**INTELIGENCIA ARTIFICIAL  
Procesamiento de grandes volúmenes de datos**

**UNIDAD 1   
BIG DATA**

* **Introducción**
* **Almacenamiento y procesamiento de grandes cantidades de datos**
* **Los problemas actuales**
* **Las soluciones y bases de datos propuestas**
* **Bases de datos Big Data**
* **Sistema de archivos distribuidos para garantizar escalabilidad**

**UNIDAD 2  
PREPARANDO NUESTRO ENTORNO DE TRABAJO**

* **Bash Shell**
* **Concepto de virtualización**
* **Concepto de contenedores e imágenes**
* **Diferencias entre virtualización y contenedores**
* **VirtualBox, instalación, configuración y ejecución**
* **Docker, instalación, configuración y ejecución**

**UNIDAD 3  
BASES DE DATOS NoSQL –TEMA IMPORTANTE**

* **Diferencias con las bases de datos SQL o relacionales**
* **Tipos de bases de datos NoSQL**
* **Base de datos documentales, orientadas a grafos, clave valor**
* **Base de datos orientadas a objetos, tabulares, orientadas a Arrays**
* **Ventajas y desventajas de las bases de datos NoSQL**
* **MongoDB, instalación, configuración y utilización**
* **El servicio MongoDB y el cliente MongoDB**
* **Terminología, colecciones, documentos, operaciones**
* **Modulo PyMongo de Python para utilizar MongoDB**
* **Neo4j**

**UNIDAD 4  
APACHE HADOOP**

* **Qué es Apache Hadoop**
* **El Ecosistema Hadoop**
* **Gestión de datos**
* **Acceso a datos, Hive, Pig, Hbase, Mahout**
* **Gobernabilidad e integración de datos, Flume, Sqoop**
* **Seguridad, Knox**
* **Operaciones, Ambari, Zookeeper**
* **Conceptos básicos Hadoop. Instalación. Configuraciones básicas.**
* **Almacenamiento y procesamiento distribuído.**
* **Capa de Almacenamiento HDFS (Hadoop Distributed File System).**
* **Estructura de datos distribuidos.**
* **Capa de Procesamiento MapReduce.**
* **Estructura de procesamiento distribuido.**
* **Map y Reduce.**

**UNIDAD 5  
APACHE SPARK**

* **Apache Spark y Python para Big Data y Machine Learning**
* **Instalación de Apache Spark**
* **Pyspark y los RDDs**
* **Cargando y explorando datos**
* **Pre-procesamiento de datos**
* **Construir un modelo de Machine Learning con Spark ML**
* **Evaluación del modelo**

**UNIDAD 6  
AMAZON WEB SERVICES**

* **Capa gratuita de Amazon**
* **Creación de una cuenta.**
* **AWS Console**
* **Amazon Elastic Compute Cloud (EC2)**
* **Creación de un servidor virtual e instalación S.O.**
* **Amazon DynamoDb**