



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE INGENIERÍA



TEMAS SELECTOS DE
INGENIERÍA EN COMPUTACIÓN I
(Ingeniería de datos)

2956

10

6

Asignatura

Clave

Semestre

Creditos

INGENIERÍA ELÉCTRICA

COMPUTACIÓN

INGENIERÍA EN COMPUTACIÓN

División

Departamento

Carrera(s) en que se imparte

Asignatura:

Obligatoria

Horas/semana:

Teóricas

3.0

Prácticas

0.0

Optativa

X

Total

3.0

Horas/semestre:

Teóricas

48.0

Prácticas

0.0

Total

48.0

Modalidad:

Curso teórico

Seriación obligatoria antecedente:

Ninguna

Seriación obligatoria consecuente:

Ninguna

Descripción del curso:

El curso abordará todas las prácticas alrededor de la ingeniería de datos, es decir, el procesamiento e integración de datos en un mismo repositorio, para su análisis, a partir de datos provenientes de diversas fuentes - por ejemplo, bases de datos transaccionales, archivos de texto, NoSQL, o datos no estructurados - y así apoyar la toma de decisiones basadas en datos dentro de una empresa u organización (pública o privada), con fines de lucro o de investigación; o bien, la creación de modelos de aprendizaje automático que proyectan posibles sucesos futuros. Para ello se analizarán conceptos clave como datos masivos, entrenamiento de modelos, ciencia de datos, etc.

Objetivo(s) del curso:

El alumno profundizará sus conocimientos en todo lo relacionado con sistemas analíticos, volviéndose un ingeniero de datos altamente calificado al manipular toda clase de información. Será capaz de crear canalizaciones de datos a cualquier escala y afrontar proyectos con diferente infraestructura en instalaciones propias y en la nube. Aprenderá a analizar el estado del arte para proponer arquitecturas de datos robustas, utilizando tecnologías de frontera, adecuadas a cada caso de uso y resolver problemas de ingeniería de datos de cualquier industria: farmacéutica, automotriz, banca, seguros, así como de investigación científica; una actividad altamente demandada y valorada en el mercado laboral actual.

Temario

NÚM.	NOMBRE	HORAS
1	¿Que es la ingeniería de datos?	3.0
2	Almacenamiento de datos	6.0
3	Integración de datos	7.5
4	Procesamiento de datos masivos	9.0
5	Analítica de datos tradicional y avanzada	6.0
6	Cómputo en la nube	9.0
7	Estado del arte: Arquitectura y otras prácticas de datos	7.5

Actividades prácticas	0.0
Total	48.0

TEMARIO DESGLOSADO

1. ¿Qué es la ingeniería de datos?

Objetivo: El alumno comprenderá el concepto de ingeniería de datos y las actividades relacionadas con esta práctica.

- 1.1. Concepto de ingeniería de datos
- 1.2. Ciclo de vida de los datos
- 1.3. Habilidades y conocimientos básicos de un ingeniero de datos
- 1.4. Lenguajes de programación, APIs, marcos de trabajo y metodologías de desarrollo

2. Almacenamiento de datos

Objetivo: El alumno profundizará en temas de bases de datos: relacionales, analíticas, distribuidas, etc., así como los modelos de datos lógicos y organización interna que las rigen.

- 2.1. Tipos de bases de datos y repositorios
- 2.2. Almacenes de datos: esquema estrella vs crossed tables
- 2.3. Lagos de datos: sistemas de archivos y almacenes de archivos blob
- 2.4. Formatos de archivos: separados por caracter, ancho fijo, Avro, Parquet, ORC

3. Integración de datos

Objetivo: El alumno aprenderá que es una canalización de datos, que herramientas necesita y su importancia dentro de una empresa o en la investigación científica.

- 3.1. Procesos de extracción, transformación y carga con herramientas tradicionales
- 3.2. Calidad y limpieza de datos
- 3.3. SQL para procesamiento de datos
- 3.4. Orquestación de procesos por lotes, tiempo real y micro lotes

4. Procesamiento de datos masivos

Objetivo: El alumno descubrirá la manera más eficiente de procesamiento de datos, utilizando los entornos de desarrollo más avanzados disponibles que permiten integrar volúmenes de datos extremadamente grandes.

- 4.1. Datos masivos, cómputo distribuido, arquitecturas propietarias, computadoras genéricas
- 4.2. Proyecto Apache Hadoop
- 4.3. Apache Spark: Arquitectura y estructuras de datos
- 4.4. DAG, ejecución perezosa, acciones y transformaciones
- 4.5. Contenedores y entornos DataOps

5. Analítica de datos tradicional y analítica avanzada

Objetivo: El alumno obtendrá nuevos conceptos ligados a la presentación de información ejecutiva que coadyuba al entendimiento del negocio dentro de una compañía para la toma de decisiones.

- 5.1. Inteligencia de negocios
- 5.2. KPI's, reportes y gráficos
- 5.3. Herramientas de reporte y creación de tableros de control: Gartner y Forrester Wave
- 5.4. Ingeniería de datos aplicada al aprendizaje máquina y ciencia de datos

6. Cómputo en la nube

Objetivo: El alumno sabrá distinguir las ventajas y desventajas de los entornos de desarrollo en la nube frente a los entornos en infraestructura local de una empresa.

- 6.1. ¿Qué es la nube? Nubes privadas y publicas

- 6.2. Ventajas y desventajas del uso de la nube
- 6.3. IaaS, PaaS y SaaS
- 6.4. Infraestructura automatizada
- 6.5. Operación en entornos de nube
- 6.6. Principales servicios de ingeniería de datos en la nube

7. Estado del arte: Arquitectura y otras prácticas de datos

Objetivo: Con el conocimiento adquirido en los temas anteriores, el alumno aprenderá a diseñar sus propias soluciones de arquitectura de forma integral, aprendiendo a colaborar con profesionales de otras disciplinas relacionadas con la ingeniería de datos.

- 7.1. Seguridad de datos, enmascaramiento y encriptación de datos
- 7.2. Gobierno de datos, catálogos de datos, mallas y tejidos de datos
- 7.3. Arquitectura de datos, diseño y planeación, estimación de tiempos, recursos y costos
- 7.4. Implementación y gestión de proyectos
- 7.5. Certificaciones

....
u

Bibliografía básica	Temas para los que se recomienda
Laia Subirats M., Diego O. Pérez T., y Mireia Calvo G. 2019. <i>Introducción al ciclo de vida de los datos</i> (1ra. ed.). FUOC, España.	1
William H. Inmon. 2005. <i>Building the Data Warehouse</i> (4th. ed.). Wiley Publishing, USA.	1, 2
Denny Lee, Tristen Wentling, Scott Haines and Prashanth Babu. 2024. <i>Delta Lake: The Definitive Guide: Modern Data Lakehouse Architectures with Data Lakes</i> (1st. ed.). O’Reilly Media, Inc., USA.	2, 7
Ralph Kimball and Joe Caserta. 2004. <i>The Data Warehouse ETL Toolkit Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data</i> (1st. ed.). Wiley Publishing, USA.	3
Luis J. Aguilar. 2013. <i>Big Data: Análisis de grandes volúmenes de datos en organizaciones</i> (1st. ed.). Alfaomega, México.	4
Jules S. Damji, Brooke Wenig, Tathagata Das, and Denny Lee. 2020. <i>Learning Spark</i> . O’Reilly Media, Inc., USA.	4
Luis J. Aguilar. 2019. <i>Inteligencia de negocios y analítica de datos</i> (1st. ed.). Alfaomega, México.	4, 5
Valliappa Lakshmanan. 2018. <i>Data science on the Google Cloud Platform</i> (1st. ed.). O’Reilly Media, Inc., USA.	5, 6
Walter E. Calcagno L. 2024. <i>Arquitectura e ingeniería de datos</i> (1ra. ed.). Anaya, España.	1, 2, 3, 7

Bibliografía complementaria	Temas para los que se recomienda
Dan Sullivan. 2020. <i>Official Google Cloud Certified Professional Data Engineer Study Guide</i> . John Wiley & Sons, Inc., USA.	7

Alvaro Alea F. <i>Manual de Linux</i> . Universidad de Santiago de Compostela, España.	1, 3, 4, 6
Bill Chambers and Matei Zaharia. 2018. <i>Spark: The Definitive Guide</i> (1st. ed.). O'Reilly Media, Inc., USA.	4
Adoración De Miguel M., Mario Piattini, y Marcos Esperanza. 2000. Diseño de bases de datos relacionales. Alfaomega, México.	1
Paul Deitel and Harvey Deitel. 2020. Intro to Python for Computer Science and Data Science (1st. ed.). Pearson Education, Inc., USA.	4, 5

Materiales del curso	
Presentaciones, lecturas, artículos científicos, documentación oficial, lenguajes de programación Python y Scala usando IntelliJ Idea, PyCharm, Visual Studio Code, POSIX, y software libre y de código abierto como distribuciones Linux, máquinas virtuales y contenedores.	
Marcos de trabajo para manejo de datos masivos del proyecto Apache Hadoop, distribución Cloudera, Pentaho, IPC y Tableau	
Servicios de nubes públicas: Google Cloud Platform, Amazon AWS y Microsoft Azure	

Evaluación	
Proyecto	40%
Exámenes	40%
Tareas	20%

Información del profesor

Nombre completo:

Kevin Axel Romero Hernández

Correo electrónico institucional:

Horario de la clase:

Semblanza corta del profesor.

Arquitecto e ingeniero de datos con experiencia en el sector bancario, seguros, retail y organismos autónomos de gobierno de amplio prestigio. Louis Vuitton, Bancomer, UPS, Banorte, GNP, Telcel, INE y UNAM, participando en proyectos multinacionales en México, Estados Unidos, España, Perú y Colombia.