

Programa detallado curso:

**"Big Data y Supercomputación: Transformando datos en conocimiento"**

Duración: 23 horas (20 horas de sesiones + 3 horas de casos de éxito).

Horario: del 8 al 14 de junio, de 9 a 14 horas.

Lugar: Laboratorio I de Informática del Centro Universitario de Mérida.

**Bloque 1: Presentación** (1 hora)

- Presentación de LUSITANIA I y II.
- Entorno de trabajo.
- Proyectos de la Fundación.
- Introducción a *big data* y ejemplos para mostrar la problemática del *big data*.
- Retos al procesar datos *big data*.

**Bloque 2: Supercomputación y framework de Hadoop** (11 horas)

- MV de Cloudera a través de OpenNebula.
- Hadoop: Justificación y características.
- HDFS: Características, arquitectura, rendimiento, configuración, lectura y escritura.
- Práctica de Hadoop: clúster HDFS y uso y administración del mismo.
- Módulos básicos del ecosistema de Apache para Hadoop.
- Casos prácticos de recolección de datos para su almacenamiento en el clúster.
- Práctica de consultas a datos del clúster HDFS.
- Procesamiento de datos en tiempo real.
- Spark: motivación y servicios.
- Transformaciones y Acciones en Spark.
- Planificador de DAGs (Directed Acyclic Graph).
- Administración del HistoryServer y servicios de Spark.

**Bloque 3: Data Science** (8 horas)

- Metodología y buenas prácticas para el procesamiento de datos científicos.
- Herramientas para la exploración de datos.
- Detección de patrones en conjuntos de datos científicos.
- Práctica con Spark y estrategias de definición de transformaciones.
- Machine learning: conceptos básicos.
- Práctica: ML en Spark con MLib (Machine Learning Library) y bootstrapping.

**Bloque 4: Casos de éxito con Big Data** (3 horas)

- CénitS: ¿Pig Data, cómo que Pig Data?, ¿pero no era Big Pata?
- CénitS: Proyecto ConSumar.
- Analytics: Optimización de inversión de Capex para implantar 4G en función de patrones de movilidad de los clientes de telefonía de la ciudad.

Programa resumido curso:

**"Big Data y Supercomputación: Transformando datos en conocimiento"**

Duración: 23 horas (20 horas de sesiones + 3 horas de casos de éxito).

Horario: del 8 al 14 de junio, de 9 a 14 horas.

Lugar: Laboratorio I de Informática del Centro Universitario de Mérida.

**Bloque 1: Presentación** (1 hora)

**Bloque 2: Supercomputación y *framework* de Hadoop** (11 horas)

- Clúster de Hadoop con Cloudera y OpenNebula.
- Flujo de trabajo con *big data* a través de distintos módulos del ecosistema de Apache.

**Bloque 3: *Data Science*** (8 horas)

- Buenas prácticas para el procesamiento de grandes volúmenes de datos.
- Obtención de conocimiento a partir del procesamiento de grandes volúmenes de datos.
- Machine Learning y Spark.

**Bloque 4: Casos de éxito con *Big Data*** (3 horas)

- CénitS: ¿Pig Data, cómo que Pig Data?, ¿pero no era Big Pata?
- CénitS: Proyecto ConSumar.
- Analytics: Optimización de inversión de Capex para implantar 4G en función de patrones de movilidad de los clientes de telefonía de la ciudad.