

CORSO INTRODUZIONE AI BIG DATA

La ditta STECCAPARAPETUTTI S.R.L. ha un sistema di contabilità che memorizza tutti gli ordini fatti (per ordini si intendono Fatture, Buoni Prelievo, Offerte...). In previsione dell'anno successivo, la ditta vorrebbe sapere il suo andamento di vendita.

Scaricato il file Ordini.csv strutturato in questa maniera:

Tipo-documento,data(aaaaammgg),costo(€)

FATTURA, 20160104,139.8

FATTURA, 20160104,169.2

FATTURA, 20160104,65.3

.....

Progettare e realizzare per ogni anno i seguenti job:

1. Calcola la media di vendita per ogni mese di ogni anno
2. Calcola la varianza di vendita per ogni mese di ogni anno
3. Identifica il mese di ogni anno con maggiore vendita
4. Identifica il mese di ogni anno con minore vendita

Per ciascun job bisogna illustrare e documentare in un rapporto finale:

- Implementazione R
- Implementazione del paradigma mapReduce di HADOOP.
Nel caso in cui si verificassero problemi con l'installazione e il funzionamento di HADOOP con R allora si può utilizzare il TOOL.

Deve essere consegnato solamente il rapporto finale scritto come una relazione. Con i codici allegati.

I Risultati ottenuti devono essere illustrati tramite dei grafici.

Suggerimento: i documenti tipo "offerta" sono da considerarsi come una vendita effettiva?