
Formulario de Aprobación Curso de Posgrado

Asignatura: Gestión y Visualización de Grandes Volúmenes de Datos

(Si el nombre contiene siglas deberán ser aclaradas)

Profesor de la asignatura ¹: Dr. Javier Espinosa (Investigador del Centro Nacional de Supercomputación,
Barcelona, España)

(título, nombre, grado o cargo, Instituto o Institución)

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

(título, nombre, grado, Instituto)

Otros docentes de la Facultad:

(título, nombre, grado, Instituto)

Docentes fuera de Facultad: Dra. Genoveva Vargas Solar (Investigadora Titular CNRS, Laboratorio LIG, Grenoble,
Francia)

(título, nombre, cargo, Institución, país)

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)

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 analizar los problemas de visualización y análisis de datos
- Diseñar y aplicar técnicas de limpieza de datos (data cleansing)
- Diseño de soluciones para administrar colecciones de datos voluminosas.
- Uso de técnicas de visualización para la comprobación de hipótesis y la exploración de datos.

Conocimientos previos exigidos: Bases de datos; sistemas distribuidos.

Conocimientos previos recomendados:

Metodología de enseñanza:

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

- Horas clase (teórico): 12
- Horas clase (práctico): 12
- Horas clase (laboratorio): 0
- Horas consulta: 6
- Horas evaluación: 0
 - Subtotal horas presenciales: 30
- Horas estudio: 30
- Horas resolución ejercicios/prácticos: 0
- Horas proyecto final/monografía: 30
 - Total de horas de dedicación del estudiante: 90

Forma de evaluación:

Proyecto.

Temario:

1. Data centric sciences new challenges
 - 1.1. Beyond Big Data
 - 1.2. Data exploitation from different perspectives
2. Curating data collections
 - 2.1. Curation what for ?
 - 2.2. From harvesting to maintenance
 - 2.3. Elements of data lakes and quality
 - 2.4. Conjectures and properties beyond data archival
3. Visualizing and exploring data collections
 - 3.1. Which metaphor for which perspective?
 - 3.2. From simple graphics to 3D life environments
 - 3.3. Technical challenges

Bibliografía:

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

- Introduction to Information Visualization. *Ricardo Mazza. Springer-Verlag, 2009.*
- The Visual Display of Quantitative Information(2nd Edition). *E. Tufte. Graphics, 2001.*
- 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.
- 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.
- 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.
 - J. D. Ullman, "Designing good MapReduce algorithms," XRDS Crossroads, ACM Mag. Students, vol. 19, no. 1, Sep. 2012.

Datos del curso

Fecha de inicio y finalización: Lunes 28 de noviembre al miércoles 7 de diciembre 2016

Horario y Salón: 17 a 20 horas, Salón a confirmar
