

**Universidad Católica San Pablo (UCSP)**  
**Escuela Profesional de**  
**Ciencia de la Computación**  
**SILABO**



**CS370. Big Data (Obligatorio)**

**1. Información general**

1.1 Escuela	:	Ciencia de la Computación
1.2 Curso	:	CS370. Big Data
1.3 Semestre	:	9 <sup>no</sup> Semestre.
1.4 Prerrequisitos	:	<ul style="list-style-type: none"><li>• CS272. Bases de Datos II. (5<sup>to</sup> Sem)</li><li>• CS3P1. Computación Paralela y Distribuída. (8<sup>vo</sup> Sem)</li></ul>
1.5 Condición	:	Obligatorio
1.6 Modalidad de aprendizaje	:	Presencial
1.7 horas	:	1 HT; 2 HP; 2 HL;
1.8 Créditos	:	3

**2. Profesores**

**3. Fundamentación del 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.

**4. Resumen**

1. Introducción a Big Data 2. Hadoop 3. Procesamiento de Grafos en larga escala

**5. Objetivos Generales**

- 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. Contribución a los resultados (Outcomes)**

Esta disciplina contribuye al logro de los siguientes resultados de la carrera:

- 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. Contenido**

<b>UNIDAD 1: Introducción a Big Data (15)</b>	
<b>Competencias: a,b,i</b>	
<b>Contenido</b>	<b>Objetivos Generales</b>
<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>
<b>Lecturas:</b> Coulouris et al. (2011)	

<b>UNIDAD 2: Hadoop (15)</b>	
<b>Competencias: a,b,i</b>	
<b>Contenido</b>	<b>Objetivos Generales</b>
<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>
<b>Lecturas:</b> Hwang, Dongarra, and Fox (2011), Buyya, Vecchiola, and Selvi (2013)	

<b>UNIDAD 3: Procesamiento de Grafos en larga escala (10)</b>	
<b>Competencias: a,b,i</b>	
<b>Contenido</b>	<b>Objetivos Generales</b>
<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>
<b>Lecturas:</b> Low et al. (2012), Malewicz et al. (2010), Baluja et al. (2008)	

<b>8. Metodología</b>
<p>El profesor del curso presentará clases teóricas de los temas señalados en el programa propiciando la intervención de los alumnos.</p> <p>El profesor del curso presentará demostraciones para fundamentar clases teóricas.</p> <p>El profesor y los alumnos realizarán prácticas</p> <p>Los alumnos deberán asistir a clase habiendo leído lo que el profesor va a presentar. De esta manera se facilitará la comprensión y los estudiantes estarán en mejores condiciones de hacer consultas en clase.</p>

**Evaluación Continua 1** : 20 %

**Examen parcial** : 30 %

**Evaluación Continua 2** : 20 %

**Examen final** : 30 %

## References

- Baluja, Shumeet et al. (2008). “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. ACM: Beijing, China, pp. 895–904. ISBN: 978-1-60558-085-2. DOI: 10.1145/1367497.1367618.
- Buyya, Rajkumar, Christian Vecchiola, and S. Thamarai Selvi (2013). *Mastering Cloud Computing: Foundations and Applications Programming*. 1st. Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA. ISBN: 9780124095397, 9780124114548.
- Coulouris, George et al. (2011). *Distributed Systems: Concepts and Design*. 5th. Addison-Wesley Publishing Company: USA. ISBN: 0132143011, 9780132143011.
- Hwang, Kai, Jack Dongarra, and Geoffrey C. Fox (2011). *Distributed and Cloud Computing: From Parallel Processing to the Internet of Things*. 1st. Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA. ISBN: 0123858801, 9780123858801.
- Low, Yucheng et al. (Apr. 2012). “Distributed GraphLab: A Framework for Machine Learning and Data Mining in the Cloud”. In: *Proc. VLDB Endow.* 5(8), pp. 716–727. ISSN: 2150-8097. DOI: 10.14778/2212351.2212354.
- Malewicz, Grzegorz et al. (2010). “Pregel: A System for Large-scale Graph Processing”. In: *ACM SIGMOD Record*. SIGMOD '10, pp. 135–146. DOI: 10.1145/1807167.1807184.