

# Plan Académico 2018-2019

Máster en Big Data & Business  
Analytics (MBDA)

*Norberto Díaz-Díaz*  
*Director Académico MBDA*



## Módulo I. Introducción

---

### Asignatura: Introducción a Big Data

En esta asignatura se introduce al estudiante al paradigma Big Data. Así mismo se aborda a qué se debe el incremento de la cantidad e importancia de los datos y cómo se pueden abordar éstos mediante técnicas inteligentes.

- **Introducción a Big Data**  
¿Qué es el Big Data? Cuáles son las tecnologías que emplea y cuáles son las categorías de casos de uso principales
- **Digitalización del Front Office**  
Dada la re-Evolución Digital a la que estamos asistiendo, tanto las personas como las empresas, es necesario reflexionar sobre las fuerzas digitales que provocan esa re-evolución, con el objetivo de para entender el impacto al que se ven sometidas las organizaciones en su transformación digital.
- **Knowledge Data Discovery**  
Se analizarán cada una de las fases del proceso de descubrimiento de conocimiento, analizando tanto el aprendizaje supervisado como el no supervisado.

## Módulo II. Tecnologías Big Data

---

### Asignatura: Gestión y Visualización de Datos Masivos

En la asignatura se abordarán tanto los conceptos básicos como las tecnologías específicas para el manejo de grandes conjuntos de datos, su análisis preliminar, su acceso y visualización, e incluso las cuestiones legales que afectan a este tipo de almacenes.

- **Gestión de datos masivos**  
Principios básicos en los almacenes de datos distribuidos: Teorema CAP, BaSE , replicación y *sharding*.
- **Tecnologías básicas de procesamiento**  
Introducción al ecosistema Hadoop, al paradigma Map-Reduce y práctica con HDFS.
- **Motores de datos**  
Introducción a las bases de datos NoSQL incluyendo una amplia variedad de modelos (almacenes clave-valor, documentales, orientadas a columnas, grafos y bases de datos *in-memory*), así como casos prácticos que permitirán aproximarse a Apache Cassandra, MongoDB, Neo4J y HBase.
- **Visualización de datos**  
Técnica y herramientas para el análisis exploratorio de datos gracias a la visualización, incluyendo prácticas con Gephi y R.
- **Recuperación de información desde almacenes de datos masivos**  
Fundamentos de la recuperación de información, su importante, así como casos prácticos con Apache Solr.
- **Cuestiones legales en el manejo de datos masivos**  
Tomar conciencia de la normativa que se aplica dentro del ámbito del Big Data tanto a nivel nacional como europeo, los límites que se imponen y las obligaciones que conlleva el manejo de datos masivos.

### Asignatura: Plataformas Tecnológicas

En la presente asignatura se abordarán las distintas herramientas y frameworks que servirán para sacar el máximo partido a los diferentes gestores de datos masivos vistos en la asignatura anterior. La comprensión de todas las herramientas existentes le dará una visión global al estudiante de las distintas posibilidades de procesamiento, así como a manejar entornos empresariales Cloud. Gracias al gran abanico de tecnologías abordar, será posible conocer el horizonte tecnológico existente y optar por aquellas herramientas adecuadas al afrontar los retos de proyectos e infraestructuras Big Data según los requerimientos precisados. Además, servirán como punto de partida para comprender aquellas tecnologías y algoritmos de la asignatura de "Tecnologías Avanzadas".

- **Tecnologías Open Source**

Comprensión y manejo de Framework de desarrollo como Apache Hadoop y Apache Spark. Además, se tratarán distintas distribuciones que ofrecen empresas distribuidoras y líderes de servicios Big Data; como Hortonworks o Cloudera. Por cada framework, se trabajará cada una de las partes esenciales y otras tecnologías afines que han sido desarrolladas para sus respectivas evoluciones.

Por otro lado, para eventos de baja latencia y como complemento a los anteriores frameworks se abordaran las tecnologías Apache Flume y Apache Kafka. Gracias a estas tecnologías open Source podremos afrontar distintos escenarios gestionando masivas colas de mensajes u otorgando la capacidad de agregar, mover o recopilar ingentes cantidades de datos dentro de nuestro sistema Big Data.

- **BigInsights**

Dentro de las distintas distribuciones de Hadoop del mercado se analiza BigInsights. Se describen los componentes más importantes (e.g. Hadoop, BigSQL, Hbase, y como se pueden utilizar para el análisis de Big Data. Cada descripción de cada componente vendrá acompañada de casos de uso.

- **Stream Computing Frameworks**

Descripción de lo que es Stream Computing para analizar los datos en tiempo real, según se están produciendo sin esperar a almacenarlos. Se muestra distintos casos de uso que pueden ser muy variado: desde analizar el tráfico de una ciudad (Smart Traffic) hasta su utilización en hospitales para detectar situaciones críticas en pacientes. En este aspecto, se trabajarán con las siguientes tecnologías: Spark Streaming, Apache Storm junto con tu extensión Apache Trident y, por último, Apache Flink.

- **Cloud Computing aplicado a Big Data**

El paradigma Cloud Computing es un elemento que cada vez tiene más fuerza dentro del mundo del Big Data. En este bloque se trabajará en detallar y analizar qué es el Cloud Computing, los tipos de nube existentes y las tecnologías disponibles de dos de los grandes proveedores de este servicio: IBM y Amazon. Para la parte de los servicios de IBM se trabajará con la herramienta BlueMix: Software como Servicio y análogamente, se presentará y trabajará con la suite de Amazon.

- **Plataforma Data Science Experience**

IBM Data Science Experience (DSx) es un entorno cloud interactivo y colaborativo donde los data scientists pueden usar multiples herramientas y recursos como RStudio, Jupyter, Python, Scala, Spark, IBM Watson Machine Learning. En este bloque se explicarán los conceptos básicos, algunas de las herramientas disponibles y cómo éstas pueden utilizarse en el entorno de Big Data.

### Asignatura: Tecnológicas Avanzadas

- **Computación Cognitiva**

Los sistemas cognitivos son una nueva forma de máquinas inteligentes que aprenden de la interacción con los usuarios.

Watson es el máximo exponente de la cognición digital: entiende el lenguaje natural, formula hipótesis, se mueve en la ambigüedad propia del ser humano, ofrece respuestas basándose en evidencias, asigna niveles de confianza a sus respuestas, maneja ingentes cantidades de datos (Big Data), está al día y aprende.

Se mostrará cómo funciona Watson y sus principales aplicaciones actuales.

- **Watson Analytics**  
"Converse con sus datos", Watson Analytics utiliza el lenguaje natural, la automatización de la preparación de datos, modelos predictivos, visualizaciones avanzadas, que permiten a las líneas de negocio tener control de sus análisis sin depender de los departamentos de IT.

## Módulo III. Análisis de Negocio

---

### Asignatura: Big Data Analytics

Una vez que sabemos cómo gestionar y almacenar la ingente cantidad de información que es generada, es el momento de analizarlos. Para ello, en esta asignatura se verán técnicas y procedimientos de análisis para datos masivos.

- **Knowledge Data Discovery**  
Se analizarán cada una de las fases del proceso de descubrimiento de conocimiento, analizando tanto el aprendizaje supervisado como el no supervisado.
- **Minería de Texto**  
Descripción de Minería de Texto y su aplicación dentro del mundo de Big Data. Para su entendimiento se muestra una herramienta (Text Analytics) que permite realizar análisis de grandes volúmenes de texto utilizando Hadoop.
- **Social Intelligence**  
En el proceso de Transformación Digital, es absolutamente relevante comprender el disruptivo carácter digital que juegan las personas, así como las posibilidades que tienen a su disposición para entender/conocer/comprender y valorar la oferta y demanda existentes.
- **Prácticas de Data Mining con SPSS Modeler**  
Explicación de la metodología CRISP-DM junto con el panel de desarrollo de modelos de SPSS Modeler. Explicación de cómo diseñar un diagrama; explicación de nodos de fuentes y destinos de datos; explicación de nodos de manipulación de filas y columnas; explicación de nodos de descripción estadística y de gráficos; explicación de nodos de modelización: clustering (K-Means, Bietapico, Kohonen), clasificación/estimación (regresión, logística, árboles CART/CHAID/C5.0, redes neuronales) reglas de asociación)  
Realización de ejercicios asistidos por el profesor; realización de ejercicios enunciados con resultados esperados; taller de realización de un modelo de credit risk scoring

### Asignatura: Business Analytics

- **Modelo de Lifetime value y abandono de clientes**  
Explicación detallada de los modelos de valoración del potencial de un cliente (*lifetime value*) y de abandonos de clientes y su empleo por los departamentos de Marketing para definir las estrategias de fidelización de clientes. Detalle matemático del cálculo del lifetime value a través de una cadena de Markov que proyecta la rentabilidad actual de un cliente a un horizonte probable. El modelo de abandonos será un modelo dinámico que aprenderá de las clasificaciones del pasado. Se realizará una práctica de valoración con un ejercicio en hoja de cálculo, y una definición de las estrategias de gestión de clientes una vez determinados los niveles de potencial y abandono de un cliente.
- **Marketing Contextual I**  
Se debate sobre el Caso Re-elección Obama: Big Data en la captación de 150 millones de votantes  
Se explica el papel y valor del Big Data en el Marketing "extremo-a-extremo", viendo casos concretos como "La vida de un impacto publicitario" y su destilado BigData.

Asimismo, se incluye un caso de Hiper-personalización con modelos de recomendación.

- **Marketing Contextual II**  
Se explica una activación de marketing para un modelo de retención de clientes, basado en un score de fuga y calculando un score de Engagement de cliente. Asimismo, se explica el Big Data de Engagement de Cliente que se pone en juego.  
Todo ello en una clase práctica basada en IBM Campaign Automation donde se realiza una campaña en canales digitales, desde la definición del público objetivo hasta la personalización/recomendación de las comunicaciones, incluyendo la maquetación de las mismas.
- **Optimización de Clientes**  
Se explica cómo optimizar el valor de la cartera de clientes mediante la aplicación la gestión de carteras financieras de Markovich al mundo del Marketing. Se analiza el valor potencial (lifetime value) de cada uno de los clientes y, la evolución conjunta de cada uno de ellos mediante el empleo de una cadena de markov que optimice el valor de la cartera de clientes.
- **Asistentes Cognitivos**  
La evolución de la inteligencia artificial o cognitiva está poniendo de relevancia un nuevo tipo de medio/canal: los chatbots o asistentes virtuales. El Big Data subyacente adquiere un protagonismo relevante para el reconocimiento del lenguaje natural o la expansión del conocimiento.  
Se introduce una tipología de chatbots y se describen varios ejemplos. Incluye el desarrollo práctico de un chatbot con IBM Watson Conversations.
- **Cuadro de Mando Integral / Balanced Scorecard/Mapas Estratégicos**  
Descripción de las metodologías de medición estratégica de Kaplan y Norton: Cuadro de Mando Integral, Balanced Scorecard y Mapas Estratégicos. Las 7 relaciones causa-efecto. Desarrollo de un Proyecto de Cuadro de Mando Integral. Ejemplos reales
- **CU: Retail**  
Descripción de un caso de transformación digital en la industria de Retail, con aproximación a modelos de desintermediación como estrategia para consolidar la relación con el cliente.
- **CU: IOT aplicado al sector del automóvil**  
En este módulo se verán las tecnologías de internet de las cosas, se hará un breve resumen de las procesos existentes, para terminar con la aplicación de las mismas al sector de la automoción, donde las aplicaciones son numerosas y de gran importancia cara a futuro. También se hará una introducción general al machine learning con la aplicación de un caso práctico aplicado a la automoción.