

# CLDR-2 - CLUDERA DATA DEVELOPER

Categoria: **Cloudera**

## INFORMAZIONI SUL CORSO



**Durata:**  
5 Giorni



**Categoria:**  
Cloudera



**Qualifica Istruttore:**  
Docente Senior (min.  
5 anni)



**Dedicato a:**  
Sviluppatore



**Produttore:**  
PCSNET

## OBIETTIVI

Al termine del corso gli allievi avranno acquisito tutti gli elementi necessari all'implementazione di piattaforme Big Data per l'analisi ed il processamento dei dati.

## PREREQUISITI

- Buona conoscenza della programmazione strutturata derivata dall'uso di un qualsiasi linguaggio di programmazione
- Buona conoscenza dei database e del SQL
- Buona conoscenza del sistema operativo Linux

## CONTENUTI

### Data Ingestion

- Essere in grado di trasferire i dati tra sistemi esterni ed il tuo cluster
- Importare dati da Mysql dentro HDFS usando Sqoop
- Esporta i dati su database Mysql usando Sqoop
- Cambiare le strutture dei file durante gli import usando Sqoop
- Fare ingestion dei dati dentro HDFS con streaming real-time e near real time
- Processare streaming di dati come questi vengono caricati sul cluster
- Caricare i dati dentro e fuori hdfs usando i comandi shell

### Trasformare, memorizzare e strutturare i dati

- Convertire un set di dati in un dato formato memorizzato in hdfs dentro un nuovo formato e scriverlo sempre in hdfs
- Caricare i dati RDD da hdfs usando Spark
- Scrivere i dati RDD in hdfs usando Spark

### Leggere e scrivere file in vari modi

- Performare flussi di estrazioni, trasformazioni e caricamento sui dati per fare Data Analysis
- Usare SparkSQL per interagire con il metastore programmaticamente nelle tue applicazioni
- Capire i fondamenti di interrogare i dataset con Spark

- Filtrare i dati usando Spark
- Scrivere interrogazioni che creano statistiche di dati aggregati
- Incrociare tipologie di database differenti usando Spark
- Produrre dati ordinati e classificati

### **Configurazione**

- Usare opzioni su linea di comando per cambiare la tua configurazione
- Come incrementare la memoria disponibile

## INFO

**Materiale didattico:** Materiale didattico e relativo prezzo da concordare

**Costo materiale didattico:** NON incluso nel prezzo del corso

**Natura del corso:** Operativo (previsti lab su PC)