

San Pablo Catholic University (UCSP)
Undergraduate Program in
Computer Science
SILABO



CS3P2. Cloud Computing (Mandatory)

1. General information

1.1 School	:	Ciencia de la Computación
1.2 Course	:	CS3P2. Cloud Computing
1.3 Semester	:	10 ^{mo} Semestre.
1.4 Prerequisites	:	CS370. Big Data. (9 th Sem)
1.5 Type of course	:	Mandatory
1.6 Learning modality	:	Face to face
1.7 Horas	:	1 HT; 2 HP; 2 HL;
1.8 Credits	:	3

2. Professors

3. Course foundation

In order to understand the advanced computational techniques, the students must have a strong knowledge of the various discrete structures, structures that will be implemented and used in the laboratory in the programming language.

4. Summary

1. Distributed Systems 2. Cloud Computing 3. Centros de Procesamiento de Datos 4. Cloud Computing 5. Cloud Computing 6. Modelos de Programación

5. Generales Goals

- That the student is able to model computer science problems using graphs and trees related to data structures.
- That the student apply efficient travel strategies to be able to search data in an optimal way.

6. Contribution to Outcomes

This discipline contributes to the achievement of the following outcomes:

- a) An ability to apply knowledge of mathematics, science. (**Usage**)
- b) An ability to design and conduct experiments, as well as to analyze and interpret data. (**Usage**)
- i) An ability to use the techniques, skills, and modern computing tools necessary for computing practice. (**Usage**)
- j) Apply the mathematical basis, principles of algorithms and the theory of Computer Science in the modeling and design of computational systems in such a way as to demonstrate understanding of the equilibrium points involved in the chosen option. (**Usage**)

7. Content

UNIT 1: Distributed Systems (15)	
Competences: a,b	
Content	Generales Goals
<ul style="list-style-type: none"> • Faults (cross-reference OS/Fault Tolerance) <ul style="list-style-type: none"> – Network-based (including partitions) and node-based failures – Impact on system-wide guarantees (e.g., availability) • Distributed message sending <ul style="list-style-type: none"> – Data conversion and transmission – Sockets – Message sequencing – Buffering, retrying, and dropping messages • Distributed system design tradeoffs <ul style="list-style-type: none"> – Latency versus throughput – Consistency, availability, partition tolerance • Distributed service design <ul style="list-style-type: none"> – Stateful versus stateless protocols and services – Session (connection-based) designs – Reactive (IO-triggered) and multithreaded designs • Core distributed algorithms <ul style="list-style-type: none"> – Election, discovery 	<ul style="list-style-type: none"> • Distinguish network faults from other kinds of failures [Familiarity] • Explain why synchronization constructs such as simple locks are not useful in the presence of distributed faults [Familiarity] • Write a program that performs any required marshalling and conversion into message units, such as packets, to communicate interesting data between two hosts [Usage] • Measure the observed throughput and response latency across hosts in a given network [Usage] • Explain why no distributed system can be simultaneously consistent, available, and partition tolerant [Familiarity] • Implement a simple server – for example, a spell checking service [Usage] • Explain the tradeoffs among overhead, scalability, and fault tolerance when choosing a stateful v stateless design for a given service [Familiarity] • Describe the scalability challenges associated with a service growing to accommodate many clients, as well as those associated with a service only transiently having many clients [Familiarity] • Give examples of problems for which consensus algorithms such as leader election are required [Usage]
Readings: Coulouris et al. (2011)	

UNIT 2: Cloud Computing (15)	
Competences: a,b	
Content	Generales Goals
<ul style="list-style-type: none"> • Visión global de <i>Cloud Computing</i>. • Historia. • Visión global de las tecnologías que envuelve. • Beneficios, riesgos y aspectos económicos. • Cloud services <ul style="list-style-type: none"> – Infrastructure as a service <ul style="list-style-type: none"> * Elasticity of resources * Platform APIs – Software as a service – Security – Cost management • Internet-Scale computing <ul style="list-style-type: none"> – Task partitioning – Data access – Clusters, grids, and meshes 	<ul style="list-style-type: none"> • Explicar el concepto de Cloud Computing. [Familiarity] • Listar algunas tecnologías relacionadas con Cloud Computing. [Familiarity] • Explain strategies to synchronize a common view of shared data across a collection of devices [Familiarity] • Discutir las ventajas y desventajas del paradigma de Cloud Computing. [Familiarity] • Expresar los beneficios económicos así como las características y riesgos del paradigma de Cloud para negocios y proveedores de cloud. [Familiarity] • Diferenciar entre los modelos de servicio. [Usage]
Readings: Hwang, Dongarra, and Fox (2011), Buyya, Vecchiola, and Selvi (2013)	

UNIT 3: Centros de Procesamiento de Datos (10)	
Competences: i	
Content	Generales Goals
<ul style="list-style-type: none"> • Visión global de un centro de procesamiento de datos. • Consideraciones en el diseño. • Comparación de actuales grandes centros de procesamiento de datos. 	<ul style="list-style-type: none"> • Describir la evolución de los Data Centers. [Familiarity] • Esbozar la arquitectura de un data center en detalle. [Familiarity] • Indicar consideraciones de diseño y discutir su impacto. [Familiarity]
Readings: Hwang, Dongarra, and Fox (2011), Buyya, Vecchiola, and Selvi (2013)	

UNIT 4: Cloud Computing (20)	
Competences: i,j	
Content	Generales Goals
<ul style="list-style-type: none"> • Virtualization <ul style="list-style-type: none"> – Shared resource management – Migration of processes • Seguridad, recursos y aislamiento de fallas. • Almacenamiento como servicio. • Elasticidad. • Xen y Wmware. • Amazon EC2. 	<ul style="list-style-type: none"> • Virtualization <ul style="list-style-type: none"> – Shared resource management – Migration of processes . [Familiarity] • Explain the advantages and disadvantages of using virtualized infrastructure. [Familiarity] • Identificar las razones por qué la virtualización está llegando a ser enorme útil, especialmente en la cloud. [Familiarity] • Explicar diferentes tipos de aislamiento como falla, recursos y seguridad proporcionados por la virtualización y utilizado por la cloud. [Familiarity] • Explicar la complejidad que puede tener el administrar en términos de niveles de abstracción y interfaces bien definidas y su aplicabilidad para la virtualización en la cloud. [Familiarity] • Definir virtualización y identificar diferentes tipos de máquinas virtuales. [Familiarity] • Identificar condiciones de virtualización de CPU, reconocer la diferencia entre <i>full virtualization</i> y <i>paravirtualization</i>, explicar emulación como mayor técnica para virtualización del CPU y examinar planificación virtual del CPU en Xen. [Familiarity] • Esbozar la diferencia entre la clásica memoria virtual del SO y la virtualización de memoria. Explicar los múltiples niveles de mapeamiento de páginas en oposición a la virtualización de la memoria. Definir memoria <i>over-commitment</i> e ilustrar sobre Wmware <i>memory ballooning</i> como técnica de reclamo para sistemas virtualizados con memoria <i>over-committed</i>. [Familiarity]
Readings: Hwang, Dongarra, and Fox (2011), Buyya, Vecchiola, and Selvi (2013)	

UNIT 5: Cloud Computing (12)	
Competences: i,j	
Content	Generales Goals
<ul style="list-style-type: none"> • Cloud-based data storage <ul style="list-style-type: none"> – Shared access to weakly consistent data stores – Data synchronization – Data partitioning – Distributed file systems – Replication • Visión global sobre tecnologías de almacenamiento. • Conceptos fundamentales sobre almacenamiento en la cloud. • Amazon S3 y EBS. • Sistema de archivos distribuidos. • Sistema de bases de datos NoSQL. 	<ul style="list-style-type: none"> • Describir la organización general de datos y almacenamiento. [Familiarity] • Identificar los problemas de escalabilidad y administración de la big data. Discutir varias abstracciones en almacenamiento. [Familiarity] • Comparar y contrastar diferentes tipos de sistema de archivos. Comparar y contrastar el Sistema de Archivos Distribuido de Hadoop (HDFS) y el Sistema de Archivos Paralelo Virtual (PVFS). [Usage] • Comparar y contrastar diferentes tipos de bases de datos. Discutir las ventajas y desventajas sobre las bases de datos NoSQL. [Usage] • Discutir los conceptos de almacenamiento en la cloud. [Familiarity]
Readings: Hwang, Dongarra, and Fox (2011), Buyya, Vecchiola, and Selvi (2013)	

UNIT 6: Modelos de Programación (12)	
Competences: j	
Content	Generales Goals
<ul style="list-style-type: none"> • Visión global de los modelos de programación basados en cloud computing. • Modelo de Programación MapReduce. • Modelo de programación para aplicaciones basadas en Grafos. 	<ul style="list-style-type: none"> • Explicar los aspectos fundamentales de los modelos de programación paralela y distribuida. [Familiarity] • Diferencias entre los modelos de programación: MapReduce, Pregel, GraphLab y Giraph. [Usage] • Explicar los principales conceptos en el modelo de programación MapReduce. [Usage]
Readings: Hwang, Dongarra, and Fox (2011), Buyya, Vecchiola, and Selvi (2013), Low et al. (2012), Malewicz et al. (2010), Baluja et al. (2008)	

8. Methodology
<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>

9. Assessment

Continuous Assessment 1 : 20 %

Partial Exam : 30 %

Continuous Assessment 2 : 20 %

Final exam : 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: *Proc. ACM SIGMOD*. SIGMOD '10, pp. 135–146. DOI: 10.1145/1807167.1807184.