



**.UBA40**<sup>∞</sup>  
AÑOS DE  
DEMOCRACIA

## **Resolución Consejo Directivo**

**Número:**

**Referencia:** EX-2023-06349795- -UBA-DMESA#FCEN - POSTGRADO - Sesión  
11/12/2023

---

**VISTO:**

La nota presentada por la Dirección del Departamento de Computación, mediante la cual eleva la información del curso de posgrado Acerca de la Ingeniería de Datos para Aprendizaje Automático para el año 2024,

**CONSIDERANDO:**

lo actuado por la Comisión de Doctorado,

lo actuado por este Cuerpo en la sesión realizada el día 11 de diciembre de 2023,

en uso de las atribuciones que le confiere el Artículo 113° del Estatuto Universitario,

**EL CONSEJO DIRECTIVO DE LA FACULTAD  
DE CIENCIAS EXACTAS Y NATURALES**

## **R E S U E L V E:**

**ARTÍCULO 1°:** Aprobar el nuevo curso de posgrado Acerca de la Ingeniería de Datos para Aprendizaje Automático de 32 horas de duración, que será dictado por el Dr. Leandro Nahabedian con la colaboración del Dr. Ricardo Oscar Rodríguez.

**ARTÍCULO 2°:** Aprobar el programa del curso de posgrado Acerca de la Ingeniería de Datos para Aprendizaje Automático que como anexo forma parte de la presente Resolución, para su dictado en el primer bimestre de 2024.

**ARTÍCULO 3°:** Aprobar un puntaje máximo de uno y medio (1,5) puntos para la Carrera del Doctorado.

**ARTÍCULO 4°:** Establecer un arancel de CATEGORÍA 2, estableciendo que dicho arancel estará sujeto a los descuentos y exenciones estipulados mediante la Resolución CD N.º 1072/19. Disponer que los fondos recaudados ingresen en la cuenta presupuestaria habilitada para tal fin, y sean utilizados de acuerdo a la Resolución 072/03

**ARTÍCULO 5°:** Disponer que, de no mediar modificaciones en el programa, la carga horaria y el arancel, el presente Curso de Posgrado tendrá una vigencia de cinco (5) años a partir de la fecha de la presente Resolución.

**ARTÍCULO 6°:** Comuníquese a todos los Departamentos Docentes, a la Dirección de Estudiantes y Graduados, a la Biblioteca de la FCEyN y a la Secretaría de Posgrado con copia del programa incluida. Cumplido, pase COMPUTACION#FCEN y resérvese.

## ANEXO

### PROGRAMA

- Clase 1: Presentación. ¿Cuál es el rol del data engineer? Proveedores cloud: AWS GCP Azure. Generalidades de los tres proveedores. Revisión general de todos los componentes. Almacenamiento
- Clase 2: Conceptos básicos de almacenamiento de datos. Tipos de datos a almacenar. Base de datos relacionales vs no-relacionales. Buckets. Graph databases. Caching.
- Clase 3: Datos en real-time. Procesamiento de datos en streaming. Arquitectura de streaming con retries. Pub/Sub vs Stream Analytics vs Kinesis vs Kafka Procesado
- Clase 4: ETL vs ELT. Orquestadores (Airflow). Stateless apps.
- Clase 5: Spark (pySpark). Databricks. Soluciones Dockerizadas Arquitectura
- Clase 6: Misc: Infrastructure as code. CI/CD. Monitoreo y Reportes.
- Clase 7: Soluciones a alto nivel utilizando los componentes vistos. Soluciones Batch. Soluciones Real-time. Arquitectura Lambda. Repaso para el examen.
- Clase 8: Examen

### BIBLIOGRAFÍA

La bibliografía es principalmente la documentación de estos tres proveedores cloud y sus learning paths: AWS, GCP y Azure. Como bibliografía adicional se destacan algunos papers que se listan a continuación:

- Polyzotis, Neoklis, Martin Zinkevich, Sudip Roy, Eric Breck, and Steven Whang. “Data validation for machine learning”. Proceedings of Machine Learning and Systems 1 (2019).
- Polyzotis, Neoklis, Sudip Roy, Steven Euijong Whang, and Martin Zinkevich. “Data

Management Challenges in Production Machine Learning”. In Proceedings of the 2017 ACM International Conference on Management of Data (2017).

● Sambasivan, Nithya, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, and Lora M. Aroyo. “Everyone wants to do the model work, not the data work: Data Cascades in High-Stakes AI”. In Proceedings of the Conference on Human Factors in Computing Systems, (2021).