
Curso Académico: (2024 / 2025)**Fecha de revisión: 24-04-2024**

Departamento asignado a la asignatura: Departamento de Ingeniería Telemática**Coordinador/a: DIAZ SANCHEZ, DANIEL****Tipo: Optativa Créditos ECTS : 6.0****Curso : 4 Cuatrimestre :**

REQUISITOS (ASIGNATURAS O MATERIAS CUYO CONOCIMIENTO SE PRESUPONE)

Arquitectura de Sistemas (o tener familiaridad con el trabajo en el laboratorio)
Aplicaciones Telemáticas (o tener conocimientos de protocolos superiores, concretamente HTTP)

OBJETIVOS

El objetivo de este curso es que el estudiante adquiera conocimientos para entender el funcionamiento y hacer uso de los sistemas modernos de computación distribuida que posibilitan muchos de los servicios populares que son ofrecidos y consumidos por millones de usuarios en todo el mundo, como pueden ser Dropbox, Spotify, Youtube, Google Search Engine

La filosofía de esta asignatura es facilitar el aprendizaje práctico de la computación distribuida a partir del uso de interfaces reales de servicios conocidos y actuales. Para lograr este objetivo, el alumno debe adquirir una serie de conocimientos y capacidades.

Conocimientos

Los conocimientos adquiridos al cursar esta asignatura son:

1. Conocer de la estructura de un sistema de comunicaciones distribuidas modernas, sus características internas y sus interfaces con las aplicaciones móviles, de escritorio y otros sistemas
2. Conocer los mecanismos básicos de comunicación en sistemas distribuidos mediante protocolos usados actualmente para interactuar con ellos como HTTP para interfaces REST o WebServices
3. Conocer los sistemas modernos de distribución de carga desde Enterprise Service Bus a los modernos sistemas de Map Reduce usados, por ejemplo, por Google, Amazon y muchos otros.
4. Conocer escenarios reales para aplicaciones web con muchas visitas (amazon, periódicos), arquitecturas de cloud para el desarrollo de aplicaciones y/o almacenamiento.
5. Conocer las ventajas y desventajas de estos sistemas y cuando una determinada aplicación debe ser implementada de forma distribuida (escalado de recursos y economía de centro de datos). Conocer casos reales de data sets que necesitan procesamiento distribuido.

Capacidades específicas

Las capacidades específicas que adquirirá el alumno al finalizar el curso:

1. Utilizar mecanismos básicos de comunicación en sistemas distribuidos.
2. Conocer aplicaciones centralizadas y distribuidas en red mediante prácticas demostrativas
3. Modelar y desplegar un sistema distribuido y hacer uso de aplicaciones y clouds existentes

Capacidades generales

En cuanto a las capacidades generales que adquirirá:

1. Habilidad para aplicar conocimientos de tecnologías de telecomunicación e ingeniería. Esta capacidad se trabajará especialmente en las prácticas de laboratorio así como en la resolución de ejercicios en las clases teóricas.
2. Habilidad para usar técnicas y herramientas de ingeniería necesarias para el ejercicio profesional
3. Habilidad para comunicarse de forma efectiva tanto de manera oral, escrita o gráfica tanto en español como en inglés a lo largo del desarrollo de las actividades propuestas en la asignatura (ejercicios, trabajos sobre nuevas tecnologías, etc.).
4. Reconocimiento de la necesidad de un aprendizaje continuo y la habilidad de obtener y aplicar la información requerida accediendo a literatura técnica relacionada con el ámbito de la asignatura tanto en español como en inglés
5. Conocimiento de nuevas tecnologías y tendencias en el campo de estudio

DESCRIPCIÓN DE CONTENIDOS: PROGRAMA

Resumen

La asignatura se divide aproximadamente a la mitad entre laboratorio y teoría. Se contemplan tres bloques de contenido:

- 1) Cloud: en la que se analizan las arquitecturas que han llevado a la nube, los modelos actuales. En el laboratorio se usa AWS, Google Cloud
- 2) Comunicaciones y la nube: representación de datos, uso de REST, colas y otros protocolos para la interacción, envío para persistencia de datos y extracción de datos (construcción de datasets). En el laboratorio se usa AWS, Google Cloud, Spotify u otros sistemas reales.
- 3) Big data: se analizan las necesidades, el tipo de nodo (hardware) necesario, los sistemas de almacenamiento distribuido, dimensionamiento y arquitectura de centros de datos. En el laboratorio se usa HDFS y Hadoop para entender los principios del Location Aware y MapReduce (una sola ronda - nodos de cómputo y almacenamiento), Spark para introducir el uso de múltiples rondas de MapReduce, inmutabilidad y restricciones que impone la distribución sobre la manipulación de los datos (nodos de cómputo y RAM y porqué técnicamente facilita Machine Learning), Spark para streaming (nodos de cómputo y RAM y la necesidad de un broker a la entrada - kafka u otros por escalabilidad), uso de programación de alto nivel soportada por un cluster

Programa Computación en la Nube

Introducción a la computación distribuida: Evolución de la computación, Sistemas legados y evolución a sistemas distribuidos, Sistemas Distribuidos y Cloud Computing, Modelos de distribución de cómputo, ¿Qué es cloud computing?, Arquitecturas respecto al despliegue (IaaS, PaaS, SaaS), Aspectos de seguridad y localización en Cloud, Retos y oportunidades, casos de uso, LABS: AWS EC2 (IaaS), Google Cloud Compute Engine (IaaS), Google App Engine (PaaS), AWS Lambda (PaaS-serverless)

Comunicaciones y la nube: Protocolos legados, Protocolos actuales (síncronos, asíncronos, mensajería, colas), Representación de los datos, Captura y extracción de datos, Consideraciones arquitectónicas y de red, Retos y oportunidades, casos de uso, LABS: Introducción a REST con FLASK (Google Cloud), Extracción programática vía API (Twitter, Spotify), Extracción forzada de datos vía Scrapping (local), MQTT y otros protocolos IoT/M2M (local)

Big Data: Arquitectura de sistemas de Big Data, Análisis de características hardware de procesamiento/almacenamiento, Anatomía de un centro de datos

Sistemas modernos de almacenamiento distribuido, Sistemas modernos de procesado en lotes, Sistemas modernos de procesado de flujos, Sistemas altamente distribuidos, Retos y oportunidades, casos de uso, LABS: HDFS/HADOOP (cluster local), Big data spark batch (cluster local), Big data spark stream (cluster local), Algoritmos en spark, Spark y cuadernos (pySpark)

ACTIVIDADES FORMATIVAS, METODOLOGÍA A UTILIZAR Y RÉGIMEN DE TUTORÍAS

Las actividades utilizadas para verificar las competencias y destrezas en el curso son :

- Durante la primera mitad del curso se realizarán prácticas guiadas donde se modificará código proporcionado por el profesorado para adquirir de forma gradual las distintas destrezas necesarias, todas ellas relacionadas con el estudio de un caso de estudio presentado al principio de la asignatura para conocer los sistemas Cloud modernos como los proporcionados por Google y Amazon
- Búsqueda de documentos auxiliares para completar la información que se estudia en un tema. En su informe final, deben mencionar las fuentes de información utilizadas .
- Uso de diferentes herramientas para la construcción de su proyecto software: máquinas virtuales, IDEs y controlador de versiones en las sesiones de laboratorio.
- Ejercicios sobre los temas vistos durante el curso

Durante estas actividades la plantilla docente revisa el trabajo de los estudiantes en la clase, supervisa las sesiones de laboratorio, responde a las preguntas en el foro del curso, mantiene tutorías semanales en el despacho y convoca tutorías con múltiples alumnos cuando lo considera oportuno.

SISTEMA DE EVALUACIÓN

Peso porcentual del Examen Final:	0
Peso porcentual del resto de la evaluación:	100

La evaluación de la adquisición de competencias se llevará a cabo mediante diversas pruebas durante el periodo lectivo.

Evaluación teórica (sobre 2 puntos):

- Examen escrito que evaluará la adquisición de la parte de conceptos teóricos necesarios para la adquisición de las competencias enumeradas anteriormente y su madurez a la hora de exponer y presentar sus conocimientos de forma apropiada y con un vocabulario correcto y depurado.

Evaluación práctica (sobre 8 puntos):

- Trabajo en clase (2 puntos) cuya estructura estará diseñada para, al mismo tiempo, guiar a los alumnos en la adquisición de las competencias necesarias, realizar un seguimiento de su aprendizaje y evaluar su conocimiento de la parte procedimental y así poder atestiguar la adquisición de dichas competencias. Se valorarán presentaciones o trabajos en grupo para discutir aplicaciones de la tecnología presentada.
- Proyecto de laboratorio en parejas (6 puntos) donde los alumnos deberán demostrar que dominan y que pueden aplicar los conceptos prácticos de la asignatura. Se evaluará mediante presentación y defensa durante la semana de entrega de trabajos (al finalizar el cuatrimestre) o el día del examen.

BIBLIOGRAFÍA BÁSICA

- Tom White Hadoop : the definitive guide, O'Reilly, 2009
- George F. Coulouris Distributed systems : concepts and design, Addison-Wesley, 2005

RECURSOS ELECTRÓNICOS BÁSICOS

- Chambers, Bill ; Zaharia, Matei . Spark: Big Data Processing Made Simple:
<https://learning.oreilly.com/library/view/spark-the-definitive/9781491912201/?ar>

- Mike Ryan, Federico Lucifredi . AWS System Administration: <https://learning.oreilly.com/library/view/aws-system-administration/9781449342562/?ar>

- Tom White . Hadoop : the definitive guide: <https://www.oreilly.com/library/view/hadoop-the-definitive/9780596521974/?ar>