

Big Analytics: de la información al conocimiento

Programa

1. Introducción
2. Almacenamiento
3. Arquitectura e Implementación
4. Ingesta de datos
5. Lenguajes de Programación
6. Bases de Datos
 - a. Relacionales
 - b. No Relacionales
 - c. Indexadores
7. Procesamiento y Transformación del Dato
8. Analytics



Big Analytics: de la información al conocimiento

1.- Introducción

Arquitectura:

Ecosistema y piezas principales

Tecnologías, batch layer, speed layer, ingesta, procesamiento, analítica.

Que es, como funciona, origen y evolución.

Arquitectura Lambda:

Batch Layer y Speed Layer.

Sources, ingesta, almacenamiento, procesamiento, explotación y analítica.

Piezas cross: Gobierno del dato, seguridad, monitorización, trazabilidad, Business Intelligence, ETL.

Ejemplos reales y casos de uso:

Amazon, banca, smart cities, IoT.

Arquitectura: Ecosistema y piezas principales, tecnologías, batch layer, speed layer, ingesta, procesamiento, analítica.

2.- Almacenamiento

Almacenamiento y procesamiento básico:
HDFS, MR, Hive, Pig, Solr, Impala

3.- Arquitectura e Implementación

Open Source vs Distribuciones comerciales (Cloudera, Hortonworks, MapR, IBM BigInsights...). Instalación on premise (arquitectura física) vs Cloud (Ej: AWS). Arquitectura y componentes de AWS. Ejemplos sobre AWS.



4.- Ingesta de datos

Ejemplo Máquina virtual Cloudera.

Instalar máquina virtual y ejemplo guiado sencillo.

Ingesta batch: Kafka con ejemplos.

Big Analytics: de la información al conocimiento

5.- Lenguajes de Programación

R

Python

6.- Bases de Datos Relacionales

SQL

7.- Bases de Datos No Relacionales

MongoDB

Neo4j, Hbase

Cassandra y/o otras.

Teoría y diferencias. Ejemplos de Hbase y de MongoDB y/o otras.

8.- Bases de Datos: Indexadores

Elasticsearch y ejemplos.

9.- Procesamiento y Transformación del Dato

Spark: Spark en Python. Batch y Real Time (Spark Streaming).

Spark: Spark en Python. Machine Learning en Spark: ML y MLib.

Spark: SparkR

Análisis de redes sociales. Spark GraphX y Gephi.

10.- Analytics

Introducción al análisis masivo de datos: descriptivos y visualización de Big Data

Técnicas no supervisadas de análisis de datos

Aprendizaje supervisado de datos: Técnicas de reducción de datos y de clasificación I.

Técnicas de clasificación 2: Técnicas de combinación de clasificadores

Máquinas de vectores soporte (SVM) y algoritmos genéticos.

Técnicas de Regresión.

Modelos probabilísticos gráficos.

