

## 1. CURSO

CS370. Big Data (Obligatorio)

## 2. INFORMACIÓN GENERAL

- 2.1 Créditos : 3
- 2.2 Horas de teoría : 1 (Semanal)
- 2.3 Horas de práctica : 2 (Semanal)
- 2.4 Duración del periodo : 16 semanas
- 2.5 Condición : Obligatorio
- 2.6 Modalidad : Presencial
- 2.7 Prerrequisitos :
  - CS272. Bases de Datos II. (5<sup>to</sup> Sem)
  - CS3P1. Computación Paralela y Distribuída. (8<sup>vo</sup> Sem)

## 3. PROFESORES

Atención previa coordinación con el profesor

## 4. INTRODUCCIÓN AL CURSO

En la actualidad conocer enfoques escalables para procesar y almacenar grandes 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

- a) Aplicar conocimientos de computación y de matemáticas apropiadas para la disciplina. (**Usar**)
- b) Analizar problemas e identificar y definir los requerimientos computacionales apropiados para su solución. (**Usar**)
- i) Utilizar técnicas y herramientas actuales necesarias para la práctica de la computación. (**Usar**)
- j) Aplicar la base matemática, principios de algoritmos y la teoría de la CS en el modelamiento y diseño de sistemas. (**Usar**)

## 7. COMPETENCIAS ESPECÍFICAS

■NoSpecificOutcomes■

## 8. TEMAS

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

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

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

## 9. PLAN DE TRABAJO

### 9.1 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.

### 9.2 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.

### 9.3 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.

## 10. SISTEMA DE EVALUACIÓN

\*\*\*\*\* EVALUATION MISSING \*\*\*\*\*

## 11. 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>.