

**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.

**Competencias Genéricas:**

No aplican

**Competencias Específicas:**

No aplican

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

No aplican

---

**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

**Insercion de la Asignatura en el plan de Estudios:**

La asignatura aprovecha los fundamentos incorporados en materias previas el futuro profesional, porque vincula los conceptos y fundamentos vistos en “Bases de Datos” y en “Bases de datos multidimensionales”. En dichas materias el futuro profesional interiorizó los fundamentos de los procesos transaccionales y de los procesos de soporte a la toma de decisiones en contextos de producción. Del análisis de las funciones y problemas que debe afrontar como profesional informático en cualquier tipo de organización, se desprende que hará uso de los conceptos aprendidos en las materias antes mencionadas.

En el cursado de la asignatura el alumno emplea conceptos matemáticos adquiridos en “Algebra y Cálculo” y “Probabilidad y estadística”. La asignatura “Bases de Datos” provee los conocimientos para comprender como se estructuran los datos en las bases de datos y ayudan a resolver problemas de modelado e implementación de programas y datos, y luego ampliados en “Bases de Datos Multidimensionales”. Los conceptos de Técnicas y Herramientas para Datos Masivos no solo son importantes en la vida profesional del egresado sino también en el desarrollo de la autonomía del alumno al integrar conceptos previos con nuevos modelos y tecnologías (por ej: Bases de Datos NoSQL y Cloud).

**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).

**Metodología Didáctica:**

Clases basadas en problemas

Teniendo en cuenta que en esta etapa final de la carrera los alumno/as cuentan con conocimiento profundo de los procesos transaccionales y los de toma de decisiones el enfoque de la materia está basado en el abordaje de problemas a partir de diferentes estrategias de implementación

Las primeras clases, tienen como objetivo principal llegar desde un problema general a la formalización de un problema de implementación o de investigación según el tema. Complementariamente se realiza una 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 como la pizarra para los desarrollos y el proyector 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. Durante la clase cada uno de los grupos dispondrá de tiempo para realizar las actividades de investigación y/o implementación y en las cuales se hace un seguimiento continuo. Como complemento de manera asincrónica se proponen lecturas complementarias y ejercicios conceptuales para profundizar en algunos temas.

**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 clase. Se promueve el trabajo en grupo, motivando el razonamiento para aplicar los conceptos discutidos en las clases y un vínculo con problemas que este resolviendo el grupo.

En lo que refiere a las etapas de formación experimental y diseño estas consisten en el desarrollo e implementación en un Sistema de Gestión de Base de Datos No Relacional en relación a un problema concreto. A los alumnos se les provee los requerimientos del problema que deben desarrollar y ellos deberán proponer el diseño e implementación de la misma. En lo temas vinculados a almacenamiento e la nube se combinan actividades prácticas de implementación con trabajos de investigación. En todos los casos se promueve el trabajo grupal de los alumnos, dado que en conjunto deben resolver un problema de diseño.

**Listado de Actividades de Formación Práctica:**

TP 1 – Bases de Datos NoSQL

TP 2 – Procesamiento Distribuido y Almacenamiento en la nube

**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 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



**Cronograma de parciales durante el primer Cuatrimestre:**

---

**Cronograma de parciales durante el segundo Cuatrimestre:**

**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)

**Equipo de Cátedra:**

Prof. Titular Alejandro Hadad: Funciones: planificación y coordinación del dictado.

Prof. Adjunto Walter Elías: Preparación y dictado de clases basadas en problemas.

JTP Maximiliano Fernandez: Colaboración en temas puntuales.

**Actividades de Investigación Gestión y Extensión:**

VER documento de planificación de la asignatura Bases de Datos de la Lic. en Bioinformática

---

**Requisitos de admisión para alumnos oyentes:**

Ser alumno de la Facultad de Ingeniería.

---

**Infraestructura, equipamiento y recursos necesarios:**

Laboratorio con computadoras personales.

**Otros:**