



## Universidad Nacional de Ingeniería (UNI)

Programa Profesional de  
Inteligencia Artificial  
Sílabo 2024-I

### 1. CURSO

CS370. Big Data (Obligatorio)

### 2. INFORMACIÓN GENERAL

- 2.1 Curso : CS370. Big Data
- 2.2 Semestre : 9<sup>no</sup> 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<sup>to</sup> Sem)
  - CS3P1. Computación Paralela y Distribuida. (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. RESULTADOS DEL ESTUDIANTE

- 1) Analizar un problema computacional complejo y aplicar los principios computacionales y otras disciplinas relevantes para identificar soluciones. (**Usar**)
- 6) Aplicar la teoría de la computación y los fundamentos del desarrollo de software para producir soluciones basadas en computación. (**Usar**)

### 7. TEMAS

Unidad 1: Introducción a Big Data (15 horas)	
Resultados esperados: 1,6	
Temas	Objetivos de Aprendizaje
<ul style="list-style-type: none"><li>• Introducción a Big Data</li><li>• Visión global sobre Big Data</li></ul>	<ul style="list-style-type: none"><li>• Entender los conceptos relacionados de Big Data[Familiarizarse]</li></ul>
Lecturas : [Cou+11]	

<b>Unidad 2: Sistemas Distribuidos (15 horas)</b>	
<b>Resultados esperados: 1,6</b>	
<b>Temas</b>	<b>Objetivos de Aprendizaje</b>
<ul style="list-style-type: none"> <li>• Sistemas Distribuidos.</li> <li>• Sistemas de Archivos Distribuidos.</li> <li>• Ambiente de Programación en Unix.</li> </ul>	<ul style="list-style-type: none"> <li>• Entender los conceptos relacionados a los Sistemas Distribuidos. [Familiarizarse]</li> <li>• Entender los conceptos relacionados a los Sistemas de Archivos Distribuidos. [Familiarizarse]</li> <li>• Entender los conceptos relacionados a la gestión de datos usando la programación en Unix. [Usar]</li> </ul>
<b>Lecturas :</b> [HDF11], [BVS13]	

<b>Unidad 3: Procesamiento de Big Data (10 horas)</b>	
<b>Resultados esperados: 1,6</b>	
<b>Temas</b>	<b>Objetivos de Aprendizaje</b>
<ul style="list-style-type: none"> <li>• Introducción al framework Hadoop.</li> <li>• Sistema de Archivo Distribuido de Hadoop.</li> <li>• Introducción al modelo de programación MapReduce.</li> </ul>	<ul style="list-style-type: none"> <li>• Entender los conceptos relacionados al framework Hadoop. [Familiarizarse]</li> <li>• Entender los conceptos relacionados al Sistema de Archivo Distribuido de Hadoop. [Familiarizarse]</li> <li>• Entender y aplicar el modelo de programación MapReduce. [Usar]</li> </ul>
<b>Lecturas :</b> [yarn]	

<b>Unidad 4: Procesamiento de Big Data II (10 horas)</b>	
<b>Resultados esperados: 1,6</b>	
<b>Temas</b>	<b>Objetivos de Aprendizaje</b>
<ul style="list-style-type: none"> <li>• Introducción al framework Spark.</li> <li>• Spark SQL.</li> <li>• Spark ML.</li> </ul>	<ul style="list-style-type: none"> <li>• Entender los conceptos relacionados al framework Spark. [Familiarizarse]</li> <li>• Entender y aplicar el módulo del framework Spark SQL. [Usar]</li> <li>• Entender y aplicar el módulo del framework Spark ML. [Usar]</li> </ul>
<b>Lecturas :</b> [spark]	

<b>Unidad 5: Procesamiento de Stream (10 horas)</b>	
<b>Resultados esperados: 1,6</b>	
<b>Temas</b>	<b>Objetivos de Aprendizaje</b>
<ul style="list-style-type: none"> <li>• Visión global de herramientas para procesamiento de stream.</li> <li>• Procesamiento de stream y la nube.</li> </ul>	<ul style="list-style-type: none"> <li>• Entender los conceptos relacionados al procesamiento de stream. [Familiarizarse]</li> <li>• Conocer las herramientas para procesamiento de stream. [Usar]</li> </ul>
<b>Lecturas :</b> [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

- [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.