

1. Datos Generales de la asignatura

Nombre de la asignatura:	Grandes Datos y NoSQL
Clave de la asignatura:	CDD-2103
SATCA¹:	2-3-5
Carrera:	Ingeniería en Sistemas Computacionales.

2. Presentación

Caracterización de la asignatura
<p>Una de las tecnologías más importantes y recientes en el mundo del manejo de datos es lo que se conoce como Grandes Datos.</p> <p>Grandes datos, <i>Big Data</i>, grandes volúmenes de datos o macrodatos, como uno le quiera llamar, están constituidos por la avalancha de datos generados por una multitud de tendencias tecnológicas: dispositivos móviles, redes sociales, internet, computación en la nube, almacenes de datos, entre otras.</p> <p>Los <i>Big Data</i> crecen de manera espectacular cada año, teniendo en algunos casos, empresas que almacenan datos cuyo tamaño alcanzan los petabytes.</p> <p>Esta asignatura aporta al perfil del Ingeniero en Sistemas Computacionales los conocimientos necesarios para almacenar, recuperar y analizar datos en un ambiente de <i>Big Data</i>, haciendo uso de herramientas actuales para dichas tecnologías.</p>
Intención didáctica
<p>El contenido temático se organiza en cuatro unidades.</p> <p>La primera unidad es una introducción al tema de <i>Big Data</i>. En ésta se exponen los conceptos básicos de los Grandes datos, la importancia de los mismos, así como sus áreas de aplicación.</p>

¹ Sistema de Asignación y Transferencia de Créditos Académicos

En la segunda unidad se ponen en práctica los elementos básicos de una arquitectura de *Big Data*: almacenamiento, procesamiento y reducción en paralelo de datos distribuidos.

La tercera unidad se aplica lo que se conoce como Analítica de Datos, tema que permite al estudiante reafirmar las bases para fungir como un Científico de Datos.

La última unidad trata un tema de gran interés: Computación en la nube, pero enfocado a Grandes Datos.

Finalmente, es de suma importancia que las actividades prácticas promuevan, mediante la utilización de herramientas comerciales vigentes en el mercado, el desarrollo de habilidades para el análisis, modelado, diseño y desarrollo de sistemas de bases de datos móviles empresariales. Asimismo, es de suma relevancia propiciar la implementación de casos de estudio reales que ofrezcan escenarios distintos, mediante suficientes prácticas que permitan la aplicación de los conceptos y técnicas de programación vistas en clase con el fin de que el aprendizaje sea más significativo para el desarrollo de las competencias.

3. Participantes en el diseño y seguimiento curricular del programa

Lugar y fecha de elaboración o revisión	Participantes	Observaciones
Instituto tecnológico de Zacatepec. Departamento de sistemas y Computación Academia de sistemas computación. Reunión para el desarrollo de Especialidades de Noviembre 2020 a Marzo 2021.	Tecnológico Nacional de México campus Instituto Tecnológico de Zactepec DR. Francisco Javier Cartujano Escobar M.C. Claudia Noguérón González Lic. Estela Rodríguez Zavaleta M.T.I. Laura Villavicencio Gómez M.C. Boris Antonio Aranda Benitez M.C. José Francisco Carpio Tovilla M.C.I. Misael López Nava M.C. Leticia Santa Olalla Ocampo Dra. Ana Celia Campos Hernández	Programa elaborado por profesores del departamento de Sistemas y Computación, presentado y aprobado en el pleno de la Academia. Esta materia forma parte de la especialidad Ciencia de Datos para la Toma de Decisiones , para la carrera Ingeniería en Sistemas Computacionales, Plan de estudios 2010

	MT.I. Claudia Gabriela Bustillos Gaytan	
--	---	--

4. Competencia a desarrollar

Implementa arquitecturas de Grandes Datos (Big Data) con la capacidad de aplicar técnicas de analítica de datos y con la capacidad de almacenar y procesar datos en la nube.

5. Competencias previas

- Instalar y configurar redes de área local y redes inalámbricas.
- Manipular bases de datos NoSQL.
- Dominar herramientas ETL.
- Codificar aplicaciones en Lenguaje Java y Python
- Operar sistemas operativos, principalmente la familia LINUX.
- Usar manejadores de base de datos mediante SQL

6. Temario

No.	Temas	Subtemas
1	Introducción a Grandes Datos (<i>Big Data</i>)	1.1 Caracterización de los Grandes Datos. 1.2 Retos en el procesamiento de grandes volúmenes de datos 1.3 Analítica de Grandes Datos en las organizaciones. 1.4 Arquitectura de los Grandes Datos. 1.5 Áreas de aplicación de los Grandes Datos

2	Implementación de una Arquitectura de Grandes Datos	2.1 Ecosistema de los Grandes datos 2.2 Distribuciones para la implementación de un ecosistema de Grandes Datos 2.3 Almacenamiento, distribución y replicación de datos en ambientes de Grandes Datos 2.4 Importación / Exportación de datos en ambientes de Grandes datos 2.5 Procesamiento de datos en paralelo utilizando el marco de trabajo MapReduce 2.6 Monitoreo del procesamiento de datos en paralelo
3	Análítica de los Grandes Datos	3.1 Importancia de la analítica de los Grandes Datos 3.2 Ciencia de los datos 3.3 Lenguajes para la analítica de los grandes datos 3.4 Desarrollo de aplicaciones para analítica de los grandes datos
4	Grandes Datos en la nube	4.1 Conceptualización de Grandes Datos en la nube. 4.2 Proveedores de Grandes Datos en la nube 4.3 Almacenamiento y procesamiento de Grandes Datos en la nube

7. Actividades de aprendizaje de los temas

Unidad 1: Introducción a grandes datos (<i>Big Data</i>)	
Competencias	Actividades de aprendizaje
<p>Específica:</p> <p>Entender el propósito y beneficios de implementar una arquitectura de Big Data.</p> <p>Genéricas:</p> <ul style="list-style-type: none"> • Habilidades para buscar, procesar y analizar información procedente de diversas fuentes. 	<ul style="list-style-type: none"> • Investigar conceptos relacionados al tema en internet, libros y artículos. • Acceder foros de discusión sobre uso, tendencias y convergencia de la tecnología. • Elaboración de mapa mental.

<ul style="list-style-type: none"> • Capacidad de abstracción, análisis y síntesis • Capacidad de aprender por cuenta propia. • Capacidad de comunicación oral y escrita. • Capacidad de trabajo en equipo. 	
Unidad 2: Implementación de una Arquitectura de Grandes Datos	
Competencia	Actividades de aprendizaje
<p>Específica</p> <ul style="list-style-type: none"> • Implementar una arquitectura de Grandes Datos para procesar y reducir en paralelo grandes volúmenes de datos, principalmente las proporcionadas por el ecosistema Hadoop. <p>Genéricas:</p> <ul style="list-style-type: none"> • Habilidades para buscar, procesar y analizar información procedente de diversas fuentes. • Capacidad de abstracción, análisis y síntesis • Capacidad de aprender por cuenta propia. • Capacidad de comunicación oral y escrita. • Capacidad de trabajo en equipo. 	<ul style="list-style-type: none"> • Instalar los programas de cómputo necesarios para implementar una arquitectura de Grandes Datos. • Investigar los temas en internet, libros y manuales. • Desarrollar pequeñas aplicaciones para la automatización de procesamiento y reducción de datos. • Avanzar en el desarrollo del proyecto final integrador
Unidad 3: Analítica de Grandes Datos	
Competencia	Actividades de aprendizaje
<p>Específica</p> <ul style="list-style-type: none"> • Utilizar herramientas para analizar y obtener conocimiento significativo que sea relevante para toma de decisiones en un ambiente corporativo. <p>Genéricas:</p>	<ul style="list-style-type: none"> • Instalar las herramientas de desarrollo. • Investigar los temas en internet, libros y manuales. • Realizar prácticas que utilicen herramientas actuales para analítica de datos.

<ul style="list-style-type: none"> • Habilidades para buscar, procesar y analizar información procedente de diversas fuentes. • Capacidad de abstracción, análisis y síntesis • Capacidad de aprender por cuenta propia. • Capacidad de comunicación oral y escrita. • Capacidad de trabajo en equipo. 	<ul style="list-style-type: none"> • Avanzar en el desarrollo del proyecto final integrador
Unidad 4: Grandes Datos en la nube	
Competencia	Actividades de aprendizaje
<p>Específica</p> <ul style="list-style-type: none"> • Entender el proceso de almacenar y procesar Grandes Datos en la nube <p>Genéricas:</p> <ul style="list-style-type: none"> • Habilidades para buscar, procesar y analizar información procedente de diversas fuentes. • Capacidad de abstracción, análisis y síntesis • Capacidad de aprender por cuenta propia. • Capacidad de comunicación oral y escrita. • Capacidad de trabajo en equipo. 	<ul style="list-style-type: none"> • Investigar los temas en internet, libros y manuales. • Utilizar herramientas para almacenar y procesar Grandes Datos en la nube. • Concluir y exponer el proyecto final integrador

8. Prácticas

<ul style="list-style-type: none"> • Investigar las tecnologías involucradas en la implementación de una arquitectura de Big Data. • Investigar áreas de aplicación de la tecnología de Big Data. • Investigar las diferentes distribuciones para la implementación de una arquitectura de Grandes Datos. • Instalar, configurar y actualizar el ecosistema Hadoop. • Investigar características y uso de HDFS.
--

- Implementar procesos de exportación/importación en ambiente de Hadoop
- Implementar procesos que accedan y reduzcan (transformen) Grandes Datos utilizando principalmente la herramienta MapReduce de Hadoop.
- Desarrollar procesos de Analítica de Datos utilizando Spark-SQL, Pig y Hive.
- Investigar otras herramientas para la Analítica de Datos
- Investigar herramientas para el almacenamiento y manipulación de Datos en la nube.
- Desarrollar procesos de Analítica de Datos en la nube.

9. Proyecto de asignatura

Un elemento esencial en la asignatura es el desarrollo de un proyecto integrador a realizar por el estudiante. El objetivo del proyecto que planteé el docente 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 a) planificar un proceso de intervención empresarial, científico o social; b) el diseño de un modelo, entre otros, según el tipo de proyecto; c) las actividades a realizar; d) los recursos requeridos y e) 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 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 o de investigación. Ésta se debe realizar a través del reconocimiento de logros y aspectos a mejorar. Se deberá promover el concepto de “evaluación para la mejora continua”, la metacognición, el desarrollo del pensamiento crítico y reflexivo en los estudiantes.

10. Evaluación por competencias

La evaluación debe ser continua, formativa, flexible e integral, por lo que se debe considerar el desempeño en cada una de las actividades de aprendizaje, haciendo especial énfasis en lo siguiente:

- Realizar una evaluación diagnóstica al inicio del semestre, para partir de saberes previos, expectativas e intereses que tengan los estudiantes.
- Durante el desarrollo del curso debe llevarse a cabo una evaluación formativa que permita realimentar el proceso de aprendizaje y establecer las estrategias para el logro de los objetivos establecidos. Realizar evaluaciones mediante: exámenes teóricos y prácticos, desarrollo de prácticas de laboratorio, tarea y ejercicios.
- Al finalizar el curso debe realizarse una evaluación sumativa que se vincula con aquellas acciones que se orientan a dar cuenta de productos, saberes, desempeños y actitudes que se deben considerar para la calificación.
- Se sugiere utilizar como herramienta de evaluación el portafolio de evidencias, y como instrumento la lista de cotejo y la rúbrica.

Algunos productos sugeridos para la evaluación son:

- Exámenes escritos para comprobar el manejo de aspectos teóricos y declarativos.
- Exámenes prácticos que demuestren el conocimiento adquirido en relación a la inteligencia del negocio.
- Portafolio de evidencias de las prácticas realizadas en el laboratorio.
- Realización de tareas-ejercicios fuera de clase.
- Reportes escritos de las observaciones hechas durante las actividades, así como avances del proyecto integrador.
- Información obtenida durante las investigaciones solicitadas plasmada en documentos escritos o digitales.
- Presentación de información mediante mapas mentales y conceptuales.
- Realización de un proyecto integrador donde se plasmen los conocimientos adquiridos durante la materia y su relación con otras.

11. Fuentes de información

- a. Pérez, M. 2015. **Big Data: Técnicas, herramientas y aplicaciones.** Editorial Alfaomega.
- b. Chambers, B., & Zaharia, M. 2018. **Spark: The definitive Guide, Big Data processing Made Simple.** Editorial O'Reilly.
- c. Bruce, P., & Bruce, A. 2017. **Practical Statistics for Data Scientists.** Editorial O'Reilly.
- d. Tomcy, J., & Pankaj, M. 2017. **Data Lake for Enterprises.** Editorial Packt.
- e. Joyanes, L. 2016. **Big Data: Análisis de grandes volúmenes de datos en organizaciones.** Alfaomega.
- f. A. Jorgesen, J., Rowland-Jones, J., Welch, D., Clark, C., & Price, B. Mitchell. 2014. **Microsoft Big Data Solutions.** Editorial John Wiley and Sons.
- g. Lublinsky, B., Smith, K., & Yakubovich, A. 2013. **Professional Hadoop Solutions.** Editorial WROX.
- h. Tigani, J. & Naidu, S. 2014. **Google BigQuery Analytics.** Editorial John Wiley and Sons.
- i. María Pérez Marquéz. **Microsoft SQL Azure: Administración y desarrollo en la nube.** Editorial Alfaomega, 2012