

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:

- Bloques 1, 2 y 3: Lunes 12, martes 13, miércoles 14 y lunes 19 de 16 a 21 horas.
- Bloque 4: Jueves 15 (de 16 a 19 horas, durante las 8ª Jornadas CénitS).

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 / Puesta en valor del servicio (9 horas)

- *Cloud computing* y OpenNebula.
- Hadoop: Justificación y características.
- Módulos básicos del ecosistema de Apache para Hadoop.
- HDFS: Características, arquitectura, rendimiento, configuración, lectura y escritura.
- MapReduce / YARN: Flujo de información y estrategias de definición de *map* y *reduce*.
- Práctica: Adición de nuevo nodo al clúster de Hadoop.
- Casos prácticos de recolección de datos para su almacenamiento en el clúster.
- Procesamiento de datos en tiempo real.

Bloque 3: *Data Science* (10 horas)

- Metodología y buenas prácticas para el procesamiento de datos científicos.
- Programación con R.
- Modelo de programación *map / reduce* con R.
- Herramientas para la detección de patrones en conjuntos de datos científicos.
- *Machine learning*: conceptos básicos.
- Implementación de algoritmos de *machine learning* en Hadoop.

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

- Nuk Consultants: Big Data en el sector turístico.
- ViewNext: Detección de Fraude a la Seguridad Social: Empresas Ficticias.
- CénitS: Aplicación de técnicas Big Data a la predictibilidad del consumo energético. Proyecto ConSumar.

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:

- Bloques 1, 2 y 3: Lunes 12, martes 13, miércoles 14 y lunes 19 de 16 a 21 horas.
- Bloque 4: Jueves 15 (de 16 a 19 horas, durante las 8ª Jornadas CénitS).

Bloque 1: Presentación (1 hora)

Bloque 2: Supercomputación y *framework* de Hadoop / Puesta en valor del servicio (9 horas)

- Diseño, configuración y despliegue de servicios de un clúster de Hadoop.
- OpenNebula para el despliegue de servicios de procesamiento de datos *big data*.
- Flujo de trabajo con *big data* a través de distintos módulos del ecosistema de Apache.

Bloque 3: *Data Science* (10 horas)

- Buenas prácticas para el procesamiento de datos científicos.
- *Data mining* para descubrir patrones en conjuntos de datos científicos con R.
- Adaptación de algoritmos de *data mining* para datos en clústers de Hadoop.

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

- Nuk Consultants: Big Data en el sector turístico.
- ViewNext: Detección de Fraude a la Seguridad Social: Empresas Ficticias.
- CénitS: Aplicación de técnicas Big Data a la predictibilidad del consumo energético. Proyecto ConSumar.