

FUNDAMENTOS DE PROYECTOS BIG DATA

Nº de Créditos: **3 ECTS**
Primer Semestre
Primer Curso

EQUIPO DOCENTE

Gallego, Jenaro

Arquitecto de Tecnologías de
Información en Big Data y
Analítica de Negocio
IBM

Coordinador

Reina, Luis

Especialista en Big Data y
Analytics
IBM

Moreno, Víctor

Escuela Politécnica Superior
UAM

González, Iván

Escuela Politécnica Superior
UAM

Sánchez, Juan José

Escuela Politécnica Superior
UAM

Sirvent, Pilar

Arquitecto de Tecnologías de
Información en Big Data y
Analítica de Negocio
IBM

Canivell, Cristian

Especialista en Tecnologías de
Información
IBM

Martín Prado, Alberto

Arquitecto de Tecnologías de
Información. Client Technical
Advisor
IBM

OBJETIVOS

Los proyectos de Big Data suponen un reto para las empresas involucradas tanto las que van a actuar como usuarios finales como para aquellas que llevan a cabo la realización del proyecto. En muchos casos, además, es una única la que desempeña ambos papeles. Por un lado el desarrollo y despliegue de la solución exige una metodología de proyecto muy flexible y ágil (Metodología Ágil) y por otro, el volumen y variabilidad de la información ponen a prueba los procesos que la empresa tiene en la gestión de los

datos: conocer con qué información se cuenta en todo momento y en qué estado de su ciclo de vida (Gobierno de la Información).

Adicionalmente a estos aspectos la tecnología es otra dimensión con una especial relevancia en estos proyectos. Es innegable el peso específico tan importante que está teniendo el Código Libre. No todos los proyectos de Big Data necesitan involucrar Código Libre pero, por otro lado, es innegable que muchos de los procesos que se tienen que desarrollar pueden ser cubiertos con componentes de Código Libre. El más representativo en todo este entorno es Hadoop y su ecosistema (decenas de componentes de software compatibles con Hadoop). En esta asignatura comenzaremos a conocer toda esta tecnología.

PROGRAMA DETALLADO

- Metodología Ágil para el desarrollo de software
- Ciclo de vida del dato
- Proyecto Apache Hadoop
 - HDFS
 - Modelo de Programación MapReduce
- Desarrollo de aplicaciones (Pig, Jaql, Scala, Python)
 - Programación básica en Java con Hadoop
 - Programación básica en Python con Hadoop
 - Programación desde lenguajes de alto nivel (Pig, Hive)
 - Programación con R en entorno hadoop (BigR)
- Examen

BIBLIOGRAFÍA

1. Metodología Agile for Dummies (https://www-01.ibm.com/marketing/iwm/iwm/web/signup.do?source=swg-rtl-sd-wp&S_PKG=500026415)
2. Information Governance (<http://www.redbooks.ibm.com/Redbooks.nsf/RedbookAbstracts/sg248165.html?Open>)
3. Wikipedia Apache Hadoop. (<http://wiki.apache.org/hadoop/>)

MÉTODOS DOCENTES

- Lección magistral
- Resolución de problemas
- Prácticas de laboratorio
- Estudio de casos

MÉTODOS DE EVALUACIÓN

- Asistencia a clase: **10%**
- Evaluación continua: **40%**
- Examen final: **50%**