



## WORKSHOP

# BIG DATA

Las organizaciones poseen un activo fundamental para obtener ventajas competitivas, este activo es cada vez más abundante y se genera a altas velocidades, lo cual se conoce como stormdata. El reto de las organizaciones es contar con tecnologías y profesionales que sean capaces de trabajar estos datos y obtener conclusiones de impacto para el negocio. La presente secuencia de contenidos pretende formar profesionales de datos, con bases teóricas sólidas y habilidades para el manejo de las tecnologías avanzadas de Big data.

### • Conocimientos Previos

Experiencia en programación básica  
Experiencia trabajando datos con SQL

### • El participante al final del curso sera capaz de :

Entender e interactuar con las tecnologías más avanzadas de Big Data sobre plataformas de nubes públicas como Azure, GCP y AWS. Poder construir un storage Delta Lake y respectivos data pipelines sobre la plataforma Databricks utilizando scala y pyspark como lenguaje de programación. Implementar una arquitectura unificada Lambda para incluir procesos en Real time.

- **Dirigido a:** profesionales y/o estudiantes de últimos ciclos con o sin experiencia en el análisis de datos o implementación de proyectos de Inteligencia de Negocios.

- **Duración del curso: 36 Hrs. / 6 sesiones**

### Evaluación

Será totalmente práctica. Se realizarán entre 4 o 5 prácticas de las cuales se eliminará la nota más baja y se obtendrá un promedio (PP). Durante la última sesión se realizará un examen final (EF), el cual se promediará con la nota de prácticas y de esta manera se tendrá la calificación final

PROMEDIO DE PRÁCTICAS:

$PP = (PR1 + PR2 + PR3 + PR4 - \text{MENOR (PR)})$

NOTA FINAL:

$NF = PP + EF$

### Modalidad Online

requisitos minimos para la PC:

- Procesador - 2.0 GHZ.
- 4GB RAM.
- Espacio Libre 16GB en el disco duro.
- Conexión a internet estable.

Para más información:

E-mail: [sisuni.info@uni.edu.pe](mailto:sisuni.info@uni.edu.pe)

Tlf.: 200 - 9060 Opción 1 / WhatsApp: 970-063-319 / 943-229-860

Conéctate a nuestras  
redes sociales



# “Aumenta tus conocimientos, desarrolla nuevas habilidades y construye hoy tu futuro”.

## 1. BIG DATA FUNDAMENTOS

- Conceptos Core.
- Plan de Infraestructura Cloud y Onprem.
- Servicios Cloud AWS, AZURE y GCP.
- Arquitecturas de Big Data.
- Preparación de un cluster de procesamiento.
- Preparación de un sistema de almacenamiento.
- Lectura On memory.

## 2. PATRONES DE PROCESAMIENTO

- Sistema de archivos en cloud
- Hive como motor de procesamiento estructurado
- Esquema OnRead y creación de Tablas externas
- Binarización de datos delta
- Procesamiento Estructurado
- Procesamiento Funcional con Spark
- Scala Transformations
- Almacenamiento de DataFrames

## 3. IMPLEMENTACIÓN DE UN DELTA LAKE

- Definición de la arquitectura.
- Definición de la estrategia de procesos.
- Despliegue de taxonomías.
- Poblado de la capa Bronze.
- Modelos de Industria y la Capa Silver.
- Poblado estructurado y Semi estructurado.
- Soluciones BI y IA sobre la capa Gold.

## 4. PROCESOS EN REAL TIME

- Stomr data y casos de uso.
- Arquitectura LAMBDA.
- Implementando una infraestructura Kafka.
- Activando el Servicio Zookeeper.
- Implementando una ingesta En Real time.
- Implementación de Real Time en modo PaaS.