

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

Resolución N° 19/13/12-00 Acta N° 1054/08/07/2019 - ANEXO 01

I. IDENTIFICACIÓN

| | |
|------------------------------------|-------------------------|
| 1. Asignatura | : Electiva 7 - Big Data |
| 2. Semestre | : Décimo |
| 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 teóricas | : 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 GENERALES

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

IV. OBJETIVOS ESPECIFICOS

A. Conocimientos

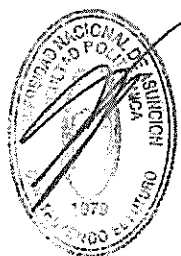
1. Identificar plataformas y sus arquitecturas para almacenamiento viable de datos a escala Big Data.
2. Identificar herramientas para el procesamiento de datos almacenados en plataformas Big Data.
3. Distinguir paradigmas de programación aplicables a plataformas orientadas a Big Data.
4. Aplicar algoritmos de Data mining típicamente aplicados a datos
5. Implementar algoritmos de Data mining sobre plataformas Big Data.
6. Describir lo que se puede extraer desde las distintas fuentes de datos Big Data y sus aplicaciones.

B. Habilidades

1. Evaluar y seleccionar las herramientas y plataformas más adecuadas según los requerimientos de almacenamiento, análisis de datos, integración, costos y gobierno de la infraestructura.
2. Dimensionar, costear y diseñar la arquitectura de Big Data en función a los requerimientos de almacenamiento y procesamiento
3. Diseñar procesos de obtención de datos y almacenamiento masivo en plataformas orientadas a Big Data.
4. Diseñar y programar procesos de análisis de datos a escala Big Data con altos niveles de seguridad de la información.
5. Diagnosticar y corregir problemas de rendimiento en procesos de análisis en la plataforma Big Data.
6. Detectar vulnerabilidades de seguridad en la plataforma de Big Data, y proponer acciones paliativas.

C. Competencias

1. Capacidad de aplicar los conocimientos en la práctica.
2. Disposición para el trabajo en equipo.
3. Capacidad de abstracción, análisis y síntesis y presentaciones orales.
4. Habilidades para buscar, procesar y analizar información procedente de fuentes diversas.
5. Capacidad para identificar, plantear y resolver problemas
6. Capacidad de comunicación oral y escrita.



V. PRE – REQUISITO

- Para cursar, el alumno debe haber aprobado el 80 % de los créditos.

VI. CONTENIDO**6.1. Unidades programáticas**

- 1- Los datos y la cuarta revolución industrial
- 2- Introducción a Big data
- 3- Programación con Python y Spark para Big data
- 4- Extracción de conocimiento (KDD)
- 5- Gestión de datos masivos
- 6- Tecnologías básicas de procesamiento
- 7- Motores de datos
- 8- Conceptos de Business Intelligence - BI
- 9- Recuperación de la información desde repositorios de datos masivos
- 10- Cuestiones Legales en el manejo de datos masivos
- 11- Presentación de un Proyecto Final

6.2. Desarrollo de las unidades programáticas**1. Los datos y la cuarta revolución industrial**

- 1.1. Re evolución digital
- 1.2. Fuerzas digitales
- 1.3. Impacto de la transformación digital

2. Introducción a Big data

- 2.1. Concepto y características de Big data
- 2.2. Tecnologías utilizadas
- 2.3. Categorías de casos de usos principales

3. Programación con Python y Spark para Big data

- 3.1. Concepto de Programación
- 3.2. Aplicaciones usando el lenguaje Python

4. Extracción de conocimiento (KDD)

- 4.1. Análisis de las fases del proceso KDD
- 4.2. Análisis del aprendizaje supervisado y no supervisado sobre plataformas de Big data

5. Gestión de datos masivos

- 5.1. Principios básicos de almacenamiento de datos distribuidos: Teorema CAP, BaSE, replicación y *sharding*.

6. Tecnologías básicas de procesamiento

- 6.1. Introducción al ecosistema, Spark, Hadoop
- 6.2. Introducción al Paradigma Map- Reduce y Practica con HDFS

7. Motores de datos

- 7.1. Introducción a bases de datos NoSQL
- 7.2. Casos prácticos para aproximación a Apache, Cassandra, Mongo DB, Neo4J y HBase

8. Conceptos de Business Intelligence - BI

- 8.1. Conceptos y objetivos de BI, software de BI. Casos reales
- 8.2. Técnicas y herramientas para el análisis exploratorio de datos por medio de la visualización.

9. Recuperación de la información desde repositorios de datos masivos

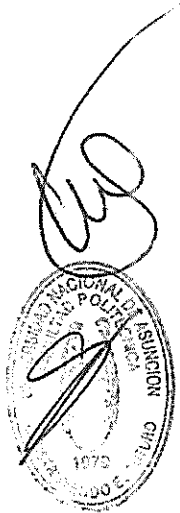
- 9.1. Fundamentos de la recuperación de información. Importancia.
- 9.2. Casos prácticos

10. Cuestiones Legales en el manejo de datos masivos

- 10.1. Examinar las normativas que se aplican en el ámbito de Big data a nivel nacional e internacional.
- 10.2. Límites que se imponen y obligaciones que conlleva el manejo de datos masivos.

11. Presentación de un Proyecto Final**VII. ESTRATEGIAS METODOLÓGICAS**

1. Exposición de fundamentos teóricos apoyados por medios audiovisuales.
2. Trabajos de investigación y evaluación de herramientas y técnicas.
3. Diseño y Programación de un Proyecto de Big Data
4. Trabajos en Laboratorio.



VIII. MEDIOS AUXILIARES

1. Pizarras acrílicas.
2. Marcadores.
3. Borrador de pizarra acrílica.
4. Computadoras.
5. Equipo multimedia
6. Material bibliográfico.
7. Plataforma virtual "EDUCA".
8. Sala de laboratorio equipada para las prácticas.

IX. EVALUACIÓN

1. Es imprescindible la entrega de todas las prácticas
2. Según Reglamentación y Normativas vigentes en la FPUNA

X. 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 Hadoop. 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/>
- Python 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 minigñ = 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.

