

CARRERA: Licenciatura en Ciencias de Datos	CURSO LECTIVO: 2024
CÁTEDRA: Introducción a la Ingeniería de Datos	CURSO: 1º año - 1º semestre
DURACIÓN: Semestral	Hs. TOTALES: 48 Hs. Reloj Totales
SEMANAS: 16	Hs. TEÓRICAS: 32 Hs. Reloj Totales Hs. PRÁCTICAS: 16 Hs. Reloj Totales

PROFESORES ASISTENTES: CRESPO COLINA, DILIANA / DEVRIENT GOMEZ, CRISTIAN

1. OBJETIVOS DE LA ASIGNATURA

Que los alumnos logren:

- Comprender la importancia del procesamiento de datos en el marco de las ciencias de datos.
- Formar una idea de la perspectiva histórica del procesamiento de datos para comprender el presente de la práctica y las tendencias a futuro.
- Incorporar el alcance de la ingeniería de datos.
- Reconocer la infraestructura necesaria para el procesamiento de datos.
- Familiarizarse con el concepto de plataforma de datos.
- Introducir el concepto de Cloud Computing desde la perspectiva del procesamiento de datos.

2. UNIDADES TEMÁTICAS

1. Introducción.

Distinción entre Ciencia e Ingeniería. Definición de "Ingeniería de Datos". Sistemas Informáticos. Necesidad de una "Ingeniería de Sistemas". Necesidad de una "Ingeniería de Datos". Rol del Ingeniero de Datos. Big Data y patrones de diseño de plataformas de procesamiento de datos.

2. Infraestructura.

Data Centers. Virtualización de recursos y containers. Networking. Seguridad. Continuidad de negocio. Servicios on prem vs on cloud.

3. Sistemas Operativos.

Definición. Historia. Conceptos Básicos. Linux. Windows. Hipervisores. Comandos básicos. Nociones básicas de los lenguajes de scripting. Shell Script.

4. Bases de Datos.

Bases de Datos. Historia de las Bases de Datos. Tipos de Bases de Datos. Bases de

datos transaccionales (OLTP) y analíticas (OLAP). Bases de Datos Relacionales.
Bases de Datos

Orientadas a Documentos. Bases de Datos Clave-Valor. Bases de Datos Orientadas a Grafos. Bases de Datos de Series de Tiempo.

5. **SQL.**

Introducción a SQL. Configuración y utilización del RDBMS. Lenguaje de Definición de datos (DDL). Lenguaje de Manipulación de Datos (DML).

6. **Flujos de Datos.**

Flujos de datos. Filtrado, Homogeneización y Transformación de datos. Recolección y Administración de Datos. Datos en Plataformas. Cloud Data Platform.

7. **Gobierno de Datos.**

Data Governance. Necesidad y justificación de un gobierno de datos. La organización orientada a datos. Seguridad. Linaje. Actores asociados a datos en las organizaciones.

8. **Calidad y Observabilidad.**

Data Quality. Data Observability. Curación de Datos. Estrategia de calidad de datos. Herramientas de calidad de Datos. Herramientas de Observabilidad de Datos.

3. BIBLIOGRAFÍA

3.1 BIBLIOGRAFÍA GENERAL OBLIGATORIA

- “The data engineering cookbook” – Andreas Kretz - Edición online: <https://github.com/andkret/Cookbook>
- “Big Data: Principles and best practices of scalable realtime data systems” – N. Marz, J. Warren – Ed. Manning – 2015 - ISBN 9781617290343.
- “Data Pipelines with Apache Airflow” – B.P. Harenslak & J. Rutger de Ruiter – Ed. Manning - 2021 - ISBN 9781617296901.
- “Designing Cloud Data Platforms” – D. Zburivsky & L. Partner – Ed. Manning - 2021 - ISBN 9781617296444.
- “Data Governance: The Definitive Guide” – E. Eryurek, U. Gilad, V. Lakshmanan, A. Kibunguchy-Grant & J. Ashdown - O'Reilly Media, Inc. - 2021 - ISBN: 9781492063490.

4. METODOLOGÍA

El curso está organizado en 8 unidades temáticas divididas en encuentros de 3hs reloj semanales, a realizarse en formato presencial. La modalidad adoptada para el dictado será **teórico-práctica**. En las clases se presentarán los temas de cada unidad, proponiendo espacios de intercambio con el docente y entre los estudiantes a partir de consignas específicas. Se facilitará material de lectura obligatoria y complementaria para complementar la comprensión de las unidades.

5. EVALUACIONES Y CRITERIOS PARA LA APROBACIÓN

La aprobación de la materia estará supeditada al cumplimiento de la condición de asistencia exigida por la Universidad, la aprobación de todas las actividades prácticas y la aprobación del examen integrador.

Los trabajos prácticos podrán ser individuales o grupales, debiéndose cargar a través de la plataforma de Entornos Virtuales de Aprendizaje en tiempo y forma, otorgándole una única instancia de revisión y recuperación. Las actividades prácticas deberán contar con su aprobación para acceder a la instancia de evaluación final.

Para los trabajos prácticos y la evaluación final se realizarán sesiones de consultas individuales y grupales, haciendo además puesta en común general si el caso lo requiera. A los estudiantes que presenten dificultades se les observará y se los guiará para resolver el conflicto.

La instancia de recuperación está prevista para estudiantes que no hayan aprobado el examen integrador o que hayan estado ausentes.

Criterios de Evaluación:

- Respeto de las consignas presentadas.
- Resolución correcta de los problemas planteados.
- Adecuada respuesta a los contenidos teóricos.

6. CRITERIOS y MODALIDAD PARA LA EVALUACIÓN DEL EXAMEN FINAL

El examen final consiste en una evaluación oral y escrita, presencial e individual, donde el alumno deberá demostrar conocimientos teóricos y prácticos. El examen final se diferencia en que abarca todos los temas del programa. Los ejercicios prácticos tendrán un carácter integrador, articulando los distintos contenidos vistos en la materia. Finalmente, en las preguntas teóricas se pretende que el alumno demuestre un conocimiento profundo de los temas, relacionando conceptos entre sí.

Criterios de Evaluación:

- Respeto de las consignas presentadas.
- Adecuada respuesta a los contenidos teóricos.
- Relación de conceptos pertinente.
- Resolución correcta de los problemas planteados.
- Fundamentación bibliográfica de los temas.