

Data-intensive computing systems



Esempi di progetti

University of Verona
Computer Science Department

Damiano Carra

Generalità

- ❑ Lista con indicazione sommaria del progetto
 - Serve per capire il tema
- ❑ Il progetto vero e proprio va poi concordato / discusso con il docente nei suoi dettagli
 - Alcune implementazioni potrebbero richiedere successivamente uno studio di performance
 - es., impatto di alcuni parametri (numero di reducers, numero di macchine, dimensione dell'input) sul tempo di analisi
- ❑ 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
 - Funzionante! → Il codice verrà testato dal docente



Generalità (cont'd)

- ❑ I progetti possono essere svolti da soli o in coppia
 - Nel caso di due persone, la complessità ovviamente aumenta

- ❑ Lo studente 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 dettagliato

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

- ❑ Algoritmi EM e algoritmi di ottimizzazione gradient based
 - Cap 7 del libro di Jimmy Lin
 - Implementazione degli algoritmi e loro applicazione su dataset forniti dal docente

- ❑ 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

5



Progetti / 3

- ❑ 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

- ❑ 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 la [Dott.ssa Giugno](#), ospite del Dipartimento fino alla prossima estate
 - Esempi: Graph matching, Graph Centrality, Graph Clustering

difficile!

6



Sulla valutazione della didattica



7

Iscrizione all'esame

- Indipendentemente da quando consegnerete il progetto

ISCRIVETEVI AD UNO DEGLI APPELLI DI FEBBRAIO

(e passate parola a tutti coloro che oggi non sono presenti)

- Motivo

- Serve per compilare il questionario della didattica
- Solo se ci saranno un numero sufficientemente alto di compilazioni il questionario avrà valore (statistico)



8

Valutazione della didattica

❑ Estremamente importante!

- Feedback ufficiale dato dagli studenti
- Aiuta a migliorare l'offerta formativa

❑ Tuttavia, ci sono dei limiti sulla sua utilità

- Una valutazione numerica indica che qualcosa è andato male o bene, ma non il *perché*
- Esempio di domanda:
 - “Le modalità d'esame sono state definite in modo chiaro dal docente?”
- Che significato hanno i voti 2 o 3 (su una scala da 1 a 4)?

9



Due richieste

1. **Prima** di iniziare a compilare il questionario

- Pensate ad almeno 4 cose che migliorereste del corso
 - Ad es., argomenti del corso che non sono state chiari, oppure “inutili”, oppure che avreste voluto fossero approfonditi
 - Pensate in maniera *costruttiva*!
- Scrivete tali migliorie su un foglio da tenere vicino a voi

2. **Mentre** state compilando il questionario

- Per ogni voce del questionario a cui date un voto insufficiente, scrivete sul foglio che vi state tenendo a fianco le motivazioni del voto assegnato

→ Nel testo libero che si trova alla fine del questionario, riportate le annotazioni che vi siete fatti sul foglio

→ PS: Eventualmente annotate anche cosa vi è piaciuto

10

