

Informe Técnico: Clase 09 de septiembre 2023

Cristian Rivera

Faculta de Ingeniería y Ciencias Básicas,
Universidad Central
Maestría Analítica de Datos
Automatizacion e Integración de Datos para la IA
Bogotá, Colombia
{criveraa}@ucentral.edu.co

September 23, 2023

1 Resumen ejecutivo

El 09 de septiembre de 2023, se llevó a cabo una clase sobre "Automatización e Integración de Datos para la IA". Durante esta sesión, se abordaron temas relacionados con tecnologías emergentes, se discutieron herramientas específicas como Oracle Database Express Server y Apache Hadoop, y se llevó a cabo un ejercicio práctico de manejo de datos con Python. También se proporcionó orientación sobre la creación y gestión de cuentas en Oracle Cloud y la integración con MongoDB.

2 Introducción

El propósito de esta clase fue proporcionar una visión integral de la gestión de datos en el contexto de la Inteligencia Artificial. Se abordaron temas desde la selección de herramientas hasta el manejo práctico de datos con códigos de Python.

3 Contexto y antecedentes

Con el aumento de la Inteligencia Artificial (IA), la capacidad para gestionar, integrar y automatizar datos se ha vuelto esencial. Las organizaciones buscan constantemente herramientas y metodologías para hacerlo de manera eficiente.

4 Metodología

La clase se dividió en segmentos teóricos y prácticos. Los aspectos teóricos abordan la selección y uso de herramientas, mientras que la parte práctica se centra en la ejecución y aplicación de estas herramientas.

5 Desarrollo de la clase

5.1 Repaso de noticias tecnológicas actuales

La clase comenzó con una revisión de las principales noticias en el mundo tecnológico. Esta actividad permitió contextualizar la importancia de la automatización y la integración de datos en el panorama tecnológico actual, abordando tendencias emergentes y sus implicaciones para la industria y la investigación.

5.2 Presentación de Jesús Moreno sobre Oracle Database Express Server

Jesús Moreno llevó a cabo una exposición detallada sobre Oracle Database Express Server, subrayando sus características principales, ventajas y aplicaciones prácticas. Destacó cómo esta herramienta puede ser esencial para organizaciones que buscan una solución de base de datos robusta y confiable.



Figure 1: Oracle Servidor express

5.3 Exposición de Daniel Soto sobre Apache Hadoop

Daniel Soto profundizó en Apache Hadoop, una herramienta ampliamente utilizada para el procesamiento de grandes conjuntos de datos. Abordó la arqui-

itectura fundamental del sistema, sus componentes clave como HDFS y MapReduce, y cómo se integra con otros sistemas y herramientas para proporcionar soluciones de big data eficientes.

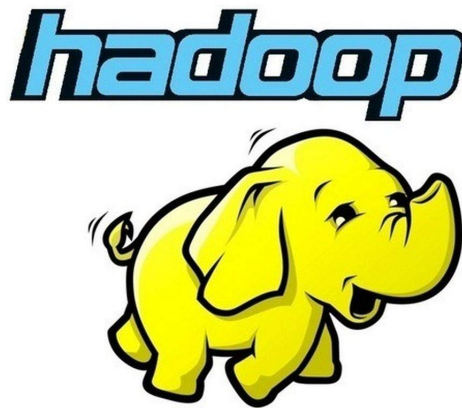


Figure 2: Apache Hadoop

5.4 Discusión sobre fuentes de datos

Se exploraron en profundidad las distintas fuentes de datos disponibles para las organizaciones. Desde archivos convencionales en formatos como xls, pdf, csv y txt hasta bases de datos más avanzadas como SQL, NoSQL, data lakes y data warehouses. Se enfatizó en cómo cada fuente tiene sus propios desafíos y ventajas, y la necesidad de elegir la fuente adecuada según el proyecto o aplicación en cuestión. También se mencionó la creciente relevancia de los servicios web y cómo los formatos como JSON están revolucionando la manera en que se intercambian datos en la web.

5.5 Ejercicio práctico en Python

En esta sección, se realizó un ejercicio interactivo para los estudiantes, donde se les mostró cómo gestionar y validar grandes volúmenes de datos utilizando Python. El ejercicio resaltó la importancia de la validación de datos para asegurar su integridad y calidad. Además, se discutió la importancia de comprimir archivos para optimizar el almacenamiento, especialmente cuando se trabaja con grandes conjuntos de datos.

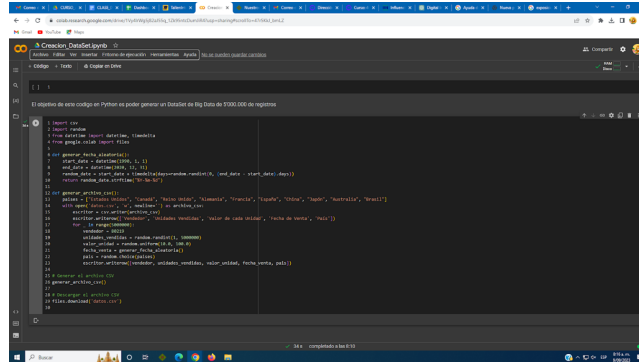


Figure 3: Ejercicio práctico en Python

5.6 Demostración en Oracle Cloud

Se llevó a cabo una demostración paso a paso sobre cómo crear una cuenta en Oracle Cloud, configurar un repositorio de bases de datos y cómo conectarlo a herramientas como Oracle Developer. Esta demostración sirvió para resaltar las capacidades de Oracle Cloud en cuanto a la gestión de datos y cómo los profesionales pueden aprovechar estas herramientas para sus proyectos.

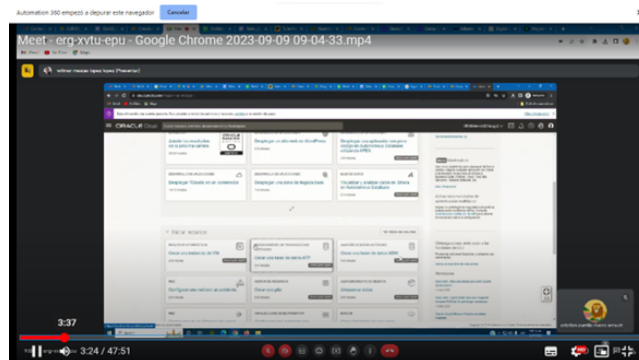


Figure 4: Demostración en Oracle Cloud

5.7 Integración con MongoDB

Finalmente, se mostró cómo cargar datos previamente gestionados en MongoDB, una popular base de datos NoSQL. Además, se explicó cómo conectar MongoDB con Oracle Cloud, permitiendo a los usuarios acceder a datos no relacionales y almacenarlos en la nube, optimizando recursos y aprovechando la capacidad de Oracle Cloud.

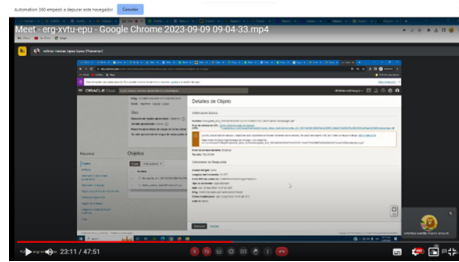


Figure 5: Conexión de Oracle Cloud a MongoDB

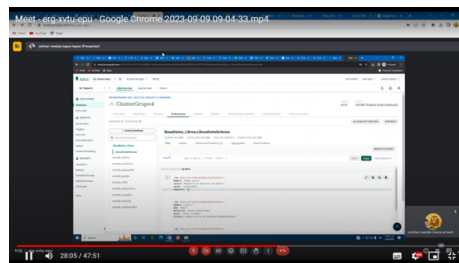


Figure 6: Integración con MongoDB

6 Resultados y análisis

Los estudiantes adquirieron conocimientos tanto teóricos como prácticos. Comprendió la importancia de gestionar adecuadamente los datos, desde su obtención hasta su almacenamiento, especialmente en contextos de IA.

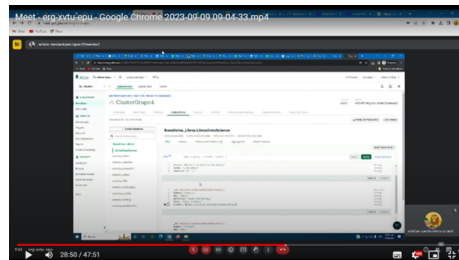


Figure 7: Resultados y análisis

7 Conclusiones

La gestión adecuada de los datos es crucial para las aplicaciones de IA. Las herramientas y técnicas discutidas en la clase son esenciales para cualquier profesional en el campo.