

CIENCIA DE LA COMPUTACIÓN

BASE DE DATOS II

3 CRÉDITOS



ÍNDICE

ASIGNATURA	3
DATOS GENERALES	3
Ciclo: 5º	3
Créditos: Tres (3) créditos	3
Horas de teoría: una (1) semanal	3
Horas de práctica: cuatro (4) semanales	3
Duración del período: dieciséis (16) semanas	3
Condición:	3
Modalidad: Virtual	3
Requisitos:	3
PROFESORES	3
Profesor coordinador del curso	3
Profesor(es) instructor(es) del curso	3
INTRODUCCIÓN AL CURSO	3
OBJETIVOS	4
COMPETENCIAS Y CRITERIOS DE DESEMPEÑO	5
RESULTADOS DE APRENDIZAJE	5
TEMAS	5
PLAN DE TRABAJO	6
Metodología	6
Sesiones de teoría	7
Sesiones de práctica (laboratorio o taller)	7
SISTEMA DE EVALUACIÓN	7
REFERENCIAS BIBLIOGRÁFICAS	8

UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA
SILABO 2021-1

1. ASIGNATURA

CS2702 - Bases de Datos II

2. DATOS GENERALES

2.1 Ciclo: 5º

2.2 Créditos: Tres (3) créditos

2.3 Horas de teoría: una (1) semanal

2.4 Horas de práctica: cuatro (4) semanales

2.5 Duración del período: dieciséis (16) semanas

2.6 Condición:

- Obligatorio para ciencia de la computación

2.7 Modalidad: Virtual

2.8 Requisitos:

- CS2701 – Base de Datos I. (4to Sem)

3. PROFESORES

3.1 Profesor coordinador del curso

Heider Sanchez Enriquez (hsanchez@utec.edu.pe)

Horario de atención: miércoles de 8:00 a 10:00 AM

3.2 Profesor(es) instructor(es) del curso

Heider Sanchez Enriquez (hsanchez@utec.edu.pe)

Horario de atención: miércoles de 8:00 a 10:00 AM

4. INTRODUCCIÓN AL CURSO

La gestión de la información (IM) 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.

5. OBJETIVOS

Sesión 1: Explicar los conceptos de registro, tipos de registro, y archivos, así como las diversas técnicas para colocar registros de archivos en un disco.

Sesión 2: Analizar y aplicar los principales métodos de organización de archivos. Describir la diferencia entre un índice no denso y uno denso. Evaluar el costo y beneficio de diversos esquemas de indexación en memoria secundaria.

Sesión 3: Explicar la teoría y la aplicación de técnicas de hash estática y dinámica, y describe las relaciones entre hashing, compresión, y búsquedas eficientes en bases de datos. Implementar índices de multinivel dinámicos usando árboles-B.

Sesión 4: Describir los problemas específicos para la ejecución de una transacción eficiente. Explicar el efecto de diferentes niveles de aislamiento sobre los mecanismos de control de concurrencia.

Sesión 5: Describir técnicas específicas para una recuperación eficiente de la base de datos frente a fallos de sistema.

Sesión 6: Explicar los conceptos básicos de recuperación de la información basado en el contenido. Analizar y diseñar técnicas eficientes de recuperación de documentos de texto: modelo booleano e índice invertido.

Sesión 7: Analizar y diseñar modelos de recuperación de documentos de texto por ranking en función de la similitud entre objetos.

Sesión 8: Describir algunas técnicas básicas relacionadas al archivamiento, preservación y recuperación de la información en una librería digital multimedia.

Sesión 9: Diseñar modelos avanzados para búsqueda eficiente en base de datos multimedia. Aplicar el índice multidimensional R-Tree en una solución real de recuperación de imágenes.

Sesión 10: Describir las características de los sistemas distribuidos.

Sesión 11: Explicar las técnicas usadas para la fragmentación de datos, replicación, y la asignación durante el proceso de diseño de base de datos distribuida.

Sesión 12: Evaluar estrategias simples para la ejecución de una consulta distribuida para seleccionar una estrategia que minimice la cantidad de transferencia de datos.

Sesión 13: Evaluar estrategias simples para la optimización de costos de asignación y costos de ejecución de consultas.

Sesión 14: Explicar cómo el protocolo de dos fases de commit es usado para resolver problemas de transacciones que acceden a bases de datos almacenadas en múltiples nodos.

Sesión 15: Explicar el uso del modelo de procesamiento MapReduce. Revisar las principales bases de datos NoSQL distribuidas.

6. COMPETENCIAS Y CRITERIOS DE DESEMPEÑO

Los criterios de desempeño que se van a trabajar en este curso son:

2.2. Formula problemas complejos de computación y otras disciplinas relevantes en el dominio (*nivel 2*).

3.2. Diseña, implementa y evalúa soluciones a problemas complejos de computación (*nivel 2*).

4.1. Crea, selecciona, adapta y aplica técnicas, recursos y herramientas modernas para la práctica de la computación y comprende sus limitaciones (*nivel 3*).

7. RESULTADOS DE APRENDIZAJE

Al final del curso de Base de datos II se espera que el estudiante sea capaz de:

RA1. Aplicar los conceptos de fragmentación, asignación y optimización de consultas distribuidas para resolver problemas de escalabilidad.

RA2. Simular un sistema gestor de base de datos enfocado en los diferentes métodos de organización de archivos y acceso concurrente.

RA3. Construir soluciones de recuperación de la información basado en el contenido usando técnicas de indexación para consultas eficientes y herramientas como MongoDB, Twitter API, Python Anaconda.

8. TEMAS

1. Diseño Físico de Bases de Datos

1.1. Almacenamiento y estructura de archivos

1.2. Organizando registros en un archivo

- 1.2.1. Registros de longitud fija
- 1.2.2. Registros de longitud variable
- 1.3. Métodos de organización de archivos básicos
 - 1.3.1. Archivos tipo Heap
 - 1.3.2. Archivos secuenciales / ordenados
 - 1.3.3. Archivos aleatorios / acceso directo
 - 1.3.4. Archivos secuenciales indexados
 - 1.3.5. Archivos Hash
 - 1.3.6. Árboles B+

2. Transacciones, Recuperación y Control de Concurrencia

- 2.1. Control de concurrencia
- 2.2. Prevención de interbloqueos
- 2.3. Recuperación de Base de Datos

3. Recuperación de Información basado en el contenido

- 3.1. Búsqueda y Recuperación de la Información
- 3.2. Recuperación de información documental
 - 3.2.1. Modelo booleano
 - 3.2.2. Índice Invertido
 - 3.2.3. Recuperación por ranking
- 3.3. Base de datos multimedia
 - 3.3.1. Conceptos de búsqueda por similitud
 - 3.3.2. Modelos de similitud para Objetos Multimedia
 - 3.3.3. Índices multidimensionales

4. Bases de Datos Distribuidas

- 4.1. Sistemas distribuidos
- 4.2. Diseño de base de datos distribuidas
 - 4.2.1. Fragmentación Horizontal
 - 4.2.2. Fragmentación Vertical
- 4.3. Procesamiento de consultas distribuidas
 - 4.3.1. Descomposición
 - 4.3.2. Localización
 - 4.3.3. Optimización
- 4.4. Transacciones y control de concurrencia distribuida
- 4.5. MapReduce
- 4.6. Base de datos NoSQL
 - 4.6.1. MongoDB
 - 4.6.2. Cassandra

9. PLAN DE TRABAJO

9.1 Metodología

Fecha de actualización: 09/04/2021

Revisado y aprobado por el Centro de Excelencia en Enseñanza y Aprendizaje y la Dirección de Ciencia de la Computación

Se fomenta la participación individual y en equipo para exponer sus ideas, motivándolos con puntos adicionales en las diferentes etapas de la evaluación del curso.

A lo largo del curso se proporcionan diferentes lecturas, las cuales podrían ser evaluadas. El uso de herramientas Online permite a cada estudiante acceder a la información del curso, e interactuar fuera de aula con el profesor y con los otros estudiantes.

9.2 Sesiones de teoría

Las sesiones de teoría se llevan a cabo en clases magistrales donde se realizarán actividades que propicien un aprendizaje activo, con dinámicas que permitan a los estudiantes interiorizar los conceptos.

Se desarrollará un proyecto enfocado a entender y usar una base de datos NoSQL (no solo SQL) y sus características distribuidas aplicado a un dominio particular de datos. Además, se debe realizar una implementación demostrativa de la BD asignada.

9.3 Sesiones de práctica (laboratorio o taller)

Las sesiones prácticas se llevan a cabo en clase donde se desarrollan una serie de ejercicios y/o conceptos prácticos mediante planteamiento de problemas, la resolución de problemas, ejercicios puntuales y/o en contextos aplicativos.

Cada tópico tiene dos laboratorios son del tipo individual cuando la actividad requiere el logro de destreza en el manejo de herramientas, y del tipo grupal cuando la actividad requiere de la participación de dos o tres personas para la concreción de un trabajo.

Las herramientas principales a utilizar son los siguientes:

- C++, para implementar los métodos de organización de archivos.
- Python V3, para implementar las técnicas de recuperación de la información.
- PostgreSQL V12, para la experimentación e implementación de base de datos distribuidas.

Se desarrollarán dos proyectos orientados a la aplicación de las técnicas aprendidas en escenarios reales. El primero, enfocado a la construcción de un mini-sistema gestor de base de datos implementando los métodos de organización de archivos. El segundo, enfocado a la construcción de un índice invertido para búsquedas eficientes en documentos de texto, y la construcción de un índice multidimensional para búsqueda eficiente de imágenes.

10. SISTEMA DE EVALUACIÓN

Sesiones Teóricas:

Las sesiones de teoría se llevan a cabo en clases magistrales donde se realizarán actividades que propicien un aprendizaje activo, con dinámicas que permitan a los estudiantes interiorizar los conceptos.

Sesiones de Laboratorio:

Para verificar que los alumnos hayan alcanzado el logro planteado para cada una de las unidades de aprendizaje, realizarán actividades que les permita aplicar los conocimientos adquiridos durante las sesiones de teoría y se les propondrá retos que permitan evaluar el desempeño de los alumnos.

Exposiciones individuales o grupales:

Se fomenta la participación individual y en equipo para exponer sus ideas, motivándolos con puntos adicionales en las diferentes etapas de la evaluación del curso.

Lecturas:

A lo largo del curso se proporcionan diferentes lecturas, las cuales podrían ser evaluadas. El uso de herramientas Online permite a cada estudiante acceder a la información del curso, e interactuar fuera de aula con el profesor y con los otros estudiantes.

Sistema de Evaluación:

La nota final NF se obtiene a través de:

EVALUACIÓN	TEORÍA	PRÁCTICA Y/O LABORATORIO
	Practica Calificada PC1 (13%) Practica Calificada PC2 (13%) Practica Calificada PC3 (14%)	Evaluación Continua C1 (10%) Evaluación Continua C2 (10%) Proyecto P1 (15%) Proyecto P2 (15%) Proyecto P3 (10%)
	40%	60%
	100%	

La ponderación de la evaluación se hará si ambas partes están aprobadas o siguiendo los parámetros decididos por la dirección de la carrera.

Las rúbricas que permitirán medir las actividades más significativas del curso y que, además se relacionan con la evaluación de las competencias del estudiante son: [enlace](#)

11. REFERENCIAS BIBLIOGRÁFICAS

Michael J. Folk; Bill Zoellick; Greg Riccardi. File Structures : An Object-Oriented Approach with C++, 1997.

Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze. Introduction to information retrieval, 2008.

Patrick Valduriez M. Tamer Ozsü. Principles of Distributed Database Systems, Second Edition. Prentice Hall, 1999.

Eric Newcomer Philip A. Bernstein. Principles of Transaction Processing, First Edition. Morgan Kaufmann, 1997.

Shamkant B. Navathe Ramez Elmasri. Fundamentals of Database Systems, Fourth Edition. Addison Wesley, 2004.

Donald K. Burleson. Physical Database Design Using Oracle. CRC Press, 2004