



MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN PROPUESTA DE CURSO DE POSGRADO

1- DATOS GENERALES DE LA ACTIVIDAD CURRICULAR					
1.1 Título del Curso		ALMACENAMIENTO Y PROCESAMIENTO PARA BIG DATA			
1.2 Área temática ¹		INGENIERÍA DE SOFTWARE			

2- COMPOSICION DEL EQUIPO DOCENTE					
2.1 Responsable a cargo de la actividad curricular		Dra. Agustina Buccella			
2.2 Docentes					

3- CARGA HORARIA						
Carga horaria teórica		15 hs				
Carga horaria práctica		25 hs				
Carga horaria total		40 hs				
Distribución horaria semanal		Lu	Ma	Mie	Jue	Vie
Fecha de inicio sugerida						

4- BREVE RESUMEN DE CONTENIDOS (hasta 400 palabras)					
Fundamentos de Big Data. Modelos y Frameworks para Big Data. Sistemas de almacenamiento distribuido. Procesamiento para grandes volúmenes de datos.					

5- CONOCIMIENTOS PREVIOS REQUERIDOS					
Cursos aprobados: 01, 02, 03, 04					

6- OBJETIVOS					
Conocer y profundizar la base metodológica para el desarrollo de aplicaciones de Big Data junto con las tendencias tecnológicas y conceptuales basadas en el almacenamiento y procesamiento de los datos. La asignatura tendrá por objetivo comprender y aplicar, con las tecnologías brindadas, las técnicas utilizadas en las actividades principales de las metodologías de Big Data que hoy son relevantes.					

7- CONTENIDOS (organizados en unidades, ejes, módulos, otros)					
--	--	--	--	--	--

¹ Corresponde a uno de los siguientes tópicos: Algoritmos y Lenguajes; Teoría de la Computación; Ingeniería de Software, Bases de Datos y Sistemas de Información; Arquitecturas, Sistemas Operativos y Redes.



PROGRAMA ANALÍTICO:

Unidad I: INTRODUCCIÓN A BIG DATA

Concepto de Big Data y Analíticas. Terminología. Características: Volumen, Velocidad, Variedad, Veracidad, Valor. Ciclo de Vida. Analítica de Datos: descriptiva, diagnóstica, predictiva y prescriptiva. Ejemplos. Casos de Estudio

Unidad II: ALMACENAMIENTO DE DATOS

Conceptos del almacenamiento de grandes volúmenes de datos. Diseño de almacenamiento distribuido. Tipos de fuentes y tipos de almacenamiento. Lago de Datos. Diseño de los Lagos de Datos. Tecnologías asociadas.

Unidad III: PROCESAMIENTO DE GRANDES VOLÚMENES DE DATOS

Conceptos y tipos de procesamiento: paralelo, distribuido, batch (map-reduce), transaccional, tiempo real. Operaciones distribuidas. Aplicaciones. Tecnologías asociadas.

Tecnologías para Big Data:

Unidad II: HDFS, MongoDB.

Unidad III: Apache Spark (RDD), Hadoop-MapReduce, YARN

8- PROPUESTA DIDÁCTICA (metodología de trabajo de clases teóricas y prácticas)

La asignatura se organiza en torno a clases teórico/prácticas. Las mismas inician desde conceptos teóricos generales hasta aquellos más específicos y complejos. Para las clases, los/as estudiantes deben disponer con anterioridad de la documentación básica que se explicará, así como de los ejercicios o prácticas asociadas. En general cada material teórico brindado posee una guía de los temas que se van dictando en cada clase, así como de la bibliografía utilizada.

Para fomentar la participación y facilitar la comprensión de temas, se propone realizar al finalizar cada contenido, ejercicios prácticos de laboratorio en donde los alumnos puedan aplicar conceptos vistos y participar activamente.

En cuanto a la práctica, cada unidad temática tiene asociado uno o más trabajos prácticos de elaboración individual o grupal. Las clases se centran en atención de consultas y resolución de ejercicios con participación activa de estudiantes.

Se plantean 3 trabajos prácticos para plasmar los conceptos teóricos vistos. El Práctico 1 está orientado al análisis de casos de estudio reales donde se plantean proyectos de big data detectando todo su ciclo de vida y las actividades realizadas. Luego el Práctico 2 se centra en los diferentes sistemas de almacenamiento de datos específicos para manipular grandes volúmenes de datos analizando en forma práctica las formas de manipular la información analizada junto con los objetivos, ventajas y desventajas de cada uno. Por último, en el Práctico 3 se trabaja con los conceptos de procesamiento distribuido aplicando ejercicios que muestren la dinámica de dividir y procesar la información en varios nodos.

A su vez, como los alumnos deben presentar reportes teórico/prácticos, los trabajos prácticos realizados conducen a la elaboración de dichos entregables.

9- MODALIDAD DE EVALUACIÓN Y CONDICIONES DE ACREDITACIÓN²

Mediante la realización de los trabajos prácticos propuestos en clase y la elaboración de un trabajo final que deberá ser realizado una vez entregado al finalizar el curso (cuyo tema y fecha de entrega se acordará con los participantes).

² Son condiciones mínimas para la aprobación de todos los cursos: cumplir con un mínimo del 80% de asistencia a las clases, realizar las tareas y aprobar las evaluaciones que se hayan propuesto en el programa, con una calificación no menor a 7 (puntos). Los trabajos de evaluación pautados y la calificación de los alumnos deberán realizarse dentro de los 60 días posteriores a la finalización del curso.



10- BIBLIOGRAFÍA DE LECTURA OBLIGATORIA CORRESPONDIENTE A CADA UNIDAD Y GENERAL

1. Big Data Fundamentals: Concepts, Drivers & Techniques. Thomas Erl, Wajid Khattak and Paul Buhler. 1st Edición. ISBN: 978-0134291079. Prentice Hall 2016
2. Big Data Analytics: A Hands-On Approach. Arshdeep Bahga and Vijay Madisetti. Published by Arshdeep Bahga & Vijay Madisetti. ISBN-10: 9781949978001. 2019

Herramientas:

<https://spark.apache.org/>

<https://hadoop.apache.org/>

<https://www.mongodb.com/>

11- INFRAESTRUCTURA E INSUMOS REQUERIDOS³

Proyector multimedia, pantalla, acceso a Internet

12 – OTRA INFORMACIÓN RELEVANTE

³Deberá constar aquí si la realización del curso requiere contar con instalaciones especiales (laboratorio, sala de informática, equipamiento audiovisual, etc). Explicitar si se estima que el curso debe tener un número máximo determinado de asistentes para poder ser dictado.