



Unidad de Enseñanza Aprendizaje
E Impartición de Cátedra

**Temas selectos de Minería
de Datos en Big Data**

Reforma Curricular UAT 2023

Dependencia Académica: Facultad de Ingeniería y Ciencias

Programa Académico: Maestría en Ciencias e Ingeniería de Datos

Tipo: Único

Unidad de Enseñanza Aprendizaje e Impartición de Cátedra

PERFIL DOCENTE DE LA UNIDAD DE ENSEÑANZA-APRENDIZAJE		
FORMACIÓN PROFESIONAL	COMPETENCIAS PEDAGÓGICAS	EXPERIENCIA LABORAL
-Maestría en el área disciplinar, en computación o en ciencia de datos -Doctorado en el área disciplinar, en computación o en ciencia de datos	Diseña, organiza y propone actividades didácticas y problemas para su solución por parte de los educandos. Conoce y aplica distintas estrategias y formas de evaluación sobre el proceso educativo. Establece un clima de relación en el grupo que favorece actitudes de confianza, autoestima, respeto, disciplina, creatividad por el estudio. Conoce los materiales de enseñanza y los recursos didácticos disponibles.	Mínimo 1 año de experiencia en la impartición de cátedra a nivel posgrado en el desarrollo de habilidades para aprender

UNIDAD DE ENSEÑANZA-APRENDIZAJE						
Datos Generales						
ASIGNATURA	Tópicos selectos de Minería de Datos en Big Data				TIPO DE ASIGNATURA	Optativa
CLAVE	HTC	HTI	TH	TC	ASIGNATURA ANTECEDENTE	PERIODO ESCOLAR
MCID22	4(56)	3(42)	7(98)	6	-	3ero
CONTRIBUCIÓN DE LA ASIGNATURA EN EL PERFIL DE EGRESO INSTITUCIONAL		Formar profesionistas con capacidad de aprender, adaptarse y solucionar problemas inherentes al ejercicio de su profesión, con ética y actitud positiva ante el trabajo.				
CONTRIBUCIÓN DE LA ASIGNATURA EN EL PERFIL DE EGRESO DEL PROGRAMA ACADÉMICO		Conocimientos básicos generales y específicos de la profesión.				
OBJETIVO GENERAL DE LA UEA		El objetivo de la materia es preparar al alumno para producir rápidamente inteligencia procesable de fuentes de datos de gran tamaño. De tal manera, que se permita descubrir o extraer información relevante que permita al usuario final tomar una decisión.				

ATRIBUTOS DE LA UEA			
SABER	SABER HACER	SABER SER	SABER CONVIVIR
Conocimientos El alumno conocerá los conceptos básicos relacionados con la minería de datos en grandes cantidades de información.	Habilidades y aptitudes El alumno podrá implementar algoritmos de minería de datos en bases de datos de gran tamaño de tal manera que pueda extraer nueva información no trivial de información implícita, previamente desconocida y potencialmente útil.	Actitudes y valores Implementas soluciones utilizando técnicas de minería de datos en el contexto de grandes volúmenes de datos.	Interrelación pacífica, justa y respetuosa Comunicación efectiva Ética profesional

CRITERIOS DE EVALUACIÓN	VALOR EN PORCENTAJE
PARTICIPACIÓN INDIVIDUAL	5
PARTICIPACIÓN GRUPAL (EXPOSICIONES)	5
EVALUACIONES PARCIALES	50
ACTIVIDADES REALIZADAS	20
PRODUCTO INTEGRADOR	20
TOTAL	100

Nota: los criterios de evaluación que se enlistan son probables, el profesor decidirá cuáles incluir, debiendo considerar al menos 4 y los porcentajes para cada uno de ellos dando una suma total de 100%

PRODUCTO INTEGRADOR DE LA UEA		
DESCRIPCIÓN		Prototipo de software donde haga uso de técnicas de minería de datos utilizando conjunto de datos de gran tamaño.
NIVELES DE DESEMPEÑO		CRITERIOS DE DESEMPEÑO
10	Excelente	La aplicación cumple con el 100 % de los requerimientos solicitados y presentación del proyecto ante el grupo.
9	Bueno	La aplicación cumple con al menos el 90 % de los requerimientos solicitados y presentación del proyecto ante el grupo.
8	Regular	La aplicación cumple con al menos el 80 % de los requerimientos solicitados y presentación del proyecto ante el grupo.
7	Básico	La aplicación cumple con al menos el 70 % de los requerimientos solicitados y presentación del proyecto ante el grupo.
6	Elemental	No alcanza el mínimo aprobatorio.
NA	Aún no competente	No alcanza el mínimo aprobatorio.

UNIDAD O BLOQUE TEMÁTICO	
NÚMERO Y NOMBRE	1. Fundamentos de Big Data
OBJETIVO DE LA UNIDAD O BLOQUE TEMÁTICO	Introducir los conceptos básicos de minería de datos y big data
TIEMPO/DURACIÓN	Horas: 8 Semanas: 2
RECURSOS EDUCATIVOS:	Proyector, equipo de cómputo
DESGLOSE DE CONTENIDOS ESPECÍFICOS	
1.1. Fundamentos de Big Data 1.2. Datos estructurados y no estructurados 1.3. Bases de datos No SQL 1.4. Bases de datos relacionales	

UNIDAD O BLOQUE TEMÁTICO	
NÚMERO Y NOMBRE	2. Herramientas para Procesamiento de Big Data (Parte 1)
OBJETIVO DE LA UNIDAD O BLOQUE TEMÁTICO	Conocer e identificar las herramientas de software para el procesamiento de grandes volúmenes de datos.
TIEMPO/DURACIÓN	Horas: 20 Semanas: 5
RECURSOS EDUCATIVOS:	Proyector, equipo de cómputo
DESGLOSE DE CONTENIDOS ESPECÍFICOS	

- 2.1. Fundamentos de sistemas distribuidos
- 2.2. Teorema CAP
- 2.3. Clusters para programación masivamente paralela (MPP)
- 2.4. Virtualización de clusters y data centers
- 2.5. Arquitecturas cloud. El paradigma MapReduce

UNIDAD O BLOQUE TEMÁTICO	
NÚMERO Y NOMBRE	3. Herramientas para Procesamiento de Big Data (Parte 2)
OBJETIVO DE LA UNIDAD O BLOQUE TEMÁTICO	Conocer e identificar las herramientas de software para el procesamiento de grandes volúmenes de datos.
TIEMPO/DURACIÓN	Horas: 20 Semanas: 5
RECURSOS EDUCATIVOS:	Proyector, equipo de cómputo
DESGLOSE DE CONTENIDOS ESPECÍFICOS	
3.1. Hadoop File System 3.2. Evolución de Hadoop (YARN) 3.3. Arquitectura, componentes 3.4. Lenguajes de alto nivel (HiveQL, Pig Latin) 3.5. Análisis de datos usando Hadoop y Hive 3.6. Programación con Apache Spark	

IMPARTICIÓN DE CÁTEDRA

Semana No. 1-2

NÚMERO Y NOMBRE DE LA UNIDAD O BLOQUE TEMÁTICO:	1. Fundamentos de Big Data
OBJETIVO DE LA UNIDAD O BLOQUE TEMÁTICO:	Introducir los conceptos básicos de minería de datos y big data

CONTENIDO TEMÁTICO	MOMENTOS DE LA CÁTEDRA		
	Apertura	Desarrollo	Cierre
1.1 Fundamentos de Big Data 1.2 Datos estructurados y no estructurados 1.3 Bases de datos No SQL 1.4 Bases de datos relacionales	El profesor expone el objetivo de la Unidad y cómo impacta en el objetivo del curso. El profesor introduce el tema y valora los conocimientos previos del alumno mediante una interacción entre alumnos (lluvia de ideas).	El profesor explica los conceptos y elementos para implementar y gestionar grandes volúmenes de datos. El profesor ejemplifica diseños de bases de datos no relacionales.	Retroalimentación de los conceptos adquiridos por el alumno. El profesor ilustra escenarios para aclarar las dudas de los alumnos.

IMPARTICIÓN DE CÁTEDRA

Semana No. 3-4-5-6-7

NÚMERO Y NOMBRE DE LA UNIDAD O BLOQUE TEMÁTICO:	2. Herramientas para Procesamiento de Big Data (Parte 1)
OBJETIVO DE LA UNIDAD O BLOQUE TEMÁTICO:	Conocer e identificar las herramientas de software para el procesamiento de grandes volúmenes de datos.

CONTENIDO TEMÁTICO	MOMENTOS DE LA CÁTEDRA		
	Apertura	Desarrollo	Cierre
2.1. Fundamentos de sistemas distribuidos 2.2. Teorema CAP 2.3. Clusters para programación masivamente paralela (MPP) 2.4. Virtualización de clusters y data centers 2.5. Arquitecturas cloud. El paradigma MapReduce	El profesor expone el objetivo de la Unidad y cómo beneficia en el objetivo del curso. El profesor introduce el tema y valora los conocimientos previos del alumno mediante una interacción entre alumnos (lluvia de ideas).	El profesor explica conceptos sobre programación paralela y virtualización de clusters. El profesor ejemplifica soluciones basadas en map reduce.	Retroalimentación de los conceptos adquiridos por el alumno. El profesor ilustra escenarios para aclarar las dudas de los alumnos.

IMPARTICIÓN DE CÁTEDRA	
Semana No. _8-9-10-11-12	
NÚMERO Y NOMBRE DE LA UNIDAD O BLOQUE TEMÁTICO:	3. <i>Herramientas para Procesamiento de Big Data (Parte 2)</i>
OBJETIVO DE LA UNIDAD O BLOQUE TEMÁTICO:	<i>Conocer e identificar las herramientas de software para el procesamiento de grandes volúmenes de datos.</i>

CONTENIDO TEMÁTICO	MOMENTOS DE LA CÁTEDRA		
	Apertura	Desarrollo	Cierre
3.1. Hadoop File System 3.2. Evolución de Hadoop (YARN) 3.3. Arquitectura, componentes 3.4. Lenguajes de alto nivel (HiveQL, Pig Latin) 3.5. Análisis de datos usando Hadoop y Hive 3.6. Programación con Apache Spark	El profesor expone el objetivo de la Unidad y cómo beneficia en el objetivo del curso. El profesor introduce el tema y valora los conocimientos previos del alumno mediante una interacción entre alumnos (lluvia de ideas).	El profesor explica el funcionamiento y el diseño de soluciones basados en Hadoop. El profesor ejemplifica implementaciones utilizando varias arquitecturas.	Retroalimentación de los conceptos adquiridos por el alumno. El profesor ilustra escenarios para aclarar las dudas de los alumnos.

Nota: la planeación de "IMPARTICIÓN DE CÁTEDRA (Desglose del Bloque Temático)" se desarrolla por semana, por lo que en los periodos de clases de "primavera y otoño" el total serán 16 tablas.

REFERENCIAS BIBLIOGRÁFICA Y ELECTRÓNICA (APA)	
Básica	N. Marz and J. Warren, <i>Big Data: Principles and best practices of scalable realtime data systems</i> , 1era. Ed., Manning Publications, 2015, pp. 328.
Complementaria	J. Leskovec, A. Rajaraman, and J. D. Ullman, <i>Mining of Massive Datasets</i> , 2nd ed., Cambridge University Press, 2014, pp. 476.
	H. Karau and R.I Warren, <i>High Performance Spark: Best Practices for Scaling and Optimizing Apache Spark</i> , O'Reilly Media, 2017, pp. 358.
	M. Grover, T. Malaska, J. Seidman, and G.Shapira, <i>Hadoop Application Architectures: Designing Real-World Big Data Applications</i> O'Reilly Media, 2015, pp. 400.

Nota: las referencias bibliográficas (básicas y complementarias) se colocan solo al final de la UEA e IC. Se sugiere colocarlas de acuerdo con los criterios de la APA.

CRÉDITOS
Elaborado por: Edgar Tello Leal Fecha: 13/08/2017
Actualizado por: Edgar Tello Leal Fecha: 31/05/2023
Autorizado por: Fecha