

**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; 4 HP;
1.8 Créditos	:	3
1.9 Plan	:	Plan Curricular 2016

**2. Profesores**

**Titular**

- Alvaro Henry Mamani-Aliaga <ahmamani@ucsp.edu.pe>
  - Doctor en Ciencia de la Computación, UNSA, Perú, 2019.
  - Master en Ciencia de la Computación, IME-USP, Brasil, 2011.

**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. Sistemas Distribuidos 3. Procesamiento de Big Data 4. Procesamiento de Big Data II 5. Procesamiento de Stream

**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:

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

## 7. Contenido

UNIDAD 1: Introducción a Big Data (15)	
Resultados del estudiante: 1,6	
Contenido	Objetivos Generales
<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>
<b>Lecturas:</b> Coulouris et al. (2011)	

UNIDAD 2: Sistemas Distribuidos (15)	
Resultados del estudiante: 1,6	
Contenido	Objetivos Generales
<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> Hwang, Dongarra, and Fox (2011), Buyya, Vecchiola, and Selvi (2013)	

UNIDAD 3: Procesamiento de Big Data (10)	
Resultados del estudiante: 1,6	
Contenido	Objetivos Generales
<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	

UNIDAD 4: Procesamiento de Big Data II (10)	
Resultados del estudiante: 1,6	
Contenido	Objetivos Generales
<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	

UNIDAD 5: Procesamiento de Stream (10)	
Resultados del estudiante: 1,6	
Contenido	Objetivos Generales
<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: spark</b>	

## 8. Metodología

1. El profesor del curso presentará clases teóricas de los temas señalados en el programa propiciando la intervención de los alumnos.
2. El profesor del curso presentará demostraciones para fundamentar clases teóricas.
3. El profesor y los alumnos realizarán prácticas
4. 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.

## 9. Evaluar 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.

## 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.

## Sistema de Evaluación:

La nota final se obtiene a través de:

EVALUACIONES PERMANENTES	EVALUACIONES
<b>Evaluación Permanente 1</b> : 24 %	<b>Evaluación Parcial</b> : 20 %
<b>Evaluación Permanente 2</b> : 36 %	<b>Evaluación Final</b> : 20 %
60%	40%

Donde:

Evaluación Permanente: Comprende trabajos grupales, participación activa en clase, test de ejercicios.

- Permanente 1 (Semanas 1 - 9)
- Permanente 2 (Semanas 10 - 17)

Para aprobar el curso, el alumno debe obtener 11.5 o más en la nota final.

## References

- 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.