

# Fuentes de Información

**Abril – junio 2020**

---

## OBJETIVOS

La gestión y análisis de grandes volúmenes de información (Big Data and Analytics) comienza por identificar y coleccionar los datos que son relevantes para la organización. Este curso presenta tecnologías alternas a los sistemas empresariales, las cuales juegan un papel cada vez más importante como fuentes generadoras de grandes volúmenes de datos. Se trata de tecnologías con flujos de datos a gran velocidad, de diversa índole, que deben ser capturados y analizados en poco tiempo, en ocasiones en tiempo real.

Al final del curso, el alumno será capaz de comprender las características más relevantes de distintas fuentes de datos y los fundamentos de las herramientas que pueden ser utilizadas para su captura y análisis.

## METODOLOGÍA

La dinámica estará basada principalmente en las exposiciones teóricas por parte del profesor, las cuales estarán complementadas por lecturas de artículos de actualidad.

La mayoría de los temas tendrá un componente práctico en el que se realizarán proyectos en maquetas de experimentación para afianzar los conceptos teóricos.

Se fomentará la discusión de temas en clase, en foros electrónicos y a través de la exposición de proyectos.

## TEMARIO TENTATIVO

- **Grandes volúmenes de información**
  - Big Data (y Analítica). Conceptos básicos
  - Volumen, **VARIEDAD**, Velocidad, Veracidad
  - Conceptos básicos de Hadoop y el modelo MapReduce
  - Pig y Hive. Conceptos básicos
- **Stream processing**
  - Procesamiento de datos en tiempo real
  - Arquitecturas y modelos: Streams, S4, Storm, Spark
- **Redes**
  - Conceptos básicos. Grados de separación, de conectividad. Erdos-Renyi, Barabasi-Albert
  - Modelos. Búsqueda descentralizada
- **Texto**
  - Textos no estructurados y semi-estructurados (documentos, correo electrónico, micro blogs, ...)
  - Bases no estructuradas. Coexistencia datos no estructurados y estructurados
  - Principios básicos de análisis de texto
- **Redes sociales en Internet**
  - Facebook, Twitter
  - Análisis en redes sociales (sentimiento, propagación de información)
- **Internet de las cosas**
  - Redes M2M, CDR, RFID

- Redes de sensores
- Datos geoespaciales
- Plataformas y tecnologías para IoT
- El ecosistema Fiware

## EVALUACIÓN

**Calificación** = 0.25\*Examen parcial  
 + 0.25\*Examen final  
 + 0.25\*Controles, cuestionarios, tareas  
 + 0.25\*Proyecto final

## BIBLIOGRAFÍA

- Inmon, W., Nesavich, A., Tapping into Unstructured Data, Prentice Hall, 2008
- 
- Kraut, R. E. & Resnick, P. (2012). Building successful online communities: Evidence-based social design. Cambridge, MA: MIT Press.
- Easley, D & Kleinberg, J. (2010), Networks, Crowds, and Markets: Reasoning About a Highly Connected World, Cambridge University Press
- Newman, M.E.J. (2010), Networks: An Introduction, Oxford University Press
- Barabasi, A. (2003), Linked, Plum,
- Faludi, R. (2010), Building wireless sensor networks with ZigBee, Arduino and Processing, O'Reilly
- White, T. (2012), Hadoop. The definitive guide. O'Reily
- InfoSphere BigInsights Essentials using Apache Hadoop
- Zikopoulos, Deroos, Parasuraman, Deutsch, Corrigan, Giles (2013) Harness the Power of Big Data.McGraw Hill