

Universidad Católica San Pablo
Facultad de Ingeniería y Computación
Escuela Profesional de
Ciencia de la Computación
SILABO



CS271T. Bases de Datos II (Obligatorio)

2017-II

1. DATOS GENERALES

1.1 CARRERA PROFESIONAL	:	Ciencia de la Computación
1.2 ASIGNATURA	:	CS271T. Bases de Datos II
1.3 SEMESTRE ACADÉMICO	:	5 ^{to} Semestre.
1.4 PREREQUISITO(S)	:	CS270T. Bases de Datos I. (4 ^{to} Sem)
1.5 CARÁCTER	:	Obligatorio
1.6 HORAS	:	1 HT; 2 HP; 2 HL;
1.7 CRÉDITOS	:	3

2. DOCENTE

Dr. Javier Leandro Tejada Cárcamo

- Dr. Ciencia de la Computación, IPN, México, 2009.

3. FUNDAMENTACIÓN DEL CURSO

La Gestión de la Información (*IM-Information Management*) juega un rol principal en casi todas las áreas donde los computadores son usados. Esta área incluye la captura, digitalización, representación, organización, transformación y presentación de información; algoritmos para mejorar la eficiencia y efectividad del acceso y actualización de información almacenada, modelamiento de datos y abstracción, y técnicas de almacenamiento de archivos físicos. Este también abarca la seguridad de la información, privacidad, integridad y protección en un ambiente compartido. Los estudiantes necesitan ser capaces de desarrollar modelos de datos conceptuales y físicos, determinar que métodos de IM y técnicas son apropiados para un problema dado, y ser capaces de seleccionar e implementar una apropiada solución de IM que refleje todas las restricciones aplicables, incluyendo escalabilidad y usabilidad.

4. SUMILLA

1. IM/Diseño Físico de Bases de Datos.2. IM/Procesamiento de Transacciones.3. IM/Almacenamiento y Recuperación de Información.4. IM/Bases de Datos Distribuidas.

5. OBJETIVO GENERAL

- Hacer que el alumno entienda las diferentes aplicaciones que tienen las bases de datos, en las diversas áreas de conocimiento.
- Mostrar las formas adecuadas de almacenamiento de información basada en sus diversos enfoques y su posterior recuperación de información.

6. CONTRIBUCIÓN A LA FORMACIÓN PROFESIONAL Y FORMACIÓN GENERAL

Esta disciplina contribuye al logro de los siguientes resultados de la carrera:

- b) Analizar problemas e identificar y definir los requerimientos computacionales apropiados para su solución. [Nivel Bloom: 4]
- d) Trabajar efectivamente en equipos para cumplir con un objetivo común. [Nivel Bloom: 3]
- e) Entender correctamente las implicancias profesionales, éticas, legales, de seguridad y sociales de la profesión. [Nivel Bloom: 3]
- i) Utilizar técnicas y herramientas actuales necesarias para la práctica de la computación. [Nivel Bloom: 3]
- j) Aplicar la base matemática, principios de algoritmos y la teoría de la Ciencia de la Computación en el modelamiento y diseño de sistemas computacionales de tal manera que demuestre comprensión de los puntos de equilibrio involucrados en la opción escogida. [Nivel Bloom: 3]

7. CONTENIDOS

UNIDAD 1: IM/Diseño Físico de Bases de Datos.(10 horas)

Nivel Bloom: 4

OBJETIVO GENERAL

CONTENIDO

- Explicar los conceptos de registros, tipos de registros y archivos, así como también las diferentes técnicas para colocar registros de archivos en un disco.
- Dar ejemplos de aplicaciones de índices primarios, secundarios y clusterizados.
- Distinguir entre un índice denso y uno no denso.
- Implementar índices multinivel dinámicos usando árboles B.
- Explicar la teoría y aplicación de técnicas *hash* interno y externo.
- Usar dispersión (*hash*) para facilitar la expansión de archivos dinámicos.
- Describir las relaciones entre compresión, dispersión (*hash*) y búsquedas eficientes en base de datos.
- Evaluar costos y beneficios de diferentes esquemas de dispersión.
- Explicar cómo el diseño la base de datos físicas afecta a la eficiencia de las transacciones en base de datos..

- Almacenamiento y estructura de archivo.
- Archivos indexados.
- Archivos *hashed*.
- Archivos de firma (*signature*).
- Árboles B.
- Archivos con índice denso.
- Archivos con registros de longitud de variable.
- Eficiencia de la base de datos y afinamiento (*tuning*).

Lecturas: [Burlson, 2004], [Date, 2005], [Celko, 2005]

UNIDAD 2: IM/Procesamiento de Transacciones.(12 horas)	
Nivel Bloom: 5	
OBJETIVO GENERAL	CONTENIDO
<ul style="list-style-type: none"> ▪ Crear una transacción mediante la incrustación de SQL en un programa de aplicación. ▪ Explicar el concepto de compromiso implícito. ▪ Describir los temas específicos para la ejecución de transacciones eficientes. ▪ Explicar cuándo y por qué el <i>rollback</i> es necesario y cómo el <i>logging</i> asegura un <i>rollback</i> apropiado. ▪ Explicar los efectos de los diferentes niveles de aislamiento sobre los mecanismos de control de concurrencia. ▪ Escoger el nivel de aislamiento apropiado para implementar un protocolo de transacción especificado. 	<ul style="list-style-type: none"> ▪ Transacciones. ▪ Recuperación y falla. ▪ Control de concurrencia.
Lecturas: [Bernstein and Newcomer, 1997], [Elmasri and Navathe, 2004]	

UNIDAD 3: IM/Almacenamiento y Recuperación de Información.(10 horas)	
Nivel Bloom: 3	
OBJETIVO GENERAL	CONTENIDO
<ul style="list-style-type: none"> ▪ Explicar el almacenamiento de información básico y conceptos de recuperación.. ▪ Describir qué temas son específicos para la recuperación de información eficiente. ▪ Dar aplicaciones de estrategias de búsqueda alternativa y explicar por qué la estrategia de búsqueda particular es apropiada para la aplicación. ▪ Realizar investigación basada en Internet. ▪ Diseñar e implementar un sistema de almacenamiento y recuperación de tamaño pequeño a medio. 	<ul style="list-style-type: none"> ▪ Caracteres, cadenas, códigos, texto. ▪ Documentos, publicación electrónica, marcado <i>markup</i> y lenguaje de marcado. ▪ Árboles digitales, archivos invertidos, árboles PAT, archivos <i>signature</i>, indexación. ▪ Análisis morfológico, extracción de la raíz de una palabra (<i>stemming</i>), frases, <i>stop list</i>. ▪ Distribución de la frecuencia de términos, incerteza, difusibilidad, por peso. ▪ Espacio vectorial, probabilísticos, lógico y modelos avanzados. ▪ Necesidad de información, relevancia, evaluación, efectividad. ▪ Tesauro, ontologías, clasificación y categorización, metadata. ▪ Información bibliográfica, bibliometría, citas. ▪ Ruteo y filtrado (en comunidad). ▪ Búsqueda y estrategias de búsqueda, comportamiento de búsqueda de información, modelamiento de usuario, retroalimentación. ▪ Sumarización y visualización de información. ▪ Integración de citas, palabras clave, esquemas de clasificación y otros términos. ▪ Sistemas y protocolos (incluyendo Z39.50, OPACs, motores WWW, sistemas de investigación).
Lecturas: [Brusilovsky et al., 1998], [Elmasri and Navathe, 2004]	

UNIDAD 4: IM/Bases de Datos Distribuidas.(36 horas)	
Nivel Bloom: 4	
OBJETIVO GENERAL	CONTENIDO
<ul style="list-style-type: none"> ▪ Explicar las técnicas usadas para la fragmentación, replicación, alocaación de datos durante el proceso de diseño de bases de datos distribuidas. ▪ Evaluar estrategias simples para ejecutar una consulta distribuida para seleccionar la estrategia que minimice la cantidad de transferencia de datos. ▪ Explicar cómo el protocolo de <i>commit</i> en dos fases es usado para tratar una transacción que accese a una base de datos almacenada en múltiples nodos. ▪ Describir el control de concurrencia distribuido basado en la distinción de técnicas de copiado y el método de voto. ▪ Describir los tres niveles de software en el modelo cliente-servidor. 	<ul style="list-style-type: none"> ▪ Almacenamiento de datos distribuido. ▪ Procesamiento de consultas distribuidas. ▪ Modelo de transacción distribuido. ▪ Control de concurrencia. ▪ Soluciones heterogéneas y homogéneas. ▪ Cliente-servidor.
Lecturas: [Ozsu and Valduriez, 1999], [Date, 2005]	

8. METODOLOGÍA
<p>Evaluación Permanente 1 : 20 %</p> <p>Examen Parcial : 30 %</p> <p>Evaluación Permanente 2 : 20 %</p> <p>Evaluación Final : 30 %</p> <p style="padding-left: 40px;">Trabajo Final : 50 %</p> <p style="padding-left: 40px;">Examen Final : 50 %</p>

9. EVALUACIONES
<p>Evaluación Permanente 1 : 20 %</p> <p>Examen Parcial : 30 %</p> <p>Evaluación Permanente 2 : 20 %</p> <p>Examen Final : 30 %</p>

Referencias

- [Bernstein and Newcomer, 1997] Bernstein, P. A. and Newcomer, E. (1997). *Principles of Transaction Processing, First Edition*. Morgan Kaufmann.
- [Brusilovsky et al., 1998] Brusilovsky, P., Kobsa, A., and Vassileva, J. (1998). *Adaptive Hypertext and Hypermedia, First Edition*. Springer.
- [Burlleson, 2004] Burlleson, D. K. (2004). *Physical Database Design Using Oracle*. CRC Press.

- [Celko, 2005] Celko, J. (2005). *Joe Celko's SQL Programming Style*. Elsevier.
- [Date, 2005] Date, C. (2005). *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition*. Elsevier.
- [Elmasri and Navathe, 2004] Elmasri, R. and Navathe, S. B. (2004). *Fundamentals of Database Systems, Fourth Edition*. Addison Wesley.
- [Ozsu and Valduriez, 1999] Ozsu, M. T. and Valduriez, P. (1999). *Principles of Distributed Database Systems, Second Edition*. Prentice Hall.