

Curso gratuito de

CIENCIA DE DATOS II

DESCRIPCIÓN

El Curso de Ciencia de Datos II

Este curso te proporcionará un conocimiento más profundo y práctico de Big Data, Hadoop, Spark y PySpark.

A través de una combinación de teoría y práctica, explorarás una amplia gama de temas clave que incluyen el procesamiento de grandes volúmenes de datos, la administración eficiente de conjuntos de datos masivos y el análisis avanzado mediante Spark y PySpark.

A medida que avances en el curso, desarrollarás una comprensión sólida de cómo utilizar estas herramientas para extraer información valiosa y realizar análisis de datos a gran escala.

Nuestro enfoque práctico te permitirá aplicar tus conocimientos en escenarios del mundo real, resolviendo desafíos y problemas comunes que puedes encontrar en tu carrera profesional.

Además, trabajarás en proyectos prácticos que te ayudarán a fortalecer tus habilidades y a construir un portafolio sólido.



INFORMACIÓN GENERAL



Horario*

Colocar los días y horas



Modalidad

Virtual (zoom)



Duración

6 Sesiones



Horas

24 horas

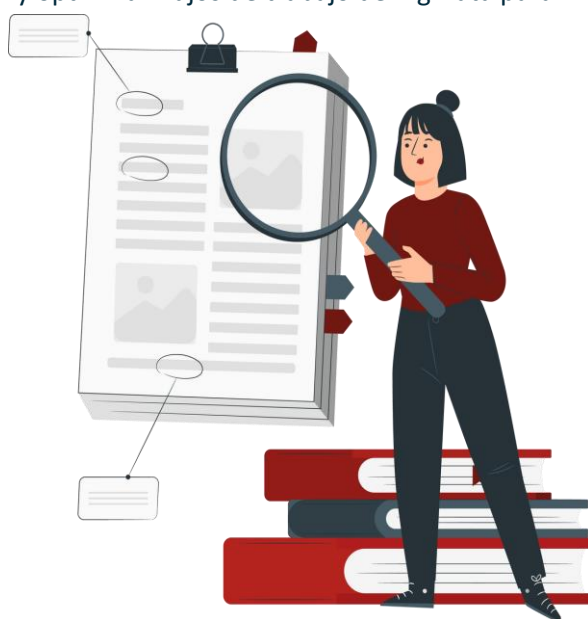
¿QUÉ APRENDERÁS EN ESTE CURSO?



Al finalizar el curso, estarás preparado para abordar desafíos complejos en el ámbito del Big Data y aplicar eficientemente las tecnologías de Hadoop, Spark y PySpark. Tendrás la capacidad de diseñar y gestionar sistemas de almacenamiento y procesamiento de grandes volúmenes de datos, implementar soluciones eficientes para el procesamiento paralelo y distribuido de datos utilizando Hadoop, utilizar Spark y PySpark para realizar análisis avanzados y procesamiento de datos en tiempo real, aplicar técnicas y algoritmos de procesamiento de datos masivos para obtener información valiosa, resolver problemas complejos relacionados con la manipulación y análisis de grandes conjuntos de datos, desarrollar y optimizar flujos de trabajo de Big Data para aplicaciones del mundo real. Este curso te proporcionará las habilidades esenciales y la experiencia práctica necesaria para destacarte en el campo del Big Data y enfrentar proyectos desafiantes con confianza.



Los materiales del curso los podrán encontrar en UNI Virtual, así como los links de las grabaciones de las clases.



DIRIGIDO A

Este curso está dirigido a alumnos UNI de pregrado, matriculados en el semestre 2023-2, con conocimientos básicos de cálculo, álgebra lineal y estadística.

BENEFICIOS

- ☒ Acceso al aula virtual (UNI Virtual).
- ☒ Grabaciones de las clases.
- ☒ Materiales descargables.

Certificado

Al haber aprobado el curso con un promedio ponderado no menor a 12 se le otorga al participante un certificado a nombre de la Universidad Nacional de Ingeniería.

Los alumnos que no hayan aprobado el curso, podrán obtener una constancia de asistencia del curso, a nombre de la Universidad Nacional de Ingeniería, siempre y cuando hayan asistido al 75% de las clases.



TEMARIO

DEL CURSO

SESIÓN 1:

Introducción al Big Data y Contexto Actual

1. Definición y características de Big Data
2. Importancia y aplicaciones en la actualidad
3. Introducción a las tecnologías de Hadoop y Spark
4. Práctica 1: Configuración del entorno

SESIÓN 2:

Fundamentos de Hadoop y su Arquitectura

1. Estructura y componentes esenciales de Hadoop
2. Funciones de Job Tracker y Name Node
3. Implementación de Balance de Carga
4. Transición a YARN (Yet Another Resource Negotiator)
5. Práctica 2: Configuración y operación básica

SESIÓN 3:

MapReduce en Profundidad

1. Detalles del proceso de MapReduce
2. Roles y funciones de Mapper y Reducer
3. Estrategias de optimización
4. Práctica 3: Desarrollo de aplicaciones MapReduce

SESIÓN 4:

Spark: Potenciando el Procesamiento de Datos

1. Ventajas y características clave de Apache Spark
2. Componentes esenciales y su función
3. Arquitectura de procesamiento distribuido
4. Práctica 4: Ejecución de tareas básicas en Spark

SESIÓN 5:

Programación Efectiva en Spark con PySpark

1. Trabajo con RDD y Agregaciones en Spark
2. Uso de Merging y Cross Joins en PySpark
3. Práctica 5: Desarrollo de aplicaciones prácticas.

SESIÓN 6:

Introducción a PySpark

1. Comparación con Scala y beneficios
2. Configuración y entorno PySpark
3. Práctica 6: Migración de aplicaciones de Scala a PySpark

DOCENTE

Juan Chipoco

Docente de Ciencia de Datos en el Oficina de tecnologías de la información de la Universidad Nacional de Ingeniería - OTI UNI.

Data Scientist con amplia trayectoria, especializado en el diseño de soluciones de ciencia de datos y con experiencia global en la ejecución exitosa de proyectos en diversos entornos. Ha liderado iniciativas en Perú, Bolivia, Inglaterra y Bahrein, en la implementación de soluciones innovadoras.





**UNIVERSIDAD
NACIONAL DE
INGENIERÍA**

OTI  UNI



**TRANSFORMACIÓN
digital**