

ECOSISTEMA DE SOPORTE A PROYECTOS DE BIG DATA

Nº de Créditos: **3 ECTS**
Primer Cuatrimestre
Segundo Curso

EQUIPO DOCENTE

Gallego Gómez, Jenaro

Arquitecto de Tecnologías de
Información

IBM

Coordinador

Reina Julia, Luis

Arquitecto en Big Data y
Analytics

IBM

Alonso Delgado, Ignacio

Especialista en Big Data y
Analytics

IBM

Cano Bailén, Francisco Javier

Especialista en Big Data y
Analytics

IBM

Moreno Navarro, Isaac

Especialista en Big Data y
Analytics

IBM

OBJETIVOS

- Los proyectos de Big Data suponen considerar todos los pasos del ciclo de vida analítico del dato: Aprovisionar, Persistir, Analizar y Consumir o explotar el resultado. En esta asignatura se revisarán cada uno de los pasos tomando como ejemplos algunos de los componentes más representativos de la industria alrededor del entorno Hadoop.
- Adicionalmente, los entornos analíticos en tiempo real presentan una casuística especial de cada uno de los pasos mencionados. Se revisará en especial la solución InfoSphere Streams de IBM.

PROGRAMA DETALLADO

- Aprovisionamiento de datos:
 - En entorno Hadoop: Apache Flume, Apache Sqoop
 - Kafka
- Persistencia y explotación
 - Repaso de Hive. Interfaz SerDe
 - HBase
 - Avro
- Planificación de trabajos en entorno Hadoop.
 - Planificación con Oozie
- Análisis de datos estáticos.
 - Entorno de desarrollo y colaboración IBM Data Science Experience
- Consumo y Explotación
 - Buscadores, visualizadores: SolR, ElasticSearch
 - Otras soluciones
- Análisis de datos en movimiento
 - Stream Computing
 - Infosphere Streams

BIBLIOGRAFÍA

- Open source Big Data for the impatient. Hadoop example: Hello World with Java, Pig, Hive, Flume, Fuse, Oozie, and Sqoop with Informix, DB2, and MySQL.
<https://www.ibm.com/developerworks/data/library/techarticle/dm-1209hadoopbigdata/>
- Repaso de conocimiento de Hive. Build a data library with Hive.
<https://www.ibm.com/developerworks/library/bd-hivelibrary/index.html>
- Labs de introducción a HBase: <https://developer.ibm.com/hadoop/docs/getting-started/tutorials/hbase-intro-lab/hbase-intro-lab-1-overview/>
- Big data serialization using Apache Avro with Hadoop:
<https://www.ibm.com/developerworks/library/bd-avrohadoop/index.html>
- Oozie workflow scheduler with Hadoop:
<https://www.ibm.com/developerworks/library/bd-ooziehadoop/index.html>
- Quick Overview IBM Datascience eXperience:
<https://datascience.ibm.com/blog/welcome-to-the-data-science-experience/>
- Lucene Apache Solr. <http://lucene.apache.org/solr/>
- An Introduction to IBM InfoSphere Streams:
<https://www.ibm.com/developerworks/library/bd-streamsintro/index.html>

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%**