

BigData

Guía de asignatura

Última actualización: enero de 2023

1. Información general

Nombre de la asignatura	Big Data
Código	11310067
Tipo de asignatura	Electivo
Número de créditos	3AB
Tipo de crédito	2A + 1B (teórico-práctica)
Horas de trabajo con acompañamiento directo del profesor	80
Horas de trabajo independiente del estudiante	64
Prerrequisitos	Análisis Estadístico de Datos, Redes de Computadores
Correquisitos	Ninguno
Horario	M y J 13:00 – 15:00am
Líder de área	Pedro Wightman pedro.wightman@urosario.edu.co
Salón	Hipatia Claustro Torre 2 piso %

2. Información del profesor y monitor

Nombre del profesor	Ferney A Beltrán Molina
Perfil profesional	Ingeniero Electrónico de la Universidad Nacional de Colombia. Magister en TIC de la Universidad Rey Juan Carlos y candidato Doctor de la misma Universidad. Director del Centro de investigación e innovación CEINTECCI. Miembro de la junta directiva de la Asociación Colombiana para el Avance de la Ciencia, Avanciencia. Las investigaciones se enmarcan en el área de procesamiento y análisis de datos basadas en IA., simulación y modelado por computación, la optimización de los sistemas de procesamiento, tanto en hardware como en software y el diseño de sistemas electrónicos reconfigurables

Correo electrónico institucional	ferney.beltran@urosario.edu.co
Lugar y horario de atención	
Página web	
Nombre profesor auxiliar o monitor	
Correo electrónico institucional	
Lugar y horario de atención	
Página web, Skype u otros medios	

3. Resumen y propósitos del curso

Este curso presenta los fundamentos de las plataformas de Big Data, sus principios de diseño, implementación y operación. El curso busca que el estudiante esté en capacidad de diseñar, implementar y operar plataformas de Big Data para el consumo, almacenamiento y procesamiento de grandes cantidades de datos. El curso busca tanto que el estudiante comprenda los principios de diseño de las plataformas de Big Data, como que esté en capacidad de instalar plataformas ampliamente usadas actualmente incluyendo todas las capas de software asociadas.

El curso busca desarrollar habilidades en el estudiante para implementar plataformas de Big Data, incluyendo capacidades de análisis de datos en lotes y en línea. Se busca que el estudiante comprenda tanto los principios de diseño de una plataforma de Big Data como que esté en capacidad de incorporar este tipo de plataformas en productos de ciencia de datos.

4. Conceptos fundamentales

1. Estructura de almacenamiento y manejo de gran cantidad de datos.
2. Sistema de archivos distribuidos para Bigdata.
3. Análisis de datos masivos en lotes y streaming.
4. Procesamiento de datos masivos.
5. Conceptos de aprendizaje automático y bigData.

5. Resultados de aprendizaje esperados (RAE)

1. Comprender los principios de diseño de una plataforma de Big Data.
2. Instalar y configurar las capas fundamentales de una plataforma de Big Data.
3. Instalar y configurar las capas superiores de una plataforma de Big Data, asociadas a la manipulación y acceso de datos.
4. Instalar y configurar una arquitectura lambda para consumir y procesar datos en línea.
5. Conectar la plataforma de Big Data con otros servicios que permitan generar productos de analítica.

6. Modalidad del curso

Presencial. Todos sus estudiantes estarán presentes en el salón de clase. No se tendrá opción de conexión remota a las clases.

7. Estrategias de aprendizaje

La secuencia de aprendizaje que se trabajará en el curso es la siguiente:

1. Antes de clase: El estudiante debe estudiar el material propuesto por el profesor: Videos, notas de clase, capítulos de libro, documentos, etc.
2. Durante la clase: El profesor hace un resumen del tema durante el cual los estudiantes pueden resolver las dudas generadas durante el estudio del material previo o durante la explicación. Se desarrollan ejercicios prácticos relacionados con el tema que los estudiantes resolverán de manera individual o grupal según la guía de trabajo.
3. Después de clase: Los estudiantes resuelven en grupo o individualmente un ejercicios prácticos o talleres relacionado con el tema visto.
4. Proyecto de semestre, en el que los estudiantes, por grupos, aplican los conocimientos adquiridos en la solución de un problema de Bigdata.

8. Actividades de evaluación

Actividad de evaluación		Porcentaje
Evaluación 1	Evaluación teórica - práctica	15
	Autoevaluación y Coevaluación	5
Evaluación 2	Evaluación teórica - práctica	15
	Autoevaluación y Coevaluación	5
Evaluación 3	Evaluación teórico - práctica	10
	Autoevaluación y Coevaluación	5
Evaluación 4	Evaluación teórica - práctica	15
	Autoevaluación y Coevaluación	5
Proyecto Final	hetero-evaluación	15
	Autoevaluación y Coevaluación	10

9. Programación de actividades

Fecha	Tema	Evaluación
31-ene	Introducción a Big Data: conceptos fundamentales	
2-feb	Repaso Unix Linux	
7-feb	MapReduce	
9-feb	Laboratorio de MapReduce	Modelo de programación MapReduce
14-feb	introducción Hadoop y HDFS	

16-feb	Laboratorio Apache Hadoop y HDFS	Acceder a HDFS para manejar de datos distribuidos
21-feb	Repaso	
23-feb	Evaluación 1	Problemas en MapReduce para BigData
28-feb	Hadoop y YARN	
2-mar	laboratorio Hadoop YARN	Gestión de Recursos
7-mar	Conceptos y Programación Spark	
9-mar	Laboratorio Spark 1	Flujo interactivo para procesamiento
14-mar	Spark API de bajo nivel y API de alto nivel	
16-mar	Laboratorio Spark 2	Flujo interactivo para procesamiento
21-mar	Concepto de streaming	
23-mar	Laboratorio de sparkStreaming	Datos en tiempo real
28-mar	Repaso	
30-mar	Evaluación 2	Problema de Bigdata con spark
4-abr	SEMANA SANTA	
6-abr		
11-abr	introducción Cloud y Bigdata	
13-abr	Laboratorio Cloud 1	
18-abr	Tecnologías y Servicios en Cloud para BigData	
20-abr	Laboratorio Cloud 2	
25-abr	Repaso	
27-abr	Evaluación 3	Problemas de BigData en Cloud
2-may	Introducción a Machine Learning y BigData	
4-may	Laboratorio de Machine Learning	Implementación de ML en Bigdata
9-may	Machine learning en Spark	
11-may	Laboratorio de Machine learning en Spark	Implementacion de Mlib
16-may	Mlops	
18-may	Laboratorio de Machine Learning Mlops	
23-may	Repaso	
25-may	Evaluación 4	
30-may	Semana de finales	Presentación del proyecto
1-jun		

10. Factores de éxito para este curso

A continuación, se sugieren una serie de acciones que pueden contribuir, de manera significativa, con el logro de metas y consecuentemente propiciar una experiencia exitosa en este curso:

1. Planificar y organizar el tiempo de trabajo individual que le dedicará al curso

2. Organizar el sitio y los materiales de estudios
3. Tener un grupo de estudio, procurar el apoyo de compañeros
4. Cultivar la disciplina y la constancia, trabajar semanalmente, no permitir que se acumulen temas ni trabajos
5. Realizar constantemente una autoevaluación, determinar si las acciones realizadas son productivas o si por el contrario se debe cambiar de estrategias
6. Asistir a las horas de consulta del profesor, participar en clase, no quedarse nunca con la duda
7. Utilizar los espacios destinados para consultas y resolución de dudas, tales como la Sala Knuth
8. Propiciar espacios para el descanso y la higiene mental, procurar tener buenos hábitos de sueño
9. Tener presente en todo momento valores como la honestidad y la sinceridad, al final no se trata solo de aprobar un examen, se trata de aprender y adquirir conocimientos. El fraude es un autoengaño

11. Bibliografía y recursos

1. Erl, T. (2013). Big Data Fundamentals: Concepts, Drivers & Techniques. Prentice Hall.
2. White, Tom. (2015) Hadoop: The Definitive Guide. 4th edition. O'Reilly Media.
3. Chambers, B., & Zaharia, M. (2018). Spark: The Definitive Guide. O'Reilly Media.

12. Bibliografía y recursos complementarios

1. Marz, N., Warren, J. Big Data: Principles and best practices of scalable realtime data systems. Manning Publications. 2015.

13. Acuerdos para el desarrollo del curso

- No está permitido comer o usar dispositivos móviles dentro de clase
- Los estudiantes deben asistir al salón de clase en el horario establecido
- Los estudiantes podrán hacer intervenciones levantando la mano y deben respetar el orden establecido
- Las sesiones de clase podrán ser grabadas. Se contará con material de consulta y repaso en e-aulas que no pretende reemplazar la participación de los estudiantes en las clases

Todas las sesiones y actividades del curso son de carácter teórico-práctico; es decir, incluyen clases magistrales, discusiones, ejercicios o talleres.

Para todas las sesiones se espera que el estudiante realice un trabajo independiente previo que permite un avance continuo en los temas y facilite el entendimiento de los mismos.

No se realizará aproximación de notas al final del semestre. Las notas solo serán cambiadas con base en reclamos OPORTUNOS dentro de los límites de tiempo determinados por el Reglamento

Académico. Si por motivos de fuerza mayor el estudiante falta a algún parcial o quiz, deberá seguir el procedimiento regular determinado por el Reglamento Académico para presentar supletorios. No habrá acuerdos informales al respecto. No se eximirá a ningún estudiante de ningún examen.

ASISTENCIA AL CURSO

Con el propósito de afianzar el modelo pedagógico contemplado en el Proyecto Educativo Institucional y promover un rendimiento académico óptimo, es necesario asegurar un espacio de interacción entre estudiantes y profesores que facilite la reflexión y el debate académico en torno al conocimiento. En este sentido, se valora la participación en las actividades académicas y esta se considera como un deber y un derecho del estudiante. (Artículo 48 Reglamento Académico). **De no asistir a más del 80% de las clases el 15% se pierde con 0.0.**

Si el estudiante se presenta 20 minutos luego de dar inicio a alguna evaluación parcial o final, no podrá presentarla y deberá solicitar supletorio siguiendo la reglamentación institucional.

PROCESOS DISCIPLINARIOS-FRAUDE EN EVALUACIONES

Teniendo en cuenta el reglamento formativo-preventivo y disciplinario de la Universidad del Rosario, y la certeza de que las acciones fraudulentas van en contra de los procesos de enseñanza y aprendizaje, cualquier acto corrupto vinculado a esta asignatura será notificado a la secretaría académica correspondiente de manera que se inicie el debido proceso disciplinario. Se recomienda a los estudiantes leer dicho reglamento para conocer las razones, procedimientos y consecuencias que este tipo de acciones pueden ocasionar, así como sus derechos y deberes asociados a este tipo de procedimientos.

La asignatura no tiene ningún tipo de bono.

PROTOCOLO PARA LAS ACTIVIDADES CON EVALUACIÓN

Los siguientes lineamientos serán seguidos de forma estricta y sin excepción en actividades como quices, tareas y evaluaciones parciales.

1. Solamente será posible tener acceso a e-aulas.urosario.edu.co y a los sitios web correspondientes a la documentación correspondiente a cada temario de BigData dispuestos por el profesor.
2. Maletas, morrales, bolsos, etc. deben estar ubicados al frente del salón.
3. Celulares y otros dispositivos electrónicos deben estar apagados y ser guardados dentro de las maletas antes de ser ubicadas en su respectiva posición.
4. El estudiante no debe intentar ocultar ningún código que no sea propio en la solución a la actividad.
5. El estudiante solo podrá disponer de hojas en blanco como borrador de apuntes (opcional).
6. El estudiante puede tener una hoja manuscrita de resumen (opcional). Esta hoja debe estar marcada con nombre completo.

7. Los grupos pueden consultar entre ellos *las ideas básicas*; sin embargo, la solución y detalles del ejercicio debe realizarse *individualmente*. Cualquier tipo de fraude o plagio es causa de anulación directa de la evaluación y correspondiente proceso disciplinario.
8. El grupo de trabajo debe indicar en su entrega de la solución a la actividad cualquier asistencia que haya recibido.
9. El grupo no debe consultar ninguna solución de la solución a la actividad que no sea la suya.
10. El grupo no debe intentar ocultar ningún código que no sea propio en la solución a la actividad.
11. E-aulas se cerrará a la hora en punto acordada para el final de la evaluación. La solución de la actividad debe ser subida antes de esta hora. El material entregado a través de e-aulas será calificado tal como está. Si ningún tipo de material es entregado por este medio, la nota de la evaluación será 0.0.

Se aconseja subir a e-aulas versiones parciales de la solución a la actividad.

- Todas las evaluaciones serán probadas en la herramienta que defina el docente. Si el estudiante usa una herramienta diferente, se atiene a la existencia de problemas con librerías, compiladores, etc.
- Todas las entregas están sujetas a herramientas automatizadas de detección de plagio en códigos.
- La evaluación debe presentarse exclusivamente en uno de los computadores ubicados en el salón de clase y a la hora acordada. Presentar la evaluación desde otro dispositivo o en otro horario diferente al estipulado es causa de anulación, a no ser que sea autorizado por el docente.

14. Respeto y no discriminación

Si tiene alguna discapacidad, sea este visible o no, y requiere algún tipo de apoyo para estar en igualdad de condiciones con los(as) demás estudiantes, por favor informar a su profesor(a) para que puedan realizarse ajustes razonables al curso a la mayor brevedad posible. De igual forma, si no cuenta con los recursos tecnológicos requeridos para el desarrollo del curso, por favor informe de manera oportuna a la Secretaría Académica de su programa o a la Dirección de Estudiantes, de manera que se pueda atender a tiempo su requerimiento.

Recuerde que es deber de todas las personas respetar los derechos de quienes hacen parte de la comunidad Rosarista. Cualquier situación de acoso, acoso sexual, discriminación o matoneo, sea presencial o virtual, es inaceptable. Quien se sienta en alguna de estas situaciones puede denunciar su ocurrencia contactando al equipo de la Coordinación de Psicología y Calidad de Vida de la Decanatura del Medio Universitario (Teléfono o WhatsApp 322 2485756).