CTIC UNI

UNIVERSIDAD NACIONAL DE INGENIERÍA

OFICINA DE TECNOLOGÍAS DE LA INFORMACIÓN

Programa de Iniciación Tecnológica PIT - 2025

CERTIFICADO

Otorgado a

ARLES MELVIN CARRASCO CHOQUE

Por haber aprobado el curso de

CIENCIA DE DATOS 2: BIG DATA

Realizado del 14 de febrero al 10 de marzo del presente, con una duración de 16 horas.

Lima, marzo de 2025

Mag. Ing. Rubén Arturo Borja Rosales
JEFE OTI - UNI

N° Certificado: 017 - 0027546

FIRMA DIGITAL

Firmado digitalmente por: BORJA ROSALES Ruben Arturo FAU 20169004359 hard
Motivo: En señal de conformidad
Fecha: 01/04/2025 23:20:53-0500

UNIVERSIDAD NACIONAL DE INGENIERÍA

OFICINA DE TECNOLOGÍAS DE LA INFORMACIÓN

CTIC UNI

Temario del Curso:

CIENCIA DE DATOS 2: BIG DATA

- Etapa definir
- Introducción al big data y contexto actual
- Definición y características de big data
- Importancia y aplicaciones en la actualidad
- Introducción a las tecnologías de hadoop y spark
- Fundamentos de hadoop y su arquitectura
- Estructura y componentes esenciales de hadoop
- Funciones de job tracker y name node
- Implementación de balance de carga
- Transición a yarn (yet another resource negotiator)
- Mapreduce en profundidad
- Detalles del proceso de mapreduce
- Roles y funciones de mapper y reducer
- Estrategias de optimización
- Spark: potenciando el procesamiento de datos
- Ventajas y características clave de apache spark
- Componentes esenciales y su función
- Arquitectura de procesamiento distribuido
- Programación efectiva en spark con pyspark
- Trabajo con rdd y agregaciones en spark
- Uso de merging y cross joints en pyspark
- Introducción a pyspark
- Comparación con scala y beneficios
- Configuración y entorno pyspark

Nota final: 19 (Diecinueve)

N° Certificado: 017 - 0027546