APACHE SPARK CON JAVA

Objetivos:

El alumno al final del curso aprenderá:

- Una descripción general de la arquitectura de Apache Spark.
- Abstracción primaria de Apache Spark,
- Conjuntos de datos distribuidos resilientes (RDD) para procesar y analizar grandes conjuntos de datos.
- Desarrollar aplicaciones Apache Spark 2.0 utilizando transformaciones y acciones RDD y Spark SQL.
- Escalar aplicaciones Spark en un clúster Hadoop YARN a través del servicio Elastic MapReduce de Amazon.
- Analizar datos estructurados y semiestructurados utilizando conjuntos de datos y marcos de datos,
- Desarrollar una comprensión profunda de Spark SQL.
- Compartir información entre diferentes nodos en un clúster de Apache Spark mediante acumuladores y variables de transmisión.
- Técnicas avanzadas para optimizar y ajustar los trabajos de Apache Spark mediante la partición, el almacenamiento en caché y la persistencia de RDD.

Duración: 20 Horas

Dirigido a:

- Cualquiera que quiera comprender completamente cómo funciona la tecnología Apache Spark y aprender cómo se utiliza Apache Spark en el campo.
- Ingenieros de software que quieran desarrollar aplicaciones Apache Spark 2.0 usando Spark Core y Spark SQL.
- Científicos de datos o ingenieros de datos que quieran avanzar en su carrera mejorando sus habilidades de procesamiento de big data..

Requisitos: Habilidades y conocimientos previos de programación Java

APACHE SPARK CON JAVA

Grupoloyal

Contenido:

- Comenzando con Apache Spark
  - Introducción a Spark
- Spark Arquitectura y Componentes
  - Arquitectura de Spark
  - Componentes de Spark
- Tópicos avanzados de Spark.
  - Acumuladores
  - Soluciones a problemas de StackOverflow Survey Follow-up
  - Variables de difusión
- Spark SQL y Dataframe
  - Introducción a Spark SQL
  - DataSet
    - Crear un Dataset
    - Tipos de Dataset
    - Operar con Dataset
    - Filtros usando Expresiones / Lambdas / columnas
    - Usando Dataset o RDD
    - Conversión de Dataset y RDD
    - Ajuste del rendimiento de Spark SQL
    - Motor SQL distribuido
  - DataFrame
    - Creación de Dataframe