

Formulario de aprobación de curso de posgrado/educación permanente

Asignatura: Modelado y Procesamiento de Grandes Volúmenes de Datos

(Si el nombre contiene siglas deberán ser aclaradas)

Modalidad:

(posgrado, educación permanente o ambas)

Posgrado

Educación permanente

Profesor de la asignatura ¹: Dr. Adriana Marotta, grado 4, Instituto de Computación
(título, nombre, grado o cargo, instituto o institución)

Profesor Responsable Local ¹:

(título, nombre, grado, instituto)

Otros docentes de la Facultad: Ing. Sebastián García Parra, grado 2, Instituto de Computación
(título, nombre, grado, instituto)

Docentes fuera de Facultad: Ing. Daniel Callejas, Ingeniero de Datos ,IDATHA, Uruguay
(título, nombre, cargo, institución, país)

¹ Agregar CV si el curso se dicta por primera vez.

(Si el profesor de la asignatura no es docente de la Facultad se deberá designar un responsable local)

[Si es curso de posgrado]

Programa(s) de posgrado: Especialización en Sistemas de Información y tecnologías de gestión de datos, Maestría en Sistemas de Información y tecnologías de gestión de datos

Instituto o unidad: Instituto de Computación

Departamento o área:

Horas Presenciales: 50

(se deberán discriminar las horas en el ítem Metodología de enseñanza)

Nº de Créditos: 6

[Exclusivamente para curso de posgrado]

(de acuerdo a la definición de la UdelaR, un crédito equivale a 15 horas de dedicación del estudiante según se detalla en el ítem Metodología de enseñanza)

Público objetivo: Estudiantes de posgrado y/o profesionales interesados en el área de modelado y procesamiento de datos.

Cupos: cupo mínimo: 6, cupo máximo: 30

(si corresponde, se indicará el número de plazas, mínimo y máximo y los criterios de selección. Asimismo, se adjuntará en nota aparte los fundamentos de los cupos propuestos. Si no existe indicación particular para el cupo máximo, el criterio general será el orden de inscripción, hasta completar el cupo asignado)

Objetivos: Durante más de 30 años, los sistemas de bases de datos basados en el modelo relacional dominaron el almacenamiento y la gestión de datos. En la última década, y debido a múltiples razones, esta hegemonía se ha visto desafiada. En particular, las aplicaciones sobre la Internet, con billones de usuarios distribuidos mundialmente generando datos constantemente, configuran un nuevo escenario y plantean problemas que los sistemas de bases de datos relacionales no logran resolver adecuadamente. Como consecuencia, en la actualidad coexisten nuevos modelos y estrategias de gestión de datos. A la hora de diseñar una solución informática resulta imprescindible contar con información y hacer un análisis crítico de cada una de estas estrategias, que permita tomar decisiones fundadas más allá de las tendencias.

El objetivo de esta asignatura es brindar una visión general sobre el modelo relacional y los nuevos modelos de gestión de datos, analizando las fortalezas y debilidades de cada uno de ellos. Esto busca enriquecer la mirada y el espíritu crítico sobre las diferentes estrategias disponibles.

Durante el curso se hará énfasis en los modelos de datos subyacentes y los lenguajes de consulta asociados. Al final del curso el estudiante contará con elementos para decidir cuál es el modelo y la estrategia de gestión de datos que mejor se ajusta a su problema.

Conocimientos previos exigidos:

Conocimientos previos recomendados: Conocimientos de programación.

Metodología de enseñanza:

(comprende una descripción de la metodología de enseñanza y de las horas dedicadas por el estudiante a la asignatura, distribuidas en horas presenciales -de clase práctica, teórico, laboratorio, consulta, etc.- y no presenciales de trabajo personal del estudiante)

Descripción de la metodología: El curso se desarrollará en base a exposiciones teóricas alternadas con ejercicios prácticos de aplicación de los conceptos. Cada unidad temática finalizará con un ejercicio grupal donde se aplicarán, de forma práctica, los conceptos teóricos introducidos. En estos ejercicios se trabajará con las herramientas y lenguajes de consultas propios de cada uno de los motores de bases de datos abordados.

Algunos ejercicios serán entregables y formarán parte de la evaluación del curso.

[Obligatorio]

Detalle de horas:

- Horas de clase (teórico): 20
- Horas de clase (práctico): 10
- Horas de clase (laboratorio): 13
- Horas de consulta: 5
- Horas de evaluación: 2
 - Subtotal de horas presenciales: 50
- Horas de estudio: 20
- Horas de resolución de ejercicios/prácticos: 20
- Horas proyecto final/monografía:
 - Total de horas de dedicación del estudiante: 90

Forma de evaluación: Todos los estudiantes serán evaluados a través de entregas de ejercicios prácticos en grupos y una prueba final individual.

[Indique la forma de evaluación para estudiantes de posgrado, si corresponde]

[Indique la forma de evaluación para estudiantes de educación permanente, si corresponde]

Temario:

- Introducción
 - Perspectiva histórica y evolución de los sistemas de bases de datos.
 - Modelo conceptual
- El modelo relacional
 - Introducción al modelo relacional
 - Restricciones de integridad
 - Propiedades ACID
 - SQL
- Bases de datos de documentos
 - Introducción al modelado basados en documentos
 - Aspectos técnicos de MongoDB
 - Consultas en MongoDB
- Bases de datos de grafos
 - ¿Qué es un grafo?
 - Introducción a los modelos de grafos
 - Aspectos técnicos de Neo4j
 - Neo4j Cypher
- BigData y el ecosistema Apache Hadoop
 - Definición de BigData
 - Principales componentes del ecosistema Apache Hadoop (Spark, Cassandra, Hive, Pig)
 - Aspectos técnicos de Hadoop
 - Hadoop MapReduce

Bibliografía:

(título del libro-nombre del autor-editorial-ISBN-fecha de edición)

Fundamentals of Database Systems (6th Edition) Elmasri, R., Navathe, S., Pearson, ISBN 978-0-1360-8620-8, 2010

Next Generation Databases: NoSQL, NewSQL, and Big Data. Guy Harrison, Apress, ISBN 978-1-4842-1330-8, 2015. (disponible en el Portal Timbó)

Concise Guide to Databases: A Practical Introduction. Peter Lake and Paul Crowther, In Undergraduate Topics in Computer Science, Springer-Verlag London, ISBN 978-1-4471-5600-0, 2013. (disponible en el Portal Timbó)

Manuales de las herramientas que se usarán en el laboratorio.
