

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.-

Competencias Genéricas:

CT 1: Identificación, formulación y resolución de problemas de Ingeniería en Transporte..

CT 3: Gestión, planificación, ejecución y control de proyectos de Ingeniería en Transporte..

Competencias Específicas:**Argumentación de aportes marcados en la matriz de competencias:**

Gestión de Datos Masivos aporta a la primera competencia genérica tecnológica: "Identificación, formulación y resolución de problemas en Ingeniería en Transporte" en una primera etapa con un nivel de dominio inicial, que permite el desarrollo de conocimientos disciplinares, habilidades y destrezas que serán el fundamento para la adquisición de esta competencia, a través del entendimiento de los patrones de diseño por lo cual se aborda en un primer nivel de dominio de la competencia.

En cuanto a las competencia "Gestión, planificación, ejecución y control de proyectos en Ingeniería en Transporte" la misma se evalúa a través del seguimiento continuo de un proyecto de gestión, manipulación y análisis de datos a partir de diferentes fuentes de datos utilizando tecnologías vinculadas a Sistemas de Gestión de bases de datos. Para el desarrollo de las actividades prácticas se utilizan lenguajes específicos y entornos de trabajo usados también en el ámbito profesional. Se trata de una aproximación al campo profesional, a través del desarrollo asistido de un proyecto, por lo cual se trabaja en un segundo nivel de dominio de la competencia.

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:

2° Año

Algoritmos y Estructuras de Datos

Insercion de la Asignatura en el plan de Estudios:

La asignatura es uno de los pilares en la práctica de cualquier profesional vinculado de alguna manera a la informática, porque lo introduce en los conceptos y fundamentos que son el soporte vinculados a los modelos de datos de la mayoría de las aplicaciones informáticas que el futuro profesional debe trabajar, desarrollar, auditar y/o administrar. Del análisis de las funciones y problemas que debe afrontar como profesional en cualquier tipo de organización, se desprende que hará un uso cotidiano de los conceptos aprendidos en la asignatura.

En el cursado de la asignatura el alumno emplea conceptos matemáticos adquiridos en “Álgebra Lineal” y “Probabilidad y estadística”. La asignatura “Algoritmos y Estructuras de datos” sirve para comprender como se estructuran los datos en los archivos (y por ende en las bases de datos) y ayudan a resolver problemas de modelado e implementación de programas y datos, conceptos que fueron desarrollados en “Fundamentos de Programación”. Los conceptos de modelado y diseño desarrollados serán de mucha utilidad en relación a los métodos abordados en “Investigación Operativo II” y en “TIC y Geomática” porque, como ya fue expresado, todo sistema por computadora, en general, está ligado a un sistema de gestión de datos. Los conceptos de gestión de datos masivos también serán útiles para el diseño de arquitecturas de sistemas de información y el empleo de métricas relacionadas con los modelos 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.

Metodología Didáctica:**Clases Teóricas**

Las clases teóricas tienen una duración entre 2 hs. semanales, tienen como objetivo principal la presentación de los conceptos, su análisis y la descripción de sus principales aplicaciones. La exposición se inicia con una parte introductoria donde se plantean los objetivos y se ubica lo que se va a desarrollar en el contexto de la asignatura. Luego una parte de desarrollo donde se exponen los contenidos en forma ordenada y clara. Finalmente una parte de conclusiones que contempla la revisión de lo expuesto, recalcando los contenidos más importantes. Para el dictado de las clases se emplean, los elementos tradicionales: tiza y pizarrón, como también el proyector de cristal líquido que es esencial para presentar gráficos y/o animaciones que ayuden al alumno a comprender el contenido. Esta clase son de tipo coloquial, tendiendo a que el alumno analice y comprenda los conceptos vertidos, tratando de captar su interés por los temas tratados.

Clases Prácticas

Las clases prácticas tienen una duración de 2 hs. semanales. Estas clases son de dos tipos: de resolución de problemas que se realizan en el aula, y de laboratorio, que se llevan a cabo en las aulas equipadas con computadoras personales.

En las clases de resolución de problemas se promueve el trabajo individual de los alumnos, motivando el razonamiento para aplicar los conceptos teóricos en situaciones prácticas, contribuyendo al desarrollo de la creatividad del alumno mediante el planteo de situaciones novedosas a las originalmente propuestas, motivando la búsqueda del resultado de problemas de diversa complejidad.

Las clases de laboratorio son, en general, clases de apoyo para la realización de los trabajos prácticos. Los alumnos trabajan con la computadora directamente. El objetivo además de los mencionados en el párrafo anterior para la resolución de problemas, es que el alumno se familiarice y tome contacto con los productos comerciales existentes, aprenda sus características y desarrolle una aplicación concreta.

Formación Práctica:

La instancia de resolución de ejercicios y problemas se concreta en la solución de guías de problemas en el contexto de situaciones reales o hipotéticas cuya solución requiere la aplicación de los conocimientos vertidos en las clases teóricas. Se promueve tanto el trabajo individual como grupal de los alumnos, motivando el razonamiento para aplicar los conceptos vertidos en las clases teóricas y lograr un resultado determinado a las premisas planteadas en los problemas.

En lo que refiere a las etapas de formación experimental y diseño estas consisten en el desarrollo de modelos de datos sobre tecnologías de gestión de datos. A los alumnos se les provee los requerimientos de los modelos de datos que deben desarrollar y ellos deberán proponer el diseño e implementación de la misma. Para ello deben integrar gran parte del contenido teórico desarrollado en clases en la resolución de un problema concreto. Para este caso se promueve el trabajo grupal de los alumnos, dado que en conjunto deben resolver un problema de diseño.

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

Intensidad de la formación práctica

Detalle de la carga horaria total prevista para cada una de las siguientes actividades:

Actividades prácticas que aportan a las competencias específicas en el Nivel de dominio 1: 0 horas

Actividades prácticas que aportan a las competencias específicas en el Nivel de dominio 2: 0 horas

Actividades prácticas que aportan a las competencias específicas en el Nivel de dominio 3: 0 horas

Horas totales de actividades de formación práctica: 40 horas

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.

Cronograma de parciales durante el primer Cuatrimestre:

Primer Examen Parcial: 13 de Mayo de 2024

Recuperatorio 01: 17 de Junio de 2024

Cronograma de parciales durante el segundo Cuatrimestre:

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",

Equipo de Cátedra:

El equipo de cátedra está compuesto por:

Prof. Titular Alejandro Hadad: Funciones: planificación, preparación y dictado de las clases teóricas

JTP Maximiliano Fernandez: Preparación y dictado de clases prácticas y de laboratorio.

Sr. Hermes Cettour: Auxiliar Alumno de 2da Cat. Funciones: colaborar durante el dictado de las clases prácticas

JTP Walter Elías: Afectado al dictado de las materias Bases de Datos Multidimensionales y Técnicas y Herramientas para la Gestión de Datos Masivos. Colaboración en temas específicos

Actividades de Investigación Gestión y Extensión:

Alejandro Hadad:

Investigación:

- 1.- Director del Laboratorio de Sistemas de Información
- 2.- Al comienzo del año se trabajará en el desarrollo del informe final del PID 6146 en curso "SISTEMA DE INFORMACIÓN BASADO EN NORMA DICOM PARA APLICACIONES OFTALMOLÓGICAS ORIENTADAS A RETINOPATÍAS DEL PREMATURO (ROP)" el cual finalizará en 2023.
- 3.- Se está trabajando en la formulación de un nuevo Proyecto de Investigación y en la evaluación de nuevas líneas de investigación

Gestión

- 1- Director del Departamento Académico Informática

Requisitos de admisión para alumnos oyentes:

Ser alumno de la Facultad de Ingeniería.

Infraestructura, equipamiento y recursos necesarios:

Laboratorio con computadoras personales.

Otros: