



Proyecto de fin de curso

Bases de Datos No Relacionales

Instituto de Computación, FING, UdelAR – 2025

CC-BY Lorena Etcheverry lorenae@fing.edu.uy

Objetivos

- Estudiar problemas de la gestión de datos en el contexto de las bases no relacionales
- Experimentar metodologías
- Analizar y comparar soluciones tecnológicas y herramientas

Ideas de proyectos



Familia 1: Migración entre modelos

Idea base: Partir de una base de datos existente (relacional o no relacional) y migrarla a otro modelo no relacional, justificando las decisiones de diseño y adaptando las consultas.

Ejemplos de variantes:

- Relacional → Documental (MongoDB)
- Documental → Grafos (Neo4j)
- Columnar → Clave-valor
- CSV + metadatos → Documental

¡Hacer comparaciones justas!

EJ: Consultas
“equivalentes”



Familia 2: Evaluación de performance

Idea base: Comparar rendimiento entre dos motores no relacionales usando un mismo conjunto de datos y consultas.

Ejemplos de variantes:

- MongoDB vs Couchbase en consultas agregadas
- Neo4j vs OrientDB en *graph traversals*
- Redis vs Cassandra en operaciones de lectura/escritura

Hacer comparaciones justas.

- Mismo HW
- Consultas “equivalentes”



Familia 3: Exploración en profundidad de un motor

Idea base: Profundizar en aspectos internos, configuraciones o mecanismos de optimización de un motor no relacional.

Variantes:

- Indexación y performance en MongoDB
- Cypher tuning en Neo4j
- Replicación y consistencia en Cassandra

Familia 4: Coast to coast

Idea base: Resolver un problema real o responder preguntas relevantes dentro de un dominio específico, recorriendo todas las etapas del ciclo de vida de los datos: obtención, análisis, modelado, transformación, persistencia y explotación.

Características clave:

- Se parte de un conjunto de preguntas o necesidades en un dominio.
- Puede implicar **integrar datos de distintas fuentes y formatos** (CSV, APIs, JSON, HTML, etc.).
- Se espera que el grupo **justifique técnicamente el uso de uno o más modelos no relacionales** adecuados para representar y procesar la información.
- La solución puede ser **poliglota**, combinando distintas tecnologías según el tipo de datos o etapa del proceso.

Familia 4: Coast to coast (cont)

Formato esperado:

- Diseño detallado de la solución, Prototipo funcional de al menos una parte sustantiva del sistema, Justificación clara de las decisiones tomadas (modelado, tecnologías, integración).

Ejemplos de variantes:

- Datos de movilidad urbana (archivos CSV + API en tiempo real) → análisis documental y temporal.
- Seguimiento de actividad en redes + datos personales → modelo de grafos y documental.
- Datos de sensores ambientales + información institucional → clave-valor + documental.
- Datos educativos abiertos + indicadores institucionales → combinación documental / grafos / analítico.

Un caso coast to coast: Ecosistema de IA en Uruguay

- Datos relevados por encuestas en un Proyecto de iniciación a la investigación (CSIC, Cattivelli, Defranco, Etcheverry, Bianchi)
- Personas en ecosistemas, relaciones entre ellas
- Interesa modelarlo como grafos (multicapa con conexiones) y hacer algo de analítica.
- También interesa enriquecerlo construyendo grafo de relaciones de coautoría (para algunos investigadores dados)

Máximo 2 grupos

Familia 5: Modelado cruzado o variantes

Idea base: Tomar un mismo dominio y modelarlo en al menos dos modelos no relacionales distintos, discutiendo ventajas y desventajas.

Variantes:

- Red académica modelada en grafos y documental
- Sistema de pedidos modelado en documental y clave-valor
- Datos meteorológicos en columnar vs documental

Es importante hacer comparaciones justas.

- Consultas “equivalentes”
- Diseños “razonables”



Coast to coast : ejemplos ediciones anteriores

- Reporte sobre [el número de Abreu](#)
- Fórmula 1, NBA, otros deportes (Neo4j)
- Recomendación de películas (Neo4j)
- Viveza criolla en la musica uruguaya
- Analisis de la conversacion de Twitter en Uruguay utilizando Neo4j
- Base Neo4 base Spotify para consultas y aplicar algoritmos de Aprendizaje Automático.
- Compras estatales: mejorar eficiencia, combatir corrupción y detección de comunidades.
- Estudio de persistencia de datos del STM.
- Análisis del sistema de recolección de contenedores.
- Comportamiento termodinámico del fondo de un reactor de Producción de pulpa marrón (UPM).
- Modelado datos temporales y espaciales de meteorología.

Familia 6: Relevamiento y comparación de bases analíticas

Idea base: Evaluar y comparar soluciones orientadas a análisis de datos (e.g., Druid, ClickHouse, Apache Pinot, etc.), con foco en instalación, modelo y consultas.

Variantes:

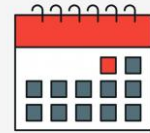
- Relevamiento de documentación + hands-on mínimo (tablas + queries)
- Comparación de performance en operaciones OLAP simples
- Caso práctico con carga desde archivos y consultas

Forma de trabajo



- Cada grupo deberá elegir una familia de problema y elaborar un plan de trabajo a realizar en 6 semanas (60 horas): template publicado. Defensas: a partir del 30/6
- Realizaremos checkpoints intermedios de avance (por zoom)
- Comunicación fluida con los docentes por EVA y correo. Instancias de consulta y seguimiento.

Cronograma



Fecha límite	Entregable / Actividad	Objetivo / Contenido esperable
Lunes 12/5	 Entrega de la primera versión del Plan de Trabajo	Objetivos claros, familia elegida, tareas planificadas, tecnologías tentativas.
Miércoles 14/5	 Devolución de planes + comienzo oficial del proyecto	Feedback y ajustes.
Lunes 19/5	 Checkpoint 1: Modelado + planificación detallada	- Diagrama/modelo no relacional- Decisiones de diseño- Plan técnico de tareas por etapa
Lunes 2/6	 Checkpoint 2: Prototipo parcial funcional + carga de datos reales	- Datos obtenidos, transformados e integrados- Funcionalidad básica demostrable- Capturas o scripts
Lunes 16/6	 Checkpoint 3: Versión preliminar del informe final	- Borrador casi completo- Elementos técnicos y de análisis- Identificación de puntos por cerrar
Miércoles 25/6	 Entrega final del informe	Informe completo, incluye: descripción, modelado, decisiones, desarrollo, resultados, dificultades.
Desde lunes 30/6	 Defensas orales	Presentación + preguntas. 8–10 minutos máximo.

Entrega final



- Un informe del trabajo realizado siguiendo el template publicado
- Código y datos utilizados accesibles en algún repositorio (GitLab de FING preferentemente)
- Una presentación/defensa de 10 minutos