



## 1. Datos Generales de la asignatura

<b>Nombre de la asignatura:</b>	Big Data y sus aplicaciones
<b>Clave de la asignatura:</b>	IAD-2407
<b>SATCA<sup>1</sup>:</b>	2-3-5
<b>Carrera:</b>	Ingeniería en Inteligencia Artificial

## 2. Presentación

<p><b>Caracterización de la asignatura</b></p> <p>Esta asignatura aporta al perfil del Ingeniero en Inteligencia Artificial las siguientes habilidades:</p> <ul style="list-style-type: none"> <li>● Coordinar y participar en equipos multidisciplinarios para la aplicación de soluciones innovadoras en diferentes contextos de procesamiento masivo de datos.</li> <li>● Diseñar, implementar y gestionar bases de datos en entornos Big Data optimizando los recursos disponibles, conforme a las normas vigentes de manejo y seguridad de la información en la nube.</li> <li>● Implementar herramientas, técnicas y métodos para las soluciones Big Data que apoyen la productividad y competitividad de las organizaciones cumpliendo con estándares de calidad.</li> </ul> <p>Actualmente la cantidad de información que se genera diariamente en Internet ha tenido un crecimiento de manera exponencial, lo cual representa todo un reto los requerimientos de almacenamiento de forma distribuida, así como su procesamiento masivo al considerar grandes volúmenes de información, para su posterior análisis y explotación utilizando diversas herramientas y técnicas estadísticas. Para ello se cuentan con soluciones Big Data, es decir el procesamiento de grandes volúmenes de información mediante la implementación de soluciones basadas en los ecosistemas más comúnmente utilizados: Apache Hadoop y Spark.</p> <p>Esta asignatura promueve y desarrolla en el estudiante las competencias necesarias para integrar eficientemente soluciones de cómputo en la nube y Big Data que permita a las organizaciones nuevas herramientas de apoyo en la toma de decisiones para la mejora en la competitividad, innovación y adaptación constante a los nuevos escenarios.</p> <p>En esta asignatura se abordan las bases teóricas y prácticas referentes a cómputo en la nube, proveedores de servicios, fundamentos de Big Data, frameworks, modelos y técnicas para procesamiento masivo de datos, metodologías y tecnologías de almacenamiento distribuido, e implementación de soluciones de Big Data; aplica conocimientos de otras asignaturas, tales como: Base de Datos NoSQL, Diseño y Construcción e Implementación de Data Waterhouse y Análisis de Datos.</p> <p>El enfoque sugerido para la asignatura requiere que las actividades prácticas promuevan el desarrollo de habilidades para la configuración del ecosistema de soluciones para Big Data, utilizando las herramientas de que disponen Apache Hadoop o Spark, y con ello lograr el almacenamiento distribuido, procesamiento masivo de datos y explotación de la información.</p>
---

<sup>1</sup> Sistema de Asignación y Transferencia de Créditos Académicos



### Intención didáctica

A fin de obtener los resultados esperados, esta asignatura debe centrarse en la realización de prácticas y propiciar la participación en la implementación de un proyecto conjunto con otra asignatura(s) y/o carreras.

El contenido se organiza en cuatro temas: Fundamentos de Big Data, Procesamiento de Datos Masivos, Implementando tecnología Big Data.

También deben propiciar mediante prácticas, la implementación de casos de estudio reales que ofrezcan escenarios distintos que permitan la aplicación de los conceptos para lograr que el aprendizaje sea significativo para el desarrollo de las competencias.

En el desarrollo de la materia, se diseñarán actividades para el desarrollo de las siguientes competencias genéricas:

- Habilidades para buscar, procesar y analizar información procedente de diversas fuentes.
- Capacidad de abstracción, análisis y síntesis.
- Capacidad crítica y autocrítica.
- Capacidad de aplicar los conocimientos en la práctica.
- Que los contenidos sean abordados en su totalidad.
- Que se cuente con la infraestructura necesaria para realizar las prácticas
- Que el laboratorio de prácticas cuente con un software instalado (Apache Hadoop o Spark) para soluciones Big Data, el cual deberá utilizarse durante el desarrollo de la materia.
- Que toda práctica diseñada por el docente sea afín a los temas del programa.
- Que los estudiantes adquieran las competencias específicas de cada tema

### 3. Participantes en el diseño y seguimiento curricular del programa

Lugar y fecha de elaboración o revisión	Participantes	Observaciones
Tecnológico Nacional de México del 4 al 06 de marzo del 2024.	Representantes de los Institutos Tecnológicos de: Celaya, Chihuahua, Iztapalapa III, La Paz, Matehuala, Mérida, Minatitlán, Querétaro, Saltillo, Tijuana. Instituto Tecnológico Superior de Teziutlán. Tecnológico de Estudios Superiores de Ixtapaluca.	Propuesta sintética de la carrera de Ingeniería en Inteligencia Artificial.



<p>Tecnológico Nacional de México del 22 al 26 de abril del 2024</p>	<p>Representantes de los Institutos Tecnológicos de: Celaya, Chihuahua, Iztapalapa III, La Paz, Matehuala, Mérida, Minatitlán, Querétaro, Saltillo, Tijuana. Instituto Tecnológico Superior de Teziutlán, Tecnológico de Estudios Superiores de Ixtapaluca.</p>	<p>Diseño y/o desarrollo curricular de la carrera de Ingeniería en Inteligencia Artificial.</p>
<p>Tecnológico Nacional de México del 27 al 31 de mayo del 2024.</p>	<p>Representantes de los Institutos Tecnológicos de: Celaya, La Paz, Matehuala, Mérida, Minatitlán.</p>	<p>Consolidación curricular de la carrera de Ingeniería en Inteligencia Artificial.</p>

#### 4. Competencia(s) a desarrollar

<b>Competencia(s) específica(s) de la asignatura</b>
<ul style="list-style-type: none"> <li>● Identifica los principales modelos de infraestructura en la nube para dar soporte a soluciones de Big Data.</li> <li>● Conoce el ecosistema y los diversos frameworks para brindar soporte en el procesamiento de datos masivos.</li> <li>● Aplica modelos y herramientas para procesamiento de datos masivos basados en análisis predictivo, Mapreduce, y tecnologías de almacenamiento distribuido.</li> <li>● Implementa las etapas de preparación y análisis de datos para el desarrollo de soluciones usando tecnologías Big Data.</li> <li>● Crear un informe analítico completo utilizando diferentes herramientas para Big Data, que satisfaga las necesidades específicas de un conjunto de datos dado.</li> </ul>

#### 5. Competencias previas

<ul style="list-style-type: none"> <li>● Conoce el nuevo contexto de las bases de datos NoSQL y aprende a diferenciarlas con respecto a las bases de datos relacionales.</li> <li>● Conoce las principales técnicas de limpieza de datos.</li> <li>● Conoce y aplica las principales técnicas, procedimientos y herramientas para la extracción, transformación y carga de datos.</li> <li>● Comprende los conceptos básicos de análisis de datos y visualización, así como tener habilidades básicas en el manejo de datos utilizando al menos una herramienta de software.</li> <li>● Utiliza habilidades de resolución de problemas de forma óptima y estructurada.</li> </ul>
---



## 6. Temario

No.	Temas	Subtemas
1	Bases de datos en la nube.	1.1. Bases de Datos Relacional en la nube. 1.1.1. Aplicaciones. 1.2. Bases de Datos NoSQL. 1.2.1. Aplicaciones.
2	Fundamentos de Big Data.	2.1. Introducción al Big Data. 2.2. Ciencia de datos y científico de datos. 2.3. Ecosistema de información y sistemas de gestión de recursos. 2.3.1. Frameworks para procesamientos de datos masivos.
3	Procesamiento de datos masivos.	3.1. Introducción al procesamiento de datos masivos. 3.2. Modelos y herramientas de procesamiento de datos masivos. 3.2.1. Análisis predictivo. 3.2.2. Map-Reduce. 3.2.3. Preparación e integración de datos. 3.2.4. Almacenamiento distribuido. 3.3. Tecnologías de almacenamiento.
4	Implementando tecnología Big Data.	4.1. Etapa de preparación. 4.1.1. Fase inicial. 4.1.2. Planificación. 4.1.3. Fase de diseño. 4.1.4. Fase de implementación. 4.2. Etapa de Análisis de datos. 4.2.1. Fase de recopilación. 4.2.2. Fase de análisis. 4.2.3. Fase de agregación de valor.



5	Herramientas para Big Data	<p>5.1. Análisis de datos y generación de reportes</p> <p>5.1.1. Analítica de datos en el ambiente.</p> <p>5.1.2. Preparación de los datos para su análisis.</p> <p>5.1.3. Modelos.</p> <p>5.1.4. Visualización de datos.</p> <p>5.2. Análisis de datos y generación de reportes.</p> <p>5.2.1. Análisis de datos de redes sociales.</p> <p>5.2.2. Generación de reportes</p> <p>5.3. Analítica de datos.</p> <p>5.3.1. Introducción al espacio de trabajo.</p> <p>5.3.2. Manejo de una tabla.</p> <p>5.3.3. Trabajo con más de una tabla.</p> <p>5.3.4. Representación gráfica de un conjunto de datos.</p>
---	----------------------------	--

## 7. Actividades de aprendizaje de los temas

1. Bases de datos en la nube	
Competencias	Actividades de aprendizaje
<p><i>Específica(s):</i> Identifica los principales modelos de infraestructura en la nube para dar soporte a soluciones de Big Data.</p> <p><i>Genéricas:</i></p> <ul style="list-style-type: none"> <li>● Habilidades para buscar, procesar y analizar información procedente de diversas fuentes.</li> <li>● Capacidad de abstracción, análisis y síntesis.</li> <li>● Capacidad de comunicación oral y escrita.</li> <li>● Capacidad de trabajo en equipo.</li> <li>● Habilidad para trabajar en forma autónoma.</li> </ul>	<ul style="list-style-type: none"> <li>● Investigar los conceptos fundamentales de cómputo en la nube: Antecedentes, Modelos de implementación y Modelos de servicio.</li> <li>● Elaborar una tabla comparativa con las principales ventajas y retos de cómputo en la nube.</li> <li>● Identificar los principales proveedores de servicios y costos de cómputo en la nube.</li> <li>● Elaborar una guía para elegir un proveedor de servicios.</li> <li>● Comparar los tipos de bases de datos relacionales y NoSQL.</li> <li>● soportados en la Nube.</li> </ul>



<b>2. Fundamentos de Big Data</b>	
Competencias	Actividades de aprendizaje
<p><i>Específica(s):</i> Conoce el ecosistema y los diversos frameworks para brindar soporte en el procesamiento de datos masivos.</p> <p><i>Genérica(s):</i></p> <ul style="list-style-type: none"> <li>● Habilidades para buscar, procesar y analizar información procedente de diversas fuentes.</li> <li>● Capacidad de abstracción, análisis y síntesis</li> <li>● Capacidad de comunicación oral y escrita</li> </ul>	<ul style="list-style-type: none"> <li>● Investigar las principales ventajas y áreas de aplicación del Big Data, así como las nuevas oportunidades que se generan mediante su implementación.</li> <li>● Realizar un cuadro comparativo con los roles que se derivan de la ciencia de datos (científico de datos, analista de datos, ingeniero de datos, etc.)</li> <li>● Identificar la arquitectura y principales componentes del ecosistema para procesamiento de datos masivos.</li> <li>● Realizar un estudio comparativo de los principales frameworks para procesamientos de datos masivos, considerando dos de los principales de la industria: Hadoop y Spark, analizando las principales características, ventajas, desventajas, y arquitectura.</li> </ul>
<b>3. Procesamiento de datos masivos</b>	
Competencias	Actividades de aprendizaje
<p><i>Específica(s):</i> Aplica modelos y herramientas para procesamiento de datos masivos basados en análisis predictivo, Map-Reduce, y tecnologías de almacenamiento distribuido.</p> <p><i>Genérica(s):</i></p> <ul style="list-style-type: none"> <li>● Habilidades para buscar, procesar y analizar información procedente de diversas fuentes.</li> <li>● Capacidad de abstracción, análisis y síntesis.</li> <li>● Capacidad crítica y autocrítica.</li> <li>● Capacidad de aplicar los conocimientos en la práctica.</li> <li>● Capacidad para identificar, plantear y resolver problemas.</li> <li>● Trabajo en equipo.</li> <li>● Toma de decisiones.</li> </ul>	<ul style="list-style-type: none"> <li>● Investigar lo relacionado al procesamiento de datos masivos, identificando las principales técnicas, métodos de procesamiento y herramientas.</li> <li>● Desarrollar un caso práctico aplicando análisis predictivo (ventas, marketing, gestión de riesgos, detección de fraudes, salud, deportes, etc.)</li> <li>● Aplicando Map Reduce mediante un lenguaje de programación (p.e. Python), desarrolle un ejemplo para contar palabras de un texto determinado.</li> <li>● Identificar los pasos a seguir en la preparación e integración de datos para su análisis posterior en una solución de Big Data.</li> <li>● Aplicar técnicas y tecnologías de almacenamiento distribuido para dar soporte a una solución de Big Data.</li> </ul>



<b>4. Implementando tecnología Big Data</b>	
Competencias	Actividades de aprendizaje
<p><i>Específica(s):</i> Implementa las etapas de preparación y análisis de datos para el desarrollo de soluciones usando tecnologías Big Data.</p> <p><i>Genéricas:</i></p> <ul style="list-style-type: none"> <li>• Habilidades para buscar, procesar y analizar información procedente de diversas fuentes.</li> <li>• Capacidad de aplicar los conocimientos en la práctica.</li> <li>• Capacidad para identificar, plantear y resolver Problemas.</li> <li>• Trabajo en equipo.</li> <li>• Toma de decisiones.</li> </ul>	<ul style="list-style-type: none"> <li>• Desarrollar y documentar un caso práctico en el cual aplique cada una de las fases de la etapa de preparación: Fase inicial, Fase de planificación, Fase de diseño y Fase de implementación.</li> <li>• Desarrollar y documentar un caso práctico en el cual aplique cada una de las fases de la etapa de análisis de datos: Fase recopilación, Fase de Análisis y Fase de agregación de valor.</li> </ul>
<b>5. Herramientas para Big Data</b>	
Competencias	Actividades de aprendizaje
<p><i>Específica(s):</i> Elaborar informes analíticos detallados utilizando Microsoft Power BI, R y Python, aplicando técnicas específicas de preparación, modelado y visualización de datos en cada una de las herramientas.</p> <p><i>Genéricas:</i></p> <ul style="list-style-type: none"> <li>• Interpretar datos y transformarlos en información útil.</li> <li>• Aplicar técnicas de preparación y limpieza de datos.</li> <li>• Utilizar herramientas analíticas para generar insights.</li> <li>• Comunicar efectivamente resultados a través de visualizaciones y reportes.</li> <li>• Integrar múltiples fuentes de datos para análisis complejos.</li> </ul>	<p>Seminario introductorio:</p> <ul style="list-style-type: none"> <li>• Presentación de conceptos básicos de análisis de datos y herramientas para Big Data.</li> <li>• Demostración de Microsoft Power BI, R y Python para análisis de datos.</li> </ul> <p>Laboratorio práctico:</p> <ul style="list-style-type: none"> <li>• Ejercicio guiado sobre preparación de datos en Microsoft Power BI.</li> <li>• Ejercicio práctico sobre análisis de datos de redes sociales utilizando R.</li> <li>• Ejercicio de manipulación de datos utilizando Python.</li> </ul>



## 8. Práctica(s)

- Instalar y configurar un ambiente para Big Data mediante Hadoop o Spark.
- Crear espacios para almacenamiento distribuido mediante el framework seleccionado.
- Instalar y configurar una base de datos NoSQL para gestión de datos no estructurados.
- Realizar prácticas con el manejador de bases de datos NoSQL mediante la implementación de los procesos Map-Reduce.
- Desarrollo de casos prácticos de análisis predictivo utilizando técnicas estadísticas, abordando problemas y desafíos específicos del dominio de interés público o empresarial (por ejemplo, salud, finanzas, marketing).
- Elaboración de una propuesta de implementación de Big Data para la solución de una problemática real identificada en algún área de la industria o institución en particular, aplicando las etapas de preparación y de análisis de datos con sus respectivas fases en cada una de las etapas.
- Práctica con Microsoft Power BI: Preparación de un conjunto de datos proporcionado para su análisis.
- Creación de modelos y visualizaciones en Power BI.
- Generación de un informe detallado sobre los insights obtenidos.
- Práctica con R y Python: Recopilación de datos de una fuente específica (por ejemplo, red social) utilizando el paquete en R, análisis de datos utilizando técnicas relevantes en R, y representación gráfica de los resultados utilizando Python.

## 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 comunitario, 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

Para evaluar las actividades de aprendizaje se recomienda solicitar: Mapa conceptual, tablas comparativas, examen teórico, examen práctico, reportes escritos de investigación, reportes de prácticas, guía de proyectos.

Para verificar el nivel de logro de las competencias del estudiante se recomienda utilizar: Rubricas, Matriz de valoración, Matriz Avance de proyecto integrador

## 11. Fuentes de Información

1. Bengfort, B., Kim, J. (2016). Data Analytics with Hadoop. An Introduction for Data Scientists. Publisher: O'Reilly Media.
2. Hu, F..(2016). Big Data: Storage, Sharing, and Security. Publisher: CRC Press.
3. Karau, H., Konwinski, A., Wendell, P., Matei Zaharia, M. (2015). Learning Spark: Lightning-Fast Big Data Analysis. Publisher: O'Reilly Media Inc.
4. Loukides, M. (2011). What is Data Science?. Publisher: O'Reilly Media Inc.
5. Lublinsky B., Smith, K.T., Yakubovich, A. (2013). Professional Hadoop Solutions. Publisher: John Wiley & Sons.
6. Lublinsky, B., Smith, K.T.(2014). Hadoop: Soluciones Big Data. Editorial : Anaya Multimedia.
7. Macías, M., Gómez, M., Tous, R., Torres J.(2015). INTRODUCCIÓN A APACHE SPARK.
8. McKindsey & Company(2013). Big Data, Analytics, and the Future of Marketing & Sales. Publisher: McKinsey & Company.
9. Perera, S., Gunarathne, T. (2013). Hadoop MapReduce Cookbook. Publisher: Pack.
10. Prajapati, V. (2013). Big Data Analytics with R and Hadoop. Publisher: Packt Publishing.
10. White, T.(2012). Hadoop: The Definitive Guide, 3rd edition. Publisher: O'Reilly Media Inc.