
Formulario de Aprobación Curso de Actualización

Asignatura: Computación distribuida y procesamiento de grandes volúmenes de datos

Profesor de la asignatura: Sergio Nesmachnow, Profesor Titular, Instituto de Computación

Otros docentes de la Facultad: Néstor Rocchetti, Instituto de Computación

Instituto ó Unidad: Instituto de Computación

Departamento ó Área: Centro de Cálculo

Horas Presenciales: 35

Público objetivo y Cupos:

Público objetivo: profesionales y estudiantes en las áreas de Computación e Investigación de Operaciones interesados en los modelos de computación distribuida y procesamiento de grandes volúmenes de datos.

Curso sin cupos.

Objetivos:

Introducir al estudiante en las temáticas de computación distribuida y a las técnicas y tecnologías para resolver problemas de análisis de datos masivos.

Conocimientos previos exigidos: Fundamentos de informática.

Conocimientos previos recomendados: Fundamentos de programación

Metodología de enseñanza:

Exposiciones teórico-prácticas para presentar los principales conceptos vinculados con la computación distribuida y las metodologías de análisis y procesamiento de grandes volúmenes de datos.

- Horas clase (teórico): 20
- Horas clase (práctico): 5
- Horas clase (laboratorio): 5
- Horas consulta: 5 horas
- Horas evaluación:
 - Subtotal horas presenciales: 35

- Horas estudio: 25 horas
- Horas resolución ejercicios/prácticos: 10
- Horas trabajo final: 20
- Total de horas de dedicación del estudiante: 90 horas

Forma de evaluación: Participación oral y actividades prácticas realizadas en clases (20%). Entrega de un trabajo final (80%) que presente una aplicación de procesamiento de datos que utilice las técnicas de computación distribuida presentadas en el curso para la resolución de un problema específico.

Temario:

1. Computación distribuida y computación cloud.
2. Manejo de grandes volúmenes de datos.
3. El modelo de computación Map-Reduce.
4. El framework Hadoop y su ecosistema.
5. Almacenamiento: HDFS y HBase.
6. Aplicaciones de Map Reduce sobre Hadoop: conteo, índice invertido, filtros.
7. Procesamiento de datos en tiempo real: Apache Spark.
8. Ejemplos de aplicaciones en Spark y el lenguaje Scala.
9. Análisis de datos utilizando Spark y el lenguaje R.
10. Aplicaciones iterativas: Google Pregel y Apache Giraph.

Bibliografía:

Kai Hwang, Jack Dongarra, and Geoffrey C. Fox. 2011. Distributed and Cloud Computing: From Parallel Processing to the Internet of Things (1st ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. ISBN:0123858801 9780123858801.

George Reese. 2009. Cloud Application Architectures: Building Applications and Infrastructure in the Cloud. O'Reilly Media, Inc.. ISBN:0596156367 9780596156367.

Mario Macías, Mauro Gómez, Rubèn Tous y Jordi Torres. 2015. Introducción a Apache Spark para empezar a programar el Big Data. Editorial UOC. ISBN: 978-84-9116-037-3



Facultad de Ingeniería Comisión Académica de Posgrado

Datos del curso

Fecha de inicio y finalización: abril-mayo de 2018

Horario y Salón: a definir

Arancel: 17.820 pesos uruguayos
