

## INF340-BIG DATA

### Fundamentación:

En los últimos años, la mejora de la eficiencia empresarial tiene una relación directa con la gestión de grandes volúmenes de datos a una mayor velocidad. Es por ello que las empresas necesitan implementar soluciones de Data Management a través de una estrategia integrada que les permita describir, organizar, integrar, gestionar y utilizar datos útiles al conjunto de su estrategia y unidades de negocio.

Big data se refiere a las grandes colecciones de datos, estructurados, o no estructurados, que pueden crecer a volúmenes enormes y a un ritmo tan alto que es complejo manejarlos con las técnicas habituales de los sistemas de bases de datos y las herramientas de análisis existentes hasta hace un tiempo.

La asignatura Big Data está orientada a sacar el máximo provecho de la tecnología frente a programas que muestran lo que la tecnología disponible permite. La tecnología es un instrumento y un servicio que debe contribuir al alcance de los objetivos estratégicos de la empresa.

### Objetivos

---

#### Generales

- Entender qué es el Big Data y el Data Management, su impacto en las organizaciones y los beneficios que puede aportar su adopción.

#### Específicos

- Conocer las capacidades de los nuevos sistemas relacionales en términos de escalabilidad y rendimiento.
- Entender la tecnología Big Data, para qué sirve y como se usa.
- Conocer cómo se construye un sistema para Big Data, y algunas técnicas recientes que ayudan al desarrollo.
- Conocer las interfaces de usuario para Big Data y como se visualizan los datos.
- Entender que es el Machine learning y los algoritmos predictivos para análisis de datos

### Competencias

---

#### Genéricas

- Gestión de la información y el conocimiento
- Soluciones innovadoras basadas en el conocimiento

## Específicas

- Dominio para diseñar soluciones apropiadas a diversos dominios de aplicación, utilizando los principios y métodos propios de la ingeniería
- Habilidad para desarrollar soluciones informáticas aplicando conocimientos de Sistemas informáticos, infraestructura y comunicación.

## Contenidos

---

### **1. Introducción general y ecosistema Big Data**

#### 1.1. ¿Qué es Big Data?

- 1.1.1. La explosión de los datos.
- 1.1.2. Big Data es algo más que muchos datos. Las V's.
- 1.1.3. Oportunidades y retos del Big Data.
- 1.1.4. Análisis de datos tradicional vs Big Data. El científico de datos.
- 1.1.5. Ejemplos de aplicaciones Big Data

#### 1.2. Presente y futuro del Big Data

- 1.2.1. Conceptos fundamentales.
- 1.2.2. Ecosistema Big Data.
- 1.2.3. Introducción a HADOOP.

### **2. Minería de datos y Análisis de datos**

#### 2.1. Integración de datos

- 2.1.1. Fuentes de datos tradicionales y nuevas
- 2.1.2. Información estructurada, semi-estructurada y no estructurada
- 2.1.3. Procesos ETL y técnicas de integración
- 2.1.4. Modelado de datos

#### 2.2. Análisis de datos

- 2.2.1. Transformando datos en información.
- 2.2.2. KPIs (Key Performance Indicators).
- 2.2.3. Herramientas de Business Intelligence (BI)

#### 2.3. Minería de datos

- 2.3.1. Conceptos básicos de Data Mining
- 2.3.2. Modelos descriptivos y predictivos.
- 2.3.3. Técnicas fundamentales: reglas de asociación, clasificación, clustering, patrones.
- 2.3.4. Aplicación a problemas concretos de predicción.

### **3. Técnicas y herramientas de última generación para Big Data**

#### 3.1. Técnicas y metodologías para Big Data

- 3.1.1. Metodologías en proyectos Big Data

3.1.2. El paradigma MapReduce.

3.2. Tecnologías Big Data

3.2.1. Plataformas Big Data

3.2.2. Herramientas de última generación

#### 4. Visualización de la información

4.1. Visualización de datos y Big Data

4.1.1. Conceptos básicos de visualización.

4.1.2. Mejoras y recursos para la visualización tradicional.

4.1.3. Nuevas herramientas gráficas.

4.1.4. Tableros de información.

4.1.5. Visual Discovery.

4.1.6. Visual Analytics.

#### Bibliografía

---

##### **BASICA:**

**Fernández, E. P.** (2017). Big data: eje estratégico en la industria audiovisual. Barcelona, ESPAÑA: Editorial UOC

<http://site.ebrary.com/bibliotecadigital.idm.oclc.org/lib/biblioues21sp/reader.action?ppg=10&docID=11335772&tm=1499196198796>

**Braulio, G. N., & Curto, D. J.** (2015). Customer analytics: mejorando la inteligencia del cliente mediante los datos. Barcelona, ESPAÑA: Editorial UOC.

<http://site.ebrary.com/bibliotecadigital.idm.oclc.org/lib/biblioues21sp/reader.action?ppg=1&docID=11217039&tm=1499196616045>

**Trujillo, J.** (2013) Diseño y explotación de almacenes de datos: conceptos básicos de modelado multidimensional, Ed. ECU

<http://site.ebrary.com/bibliotecadigital.idm.oclc.org/lib/biblioues21sp/reader.action?ppg=3&docID=10751536&tm=1499458821926>

#### **Bibliografía ampliatoria**

**MINERIA DE DATOS A TRAVES DE EJEMPLOS.1ED**

Autor: Perez Editorial: ALFAOMEGA GRUPO EDITOR

<http://www.cuspide.com/9786076221747/Mineria+De+Datos+A+Traves+De+Ejemplos+1Ed/>

**Aplicación de un sistema business intelligence en un contexto big data de una empresa industrial alimentaria.**

<http://eds.a.ebscohost.com/eds/detail/detail?vid=0&sid=85b91960-6377-4e19-8c35-1074c73eb3c3%40sessionmgr4009&bdata=Jmxhbmc9ZXMmc2l0ZT1lZHMtbGl2ZQ%3d%3d#AN=122868897&db=fua>

**Big Data: una exploración de investigaciones, tecnologías y casos de aplicación.**

<http://eds.a.ebscohost.com/eds/detail/detail?vid=0&sid=b1a1b717-ac36-4ce2-9574-da0260267098%40sessionmgr4008&bdata=Jmxhbmc9ZXMmc2l0ZT1lZHMtbGl2ZQ%3d%3d#AN=123557860&db=fua>

**Macías, M., & Gómez, M. (2015).** Introducción a Apache Spark: para empezar a programar el big data. Barcelona, ESPAÑA: Editorial UOC

<http://site.ebrary.com/bibliotecadigital.idm.oclc.org/lib/biblioues21sp/reader.action?ppg=1&ocID=11217141&tm=1499196471385>

**RECURSOS:**

---

**Contenidos y materiales multimediales en plataforma.**

**Bibliografía obligatoria.**

**Buscadores vinculados a la profesión.**

**Carga Horaria**

---

La carga horaria semestral dedicada al dictado de clases teóricas y prácticas de la asignatura es de 43 horas reloj, asignándosele del total, 16 horas a las actividades prácticas.

**Metodología**

---

Las clases teóricas se organizarán tomando como base el material incluido en la bibliografía seleccionada. Las clases se desarrollarán a partir de las explicaciones relativas a los temas de la bibliografía por parte del profesor y de las intervenciones por parte de los alumnos dirigidas a aclarar conceptos o a profundizar en determinados temas que puedan resultar de interés. Se utilizará material didáctico de apoyo.

En las clases prácticas se revisarán investigaciones actuales y se discutirán los aspectos metodológicos más relevantes. En las clases prácticas el profesor asumirá la tarea de introducir y supervisar el desarrollo de la actividad programada, correspondiendo al alumno el papel protagonista al participar directamente en la revisión y discusión de las investigaciones, y en la resolución de las actividades propuestas.

### **Forma de Evaluación**

---

En esta materia se evalúa el proceso de aprendizaje del alumno a través de las tareas realizadas por ellos en las fechas estipuladas institucionalmente.

La aprobación, se determina en función al cumplimiento de los criterios de evaluación especificados. Estos criterios se basan en tres áreas claves/críticas: cumplimiento de tiempos de entrega (según cronograma), cumplimiento de las especificaciones de forma del entregable y nivel de logro de las competencias vinculadas a esta materia.

La condición de regularidad de esta materia, se define según lo establecido en el Reglamento Institucional, de acuerdo a la condición del alumno (Regular, Libre, Promocionado).