

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: LABORATORIO DE MANEJO DE DATOS MASIVOS				Ciclo de formación: Profesional Eje general de formación: En contexto 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:
LM32CP00020 2	00	02	02	02	Obligatoria	Práctica	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

<p>Presentación:</p> <p>En esta unidad de aprendizaje se realizan prácticas para el manejo y entendimiento de volumen y variedad de datos, así como practicar en la distinción entre las bases de datos operativas de las analíticas y comprender cómo se aplican en big data.</p>
<p>Propósito:</p> <p>Conozca y comprenda los fundamentos del diseño de tablas y bases de datos, mediante el desarrollo de sistemas usando SQL, para trabajar con sistemas de big data para almacenamiento y análisis con ética y de forma creativa.</p>



Competencias que contribuyen al perfil de egreso	
Competencias genéricas:	
<ul style="list-style-type: none"> • CG15 Capacidad para formular y gestionar proyectos. • CG18 Capacidad para actuar en nuevas situaciones. • CG20 Capacidad de expresión y comunicación. 	
Competencias específicas:	
<ul style="list-style-type: none"> • CE4 Analiza soluciones computacionales mediante la aplicación de fundamentos teóricos del diseño de algoritmos y estructuras de datos adecuadas para resolver problemas con pensamiento crítico. • CE9. Aplica conocimientos, habilidades y actitudes adquiridas mediante la realización de proyectos multidisciplinarios para promover aprendizajes significativos de manera constructiva y participativa. 	

CONTENIDOS

Bloques	Temas
1. Fundamentos de sistemas de datos intensivos	1.1. Diseño de estructuras de sistemas BD 1.2. Diseño de sistema tipo R: RBDMS 1.3. Ejemplos prácticos de manejo de errores de almacenamiento 1.4. Ejercicios con componentes de Almacenamiento (B-trees, row oriented; column oriented); Locking (Fine-grained; Coarse-grained); Recovery (Log data writes, Log queries)
2. Formatos de almacenamiento de datos	2.1. Ejemplos para entender jerarquías de almacenamiento 2.2. Ejemplos de combinación de dispositivos de almacenamiento 2.3. Prácticas para acceder a los datos: Locality y Searchability (Rows vs Columns stores) 2.4. Prácticas para mejora de búsquedas: Particiones (add, remove y list de archivos en directorios) 2.5. Prácticas de tipos de ordenamiento Z; C-Store (Null, Dictionary, Run-length, Bit-vector, Lempel-Ziv); Sparse vs Dense 2.6. Práctica para manejar duplicados (Handling Duplicate Keys)
3. Computación paralela	3.1. Ejercicios de procesamiento en paralelo (Data-Parallel Thinking) 3.2. Computación distribuida usando Spark 3.3. Prácticas y entendimiento de Spark
4. Gestión de Big Data en clústeres y	4.1. Ejemplos prácticos de administración de en clústeres y almacenamiento en la nube 4.2. Ejemplos de seguridad y privacidad de datos; recuperación de fallas y Concurrencia



almacenamiento en la nube	
5. Sistemas de gestión y modelado de Big Data	5.1. Ejemplos de modelado y gestión de Big Data

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)	(X)
Debate o Panel	()	Trabajos de investigación documental	()
Lectura comentada	(X)	Anteproyectos de investigación	()



Seminario de investigación	()	Discusión guiada	(X)
Estudio de Casos	()	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	(X)

CRITERIOS DE EVALUACIÓN

Crterios	Porcentaje
<ul style="list-style-type: none"> • Exámenes parciales prácticos • Realización de prácticas • 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 el área afín a la disciplina de la unidad de aprendizaje (Ciencias Computacionales)



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/>

