

Universidad de Puerto Rico en Humacao
Departamento de Matemáticas
Guía de Estudio COMP 4098 Enero – Mayo 2026

Profesor: Ollantay Medina
Correo Electrónico: ollantay.medina@upr.edu
Oficina: CNO-190
Horas de oficina: Lu,Mi,Vi: 12:30pm-2:30pm

Texto:

Watt, A., College, D., Eng, N., (2014) “Database Design – 2nd Edition”

Descripción del Curso:

Este curso transiciona de sistemas de bases de datos tradicionales a ingeniería de datos moderna y arquitecturas adaptadas para Ciencia de Datos. Los estudiantes dominarán las arquitecturas que potencian analítica moderna e IA. El currículo contrasta sistemas Transaccionales (OLTP) con motores Analíticos (OLAP), y explora un stack de persistencia políglota—incluyendo PostgreSQL, DuckDB, MongoDB y Bases de Datos Vectoriales—para manejar datos estructurados, semi-estructurados y no estructurados esenciales para aplicaciones de IA como Retrieval-Augmented Generation (RAG).

Objetivos del Curso:

Al finalizar este curso, los estudiantes podrán:

- 1. Distinguir y Diseñar Sistemas de Datos:** Diseñar soluciones usando el motor de almacenamiento apropiado (Orientado a Filas vs. Orientado a Columnas) basado en requisitos de carga de trabajo (OLTP vs. OLAP).
- 2. Dominar SQL Analítico:** Ir más allá de CRUD básico para realizar análisis de datos complejos usando Funciones de Ventana, CTEs y agregaciones para reportes de alto rendimiento.
- 3. Gestionar Diversos Modelos de Datos:** Modelar y consultar efectivamente datos usando almacenes Relacionales (PostgreSQL), de Documentos (MongoDB) y Clave-Valor (Redis).
- 4. Implementar Capas de Acceso de Datos en Producción:** Implementar Capas de Acceso a Datos (DAL) robustas usando Connection Pooling, DAOs y ORMs para desacoplar la lógica de aplicación de la infraestructura de base de datos.
- 5. Construir Pipelines de Datos para IA:** Integrar Bases de Datos Vectoriales y Embeddings para soportar búsqueda semántica y construir sistemas de Retrieval-Augmented Generation (RAG) para datos no estructurados.

TEMAS
Módulo 1: Fundamentos Relacionales y SQL Transaccional (4 semanas) Modelo relacional, llaves e integridad referencial. Modelado ER. Normalización y calidad de datos. SQL transaccional con PostgreSQL (DDL, DML, CRUD).
Módulo 2: SQL Analítico y Arquitecturas (4 semanas) OLTP vs. OLAP y almacenamiento orientado a filas vs. columnas. DuckDB para análisis. Joins complejos, agregaciones, funciones de ventana (Window Functions), CTEs. Laboratorio de rendimiento comparando PostgreSQL y DuckDB.
Módulo 3: NoSQL y Modelos de Datos Flexibles (3 semanas) Teorema CAP y sistemas distribuidos. Bases de datos de documentos (MongoDB) y bases de datos clave-valor (Redis). Modelado y consulta de datos semi-estructurados.
Módulo 4: Capa de Acceso a Datos y Arquitectura (1 semana) Arquitectura de la Capa de Acceso a Datos (DAL). Patrones de implementación: DAO, Repository, ORM. Connection Pooling e independencia de base de datos.
Módulo 5: Arquitecturas Modernas e IA (3 semanas) Datos no estructurados y embeddings vectoriales. Bases de datos vectoriales y búsqueda por similitud. Arquitectura RAG (Retrieval-Augmented Generation).
Exposición Proyecto Final: Fecha anunciada por Registraduría

Curva de Notas: 100 - 85 (A); 84 - 75 (B); 74 - 60 (C); 59 - 50 (D); 49 – 0 (F)

Evaluación:

75 % Tres Exámenes Parciales
25 % Proyecto Final

Reglas:

1. El curso está catalogado como uno presencial. Se requiere la asistencia regular y puntual a las clases. Asignaciones o pruebas cortas presenciales NO se reponen.
2. Se utilizará la plataforma de Moodle Institucional como recurso de enseñanza complementario y en caso de que surja la necesidad, la plataforma de ZOOM para ofrecer las conferencias.
3. A menos que el profesor lo autorice, NO se permite el uso de teléfonos celulares durante la clase (Cert. # 1994-95-42 de la Junta Académica de la UPRH).
4. Se presume que todo trabajo sometido ha sido hecho por el estudiante. Trabajos entregados que reflejan una infracción evidente a esta regla recibirán cero automáticamente.