

Big Data, métodos y soluciones prácticas para el análisis de datos

Curso práctico de 5 días - 35h
Ref.: BID - Precio 2025: 2 520€ sin IVA

Este curso le permitirá comprender los retos y las aportaciones de los macrodatos o Big Data, así como las tecnologías para implementarlos. Aprenderá a integrar volúmenes masivos de datos estructurados y no estructurados a través de un ETL y a analizarlos con modelos estadísticos y cuadros de control dinámicos.

OBJETIVOS PEDAGÓGICOS

Al término de la formación, el alumno podrá:

Comprender los conceptos y la contribución de los Big Data a los problemas de la empresa

Comprender el ecosistema tecnológico necesario para llevar a cabo un proyecto de Big Data

Adquirir competencias técnicas para gestionar flujos masivos de datos complejos, no estructurados

Implementar modelos de análisis estadísticos para responder a las necesidades de la empresa

Comprender una herramienta de visualización de datos para reproducir análisis dinámicos

Configurar una plataforma Hadoop y sus componentes básicos, utilizar un ETL para gestionar los datos, crear modelos de análisis y cuadros de control.

PROGRAMA

última actualización: 08/2024

1) Comprender los conceptos y los retos de los Big Data

- Orígenes y definición de Big Data.
- Las cifras clave del mercado en el mundo y en Francia.
- Los retos de los Big Data: Rendimiento de la inversión, organización y confidencialidad de los datos.
- Un ejemplo de arquitectura de Big Data.

2) Tecnologías de Big Data

- Descripción de la arquitectura y componentes de la plataforma Hadoop.
- Modos de almacenamiento (NoSQL, HDFS).
- Principios de funcionamiento de MapReduce, Spark, Storm...
- Principales distribuciones del mercado (Hortonworks, Cloudera, MapR, Elastic Map Reduce y Biginsights).
- Instalar una plataforma Hadoop.
- Las tecnologías del datascientist.
- Presentación de tecnologías específicas para Big Data (Tableau, Talend, Qlikview, etc.).

Ejercicio : Instalación de una plataforma Big Data Hadoop (a través de Cloudera QuickStart u otra).

3) Gestionar datos estructurados y no estructurados

- Principios de funcionamiento del sistema de archivos distribuido Hadoop (HDFS por su sigla en inglés).

PARTICIPANTES

Minería de datos, investigadores estadísticos, desarrolladores, gestores de proyecto y consultores de inteligencia empresarial.

REQUISITOS PREVIOS

Conocimientos básicos de modelos relacionales, estadística y lenguajes de programación. Conocimientos básicos de los conceptos de la Inteligencia Empresarial (Business Intelligence).

COMPETENCIAS DEL FORMADOR

Los expertos que imparten la formación son especialistas en las materias tratadas. Han sido validados por nuestros equipos pedagógicos, tanto en el plano de los conocimientos profesionales como en el de la pedagogía, para cada curso que imparten. Cuentan al menos con entre cinco y diez años de experiencia en su área y ocupan o han ocupado puestos de responsabilidad en empresas.

MODALIDADES DE EVALUACIÓN

El formador evalúa los progresos pedagógicos del participante a lo largo de toda la formación mediante preguntas de opción múltiple, escenificaciones de situaciones, trabajos prácticos, etc.

El participante también completará una prueba de posicionamiento previo y posterior para validar las competencias adquiridas.

MEDIOS PEDAGÓGICOS Y TÉCNICOS

- Los medios pedagógicos y los métodos de enseñanza utilizados son principalmente: ayudas audiovisuales, documentación y soporte de cursos, ejercicios prácticos de aplicación y ejercicios corregidos para los cursillos prácticos, estudios de casos o presentación de casos reales para los seminarios de formación.
- Al final de cada cursillo o seminario, ORSYS facilita a los participantes un cuestionario de evaluación del curso que analizarán luego nuestros equipos pedagógicos.
- Al final de la formación se entrega una hoja de presencia por cada media jornada de presencia, así como un certificado de fin de formación si el alumno ha asistido a la totalidad de la sesión.

MODALIDADES Y PLAZOS DE ACCESO

La inscripción debe estar finalizada 24 horas antes del inicio de la formación.

ACCESIBILIDAD DE LAS PERSONAS CON DISCAPACIDAD

¿Tiene alguna necesidad específica de accesibilidad? Póngase en contacto con la Sra. FOSSE, interlocutora sobre discapacidad, en la siguiente dirección psh-accueil@orsys.fr para estudiar de la mejor forma posible su solicitud y su viabilidad.

- Importar datos externos a HDFS.
- Realizar consultas SQL con HIVE.
- Usar PIG para procesar datos.
- El principio de ETL (Talend...).
- Gestión de streaming de datos masivo (NIFI, Kafka, Spark, Storm, etc.)

Ejercicio : Implementación de flujos de datos masivos.

4) Técnica y métodos de análisis de Big Data

- El aprendizaje automático, un componente de la inteligencia artificial.
- Descubrir las tres familias: Regresión, clasificación y agrupación.
- Preparación de datos (preparación de datos, ingeniería de características).
- Generar modelos en R o Python.
- Aprendizaje por conjuntos.
- Descubrir las herramientas disponibles en el mercado: Jupyter Notebook, Dataiku, Amazon Machine Learning, etc.

Ejercicio : Realización de análisis con una de las herramientas estudiadas.

5) Visualización de datos y casos de uso concretos

- Definir la necesidad de la visualización de datos.
- Análisis y visualización de datos.
- ¿Pueden utilizarse todos los tipos de datos en DataViz?
- Las herramientas DataViz del mercado.

Ejercicio : Instalación y utilización de una herramienta de visualización de datos para crear análisis dinámicos.

6) Conclusión

- Lo que hay que recordar.
- Síntesis de buenas prácticas.
- Bibliografía.

FECHAS

Contacto