



1. CURSO

IIS64000. Big Data y Analítica de Datos (Obligatorio)

2. INFORMACIÓN GENERAL

- 2.1 Créditos : 4
- 2.2 Horas de teoría : -
- 2.3 Horas de práctica : 2 (Semanal)
- 2.4 Horas autónomas : 128 (horas)
- 2.5 Duración del periodo : 16 semanas
- 2.6 Condición : Obligatorio
- 2.7 Modalidad : Presencial
- 2.8 Prerrequisitos :
 - SFW52090. Gerenciamiento de Datos II. (5^{to} Sem)
 - SFW62071. Computación Paralela y Distribuída. (8^{vo} Sem)

3. PROFESORES

Atención previa coordinación con el profesor

4. INTRODUCCIÓN AL CURSO

En la actualidad conocer enfoques escalables para procesar y almacenar grande volúmenes de información (terabytes, petabytes e inclusive exabytes) es fundamental en cursos de ciencia de la computación. Cada día, cada hora, cada minuto se genera gran cantidad de información la cual necesita ser procesada, almacenada, analizada.

5. OBJETIVOS

- Que el alumno sea capaz de crear aplicaciones paralelas para procesar grandes volúmenes de información.
- Que el alumno sea capaz de comparar las alternativas para el procesamiento de big data.
- Que el alumno sea capaz de proponer arquitecturas para una aplicación escalable.

6. COMPETENCIAS

- 3) Justifica un desempeño individual, como parte de equipos de trabajo o como líder de proyectos de grupos multidisciplinarios en entornos globales con el fin de asegurar la calidad de software, aplicando normas, configuraciones, regulaciones y métricas. (**Evaluar**)
- 4) Diseña soluciones de Software de acuerdo a los estándares y políticas de seguridad de la información en uno o varios dominios de aplicación siendo socialmente responsables y demostrando ética profesional. (**Evaluar**)

7. TEMAS

Unidad 1: Introducción a Big Data (15 horas)	
Competencias esperadas:	
Temas	Objetivos de Aprendizaje
<ul style="list-style-type: none"> • Visión global sobre Cloud Computing • Visión global sobre Sistema de Archivos Distribuidos • Visión global sobre el modelo de programación MapReduce 	<ul style="list-style-type: none"> • Explicar el concepto de Cloud Computing desde el punto de vista de Big Data[Familiarizarse] • Explicar el concepto de los Sistema de Archivos Distribuidos [Familiarizarse] • Explicar el concepto del modelo de programación MapReduce[Familiarizarse]
Aprendizaje autónomo	
<ul style="list-style-type: none"> • Desarrollo de ejercicios prácticos 	
Lecturas : [Cou+11]	

Unidad 2: Hadoop (15 horas)	
Competencias esperadas:	
Temas	Objetivos de Aprendizaje
<ul style="list-style-type: none"> • Visión global de Hadoop. • Historia. • Estructura de Hadoop. • HDFS, Hadoop Distributed File System. • Modelo de Programación MapReduce 	<ul style="list-style-type: none"> • Entender y explicar la suite de Hadoop. [Familiarizarse] • Implementar soluciones usando el modelo de programación MapReduce. [Usar] • Entender la forma como se guardan los datos en el HDFS. [Familiarizarse]
Aprendizaje autónomo	
<ul style="list-style-type: none"> • Desarrollo de ejercicios prácticos 	
Lecturas : [HDF11], [BVS13]	

Unidad 3: Procesamiento de Grafos en larga escala (10 horas)	
Competencias esperadas:	
Temas	Objetivos de Aprendizaje
<ul style="list-style-type: none"> • Pregel: A System for Large-scale Graph Processing. • Distributed GraphLab: A Framework for Machine Learning and Data Mining in the Cloud. • Apache Giraph is an iterative graph processing system built for high scalability. 	<ul style="list-style-type: none"> • Entender y explicar la arquitectura del proyecto Pregel. [Familiarizarse] • Entender la arquitectura del proyecto GraphLab. [Familiarizarse] • Entender la arquitectura del proyecto Giraph. [Familiarizarse] • Implementar soluciones usando Pregel, GraphLab o Giraph. [Usar]
Aprendizaje autónomo	
<ul style="list-style-type: none"> • Desarrollo de ejercicios prácticos 	
Lecturas : [Low+12], [Mal+10], [Bal+08]	

8. PLAN DE TRABAJO

8.1 Metodología

Se fomenta la participación individual y en equipo para exponer sus ideas, motivándolos con puntos adicionales en las diferentes etapas de la evaluación del curso.

8.2 Sesiones Teóricas

Las sesiones de teoría se llevan a cabo en clases magistrales donde se realizarán actividades que propicien un aprendizaje activo, con dinámicas que permitan a los estudiantes interiorizar los conceptos.

8.3 Sesiones Prácticas

Las sesiones prácticas se llevan en clase donde se desarrollan una serie de ejercicios y/o conceptos prácticos mediante planteamiento de problemas, la resolución de problemas, ejercicios puntuales y/o en contextos aplicativos.

9. SISTEMA DE EVALUACIÓN

Cada uno de los rubros del esquema de evaluación y la nota final del curso son redondeados a números enteros. La nota final del curso es el promedio ponderado de los rubros correspondientes: evaluación permanente, examen parcial y examen final.

Los promedios calculados componentes del rubro 'Evaluación Permanente' mantendrán su cálculo con 2 decimales.

	%	Observaciones	Semana	Rezagable
Evaluación Continua	70%			
Práctica Calificada	70%			
Práctica Calificada ₁		Se elimina la práctica con la menor nota	4	No
Práctica Calificada ₂		Se elimina la práctica con la menor nota	8	No
Práctica Calificada ₃		Se elimina la práctica con la menor nota	12	No
Proyecto	30%		15	
Examen final	30%			

10. BIBLIOGRAFÍA BÁSICA

- [Bal+08] Shumeet Baluja et al. "Video Suggestion and Discovery for Youtube: Taking Random Walks Through the View Graph". In: *Proceedings of the 17th International Conference on World Wide Web. WWW '08*. Beijing, China: ACM, 2008, pp. 895–904. ISBN: 978-1-60558-085-2. DOI: 10.1145/1367497.1367618. URL: <http://doi.acm.org/10.1145/1367497.1367618>.
- [BVS13] Rajkumar Buyya, Christian Vecchiola, and S. Thamarai Selvi. *Mastering Cloud Computing: Foundations and Applications Programming*. 1st. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2013. ISBN: 9780124095397, 9780124114548.
- [Cou+11] George Coulouris et al. *Distributed Systems: Concepts and Design*. 5th. USA: Addison-Wesley Publishing Company, 2011. ISBN: 0132143011, 9780132143011.
- [HDF11] Kai Hwang, Jack Dongarra, and Geoffrey C. Fox. *Distributed and Cloud Computing: From Parallel Processing to the Internet of Things*. 1st. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011. ISBN: 0123858801, 9780123858801.
- [Low+12] Yucheng Low et al. "Distributed GraphLab: A Framework for Machine Learning and Data Mining in the Cloud". In: *Proc. VLDB Endow*. 5.8 (Apr. 2012), pp. 716–727. ISSN: 2150-8097. DOI: 10.14778/2212351.2212354. URL: <http://dx.doi.org/10.14778/2212351.2212354>.
- [Mal+10] Grzegorz Malewicz et al. "Pregel: A System for Large-scale Graph Processing". In: *ACM SIGMOD Record*. SIGMOD '10 (2010), pp. 135–146. DOI: 10.1145/1807167.1807184. URL: <http://doi.acm.org/10.1145/1807167.1807184>.