

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

COORDINACIÓN GENERAL DE FORMACIÓN BÁSICA

COORDINACIÓN GENERAL DE FORMACIÓN PROFESIONAL Y VINCULACIÓN UNIVERSITARIA

PROGRAMA DE UNIDAD DE APRENDIZAJE

I. DATOS DE IDENTIFICACIÓN

- 1. Unidad Académica:** Facultad de Ingeniería, Mexicali; Facultad de Ingeniería, Arquitectura y Diseño, Ensenada; y Facultad de Ciencias Químicas e Ingeniería, Tijuana.
- 2. Programa Educativo:** Ingeniero en Computación
- 3. Plan de Estudios:** 2020-1
- 4. Nombre de la Unidad de Aprendizaje:** Datos Masivos
- 5. Clave:** 36315
- 6. HC:** 02 **HL:** 00 **HT:** 02 **HPC:** 00 **HCL:** 00 **HE:** 02 **CR:** 06
- 7. Etapa de Formación a la que Pertenece:** Terminal
- 8. Carácter de la Unidad de Aprendizaje:** Optativa
- 9. Requisitos para Cursar la Unidad de Aprendizaje:** Ninguno



Equipo de diseño de PUA

Sukey Sayonara Nakasima López
Mauricio Alonso Sánchez Herrera
Félix Fernando González Navarro
Juan Pablo Torres Herrera

Vo.Bo. de subdirector(es) de Unidad(es) Académica(s)

Alejandro Mungaray Moctezuma
Humberto Cervantes de Ávila
Rocío Alejandra Chávez Santoscoy

Fecha: 17 de octubre de 2019

II. PROPÓSITO DE LA UNIDAD DE APRENDIZAJE

El propósito de la unidad de aprendizaje Datos Masivos es brindar los conocimientos y habilidades para la implementación de un entorno que permita de forma eficiente la recolección, almacenamiento, integración, seguridad y gestión, así como el análisis, visualización e inferencias de los resultados obtenidos del procesamiento de los datos masivos.

Su utilidad radica en que el estudiante puede aplicar los conceptos de datos masivos, mediante técnicas y metodologías para el diseño e implementación de un entorno propicio que cumpla con los requerimientos necesarios para su gestión y analítica de datos, con la finalidad de extraer ideas de valor que sirvan como referentes estratégicos para la toma de decisiones.

Esta asignatura se imparte en la etapa terminal con carácter optativo. Pertenece al área de conocimiento Ingeniería aplicada. Se recomienda que al momento de cursarla el estudiante se encuentre en el octavo período del plan de estudios, después de haber adquirido los conocimientos y habilidades referentes a inteligencia artificial, ciencia de datos y cómputo en la nube.

III. COMPETENCIA DE LA UNIDAD DE APRENDIZAJE

Diseñar e implementar un entorno de gestión eficiente y segura de datos masivos, con la finalidad de extraer ideas de valor que coadyuven en la toma de decisiones para la resolución de problemas complejos, a través del análisis, modelado de procesos, herramientas de integración, extracción y transformación de datos masivos, con actitud crítica, proactiva y colaborativa.

IV. EVIDENCIA(S) DE DESEMPEÑO

Elabora el modelado, diseño e implementación de un entorno para la gestión y analítica de datos masivos, generando como producto entregable un reporte técnico, que cumpla con los requerimientos planteados de casos de estudio (este último considerado como proyecto final), logrando indicadores relevantes para la toma de decisiones.

V. DESARROLLO POR UNIDADES

UNIDAD I. El rol de los datos como motor económico, social y tecnológico

Competencia:

Determinar el lugar que ocupan los datos desde una perspectiva económica, social, industrial y tecnológica, para valorar las necesidades actuales, tendencias y los beneficios de su utilización, mediante el análisis de su evolución y aplicación en los diferentes sectores, con actitud crítica, reflexiva y colaborativa.

Contenido:**Duración:** 2 horas

- 1.1 La economía de los datos: Definición, evolución y situación actual desde una perspectiva económica.
- 1.2 Industria 4.0 y las tecnologías digitales.
- 1.3 Sectores pioneros en datos masivos.
- 1.4 Evolución de los sistemas de gestión y almacenamiento de datos.

UNIDAD II. Datos masivos

Competencia:

Analizar conceptos del nuevo paradigma de datos masivos, mediante la identificación de los elementos característicos que distinguen este paradigma con respecto a los anteriores, para determinar el flujo de las actividades que permiten generar valor e ideas útiles en la toma de decisiones, de forma proactiva, coherente y crítica.

Contenido:

Duración: 4 horas

2.1 Datos masivos (Big Data)

2.1.1 Desmitificación de este nuevo paradigma

2.1.2 Dimensionalidad y características

2.1.2.1 Volumen

2.1.2.2 Velocidad

2.1.2.3 Variedad

2.1.2.4 Veracidad

2.1.2.5 Valor

2.2 Tipos de datos

2.2.1 Estructurados, no estructurados, semi-estructurados y meta-data

2.3 Transformación de sectores industriales (casos de estudio)

2.4 Cadena de valor de Datos Masivos: descripción del flujo de una serie de pasos necesarios para generar valor e ideas útiles desde los datos

2.4.1 Adquisición de datos

2.4.2 Análisis de datos

2.4.3 Curación de datos

2.4.4 Almacén de datos

2.4.5 Uso de los datos

2.5 Cronología del origen y el estado actual de los datos masivos a través de: Documentos de divulgación y desarrollos basados en ellos, empresas, herramientas.

UNIDAD III. Ecosistema técnico para datos masivos

Competencia:

Explicar la composición de un ecosistema técnico para datos masivos, mediante la identificación de sus elementos indispensables, para trabajar con datos masivos, de forma proactiva, crítica y analítica.

Contenido:

Duración: 6 horas

3.1 Arquitectura de datos masivos en capas

3.1.1 Capa de infraestructura

3.1.1.1 Cómputo en la nube (almacenamiento, distribución, paralelización)

3.1.2 Capa de cómputo

3.1.2.1 Herramientas para la integración, gestión de datos y modelo de programación

3.1.3 Capa de aplicación

3.1.3.1 Interfaces para la implementación de modelos de programación y análisis de datos.

3.2 Componentes tecnológicos

3.2.1 Sistemas de archivo y bases de datos distribuidas

3.2.2 Plataformas de cómputo en la nube

3.2.3 Sistemas de almacenamiento escalable

3.2.4 Procesamiento masivo paralelo (MPP)

3.3 Gobernabilidad en datos masivos

3.4 Roles

3.4.1 Negocios

3.4.2 Científico de datos

3.4.3 IT

UNIDAD IV. Integración de datos masivos

Competencia:

Integrar fuentes de datos a un entorno de datos masivos, mediante el modelado del proceso ETL y la implementación de un framework, para su gestión, de forma metódica, creativa y colaborativa.

Contenido:

Duración: 10 horas

4.1 Modelo de datos

4.1.1 Key-value

4.1.2 Documentos

4.1.3 Wide-column

4.1.4 Gráficos

4.2 Herramientas de extracción, transformación y carga (ETL)

4.2.1 Proceso de ETL

4.2.1.1 Elementos del proceso ETL

4.2.1.2 Uso de técnicas BPNM (Business Process Model and Notation) para modelar el proceso ETL

4.2.2 Framework para la implementación del proceso ETL

4.2.2.1 Arquitectura del framework

4.2.2.2 Implementación del framework

4.2.2.3 Integración de la(s) fuente(s) de dato(s)

4.2.2.4 API's para la manipulación de los datos integrados

UNIDAD V. Analítica de Datos Masivos

Competencia:

Aplicar la metodología de KDD para el análisis de datos, mediante bibliotecas de consulta y machine learning, para la generación de reportes de distintos tipos de análisis y visualización que coadyuven en la toma de decisiones, de forma crítica, creativa y colaborativa.

Contenido:

Duración: 10 horas

- 5.1 Tipos de Análisis
 - 5.1.1 Descriptiva
 - 5.1.2 Diagnóstica
 - 5.1.3 Predictiva
 - 5.1.4 Prescriptiva
- 5.2 Tecnologías emergentes de analítica
- 5.3 Proceso de descubrimiento de conocimiento en base de datos KDD
- 5.4 Principales métodos de minado de datos
- 5.5 Bibliotecas de consulta para el manejo de datos masivos
- 5.6 Bibliotecas de Minería de Datos y Machine Learning
- 5.7 Visualización
 - 5.7.1 Fundamentos de visualización
 - 5.7.2 Herramientas de visualización
- 5.8 Interpretación

VI. ESTRUCTURA DE LAS PRÁCTICAS DE TALLER

No. de Práctica	Competencia	Descripción	Material de Apoyo	Duración
UNIDAD I				
1	Reflexionar sobre la necesidad de los datos masivos en la actualidad, a partir de su análisis desde las perspectivas económica, social, industrial y tecnológica, para contrastar distintos puntos de vista y valorar su importancia, con respeto y actitud colaborativa.	Investiga casos de estudio que narren el surgimiento de los datos masivos, así como su importancia en diferentes sectores (económico, social, industria y tecnológico). Realiza una presentación electrónica (powerpoint o afines) y exposición frente al grupo.	Equipo de cómputo con conexión a internet, consulta de base de datos y/o libros electrónicos.	2 horas
UNIDAD II				
2	Contrastar los elementos característicos que distinguen el paradigma de datos masivos con respecto a los demás (datos estructurados y relacionales, OLAP, datawarehouse, entre otras), a través de investigaciones bibliográficas y referencias electrónicas, para identificar y relacionar los conceptos con el nuevo paradigma, de forma analítica, colaborativa y comunicativa.	Investiga en lecturas proporcionadas por el docente, así como en otras referencias electrónicas, los elementos característicos de datos masivos y sus diferencias con los paradigmas anteriores, Sintetiza la información apoyado de esquemas mentales o mapas conceptuales y realiza una presentación electrónica (powerpoint o afines) para exponerla frente al grupo.	Equipo de cómputo con conexión a internet, consulta de base de datos, libros electrónicos y lecturas proporcionadas por el docente.	1 hora
3	Examinar el flujo de datos en una organización, mediante el	Investiga en casos de estudio proporcionados por el docente, así	Equipo de cómputo con conexión a internet, consulta	1 hora

	análisis de casos de estudio, para identificar la serie de pasos que permiten generar valor e ideas útiles a partir de datos, de forma crítica, colaborativa y comunicativa.	como en otras referencias electrónicas acerca de la cadena de valor de datos masivos, Sintetiza la información apoyado de esquemas mentales o mapas conceptuales y una presentación electrónica (powerpoint o afines) para exponerla frente al grupo.	de base de datos, libros electrónicos y lecturas proporcionadas por el docente.	
4	Explicar la cronología del origen y el estado actual de los datos masivos, a partir de su esquematización y exposición, con la finalidad de conocer su evolución tecnológica y metodológica, de forma analítica, inquisitiva y colaborativa.	Esquematiza información técnica de algunas herramientas, mediante artículos de divulgación. Crea una presentación electrónica (powerpoint o afines) y expone frente al grupo.	Equipo de cómputo con conexión a internet, consulta de base de datos, libros electrónicos y lecturas proporcionadas por el docente.	2 horas
UNIDAD III				
5	Analizar la arquitectura de datos masivos, mediante su esquematización y exposición, para identificar los componentes esenciales de la estructura de su ecosistema, de forma analítica y crítica.	Realiza una investigación en internet sobre la arquitectura de datos masivos. Esquematiza la información obtenida identificando los componentes esenciales de su estructura. Crea una presentación electrónica (powerpoint o afines) y la expone frente al grupo.	Equipo de cómputo con conexión a internet, consulta de base de datos, libros electrónicos y lecturas proporcionadas por el docente.	2 horas
6	Explicar los diferentes componentes tecnológicos de un ecosistema de datos masivos, mediante la búsqueda y análisis detallado de sus elementos fundamentales, para construir un ecosistema para datos masivos, de forma crítica e	Realiza una investigación documental en referencias bibliográficas y electrónicas sobre los componentes tecnológicos de datos masivos. Analiza y compila la información obtenida y crea una presentación electrónica para exponerla frente al grupo.	Equipo de cómputo con conexión a internet, consulta de base de datos, libros electrónicos y lecturas proporcionadas por el docente.	2 horas

	inquisitiva.			
7	Examinar la gobernabilidad y los roles en datos masivos, a partir de una investigación documental de los mismos, con la finalidad de descubrir e identificar los diferentes niveles de uso y acceso que tendrán los distintos actores al ecosistema de datos masivos, de forma analítica y comunicativa.	Realiza una investigación documental en referencias bibliográficas y electrónicas respecto a la gobernabilidad y roles en datos masivos. Analiza y compila la información obtenida identificando los niveles de uso y acceso de los actores. Crea una presentación electrónica para exponerla frente al grupo.	Equipo de cómputo con conexión a internet, consulta de base de datos, libros electrónicos y lecturas proporcionadas por el docente.	2 horas
UNIDAD IV				
8	Construir un modelo de datos para datos masivos, con la finalidad de poder representar de manera conceptual la naturaleza de los datos, de forma analítica, crítica, inquisitiva, colaborativa y comunicativa	Investiga documentalmente en referencias bibliográficas y electrónicas respecto al modelado de datos en datos masivos. Analiza y compila la información obtenida y crea una presentación electrónica para exponerla frente al grupo.	Equipo de cómputo con conexión a internet, consulta de base de datos, libros electrónicos y lecturas proporcionadas por el docente.	2 horas
9	Explicar el proceso de ETL para datos masivos, mediante la distinción de sus elementos esenciales, para la integración de datos en un ecosistema de datos masivos, de forma analítica, crítica, e inquisitiva.	Realiza una investigación documental en referencias bibliográficas y electrónicas respecto al proceso de ETL para datos masivos utilizando la técnica de BPNM basado en un caso de estudio Analiza y compila la información obtenida distinguiendo sus elementos esenciales. Crea una presentación electrónica para	Equipo de cómputo con conexión a internet, consulta de base de datos, libros electrónicos y lecturas proporcionadas por el docente.	2 horas

		exponerla frente al grupo.		
10	Experimentar la configuración de un ecosistema de datos masivos que permita la integración de fuentes de datos y la aplicación de herramientas analíticas, a partir de la arquitectura y el proceso de implementación de un framework orientado a este tipo de datos, para la extracción de ideas de valor, de forma analítica, crítica e inquisitiva.	El docente brinda lecturas referentes a la arquitectura y el proceso de implementación de un framework orientado a datos masivos. Investiga documentalmente en referencias bibliográficas y electrónicas sobre la arquitectura y el proceso de implementación del framework asignado por el docente, para conocer distintos ecosistemas de datos masivos Crea una presentación electrónica para exponerla frente al grupo.	Equipo de cómputo con conexión a internet, consulta de base de datos, libros electrónicos y lecturas proporcionadas por el docente.	2 horas
11	Integrar distintas fuentes de datos que cumplan con los requerimientos de datos masivos, a través del ecosistema previamente configurado y utilización de API's para la manipulación de los datos, con la finalidad de conocer su proceso de integración y uso, de forma analítica y crítica.	El docente brinda lecturas referentes a la integración de fuentes de datos y API's para la manipulación de datos masivos. Investiga documentalmente en referencias bibliográficas y electrónicas sobre el tema, para conocer distintos ecosistemas de datos masivos Crea una presentación electrónica para exponerla frente al grupo y entrega un reporte técnico.	Equipo de cómputo con conexión a internet, consulta de base de datos, libros electrónicos y lecturas proporcionadas por el docente.	2 horas
UNIDAD V				
12	Examinar los tipos de análisis que se pueden desarrollar en ecosistemas de datos masivos y tecnologías emergentes para este fin, a partir de la reflexión sobre los diferentes niveles de información que puede generar, para coadyuvar en la	El docente brinda lecturas referentes acerca de tipos de análisis en datos masivos. Investiga documentalmente en referencias bibliográficas y electrónicas sobre el tema, para conocer distintos tipos de análisis de datos	Equipo de cómputo con conexión a internet, consulta de base de datos, libros electrónicos y lecturas proporcionadas por el docente.	1 hora

	toma de decisiones, de forma analítica, crítica e inquisitiva.	Crea una presentación electrónica para exponerla frente al grupo utilizando herramientas para sintetizar información.		
13	Examinar bibliotecas de consulta y de minería de datos básica para la manipulación y análisis de los datos, a partir de una investigación documental, con la finalidad de experimentar y compartir el aprendizaje de diferentes frameworks, de forma analítica, colaborativa y comunicativa.	Investiga en lecturas proporcionadas por el docente, así como en otras referencias electrónicas acerca de bibliotecas de consulta y minería de datos masivos, Sintetiza la información y crea una presentación electrónica (powerpoint o afines) para exponerla frente al grupo, entregando un reporte técnico al finalizar la práctica.	Equipo de cómputo con conexión a internet, consulta de base de datos, libros electrónicos y lecturas proporcionadas por el docente.	4 horas
14	Construir modelos básicos de minería de datos, a partir del proceso KDD, para el análisis de los datos integrados al ecosistema de datos masivos previamente configurado, de forma analítica, crítica y colaborativa.	Construye modelos a partir del proceso KDD, bibliotecas de minería de datos y consulta, crear una presentación electrónica (powerpoint o afines) y exposición frente al grupo, así como un reporte técnico.	Equipo de cómputo con conexión a internet, consulta de base de datos, libros electrónicos y lecturas proporcionadas por el docente.	3 horas
15	Construir distintas presentaciones de modelos, mediante la aplicación de los fundamentos y herramientas de visualización a los modelos previamente creados, con la finalidad de conocer diversas formas de presentar la información obtenida de los modelos de minería de datos, de forma analítica, crítica y comunicativa.	Construye distintas presentaciones de modelos a partir de diversas herramientas de visualización, crear una presentación electrónica (powerpoint o afines) y exposición frente al grupo, así como un reporte técnico.	Equipo de cómputo con conexión a internet, consulta de base de datos, libros electrónicos y lecturas proporcionadas por el docente.	2 horas

16	Realizar inferencias, a partir de los datos procesados en modelos de minería de datos, para la generación de ideas y valor del contexto que se esté evaluando, de forma analítica, crítica e inquisitiva,	Fundamenta y explica las interpretaciones obtenidas de los datos analizados y resumidos a partir de los procesos anteriores, siendo esta última práctica de taller la última parte del proyecto final, crear una presentación electrónica (powerpoint o afines) y exposición frente al grupo, así como un reporte técnico, desde la arquitectura del framework hasta las inferencias o interpretaciones de los datos obtenidos de los modelos de minería de datos previamente construidos.	Equipo de cómputo con conexión a internet, consulta de base de datos, libros electrónicos y lecturas proporcionadas por el docente.	2 horas
----	---	--	---	---------

VII. MÉTODO DE TRABAJO

Encuadre: El primer día de clase el docente debe establecer la forma de trabajo, criterios de evaluación, calidad de los trabajos académicos, derechos y obligaciones docente-alumno.

Estrategia de enseñanza (docente):

Presentaciones para demostración de metodologías, técnicas y plataformas diversas, casos de estudio de referencia (benchmark) para aplicación del conocimiento adquirido y desarrollo de habilidades.

Estrategia de aprendizaje (alumno):

Investigar, presentar, interactuar con plataformas, redactar reportes técnicos y desarrollar actividades de talleres, así como un proyecto final como evidencia de desempeño que integre tanto conocimiento como habilidades adquiridas durante el curso.

III. CRITERIOS DE EVALUACIÓN

La evaluación será llevada a cabo de forma permanente durante el desarrollo de la unidad de aprendizaje de la siguiente manera:

Criterios de acreditación

- Para tener derecho a examen ordinario y extraordinario, el estudiante debe cumplir con los porcentajes de asistencia que establece el Estatuto Escolar vigente.
- Calificación en escala del 0 al 100, con un mínimo aprobatorio de 60.

Criterios de evaluación

- Exámenes escritos.....30%
- Prácticas del taller.....50%
- Evidencia de desempeño.....20%
(Proyecto final)

Total..... 100%

Nota: Será obligatorio obtener una calificación aprobatoria por cada criterio de evaluación para aprobar el curso.

IX. REFERENCIAS

Básicas	Complementarias
<p>Dean, J. (2014). <i>Big data, data mining, and machine learning: value creation for business leaders and practitioners</i>. Estados Unidos: John Wiley & Sons. [clásica].</p> <p>Hurwitz, J., Nugent, A., Halper, D. F., & Kaufman, M. (2013). <i>Big Data for Dummies</i>. Estados Unidos: John Wiley & Sons Inc. [clásica].</p> <p>Kleppman, M. (2017). <i>Designing Data- Intensive Applications</i>. Estados Unidos: O'Reilly Media</p> <p>Leskovec, J., Rajaraman, A., y Ullman, J. D. (2014). <i>Mining of massive datasets</i>. Inglaterra: Cambridge University Press. [clásica]</p> <p>Marz, N., & Warren, J. (2015). <i>Big Data: Principles and best practices of scalable real-time data systems</i>. Estados Unidos: Manning Publications Co.</p> <p>O'Reilly Radar Team. (2011). <i>Big Data Now: Current Perspectives from O'Reilly Radar</i>. O'Reilly Media, Incorporated. [clásica].</p> <p>Wessler, M. (2013). <i>Big data analytics for dummies</i>. New Jersey: John Wiley&Sons. [clásica].</p> <p>White, T. (2012). <i>Hadoop: The definitive guide</i>. O'Reilly Media, Incorporated [clásica].</p>	<p>Baesens, B. (2014). <i>Analytics in a big data world: The essential guide to data science and its applications</i>. Nueva Jersey: John Wiley & Sons. [clásica].</p> <p>Cavanillas, J. M., Curry, E., y Wahlster, W. (Eds.). (2016). <i>New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe</i>. doi: 10.1007/978-3-319-21569-3</p> <p>Furht, B., y Villanustre, F. (2016). <i>Big data technologies and applications</i> (pp. 978-3). doi:10.1007/978-3-319-44550-2</p> <p>Grus, J. (2015). <i>Data science from scratch: first principles with python</i>. O'Reilly Media</p> <p>Mayer-Schönberger, V., y Cukier, K. (2013). <i>Big data: la revolución de los datos masivos</i>. España: Turner Noema. [clásica]</p> <p>Ontiveros, E. (dir), y López Sabater, V. (coord). (14 de junio de 2018). <i>Economía de los Datos Riqueza 4.0</i>. [archivo de video]. Recuperado de: www.fundaciontelefonica.com/publicaciones</p>

X. PERFIL DEL DOCENTE

El docente que imparta esta asignatura deberá poseer título de licenciatura en Ingeniero en Computación o afín. Preferentemente posgrado en ciencia de datos o afín con experiencia en tecnologías orientadas a datos masivos, así como también experiencia mínima de dos años en docencia. Que desempeñe su labor con profesionalismo, compromiso y tolerancia.