



Universidad Nacional Mayor de San Marcos
Escuela Profesional de
Ciencia de la Computación
Silabo del curso
Periodo Académico 2018-II

1. **Código del curso y nombre:** CS3P02. Cloud Computing (Obligatorio)
2. **Créditos:** 3
3. **Horas de Teoría y Laboratorio:** 1 HT; 4 HL; (15 semanas)
4. **Docente(s)**

Atención previa coordinación con el profesor

5. Bibliografía

- [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>.

6. Información del curso

- (a) **Breve descripción 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.
- (b) **Prerrequisitos:** CS3700. Big Data. (9^{no} Sem)
- (c) **Tipo de Curso:** Obligatorio
- (d) **Modalidad:** Presencial

7. Objetivos específicos del curso.

- 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.

8. Contribución a los resultados (*Outcomes*)

- 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 Ciencia de la Computación en el modelamiento y diseño de sistemas computacionales de tal manera que demuestre comprensión de los puntos de equilibrio involucrados en la opción escogida. (**Usar**)

9. Competencias (IEEE)

- C2.** Capacidad para tener una perspectiva crítica y creativa para identificar y resolver problemas utilizando el pensamiento computacional.⇒ **Outcome a**
- C4.** Una comprensión del hardware de la computadora desde la perspectiva del software, por ejemplo, el uso del procesador, memoria, unidades de disco, pantalla, etc.⇒ **Outcome b**
- C16.** Capacidad para identificar temas avanzados de computación y de la comprensión de las fronteras de la disciplina.⇒ **Outcome g,i**
- CS2.** Identificar y analizar los criterios y especificaciones apropiadas a los problemas específicos, y planificar estrategias para su solución.⇒ **Outcome i**
- CS3.** Analizar el grado en que un sistema basado en el ordenador cumple con los criterios definidos para su uso actual y futuro desarrollo.⇒ **Outcome j**
- CS6.** Evaluar los sistemas en términos de atributos de calidad en general y las posibles ventajas y desventajas que se presentan en el problema dado.⇒ **Outcome g,j**

10. Lista de temas a estudiar en el curso

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

11. Metodología y Evaluación

Metodología:

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 de Laboratorio:

Para verificar que los alumnos hayan alcanzado el logro planteado para cada una de las unidades de aprendizaje, realizarán actividades que les permita aplicar los conocimientos adquiridos durante las sesiones de teoría y se les propondrá retos que permitan evaluar el desempeño de los alumnos.

Exposiciones individuales o grupales:

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.

Lecturas:

A lo largo del curso se proporcionan diferentes lecturas, las cuales son evaluadas. El promedio de las notas de las lecturas es considerado como la nota de una práctica calificada. El uso del campus virtual UTEC Online permite a cada estudiante acceder a la información del curso, e interactuar fuera de aula con el profesor y con los otros estudiantes.

Sistema de Evaluación:

12. Contenido

| Unidad 1: Sistemas distribuidos (15) | |
|---|---|
| Competencias esperadas: C2, C4 | |
| Objetivos de Aprendizaje | Tópicos |
| <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] | <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 |
| Lecturas : [Cou+11] | |

| Unidad 2: Cloud Computing (15) | |
|---|---|
| Competencias esperadas: C2, C4 | |
| Objetivos de Aprendizaje | Tópicos |
| <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] | <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 |
| Lecturas : [HDF11], [BVS13] | |

| Unidad 3: Centros de Procesamiento de Datos (10) | |
|---|--|
| Competencias esperadas: C16 | |
| Objetivos de Aprendizaje | Tópicos |
| <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] | <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. |
| Lecturas : [HDF11], [BVS13] | |

| Unidad 4: Cloud Computing (20) | |
|---|--|
| Competencias esperadas: CS2, CS3 | |
| Objetivos de Aprendizaje | Tópicos |
| <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 VMware <i>memory ballooning</i> como técnica de reclamo para sistemas virtualizados con memoria <i>over-committed</i>. [Familiarizarse] | <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 VMware. • Amazon EC2. |
| Lecturas : [HDF11], [BVS13] | |

| Unidad 5: Cloud Computing (12) | |
|---|--|
| Competencias esperadas: CS2, CS3 | |
| Objetivos de Aprendizaje | Tópicos |
| <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] | <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. |
| Lecturas : [HDF11], [BVS13] | |

| Unidad 6: Modelos de Programación (12) | |
|---|--|
| Competencias esperadas: CS6 | |
| Objetivos de Aprendizaje | Tópicos |
| <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] | <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. |
| Lecturas : [HDF11], [BVS13], [Low+12], [Mal+10], [Bal+08] | |