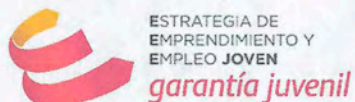




red.es



UNIÓN EUROPEA

Fondo Social Europeo  
"El FSE invierte en tu futuro"

Se otorga el presente Diploma a

**D. Manuel Montes** con D.N.I.:

Por superar con aprovechamiento el curso

## **Desarrollador Big Data Spark Cloudera**

de 230 horas totales de duración, en modalidad mixta, del 17/09/2018 al 29/10/2018.

En el marco del Exp. 2017/JUVE/ 00001751 al amparo de Resolución de 12 de enero de 2017, de la Entidad Pública Empresarial Red.es, por la que se establecen las bases reguladoras del programa de formación para el empleo juvenil en la economía digital.

Y para que así conste, expido el presente certificado.

En Madrid, a 29 de Octubre de 2018

Entidad beneficiaria

**TECH TRAIN**  
**SOLUTIONS**

El alumno

Contenidos: La formación ha sido realizada en la modalidad mixta, 230 horas de duración total ( 175 horas de forma presencial y 55 horas on line )

**Módulo nº 1 INTRODUCCIÓN A SQL ( 65 horas )**

**Contenidos teórico- prácticos:**

- Introducción
- Base de datos básica
- Estructura de base de datos
- El uso de SQL
- Creación de una base de datos
- Las consultas básicas
- Mantenimiento de la Base de datos
- Copia de tablas y modificaciones de columnas
- Índices y restricciones
- Eliminar o modificar filas de la tabla de datos
- Funciones: uso y tipos
- Funciones específicas
- Joins
- Exportación de datos, consultas y utilidades
- Importación de datos y de archivos de datos
- Importación con sentencias y utilidades
- Subconsultas generales y básicas
- Subconsulta no correlacionada y correlacionada
- Tipos de subconsultas
- Modificación de la tabla con subconsultas
- Motores de almacenamiento
- Creación de Vistas
- Transacciones
- Recuperación de Metadatos
- Conclusiones

**Módulo nº 2 PROGRAMACIÓN PYTHON ( 40 horas )**

**Contenidos teórico - prácticos:**

- Introducción
- Introducción to Python
- Variables en Python
- Colecciones en Python
- Flow Control en Python
- Estructura de Programación en Python
- Trabajo con librerías en Python
- Conclusión

**Módulo nº 3 PROGRAMACIÓN SCALA ( 40 horas )**

**Contenidos teórico – prácticos:**

- Introducción
- Scala Overview
- Introducción a Scala
- Scala's Role in Distributed
- Procesamiento de Datos
- La motivación de Scala
- Básicos Scala
- Trabajo con tipos de datos
- Agrupación de datos
- Flow Control en Scala
- Uso y creación de librerías
- Conclusión

**Módulo nº 4 CLOUDERA DESARROLLADOR PARA SPARK & HADOOP ( 85 horas)**

**Contenidos teórico – prácticos:**

- Introducción
- Introducción a Apache Hadoop y el ecosistema Hadoop
- Archivos de almacenamiento de Apache Hadoop
- Proceso de datos en un cluster de Apache Hadoop
- Relación de datos importados con Apache Sqoop
- Fundamentos de Apache Spark
- Trabajar con RDDs
- Agregación de datos en pares RDDs
- Escritura y ejecución de aplicaciones de Apache Spark
- Configuración de aplicaciones de Apache Spark
- Procesos paralelos en Apache Spark
- Persistencia de RDD
- Patrones comunes de proceso de datos en Apache Spark
- DataFrames y Spark SQL
- Procesamiento de mensajes con Apache Kafka
- Captura de datos con Apache Flume
- Integración de Apache Flume y Apache Kafka
- Apache Spark Streaming: Introducción a DStreams
- Apache Spark Streaming: Procesado de múltiples Batches
- Apache Spark Streaming: Data Sources
- Conclusión