
PROGRAMA DE ESTUDIOS: CURSO DE INTRODUCCIÓN A BIG DATA

KOICA

 **HANDONG GLOBAL
UNIVERSITY**



 **UNA**

Área: Big Data

Introducción a Big Data

□ Información básica

Información del curso

| | | |
|---|------------------|---|
| 1 | Título | <i>Introducción a Big Data</i> |
| 2 | Año | <i>2024</i> |
| 3 | Semestre | <i>Primero (Agosto a Octubre)</i> |
| 4 | Departamento | <i>Centro de Innovación TIC (FPUNA)</i> |
| 5 | Nivel | <i>Básico</i> |
| 6 | Formato de clase | <i>Online</i> |

Horario y Lugar

| | | |
|---|-----------|---|
| 1 | Días | <i>Lunes, Miércoles y Viernes</i> |
| 2 | Horario | <i>de 18:30 a 20:00 (Lunes y Miércoles) y 18:30 a 20:30 (Viernes)</i> |
| 3 | Ubicación | <i>Plataforma EDUCA</i> |

Información del instructor/a

| | | |
|---|---------------------|--|
| 1 | Nombre | <i>Prof. Ing. Richard Daniel Jiménez Riveros</i> |
| 2 | Oficina | <i>N/A</i> |
| 3 | Contacto (correo) | <i>rjimenez@pol.una.py</i> |
| 4 | Contacto (teléfono) | <i>(0982) 136-081</i> |

Perfil profesional

Egresado de la Facultad Politécnica de la Universidad Nacional de Asunción (FP-UNA). Ha realizado un curso de Capacitación en Didáctica de la Educación Superior; Diplomado en Big Data; Diplomado en Inteligencia Artificial aplicados a Productos y Servicios; y cursos de tecnologías en la nube para Big Data.

Desde el 2023 se desempeña como Docente Encargado de Cátedra del Departamento de Enseñanza de Informática de la Facultad Politécnica de la Universidad Nacional de Asunción. Y, actualmente forma parte de la firma KONECTA S.A. habiéndose desempeñado como: Líder de Proyectos; Analista Funcional; Ingeniero de Datos; I+D y Capacitador, en proyectos de TI para el sector público y privado. Además, tiene un puesto como ANALISTA-REPORTING en el departamento de SISTEMAS BIG DATA & SVA de la empresa Núcleo S.A., Personal de Paraguay.



Se desempeñó como Auxiliar de Enseñanza del Departamento de Enseñanza Informática de la Facultad Politécnica de la Universidad Nacional de Asunción; sede Campus San Lorenzo, en los periodos académicos: 2018, 2019, 2020, 2021 y 2022, en las asignaturas de Algoritmos, Lenguaje de Programación, Base de Datos y Data Mining.

También, ha participado como programador web full stack en proyectos del Clúster de Investigación Aplicada (CIA-FPUNA) y otros proyectos de TI del sector público y privado.

Cuenta con una publicación científica sobre algoritmos de Optimización basados en simetrías de las permutaciones, en 35th International Conference of the Chilean Computer Science (SCCC), y participación en la XLII Conferencia Latinoamericana de Informática (CLEI) del 2016.

Notas adicionales

- ✓ Previa coordinación con el instructor los estudiantes podrán solicitar sesiones de consultas en horario de oficina de lunes a jueves.

□ Prerrequisitos

Ninguno.

□ Descripción del curso

Introducción a Big Data es un curso fundamental diseñado para presentar a los estudiantes el vasto y en evolución campo del análisis de big data. En una era en la que la generación de datos es ubicua y exponencial, la capacidad de gestionar, procesar y extraer valor de forma eficaz a partir de grandes conjuntos de datos es invaluable en todos los sectores. Este curso ofrece una inmersión profunda en los conceptos fundamentales de big data, incluida su definición, características y los desafíos que presenta. A través de una combinación de conocimientos teóricos y aplicaciones prácticas, los estudiantes explorarán cómo los big data están transformando las industrias, impulsando la innovación y dando forma a nuestra comprensión del mundo.

El plan de estudios está estructurado para proporcionar una descripción general completa de las tecnologías y metodologías de big data. Comenzando con los conceptos básicos de almacenamiento y gestión de datos, los estudiantes aprenderán sobre marcos informáticos distribuidos como Hadoop y Spark, que permiten el procesamiento eficiente de grandes conjuntos de datos. El curso cubrirá aspectos clave del análisis de datos, incluida la minería de datos, algoritmos de aprendizaje automático y métodos estadísticos diseñados para entornos de big data. Se pone énfasis en las habilidades prácticas, y los estudiantes participan en ejercicios y proyectos prácticos que simulan escenarios de análisis de datos del mundo real.

Una característica importante de este curso es su enfoque en las herramientas y plataformas que son fundamentales en el ecosistema de big data. Los estudiantes adquirirán competencia en el uso de bases de datos NoSQL para almacenamiento escalable, Apache Spark para procesamiento de big data y herramientas de visualización de datos para comunicar conocimientos. Estas habilidades son fundamentales para afrontar los desafíos técnicos de big data y aprovechar su potencial para obtener conocimientos prácticos. El curso brindará a los estudiantes acceso a estas tecnologías, ofreciendo un entorno de pruebas para la exploración y el aprendizaje.

Más allá de las habilidades técnicas, el curso aborda las consideraciones éticas y de privacidad inherentes al análisis de big data. Los estudiantes participarán en discusiones y estudios de casos que resaltan las implicaciones sociales de la recopilación y el análisis de datos, incluidas cuestiones de seguridad de datos, derechos de privacidad y uso ético de los datos. Este enfoque holístico garantiza que los graduados no sólo sean técnicamente competentes sino también éticamente conscientes, preparados para navegar las complejidades de la era digital con integridad.

"Introducción al Big Data" está dirigido a estudiantes y profesionales deseosos de profundizar en el campo de la ciencia de datos, así como a cualquier persona interesada en comprender el impacto del big data en la sociedad y los negocios. Al final del curso, los participantes tendrán una base sólida en los principios y técnicas de big data, estarán equipados con las habilidades para analizar grandes conjuntos de datos y contribuir a los procesos de toma de decisiones basados en datos. Este curso allana el camino para estudios adicionales en ciencia y análisis de datos, ofreciendo un trampolín hacia temas avanzados y áreas especializadas dentro del campo.

□ Objetivos del curso

Al completar con éxito este curso, los estudiantes podrán:

1. Comprender los conceptos clave y los desafíos del Big Data.
2. Identificar y aplicar tecnologías big data apropiadas para el análisis y procesamiento de datos.
3. Utilizar herramientas de análisis de big data para analizar grandes conjuntos de datos.
4. Desarrollar y ejecutar proyectos de big data desde la recopilación de datos hasta el conocimiento.
5. Evaluar las consideraciones éticas y de privacidad en el análisis de big data.

1. Comprender los conceptos clave y los desafíos de Big Data: los estudiantes obtendrán una comprensión profunda de lo que constituye Big Data, incluidas sus características definitorias, como volumen, velocidad, variedad y veracidad. Explorarán las complejidades y los desafíos asociados con la gestión de big data, como los problemas de almacenamiento, el procesamiento de datos y la gestión de la calidad de los datos. Esta base permitirá a los estudiantes apreciar el alcance de los big data y sus implicaciones en diversos dominios.

2. Identificar y aplicar tecnologías de Big Data apropiadas para el análisis y procesamiento de datos: este objetivo se centra en equipar a los estudiantes con el conocimiento y las habilidades para seleccionar e implementar las tecnologías adecuadas para tareas específicas de Big Data. Los estudiantes aprenderán sobre marcos informáticos distribuidos (por ejemplo, Hadoop y Spark), soluciones de almacenamiento de big data (por ejemplo, bases de datos NoSQL) y modelos de procesamiento de datos. Comprenderán las fortalezas y limitaciones de cada tecnología, lo que les permitirá tomar decisiones informadas basadas en los requisitos de sus proyectos.

3. Utilizar herramientas de análisis de Big Data para analizar grandes conjuntos de datos: los estudiantes recibirán capacitación para utilizar una variedad de herramientas y plataformas para el análisis de datos, incluidas herramientas de visualización de datos, software estadístico y bibliotecas de aprendizaje automático. Este objetivo enfatiza la experiencia práctica, y los estudiantes realizan ejercicios prácticos para aplicar estas herramientas en el análisis de conjuntos de datos del mundo real. El objetivo es desarrollar competencias para extraer información útil a partir de conjuntos de datos complejos, una habilidad muy valorada en diversas industrias.

4. Desarrollar y ejecutar proyectos de Big Data desde la recopilación de datos hasta las ideas: este objetivo integral tiene como objetivo guiar a los estudiantes a través del ciclo de vida completo de un proyecto de Big Data. Desde la recopilación inicial de datos y el preprocesamiento hasta el análisis y la obtención de conocimientos, los estudiantes aprenderán a gestionar proyectos de forma eficaz. Esto incluye formular preguntas de investigación, seleccionar metodologías apropiadas, realizar análisis e interpretar los resultados. El trabajo del proyecto fomentará las habilidades de resolución de problemas, la creatividad y la capacidad de comunicar información compleja con claridad.

5. Evaluar las consideraciones éticas y de privacidad en el análisis de Big Data: un componente esencial del curso, este objetivo aborda el panorama ético de Big Data. Los estudiantes explorarán los dilemas éticos que plantea la recopilación y el análisis de datos a gran escala, incluidas las preocupaciones sobre la privacidad, la seguridad de los datos y el potencial de sesgo en las decisiones basadas en datos. Las discusiones y los estudios de casos fomentarán el pensamiento crítico sobre las responsabilidades de los científicos de datos y el impacto de su trabajo en los individuos y la sociedad.

Al lograr estos objetivos, los estudiantes saldrán del curso con una comprensión completa de big data, equipados con las habilidades técnicas para manejar herramientas y tecnologías de big data, las habilidades de gestión de proyectos para supervisar iniciativas de big data y el marco ético para navegar las complejas cuestiones que rodean la privacidad y la seguridad de los datos. Esta fundación los preparará para estudios adicionales en ciencia de datos o carreras en industrias donde el big data juega un papel fundamental.

□ Método de evaluación

La calificación del curso se estructura de la siguiente manera para evaluar la comprensión, el compromiso y la aplicación práctica del material del curso por parte de los estudiantes:

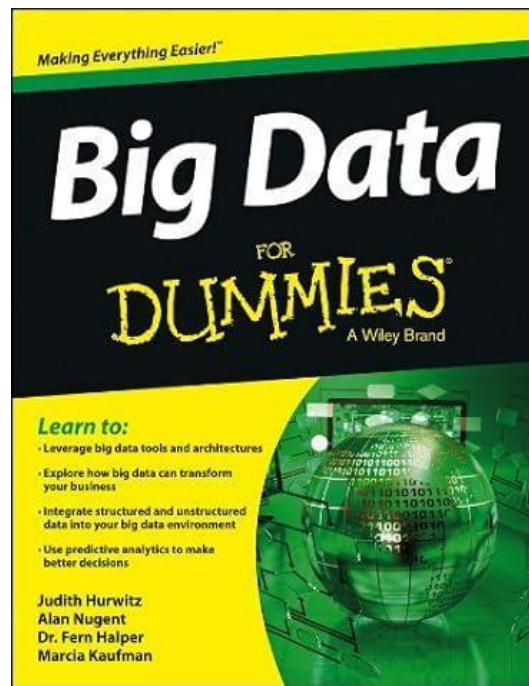
| Componentes de la nota final | Porcentaje de la nota final | Escala | Calificación (*) |
|------------------------------|-----------------------------|------------|------------------|
| Asistencia a las clases. | 20 % | 0% – 70% | NO APROBADO |
| Tareas. | 50 % | | |
| Examen Parcial. | 15 % | 71% – 100% | APROBADO |
| Examen / Proyecto Final. | 15 % | | |
| Puntaje, total | 100 % | | |

(*) Requisito general para expedición de certificados: aprobar con > 70%.

□ Libros de texto y otros materiales

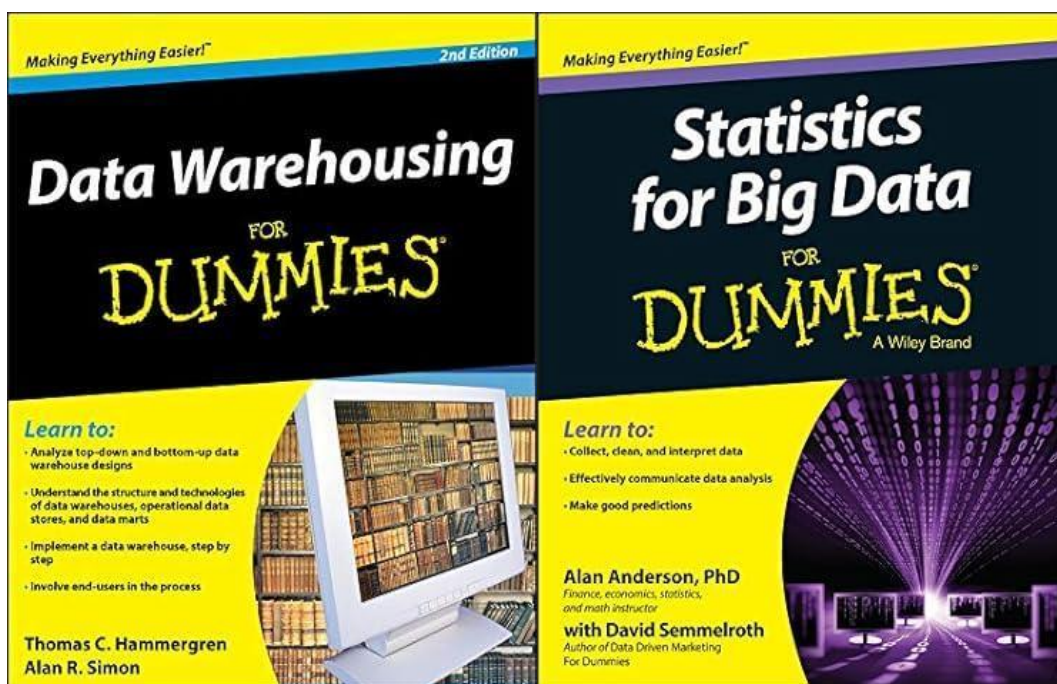
Principal:

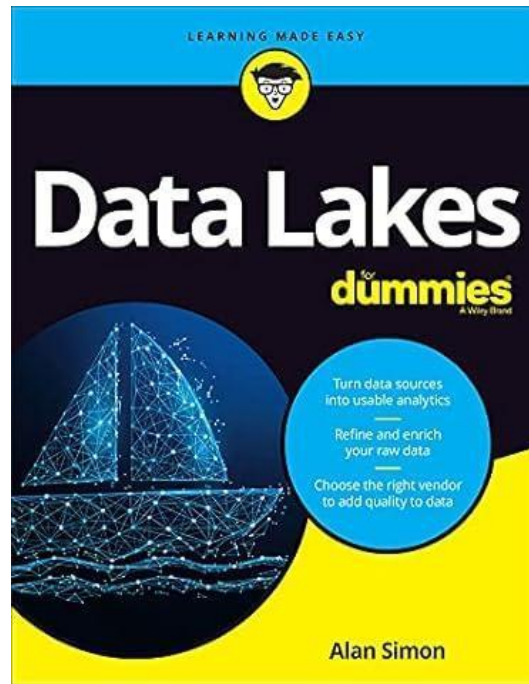
- ✓ Hurwitz, J., Nugent, A., Halper, F., & Kaufman, M. (2013). *Big Data for Dummies*. John Wiley & Sons, Inc. ISBN 978-1-118-50422-2.
- ✓ Marz, N., & Warren, J. (2015). *Big Data: Principles and Best Practices of Scalable Realtime Data Systems*. Manning.
- ✓ Ryza, S., Laserson, U., Owen, S., & Wills, J. (2017). *Advanced analytics with spark: patterns for learning from data at scale*. O'Reilly Media, Inc.
- ✓ Lin, J., & Dyer, C. (2010). *Data-Intensive Text Processing with MapReduce*. Morgan & Claypool Publishers.
- ✓ Apache Software Foundation (jun. de 2019). Apache Hadoop. Disponible en <https://hadoop.apache.org/>
- ✓ Dean, J. & Ghemawat, S. (2004). MapReduce: Simplified Data Processing on Large Clusters. Goole, Inc. Disponible en <http://research.google.com/archive/mapreduce.html>
- ✓ Lightning-fast unified analytics engine (jun. de 2019). Apache Spark TM. Disponible en <http://spark.apache.org/>
- ✓ Python TM. Programming Language. (jun. de 2019). Disponible en <https://www.python.org/>



Complementaria:

- ✓ Hammergren, Thomas C.. Data Warehousing For Dummies . Wiley.
- ✓ Simon, Alan R.. Data Lakes For Dummies (English Edition). Wiley.
- ✓ Anderson, Alan. Statistics for Big Data For Dummies (English Edition). Wiley.
- ✓ Mohanty, S., Jagadeesh, M., & Srivatsa, H. (2013). *Big Data Imperatives: Enterprise Big Data Warehouse, BI Implementations and Analytics*. Apress. ISBN 978-1430248729.
- ✓ Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt.
- ✓ Daza Vergaray, A. (2017). *Data mining: Minería de datos*. Lima: Macro EIRL.
- ✓ Larose, D. T. & Larose, C. D. (2018). *Data mining and predictive analytics*. (2° Ed.). Nueva Delhi: Wiley India.





Otros recursos:

- *Sitio Oficial de MongoDB Atlas:*
 - <https://www.mongodb.com/es>
- *Curso Introducción a MongoDB Atlas by MongoDB Inc.:*
 - <https://www.coursera.org/learn/introduction-to-mongodb/home/welcome>
- *Sitio Oficial de PostgreSQL:*
 - <https://www.postgresql.org/>
- *Latest Pentaho Data Integration (aka Kettle) Documentation:*
 - <https://pentaho-public.atlassian.net/wiki/spaces/EAI/overview?homepageId=363267360>
- *Kaggle Datasets:*
 - <https://www.kaggle.com/>
- *Opendata Paraguay:*
 - <https://www.datos.gov.py/>

□ Tareas y exámenes

Tareas

Objetivo: Las tareas están diseñadas para reforzar los conceptos discutidos en conferencias y lecturas, brindando a los estudiantes experiencia práctica en análisis de datos, uso de herramientas de big data y habilidades de resolución de problemas.

Estructura:

- Conjuntos de problemas semanales: cubriendo aspectos teóricos y prácticos de los temas semanales. Estos pueden incluir tareas de manipulación de datos, ensayos breves sobre los desafíos de big data y ejercicios con herramientas de big data como Hadoop o Spark.
- Tareas de análisis de datos: los estudiantes recibirán conjuntos de datos para limpiar, analizar y obtener información, utilizando herramientas como la biblioteca pandas de Python, Apache Spark o consultas SQL en bases de datos NoSQL.
- Evaluación: Las tareas se calificarán en función de la precisión, la creatividad en la resolución de problemas y la capacidad de aplicar conocimientos teóricos a tareas prácticas. Los comentarios se centrarán en mejorar la comprensión y la aplicación de las tecnologías de big data.

Exámenes parciales y finales

Objetivo: Los exámenes evalúan la comprensión del material del curso por parte de los estudiantes, incluida la comprensión de conceptos clave, la capacidad para analizar problemas de big data y el conocimiento de las tecnologías y herramientas apropiadas para abordar estos problemas.

Estructura:

- Examen Parcial: una combinación de preguntas de opción múltiple, respuesta corta y resolución de problemas que cubren la primera mitad del curso. Esto puede incluir preguntas sobre los fundamentos de big data, análisis de datos básicos y uso introductorio de herramientas de big data.
- Examen Final: es comprensivo, abarca todo el curso. Incluye una combinación de tipos de preguntas, con un enfoque en la integración del conocimiento en diferentes temas, como la elección de tecnologías apropiadas para escenarios de big data determinados y consideraciones éticas en el análisis de datos.
- Evaluación: los exámenes se calificarán según la profundidad de comprensión, la precisión de las respuestas y la capacidad de integrar y aplicar los conceptos del curso a situaciones nuevas.

Proyecto final

Objetivo: El proyecto final permite a los estudiantes aplicar lo que han aprendido en un escenario integral del mundo real. Los proyectos implicarán la recopilación, el procesamiento, el análisis y la presentación de resultados de datos, haciendo hincapié en el desarrollo de una solución completa de big data.

Estructura:

- Propuesta: los estudiantes envían una propuesta de proyecto que describe el problema elegido, las fuentes de datos, el plan de análisis y los resultados esperados.
- Informes de Progreso: actualizaciones periódicas para realizar un seguimiento del progreso, los desafíos encontrados y los resultados preliminares.

- **Presentación:** una presentación formal del proyecto a la clase, destacando el problema, la metodología, el análisis y los conocimientos derivados.
- **Informe Final:** un informe detallado que incluye antecedentes de investigación, metodología, análisis, resultados y conclusiones. El informe también debe reflejar el proceso del proyecto y cualquier consideración ética abordada.
- **Evaluación:** los proyectos se evaluarán en función de su originalidad, complejidad, aplicación de técnicas de big data, profundidad del análisis, calidad de los conocimientos y presentación general. Se pondrá especial énfasis en la aplicación práctica de herramientas y tecnologías, así como en la capacidad de superar desafíos típicos de proyectos de big data.

La incorporación de estos componentes en el curso garantiza una evaluación equilibrada de los conocimientos teóricos y las habilidades prácticas de los estudiantes, preparándolos para los desafíos del mundo real en el campo de big data.

□ **Actividades del curso**

Incorporar actividades diversas y atractivas en un curso de "Introducción a Big Data" es esencial para mejorar el aprendizaje, fomentar la participación y brindar experiencia práctica. A continuación, se presentan descripciones detalladas de diversas actividades del curso diseñadas para alcanzar estos objetivos:

1. Clases Interactivas

Objetivo: Introducir y explicar conceptos, tecnologías y desafíos centrales en big data. Las conferencias están diseñadas para ser interactivas, fomentando preguntas, debates y comentarios en tiempo real.

2. Laboratorios Prácticos

Objetivo: Proporcionar experiencia práctica con herramientas y técnicas de big data. Los laboratorios están estructurados para complementar los temas de las conferencias, lo que permite a los estudiantes aplicar lo que han aprendido en un entorno controlado.

3. Proyectos Grupales

Objetivo: Desarrollar el trabajo en equipo, la gestión de proyectos y las habilidades técnicas trabajando en un proyecto integral de big data desde su concepción hasta su presentación.

□ Cronograma del curso

| Semana | Unidades | Tipo de clases | Materiales |
|--------|--|---------------------------|--|
| 1 | Introducción al Big Data | <i>Clase síncronas</i> | Fuentes bibliográficas Presentaciones |
| 2 | Business Intelligence (BI) y Data Warehousing (DW) | <i>Clase síncronas</i> | Fuentes bibliográficas Presentaciones |
| 3 | Fundamentos tecnológicos para el Big Data | <i>Clase síncronas</i> | Fuentes bibliográficas Presentaciones |
| 4 | Gestión de Big Data | <i>Clase síncronas</i> | Fuentes bibliográficas Presentaciones |
| 4 | Examen Parcial | <i>Evaluación parcial</i> | Formulario con 15 preguntas de selección múltiples. Contenidos desarrollados en la semana 1, 2, 3 y 4. |
| 5 | Analítica y Big Data | <i>Clase síncronas</i> | Fuentes bibliográficas Presentaciones |
| 6 | Introducción a los Data Lakes | <i>Clase síncronas</i> | Fuentes bibliográficas Presentaciones |
| 7 | Implementación de Big Data | <i>Clase síncronas</i> | Fuentes bibliográficas Presentaciones |
| 8 | Introducción a MongoDB Cloud | <i>Clase síncronas</i> | Fuentes bibliográficas Presentaciones |
| 8 | Examen final (Proyecto Final) | <i>Evaluación final</i> | Aplicar los conceptos aprendidos durante el curso en un caso práctico básico. |

□ Contenidos del curso

El desarrollo de los contenidos del programa de estudio se basa en la bibliografía principal “*Big Data for dummies*”, y se complementan con las siguientes fuentes bibliográficas: “*Data Warehousing for dummies*”; “*Data Lakes for dummies*” y “*Statics for Big Data for dummies*”.

Curso básico: Introducción a Big Data.

Duración: 8 semanas

Horas semanales: 5 horas.

Total de horas disponibles: 40 horas.

Clases teóricas: 30 horas.

Clases prácticas: 10 horas.

| | | |
|--|--|--|
| Semana 1: Introducción al Big Data. | <u>Clase 1</u> Síncrona y teórica. (1,5 hs) | Tema 1: Introducción a los fundamentos de Big Data. <ul style="list-style-type: none"> ● La evolución de la gestión de datos. ● La importancia de la gestión de datos. <ul style="list-style-type: none"> ○ Creación de estructuras de datos manejables. ○ Gestión web y de contenidos. ○ Gestión de grandes volúmenes de datos. ● Definición de Big Data. ● Cómo crear una arquitectura de gestión de Big Data exitosa. <ul style="list-style-type: none"> ○ Comenzando con capturar, organizar, integrar, analizar y actuar sobre los datos. ○ Estableciendo las bases arquitectónicas. ○ El rendimiento importa. ○ Análisis tradicional y avanzado. |
| | <u>Clase 2</u> Síncrona y teórica. (1,5 hs) | Tema 2: Examinando los tipos de Big Data. <ul style="list-style-type: none"> ● Definición de datos estructurados. ● Definición de datos no estructurados. ● Analizando los requisitos en tiempo real y no en tiempo real. ● Gestión de diferentes tipos de datos. ● Integración de tipos de datos en un entorno de Big Data. |
| | <u>Clase 3</u> Síncrona y teórica / práctica. (2 hs) | Tema 3: Introducción a la computación distribuida. <ul style="list-style-type: none"> ● Una breve historia de la computación distribuida. ● Comprender los conceptos básicos de la computación distribuida. <ul style="list-style-type: none"> ○ ¿Por qué necesitamos computación distribuida para Big Data? ○ La cambiante economía de la informática. ○ El problema de la latencia. ● Obtener el rendimiento correcto. |
| Semana 2: Business Intelligence (BI) y Data Warehousing (DW). | <u>Clase 4</u> Síncrona y teórica. (1,5 hs) | TAREA 1 (30%): Integración de Datos con Pentaho Kettle. <ul style="list-style-type: none"> ● Utilizar el conjunto de datos de la nómina de funcionarios públicos de la SFP disponible en la web para la extracción, transformación y carga en un modelo Entidad-Relacional para un sistema de bases de datos tradicional. |
| | | Tema 4: Fundamentos de Business Intelligence y Data Warehousing. <ul style="list-style-type: none"> ● Introducción a Business Intelligence (BI). ● ¿Qué es un Data Warehouse? ● Clasificación del Data Warehouse. ● ¿Qué es un Data Mart? ● Enfoques arquitectónicos para los Data Marts ● ¿Data mart o data warehouse? |

| | | |
|---|---|---|
| | | <ul style="list-style-type: none"> ● Modelado de datos multidimensional. ● Aplicaciones y casos de uso. |
| | <p><u>Clase 5</u> Síncrona y teórica. (1,5 hs)</p> | <p>Tema 5: Arquitecturas para Business Intelligence y Data Warehousing.</p> <ul style="list-style-type: none"> ● Principales categorías de la inteligencia de negocio. <ul style="list-style-type: none"> ○ Querying and reporting. ○ Business Analysis (OLAP) ○ Data mining. ○ Dashboards and Scorecards. ● Tecnologías para Data Warehousing. ● Pentaho Kettle para la Integración de Datos (ETL). <ul style="list-style-type: none"> ○ Extracción de datos. ○ Limpieza y transformación de datos. ○ Carga y reducción de datos. |
| | <p><u>Clase 6</u> Síncrona y práctica. (2 hs)</p> | <p>Tema 6: Bases de datos relaciones y Data Warehousing.</p> <ul style="list-style-type: none"> ● Instalación y Configuración de PostgreSQL. ● Instalación y Configuración de Pentaho Data Integration. ● Explicación del caso práctico. ● Implementación de un caso práctico de Data Warehouse con Pentaho Data Integration y PostgreSQL. |
| <p>Semana 3: Fundamentos tecnológicos para el Big Data.</p> | <p><u>Clase 7</u> Síncrona y teórica. (1,5 hs)</p> | <p>Tema 7: Profundizando en los componentes tecnológicos de Big Data.</p> <ul style="list-style-type: none"> ● Explorando la pila de Big Data. ● Capa 0: Infraestructura física redundante. ● Capa 1: Infraestructura de seguridad. ● Capa 2: Bases de datos operativas. ● Capa 3: Organización de servicios y herramientas de datos. ● Capa 4: Almacenes de datos analíticos. |
| | <p><u>Clase 8</u> Síncrona y teórica. (1,5 hs)</p> | <p>Tema 8: Virtualización y cómo soporta la computación distribuida.</p> <ul style="list-style-type: none"> ● Comprender los conceptos básicos de la virtualización. <ul style="list-style-type: none"> ○ La importancia de la virtualización para el Big Data. ○ Virtualización de servidores. ○ Virtualización de aplicaciones. ○ Virtualización de red. ○ Virtualización de procesadores y memoria. ○ Virtualización de datos y almacenamiento. |
| | <p><u>Clase 9</u> Síncrona y teórica. (2 hs)</p> | <p>Tema 9: Examinando la nube y el Big Data.</p> <ul style="list-style-type: none"> ● Definición de la nube en el contexto del Big Data. ● Comprensión de los modelos de implementación y distribución en la nube. ● La nube como imperativo para el Big Data. ● Proveedores en el mercado de Big Data Cloud. |
| <p>Semana 4: Gestión de Big Data.</p> | <p><u>Clase 10</u> Síncrona y teórica. (1,5 hs)</p> | <p>Tema 10: Bases de Datos Operacionales.</p> <ul style="list-style-type: none"> ● Los SGBDR (RDBMS) son importantes en un entorno de Big Data. <ul style="list-style-type: none"> ○ Base de datos relacional PostgreSQL. ● Bases de datos no relacionales. ● Bases de datos de pares clave-valor. ● Bases de datos de documentos. <ul style="list-style-type: none"> ○ MongoDB. ● Bases de datos en columnas. <ul style="list-style-type: none"> ○ HBase. ● Bases de datos de grafos. <ul style="list-style-type: none"> ○ Neo4J. ● Bases de datos espaciales. <ul style="list-style-type: none"> ○ PostGIS/OpenGEO Suite. |

| | | |
|---|--|---|
| | | <p>TAREA 2 (10%): Investigar y escribir un ensayo sobre el tema "Ética en Big Data".</p> <ul style="list-style-type: none"> • Consideraciones éticas en la recopilación, almacenamiento y análisis de grandes volúmenes de datos. |
| | <p><u>Clase 11</u> Síncrona y teórica. (1,5 hs)</p> | <p>Tema 11: Fundamentos de MapReduce.</p> <ul style="list-style-type: none"> • Antecedentes históricos sobre los orígenes de MapReduce. • Entendiendo la función de mapeo. • Añadiendo la función de reducción. • ¿Cómo trabajan juntos las funciones de mapeo y reducción? • Optimización de tareas de MapReduce. <ul style="list-style-type: none"> ○ Topología de Hardware/network. ○ Sincronización. ○ Sistema de archivos. |
| | <p><u>Clase 12</u> Síncrona y teórica. Examen Parcial (2 hs)</p> | <p>Tema 12: Explorando el mundo de Hadoop.</p> <ul style="list-style-type: none"> • Explorando Hadoop. • Comprendiendo Hadoop Distributed File System (HDFS) • Introducción a Hadoop MapReduce. <p>EXAMEN PARCIAL (15%):</p> <ul style="list-style-type: none"> • Formulario con 15 preguntas de selección múltiples. Contenidos desarrollados en la semana 1, 2, 3 y 4. |
| <p>Semana 5: Analítica y Big Data.</p> | <p><u>Clase 13</u> Síncrona y teórica. (1,5 hs)</p> | <p>Tema 13: Dispositivos y almacenes de Big Data.</p> <ul style="list-style-type: none"> • Integración de Big Data con el almacén de datos tradicional. <ul style="list-style-type: none"> ○ Optimización del almacén de datos. ○ Diferenciar las estructuras de big data de los datos del almacén de datos. • Análisis de Big Data y el almacén de datos. <ul style="list-style-type: none"> ○ Repensando la extracción, la transformación y la carga. • Cambiando el rol del almacén de datos. • Cambios en los modelos de implementación en la era del Big Data. <ul style="list-style-type: none"> ○ El modelo en la nube. • Analizando el futuro de los almacenes de datos. |
| | <p><u>Clase 14</u> Síncrona y teórica. (1,5 hs)</p> | <p>Tema 14: Definición de análisis de Big Data.</p> <ul style="list-style-type: none"> • Utilizando Big Data para obtener resultados. <ul style="list-style-type: none"> ○ Analítica básica. ○ Analítica avanzada. ○ Análisis operativo. ○ Análisis de monetización. • Modificación de productos de Business Intelligence para manejar Big Data. <ul style="list-style-type: none"> ○ Datos. ○ Algoritmos analíticos. ○ Soporte de infraestructura. • Estudiando ejemplos de análisis de Big Data. • Soluciones de análisis de Big Data. |
| | <p><u>Clase 15</u> Síncrona y práctica. (2 hs)</p> | <p>Tema 15: Introducción a la visualización de datos con Tableau.</p> <ul style="list-style-type: none"> • Primeros Pasos con Tableau. <ul style="list-style-type: none"> ○ Instalación de Tableau. ○ Conexión a fuentes de datos. • Principales elementos de una visualización. <ul style="list-style-type: none"> ○ Medidas y dimensiones. ○ Marcas: Color, Tamaño, Etiquetas, Descripción Emergentes. ○ Dando formato a nuestras medidas. ○ Filtros. • Cálculos en Tableau. |

| | | |
|--|---|---|
| | | <ul style="list-style-type: none"> ○ Cálculos de Tabla. ○ Campos Calculados. ○ Agrupaciones. ○ Conjuntos. ○ Parámetros. ○ Agregación en los campos calculados. ● Combinando información de diferentes fuentes de datos. ● Ejemplos básicos de Dashboards. |
| | | <p>Proyecto Final: Presentación para iniciar su elaboración en equipo.</p> <ul style="list-style-type: none"> ● Implementación y Casos de Estudio. <ul style="list-style-type: none"> ○ Desarrollo de un Proyecto de Big Data. ○ Definición del problema. ○ Selección de tecnologías. ○ Implementación y despliegue. |
| <p>Semana 6: Introducción a los Data Lakes.</p> | <p><u>Clase 16</u> Síncrona y teórica. (1,5 hs)</p> | <p>Tema 16: Data Lakes y Big Data.</p> <ul style="list-style-type: none"> ● ¿Qué es un Data Lake? ● Data Lakes y Big Data. ● Diferencias entre un Data Warehouse y un Data Lake. <ul style="list-style-type: none"> ○ Tipos de datos admitidos. ○ Volúmenes de datos. ○ Diferentes modelos de datos internos. ○ Arquitectura y topología. ○ Comparación entre ETL y ELT. ○ Latencia de datos. ○ Usos analíticos. ○ Incorporación de nuevas fuentes de datos. ○ Comunidades de usuarios. ○ Hosting. |
| | | <p>TAREA 3 (10%): Investigar y escribir un ensayo sobre el tema "Futuro del Big Data".</p> <ul style="list-style-type: none"> ● Tendencias emergentes y el futuro del Big Data en la tecnología y los negocios. |
| | | <p>Proyecto Final: Consultas y Tutoría a los equipos de trabajo.</p> |
| | <p><u>Clase 17</u> Síncrona y teórica. (1,5 hs)</p> | <p>Tema 17: Construyendo un Data Lake en Google Cloud.</p> <ul style="list-style-type: none"> ● Introducción a los Data Lakes en Google Cloud Platform (GCP). ● Opciones de Data Storage y ETL en GCP. ● Construyendo un Data Lake mediante Cloud Storage. ● Seguridad de Cloud Storage. ● Guardar todo tipo de tipos de datos. ● Cloud SQL como un Data Lake relacional. |
| | | <p>Proyecto Final: Consultas y Tutoría a los equipos de trabajo.</p> |
| | <p><u>Clase 18</u> Síncrona y teórica. (2 hs)</p> | <p>Tema 18: Mejores prácticas en la preparación de datos.</p> <ul style="list-style-type: none"> ● Comprobar formatos de datos. ● Verificar tipos de datos. ● Grafica tus datos. ● Verificar la precisión de los datos. ● Identificar valores atípicos (Outliers) ● Cómo lidiar con los valores faltantes (Missing Values) ● Verifique sus suposiciones sobre cómo se distribuyen los datos. ● Realice copias de seguridad y documente todo lo que hace. |
| | | <p>Proyecto Final: Consultas y Tutoría a los equipos de trabajo.</p> |
| <p>Semana 7: Implementación de Big Data.</p> | <p><u>Clase 19</u> Síncrona y teórica. (1,5 hs)</p> | <p>Tema 19: Integración de fuentes de datos.</p> <ul style="list-style-type: none"> ● Identificación de los datos que necesita. <ul style="list-style-type: none"> ○ Etapa exploratoria. ○ Etapa de codificación. ○ Etapa de integración e incorporación. |

| | | |
|--|---|--|
| | | <ul style="list-style-type: none"> ● Comprendiendo los fundamentos de la integración de Big Data. ● Definición de ETL tradicional. <ul style="list-style-type: none"> ○ Transformación de datos. ● Comprendiendo los Procesos de ELT: extracción, carga y transformación. ● Priorizando la calidad del Big Data. ● Mejores prácticas para la integración de datos en un mundo de big data. |
| | <u>Clase 20</u> Síncrona y teórica. (1,5 hs) | <p>Proyecto Final: Consultas y Tutoría a los equipos de trabajo.</p> <p>Tema 20: Aplicación de Big Data en su organización.</p> <ul style="list-style-type: none"> ● Calculando la economía de Big Data. <ul style="list-style-type: none"> ○ Identificación de tipos y fuentes de datos. ○ Modificaciones de procesos de negocio o creación de nuevos procesos. ○ El impacto tecnológico de los flujos de trabajo de Big Data. ○ Encontrar el talento para apoyar proyectos de Big Data. ● Gestión de datos empresariales y Big Data. <ul style="list-style-type: none"> ○ Definición de gestión de datos empresariales. ● Creación de una hoja de ruta para la implementación de Big Data. <ul style="list-style-type: none"> ○ Entendiendo la urgencia empresarial. ○ Proyectar la cantidad adecuada de capacidad. ○ Cómo seleccionar la metodología de desarrollo de software adecuada. ○ Equilibrar presupuestos y conjuntos de habilidades. ○ Identificación de riesgos. |
| | <u>Clase 21</u> Síncrona y teórica. (2 hs) | <p>Proyecto Final: Consultas y Tutoría a los equipos de trabajo.</p> <p>Tema 21: Seguridad y gobernanza para entornos de Big Data.</p> <ul style="list-style-type: none"> ● Seguridad en el contexto del Big Data. <ul style="list-style-type: none"> ○ Evaluación del riesgo para el negocio. ○ Los riesgos que acechan en el Big Data. ● Comprender las opciones de protección de datos. ● El desafío de la gobernanza de datos. <ul style="list-style-type: none"> ○ Auditando su proceso de Big Data. ○ Identificación de las partes interesadas clave. ● Cómo establecer la estructura organizativa adecuada. <ul style="list-style-type: none"> ○ Preparación para la gestión y administración de riesgos. ○ Establecer políticas adecuadas de gobernanza y calidad. |
| Semana 8: Introducción a MongoDB Cloud. | <u>Clase 22</u> Síncrona y práctica. (1,5 hs) | <p>Proyecto Final: Consultas y Tutoría a los equipos de trabajo.</p> <p>Tema 22: Introducción a MongoDB Atlas (DBSaaS)</p> <ul style="list-style-type: none"> ● Introducción a MongoDB. ● El modelo documental de MongoDB. ● Creación y Despliegue de un clúster Atlas. ● Gestión de bases de datos, colecciones y documentos en el explorador de datos Atlas. ● Conexión a una base de datos MongoDB. |
| | <u>Clase 23</u> Síncrona y práctica. | <p>Proyecto Final: Consultas y Tutoría a los equipos de trabajo.</p> <p>Tema 23: Operaciones CRUD de MongoDB: Insertar y Buscar Documentos.</p> <ul style="list-style-type: none"> ● Inserción de documentos en una colección MongoDB. ● Encontrar documentos en una colección MongoDB. ● Búsqueda de documentos mediante operadores de comparación. ● Consultas sobre elementos de matrices en MongoDB. ● Búsqueda de documentos mediante operadores lógicos. <p>Proyecto Final: Consultas y Tutoría a los equipos de trabajo.</p> |

| | | |
|--|---|--|
| | <p><u>Clase 24</u> Síncrona y práctica.</p> | <p>Tema 24: Operaciones CRUD de MongoDB: Reemplazar y Eliminar Documentos.</p> <ul style="list-style-type: none">● Sustitución de un documento en MongoDB.● Actualización de documentos MongoDB mediante updateOne()● Actualización de documentos MongoDB mediante findAndModify()● Actualización de documentos MongoDB mediante updateMany()● Eliminación de documentos en MongoDB. <p><u>EXAMEN FINAL (15%)</u></p> <ul style="list-style-type: none">● Entrega de un Informe del Proyecto Final: un informe detallado que incluye antecedentes de investigación, metodología, análisis, resultados y conclusiones. El informe también debe reflejar el proceso del proyecto y cualquier consideración ética abordada. |
|--|---|--|

