



Introducción a la Ciencia de
Datos Big Data

27 Horas

Código Sence 12-37-96-38-02

OBJETIVO GENERAL

Al finalizar el curso, el participante será capaz de:

- Conocer y entender los fundamentos de la Ciencia de Big Data
- Evaluar distintas estructuras de datos que componen lo que se denomina Big Data, para el diseño de proyectos de Big Data y aplicación de modelos y herramientas analíticas aplicadas que apoyen la toma de decisiones en propias en negocios
- Entender el Big Data y desarrollar habilidades orientadas a conocer los desafíos de la digitalización en las organizaciones
- Diseñar sistemas de Big Data para desarrollar análisis de datos orientados a tomar decisiones estratégicas y tácticas en la empresa



DIRIGIDO A

Profesionales de distintas áreas del conocimiento como: psicólogos, sociólogos, ingenieros comerciales, ingenieros civiles, estadísticos, etc. que trabajan o pretenden trabajar como asistentes o jefes de proyectos en empresas que enfrentan desafíos como la digitalización, el Big Data. Así también está dirigido a analistas, planificadores estratégicos y profesionales de empresas, agencias de marketing y publicidad, consultoras, organismos públicos y privados, con o sin fines de lucro, que requieran y hagan uso de análisis de datos respecto de clientes, consumidores, personas y máquinas para la toma de decisiones estratégicas y tácticas.

FUNDAMENTACION TÉCNICA

Cuando hablamos de Big Data nos referimos a conjuntos de datos o combinaciones de conjuntos de datos cuyo tamaño (volumen), complejidad (variabilidad) y velocidad de crecimiento (velocidad) dificultan su captura, gestión, procesamiento o análisis mediante tecnologías y herramientas convencionales, tales como bases de datos relacionales y estadísticas convencionales o paquetes de visualización, dentro del tiempo necesario para que sean útiles.

Lo que hace que Big Data sea tan útil para muchas empresas es el hecho de que proporciona respuestas a muchas preguntas que las empresas ni siquiera sabían que tenían. En otras palabras, proporciona un punto de referencia. Con una cantidad tan grande de información, los datos pueden ser moldeados o probados de cualquier manera que la empresa considere adecuada. Al hacerlo, las organizaciones son capaces de identificar los problemas de una forma más comprensible.

METODOLOGÍA

- El curso se imparte en modalidad presencial, por medio de charlas interactivas apoyadas por medios electrónicos y audiovisuales.
- Metodología participativa y con énfasis práctico.
- Se espera que el alumno incorpore los conocimientos y competencias, identificando y relacionando las prácticas habituales de su rol en la organización con los contenidos de la capacitación, dando énfasis a los elementos prácticos y casuísticos conforme a las responsabilidades de su cargo.

CONTENIDOS DEL CURSO

| MÓDULOS | CONTENIDOS |
|--|--|
| Módulo 1: : Introducción al Big Data | <ul style="list-style-type: none">1.1.- ¿Qué es data science?1.2.- Historia y concepto1.3.- Contexto Big Data1.4.- Relación entre Big Data y Business Intelligence (BI) |
| Módulo 2: Introducción al Data Analytics Thinking | <ul style="list-style-type: none">2.1.- La ubicuidad de las oportunidades con datos2.2.- Data Science, Ingeniería y toma de decisiones basadas en datos2.3.- Procesamiento de datos y "Big Data"2.4.- Datos y Data-Science capability as a strategic asset2.5.- Data mining y Data science |
| Módulo 3: Problemas de Negocios y Soluciones desde la Ciencia de Datos | <ul style="list-style-type: none">3.1.- Conceptos fundamentales3.2.- De los problemas del negocio a las metas de la minería de datos3.3.- Métodos supervisados versus no supervisados3.4.- El proceso de la minería de datos3.5.- Entendiendo el negocio3.6.- Entendiendo los datos3.7.- Preparando los datos3.8.- Modelamiento3.9.- Evaluación3.10.- Despliegue3.11.- Implicaciones de manejar un equipo de científicos de datos3.12.- Otras técnicas y tecnologías3.13.- Estadísticas3.14.- Data3.15.- Warehousing |

| | |
|--|---|
| Módulo 4: Escalando la Manipulación de Datos | <ul style="list-style-type: none">4.1.- Bases de datos y álgebra relacional4.2.- Bases de datos en paralelo4.3.- El procesamiento de consultas en paralelo4.4.- El análisis en la base de datos |
| Módulo 5: Herramientas | <ul style="list-style-type: none">5.1.- Hadoop5.2.- Concepto de MapReduce5.3.- Relaciones en la base de datos5.4.- Algoritmo5.5.- Exenciones5.6.- Lenguajes5.7.- Base de datos no relacionales5.8.- Clave valor y NoSQL5.9.- Tradeoffs SQL y NoSQL |
| Módulo 6: Análisis | <ul style="list-style-type: none">6.1.- Los temas de modelización estadística6.2.- Conceptos básicos6.3.- Diseño de experimentos6.4.- Las trampas6.5.- Machine Learning6.6.- Aprendizaje supervisado (reglas, árboles, bosques, el vecino más cercano, la regresión)6.7.- Optimización (descenso de gradiente y variantes)6.8.- Aprendizaje no supervisado |
| Módulo 7: Comunicando Resultados | <ul style="list-style-type: none">7.1.- Visualización7.2.- Productos de datos7.3.- Análisis de datos visuales7.4.- Procedencia7.5.- Privacidad7.6.- Ética7.7.- Gobierno |

Módulo 8: Temas Especiales

- 8.1.- Análisis del gráfico
- 8.2.- Estructura
- 8.3.- Recorridos
- 8.4.- PageRank
- 8.5.- Detección de la comunidad
- 8.6.- Consultas recursivas
- 8.7.- Web semántica