UNIVERSIDAD NACIONAL DE ASUNCIÓN
FACULTAD POLITÉCNICA
CARRERA INGENIERÍA EN INFORMÁTICA
PLAN 2008
PROGRAMA DE ESTUDIO

Resolución 25/07/09-00 Acta 1215/07/04/2025
ANEXO 05

I. IDENTIFICACIÓN

1. Asignatura : Electiva - Big Data
2. Semestre : Según la alternativa en la que se inscribe el estudiante
3. Horas semanales : 7 horas
   - 3.1. Clases teóricas : 3 horas
   - 3.2. Clases prácticas : 4 horas
4. Total real de horas disponibles : 112 horas
   - 4.1. Clases teorías : 48 horas
   - 4.2. Clases prácticas : 64 horas

II. JUSTIFICACIÓN

Las problemáticas que surgen a causa del aumento del volumen, velocidad y variabilidad de los datos obtenidos internamente en la empresa, así como cuestiones relativas a su veracidad y valor, nos conducen al paradigma Big Data. Este paradigma estudia nuevos abordajes sobre los datos mediante la aplicación de técnicas inteligentes provenientes de la integración de conocimientos de diversas áreas científicas y sectores como industria, gobierno, medios sociales, marketing y comercio electrónico, finanzas, energía y otros.

En este contexto, se observa que la demanda por profesionales con formación de Analista en Big Data es creciente en el eje ciencia/industria/gobierno. Este profesional tiene una expectativa de formación interdisciplinaria que reúne sólidas habilidades en ciencia de la computación y aplicaciones, modelado, estadística, analítica y matemática, además del conocimiento del dominio de aplicación.

Con este enfoque, en este curso el estudiante aprenderá conceptos, técnicas y herramientas para la captura, el almacenamiento, el procesamiento y la generación de información a partir de datos, que por su volumen y complejidad estructural, son considerados Big Data.

III. OBJETIVOS

3.1. Introducir el concepto de Big data, presentando las técnicas, metodologías, tecnologías para el análisis y extracción de informaciones.
3.2. Adquirir habilidades en el manejo de datos estructurados y no estructurados para la construcción de bases de datos.
3.3. Aplicar técnicas de análisis y extracción de información para grandes bases de datos.
3.4. Discutir las problemáticas con los alumnos y especialistas en "Big data".

IV. PRE-REQUISITOS

Para cursar esta asignatura, que se ofrece en la carrera como Electiva, el estudiante deberá cumplir con los prerrequisitos establecidos según la electiva en la que se inscribe, de acuerdo con la siguiente tabla:

Alternativa	Porcentaje de créditos aprobados	Cantidad de créditos requeridos
Electiva 1	55%	184
Electiva 2, Electiva 3, Electiva 4, Electiva 5	70%	235
Electiva 6, Electiva 7	80%	268

V. CONTENIDO

5.1. Unidades programáticas

5.1.1. Los datos y la cuarta revolución industrial.
5.1.2. Introducción a Big data.
5.1.3. Programación con Python y Spark para Big data.
5.1.4. Extracción de conocimiento (KDD).
5.1.5. Gestión de datos masivos.
5.1.6. Tecnologías básicas de procesamiento.
5.1.7. Motores de datos.
5.1.8. Conceptos de Business Intelligence – BI.
5.1.9. Recuperación de la información desde repositorios de datos masivos.
5.1.10. Cuestiones Legales en el manejo de datos masivos.
5.1.11. Presentación de un Proyecto Final.

5.2. Desarrollo de las unidades programáticas

5.2.1. Los datos y la cuarta revolución industrial.

5.2.1.1 Re evolución digital.
5.2.1.2 Fuerzas digitales.
5.2.1.3 Impacto de la transformación digital.

5.2.2. Introducción a Big data
5.2.2.1. Concepto y características de Big data
5.2.2.2. Tecnologías utilizadas
5.2.2.3. Categorías de casos de usos principales

5.2.3. Programación con Python y Spark para Big data
5.2.3.1. Concepto de Programación
5.2.3.2. Aplicaciones usando el lenguaje Python

5.2.4. Extracción de conocimiento (KDD)
5.2.4.1. Análisis de las fases del proceso KDD
5.2.4.2. Análisis del aprendizaje supervisado y no supervisado sobre plataformas de Big data

5.2.5. Gestión de datos masivos
5.2.5.1. Principios básicos de almacenamiento de datos distribuidos: Teorema CAP, BaSE , replicación y sharding.

5.2.6. Tecnologías básicas de procesamiento
5.2.6.1. Introducción al ecosistema, Spark, Hadoop
5.2.6.2. Introducción al Paradigma Map- Reduce y Practica con HDFS

5.2.7. Motores de datos
5.2.7.1. Introducción a bases de datos NoSQL
5.2.7.2. Casos prácticos para aproximación a Apache, Cassandra, Mongo DB, Neo4J y HBase

5.2.8. Conceptos de Business Intelligence - BI
5.2.8.1. Conceptos y objetivos de BI, software de BI. Casos reales
5.2.8.2. Técnicas y herramientas para el análisis exploratorio de datos por medio de la visualización.

5.2.9. Recuperación de la información desde repositorios de datos masivos
5.2.9.1. Fundamentos de la recuperación de información. Importancia.
5.2.9.2. Casos prácticos

5.2.10. Cuestiones Legales en el manejo de datos masivos
5.2.10.1. Examinar las normativas que se aplican en el ámbito de Big data a nivel nacional e internacional.
5.2.10.2. Límites que se imponen y obligaciones que conlleva el manejo de datos masivos.

5.2.11. Presentación de un Proyecto Final

VI. ESTRATEGIAS METODOLÓGICAS

6.1. Exposición de fundamentos teóricos apoyados por medios audiovisuales.
6.2. Trabajos de investigación y evaluación de herramientas y técnicas.
6.3. Diseño y Programación de un Proyecto de Big Data
6.4. Trabajos en Laboratorio.

VII. MEDIOS AUXILIARES

7.1. Pizarras acrílicas.
7.2. Marcadores.
7.3. Borrador de pizarra acrílica.
7.4. Computadoras.
7.5. Equipo multimedia
7.6. Material bibliográfico.
7.7. Plataforma virtual "EDUCA".
7.8. Sala de laboratorio equipada para las prácticas.

VIII. EVALUACIÓN

La evaluación sobre el aprendizaje y conocimiento adquiridos por el alumno se realizará de acuerdo a lo establecido en el reglamento de la Facultad Politécnica de la UNA.

IX. BIBLIOGRAFÍA

- Marz, N., Warren, J. (2015). *Big Data: Principles and best practices of scalable realtime data systems*. United States: 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.

ENLACES WEB

- Apache Hive TM (jun. de 2019). Disponible en https://hive.apache.org/
- Apache Software Foundation (jun. de 2019). Apache Headoop. 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/
- Pyton TM. Programming Language. (jun. de 2019). Disponible en https://www.python.org/

DISPONIBLES EN LA COLECCIÓN DE LA BIBLIOTECA DE LA FACULTAD POLITÉCNICA

- Cuevas Álvarez, A. (2017). *Python 3: curso práctico*. Bogotá: RA-MA
- Daza Vergaray, A. (2017). *Data minign = Minería de datos*. Lima: Macro EIRL.
- Fernández Huerta, I. & Díez, P. E. (2018). *Fundamentos básicos de programación: aplicación práctica con Scratch y Python*. Madrid: Delta publicaciones.

Larose, D. T. & Larose, C. D. (2018). *Data mining and predictive analytics*. (2° Ed.). Nueva Delhi: Wiley India.