

|  |  |
| --- | --- |
| **Nombre:** Coyago Arce Eunice Pamela | **Carrera:** Tecnologías de la Información |
| **Asignatura:** Modelado Avanzado de datos | **NRC:** 10052 |
| **Tema:** Anotaciones | **Fecha de Entrega:** |

Contenido

[Clase 1: 02/05/2023 2](#_Toc134102698)

[Para la próxima clase consultar 2](#_Toc134102699)

[Looker Studio 2](#_Toc134102700)

[Qué es? 2](#_Toc134102701)

[Visualización de datos 2](#_Toc134102702)

[Tableau 2](#_Toc134102703)

[Replica sheets 3](#_Toc134102704)

[Temas de ETL, ELT, Apache airflow 3](#_Toc134102705)

[Bloques funcionales 3](#_Toc134102706)

[El proyecto del 1er parcial, será un CRUD en mongo 3](#_Toc134102707)

[3 tipos de evaluación. 3](#_Toc134102708)

[Apuntes en github 4](#_Toc134102709)

[CONSULTA 1 4](#_Toc134102710)

[SQL y NoSQL 4](#_Toc134102711)

[Información dispersa 4](#_Toc134102712)

[Bases de datos relacionales y no relacionales 4](#_Toc134102713)

[Que bases de datos encontramos como SQL 4](#_Toc134102714)

[Origen y calidad de los datos 4](#_Toc134102715)

[Funcionalidad de los datos 4](#_Toc134102716)

[Tipos, usos 4](#_Toc134102717)

[Tipos de bases de datos SQL 4](#_Toc134102718)

[Bibliografía 5](#_Toc134102719)

# Clase 1: 02/05/2023

* Realizar cursos y obtener Certificados de mongo university sobre modelado de datos (udemy, platzi)
* Evidencias de hacer ejercicio

8pts. proyecto

8pts. examen

**Modelo ABP** -> Aplicaciones Basado en Proyectos

Mongo DB local y en la nube

## Para la próxima clase consultar

* SQL y noSQL
* Información dispersa
* Bases de datos relacionales y no relacionales
* Que bases de datos encontramos como SQL
* Origen y calidad de los datos
* Funcionalidad de los datos
* Tipos, usos
* Tipos de bases de datos SQL

Looker Studio -> <https://lookerstudio.google.com/navigation/reporting>



Qué es? **->** es una plataforma de inteligencia empresarial que permite a las empresas analizar y visualizar datos. Parte de sus ventajas consisten en su capacidad para unir y transformar datos de múltiples fuentes, lo que significa que permite a los usuarios acceder a información integral en tiempo real.

### Visualización de datos

La visualización de datos es una de las características más destacadas de Looker Studio, ya que permite a los usuarios crear gráficos y tablas interactivas para representar los datos de una manera fácil de entender y de analizar.

Dashboard -> es una herramienta visual que proporciona una vista de la información e indicadores de rendimiento de una empresa o proyecto.

Tableau -> <https://www.tableau.com/> es una plataforma de visualización y análisis de datos, permite a los usuarios conectarse a diversas fuentes de datos, limpiar y transformar los datos, y crear visualizaciones interactivas y tableros de control.

Replica sheets **->** es una función de Tableau que permite crear copias de una hoja de trabajo existente con todos los cálculos y filtros aplicados, ayuda a ahorrar tiempo, reducir errores y personalizar hojas de trabajo sin afectar la original.

## Temas de ETL, ELT, Apache airflow

* **ETL (Extract, Transform, Load):** es un proceso utilizado para extraer datos de múltiples fuentes, transformarlos en un formato común y cargarlos en un destino final. El objetivo principal de ETL es hacer que los datos sean consistentes y fáciles de analizar.
* **ELT (Extract, Load, Transform):** es similar a ETL, pero en lugar de transformar los datos antes de cargarlos en el destino final, los datos se cargan primero y luego se transforman.
* **Apache Airflow:** es una plataforma de flujos de trabajo que permite a los equipos de datos crear, programar y monitorear flujos de trabajo de datos complejos, facilita la automatización y el monitoreo de los flujos de trabajo.

## Bloques funcionales

Son componentes modulares que se utilizan para construir sistemas complejos a partir de unidades más simples. Cada bloque funcional es una entidad independiente que realiza una función específica y se puede conectar a otros bloques para formar un sistema más grande.

* **Revisar catálogos de datos**
* **Que data se va a trabajar? Tamaño, soporte**

## El proyecto del 1er parcial, será un CRUD en mongo

### 3 tipos de evaluación.

1. Carga de subir
2. Carga de interfaz (cargar en bloque)
3. Carga con comandos

* **Curar la data** -> proceso de verificar y corregir la calidad de los datos para asegurarse de que sean precisos, completos y consistentes antes de ser utilizados, elimina datos duplicados, corrige errores y estandariza formatos.
* **Csv a JSON** -> transformación de formatos de los datos.
* Limpiar, transformar y subir.
* Mediante consola.
* Replicación de base de datos.
* Seguridad de base de datos de mongo.

## Apuntes en github

* **Nombre del repositorio:** Modelado avanzado de datos
* Clase 1. MV

**La data no puede ser dummy?** -> Porque es dispersa, no hay una congruencia, es información de prueba, son datos simulados.

* Data de España?
* Data Channel ?
* La data más apropiada, porque con ello se trabajará todo el semestre.
* Entregas visuales.
* La data no puede abrirse en Excel.

## CONSULTA 1

### SQL y NoSQL

* **SQL** -> "Structured Query Language" = (Lenguaje de Consulta Estructurada), es un lenguaje utilizado para acceder y manipular datos almacenados en bases de datos relacionales. Las bases de datos relacionales se basan en tablas con filas y columnas, y utilizan SQL para realizar consultas y manipular los datos almacenados en estas tablas.
* **NoSQL** -> "Not Only SQL" = (No Solo SQL), se refiere a una amplia categoría de sistemas de bases de datos que no utilizan tablas relacionales, almacena datos en formatos flexibles como documentos, gráficos o clave-valor.

|  |  |  |
| --- | --- | --- |
| Característica | SQL | NoSQL |
| Estructura de Datos | Estructura de datos fija | Estructura de datos flexible |
| Escalabilidad | Escalabilidad vertical | Escalabilidad horizontal |
| Flexionar | Menos flexible en cambios | Más flexible en cambios |
| Transacciones | Soporta transacciones ACID | Soporte transaccional no ACID |
| Modelado de Datos | Basado en tablas | Basado en documentos, grafos, clave-valor, etc. |
| Facilidad de Uso | Fácil de aprender y usar | No tan fácil de aprender y usar |
| Consultar | Soporta consultas complejas | No es así |
| Escalabilidad web | No está optimizado para aplicaciones web escalables | Optimizado para aplicaciones web escalables |
| Escritura | Lento en escrituras masivas | Rápido en escrituras masivas |
| Ejemplos de Bases de Datos | MySQL, PostgreSQL, Oracle | MongoDB, Cassandra, Sofá |

### Bases de datos relacionales y no relacionales

* **Bases de datos relacionales ->** (RDBMS, Relational Database Management System) son sistemas de gestión de bases de datos que almacenan los datos en tablas relacionales. Las tablas relacionales consisten en filas y columnas, donde cada columna representa un atributo o característica de los datos, y cada fila representa una instancia o registro de esos datos, se basan en el modelo relacional, establece reglas para garantizar la integridad de los datos, lo que significa que los datos almacenados en las tablas deben ser precisos y consistentes en todo momento.
* **Bases de datos no relacionales ->** (NoSQL, Not Only SQL) son sistemas de gestión de bases de datos que se diferencian de las bases de datos relacionales en la forma en que almacenan y organizan los datos. A diferencia de las bases de datos relacionales, que utilizan tablas relacionales, las bases de datos NoSQL utilizan modelos de datos no estructurados, como documentos, gráficos o pares clave-valor, surgieron como una respuesta a la necesidad de manejar grandes volúmenes de datos y procesarlos de manera más eficiente que las bases de datos relacionales.

## Transacciones ACID

Son un conjunto de propiedades que garantizan la integridad de los datos en las operaciones de una base de datos.

* **Atomicidad (Atomicity):** Si se produce un error durante la transacción, se anula toda la transacción y se devuelven los datos al estado anterior a la transacción.
* **Consistencia (Consistency):** Las reglas y restricciones de la base de datos deben aplicarse antes y después de la transacción.
* **Aislamiento (Isolation):** Cada transacción debe ser aislada de otras transacciones, no pueden interferir entre sí.
* **Durabilidad (Durability):** Una vez que se completa una transacción, los cambios realizados en la base de datos deben ser permanentes y persistir a pesar de fallas del sistema, energía o errores de hardware.

### Información dispersa

Es información que se encuentra distribuida en diferentes lugares, formatos o sistemas, lo que puede generar incompletitud y desactualización. Esto se debe a la falta de una gestión centralizada de los datos y la ausencia de estándares de formato y herramientas de integración de datos adecuadas.

### Que bases de datos encontramos como SQL y NoSQL

* **SQL**
  + MySQL
  + Oracle
  + Microsoft SQL Server
  + PostgreSQL
  + SQLite
* **NoSQL**
  + MongoDB
  + Cassandra
  + Couchbase
  + Amazon DynamoDB
  + Apache HBase

### 

### Origen y calidad de los datos

El origen de los datos se refiere a su procedencia, mientras que la calidad de los datos se refiere a su precisión, integridad, consistencia, actualidad y relevancia. Es importante conocer el origen de los datos para entender su contexto y relevancia, y asegurarse de que los datos sean precisos, completos y relevantes para garantizar su calidad.

### Funcionalidad de los datos

Se refiere a su capacidad para ser utilizados en la realización de tareas específicas, como la toma de decisiones y la automatización de procesos. Para que los datos sean funcionales, deben ser relevantes, precisos, completos, actualizados y consistentes, ser procesados y analizados con herramientas y técnicas adecuadas. La funcionalidad de los datos es esencial para la toma de decisiones informadas y la mejora de la eficiencia operativa en una organización.

### Tipos, usos de los datos

* **Tipos**
  + **Datos numéricos:** valores numéricos.
  + **Datos alfanuméricos:** combinan letras y números para representar información.
  + **Datos categóricos**: representan categorías o clasificaciones de información.
  + **Datos espaciales**: representan información en un contexto geográfico o espacial.
  + **Datos temporales**: representan información en un contexto temporal, como fechas y horarios.
* **Uso**
  + **Análisis de datos**: para extraer información y conocimiento útil de los datos.
  + **Toma de decisiones**: para respaldar la toma de decisiones informadas en una organización.
  + **Automatización de procesos**: para optimizar y automatizar procesos de negocio.
  + **Mejora de la experiencia del usuario**: para personalizar y mejorar la experiencia del usuario en un sitio web o aplicación.
  + **Investigación y desarrollo**: para realizar investigaciones y desarrollar nuevos productos o servicios.

### Tipos de bases de datos SQL

* **Bases de datos de servidor:** son bases de datos que se ejecutan en un servidor y se accede a ellas a través de una red.
* **Bases de datos de escritorio:** son bases de datos que se ejecutan en un único equipo y se accede a ellas localmente.
* **Bases de datos de cliente-servidor:** son bases de datos que se ejecutan en un servidor y se accede a ellas a través de una aplicación cliente.
* **Bases de datos en la nube:** son bases de datos que se ejecutan en la nube y se accede a ellas a través de una conexión a Internet.
* **Bases de datos embebidas**: son bases de datos que se integran en una aplicación y se accede a ellas a través de la aplicación.

### Tipos de bases de datos NoSQL

* **Bases de datos de documentos:** almacenan datos en documentos estructurados, como JSON o XML, y se pueden acceder a través de claves únicas.
* **Bases de datos de grafos:** almacenan datos como nodos y bordes para representar las relaciones entre ellos, y se utilizan comúnmente para aplicaciones que involucran redes sociales y recomendaciones.
* **Bases de datos de llave-valor:** almacenan datos en una estructura simple de pares de llave-valor y se utilizan comúnmente para aplicaciones que requieren alta escalabilidad.
* **Bases de datos de columnas:** almacenan datos en columnas en lugar de filas, lo que permite una recuperación de datos más rápida y eficiente, y se utilizan comúnmente en aplicaciones analíticas.

# Clase 2: 04/05/2023

# Bibliografía

Bases de datos: qué tipos existen y cómo funcionan. (2017, October 20). Platzi. https://platzi.com/blog/bases-de-datos-que-son-que-tipos-existen/

Bases de datos SQL - Definición y Ejemplos. (n.d.). Tecnologias-informacion.com. Retrieved May 4, 2023, from https://www.tecnologias-informacion.com/sql.html

Help and training community. (n.d.). Salesforce. Retrieved May 4, 2023, from https://help.salesforce.com/s/articleView?id=sf.siteforce\_data\_function.htm&type=5

La importancia de la Calidad del Dato Origen. (n.d.). Powerdata.es. Retrieved May 4, 2023, from https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/353201/La-importancia-de-la-Calidad-del-Dato-Origen

La información dispersa que guardan las empresas aportaría más valor si se integrara. (n.d.). Dataprix.com. Retrieved May 4, 2023, from https://www.dataprix.com/es/blog-it/sage/informacion-dispersa-guardan-empresas-aportaria-mas-valor-si-se-integrara

Looker studio overview. (n.d.). Google.com. Retrieved May 4, 2023, from https://lookerstudio.google.com/navigation/reporting

Monster Digital Agency. (2021, August 2). Base de datos: Usos, aplicaciones y tipos de bases de datos. Epitech Spain. https://www.epitech-it.es/base-de-datos/

Orlando. (2019, June 4). Bases de Datos NoSQL. GraphEverywhere; Graph Everywhere SL. https://www.grapheverywhere.com/bases-de-datos-nosql-marcas-tipos-ventajas/

¿Qué es el modelo ACID en bases de datos? (2022, April 19). KeepCoding Bootcamps. https://keepcoding.io/blog/que-es-acid-bases-datos/

Rendón, Y. A. (n.d.). Bases de datos relacionales vs. no relacionales. Com.co. Retrieved May 4, 2023, from https://www.pragma.com.co/academia/lecciones/bases-de-datos-relacionales-vs.-no-relacionales

Tablado, F. (2020, September 7). Conoce las bases de datos SQL a fondo. Ayuda Ley Protección Datos; AyudaLeyProteccionDatos. https://ayudaleyprotecciondatos.es/bases-de-datos/sql/

Tableau: Business intelligence and analytics software. (n.d.). Tableau. Retrieved May 4, 2023, from https://www.tableau.com/

Torres, A. (2021, November 22). ¿Qué es SQL? ¿Qué es una base de datos? Sistemas de bases de datos relacionales explicados de forma simple. freecodecamp.org. https://www.freecodecamp.org/espanol/news/que-es-sql-que-es-una-base-de-datos-sistemas-de-datos-relacionales-explicados-de-forma-simple/

Usos y aplicaciones de una base de datos. (2022, May 5). Blog de Ayuware; Ayuware. https://www.ayuware.es/blog/usos-de-una-base-de-datos/

Vergara, A. (2015, October 26). SQL vs NoSQL ¿Cuál debo usar? - Tech blog for developers. Tech blog for developers | Facilcloud. https://www.facilcloud.com/noticias/sql-vs-nosql-which-one-should-i-use/

(N.d.). Amazon.com. Retrieved May 4, 2023, from https://aws.amazon.com/es/nosql/