

Universidad Católica San Pablo (UCSP)
Escuela Profesional de
Ciencia de la Computación
SILABO



CS3P2. Cloud Computing (Obligatorio)

1. Información general

1.1 Escuela	:	Ciencia de la Computación
1.2 Curso	:	CS3P2. Cloud Computing
1.3 Semestre	:	10 ^{mo} Semestre.
1.4 Prerrequisitos	:	CS370. Big Data. (9 ^{no} Sem)
1.5 Condición	:	Obligatorio
1.6 Modalidad de aprendizaje	:	Virtual
1.7 horas	:	1 HT; 2 HP; 2 HL;
1.8 Créditos	:	3

2. Profesores

3. Fundamentación del curso

Para entender las técnicas computacionales avanzadas, los estudiantes deberán tener un fuerte conocimiento de las diversas estructuras discretas, estructuras que serán implementadas y usadas en laboratorio en el lenguaje de programación.

4. Resumen

1. Sistemas distribuidos 2. Cloud Computing 3. Centros de Procesamiento de Datos 4. Cloud Computing 5. Cloud Computing 6. Modelos de Programación

5. Objetivos Generales

- Que el alumno sea capaz de modelar problemas de ciencia de la computación usando grafos y árboles relacionados con estructuras de datos.
- Que el alumno aplicar eficientemente estrategias de recorrido para poder buscar datos de una manera óptima.

6. Contribución a los resultados (*Outcomes*)

Esta disciplina contribuye al logro de los siguientes resultados de la carrera:

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

7. Contenido

UNIDAD 1: Sistemas distribuidos (15)	
Competencias:	
Contenido	Objetivos Generales
<ul style="list-style-type: none"> • Fallos: <ul style="list-style-type: none"> – Fallos basados en red (incluyendo particiones) y fallos basados en nodos – Impacto en garantías a nivel de sistema (p.e., disponibilidad) • Envío de mensajes distribuido: <ul style="list-style-type: none"> – Conversión y transmisión de datos – Sockets – Secuenciamiento de mensajes – Almacenando <i>Buffering</i>, reenviando y desechando mensajes • Compensaciones de diseño para Sistemas Distribuidos: <ul style="list-style-type: none"> – Latencia versus rendimiento – Consistencia, disponibilidad, tolerancia de particiones • Diseño de Servicio Distribuido: <ul style="list-style-type: none"> – Protocolos y servicios Stateful versus stateless – Diseños de Sesión (basados en la conexión) – Diseños reactivos (provocados por E/S) y diseños de múltiples hilos • Algoritmos de Distribución de Núcleos: <ul style="list-style-type: none"> – Elección, descubrimiento 	<ul style="list-style-type: none"> • Distinguir las fallas de red de otros tipos de fallas [Familiarizarse] • Explicar por qué estructuras de sincronización como cerraduras simples (<i>locks</i>) no son útiles en la presencia de fallas distribuidas [Familiarizarse] • Escribir un programa que realiza cualquier proceso de <i>marshalling</i> requerido y la conversión en unidades de mensajes, tales como paquetes, para comunicar datos importantes entre dos <i>hosts</i> [Usar] • Medir el rendimiento observado y la latencia de la respuesta a través de los <i>hosts</i> en una red dada [Usar] • Explicar por qué un sistema distribuido no puede ser simultáneamente Consistente (<i>Consistent</i>), Disponible (<i>Available</i>) y Tolerante a fallas (<i>Partition tolerant</i>). [Familiarizarse] • Implementar un servidor sencillo - por ejemplo, un servicio de corrección ortográfica [Usar] • Explicar las ventajas y desventajas entre: <i>overhead</i>, escalabilidad y tolerancia a fallas entre escoger un diseño sin estado (<i>stateless</i>) y un diseño con estado (<i>stateful</i>) para un determinado servicio [Familiarizarse] • Describir los desafíos en la escalabilidad, asociados con un servicio creciente para soportar muchos clientes, así como los asociados con un servicio que tendrá transitoriamente muchos clientes [Familiarizarse] • Dar ejemplos de problemas donde algoritmos de consenso son requeridos, por ejemplo, la elección de líder [Usar]
Lecturas: Coulouris et al. (2011)	

UNIDAD 2: Cloud Computing (15)	
Competencias:	
Contenido	Objetivos Generales
<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. ● Servicios en la nube. <ul style="list-style-type: none"> – Infraestructura como servicio <ul style="list-style-type: none"> * Elasticidad de recursos * APIs de la Plataforma – Software como servicio – Seguridad – Administración del Costo ● Computación a Escala de Internet: <ul style="list-style-type: none"> – Particionamiento de Tareas – Acceso a datos – Clusters, grids y mallas 	<ul style="list-style-type: none"> ● Explicar el concepto de Cloud Computing. [Familiarizarse] ● Listar algunas tecnologías relacionadas con Cloud Computing. [Familiarizarse] ● Explicar las estrategias para sincronizar una vista común de datos compartidos a través de una colección de dispositivos [Familiarizarse] ● Discutir las ventajas y desventajas del paradigma de Cloud Computing. [Familiarizarse] ● Expresar los beneficios económicos así como las características y riesgos del paradigma de Cloud para negocios y proveedores de cloud. [Familiarizarse] ● Diferenciar entre los modelos de servicio. [Usar]
Lecturas: Hwang, Dongarra, and Fox (2011), Buyya, Vecchiola, and Selvi (2013)	

UNIDAD 3: Centros de Procesamiento de Datos (10)	
Competencias:	
Contenido	Objetivos Generales
<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. [Familiarizarse] ● Esbozar la arquitectura de un data center en detalle. [Familiarizarse] ● Indicar consideraciones de diseño y discutir su impacto. [Familiarizarse]
Lecturas: Hwang, Dongarra, and Fox (2011), Buyya, Vecchiola, and Selvi (2013)	

UNIDAD 4: Cloud Computing (20)	
Competencias:	
Contenido	Objetivos Generales
<ul style="list-style-type: none"> • Virtualización. <ul style="list-style-type: none"> – Gestión de recursos compartidos – Migración de procesos • Seguridad, recursos y aislamiento de fallas. • Almacenamiento como servicio. • Elasticidad. • Xen y Wmware. • Amazon EC2. 	<ul style="list-style-type: none"> • Virtualización. <ul style="list-style-type: none"> – Gestión de recursos compartidos – Migración de procesos . [Familiarizarse] • Explicar las ventajas y desventajas de usar una infraestructura virtualizada. [Familiarizarse] • Identificar las razones por qué la virtualización está llegando a ser enormemente útil, especialmente en la cloud. [Familiarizarse] • Explicar diferentes tipos de aislamiento como falla, recursos y seguridad proporcionados por la virtualización y utilizado por la cloud. [Familiarizarse] • 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. [Familiarizarse] • Definir virtualización y identificar diferentes tipos de máquinas virtuales. [Familiarizarse] • 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. [Familiarizarse] • 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>. [Familiarizarse]
Lecturas: Hwang, Dongarra, and Fox (2011), Buyya, Vecchiola, and Selvi (2013)	

UNIDAD 5: Cloud Computing (12)	
Competencias:	
Contenido	Objetivos Generales
<ul style="list-style-type: none"> • Almacenamiento de datos en la nube: <ul style="list-style-type: none"> – Acceso compartido a data stores de consistencia débil – Sincronización de datos – Particionamiento de datos – Sistemas de Archivos Distribuidos – Replicación • 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. [Familiarizarse] • Identificar los problemas de escalabilidad y administración de la big data. Discutir varias abstracciones en almacenamiento. [Familiarizarse] • 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). [Usar] • Comparar y contrastar diferentes tipos de bases de datos. Discutir las ventajas y desventajas sobre las bases de datos NoSQL. [Usar] • Discutir los conceptos de almacenamiento en la cloud. [Familiarizarse]
Lecturas: Hwang, Dongarra, and Fox (2011), Buyya, Vecchiola, and Selvi (2013)	

UNIDAD 6: Modelos de Programación (12)	
Competencias:	
Contenido	Objetivos Generales
<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. [Familiarizarse] • Diferencias entre los modelos de programación: MapReduce, Pregel, GraphLab y Giraph. [Usar] • Explicar los principales conceptos en el modelo de programación MapReduce. [Usar]
Lecturas: Hwang, Dongarra, and Fox (2011), Buyya, Vecchiola, and Selvi (2013), Low et al. (2012), Malewicz et al. (2010), Baluja et al. (2008)	

8. Metodología
<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. Evaluar

Evaluación Continua 1 : 20 %

Examen parcial : 30 %

Evaluación Continua 2 : 20 %

Examen final : 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.