Cognome e Nome _____

Matricola _____

Big data – Graph Mining
A.A. 2022/2023

Esonero del 15 dicembre 2022

Esercizio 1 (9 punti)

Siano dati in input delle quadruple **(id_transazione, id_carta_debitore, id_esercente, importo)** che rappresentano delle transazioni effettuate con POS, in cui la transazione di Euro **importo** con identificativo univoco **id_transazione** avviene con la carta di pagamento **id_carta_debitore** in favore dell'esercente **id_esercente** *(che varia tra 0 e n-1)*.

a. Si progettino (eventualmente utilizzando più step di computazione tra loro concatenati) le funzioni *Map e Reduce* per dare in output tutti gli **id_esercente** corrispondenti ad attività commerciali che hanno incassato (come somma di tutte le transazioni) almeno € 300.000,00.
b. Si progettino (eventualmente utilizzando più step di computazione tra loro concatenati) le funzioni *Map e Reduce* per calcolare il massimo incasso ottenuto da un'attività commerciale, utilizzando spazio locale $O(n^{2/3})$.

Esercizio 2 (8 punti)

Si consideri l'algoritmo *Park Chen Yu* (con singola tabella hash) nell'ambito del problema degli *itemset frequenti*, e se ne descriva il funzionamento.

Si dica inoltre, in modo **sintetico e puntuale, giustificando la risposta**:

a. Quale è il motivo per cui l'algoritmo PCY si concentra sulle coppie di itemset frequenti.
b. Cosa si può affermare quando un contatore della tabella hash raggiunge la soglia di frequenza, e cosa si può dire quando non la raggiunge.
c. Perché l'algoritmo PCY impiega il metodo delle triple per memorizzare le coppie di itemset candidate ad essere frequenti.

Regole per lo svolgimento della prova scritta:

- Per svolgere il compito si hanno a disposizione **75** minuti.
- Scrivere **subito** nome, cognome, matricola su OGNI FOGLIO (**compreso questo**).
- Durante la prova scritta **non** è possibile abbandonare l'aula.
- Non è ammesso **per nessun motivo** comunicare in qualsiasi modo con altre persone
- Non è possibile consultare appunti, libri e dispense.
- Qualsiasi strumento elettronico di calcolo o comunicazione (telefoni cellulari, calcolatrici, palmari, computer, etc…) deve essere **completamente disattivato** e **depositato in vista sulla cattedra**
- Mettere in vista sul banco un valido documento di identità.