

Data-intensive computing systems



Modalità d'esame

University of Verona
Computer Science Department

Damiano Carra

Modalità d'esame

☐ Progetto + Discussione orale

- Fase 1: si concorda il progetto
- Fase 2: si esegue il progetto e si consegna codice + relazione
- Fase 3: il progetto viene valutato
- Fase 4: discussione orale sul progetto e su tutto il programma

☐ I progetti possono essere svolti da soli o in coppia

☐ Non ci sono tempi predefiniti per la consegna

- Gli appelli servono solo per la verbalizzazione

☐ Valutazione «one-shot»

- In caso di rifiuto del voto, si deve sviluppare un nuovo progetto



Fase 1 - Scelta del progetto

☐ Nelle prossime slide segue una lista con esempi

- Si può scegliere uno qualsiasi dei progetti proposti
 - Anche se altri studenti hanno già scelto lo stesso progetto
 - Si assume che gli studenti siano sufficientemente maturi per non “copiare”
 - in ogni caso bisogna consegnare codice e relazione, su cui viene fatto un controllo
- Gli studenti possono proporre in autonomia un progetto diverso da quelli elencati

☐ I dettagli vanno concordati con il docente

3



Progetti / 1

☐ Inverted indexing

- Cap 4 del libro di Jimmy Lin
- Implementazione completa dell'algoritmo e test su un sottoinsieme di pagine di Wikipedia
 - L'output deve contenere, per ogni parola, la lista di pagine di wikipedia che contengono quella parola

☐ Full PageRank

- Cap. 5 del libro di Jimmy Lin
- Implementazione completa dell'algoritmo di page rank
 - include il jump factor (bisogna prima calcolare il numero totale di nodi)
 - include la gestione dei nodi-sink
 - Test basato su grafi con diverse caratteristiche (small world, random, regular)

4



Progetti / 2

- ❑ Implementazione di algoritmi descritti nel libro "Mining Massive Datasets"
 - Frequent Itemset
 - Sez. 6.4.4: The SON Algorithm and Map-Reduce
 - Counting Triangles in Social Graphs
 - Sez. 10.6.4: Finding Triangles Using Map-Reduce
- ❑ Confronto prestazioni tra Pig e MapReduce
 - Implementare l'analisi di un dataset (ad es., arrivi e partenze negli aeroporti) in Pig e MapReduce e confrontare le prestazioni

5



Progetti / 3

- ❑ Use case con tool BigData non visti a lezione:
 - Flink, Storm, Samza, Flume, ...
- ❑ Progetti interdisciplinari
 - Se state seguendo altri corsi e volete applicare ciò che avete visto in questo corso, proponete pure!
 - Es: Metodi di calcolo non convenzionale: il progetto è parallelizzabile?
 - Analisi varie su grafi
 - In collaborazione con altri docenti del Dipartimento
 - Esempi: Graph matching, Graph Centrality, Graph Clustering

6



Fase 2 - Sviluppo e consegna

- ❑ Si terrà conto dell'autonomia nella fase di sviluppo
 - Considerando le comunicazioni con il docente nella risoluzione dei problemi

- ❑ L'output del progetto dovrà essere:
 - Una relazione, che spiega nel dettaglio l'implementazione fatta e riporta i risultati dello studio di performance
 - Il codice sviluppato

7



Fase 3 e 4 - Valutazione e discussione

- ❑ Una volta valutato il progetto, seguirà un colloquio
 - Discussione sul progetto stesso (ad esempio, scelte implementative)
 - Discussione sulla connessione tra il progetto e le tematiche svolte nel corso
 - Domande aperte sui temi svolti durante il corso

8

