

# Plan de Trabajo del Proyecto Final

Bases de datos no relacionales - 2025

## 1. Identificación del grupo

- Número del grupo: 23
- Integrantes:
  - Guido Pereyra (65072660-4)
  - Antonio López (1885972-4)

## 2. Familia de proyecto elegida

- Familia 5/6: Relevamiento y comparación de bases analíticas / Modelado cruzado

## 3. Objetivos del proyecto

Investigar y comparar las bases teóricas, arquitecturas y principios del funcionamiento de dos diferentes soluciones de bases de datos para el manejo de series temporales :

- ClickHouse : real-time analytics database (base analítica)
- MongoDB : time series collections (base no analítica)

Ambos motores proponen estrategias diferentes para manejar el mismo tipo de información:

- ClickHouse es un motor columnar
- MongoDB es un motor documental

Se identificará un dataset que se pueda cargar en ambos motores y que tenga suficiente volumen para que se puedan realizar comparaciones de performance implementando las mismas consultas en ambas bases.

## 4. Descripción general del proyecto

El proyecto consiste en un estudio comparativo de diferentes arquitecturas de bases de datos para un caso de uso muy específico y usado actualmente como son las series temporales. Se realizará un análisis de sus arquitecturas, modelos de datos y mecanismos de optimización, finalizando con una implementación práctica básica para verificar conceptos claves aprovechando la naturaleza open source de ambas soluciones. La motivación del proyecto es entender cómo estas

soluciones tan diferentes abordan los desafíos específicos del análisis de series temporales y que ventajas/desventajas presenta cada enfoque.

## 5. Actividades previstas y cronograma tentativo

| Actividad                         | Breve descripción                                                       | Semana(s) prevista(s) |
|-----------------------------------|-------------------------------------------------------------------------|-----------------------|
| Investigación Inicial             | Recopilar documentación, artículos y recursos de las tres tecnologías.  | Semana 1 (10 horas)   |
| Análisis de arquitecturas         | Estudiar la arquitectura, modelo de datos y optimizaciones de cada base | Semana 2 (10 horas)   |
| Preparación de entorno de trabajo | Configurar entorno y bases de datos                                     | Semana 3 (8 horas)    |
| Preparación de los datos          | Identificar y preparar dataset de series temporales para las pruebas    | Semana 3 (4 horas)    |
| Implementación ClickHouse         | Crear modelo, cargar datos y diseñar consultas                          | Semana 4 (8 horas)    |
| Implementación MongoDB            | Crear modelo, cargar datos y diseñar consultas                          | Semana 5 (4 horas)    |
| Pruebas y comparación             | Ejecutar consultas y comparar resultados                                | Semana 5 (8 horas)    |
| Documentación y análisis          | Documentar hallazgos, ventajas y desventajas de cada base               | Semana 6 (8 horas)    |

## 6. Base teórica o artículos relacionados

- Documentación
  - [The Best Time-Series Databases Compared | Timescale](#)
  - [Working with Time Series Data in ClickHouse](#)
  - [Time Series - Database Manual v8.0 - MongoDB Docs](#)
- Papers
  - [PERFORMANCE STUDY OF TIME SERIES DATABASES](#)
  - [Comparative Analysis of Time Series Databases in the Context of Edge Computing for Low Power Sensor Networks - PMC](#)

## 7. Datos y entornos

- Juegos de datos a utilizar:
  - [Tutorials and Example Datasets | ClickHouse Docs](#)
- Entornos de prueba y herramientas:
  - Lenguaje de programación
    - Python
  - Bases de datos
    - ClickHouse
    - MongoDB
  - Documentación y visualización de resultados
    - Jupyter Notebook
  - Control de versiones
    - [gitlab.fing.edu.uy](https://gitlab.fing.edu.uy)

## 8. Otras observaciones

- Las limitaciones de recursos del entorno de desarrollo local podrían no reflejar el rendimiento real en entornos de producción a gran escala.
- Sería interesante explorar la integración con herramientas de visualización como Grafana si el tiempo lo permite, ya que es un caso de uso común para datos de series temporales.