

IDENTIFICACIÓN DE LA UNIDAD DE APRENDIZAJE

Unidad académica: Instituto de Investigación en Ciencias Básicas y Aplicadas							
Plan de estudios: Licenciatura en Inteligencia Artificial							
Unidad de aprendizaje: MANEJO DE DATOS MASIVOS				Ciclo de formación: Profesional Eje general de formación: Teórico-Técnica Área de conocimiento: Fundamentos de la computación aplicada Semestre: 4º			
Elaborada por: Dr. Mauricio Rosales Rivera				Fecha de elaboración: Abril, 2021			
Clave:	Horas teóricas :	Horas prácticas :	Horas totales :	Créditos :	Tipo de unidad de aprendizaje :	Carácter de la unidad de aprendizaje :	Modalidad:
MD30CP03000 6	03	00	03	06	Obligatoria	Teórica	Escolarizada
Plan (es) de estudio en los que se imparte: A partir de todos los programas impartidos por el Instituto de Investigación en Ciencias Básicas y Aplicadas							

ESTRUCTURA DE LA UNIDAD DE APRENDIZAJE

Presentación: En esta unidad de aprendizaje se estudia las diferencias entre volumen y variedad de datos, así como distinguir entre las bases de datos operativas de las analíticas y comprender cómo se aplican en big data.
Propósito: Conozca y comprenda los fundamentos del diseño de tablas y bases de datos, al finalizar la unidad de aprendizaje, mediante el uso de SQL, para trabajar con sistemas de big data, para almacenamiento y análisis ética y responsabilidad social.
Competencias que contribuyen al perfil de egreso



Competencias genéricas:

- CG10. Habilidades en el uso de la tecnología de la información y de la comunicación.
- CG11. Habilidades para buscar, procesar y analizar información.
- CG16. Capacidad para identificar, plantear y resolver problemas.
- CG23. Capacidad de trabajo en equipo.

Competencias específicas:

- CE9. Aplica conocimientos, habilidades y actitudes adquiridas mediante la realización de proyectos multidisciplinarios para promover aprendizajes significativos de manera constructiva y participativa
- CE10. Opera equipo de cómputo y software de manera efectiva mediante la práctica constante para el correcto desarrollo de proyectos en compromiso con su medio sociocultural.

CONTENIDOS

Bloques	Temas
1. Fundamentos de sistemas de datos intensivos	1.5. ¿Qué son los sistemas de datos intensivos? 1.6. Estructuras de sistemas BD 1.7. Diseño de sistema tipo R: RBDMS 1.8. Manejo de errores de almacenamiento 1.9. Componentes: Almacenamiento (B-trees, row oriented; column oriented); Locking (Fine-grained; Coarse-grained); Recovery (Log data writes, Log queries) 1.10. Arquitecturas alternativas: Processing engines (Map Reduce, Apache Spark, Apache Storm, Lucene, TensorFlow); Open storage (Parquet, JSON, Delta Lake, Hive); Large-scale (GFS, Hadoop HDFS, Amazon S3) 1.11. Hardware de almacenamiento: Costos, Tendencias, Desempeño
2. Formatos de almacenamiento de datos	2.1. Jerarquías de almacenamiento 2.2. Combinación de dispositivos de almacenamiento 2.3. Diseño de formatos de almacenamiento 2.4. Acceso eficiente a los datos: Locality y Searchability (Rows vs Columns stores) 2.5. Mejora de búsquedas: Particiones (add, remove y list de archivos en directorios) 2.6. Ordenamiento Z 2.7. C-Store (Null, Dictionary, Run-length, Bit-vector, Lempel-Ziv) 2.8. Tradeoff: Sparse vs Dense 2.9. Manejo de duplicados (Handling Duplicate Keys)



3. Computación paralela	3.1. ¿Por qué paralelismo? 3.2. Abstracciones y conceptos básicos 3.3. Pensamiento paralelo (Data-Parallel Thinking) 3.4. Computación distribuida usando Spark 3.5. Consistencia y coherencia: Cache, Memoria 3.6. Memoria de transacción 3.7. Bases de datos distribuidas
4. Gestión de Big Data en clústeres y almacenamiento en la nube	4.1. Orientación y administración a los datos en clústeres y almacenamiento en la nube 4.2. Seguridad y privacidad de datos 4.3. Optimización de queries 4.4. Recuperación de fallas y Concurrencia
5. Sistemas de gestión y modelado de Big Data	5.1. Introducción al modelado y gestión de Big Data 5.2. Modelado y recuperación Big Data 5.3. Integración y procesamiento de Big Data 5.4. Ejemplos prácticos

ESTRATEGIAS DE ENSEÑANZA - APRENDIZAJE

Estrategias de aprendizaje sugeridas (Marque X)			
Aprendizaje basado en problemas	(X)	Nemotecnia	()
Estudios de caso	()	Análisis de textos	()
Trabajo colaborativo	(X)	Seminarios	()
Plenaria	()	Debate	(X)
Ensayo	()	Taller	()
Mapas conceptuales	()	Ponencia científica	()
Diseño de proyectos	(X)	Elaboración de síntesis	()
Mapa mental	()	Monografía	()
Práctica reflexiva	()	Reporte de lectura	()
Trípticos	()	Exposición oral	(X)
Otros			
Estrategias de enseñanza sugeridas (Marque X)			



Presentación oral (conferencia o exposición) por parte del docente	(X)	Experimentación (prácticas)	()
Debate o Panel	()	Trabajos de investigación documental	()
Lectura comentada	(X)	Anteproyectos de investigación	()
Seminario de investigación	()	Discusión guiada	(X)
Estudio de Casos	(X)	Organizadores gráficos (Diagramas, etc.)	()
Foro	()	Actividad focal	()
Demostraciones	()	Analogías	()
Ejercicios prácticos (series de problemas)	(X)	Método de proyectos	()
Interacción la realidad (a través de videos, fotografías, dibujos y software especialmente diseñado).	()	Actividades generadoras de información previa	()
Organizadores previos	()	Exploración de la web	()
Archivo	()	Portafolio de evidencias	()
Ambiente virtual (foros, chat, correos, ligas a otros sitios web, otros)	()	Enunciado de objetivo o intenciones	()

CRITERIOS DE EVALUACIÓN

Criterios	Porcentaje
<ul style="list-style-type: none"> • Exámenes parciales • Realización de práctica • Búsqueda de información • Participación en clase • Reportes 	<p>30%</p> <p>30%</p> <p>20%</p> <p>10%</p> <p>10%</p>
Total	100 %



PERFIL DEL PROFESORADO

Licenciatura, Maestría o Doctorado en ciencias computacionales o ingeniería en áreas afines a las ciencias computacionales, con experiencia docente en el área.

REFERENCIAS

Básicas:

- Wu, Y., Hu, F., Min, G., & Zomaya, A. Y. (Eds.). (2017). *Big data and computational intelligence in networking*. CRC Press.
- Wu, D., Sakr, S., & Zhu, L. (2017). *Big data storage and data models*. In *Handbook of big data technologies* (pp. 3-29). Springer, Cham.
- Malik, U., Goldwasser, M., & Johnston, B. (2019). *SQL for Data Analytics: Perform fast and efficient data analysis with the power of SQL*. Packt Publishing Ltd.
- Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). *Big Data and cloud computing: innovation opportunities and challenges*. *International Journal of Digital Earth*, 10(1), 13-53.

Complementarias:

- Programa especializado: Modern Big Data Analysis with SQL:
<https://www.coursera.org/specializations/cloudera-big-data-analysis-sql>
- Programa especializado: Principles of Data-Intensive Systems:
<http://web.stanford.edu/class/cs245/>

