Universidad Nacional San Cristobal de Huamanga (UNSCH)
Programa Profesional de
Ciencia de la Computación
Sílabo 2024-II

1. CURSO

CS370. Big Data (Mandatory)

2. INFORMACIÓN GENERAL

2.1 Curso	:	CS370. Big Data
2.2 Semestre	:	9^{th} Semester.
2.3 Créditos	:	3
2.4 horas	:	1 HT; 4 HP;
2.5 Duración del periodo	:	16 semanas
2.6 Condición	:	Mandatory
2.7 Modalidad de aprendizaje	:	Face to face
2.8 Prerrequisitos	:	

- CS272. Databases II. (5^{th} Sem)
- CS3P1. Parallel and Distributed Computing . (8^{th} Sem)
- CS272. Databases II. (5^{th} Sem)
- CS3P1. Parallel and Distributed Computing . (8^{th} Sem)

3. PROFESORES

Atención previa coordinación con el profesor

4. INTRODUCCIÓN AL CURSO

Nowadays, knowing scalable approaches to processing and storing large volumes of information (terabytes, petabytes and even exabytes) is fundamental in computer science courses. Every day, every hour, every minute generates a large amount of information which needs to be processed, stored, analyzed.

5. OBJETIVOS

- That the student is able to create parallel applications to process large volumes of information
- That the student is able to compare the alternatives for the processing of big data
- That the student is able to propose architectures for a scalable application

6. RESULTADOS DEL ESTUDIANTE

1) Analyze a complex computing problem and apply principles of computing and other relevant disciplines to identify solutions. (Usage)

6) Apply computer science theory and software development fundamentals to produce computing-based solutions. (Usage)

7. TEMAS

Unidad 1: Introducción a Big Data (15)	
Resultados esperados:	
Temas	**Objetivos de Aprendizaje (*Learning Outcomes*)**
• Overview on Cloud Computing • Distributed File System Overview • Overview of the MapReduce programming model	• Explain the concept of Cloud Computing from the point of view of Big Data[Familiarizarse] • Explain the concept of Distributed File System [Familiarizarse] • Explain the concept of the MapReduce programming model[Familiarizarse]
Lecturas : [Cou+11]	

Unidad 2: Hadoop (15)	
Resultados esperados:	
Temas	**Objetivos de Aprendizaje (*Learning Outcomes*)**
• Hadoop overview. • History. • Hadoop Structure. • HDFS, Hadoop Distributed File System. • Programming Model MapReduce	• Understand and explain the Hadoop suite [Familiarizarse] • Implement solutions using the MapReduce programming model. [Usar] • Understand how data is saved in the HDFS. [Familiarizarse]
Lecturas : [HDF11], [BVS13]	

Unidad 3: Procesamiento de Grafos en larga escala (10)	
Resultados esperados:	
Temas	**Objetivos de Aprendizaje (*Learning Outcomes*)**
• 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.	• Understand and explain the architecture of the Pregel project. [Familiarizarse] • Understand the GraphLab project architecture. [Familiarizarse] • Understand the architecture of the Giraph project. [Familiarizarse] • Implement solutions using Pregel, GraphLab or Giraph. [Usar]
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

********* EVALUATION MISSING ********

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. DOI: 10.1145/1367497.1367618. URL: http://doi.acm.org/10.1145/1367497.1367618.

[Mal+10] Grzegorz Malewicz et al. "Pregel: A System for Large-scale Graph Processing". In: SIGMOD '10 (2010), pp. 135–146. DOI: 10.1145/1807167.1807184. URL: http://doi.acm.org/10.1145/1807167.1807184.

[Cou+11] George Coulouris et al. *Distributed Systems: Concepts and Design*. 5th. USA: Addison-Wesley Publishing Company, 2011.

[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.

[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. DOI: 10.14778/2212351.2212354. URL: http://dx.doi.org/10.14778/2212351.2212354.

[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.