



Instituto Tecnológico
de Buenos Aires

Maestría y Especialización en Ciencia de Datos_

x

+

+

+



+

x

itba.edu.ar

La **Maestría en Ciencia de Datos y la Especialización en Ciencia de Datos** brindan a los profesionales herramientas para diseñar, preparar, analizar y manejar grandes volúmenes de información, estructurada y no-estructurada.

La Especialización hace énfasis, por un lado, en los fundamentos teóricos que otorgan al graduado la flexibilidad necesaria para adaptarse a los abruptos cambios tecnológicos; y por el otro, en estudios de casos y práctica en laboratorios con software comercial y open source.

La Maestría profundiza los conocimientos, fortalece y consolida las competencias propias del campo disciplinar.

La organización de la carrera está orientada a formar profesionales con sólidos conocimientos teóricos, pero también prácticos, de las técnicas, algoritmos, metodologías y tecnologías fundamentales aplicadas en el tratamiento de problemas que involucran grandes volúmenes de datos complejos. En estos casos es fundamental tener no solamente un conocimiento teórico de los problemas, sino también de las tecnologías más actuales que permitan implementar la mejor solución.

>> DIRECTOR DE LA CARRERA

- + Doctor en Ciencias de la Computación, Universidad de Buenos Aires (UBA).
- + Profesor e investigador en el área de bases de datos, en particular en Data Warehousing y Business Intelligence, Semantic Web y Sistemas de Información Geográfica.
- + Director del Centro de Extracción de Información del ITBA.



>> **Alejandro Vaisman**

>> TÍTULO A OTORGAR: Especialista en Ciencia de Datos.

*Acreditada en Sesión CONEAU N° 424 del 27 de julio de 2015.
Resolución Ministerial N°1753/16*

>> DURACIÓN:



OBJETIVOS DE LA CARRERA:

>> Objetivo general

-Formar recursos humanos de nivel superior que permitan cubrir la demanda de profesionales, actualmente insatisfecha en el campo del análisis y explotación de grandes volúmenes de datos complejos.

-Lograr una vinculación efectiva entre industria y academia que fomente la investigación y desarrollo de tecnologías de análisis y explotación de grandes datos ("Big Data") en el país.

>> Objetivos Específicos

-Proveer a los profesionales conocimientos específicos que no hayan sido alcanzados por sus carreras de grado.

-Proporcionar las herramientas que permitan al graduado diseñar, preparar, analizar y manejar grandes volúmenes de información compleja, estructurada y no-estructurada. Para ello, se enfatizará tanto en los fundamentos teóricos que le brinden al graduado la flexibilidad necesaria para adaptarse a los abruptos cambios de tecnologías, como también en estudios de casos y práctica en laboratorios con software comercial y open source.

-Desarrollar las capacidades de los profesionales para trabajar en equipos multidisciplinares, por lo que se pondrá énfasis no sólo en las habilidades técnicas sino también en las comunicacionales

PLAN DE ESTUDIOS:

La Especialización en Ciencia de Datos tiene una duración de dos semestres, con una carga horaria de 369 horas. Se compone de **ocho cursos teórico-prácticos, un seminario intensivo** con profesionales e investigadores invitados, y **un taller que tiene como finalidad la definición del Trabajo Final Integrador**.

Además, al comienzo de la carrera se dictará un curso de nivelación en Estadísticas, Bases de Datos y Algoritmos. Este curso será de carácter obligatorio para los estudiantes que las autoridades determinen durante el proceso de admisión a la carrera, de acuerdo a la formación académica y/o experiencia profesional.

La Especialización se dicta bajo la modalidad presencial.

Debido a la naturaleza multidisciplinar del área, y al hecho de que los problemas en general son analizados y resueltos por un equipo y rara vez en forma individual, se pretende desarrollar y fomentar el trabajo en equipo. El régimen de evaluación de las distintas asignaturas será definido por el profesor a cargo de cada asignatura de acuerdo a las características de la misma, e incluirá como mínimo, el desarrollo de un proyecto y un examen final.

Las actividades prácticas de las materias consistirán en el desarrollo de proyectos relativos a cada asignatura. Estos proyectos se implementarán utilizando diversas herramientas de software, definidas conjuntamente entre el docente y la Dirección de la Carrera. Cabe destacar que la Universidad tiene convenios con las empresas más importantes del área, con lo que se garantiza que los alumnos accedan a software comercial de última generación, además, de ser expuestos a software de tipo open source, brindando de esta manera, un panorama amplio de las tecnologías disponibles, con las cuales aplicar los conocimientos adquiridos durante las clases teóricas.

La infraestructura de base para la realización de las actividades prácticas en el desarrollo de la carrera de Especialización en Ciencia de Datos estará brindada por laboratorios existentes en el Departamento de Ingeniería Informática, y por la infraestructura y el laboratorio móvil correspondiente a Maestrías y Especializaciones.

Módulo 1: Nivelación

Esta asignatura se divide en tres sub-módulos: Estadística, Algoritmos, y bases de datos. Dado que la temática de Big Data es esencialmente multidisciplinaria, nuestra intención es que la Especialización atraiga profesionales y graduados de distintas áreas. Es necesario, por lo tanto, que aquellos profesionales de carreras no informáticas adquieran los conceptos imprescindibles de análisis de algoritmos y fundamentos de bases de datos, que les permitan seguir adecuadamente las asignaturas siguientes.

Por otro lado, consideramos conveniente que, aunque los candidatos hayan cursado alguna materia de Estadísticas, tomen este curso para refrescar conocimientos. La decisión de requerir o no la aprobación de este

curso es potestad de la Dirección de la carrera y del Comité Asesor, y se tomará en forma individual estudiando los antecedentes académicos y profesionales de cada candidato.

CONTENIDOS MÍNIMOS:

Fundamentos de estadística (12hs). Noción de matriz de datos. Variables cualitativas, cuantitativas. Estadística descriptiva: tablas de simple y doble entrada. Histogramas, gráficos de tortas, polígonos, nubes de puntos. Concepto de probabilidad. Distribuciones clásicas: binomial, normal, etc. Test de hipótesis.

Principios de algoritmos y programación (12hs). Tipos de datos: primitivos, abstractos y algebraicos. Tipos abstractos de datos. Secuencias. Pilas. Colas. Arreglos. Árboles binarios. Hashing. Conceptos básicos de programación imperativa. Estructuras de control. Variables. Funciones y procedimientos. Técnicas de diseño de algoritmos: dividir y conquistar, algoritmos greedy. Métodos de Sorting. Complejidad de algoritmos. Grafos. Definiciones básicas. Modelar problemas usando grafos. Algoritmos para recorrer grafos. Implementación de algoritmos simples en Java.

Principios de bases de datos (12hs). El concepto de base de datos. Distintos tipos de DBMS (database management systems). Modelos de Datos. Modelo Entidad/Relación (conceptual). Diseño de una base de datos. Modelo lógico. Modelo relacional. Pasaje del modelo E/R al relacional.

El lenguaje SQL. Cláusulas SELECT, FROM, WHERE, GROUP BY. Operadores avanzados. Diseño Relacional. Normalización: Tercera Forma Normal y Forma Normal de Boyce-Codd.

Módulo 2: Fundamentos de Análisis de Datos

Esta materia tiene como objetivo impartir los fundamentos estadísticos del análisis inteligente de datos. Es decir, no enfoca el punto de vista algorítmico, sino conceptual. Estos fundamentos serán utilizados en muchas de las asignaturas que componen la especialización.

CONTENIDOS MÍNIMOS:

- **Análisis exploratorio de datos (EDA).** Reducción de dimensionalidad: análisis de componentes principales. Regresión lineal simple y múltiple. Regresión logística. Análisis de varianza (ANOVA).
- Análisis de datos de encuestas. Curvas ROC, ganancia. Redes Bayesianas. Introducción al análisis de series de tiempo: Los modelos ARIMA (autoregressive integrated moving average), ARC (autoregressive conditional heterogeneity), GARCH (generalized autoregressive conditional heterogeneity).

Docente: Dra. Juliana Gambini

Módulo 3: Minería de Datos

En esta materia se abordarán las técnicas y algoritmos básicos de data mining, con especial énfasis en regresión, análisis de asociación, y clustering. Partiendo de las técnicas clásicas se discutirán los nuevos desafíos que impone Big Data.

CONTENIDOS MÍNIMOS:

Conceptos básicos de data mining. Modelos descriptivos y predictivos. Técnicas fundamentales: reglas de asociación, clasificación, clustering, patrones. Árboles de decisión. Aplicación a problemas concretos de predicción. Casos de estudio. KPIs (Key Performance Indicators). Dashboards. Herramientas comerciales y de código abierto.

Docente: Mg. Gustavo Denicolay Pacheco.

×

Módulo 4: Almacenes de Datos y Procesamiento Analítico en Línea

Se estudiará la arquitectura de los data warehouses (DW), así como su diseño conceptual, lógico y físico, y su explotación mediante herramientas de Online Analytical Processing (OLAP), mining, dashboards, etc.

CONTENIDOS MÍNIMOS:

Arquitecturas. Diseño conceptual, lógico y físico. El modelo multidimensional: estrella, snowflake y constellation. Slowly changing dimensions. Diseño físico. Particiones. On Line Analytical Processing: OLAP. OLAP vs OLTP. Lenguajes de consulta: SQL avanzado, funciones de ventana. Entornos avanzados para OLAP. Herramientas comerciales y de código abierto. Procesamiento en paralelo de consultas OLAP. Almacenamiento columnar.

Docente: Dr. Alejandro Vaisman.

Módulo 5: Herramientas de Procesamiento para Grandes Volúmenes de Datos *

En esta materia se estudian las herramientas que popularmente se identifican con Big Data: Hadoop y MapReduce, así como las arquitecturas de procesamiento masivamente paralelo con clusters de tipo commodity. Es decir, aquí el estudiante estará expuesto a un entorno real de Big Data, desde el punto de vista del hardware y del software.

CONTENIDOS MÍNIMOS:

Fundamentos de sistemas distribuidos. Modelos. El teorema CAP. Clusters para programación masivamente paralela (MPP). Virtualización de clusters y data centers. Arquitecturas cloud. Conceptos fundamentales de Big Data: Velocidad, Volumen, Variedad, Veracidad. Qué es y qué no es "Big Data". Datos estructurados y no estructurados. Bases de datos No SQL: MongoDB. El paradigma MapReduce. Hadoop Distributed File System. YARN. Arquitectura, componentes. Bases de datos columnares: Apache Cassandra, HBase. Key-value stores: Amazon DynamoDB, Redis. Lenguajes de alto nivel: HiveQL y Pig Latin. Apache Spark. Programación con Spark. Spark SQL. streaming, captura de Tweets y de datos en real time, machine learning con Spark. Procesamiento de datos con Notebooks Python.

Docente: Ing. Juan Martín Pampliega.

Módulo 6: Visualización de la información

La visualización de la información es parte fundamental del análisis de datos. En este curso se verán los fundamentos teóricos de la visualización de datos (por ejemplo, cómo representar visualmente datos con alta dimensionalidad) y las herramientas prácticas para implementarla.

CONTENIDOS MÍNIMOS:

Introducción, definiciones, antecedentes, gráficos notables. Principios de la excelencia gráfica. Observaciones y variables. Tipo de variables. Visualizando tablas, jerarquías y redes. Utilización del color. Representación eficiente de la información, sumarización y visualización de grandes volúmenes de datos. Prácticas con d3js, jit, Processing, Google Visualization API, Tableau, Fusion Tables y QGIS.

Docente: Lic. Ariel Aizemberg

Módulo 7: Técnicas y Algoritmos de Aprendizaje Automático

El objetivo de esta materia es que el estudiante adquiera los conocimientos generales y prácticos del estado del arte en Aprendizaje Automático para su aplicación en la práctica profesional, en particular en un contexto de Big Data. El estudiante, al finalizar el curso, conocerá los principales modelos y algoritmos de aprendizaje computacional, y podrá definir métodos y pruebas que le permitan seleccionar el modelo apropiado a los casos prácticos que se le presente. Esta materia complementará los modelos y algoritmos que se estudian en la asignatura "Data Mining".

CONTENIDOS MÍNIMOS:

Conceptos básicos de Machine Learning. Inferencias. Espacio de versiones. Aprendizaje como búsqueda heurística. Conceptos generales de sesgo y poda. Árboles de decisión. Extensiones a los algoritmos básicos y problemas de implementación. Generación de reglas. Aprendizaje Bayesiano. Algoritmos aglomerativos y de partición. K-Means, SVM. Características descriptivas y discriminantes. Redes neuronales. Introducción a Deep Learning.

Docente: Dra. Marcela Riccillo.

Módulo 8: Diseño e Implementación de Procesos de Extracción, Transformación y Carga de Grandes Volúmenes de Datos.

El proceso de extracción, transformación, y carga (ETL) de un data warehouse (DW) es el proceso clave en todo proyecto, ya que involucra el 80% de su presupuesto.

En un entorno de Big Data es aún más crítico, pues al volumen de datos, se suma la necesidad de un análisis en tiempo casi-real (por la velocidad de llegada de los datos), y una gran variedad de procesos de captura y adquisición, debido a la variedad de los datos, en gran medida provenientes de la Web. En esta asignatura se estudiarán estos procesos, con fuerte enfoque de aplicación.

CONTENIDOS MÍNIMOS:

El proceso de Extract, Transform-Load (ETL). Diseño Conceptual. Uso de técnicas de BPMN. Aplicación. Herramientas comerciales (MS Integration Services) y de código abierto (Talend). ETL para soporte de real-time OLAP y DW. Uso de Hadoop/MapReduce en el proceso de ETL. ETL vs ELT. Ejemplos prácticos y programación de procesos ETL.

Docente: Dra. Leticia Gómez.

Módulo 9: Análisis de Datos Científicos y Geográficos

Los Sistemas de Información Geográfica (GIS), y las aplicaciones científicas son, junto con las redes sociales, las más grandes fuentes de Big Data, y requieren tratamientos particulares para su administración, integración con otros tipos de datos, y consulta. Estos problemas se abordarán en esta asignatura.

CONTENIDOS MÍNIMOS:

- Sistemas de Información Geográfica (GIS): modelos discretos y continuos (Continuous Fields). Bases de datos espaciales. Data warehouses espaciales. OLAP sobre GIS. Análisis de datos neurocientíficos. Procesamiento de señales de sensores con técnicas de aprendizaje automático. Brain-human interface. Análisis de datos genéticos.

Docente: Dr. Alejandro Vaisman | Lic. Ariel Aizemberg | Dr. Rodrigo Ramele | Dr. Mario Rossi.

Módulo 10: Seminario Intensivo de Tópicos Avanzados en Datos Complejos

Este seminario se realiza **durante una semana a tiempo completo**, y es dictado por profesores invitados de reconocida relevancia en el área. Su objetivo es compartir la visión de otros especialistas, promover el intercambio con otras instituciones, y presentar a los estudiantes posibles temas de trabajo final. Normalmente se realiza en hacia el fin de la cursada, cuando los conocimientos adquiridos permiten obtener el mayor provecho del seminario. La fecha y horario se confirman con un mínimo de tres meses de antelación a la realización del seminario. Cabe destacar que el seminario **debe ser aprobado con condiciones similares al resto de las materias** (es decir, determinado por el docente, un proyecto y examen final, y asistencia mínima del 75%).

Docente: Experto invitado de una Universidad del Exterior. A designar en función de los temas a tratar.

Módulo 11: Taller de Trabajo Final Integrador (TFI)

Durante este seminario los estudiantes planificarán su trabajo final integrador. El objetivo del taller es contribuir a la formación de un pensamiento crítico, alcanzado a través de que el alumno logre: adquirir los conocimientos básicos del método científico y de las técnicas metodológicas; conocer cómo se lleva a cabo y qué etapas incluye la realización de una investigación; conocer los distintos tipos de investigaciones que pueden realizarse, los distintos instrumentos de recolección de datos que pueden utilizarse, y sus ventajas y limitaciones; adquirir habilidad para el análisis metodológico de trabajos de investigación; adquirir los conocimientos indispensables para participar en un trabajo de investigación, y encaminarse en la preparación de su TFI; adquirir las herramientas necesarias para elaborar el reporte final.

El trabajo final consistirá en el desarrollo de un proyecto individual relacionado con un área de aplicación (GIS, biología, etc).

El alumno propondrá un trabajo, y el Director y CA designarán al tutor del TFI, seleccionándolo dentro del grupo de tutores indicado anteriormente. La aceptación formal del tema seleccionado estará a cargo del Director de la Carrera en conjunto con el Comité Académico. El TFI deberá ser presentado como máximo dentro de los 12 meses a partir de la fecha de aprobación del último módulo. El TFI será evaluado por profesionales designados por el Director y seleccionados por su experiencia académica y profesional relevante y pertinente al objeto de estudio del trabajo del alumno. Estos deberán emitir su opinión escrita y fundamentada dentro de los sesenta (60) días de recibida, enmarcados en las pautas que indique el DC en acuerdo con el CA.

Docente: Dra. Alicia Mon.

PERFIL DEL GRADUADO:

El graduado estará capacitado para enfrentar un entorno de "Big Data", y proveer las mejores soluciones para su análisis.

Competencias del graduado:

- ⊕ Analizar el problema
- ⊕ Estudiar las distintas tecnologías disponibles para su solución
- ⊕ Diseñar esta solución, e implementarla

Asimismo, se espera que el graduado de la carrera esté en condiciones de liderar un equipo interdisciplinario altamente especializado, y coordinar su trabajo para brindar la mejor solución técnica y económica ante cada problema que se le plantee

MAESTRIA EN CIENCIA DE DATOS

TITULO A OTORGAR: Magister en Ciencia de Datos.

*Acreditada en sesión CONEAU N°566 del 17 de mayo de 2022.
Resolución Ministerial N°160/11.*

DURACIÓN:

PLAN DE ESTUDIOS:

El plan de estudios de la Maestría incluye los Módulos 1, 2, 3, 4, 5, 6, 7, 8, 9 y 10 descriptos en el plan de la Especialización e incorpora:

- Taller de Tesis.
- Taller de Tesis II.
- Electiva I - Bases de Datos de Grafos.
- Electiva II - Análisis y Tratamiento de Imágenes Satelitales.
- Electiva III - Implementación de Aplicaciones de Aprendizaje Automático en la Nube.
- Electiva IV - Neurorobótica.
- Tesis de Maestría.

REQUISITOS DE ADMISIÓN MAESTRIA Y ESPECIALIZACIÓN:

Para ingresar a la carrera es requisito ser graduado universitario de una carrera de, al menos, cuatro años de duración.

Realizar una entrevista con las autoridades de la carrera, con el fin de asegurar que la formación de grado y/o la experiencia profesional sea adecuada a las exigencias de la Especialización.

- Además, el aspirante deberá contar con conocimientos que permitan la comprensión del idioma inglés.
- Para acceder al título de "Especialista en Ciencias de Datos" los alumnos deberán cumplir las siguientes condiciones:

+ Cursar todos los espacios curriculares obligatorios del plan de estudios, cumpliendo con una asistencia mínima del 75% y **aprobar cada una de las materias con un mínimo de 6 (seis) puntos sobre 10 (diez)**

×

+ Presentar y aprobar un **Trabajo Final Integrador, de carácter individual**

+ Abonar todos los pagos correspondientes a las matrículas anuales y a los aranceles mensuales

Las formalidades de las condiciones de permanencia y graduación se rigen por las normas establecidas en el Reglamento General del ITBA, el Reglamento de la Escuela de Postgrado y las prescripciones específicas del Reglamento de esta Carrera.



+

+

MÁS INFORMACIÓN

ITBA
Maestrías y
Especializaciones

Contacto: *Florencia Imposti*
📞 (+54) 11 5499-0900

postgrado@itba.edu.ar
(+54911) 3754-4731



x



+

+

+



+



x

itba.edu.ar

ITBA
Instituto Tecnológico
de Buenos Aires

