

Database e data mining

Esame Finale

L'esame finale consiste nello sviluppo ed implementazione, mediante le API java di Hadoop e il paradigma MapReduce, degli algoritmi per la risoluzione dei seguenti problemi:

1. Indice invertito
2. Co-occorrenze

Descrizione del progetto

Il progetto è costituito da:

- A. pseudo-codice dell'algoritmo Map-reduce per l'implementazione dei due job*
- B. codice sorgente Java dell'implementazione dei due algoritmi*
- C. relazione finale*

Allo studente è richiesto la consegna del seguente materiale:

1. codice sorgente Java adeguatamente commentato
2. package jar dei due job
3. relazione in formato digitale contenente:
 - a. descrizione dello pseudo-codice dei due job
 - b. descrizione e commento dell'architettura dei job implementati mediante le API Java con riferimento al framework Hadoop
 - c. screenshot delle esecuzioni dei job
 - d. commenti finali sull'esecuzione dei job
 - e. commenti su potenziali miglioramenti e loro possibili implementazioni

Scadenze

Tutto il materiale deve essere consegnato entro la lezione del 19/12/2013. In tale data verrà effettuata anche la discussione del materiale prodotto.