

1. CURSO

CS3P2. Cloud Computing (Obligatorio)

2. INFORMACIÓN GENERAL

2.1 Créditos	:	3
2.2 Horas de teoría	:	1 (Semanal)
2.3 Horas de práctica	:	2 (Semanal)
2.4 Duración del periodo	:	16 semanas
2.5 Condición	:	Obligatorio
2.6 Modalidad	:	Presencial
2.7 Prerrequisitos	:	CS370. Big Data. (9 ^{no} Sem)

3. PROFESORES

Atención previa coordinación con el profesor

4. INTRODUCCIÓN AL 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.

5. OBJETIVOS

- 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. COMPETENCIAS

- a) Aplicar conocimientos de computación y de matemáticas apropiadas para la disciplina. (**Usar**)
- b) Analizar problemas e identificar y definir los requerimientos computacionales apropiados para su solución. (**Usar**)
- g) Analizar el impacto local y global de la computación sobre los individuos, organizaciones y sociedad. (**Usar**)
- i) Utilizar técnicas y herramientas actuales necesarias para la práctica de la computación. (**Usar**)
- j) Aplicar la base matemática, principios de algoritmos y la teoría de la CS en el modelamiento y diseño de sistemas. (**Usar**)

7. COMPETENCIAS ESPECÍFICAS

■NoSpecificOutcomes■

8. TEMAS

Unidad 1: Sistemas distribuidos (15)	
Competencias esperadas: a,b	
Temas	Objetivos de Aprendizaje
<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 : [Cou+11]	

Unidad 2: Cloud Computing (15)	
Competencias esperadas: a,b	
Temas	Objetivos de Aprendizaje
<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 <i>Cloud Computing</i>. [Familiarizarse] • Listar algunas tecnologías relacionadas con <i>Cloud Computing</i>. [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 <i>Cloud Computing</i>. [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 : [HDF11], [BVS13]	

Unidad 3: Centros de Procesamiento de Datos (10)	
Competencias esperadas: g,i	
Temas	Objetivos de Aprendizaje
<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 <i>Data Centers</i>. [Familiarizarse] • Esbozar la arquitectura de un data center en detalle. [Familiarizarse] • Indicar consideraciones de diseño y discutir su impacto. [Familiarizarse]
Lecturas : [HDF11], [BVS13]	

Unidad 4: Cloud Computing (20)	
Competencias esperadas: i,j	
Temas	Objetivos de Aprendizaje
<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 nube. [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 nube. [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 : [HDF11], [BVS13]	

Unidad 5: Cloud Computing (12)	
Competencias esperadas: i,j	
Temas	Objetivos de Aprendizaje
<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 : [HDF11], [BVS13]	

Unidad 6: Modelos de Programación (12)	
Competencias esperadas: g,j	
Temas	Objetivos de Aprendizaje
<ul style="list-style-type: none"> ● Visión global de los modelos de programación basados en <i>cloud computing</i>. ● 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 : [HDF11], [BVS13], [Low+12], [Mal+10], [Bal+08]	

9. PLAN DE TRABAJO

9.1 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.

9.2 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.

9.3 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.

10. SISTEMA DE EVALUACIÓN

***** EVALUATION MISSING *****

11. BIBLIOGRAFÍA BÁSICA

- [Bal+08] Shumeet Baluja et al. “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. Beijing, China: ACM, 2008, pp. 895–904. ISBN: 978-1-60558-085-2. DOI: 10.1145/1367497.1367618. URL: <http://doi.acm.org/10.1145/1367497.1367618>.
- [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.
- [Low+12] Yucheng Low et al. “Distributed GraphLab: A Framework for Machine Learning and Data Mining in the Cloud”. In: *Proc. VLDB Endow.* 5.8 (Apr. 2012), pp. 716–727. ISSN: 2150-8097. DOI: 10.14778/2212351.2212354. URL: <http://dx.doi.org/10.14778/2212351.2212354>.
- [Mal+10] Grzegorz Malewicz et al. “Pregel: A System for Large-scale Graph Processing”. In: *Proc. ACM SIGMOD. SIGMOD '10* (2010), pp. 135–146. DOI: 10.1145/1807167.1807184. URL: <http://doi.acm.org/10.1145/1807167.1807184>.