

Planificación de la Asignatura: Técnicas y Herramientas para Datos Masivos

Fecha: 23/10/2024 13:02

Código: T1641

Carrera: Tecnicatura Universitaria en Procesamiento y Explotación de Datos

Departamento Académico: Informática

Docente a cargo:

Correo del docente a cargo: alejandro.hadad@uner.edu.ar

Régimen de Dictado: Cuatrimestral 2º Cuatrimestre

Carga Horaria Semanal: 5 horas semanales

Carga Horaria Total: 75 horas

Contenidos Mínimos:

Bases de datos no relacionales (NoSQL). Conceptos fundamentales de Datos Masivos (Big Data). El paradigma Mapa Reducido (MapReduce). Sistema de archivos Hadoop. HadoopDB. Lenguajes de alto nivel: HiveQL y Pig Latin. Google Cloud SQL. Google Big Table. Utilización en Depósito de Datos (data warehousing). Análisis de datos con Hadoop y Hive.

Correlativas Regulares para cursar:

Minería de Datos

Visualización de la Información

Correlativas Aprobadas para cursar:

Espacio Integrador II

Inglés I

Aprendizaje Maquinal

Bases de Datos Multidimensionales

Gestión de Proyectos

Correlativas Aprobadas para promocionar o rendir el examen final:

No posee

Objetivo General:

- Desarrollar los conceptos de cómo se estructuran los datos en modelos no relacionales.
- Introducir la importancia de los modelos no relacionales 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 en la nube
- Describir los fundamentos y características de las bases de datos no relacionales, las operaciones de definición y manipulación de datos.

Objetivos Particulares:

- Estudiar los diferentes enfoques para el modelado de datos.
- Desarrollar los elementos y sentencias del lenguaje de consulta para la definición, manipulación de datos, administración de usuarios y seguridad.
- Describir los aspectos vinculados a la construcción de los procesos tradicionales en este contexto
- Describir los aspectos que hacen a la consistencia e integridad de los datos
- Describir las acciones a tomar relacionados con la seguridad física y lógica de los datos, resaltando su importancia.

Programa Analítico:

Unidad 1

Bases de datos no relacionales (NoSQL). Bases de datos documentales. Bases de datos clave-valor. Bases de Datos Columnares. Bases de datos basadas en grafos.

Unidad 2

Conceptos Fundamentales de Datos Masivos (Big Data). Evolución histórica. Análisis de las diferentes dimensiones de definen el al Big Data. Las 5 V de Big data.

Unidad 3

El paradigma Mapeo Reducido (Map Reduce). Sistema de Archivos Hadoop. Hadoop DB. Ejemplos

Unidad 4

Lenguajes de alto nivel (ej HiveQL y Pig Latin). Utilización en Depósito de Datos (data warehousing) Análisis de datos (ej: Hadoop y Hive).

Listado de Actividades de Formación Práctica:

TP 1 – Bases de Datos NoSQL

TP 2 – Procesamiento Distribuido y Almacenamiento en la nube

Metodología de Evaluación Durante el cursado:

Durante el cursado se efectúa un seguimiento continuo a través de las resolución de los problemas en clases y el desarrollo de dos trabajos prácticos (TP) grupal (2 alumnos por grupo) 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.

Metodología de Evaluación en Exámenes Finales:

El examen final para el alumno regular consiste en un examen teórico-práctico escrito el cual integra todos los temas dictados en la asignatura. Previo a dicho examen defenderá oralmente los TPs que se le plantearon durante el cursado

El examen final para el alumno libre consiste en la presentación y defensa del desarrollo de una aplicación: diseño, implementación, carga de datos y consultas, cuyo enunciado será entregado con una semana de anticipación. Superada esta instancia el alumno podrá rendir un examen escrito teórico-práctico similar al del alumno regular aunque con un ejercicio/problema de práctica adicional.

Condiciones de Regularidad :

Para adquirir la regularidad en la asignatura, los alumnos deben asistir al 75% de las clases prácticas y de laboratorio. Deben entregar en tiempo y forma los Trabajos Prácticos.

El régimen de promoción para esta asignatura exige cumplir las condiciones de regularidad y aprobar con 70% o más las defensas orales de los TPs

Bibliografía Principal:

Data warehouse Systems.Design and Implementation.2nd Edition.Alejandro Vaisman,EstebanZimányi.

ISBN10 3662651661

ISBN13 9783662651667.Editorial Springer-Verlag Berlin and Heidelberg GmbH & Co. KG. 2022

Bibliografía Complementaria:

ELMASRIR.yNAVATHES.,“SistemasdeBasesdeDatos”,Addison-WesleyIberoamericana,3raó5ta Edición,2002
(7 ejemplares disponibles en Biblioteca FI-UNER)

KORTHH.ySILBERSCHATZA.,”FundamentosdeBasesdeDatos3/e”,McGraw-Hill,1998(2 ejemplares
disponibles en Biblioteca FI-UNER).

JOYANES AGUILAR, L.. Big data : análisis de grandes volúmenes de datos en organizaciones. México
D.F. : Alfaomega - 2013 (1 ejemplar disponible en Biblioteca FI-UNER)