



# Posgrados

## CUCEA

*El mejor lugar para el talento*

## Procesamiento de Grandes Bases de Datos

### Actividad 3

Profesor: Dr. Osvaldo Guardado

Alumno: Oscar Fernando Flores  
Garcia

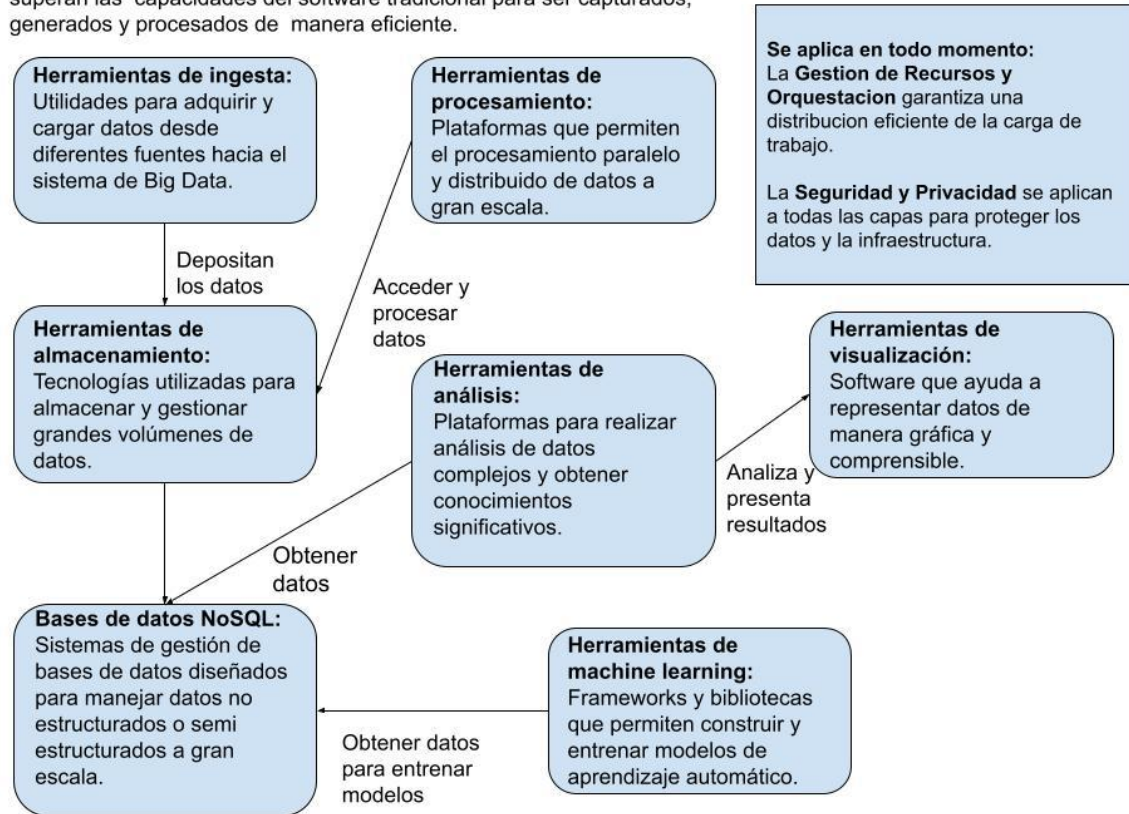
Código: 323019043

Zapopan, Jalisco a 5 de Septiembre  
del 2023

## Tema: Herramientas de Big Data

### Mapa Conceptual de Herramientas de Big Data

**Big Data:** Conjunto de datos extremadamente grandes y complejos que superan las capacidades del software tradicional para ser capturados, generados y procesados de manera eficiente.



### Conceptos Principales:

#### 1. Big Data:

- Definición: Conjunto de datos extremadamente grandes y complejos que superan las capacidades del software tradicional para ser capturados, generados y procesados de manera eficiente.
- Características: Volumen, Velocidad, Variedad, Veracidad.

#### 2. Herramientas de Almacenamiento:

- Definición: Tecnologías utilizadas para almacenar y gestionar grandes volúmenes de datos.
- Ejemplos: HDFS, Amazon S3, Google Cloud Storage.

#### 3. Herramientas de Procesamiento:

- Definición: Plataformas que permiten el procesamiento paralelo y distribuido de datos a gran escala.

- Ejemplos: Apache Hadoop, Apache Spark.

#### **4. Bases de Datos NoSQL:**

- Definición: Sistemas de gestión de bases de datos diseñados para manejar datos no estructurados o semi estructurados a gran escala.
- Tipos: Documentales, Columnares, Clave-Valor, Grafos.
- Ejemplos: MongoDB, Cassandra, Redis, Neo4j.

#### **5. Herramientas de Ingesta:**

- Definición: Utilidades para adquirir y cargar datos desde diferentes fuentes hacia el sistema de Big Data.
- Ejemplos: Apache Kafka, Apache Nifi.

#### **6. Herramientas de Análisis:**

- Definición: Plataformas para realizar análisis de datos complejos y obtener conocimientos significativos.
- Ejemplos: Apache Hive, Impala, Presto.

#### **7. Herramientas de Visualización:**

- Definición: Software que ayuda a representar datos de manera gráfica y comprensible.
- Ejemplos: Tableau, Power BI, D3.js.

#### **8. Herramientas de Machine Learning:**

- Definición: Frameworks y bibliotecas que permiten construir y entrenar modelos de aprendizaje automático.
- Ejemplos: Tensor Flow, PyTorch, scikit-learn.

#### **9. Gestión de Recursos y Orquestación:**

- Definición: Plataformas para administrar y orquestar recursos de Big Data de manera eficiente.
- Ejemplos: Apache Mesos, Kubernetes.

#### **10. Seguridad y Privacidad:**

- Definición: Soluciones para proteger los datos y garantizar el cumplimiento de regulaciones.

- Ejemplos: Ranger, Knox.

#### **Conexiones Importantes:**

- Las **Herramientas de Procesamiento** se integran con las **Herramientas de Almacenamiento** para acceder y procesar datos.
- Las **Bases de Datos NoSQL** son utilizadas tanto en **Herramientas de Almacenamiento** como en **Herramientas de Análisis**.
- Las **Herramientas de Análisis** y **Herramientas de Visualización** trabajan en conjunto para analizar y presentar los resultados.
- Las **Herramientas de Machine Learning** pueden utilizar datos procesados por las otras herramientas para entrenar modelos.
- La **Gestión de Recursos y Orquestación** garantiza una distribución eficiente de la carga de trabajo.
- La **Seguridad y Privacidad** se aplican a todas las capas para proteger los datos y la infraestructura.