

Sistemi Informativi Evoluti e Big Data – A.A. 2025-2026

Homework #4 – MapReduce

L'esercizio richiede di progettare e sviluppare un programma utilizzando il paradigma MapReduce. Nello specifico, si consideri un dataset sui passaggi ai caselli autostradali è descritto dai campi:

```
[IDVeicolo; TipoVeicolo; Tratta; Pedaggio; DataTransito;  
FasciaOraria; Provincia]
```

Il programma MapReduce deve calcolare il pedaggio medio pagato da ogni tipo di veicolo (es. camion) nell'anno 2025 e come questo è cambiato rispetto a 10 anni fa. Il campo Data è nel formato GG/MM/YYYY.

Descrizione del compito – Generare automaticamente un dataset d'esempio, utilizzando Python, progettare il programma MapReduce (valutando anche l'utilizzo della funzione Combine) e proporre due implementazioni del programma, ottenute tramite l'uso dei pacchetti `mrjob` e `PySpark` (utilizzo in locale). Fare delle considerazioni qualitative (e, se possibile, anche quantitative) sulle due diverse implementazioni.

Dettagli sulla consegna – Predisporre due file `.py` da eseguire in locale e un file PDF in cui sono riportate le considerazioni sui risultati ottenuti. Includere tutti i file in uno zip e caricare l'archivio su Moodle.