Datos de identificación			
Nombre del EE: Procesamiento masivo de datos		Área Formativa: Vocacional	
Departamento que da el servicio: Departamento de Ingeniería Industrial			
Clave:	Modalidad: Presencial		Idiomas: español
Horas totales al semestre: 64	Valor en créditos: 4		Semestre en que se cursa: IX
Carácter: Obligatoria	EE Antecedente: 200 créditos		EE subsecuente: N/A
Opciones de promoción: Calificación		Mecanismos alternativos de promoción: Equivalencia	
Presentación			
<p>Este espacio educativo proporciona los conocimientos, habilidades y/o destrezas para trabajar con Big Data. Los alumnos analizarán conceptos fundamentales y desafíos del procesamiento masivo. Aplicarán técnicas avanzadas en bases de datos distribuidas y desarrollarán habilidades en programación paralela. Diseñarán soluciones para procesamiento en tiempo real. Evaluarán casos de uso en diversas industrias, considerando aspectos éticos. Finalmente, creará n un proyecto innovador, demostrando su dominio en el campo.</p>			
Desempeños			
Competencias genéricas que se ejercitan		Unidades de competencia profesionales	
<ul style="list-style-type: none"> Utiliza con eficiencia las tecnologías digitales para la comunicación y la gestión de información académica y profesional, en un entorno de trabajo colaborativo. 		5.5. Diseñar Interfaz de usuario Que cumpla con los estándares actuales de inclusión y usabilidad 8.4. Aplicar La prueba a futuro de la solución De modo que la propuesta sea innovadora y que no pierda actualidad para evitar su discontinuidad en los próximos años.	
Resultados de Aprendizaje			
<ul style="list-style-type: none"> Analizar los conceptos fundamentales y desafíos del Big Data, evaluando diferentes arquitecturas y tecnologías para el procesamiento masivo de datos. Implementar y optimizar bases de datos distribuidas, demostrando dominio en consultas SQL avanzadas, particionamiento de tablas y procedimientos almacenados para el procesamiento eficiente de grandes volúmenes de datos. Desarrollar soluciones de programación paralela y distribuida, aplicando el paradigma MapReduce y patrones de diseño específicos para el procesamiento masivo de datos. Utilizar Apache Spark para crear aplicaciones de procesamiento masivo, incluyendo análisis en tiempo real, aprendizaje automático distribuido y procesamiento de grafos, optimizando el rendimiento de estas aplicaciones. Evaluar e integrar soluciones de procesamiento masivo de datos en diversos contextos industriales, considerando implicaciones éticas y de privacidad, y adaptándose a las tendencias emergentes en el campo. 			
Orientación didáctica			
<p>El estudiante asistirá 64 horas semestrales, (4 horas semanales) a clases presenciales, dirigidas por el profesor, donde se desarrollará los conceptos y prácticas propias del procesamiento masivo de datos.</p>			
Actividades del estudiante		Actividades del profesor	
Horas/ semestre	Actividades	Horas/ semestre	Actividades
64	<ul style="list-style-type: none"> Asistencia y participación en clase. 	64	<ul style="list-style-type: none"> Impartir clases presenciales Definir las reglas de uso de algoritmos

			generativos de IA (GitHub, Copilot, ChatGPT, Gemini, etc.) en actividades de la materia que permitan utilizarlas como complemento para incrementar la productividad, más no como la solución total de los problemas a resolver o tareas por entregar.
<i>Evaluación del aprendizaje</i>			
<i>Criterios de cumplimiento</i>	<i>Evidencias de desempeño</i>	<i>Evidencias de conocimiento</i>	
1) Cumplir con la asistencia, puntualidad (Presencial o Virtual), 2) entrega de trabajos (investigación, tareas, exámenes) y/o practicas a tiempo y siguiendo las especificaciones descritas. 3) Cumplir con los criterios acordados de desarrollo con la vinculación.	1) Realización de exámenes (en línea y / o en papel). 2) Entrega de tareas y trabajos en plataforma electrónica. 3) Realización de exposiciones en inglés sobre el tema. 4) Desarrollo de un proyecto que integre todos los conceptos y tecnologías vistos.	1) Proyecto final que refleje los conocimientos aprendidos durante el curso. 2) Material y/o diapositivas de las exposiciones. 3) El alumno responderá con ideas, conocimiento y aprendizaje a preguntas del profesor. 4) Entrega de las actividades desarrolladas.	
<i>Técnicas e instrumentos de evaluación</i>	Rúbricas para los exámenes, para tareas, prácticas e investigación y el proyecto final.		
<i>Recursos para la formación</i>			
<i>Contenidos básicos</i>		<i>Materiales</i>	
<ul style="list-style-type: none"> • Introducción al Procesamiento Masivo de Datos • Bases de Datos Distribuidas y Greenplum • Programación Paralela y Distribuida • Apache Spark para Procesamiento Masivo de Datos • Integración y Casos de Uso • Proyecto Final y Tendencias Futuras 		<ul style="list-style-type: none"> • <i>Documentos electrónicos</i> • <i>Material audiovisual</i> • <i>Equipo de proyección</i> • <i>Plumones y pintarrón</i> • <i>Recursos en la nube Computadora</i> • <i>Conexión a Internet</i> 	
<i>Bibliografía</i>			
<ul style="list-style-type: none"> • Karau, H., & Warren, R. (2023). High performance Spark: Best practices for scaling and optimizing Apache Spark. O'Reilly Media. • Parsian, M. (2021). Data algorithms with Spark: Recipes and design patterns for scaling up using PySpark. O'Reilly Media. • Deléglise, D., Schabanel, M., & Zimmermann, J. (2022). Greenplum: The definitive guide. Apress. • Chambers, B., & Zaharia, M. (2023). Spark: The definitive guide: Big data processing made simple. O'Reilly Media. 			

Perfil deseable del profesor que lo conduce o lo coordina	
Grado académico: Licenciatura. De preferencia con estudios de Maestría o Doctorado.	Área de formación: Ciencias Computacionales, Desarrollo de Software, Ingeniería en Sistemas de Información
Experiencia docente: 1 año	Experiencia profesional en el campo: 1 año
Elaboró: Dr. Federico Miguel Cirett Galán	Fecha: 30 de octubre de 2024