

Planificación de la Asignatura: Gestión de Datos Masivos

Fecha: 23/10/2024 13:02

Código: I1542

Carrera: Ingeniería en Transporte

Departamento Académico: Informática

Docente a cargo:

Correo del docente a cargo: ajhadad@bioingenieria.edu.ar

Régimen de Dictado: Cuatrimestral 1º Cuatrimestre

Carga Horaria Semanal: 4 horas semanales

Carga Horaria Total: 56 horas

Contenidos Mínimos:

Introducción al análisis de grandes volúmenes de datos. Fundamentos de bases de datos. Representación de la información. Entornos y sistemas para el análisis de grandes volúmenes de datos. Extracción de conocimiento de bases de datos. Aplicaciones.-

Correlativas Regulares para cursar:

TIC y Geomática

Correlativas Aprobadas para cursar:

Algoritmos y Estructuras de Datos

Correlativas Aprobadas para promocionar o rendir el examen final:

2º Año

Algoritmos y Estructuras de Datos

Objetivo General:

- Desarrollar los conceptos de cómo se estructura los datos en los dispositivos de almacenamiento.
- Introducir la importancia de las bases de datos y los modelos de datos para el desarrollo de Sistemas de Información.
- Enunciar y explicitar las distintas metodologías para el modelado de datos.
- Desarrollar los fundamentos y la arquitectura de los sistemas de gestión de datos masivos
- Describir los fundamentos y características de las bases de datos (y sus diferentes modelos), los lenguajes de definición, manipulación y consulta.
- Desarrollar los conceptos de bases de datos multidimensionales y los procesos de extracción de conocimiento de bases de datos.

Objetivos Particulares:

- Describir las distintas estructuras de archivos,: secuencial, indexados, directos, sus características principales, modos de accesos.
- Estudiar el modelo entidad/relación para el modelado de datos.
- Explicar los conceptos de tablas, relaciones, claves, integridad referencial, normalización.
- Desarrollar los elementos y sentencias del lenguaje de consulta estructurado (SQL) para la definición, manipulación de datos, administración de usuarios y seguridad.
- Describir los aspectos que hacen a la consistencia e integridad de los datos en una Base de Datos: transacciones, concurrencia, planificación, protocolos.
- Desarrollar los elementos y sentencias asociado a bases de datos NoSQL para la definición, manipulación de datos, administración de usuarios y seguridad.
- Desarrollar los conceptos de otros modelos de bases de datos multidimensionales: características, ventajas y desventajas.
- Desarrollar los procesos de extracción de conocimiento: fases, ventajas y desventajas.

Programa Analítico:

Unidad 1

Introducción a análisis de grandes volúmenes de datos. Datos Estructurados, Semiestructurados y no estructurados.

Unidad 2

Archivos: definición, conceptos básicos. Operaciones y Mantenimiento sobre archivos: eliminación de registros, reutilización de los espacios, fragmentación y administración del almacenamiento. Hashing: conceptos fundamentales. Métodos para resolver las colisiones..Archivos indexados:. Índices denso y escaso. Índices primario, de agregación y secundario. Índices de un solo nivel y multinivel. Métodos de acceso por árboles. Árboles B y B+.

Unidad 3

Fundamentos de sistemas de Bases de Datos: definición. Objetivos. Administrador de las bases de datos (DBMS). Arquitectura de tres niveles. Lenguajes de definición (DDL) y manipulación de datos (DML). Modelos de datos. Independencia de datos. Esquemas y subesquemas. Claves. Integridad Referencial. Transacciones. Propiedades ACID. Control de concurrencias. Bloqueo.. Protocolos basados en bloqueo.

Unidad 4

Bases de Datos Relacionales. Modelo Entidad Relación. Transformaciones a Esquema Modelo Entidad-Relación. Entidades, relaciones y atributos. Grados de una relación. Cardinalidad. Pertenencia de una relación. Diagramas Entidad-Relación. Transformación del modelo Entidad-Relación en tablas de una Base de Datos relacional

Unidad 5

Álgebra relacional: Operadores. Lenguajes relacionales: SQL. Sentencias de definición: CREATE, ALTER, DROP; manipulación: INSERT, UPDATE, DELETE; y búsqueda de datos: SELECT. Cláusulas FROM, WHERE. Cláusula GROUP By. Funciones de grupo: MIN, MAX, AVG, COUNT y SUM. Sentencias anidadas: cláusulas IN, SOME, ANY, ALL, EXISTS. Operaciones de "join": INNER, OUTER.

Unidad 6

Gestión de Datos orientadas al análisis de datos. Bases de Datos NoSQL. Propiedades. Teorema CAP.

Diferencias con el abordaje SQL. Modelos de datos utilizados. Recursos Tecnológicos Datawarehouse. Modelo multidimensional de bases de datos. Conceptos de dimensión. Introducción a “Datawarehouse” Modelos de datos. Arquitecturas. Diseño conceptual, lógico y físico. El modelo multidimensional: estrella, snowflake y constellation.

Unidad 7

Big data. Definiciones. Evolución. Plataformas de proceso de datos: sistemas paralelos, NoSQL y sistemas administradores de Big Data. El paradigma Map-Reduce. Hadoop File System. Spark.

Unidad 8

Extracción de conocimiento de bases de datos. Procesos KDD y KDP. Fases del proceso. Selección de variables y casos. Integración y limpieza de datos. Transformación de atributos. Normalización. Procesamiento. Evaluación e Interpretación.

Listado de Actividades de Formación Práctica:

- 1 - Guía de problemas de Archivos y Diagramas Entidad-Relación
- 2- Guía de Laboratorio. Sentencias del DDL
- 3- Guía de Laboratorio. Optimización de Consultas
- 4- Guía de problemas de SQL
- 5- Guía de problemas de Bases de Datos NoSQL
- 6 - Guía de problemas de Extracción de Conocimiento de Bases de Datos
- 7 - Trabajo Práctico Integrador

Metodología de Evaluación Durante el cursado:

Durante el cursado se efectúa un seguimiento continuo a través de las resoluciones de problemas en clases y el desarrollo de un trabajo práctico (TP) integrador que debe entregarse en tiempo y forma. El trabajo práctico se evalúa de dos formas, en primer lugar se los califica cuan bien se lograron las consignas y objetivos del TP, como su presentación, completitud, etc., además se realiza un coloquio de defensa de los mismos donde los alumnos deberán exponer los fundamentos, recursos y medios utilizados para la realización del TP. El TP consiste en la aplicación tecnologías relacionadas con la gestión de datos masivos en relación a los procesos de diseño, implementación, carga de datos y consultas. El tema estará relacionado aplicaciones en el área de transporte. También se evalúa mediante un examen parcial y su correspondiente recuperatorio.

Metodología de Evaluación en Exámenes Finales:

Los alumnos regulares rendirán un examen oral sobre los temas de la asignatura y sobre el trabajo práctico integrador presentado durante el cursado. El examen se considerará aprobado al alcanzar un 60% de puntuación.

Los alumnos libres rendirán inicialmente un examen escrito. Superada dicha instancia con una puntuación del 60% rendirán un examen oral en las mismas condiciones que el alumno regular.

Condiciones de Regularidad :

Para alcanzar la regularidad en la asignatura, los alumnos deben asistir al 75% de las clases prácticas y de laboratorio. Deben entregar en tiempo y forma el Trabajo Práctico y aprobar la defensa del mismo. Además deberán aprobar con 60% o más, el examen parcial ó su recuperatorio.

Para alcanzar la promoción en la asignatura, además de las condiciones de regularidad, debe aprobar con un 75% más la presentación y defensa del TP integrador.

Bibliografía Principal:

- 1- ELMASRI R. y NAVATHE S., "Sistemas de Bases de Datos", Addison-Wesley Iberoamericana, 3ra. Edición, 2002..
2. CABALLERO, MARTIN, RIESCO. Big Data con Python. Recolección, almacenamiento y proceso. Universidad Complutense de Madrid (2018)
3. BALDOMINOS GOMEZ, GONZALVEZ. Resolviendo problemas de Big Data. Un enfoque aplicado. Alfaomega (2020)
4. FERRI RAMÍREZ, HERNÁNDEZ ORALLO, RAMÍREZ QUINTANA. Introducción a la minería de datos. Pearson (2004)
5. MARZ, Nathan. Big Data. Principles and best practices of scalable real-time data systems
6. HARRISON, G. (2015). Next generation databases: NoSQL, NewSQL, and big data (1a ed.). APress.

Bibliografía Complementaria:

- 6- KORTH H. y SILBERSCHATZ A., "Fundamentos de Bases de Datos 3/e", McGraw-Hill (1998)
- 7- DATE C., "Introducción a las Bases de Datos 6e", Addison-Wesley Iberoamericana, 2000.
- 8- ROB P. y CORONEL C., "Sistemas de Bases de Datos. Diseño, implementación y administración",