

1. Datos Generales de la asignatura

Nombre de la asignatura:	BIG DATA
Clave de la asignatura:	CDD - 2303
SATCA¹:	2 - 3 - 5
Carrera:	Ingeniería en Tecnologías de Información y Comunicaciones

2. Presentación**Caracterización de la asignatura**

Con esta asignatura, el estudiante conocerá la nueva generación de tecnologías, arquitecturas y estrategias diseñadas para capturar y analizar grandes volúmenes de datos, que pueden provenir de diferentes fuentes (datos no estructurados), generarse a gran velocidad y se requiere de un alto nivel de veracidad de los mismos.

Como aportación al perfil de egresado se tiene lo siguiente:

- Identificar y resolver problemas relacionados con el paradigma Big Data
- Clasificar y almacenar los diferentes tipos de datos de este paradigma
- Poseer las habilidades metodológicas de investigación para sentar las bases de conocimiento para futuros Científicos de Datos
- Desarrollar, implementar y administrar software de sistemas o de aplicación que cumpla con los estándares de calidad con el fin de apoyar la productividad y competitividad de las organizaciones en el contexto del Big Data.
- Diseñar, desarrollar y administrar bases Big Data conforme a requerimientos definidos, normas organizacionales de manejo y seguridad de la información, utilizando tecnologías emergentes.

La asignatura presenta al alumno un panorama general y la visión actual del manejo de grandes volúmenes de datos (sobre todo datos no estructurados) lo que hoy en día genera la necesidad de contar con profesionistas capaces de analizar y procesar la información, de manera que de soluciones a la empresa u organización o bien pueda tener un mejor rendimiento.

Esta materia consiste en cuatro unidades, donde, la unidad uno presenta una introducción al Big Data, la unidad dos se refiere a la exploración y visualización de los datos mostrándose en diferentes arquitecturas o diseños, la unidad tres introduce al alumno al procesamiento de grande bases de datos mediante

¹ Sistema de Asignación y Transferencia de Créditos Académicos

técnicas de aprendizaje automático supervisado y no supervisado, finalmente la unidad cuatro presenta el análisis mediante apache hadoop y apache spark, aplicado a casos reales.

Las asignaturas con las que se relaciona esta materia son: Fundamentos de bases de datos, Taller de bases de datos, Administración de bases de datos, Gestión de proyectos de software.

Intención didáctica

En la unidad uno se investigarán los conceptos más relevantes sobre Big Data y se analizará el tipo de datos empleados, determinando si se trata de datos estructurados o no estructurados, así mismo se conocerá cuáles son las principales fuentes de estos datos y como se originan. Para ello se debe investigar los conceptos involucrados en cada uno de estos temas y comprender el panorama general de estos.

En la unidad dos se realiza la exploración y visualización de datos en Big Data, representados en distribuciones Hadoop como Cloudera, Hortonworks e IBM Biginsights, así mismo se conocen las arquitecturas de proyectos de Big Data y el almacenamiento y procesamiento de Big Data en la nube.

En la unidad tres se introduce al alumno en la búsqueda de información para procesamiento de Big Data mediante técnicas de aprendizaje automático, ya sea mediante clasificación supervisada o clasificación no supervisada, se propone aplicar ejemplos en diferentes algoritmos como Naive Bayes, el vecino más cercano (k-NN), perceptrón multicapa, entre otros.

En la unidad cuatro el alumno deberá ser capaz de procesar y analizar diferentes casos de uso de Big Data mediante herramientas de software como Apache Hadoop o Apache Spark.

Competencias Genéricas

Competencias instrumentales:

- Capacidad de análisis y síntesis.
- Conocimientos básicos de la carrera.
- Comunicación oral y escrita.
- Habilidades básicas de manejo de la computadora.
- Habilidad para buscar y analizar información proveniente de fuentes diversas.

Competencias interpersonales:

- Capacidad crítica y autocrítica.

- Trabajo en equipo
- Habilidades interpersonales

Competencias sistémicas:

- Habilidades de investigación.
- Capacidad de aprender.
- Capacidad de generar nuevas ideas (creatividad).
- Habilidad para trabajar en forma autónoma.
- Búsqueda del logro.

3. Participantes en el diseño y seguimiento curricular del programa

Lugar y fecha de elaboración o revisión	Participantes	Observaciones
Instituto Tecnológico de Toluca, septiembre 2022	M. en T. E. Luis Antonio Estrada Manuel Dr. Roberto Alejo Eleuterio Dr. Federico del Razo López Dra. Eréndira Rendón Lara	Diseño y elaboración de los programas de estudio de la especialidad para TIC

4. Competencias a desarrollar

Competencias específicas de la asignatura
<ul style="list-style-type: none"> • Conoce la nueva generación de tecnologías, arquitecturas y estrategias, con el fin de identificar, clasificar, almacenar, analizar y resolver problemas relacionados con el paradigma de Big Data.

5. Competencias previas

<p>Las adquiridas en las asignaturas de:</p> <ul style="list-style-type: none"> • Utiliza técnicas de modelado para la solución de problemas. • Comprende los conceptos de las bases de datos. • Construye modelos de bases de datos. • Domina el lenguaje de SQL. • Diseña, crea y administra esquemas de bases de datos relacionales bajo estándares. • Conoce distintas plataformas y SGBD. • Aplica las actividades que involucra cada una de las etapas del ciclo de vida de un

<p>proyecto de software.</p> <ul style="list-style-type: none"> • Analiza, diseña, desarrolla e implementa un proyecto de software. • Conoce e identifica modelos de arquitecturas de cómputo. • Identifica y aplica conceptos fundamentales de las telecomunicaciones, para aplicarlo a redes computacionales. • Selecciona, conoce y usa adecuadamente los diferentes sistemas operativos para lograr un uso más eficiente, así como diferenciar y aplicar las técnicas de manejo de recursos para el diseño, organización, utilización y optimización de los sistemas operativos. • Conoce y sabe usar técnicas y/o herramientas de administración de los sistemas operativos para la optimización de recursos existentes. • Aplica un lenguaje orientado a objetos para la solución de problemas. • Crea y manipula bases de datos utilizando distintos Gestores de Bases de Datos considerando elementos de integridad y seguridad para el tratamiento de la información en distintas plataformas.
--

6. Temario

No.	Temas	Subtemas
1	Introducción a Big Data	1.1 Definición de Big Data 1.2 Conceptos de almacenamiento distribuido 1.3 Conceptos de procesamiento distribuido 1.4 Tipos de datos (estructurados, no estructurados y semiestructurados) 1.5 Fuentes de datos Big Data
2	Exploración, procesamiento y visualización datos en Big Data	2.1 Arquitecturas de proyectos Big Data 2.2 Cloud Big Data 2.3 Distribuciones Hadoop <ul style="list-style-type: none"> 2.3.1 Cloudera 2.3.2 Hortonworks 2.3.3 IBM Biginsights
3	Machine Learning en Big Data	3.1 Clasificación supervisada 3.2 Clasificación no supervisada
4	Análisis de casos de uso de Big Data	4.1 Apache Hadoop 4.2 Apache Spark

7. Actividades de aprendizaje de los temas

Tema 1: Introducción a Big Data	
Competencias	Actividades de aprendizaje
<p>Específicas: Conoce los conceptos sobre Big Data, así como los tipos y fuentes de origen.</p> <p>Genéricas:</p> <ul style="list-style-type: none"> • Capacidad de análisis y síntesis • Capacidad de organizar y planificar • Comunicación oral y escrita en su propia lengua • Capacidad de lectura en idioma foráneo • Habilidades de gestión de información • Solución de problemas • Toma de decisiones • Capacidad crítica y autocrítica • Trabajo en equipo • Habilidades interpersonales • Capacidad de comunicarse con profesionales de otras áreas • Compromiso ético 	<ul style="list-style-type: none"> • Comprender el concepto de Big Data • Entender los conceptos de almacenamiento y procesamiento distribuido • Identificar los tipos de estructurados y no estructurados • Conocer las fuentes de Big Data
Tema 2: Exploración y visualización datos en Big Data	
Competencias	Actividades de aprendizaje
<p>Específicas: Conoce la arquitectura del Big Data así como sus representaciones en la nube y los datos en distribuciones hadoop.</p> <p>Genéricas:</p> <ul style="list-style-type: none"> • Capacidad de análisis y síntesis • Capacidad de organizar y planificar • Comunicación oral y escrita en su propia lengua • Capacidad de lectura en un idioma foráneo • Habilidades de gestión de información • Solución de problemas • Toma de decisiones • Capacidad crítica y autocrítica • Trabajo en equipo • Habilidades interpersonales • Capacidad de comunicarse con profesionales de otras áreas • Compromiso ético 	<ul style="list-style-type: none"> • Conocer la forma de representar los datos • Investigar cómo funciona el Big Data en la nube • Desarrolla representaciones de procesamiento de Big Data en Cloudera, Hortonworks e IBM Biginsights

Tema 3: Machine Learning en Big Data	
Competencias	Actividades de aprendizaje
<p>Específicas: Aplica algoritmos de aprendizaje automático a procesamiento Big Data.</p> <p>Genéricas:</p> <ul style="list-style-type: none"> • Capacidad de análisis y síntesis • Capacidad de organizar y planificar • Comunicación oral y escrita en su propia lengua • Capacidad de lectura en un idioma foráneo • Habilidades de gestión de información • Solución de problemas • Toma de decisiones • Capacidad crítica y autocrítica • Trabajo en equipo • Habilidades interpersonales • Capacidad de comunicarse con profesionales de otras áreas • Compromiso ético 	<ul style="list-style-type: none"> • Conocer los conceptos de aprendizaje automático, clasificación supervisada y clasificación no supervisada. • Aplica técnicas de clasificación supervisada a bases de datos Big Data. • Aplica técnicas de clasificación no supervisada a bases de datos Big Data.
Tema 4: Análisis de casos de uso Big Data	
Competencias	Actividades de aprendizaje
<p>Específicas: Analiza casos prácticos de procesamiento Big Data.</p> <p>Genéricas:</p> <ul style="list-style-type: none"> • Capacidad de análisis y síntesis • Capacidad de organizar y planificar • Comunicación oral y escrita en su propia lengua • Capacidad de lectura en un idioma foráneo • Habilidades de gestión de información • Solución de problemas • Toma de decisiones • Capacidad crítica y autocrítica • Trabajo en equipo • Habilidades interpersonales • Capacidad de comunicarse con profesionales de otras áreas • Compromiso ético 	<ul style="list-style-type: none"> • Desarrollar prácticas en apache hadoop • Implementar práctica con un ejemplo real en Apache Spark • Presentar el trabajo final de la práctica en Apache Spark

8. Práctica(s)

- Analizar y clasificar datos estructurados y no estructurados.
- Instalar algún software de procesamiento de bases de datos.
- Instalar cloudera, Hortonworks o IBM Biginsigts para procesar una base de datos.
- Realizar una práctica de procesamiento de Big Data aplicando un algoritmo de clasificación supervisada.
- Realizar una práctica de procesamiento de Big Data aplicando un algoritmo de clasificación no supervisada.
- Realizar una práctica de un caso real en Apache Hadoop.
- Realizar una práctica de un caso real en Apache Spark.

9. Proyecto de asignatura

El objetivo del proyecto que planteé el docente que imparta esta asignatura, es demostrar el desarrollo y alcance de la(s) competencia(s) de la asignatura, considerando las siguientes fases:

- **Fundamentación:** marco referencial (teórico, conceptual, contextual, legal) en el cual se fundamenta el proyecto de acuerdo con un diagnóstico realizado, mismo que permite a los estudiantes lograr la comprensión de la realidad o situación objeto de estudio para definir un proceso de intervención o hacer el diseño de un modelo.
- **Planeación:** con base en el diagnóstico en esta fase se realiza el diseño del proyecto por parte de los estudiantes con asesoría del docente; implica planificar un proceso: de intervención empresarial, social o comunitaria, el diseño de un modelo, entre otros, según el tipo de proyecto, las actividades a realizar los recursos requeridos y el cronograma de trabajo.
- **Ejecución:** consiste en el desarrollo de la planeación del proyecto realizada por parte de los estudiantes con asesoría del docente, es decir en la intervención (social, empresarial), o construcción del modelo propuesto según el tipo de proyecto, es la fase de mayor duración que implica el desempeño de las competencias genéricas y específicas a desarrollar.
- **Evaluación:** es la fase final que aplica un juicio de valor en el contexto laboral-profesión, social e investigativo, ésta se debe realizar a través del reconocimiento de logros y aspectos a mejorar se estará promoviendo el concepto de “evaluación para la mejora continua”, la metacognición, el desarrollo del pensamiento crítico y reflexivo en los estudiantes.

10. Evaluación por competencias

Son las técnicas, instrumentos y herramientas sugeridas para constatar los desempeños académicos de las actividades de aprendizaje.

Se recomienda evaluar los siguientes aspectos:

- Asistencia
- Participación (análisis crítico de los temas estudiados)
- Trabajo en equipo
- Prácticas
- Mapas Conceptuales y metales
- Exposiciones
- Demostraciones
- Proyecto integrador
- Investigaciones

11. Fuentes de información

Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A System for Large-scale Machine Learning. In Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, OSDI'16, USENIX Association, Savannah, GA, USA, 2–4 November 2016; pp. 265–283.

Duda, R.; Hart, P.; Stork, D. Pattern Classification, 2nd ed.; Wiley: New York, NY, USA, 2001.

Elshawi, R.; Sakr, S.; Talia, D.; Trunfio, P. Big Data Systems Meet Machine Learning Challenges: Towards Big Data Science as a Service. Big Data Res. 2018, 14, 1–11.

Evan Stubbs. Big Data, Big Innovation: Enabling Competitive Differentiation through Business Analytics (1st. ed.). Wiley Publishing. 2014.

Franklin, M.J.; et al. Apache Spark: A Unified Engine for Big Data Processing. Commun. ACM 2016, 59, 56–65.

Goodfellow, I.; Bengio, Y.; Courville, A. Deep Learning; MIT Press: Cambridge, MA, USA, 2016.

Hassan A. Karimi. Big Data: Techniques and Technologies in Geoinformatics (1st. ed.). CRC Press, Inc., USA. 2017.

Haykin, S. Neural Networks. A Comprehensive Foundation, 2nd ed.;

Pretince Hall: Upper Saddle River, NJ, USA, 1999.

Han, M. Kamber, and J. Pei, Data mining concepts and techniques, 2nd ed. Waltham, Mass.: Morgan Kaufmann Publishers, 2006.

Joyanes Aguilar, L. Big Data, análisis de grandes volúmenes de datos en organizaciones. México: Alfaomega, 2013.

Mayer-Schönberger, V., & Cukier, K. Big data. La revolución de los datos masivos (Noema). Madrid: Turner Naoma, 2013.

Mayer - Schonberger, V., & Cukier, K. Big Data: A Revolution That Will Transform How We Live, Work, and Think. London: Mariner Books, 2014.

Oussous, A.; Benjelloun, F.Z.; Lahcen, A.A.; Belfkih, S. Big Data technologies: A survey. J. King Saud Univ. Comput. Inf. Sci. 2018, 30, 431–448.

Provost, F., & Fawcett, T. Data Science for Business. Sebastopol, CA: O'Reilly & Associates Inc., 2013.

P. M. de Sá, Pattern Recognition: Concepts, Methods and Applications. Springer-Verlag Berlin Heidelberg, 2001.

Provost, F., & Fawcett, T. Data Science for Business. Sebastopol, CA: O'Reilly & Associates Inc., 2013.

Rabasa, A., & Brebbia, C. A. BIG DATA. (Vol. 1, pág. 300). NY: Wit Pr/Computational Mechanics, 2016.

Triguero, I.; Gonzalez, S.; Moyano, J.; Garcia, S.; Alcala-Fdez, J.; Luengo, J.; Fernandez, A.; del Jesus, M.; Sanchez, L.; Herrera, F. KEEL 3.0: An Open Source Software for Multi-Stage Analysis in Data Mining. Int. J. Comput. Intell. Syst. 2017, 10, 1238–1249.

Webb J. & O'Brien t. Big Data Now. United States: O'Reilly Media, Inc., 2014.