



**INVESTIGACIÓN  
CIENCIA DE DATOS**

**Realizado por:**

**Nicky Alexander Flórez Bustamante**

**Juan Andrés Menéndez Villarraga**

**David Fernando Gómez Aristizábal**

**Ficha: 2828523**

**Ciencia de Datos**

## 1) Ambiente personal de Aprendizaje

Un ambiente de aprendizaje es un entorno organizado y estructurado que facilita la adquisición de conocimientos

**Físico:** Clases en aulas

- Clases practicas
- Horas de estudio en la casa

**Social:** Trabajos en equipo para crear el proyecto

- Asistir a conferencias de computación y conocer otras personas con profesiones afines

**Emocional:** Uso de tutorías para manejar la carga estudiantil

- Grupos de apoyo

**Pedagógico:** Practicas profesionales en empresas tecnológicas

**Tecnológico:** Uso de software como Visual Studio Code o Android Studio

- Utilización de repositorios y foros con artículos sobre el tema

## 2) Métodos de aprendizaje

Los métodos de aprendizaje son estrategias, técnicas o enfoques estructurados que facilitan la adquisición, retención y aplicación de conocimientos.

Existen varios métodos de aprendizaje tales como: Aprendizaje activo, Aprendizaje basado en Proyectos, Aprendizaje cooperativo, Aprendizaje basado en problemas, aprendizaje por enseñanza, Por simulación, Autoaprendizaje o por método de casos.

En nuestro caso, podríamos combinar métodos como el aprendizaje activo, el aprendizaje basado en pruebas, el autoaprendizaje y el aprendizaje basado en problemas serían los más adecuados.

## 3) Tipos de aprendizaje

Son las diferentes formas en las que las personas adquieren, procesan y aplican conocimientos, según teorías, los principales son:

- Aprendizaje memorístico o repetitivo: Basado en la repetición para retener información sin necesariamente comprenderla en profundidad.
- Aprendizaje Cognitivo: Enfatiza procesos mentales como atención, memoria y pensamiento crítico. Relacionado con Piaget y la metacognición.
- Aprendizaje Explícito: se caracteriza porque el aprendiz tiene intención de aprender y es consciente de qué aprende.
- Aprendizaje asociativo: es un proceso por el cual un individuo aprende la asociación entre dos estímulos o un estímulo y un comportamiento.
- Aprendizaje Significativo: se caracteriza porque el individuo recoge la información, la selecciona, organiza y establece relaciones con el conocimiento que ya tenía previamente.

## 4) ¿Qué es una base de datos?

Bases de datos son sistemas estructurados que permiten almacenar, organizar y gestionar información de manera eficiente, facilitando su consulta, actualización o análisis. Funcionan mediante motores o gestores (DBMS), software especializado que administra los datos según reglas específicas.

- Tipos y motores principales:

### 1. **Relacionales (SQL):**

Estructura: Datos en tablas relacionadas mediante claves (filas y columnas).

Motores: MySQL, PostgreSQL, Oracle Database, Microsoft SQL Server.

Usos: Sistemas transaccionales (bancos, ERP), donde la integridad (ACID) es crítica.

### 2. **NoSQL:**

- Documentales (almacenan datos en formatos como JSON): motores como MongoDB, Couchbase son los ideales para apps con esquemas flexibles (redes sociales).
- Clave-valor (optimizados para velocidad): Redis (en memoria), Amazon DynamoDB. Usados en cachés o carritos de compra.
- Columnares (organizan datos por columnas): Apache Cassandra, HBase. Perfectos para análisis masivos (Big Data).
- Grafos (enfocados en relaciones entre datos): Neo4j, Amazon Neptune. Útiles en sistemas de recomendación o fraud detection.
- En memoria: Redis, Memcached. Almacenan datos en RAM para acceso ultrarrápido (sesiones de usuarios).
- Series temporales: InfluxDB, Prometheus. Capturan datos con marca de tiempo (IoT, monitoreo de servidores).
- NewSQL: CockroachDB, Google Spanner. Combina escalabilidad de NoSQL con ACID de SQL.
-

Los usos más comunes de las bases de datos son:

Web y apps: MySQL (backend de sitios), MongoDB (perfiles de usuarios).

Analítica: PostgreSQL (consultas complejas), Cassandra (procesamiento distribuido).

Empresas: Oracle (gestión corporativa), Microsoft SQL Server (reportes integrados).

Tiempo real: Redis (notificaciones instantáneas), InfluxDB (telemetría).

## 5) ¿Cuáles son los tipos de bases de datos?

Existen diversos tipos de bases de datos, cada uno de ellos enfocados a funciones específicas, las cuales son:

1. **Base de datos relacional:** Las bases de datos relacionales son las más comunes en la actualidad. *Consisten en una serie de contenedores de información denominados tablas, en los que se almacenan registros.* Los registros serían las filas de las tablas mientras que sus campos de éstos serían las columnas.

La característica más importante de las bases de datos relacionales es que los datos de las tablas pueden estar relacionados entre sí por medio de índices que nos permiten pivotar de una tabla a otra o otras.

2. **Base de datos distribuidas:** Las bases de datos distribuidas son aquellas en las que *la información no reside en un único servidor de bases de datos*, sino que existen diversos servidores, a menudo en redes distintas, que mantienen una porción de los datos.

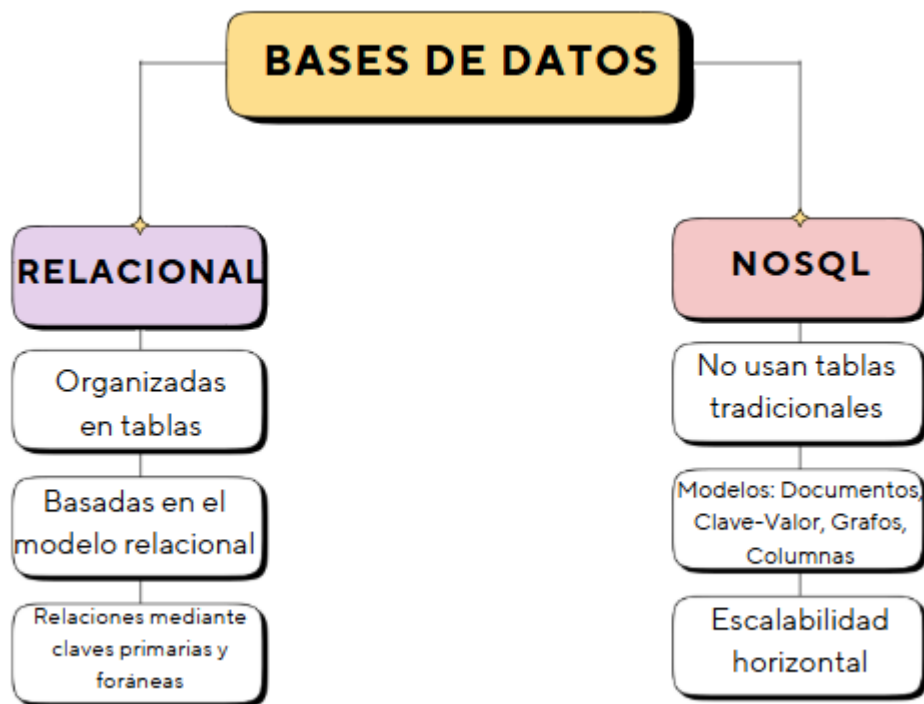
Debido a su característica distribuida, las bases de datos están alojadas en diversos nodos y se produce un trasiego de la información continuo entre ellos. Sin embargo, los clientes pueden acceder a los datos y recuperar la información sin necesidad de preocuparse de que ésta se encuentre distribuida entre varias máquinas.

3. **Base de datos orientada a objetos:** Son un modelo de bases de datos en las que la información se almacena mediante objetos. *Los objetos son conjuntos heterogéneos de datos*, tan complejos como sea necesario para modelizar aquella información que se necesite.
4. **Base de datos gráfica:** Son un tipo de bases de datos donde *se guardan datos y sus relaciones*. Sin embargo, las relaciones podemos considerarlas en este modelo de base de datos como ciudadanos de

primer nivel, ya que se almacenan tal cual en la base de datos y no mediante índices como en las relacionales.

5. **Base de datos NoSQL:** Las bases de datos NoSQL (Not only SQL) son aquellas en las que *no se usan tablas sino colecciones de elementos*. Los elementos almacenados en las colecciones pueden ser heterogéneos, de modo que en una colección podemos almacenar registros con juegos de datos distintos entre sí. Además, este tipo de base de datos no suele usar SQL para realizar consultas, sino programación funcional para hacer filtrados y otros tipos de operaciones.

## 6. Descripción BD relacionales y NoSql



## 7). Marcas y gestores de bases de datos.

### 1. Gestores de Bases de Datos Relacionales (SQL):

- **Oracle Database:** Alta disponibilidad y soporte para transacciones críticas.
- **Microsoft SQL Server:** Integración con herramientas de Microsoft y análisis de datos.
- **MySQL:** Código abierto, eficiente y popular en aplicaciones web.
- **PostgreSQL:** Soporte avanzado para JSON y aplicaciones geoespaciales.

### 2. Gestores de Bases de Datos No Relacionales (NoSQL)

- **MongoDB:** Esquema flexible, ideal para big data y aplicaciones web.
- **Apache Cassandra:** Escalabilidad lineal y tolerante a fallos.
- **Redis:** Almacenamiento en memoria para operaciones rápidas.
- **Neo4j:** Optimizado para análisis de relaciones y redes sociales.
- **Amazon DynamoDB:** Escalabilidad automática para aplicaciones serverless.

### 3. Gestores Multimodelo

- **Couchbase:** Documentos y Clave-Valor con sincronización móvil.
- **ArangoDB:** Documentos, Clave-Valor y Grafos para análisis complejos.
- **OrientDB:** Soporta múltiples modelos en una sola base.

**8) ¿Que es Python? Historia y documentación.**

Python es un lenguaje de programación el cual es multiparadigma, es decir, multipropósito ya que con este lenguaje se logra crear software, aplicaciones, páginas web, el análisis de datos, automatización de tareas y más.

Python se ha vuelto muy popular y no solo por ser multipropósito sino también porque es fácil de aprender ya que su sintaxis es basada en la indentación lo cual permite ver de forma visual y de forma clara la estructura del código. Se considera fácil ya que está orientado al lenguaje humano y no tanto al lenguaje de máquina lo que lo hace de alto nivel.

Python fue creado por **Guido Van Rossum** a finales de los 80, esto fue principalmente un proyecto personal de Guido, esto contaba con bases inspiradas de otros lenguajes como ABC y otros populares de la época. Python fue llamado así gracias al grupo de comedia Monty Python.

La primera versión oficial de Python fue lanzada en el 1991 en la versión (0.9.0) la cual contaba con manejo de excepciones y tipos de datos avanzados (Listas, Tuplas, Diccionarios, Conjuntos). Aunque en 2000 Python fue muy popular gracias a su versión 2 e en el año 2008 fue en donde se dio más a conocer, aunque tuvo una controversia ya que Python 3 no era compatible con la versión anterior. Python 3 se impuso gracias a mejoras como un manejo de texto más eficiente y una sintaxis más clara. Destacado por su simplicidad, legibilidad y versatilidad, Python se convirtió en un estándar para desarrollo web (con Django y Flask), ciencia de datos, aprendizaje automático y más.

Guido van Rossum fue el líder del proyecto hasta 2018, pero hoy Python sigue creciendo gracias a una comunidad activa.

La documentación de Python es como un manual oficial que explica cómo funciona el lenguaje y todas sus herramientas. Está disponible en línea de forma gratuita y es muy completa, incluyendo tutoriales para principiantes, guías para temas específicos (como desarrollo web o análisis de datos) y referencias detalladas sobre las funciones y bibliotecas de Python. Es un recurso invaluable tanto para quienes están empezando a programar como para desarrolladores experimentados, ya que ofrece ejemplos claros y explicaciones paso a paso. Puedes acceder a ella en el sitio oficial de Python: <https://docs.python.org>.

## 9) ¿Que son Replit y Github? Y ¿para qué son?



Replit es una plataforma web la cual permite escribir, ejecutar y compartir código desde el navegador. Es una plataforma en la cual se puede trabajar con múltiples lenguajes de programación, esto ayuda a no tener la obligación de instalar los lenguajes y/o herramientas necesarias para empezar a programar con el lenguaje deseado.

Github es una plataforma web que permite gestionar, mantener, colaborar y controlar proyectos públicos o privados principalmente de software. Github está basado en **Git** (Creado por el creador de Linux, es decir, Linus Torvalds) el cual es un control de versiones que permite controlar y rastrear cambios en un código fuente.

Replit es muy útil para aprender a programar ya que cuenta con herramientas como autocompletado o resalte de sintaxis lo cual ayuda al momento de escribir código. Además de esto permite la colaboración en tiempo real lo cual es demasiado útil para pequeños equipos o para recibir ayuda. Replit también es muy útil para probar y ejecutar alguna pequeña funcionalidad rápidamente ya que los resultados se ven en tiempo real. Al estar basado en la nube Replit permite acceder a los proyectos desde cualquier dispositivo con conexión a internet.

GitHub aparte de poder almacenar, gestionar y colaborar en proyectos de software permite rastrear cambios en el código, trabajar en ramas separadas y fusionar cambios mediante peticiones. Github es ideal para colaboración en equipo, alojar proyectos de código abierto y automatizar tareas como pruebas y despliegues. Además, sirve como portafolio digital para desarrolladores.

## 10) ¿Que es la ciencia de datos?

La ciencia de datos es ser un explorador de información. Consiste en **recopilar, organizar y analizar** datos (números, textos, base de datos, archivos.) para descubrir patrones, responder preguntas o resolver problemas. Por ejemplo, puede ayudar a predecir tendencias, entender el comportamiento de los datos o mejorar decisiones en áreas como negocios, salud o tecnología. Para lograrlo, se usan herramientas como estadísticas, programación y gráficos que transforman datos complejos en información clara y útil.

## BIBLIOGRAFÍA

<https://es.wikipedia.org/wiki/Python>

<https://replit.com>

<https://docs.replit.com>

<https://www.dongee.com/tutoriales/que-es-replit-una-guia-completa-para-principiantes/>

<https://docs.github.com/es/get-started/start-your-journey/about-github-and-git>

<https://aws.amazon.com/es/what-is/data-science/>

<https://www.arsys.es/blog/tipos-de-bases-de-datos-que-existen>

[www.mongodb.com](http://www.mongodb.com)

[Cassandra.apache.org](http://Cassandra.apache.org)

[www.oracle.com](http://www.oracle.com)