OSDA-15 - BIG DATA ARCHITECT

Categoria: **Data & Artificial Intelligence**

INFORMAZIONI SUL CORSO

Durata:	Categoria:	Qualifica Istruttore:	Dedicato a:	Produttore:
4 Giorni	Data & Artificial Intelligence	Docente Senior (min. 5 anni)	Analista	PCSNET

OBIETTIVI

Al termine del corso gli allievi dovranno aver acquisito tutti gli elementi necessari all'implementazione di piattaforme Big Data per l'analisi ed il processamento dei dati.
Nel corso sono trattati i seguenti macro-argomenti:

- -Introduzione a Big Data e Big Data Analytics
- -Il ruolo del Data Scientist
- -Database NoSQL
- -Introduzione alla piattaforma Apache Hadoop
- -Introduzione a Spark
- -Costruire un data lake

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

Introduzione a Big Data e Big Data Analytics

- -Cosa sono i Big Data
- -I concetti chiave dei Big Data e della Data Science
- -Il ciclo di vita del data analisys
  - -Descriptive Analytics
  - -Predictive Analytics
  - -Prescriptive Analytics
  - -Automated Analytics
- -Il ruolo del Data Scientist

Database NoSQL

- -Classificazione

-Caratteristiche principali
-Differenze con i DB relazionali
-Limiti dei DB relazionali
-Modellazione dei dati
-Panoramica dei principali db NoSQL

Hadoop

-Introduzione ed Ecosistema
-Introduzione ad Apache Hadoop
-Panoramica Ecosistema Hadoop
-Problematiche nei sistemi di grandi dati
-Perchè Hadoop e i suoi vantaggi
-Architettura Hadoop e HDFS
-Introduzione ad Hadoop Distribuited File Sistem
-Distrubuzione dei processi in un cluster
-Storage:conservazione dati nell'architettura HDFS
-Utilizzo e applicazioni con HDFS
-Resource Management: Architettura YARN e utilizzi

Costruire un data lake con Apache Hadoop

-Struttura di un data-lake
-Principi di base e approccio
-Affinità con i data-warehouse
-Criticità
-Utilizzare Apache Hadoop per i data-lake

Hive

-Introduzione
-Panoramica Data Storage
-Vantaggi e motivi del loro utilizzo
-Creazione database e tabelle
-Caricamento dati nelle tabelle
-Utilizzo di Sqoop in Hive

Apache Spark

-Introduzione
-Utilizzo Spark Shell
-RDDs (Resilient Distributed Datasets)
-Programmazione funzionale in Spark
-Applicazioni RDDs con Spark
-Key-Value Pair RDDs
-MapReduce
-Operazioni RDD
-Accenni di Scal in Spark
-SQL Spark
-Panoramica Spark SQL e SQL Context
-Trasformazione e Interrogazione DataFrames
-Salvataggio DataFrames

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)