

# Fundamentos de Ingeniería de Big Data

Código: ARC-107

**Propuesta de Valor:** ARCITURA

**Duración:** 24 Horas



Este curso cubre conceptos, técnicas y tecnologías relacionados con la ingeniería para el procesamiento y almacenamiento de conjuntos de datos de Big Data.

Destaca los desafíos únicos que se enfrentan cuando procesar y almacenar conjuntos de datos grandes, volátiles y dispares.

NoSQL está cubierto y el motor de procesamiento de datos MapReduce se explica en detalle como un marco base para procesamiento de datos por lotes de gran volumen.



## AUDIENCIA

- Cualquier persona que esté interesada en ampliar sus conocimientos en Big Data y deseen cubrir unas necesidades no satisfechas por las tecnologías existentes en el almacenamiento y tratamiento de grandes volúmenes de datos.



## PRE REQUISITOS

- No hay requisitos previos.



## OBJETIVOS

- Técnicas y desafíos de ingeniería de Big Data
- Big Data Storage, que incluye fragmentación, replicación, teorema CAP, ACID y BASE
- Replicación maestro-esclavo, de igual a igual, que combina la replicación con fragmentación
- Requisitos, escalabilidad, redundancia y disponibilidad de almacenamiento de Big Data
- Acceso rápido, almacenamiento a largo plazo, almacenamiento sin esquema y almacenamiento económico
- Almacenamiento en disco, incluido el sistema de archivos distribuido y las bases de datos
- Introducción a NoSQL y NewSQL
- Razones y características de NoSQL
- Tipos de bases de datos NoSQL, incluidas las bases de datos de valores clave, documentos, familias de columnas y gráficos
- Motores de procesamiento de Big Data

- Procesamiento de datos distribuido / paralelo, procesamiento de datos sin esquema
- Soporte para múltiples cargas de trabajo, escalabilidad lineal y tolerancia a fallas
- Requisitos de procesamiento de macrodatos, incluidos los modos por lotes, clúster y en tiempo real
- MapReduce para el procesamiento de macrodatos, incluido el mapa, la combinación, la partición, la reproducción aleatoria y la clasificación y reducción
- Diseño del algoritmo MapReduce
- Paralelismo de tareas, paralelismo de datos



## CERTIFICACIÓN DISPONIBLE

- Certificado oficial **ARCITURA**.
- Este curso es parte de la ruta de certificaciones de:
  - Científico Certificado en Big Data
  - Consultor de Big Data Certificado



## CONTENIDO

1. TÉCNICAS Y RETOS DE INGENIERÍA DE BIG DATA
2. ALMACENAMIENTO DE BIG DATA, INCLUYENDO FRAGMENTACIÓN, REPLICACIÓN, TEOREMA DE CAP, ACID Y BASE
3. MAESTRO-ESCLAVO, REPLICACIÓN ENTRE PARES, COMBINACIÓN ENTRE REPLICACIÓN Y FRAGMENTACIÓN
4. REQUERIMIENTOS DE ALMACENAMIENTO DE BIG DATA, ESCALABILIDAD, REDUNDANCIA Y DISPONIBILIDAD
5. ACCESO RÁPIDO, ALMACENAMIENTO A LARGO PLAZO, ALMACENAMIENTO SIN ESQUEMA Y ALMACENAMIENTO ECONÓMICO
6. ALMACENAMIENTO EN DISCO, INCLUYENDO SISTEMA DE ARCHIVOS DISTRIBUIDOS Y BASES DE DATOS
7. INTRODUCCIÓN A NOSQL Y NEWSQL
8. RACIONAL NOSQL Y CARACTERÍSTICAS
9. TIPOS DE BASES DE DATOS NOSQL, INCLUYENDO CLAVE-VALOR, DOCUMENTO, COLUMNA-FAMILIA Y BASES DE DATOS GRÁFICAS
10. MOTORES DE PROCESAMIENTO DE BIG DATA
11. PROCESAMIENTO DE DATOS DISTRIBUIDO/PARALELO, PROCESAMIENTO DE DATOS SIN ESQUEMA
12. SOPORTE MULTI CARGA DE TRABAJO, ESCALABILIDAD LINEAL Y TOLERANCIA A FALLAS
13. REQUERIMIENTOS DE PROCESAMIENTO DE BIG DATA, INCLUYENDO MODOS POR LOTES, CLUSTER Y EN TIEMPO REAL
14. MAPREDUCE PARA PROCESAMIENTO DE BIG DATA, INCLUYENDO MAPA, COMBINAR, PARTICIÓN, BARAJAR Y ORDENAR Y REDUCIR

15. DISEÑO DE ALGORITMOS MAPREDUCE

16. PARALIZACIÓN DE TAREAS, PARALIZACIÓN DE DATOS

---

## ★ BENEFICIOS

- Al finalizar el curso tendrás los conocimientos con la ingeniería para el procesamiento y almacenamiento de conjuntos de datos de Big Data.