Cognome e Nome _____

Matricola _____

Big data – Graph Mining
A.A. 2023/2024

Esonero del 15 dicembre 2023

Esercizio 1 (8 punti)

Siano date in input

- delle triple **(data, id_impianto, produzioneKWh)** che rappresentano i dati delle produzioni di impianti fotovoltaici, relativi agli ultimi 5 anni: l'impianto **id_impianto** (che varia da 0 a n-1) ha prodotto **produzioneKWh** KWh nel giorno **data**;
- delle coppie **(id_impianto, provincia)** che contengono la provincia di installazione dell'impianto fotovoltaico **id_impianto**.

1. **[3 punti]** Si progettino (eventualmente utilizzando più step di computazione tra loro concatenati) le funzioni *Map e Reduce* per dare in output l'impianto fotovoltaico che, negli ultimi 5 anni, ha prodotto più energia, utilizzando spazio locale $O(n^{1/2})$. Si discuta il *reducer size* di tutti gli step individuati.
   Suggerimento: calcolare prima il quantitativo di energia prodotto da ogni impianto fotovoltaico, per poi individuare l'impianto che ha prodotto più energia.
2. **[5 punti]** Si progettino (eventualmente utilizzando più step di computazione tra loro concatenati) le funzioni *Map e Reduce* per dare in output la provincia che, globalmente, ha prodotto più energia, utilizzando spazio locale $O(n^{1/2})$. Si discuta il *reducer size* di tutti gli step individuati.
   Suggerimento: calcolare prima il quantitativo di energia prodotto da ogni impianto fotovoltaico, passare quindi a calcolare il quantitativo di energia prodotto globalmente in ogni provincia per poi individuare la provincia che ha prodotto più energia.
   [Esercizio alternativo (dà luogo a una valutazione di **3 punti**)]: non tenere conto del vincolo di spazio locale.

Esercizio 2 (9 punti)

1. **[1,5 punti]** Si definisca il problema dei Frequent Itemset.
2. **[1,5 punti]** Si definisca cosa è il **supporto** di un itemset e quando un itemset può considerarsi **frequente**; si enunci la proprietà della **monotonia della frequenza**.
3. **[6 punti]** Far vedere sulla seguente istanza l'esecuzione passo passo dell'algoritmo A-priori per individuare tutti gli insiemi di item che abbiano supporto almeno uguale a **3**.

basket	**item**
1	1, 2, 3, 4
2	5, 6, 7, 8, 9, 10
3	1, 3, 5, 6, 7, 8
4	2, 7, 8, 9

5	7, 9
6	1, 5, 6, 7
7	3, 5, 7, 9, 10
8	7, 8, 9
9	5, 7, 8
10	2, 4, 5

Regole per lo svolgimento della prova scritta:

- Per svolgere il compito si hanno a disposizione **80** 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.