



## CORSO DI LAUREA MAGISTRALE IN INGEGNERIA INFORMATICA CLASSE LM32

Algoritmi e Strutture dati per i big data | **cfu 9** | Prof. Paola Vocca

### Etivity 3

**Svolgere i seguenti esercizi. Come risultato dell'etivity inviare il listato del programma.**

1. Installare Spark e Python ed aprire una sessione SparkSQL.
2. Creare un DataFrame
  - a. Da una lista Python
  - b. Da un file csv
  - c. Da un file json
3. Inserire la riga di intestazione
4. Visualizzare il dataset creato
5. Visualizzare la struttura
6. Selezione delle colonne.
7. Definire una colonna numerica e su cui fare operazioni di aggregazione dopo l'operazione di group by
8. Implementare le operazioni di select e filter
9. Fare un esempio di operazioni di ordinamento
10. Richiamare da SparkSQL istruzioni SQL