



Bases de Datos Avanzadas

1. Características generales

Nombre:	Bases de Datos Avanzadas
Sigla:	CI-0141
Créditos:	4
Horas lectivas:	5
Requisitos:	CI-0127 Bases de datos, CI-0128 Proyecto Integrador de Ingeniería de Software y Bases de Datos
Correquisitos:	Ninguno
Clasificación:	Curso propio
Ciclo de carrera:	I ciclo, 4to año (énfasis de Ingeniería de Software y énfasis de Ingeniería de Tecnologías de la Información) y II ciclo, 4to año (electiva del énfasis de Ciencias de la Computación)
Docente(s):	Dr. Luis Gustavo Esquivel Quirós
Datos de contacto:	Oficina: 4-4 (edificio anexo), Correo: luis.esquivel@ucr.ac.cr
Grupo:	01
Semestre y año:	I ciclo 2025
Horario y lugar de clases:	K 10 a 11:50 y V 9 a 11:50, aula 205
Modalidad:	Bajo virtual
Horario y lugar de consulta:	K 13 a 15:30, oficina 4-4 La consulta también se puede realizar utilizando Zoom. Para ello, se requiere coordinar una cita con al menos 24 horas de anticipación, para reservar el espacio y enviar los enlaces de la sesión.
Asistente:	Diego José Fernández Montealegre (diego.fernandezmontealegre@ucr.ac.cr)





2. Descripción

Las bases de datos relacionales centralizadas representan una alternativa de almacenamiento bien conocida, para la cual se pueden aplicar técnicas de mejoramiento de desempeño de consultas frente al constante crecimiento del volumen de los datos. Además, aplicando el paradigma multidimensional al almacenamiento relacional, se pueden guardar y analizar datos históricos de manera más eficiente. Sin embargo, el creciente volumen de datos y variedad de aplicaciones no tradicionales (por ejemplo, datos que por su naturaleza forman grafos o son semiestructurados o no estructurados) requieren un conocimiento nuevo sobre alternativas de modelado e implementación de bases de datos. En respuesta a estos retos, este curso se enfoca hacia diferentes mecanismos para optimizar el desempeño de consultas en bases de datos relacionales con grandes volúmenes de datos, introduce el paradigma de diseño multidimensional para almacenar y analizar datos históricos e incluye conceptos relacionados con bases de datos distribuidas y paralelas para solucionar los problemas de eficiencia y distribución geográfica de datos. Además, este curso introduce a los estudiantes a bases de datos del paradigma NoSQL tales como orientadas a documentos, orientadas a grafos y orientadas a columnas, para satisfacer las necesidades de aplicaciones no tradicionales y el manejo de *Big data*.

3. Objetivos

Objetivo general

El *objetivo general* del curso es que los estudiantes desarrollen habilidades para implementar diferentes tipos de bases de datos multidimensionales, paralelas, distribuidas y NoSQL, mediante estrategias que integren lo teórico y lo práctico, incluyendo un fuerte componente de actividades en el laboratorio, con el fin de que construyan bases de datos eficientes y acordes a las particularidades del problema a resolver.

Objetivos específicos

Durante este curso cada estudiante desarrollará habilidades para:

1. Diseñar e implementar bases de datos relacionales mediante la aplicación del paradigma multidimensional, con el propósito de construir un almacenamiento de datos históricos apto para el análisis.
2. Experimentar con bases de datos distribuidas y paralelas mediante ejercicios prácticos, para resolver problemas de eficiencia y distribución geográfica de datos.
3. Construir bases de datos no tradicionales usando las plataformas NoSQL para crear sistemas que satisfagan las necesidades de aplicaciones no tradicionales y manejo de *Big data*.
4. Identificar diferentes mecanismos de recuperación ante fallos, mediante la discusión de ejemplos, para entender el comportamiento de la base de datos ante fallos.





4. Contenidos

Objetivo específico	Eje temático	Desglose
1	Paradigma multidimensional de diseño, implementación y consultas	Diseño e implementación: Concepto de bases de datos analíticas con datos históricos. Diferencias entre bases de datos operacionales o transaccionales y bases de datos analíticas. Modelo multidimensional para el almacenamiento de datos históricos y facilidades de consultas. Aplicación del modelado multidimensional para el diseño e implementación de almacenes de datos.
		Consultas: Concepto de cubo OLAP (<i>On-Line Analytical Processing</i>) y su relación con el modelo multidimensional. Creación de cubos a partir de los almacenes de datos. Consultas automatizadas tipo <i>drill-down</i> , <i>roll-up</i> y <i>pivot</i> .
2	Bases de datos distribuidas y paralelas	Bases de datos distribuidas: arquitectura, distribución y replicación de datos, procesamiento de consultas, manejo de transacciones, control de concurrencia distribuido, fallas y recuperación del sistema.
		Bases de datos paralelas: arquitectura, distribución de datos, procesamiento paralelo de consultas y balanceo de carga.
2	Bases de datos NoSQL y su relación con Big Data	Conceptos básicos de modelado e implementación: Tipos de bases de datos NoSQL: llave-valor, orientadas a documentos, orientadas a grafos y orientadas a columnas.
		Consultas y transacciones: creación de consultas, concepto de transacción y diferentes acercamientos sobre aspectos de consistencia y disponibilidad, por ejemplo ACID (<i>Atomicity-Consistency-Isolation-Durability</i>), CAP (<i>Consistency-Availability-Partition Tolerance</i>) y BASE (<i>Basically Available – Soft State –Eventual Consistency</i>).





Objetivo específico	Eje temático	Desglose
		Uso de bases de datos NoSQL: Selección del tipo de base de datos NoSQL de acuerdo al tipo de aplicación. <i>Big data</i> y su manejo en las bases de datos NoSQL.
4	Recuperación ante fallos	Fallos y su recuperación: Tipos de fallos en bases de datos. Bitácora del sistema. Diferentes mecanismos usados por el sistema de recuperación del SGBD (protocolos <i>No-Undo/Redo</i> , <i>Undo/Redo</i> , <i>Undo/No-Redo</i> , <i>No-Undo/No-Redo</i>).

5. Metodología

Este curso busca la activa participación de las personas estudiantes en el proceso de enseñanza- aprendizaje y por esta razón requiere del compromiso constante con el curso. Se espera que las personas estudiantes se preparen para cada clase con la lectura previa del material asignado para aprovechar mejor las actividades en clase. La verificación de la preparación se hará mediante exámenes cortos y/o actividades didácticas. Para obtener una buena calificación, las personas estudiantes deben mostrar evidencia de su dominio sobre el material asignado. Se irán indicando oportunamente las lecturas por unidad temática del curso y la fecha para la cual debe realizarse cada lectura.

Los **exámenes cortos** tienen como objetivo evaluar el conocimiento adquirido por las personas estudiantes durante el curso. Los exámenes cortos realizados en clase pueden ser realizados en cualquier momento. Una vez realizado este aviso, esta forma de evaluación es válida, según la reglamentación universitaria. Por ese motivo se les recomienda estar atentos a lo solicitado en todas las clases. Los exámenes cortos realizados por medio de Mediación Virtual se realizarán con previo aviso para que las personas estudiantes puedan organizar su tiempo extra-clase. Los exámenes cortos pueden incluir materia vista en clases, en laboratorios o en el proyecto. No se hará la reposición de exámenes cortos. Los exámenes cortos son acumulativos.

El **proyecto** tiene como objetivo enfrentar y solucionar los nuevos retos que presentan las tecnologías estudiadas en el curso. El proyecto se debe realizar en grupos de tres o cuatro personas. Los detalles tanto del proyecto como de la documentación y cortes para demostrar el progreso de la ejecución del proyecto serán discutidos en clase y se entregarán como documentos aparte.

El **trabajo de laboratorio y prácticas dirigidas** permiten a las personas estudiantes practicar los conceptos vistos en clase para su mejor asimilación y para facilitar la preparación a los exámenes cortos. Los resultados de cada laboratorio y práctica





dirigida deben ser documentados por medio de un reporte en formato pdf, discutidos o comentados en cada uno de sus pasos y entregados para su evaluación, junto con la evidencia adicional de su realización (proyectos, scripts, hojas electrónicas, entre otros).

Las **tareas y actividades didácticas** sirven para verificar o complementar los conocimientos teóricos/prácticos adquiridos como resultado de las clases y el trabajo de laboratorio. Así como medios para que las personas estudiantes realicen pequeñas investigaciones o actividades sobre temas específicos. Las tareas y actividades didácticas se pueden asignar en cualquier momento y realizarse durante el horario de clase. Una vez realizado este aviso, esta forma de evaluación es válida, según la reglamentación universitaria.

Se utilizará la plataforma mediación virtual de la UCR (Metics) como apoyo a las siguientes actividades del curso: distribución de materiales y recursos, comunicación de eventos y anuncios, realización de diferentes tipos de exámenes cortos, como también especificación y entrega de tareas, proyecto y reportes de trabajo de laboratorio o prácticas dirigidas. Es responsabilidad de todo estudiante registrarse en dicha plataforma y estar pendiente de las noticias y eventos que allí se publiquen. El uso de esta plataforma virtual NO sustituye las lecciones presenciales ni las instrucciones emitidas por el docente durante las horas lectivas. Es importante recalcar que **no** todas las instrucciones dadas en clase se publicarán en la plataforma.

Políticas del curso respecto a entregas tardías de trabajo, reposiciones y otros:

- No se aceptarán trabajos después de la fecha estipulada de entrega.
- Sólo se repondrán evaluaciones bajo el procedimiento descrito en el artículo 24 de Reglamento de Régimen Académico Estudiantil de la UCR.
- Todo trabajo escrito estará sujeto a la evaluación de su redacción, ortografía y referencias.

El plagio no será tolerado en ninguno de los trabajos del curso y será castigado con una nota de cero en el trabajo correspondiente, aun cuando el plagio se haya dado en sólo una parte de este. Adicionalmente, el docente podrá elevar el caso a las instancias correspondientes de la universidad, según indica el Reglamento de Orden y Disciplina Estudiantil de la UCR.

6. Evaluación

Aspecto evaluativo	Porcentaje
Exámenes cortos	30%
Proyecto	30%
Trabajo de laboratorio y prácticas dirigidas	30%
Tareas y actividades didácticas	10%

7. Cronograma





Actividad	Fecha
Inicio de lecciones	10 de marzo
Batalla de Rivas	11 de abril
Semana Santa	14 al 18 de abril
Semana Universitaria	21 al 25 de abril
Primer corte de proyecto	Semana 8 del curso
Segundo corte de proyecto	Semana 16 del curso
Fin de lecciones	4 de julio
Exámenes cortos	Uno cada dos o tres semanas
Trabajo de laboratorio y prácticas dirigidas	Uno cada dos otras semanas

8. Bibliografía

1. Elmasri R. y Navathe S. “Fundamentos de Sistemas de Bases de Datos”, 6ta ed. Pearson-Addison Wesley, 2010.
2. Vaisman, A., & Zimányi, E. (2022). Data warehouse systems. In Data-centric systems and applications. <https://doi.org/10.1007/978-3-662-65167-4>
3. Özsü, M. T., & Valduriez, P. (2019). Principles of distributed database systems. In Springer eBooks. <https://doi.org/10.1007/978-3-030-26253-2>
4. Amazon. (2018). Amazon DynamoDB. [Online]. Available: <https://aws.amazon.com/dynamodb/> [Accessed October 2, 2018].
5. Arora R. and Aggarwal R. (2013). Modeling and querying data in MongoDB. International Journal of Scientific and Engineering Research 4(7):141–144.
6. Brewer E. (2000). Towards Robust Distributed Systems. Keynote at the ACM Symposium on Principles of Distributed Computing. [Online] Available: <http://www.cs.berkeley.edu/~brewer/cs262b-2004/PODC-keynote.pdf> [Accessed February 12, 2017].
7. Brewer E. (2012). CAP Twelve years later: How the “Rules” have changed. Computer 45(2): 23-29, IEEE Computer Society.
8. Kshemkalyani A. and Singhal M. (2008). Distributed Computing: Principles, Algorithms, and Systems. Cambridge University Press.
9. Kleppmann M. (2015). A Critique of the CAP Theorem. [Online]. Available: <https://www.cl.cam.ac.uk/research/dtg/www/files/publications/public/mk428/cap-critique.pdf> [Accessed January 13, 2018].
10. Kermarrec A.M. and van Steen M. (2007). Gossiping in distributed systems. ACM SIGOPS Operating Systems Review 42(5):2-7.
11. Kaur K. and Rani K. (2013). Modeling and querying data in NoSQL databases. In Proc. of the IEEE International Conference on BigData.
12. Demchenko, Y., Cuadrado-Gallego, J. J., Chertov, O., & Aleksandrova, M. (2024). Big Data Infrastructure technologies for data analytics. <https://doi.org/10.1007/978-3-031-69366-3>

9. Recursos estudiantiles





Para información sobre recursos estudiantiles disponibles en la UCR, incluyendo el Sistema de bibliotecas y la normativa universitaria vigente, favor visitar la página:

<https://www.ecci.ucr.ac.cr/vida-estudiantil/servicios-institucionales-para-estudiantes/guia-de-recursos-estudiantiles-de-la-ucr>

