



Universidad Nacional de Ingeniería (UNI)

Escuela Profesional de
Ciencia de la Computación
Sílabo 2024-II

1. CURSO

CS370. Big Data (Obligatorio)

2. INFORMACIÓN GENERAL

- 2.1 Curso : CS370. Big Data
- 2.2 Semestre : 9^{no} Semestre.
- 2.3 Créditos : 3
- 2.4 horas : 1 HT; 4 HP;
- 2.5 Duración del periodo : 16 semanas
- 2.6 Condición : Obligatorio
- 2.7 Modalidad de aprendizaje : Presencial
- 2.8 Prerrequisitos :
 - CS272. Bases de Datos II. (5^{to} Sem)
 - CS3P1. 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. 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 horas)	
Resultados esperados:	
Temas	Objetivos de Aprendizaje (<i>Learning Outcomes</i>)
<ul style="list-style-type: none">• Introducción a Big Data• Visión global sobre Big Data	<ul style="list-style-type: none">• Entender los conceptos relacionados de Big Data[Familiarizarse]
Lecturas : [Cou+11]	

Unidad 2: Sistemas Distribuidos (15 horas)	
Resultados esperados:	
Temas	Objetivos de Aprendizaje (<i>Learning Outcomes</i>)
<ul style="list-style-type: none"> • Sistemas Distribuidos. • Sistemas de Archivos Distribuidos. • Ambiente de Programación en Unix. 	<ul style="list-style-type: none"> • Entender los conceptos relacionados a los Sistemas Distribuidos. [Familiarizarse] • Entender los conceptos relacionados a los Sistemas de Archivos Distribuidos. [Familiarizarse] • Entender los conceptos relacionados a la gestión de datos usando la programación en Unix. [Usar]
Lecturas : [HDF11], [BVS13]	

Unidad 3: Procesamiento de Big Data (10 horas)	
Resultados esperados:	
Temas	Objetivos de Aprendizaje (<i>Learning Outcomes</i>)
<ul style="list-style-type: none"> • Introducción al framework Hadoop. • Sistema de Archivo Distribuido de Hadoop. • Introducción al modelo de programación MapReduce. 	<ul style="list-style-type: none"> • Entender los conceptos relacionados al framework Hadoop. [Familiarizarse] • Entender los conceptos relacionados al Sistema de Archivo Distribuido de Hadoop. [Familiarizarse] • Entender y aplicar el modelo de programación MapReduce. [Usar]
Lecturas : [yarn]	

Unidad 4: Procesamiento de Big Data II (10 horas)	
Resultados esperados:	
Temas	Objetivos de Aprendizaje (<i>Learning Outcomes</i>)
<ul style="list-style-type: none"> • Introducción al framework Spark. • Spark SQL. • Spark ML. 	<ul style="list-style-type: none"> • Entender los conceptos relacionados al framework Spark. [Familiarizarse] • Entender y aplicar el módulo del framework Spark SQL. [Usar] • Entender y aplicar el módulo del framework Spark ML. [Usar]
Lecturas : [spark]	

Unidad 5: Procesamiento de Stream (10 horas)	
Resultados esperados:	
Temas	Objetivos de Aprendizaje (<i>Learning Outcomes</i>)
<ul style="list-style-type: none"> • Visión global de herramientas para procesamiento de stream. • Procesamiento de stream y la nube. 	<ul style="list-style-type: none"> • Entender los conceptos relacionados al procesamiento de stream. [Familiarizarse] • Conocer las herramientas para procesamiento de stream. [Usar]
Lecturas : [spark]	

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

- [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.
- [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.