



Ministerio de Cultura y Educación
Universidad Nacional de San Luis
Facultad de Ciencias Físico Matemáticas y Naturales
Departamento: Informatica
Area: Area I: Datos

(Programa del año 2025)

I - Oferta Académica

Materia	Carrera	Plan	Año	Período
() NUEVAS TENDENCIAS EN LA GESTIÓN Y PROCESAMIENTO DE DATOS	LIC.CS.COMP.	32/12	2025	2° cuatrimestre
		026/1		
() NUEVAS TENDENCIAS EN LA GESTIÓN Y PROCESAMIENTO DE DATOS	ING. INFORM.	2-	2025	2° cuatrimestre
		08/15		

II - Equipo Docente

Docente	Función	Cargo	Dedicación
REYES, NORA SUSANA	Prof. Responsable	SEC U EX	2 Hs
BARRIONUEVO, MERCEDES DEOLINDA	Prof. Colaborador	P.Adj Exc	40 Hs
KASIAN, FERNANDO ANDRES	Prof. Colaborador	P.Adj Exc	40 Hs
LUDUEÑA, VERONICA DEL ROSARIO	Prof. Colaborador	P.Adj Exc	40 Hs
TISSERA, PABLO CRISTIAN	Prof. Colaborador	SEC F EX	2 Hs
GIL COSTA, GRACIELA VERONICA	Prof. Co-Responsable	P.Tit. Exc	40 Hs

III - Características del Curso

Credito Horario Semanal				
Teórico/Práctico	Teóricas	Prácticas de Aula	Práct. de lab/ camp/ Resid/ PIP, etc.	Total
5 Hs	Hs	Hs	Hs	5 Hs

Tipificación	Periodo
B - Teoria con prácticas de aula y laboratorio	2° Cuatrimestre

Duración			
Desde	Hasta	Cantidad de Semanas	Cantidad de Horas
04/08/2025	14/11/2025	15	75

IV - Fundamentación

En el panorama tecnológico actual, el volumen, la variedad y la velocidad de los datos han crecido exponencialmente, dando origen a la era de los grandes volúmenes de datos digitales. Esta explosión de datos ha transformado la forma en que las organizaciones operan y toman decisiones. Las bases de datos relacionales tradicionales, si bien siguen siendo fundamentales, a menudo resultan insuficientes para manejar los desafíos que presentan estos nuevos flujos de información.

Esta asignatura optativa se justifica por la necesidad de capacitar a futuros profesionales en Ciencias de la Computación e Ingenierías afines para comprender, diseñar y aplicar soluciones a los problemas de gestión y procesamiento de datos masivos. La materia se centra en las arquitecturas, tecnologías y paradigmas emergentes que abordan las limitaciones de los sistemas tradicionales, incluyendo bases de datos NoSQL, procesamiento distribuido, arquitecturas de streaming, y los desafíos de la calidad, seguridad y gobernanza de datos en entornos a gran escala.

La formación en estas nuevas tendencias es crucial para que los egresados puedan desempeñarse exitosamente en roles de alta demanda, como arquitecto de datos, ingeniero de datos, científico de datos y desarrollador de sistemas de información a gran escala.

V - Objetivos / Resultados de Aprendizaje

Al finalizar la asignatura, el estudiante estará en capacidad de:

- Comprender los principios y desafíos de gestionar y procesar grandes volúmenes de datos, así como las limitaciones de los sistemas de bases de datos relacionales en este contexto.
- Analizar y seleccionar la arquitectura de base de datos más adecuada para una problemática específica, justificando la elección en base a los requerimientos de la aplicación.
- Evaluar y aplicar diferentes paradigmas de procesamiento distribuido de datos, tanto en batch (como MapReduce) como en tiempo real (streaming).
- Analizar los desafíos de la calidad de datos y la gobernanza en sistemas de procesamiento masivo de datos.
- Desarrollar soluciones prácticas que integren diversas herramientas y tecnologías para resolver problemas reales de gestión y análisis de datos.

VI - Contenidos

Módulo 1: Fundamentos de sistemas para grandes volúmenes de datos

- Definición y características del Big Data: las "Vs" (Volumen, Velocidad, Variedad, Veracidad).
- Desafíos y oportunidades del Big Data en la industria.
- Revisión de bases de datos relacionales y sus limitaciones para Big Data.
- Introducción a las arquitecturas de datos modernas: Data Lake, Data Warehouse y Lakehouse.

Módulo 2: Bases de Datos NoSQL

- Bases de Datos Clave-Valor: Principios, modelos de datos, casos de uso. Ejemplos.
- Bases de Datos Documentales: Estructura de documentos JSON/BSON, flexibilidad del esquema, consultas. Ejemplos.
- Bases de Datos de Columnas (Wide-Column): Orientación a columnas, casos de uso. Ejemplos.
- Bases de Datos de Grafos: Nodos, aristas y propiedades. Consultas de grafos y algoritmos. Ejemplos.
- Teorema de CAP y su implicancia en la elección de una base de datos. Consistencia, disponibilidad, tolerancia a particiones.

Módulo 3: Procesamiento Distribuido de Datos

- Procesamiento en Batch:
 - * Modelo de programación MapReduce.
 - * Ejemplos de tecnologías para procesamiento en batch..
- Procesamiento de Streaming y Datos en Tiempo Real:
 - * Conceptos de streaming, eventos y micro-batches. ○ Sistemas de mensajería y colas de eventos.
 - * Procesamiento de flujos.

Módulo 4: Desafíos Transversales y el Futuro

- Calidad y Limpieza de Datos: Estrategias para la veracidad y la integridad de los datos.
- Seguridad en Big Data: Autenticación, autorización y encriptación en arquitecturas distribuidas.
- Gobernanza de Datos: Linaje de datos, catalogación y cumplimiento normativo.
- Tendencias emergentes.

Módulo 5: Casos de Estudio y Visualización

- Análisis de arquitecturas de grandes volúmenes de datos en la industria.
- Herramientas de visualización para grandes volúmenes de datos.

- Estrategias de visualización interactiva y tableros de comando.

VII - Plan de Trabajos Prácticos

Se proponen tres trabajos prácticos que abarcan los módulos de la asignatura, con el fin de consolidar los conocimientos teóricos a través de la práctica.

- TP 1 - Exploración de Bases de Datos NoSQL:

- * Implementación de un sistema de gestión de datos para un caso de uso específico utilizando alguno de los sistemas para Bases de Datos NoSQL.
- * Modelado de datos, inserción, actualización y ejecución de consultas avanzadas en sistemas que respondan a distintos modelos.

- TP 2 - Procesamiento de Datos Distribuidos:

- * Configuración y uso básico de un clúster (en modo local o en la nube).
- * Desarrollo de un pipeline de procesamiento en batch para analizar un conjunto de datos públicos.
- * Desarrollo de una aplicación simple de streaming para procesar datos en tiempo real (simulados o reales).

- TP 3 - Proyecto Integrador:

- * Diseño e implementación de una pequeña arquitectura de datos que incluya e integre los conceptos vistos en la asignatura.

VIII - Regimen de Aprobación

Condiciones para regularizar

- 1) Asistir al menos al 80% de las clases.
- 2) Entregar y aprobar los trabajos prácticos asignados por los docentes responsables, con nota mayor o igual a siete (7).

Condiciones para promocionar

- 1) Haber dado cumplimiento a las condiciones para regularizar.
- 2) Presentar y aprobar un proyecto escrito que muestre la solución a una problemática mediante la aplicación e integración de los temas vistos. O presentar y aprobar un informe escrito sobre alguna temática a investigar, en la que se integren los temas vistos. Los trabajos pueden ser grupales, entre dos personas máximo; a excepción de que el tema amerite la participación de más integrantes, y los docentes responsables presten acuerdo.

La nota mínima de aprobación sobre este punto es siete (7).

Modalidad de examen final: El examen final podrá ser oral y/o escrito. Examen Libre: No se admiten alumnos libres por la modalidad de la asignatura.

IX - Bibliografía Básica

- [1] Fundamentals of Database Systems - Elmasri, R.; Navathe, S., Sixth Edition. Addison-Wesley, 2011. ISBN-10: 0136086209 - ISBN-13: 9780136086208.
- [2] Data Architecture: A Primer for the Data Scientist, W. Inmon - D. Lindstedt, 1st ed., Morgan Kaufmann, 978-0-12-802044-9, 2015.
- [3] Mining of Massive Datasets - Rajaraman, A.; Ullman, J. Cambridge University Press, 2011. ISBN: 9781107015357.
- [4] Metric Space Indices for Dynamic Optimization in a Peer to Peer-Based Image Classification Crowdsourcing Platform - Fernando Loor, Veronica Gil-Costa, Mauricio Marín: Future Internet 16(6): 202 (2024).
- [5] An End-to-End Workflow to Efficiently Compress and Deploy DNN Classifiers on SoC/FPGA - Romina Soledad Molina, Iván René Morales, Maria Liz Crespo, Veronica Gil-Costa, Sergio Carrato, Giovanni Ramponi: IEEE Embed. Syst. Lett. 16(3): 255-258 (2024).
- [6] Feasibility of P2P-STB based crowdsourcing to speed-up photo classification for natural disasters - Fernando Loor, Manuel Manriquez, Veronica Gil-Costa, Mauricio Marín: Clust. Comput. 25(1): 279-302 (2022).
- [7] Elastic and Real-time Capacity Planning for Web Search Engines - Veronica Gil-Costa, Alonso Inostrosa-Psijas, Mauricio Marín: PDP 2020: 331-338.

- [8] Foundations of Multidimensional and Metric Data Structures - Hanan Samet. Morgan Kaufmann Publishers Inc., 2006.
- [9] Similarity Search: The Metric Space Approach (Advances in Database Systems) - Zuzula, P.; Amato, G.; Dohnal, V.; Batko, M. Springer- Verlag New York, Inc., Secaucus, NJ, USA, 2005.
- [10] Searching in Metric Spaces - Chávez, E.; Navarro, G.; Baeza-Yates, R.; Marroquín, J. L., ACM Computing Surveys, Vol. 33, No. 3, September 2001, pp. 273–321. - External memory algorithms and data structures: dealing with massive data Vitter, J. S., ACM Computing Surveys, Vol. 33, Number 2, 2001. ISSN 0360-0300, Pp 209-271,doi = <http://doi.acm.org/10.1145/384192.384193>. ACM Press.
- [11] Modern Information Retrieval: the concepts and technology behind search -
- [12] R. Baeza-Yates, B. Ribeiro-Neto, 2da. Edición, Addison-Wesley Professional; 978-0-321-41691-9, 2011.
- [13] Indexing metric spaces for exact similarity search - Lu Chen, Yunjun Gao, Xuan Song, Zheng Li, Yifan Zhu, Xiaoye Miao, and Christian S Jensen. ACM Computing Surveys, 55(6):1–39, 2022.
- [14] Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems, 1ª Edición - Martin Kleppmann, O'Reilly Media, 2017.
- [15] Spark: The Definitive Guide: Big Data Processing Made Simple - Bill Chambers y Matei Zaharia, O'Reilly Media.
- [16] Seven Databases in Seven Weeks: A Guide to Modern Databases and the NoSQL Movement - Eric Redmond y Jim R. Wilson, The Pragmatic Programmers. 2018. ISBN-13: 978-1-68050-253-4.
- [17] Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking - Foster Provost and Tom Fawcett, O'Reilly Media.
- [18] Information Retrieval: Implementing and Evaluating Search Engines - S.
- [19] Büttcher, C.L.A. Clarke, G.V. Cormack. The MIT Press, 2016.
- [20] SQL & NoSQL Databases: Models, Languages, Consistency Options and Architectures for Big Data Management - Andreas Meier, Michael Kaufmann. 2nd Ed. Springer. ISBN 978-3-031-27907-2 ISBN 978-3-031-27908-9 <https://doi.org/10.1007/978-3-031-27908-9>. 2023.
- [21] Business Database Technology: Theories and Design Process of Relational Databases, SQL, Introduction to OLAP, Overview of NoSQL Databases Shouhong Wang, Hai Wang. 2nd Edition. Universal-Publishers. ISBN: 978-1-62734- 389-3 (pbk.) ISBN: 978-1-62734-390-9 (ebk.). 2022.
- [22] Database Systems: The Complete Book - Garcia-Molina H., Ullman J, Widom J., Prentice-Hall, 2008.
- [23] Building the Data Warehouse - Inmon, W. H., 4th edition, Wiley Publishing, 2005.
- [24] The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling - Ralph Kimball y Margy Ross, 3rd Edition, Wiley, 2013.

X - Bibliografía Complementaria

[1] --

XI - Resumen de Objetivos

Al finalizar la asignatura, el estudiante estará en capacidad de:

- Comprender los principios y desafíos de gestionar y procesar grandes volúmenes de datos.
- Analizar y seleccionar la arquitectura de base de datos más adecuada para una problemática específica.
- Evaluar y aplicar diferentes paradigmas de procesamiento distribuido de datos.
- Analizar los desafíos de la calidad de datos y la gobernanza en sistemas de procesamiento masivo de datos.
- Desarrollar soluciones prácticas que integren diversas herramientas y tecnologías para resolver problemas reales de gestión y análisis de datos.

XII - Resumen del Programa

1. Grandes Volúmenes de Datos: Conceptos, Desafíos y Arquitecturas.
2. Bases de Datos NoSQL: Tipos, modelos de datos y casos de uso.
3. Procesamiento Distribuido: Arquitecturas Batch y de Streaming.
4. Calidad, Seguridad y Gobernanza de Datos.
5. Análisis y Visualización de Datos a Gran Escala.

XIII - Imprevistos

--

XIV - Otros

Contacto con la cátedra:

- Por mail: eda@unsl.edu.ar
- Por teléfono: Interno 2107 y 2108
- Personalmente: Segundo Bloque - box 7, box 8 - Primer piso.