

A promotional banner for a Big Data training course. The background shows a person in profile working at a computer in a dimly lit office. The text 'BIG DATA' is prominently displayed in white at the top right. Below it, a blue bar contains the text 'APRENDE A PROGRAMAR' in white. Underneath, the date '16 de Febrero 2019' is written in blue. Three circular logos for Hadoop, Apache Spark, and Scala are arranged horizontally. At the bottom left, a URL 'http://integratekmexico.com/training.html' is provided. The bottom of the banner features four logos: DATA INFINITY, Women.inData, Data.Keepers, and integra tek MEXICO.

BIG DATA
APRENDE A PROGRAMAR
16 de Febrero 2019

<http://integratekmexico.com/training.html>

DATA INFINITY
deep insight

Women.inData;
women, data insight

Data.Keepers;

integra
tek
MEXICO

Big Data (Nivel Básico – Intermedio)

Objetivo General

El presente curso proporciona al perfil un conocimiento básico-intermedio de lo que es el mundo de Big Data, aportara conocimiento de programación en lenguaje **Scala** con el **Framework Spark** en integración con **Cludera Distribution Hadoop (CDH)** y la utilización de **Bases de Datos Operacionales**.

Este curso está enfocado a personas del mundo de sistemas que cuenten con los conocimientos básicos-intermedios de programación y que tengan el deseo de seguir preparándose para aportar a su perfil herramientas y conocimientos de Big Data.

Requerimientos mínimos:

Deseable conocimientos de comandos Linux y nociones de programación.
Equipo de cómputo con 8GB de RAM o superior y 30GB disponibles de almacenamiento.

Curso: 20 horas

Distribución: 4 clases sabatinas.

<http://integratekmexico.com/training.html>
informacion@integratekmexico.com

TEMARIO BIG DATA (Nivel Básico - Intermedio)

Módulo 1: (5 horas)

- I. Introducción al Mundo de BigData
 - I.1 Ciencia de Datos
 - I.2 Evolución de las Bases de Datos
 - I.3 BigData

- II. Conociendo Apache Hadoop
 - II.1 Instalación de Máquina Virtual (Cloudera)
 - II.2 Configuración del ambiente en modo local

Módulo 2: (5 horas)

- III. Conociendo Apache Spark (1.6)
 - III.1 Introducción de Spark.
 - III.2 Ambiente de entrenamiento Spark.

- VI. Lenguaje de programación SCALA
 - IV.1 Introducción a Scala
 - IV.2 Aplicación de Scala en Spark

Módulo 3: (9 horas)

- V. Integración SCALA - Spark
 - V.1 Introducción a SCALA-Spark
 - V.2 Transformaciones y acciones de RDDs simple
 - V.3 Transformaciones y acciones de RDDs clave/valor
 - V.4 Lectura y escritura de Archivos (HDFS)
 - V.5 Introducción Spark-SQL
 - V.6 Spark-SQL creación de DataFrames
 - V.7 Spark-SQL operaciones con DataFrames
 - V.8 Spark-SQL con Hive
 - V.9 Spark-SQL y bases de datos operacionales (HBASE / MONGODB)

Módulo 4: (1 hora)

- VI. Otros componentes de Apache Spark
 - VI.1 Spark Streaming
 - VI.2 Spark MLlib
 - VI.3 Spark GraphX