the Cloud to cater for the typical needs of a data scientist: i.e. processing large datasets with a view to revealing meaningful patterns and relationships. Cloud implementations of state-of-the-art tools and frameworks like, e.g., MapReduce/Hadoop or TensorFlow, will be employed.

3. **Making judgements (Autonomia di giudizio).** The student will develop the ability to choose the suitable Cloud-based resource for the Data Science scenario of interest, properly estimating the ensuing costs and performance gains, as well as consciously assessing the tradeoffs involved.

4. **Communication skills (Abilità comunicative).** The student will acquire the communication skills required to express and discuss, at a rigorous technical level, the benefits and (mostly cost-related) downsides of the Cloud for Data Science applications. In addition, the student will gain the ability, for presentation purposes, to effectively highlight the features of very large datasets by means of cloud-based visualization services.

5. **Learning skills (Capacità di apprendimento).** Students will become capable of profitably consulting technical documentation concerning Data Science-oriented Cloud services, in order to concretely put them to effective use.
CONTENUTI DEL CORSO

**SQL on Google Cloud and BigQuery:** performing structured queries on BigQuery and Cloud SQL. Importing data from CSV files.

**Data acquisition into Google Cloud:** downloading selected data from a large public data set over the internet, and processing it with Google App Engine.

**Google Cloud Dataflow:** processing a real-time, real-world data set, and storing the results on the cloud. Case study: real-time geospatial data.

**Visualization with Google Data Studio:** Visualizing data stored in Google Cloud SQL. Visualizing Real Time Geospatial Data.

**Google Datalab for Data Analysis:** loading text data into Google BigQuery; rapid exploratory data analysis with Google Cloud Datalab notebooks.

**Google Cloud AI Platform:** using Google AI Platform to perform queries and present the data.

**Evaluating a Data Model:** partitioning a data set into a training set and a test set; evaluating various predictive models.

**Machine Learning with Spark on Google Cloud Dataproc.** Implementing logistic regression through machine learning on Apache Spark running on a Google Cloud Dataproc. Developing a model from a multivariable dataset.

**Machine Learning with TensorFlow:** developing and evaluating prediction models.

**MapReduce e Hadoop on Google Cloud:** exploiting parallelism and machine clusters.

---

TESTI DI RIFERIMENTO

1. Lecture notes, to be made available through the Studium portal.