
Formulario de Aprobación Curso de Posgrado 2015

Asignatura: Big data analytics: from big to smart

(Análisis de grandes volúmenes de datos: del procesamiento de grandes volúmenes al procesamiento inteligente)

Profesor de la asignatura : Genoveva Vargas Solar (Investigador Titular CNRS, Laboratorio IMAG, Grenoble, Francia)

Profesor Responsable Local ¹: Regina Motz (Grado 5, InCo)

Otros docentes de la Facultad:

(título, nombre, grado, Instituto)

Docentes fuera de Facultad: Javier Espinosa (Posdoctorado, Barcelona HPC Centre, España)

Instituto ó Unidad: Instituto de Computación

Departamento ó Area: Departamento de Programación

¹ Agregar CV si el curso se dicta por primera vez.

(Si el profesor de la asignatura no es docente de la Facultad se deberá designar un responsable local)

Fecha de inicio y finalización: 23 de noviembre al 4 de diciembre de 2015

Horario y Salón:

- Lunes 23/11: 18:00 a 21:00 horas / Salón Marrón (Nivel superior Bandejas)
- Martes 24/11: 17:30 a 20:30 horas / Salón Gris (7º piso)
- Miércoles 25/11: 17:30 a 20:30 horas / Salón Naranja (7º piso)
- Jueves 26/11: 17:30 a 20:30 horas / Salón Gris (7º piso)
- Viernes 27/11: 17:30 a 20:30 horas / Salón Marrón (Nivel superior Bandejas)

- Lunes 30/11: 17:30 a 20:30 horas / Salón Marrón (Nivel superior Bandejas)
- Martes 01/12: 17:30 a 20:30 horas / Salón Gris (7º piso)
- Miércoles 02/12: 17:30 a 20:30 horas / Salón Marrón (Nivel superior Bandejas)
- Jueves 03/12: 17:30 a 20:30 horas / Salón Gris (7º piso)
- Viernes 04/12: 17:30 a 20:30 horas / Salón Marrón (Nivel superior Bandejas)

Horas Presenciales: 30

(se deberán discriminar las mismas en el ítem Metodología de enseñanza)

Nº de Créditos: 6

(de acuerdo a la definición de la Udelar, un crédito equivale a 15 horas de dedicación del estudiante según se detalla en el ítem metodología de la enseñanza)

Público objetivo y Cupos: Estudiantes de posgrado en Informática y estudiantes avanzados de grado en Computación. Sin cupo.

(si corresponde, se indicará el número de plazas, mínimo y máximo y los criterios de selección. Asimismo, se adjuntará en nota aparte los fundamentos de los cupos propuestos. Si no existe indicación particular para el cupo máximo, el criterio general será el orden de inscripción en el Depto. de Posgrado, hasta completar el cupo asignado)

Objetivos: Presentar y experimentar técnicas de análisis de datos para generar información que permita comprender el contenido de las colecciones, extraer conocimiento y predecir situaciones a partir de los datos.

Conocimientos previos exigidos: Bases de datos, SQL, arquitecturas y programación distribuidas.

Conocimientos previos recomendados:

Metodología de enseñanza: El curso es de 2 semanas de duración. Se dictarán clases teórico-prácticas de 2 horas y media cada una. Las mismas consistirán de una parte teórica seguida de propuestas de ejercicios prácticos para realizar por parte de los estudiantes.

(comprende una descripción de las horas dedicadas por el estudiante a la asignatura y su distribución en horas presenciales -de clase práctica, teórico, laboratorio, consulta, etc.- y no presenciales de trabajo personal del estudiante)

- Horas clase (teórico): 15 hrs
- Horas clase (práctico): 15 hrs
- Horas clase (laboratorio):
- Horas consulta:
- Horas evaluación:
 - Subtotal horas presenciales: 30 hrs
- Horas estudio: 14 hrs
- Horas resolución ejercicios/prácticos : 28hs
- Horas resolución ejercicios seleccionados (a entregar) : 20hs
- Horas proyecto final/monografía:
 - Total de horas de dedicación del estudiante: 92 hrs

Forma de evaluación:

Entrega de ejercicios prácticos seleccionados en fecha a combinar.

Temario:

1. Análisis de grandes volúmenes de datos.
 - 1.1 Plataformas de hardware y recursos de redes para realizar análisis de datos
 - 1.2 Análisis de cosas
 - 1.3 Análisis cognitivo
2. Del procesamiento de grandes volúmenes al procesamiento inteligente.
 - 2.1 Panorama del tema Grandes Volúmenes de Datos
 - 2.2 Análisis de datos en -línea
 - 2.3 Análisis de datos en memoria
 - 2.4 La necesidad de exactitud
 - 2.5 La necesidad de Reproducibilidad y procedencia
3. Análisis Predictivo
 - 3.1 Análisis de los ecosistemas de datos
 - 3.2 Aplicación de técnicas de aprendizaje en profundidad

- 3.3 Tecnologías para visualización de resultados
- 3.4 Consideraciones energéticas para el análisis de grandes volúmenes de datos.

Bibliografía:

(título del libro-nombre del autor-editorial-ISBN-fecha de edición)

S. Chaudhuri, "What next?: a half-dozen data management research goals for big data and the cloud," in Proc. of the 31st PODS Symposium on Principles of Database Systems (PODS '12), 2012.

B. Grinter, "A big data confession," Interactions, vol. 20, no. 4, Jul. 2013.

A. Halevy, P. Norvig, and F. Pereira, "The Unreasonable Effectiveness of Data," IEEE Intell. Syst., vol. 24, no. 2, Mar. 2009.

P. Lyman, H. R. Varian, J. Dunn, A. Strygin, and K. Swearingen, "How Much Information?," Count. Numbers, vol. 6, no. 2, 2000.

P. Lyman, H. R. Varian, J. Dunn, A. Strygin, and K. Swearingen, "How Much Information?," Count. Numbers, vol. 6, no. 2, 2000.

R. Apps and R. Scale, Big Data Sourcebook. 2014.

J. S. Vitter, "External memory algorithms and structures: dealing with Massive Data," ACM Comput. Surv., vol. 33, no. 2, 2001.

F. Halper, "Predictive Analytics for Business Advantage," TDWI, 2014.

P. Zikopoulos, D. DeRoos, K. Parasuraman, T. Deutsch, J. Giles, and D. Corrigan, Harness the Power of Big Data. McGraw-Hill, 2013.

R. Cattell, "Scalable SQL and NoSQL data stores," SIGMOD Rec., vol. 39, no. 4, May 2011.

P. J. Sadalage and M. Fowler, NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot. 2012.

J. Kelly, D. Vellante, and D. Floyer, "Big Data Market Size and Vendor Revenues," Rep. Wikibon Proj., no. February, 2012.

A. Okcan and M. Riedewald, "Processing theta-joins using MapReduce," in Proc. of the 2011 ACM SIGMOD Int. Conference on Management of Data (SIGMOD '11), 2011.

J. D. Ullman, "Designing good MapReduce algorithms," XRDS Crossroads, ACM Mag. Students, vol. 19, no. 1, Sep. 2012.

J. Chandar, "Join Algorithms using Map / Reduce," Slides, 2010.

A. Thusoo, J. Sen Sarma, N. Jain, Z. Shao, P. Chakka, N. Zhang, S. Antony, H. Liu, and

R. Murthy, "Hive - a petabyte scale data warehouse using Hadoop," in Proc. of the 26th ICDE Int. Conference on Data Engineering (ICDE'10), 2010.