Com 2015
(2)

Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales

Planilla a completar para presentación de Cursos de Posgrado

1.- DEPARTAMENTO de COMPUTACIÓN

2.- NOMBRE DEL CURSO: Big Data: una perspectiva desde la práctica

3.- DOCENTES:

RESPONSABLE/S: Daniel Yankelevich
COLABORADORES:
AUXILIARES:

4.- CARRERA de DOCTORADO

5.- AÑO: 2015 CUATRIMESTRE/S: Primero

6.- PUNTAJE PROPUESTO PARA CARRERA DE DOCTORADO: 1 punto

7.- DURACIÓN (anual, cuatrimestral, bimestral u otra): **Menos de un mes.**

8.- CARGA HORARIA SEMANAL: 4hs.

Teóricas:
Problemas:
Laboratorio:
Seminarios:
Teórico – Práctico: .....
Salida a Campo: .....

9.- CARGA HORARIA TOTAL: 20hs.

10.- FORMA DE EVALUACIÓN: Exámen final escrito.

11.- PROGRAMA ANALÍTICO:

Dr. ESTEBAN FEUERSTEIN
DIRECTOR
Depto. COMPUTACION
FCE y N - UBA

Caso 1 - Bombas Sumergibles

- Introducción a data mining, big data, y la problemática específica que representa.
- Qué es big data. Principales desafíos.
- Modelos predictivos. Uso de ML en un ejemplo concreto.
- Variantes de

Caso 2 – Base Unificada de Datos Petroleros (BUDaP)

- Bases de datos distribuidas, los problemas de la distribución y la necesidad de la distribución.
- Sistemas de archivos distribuidos.
- Bases NoSQL, discusión de ventajas/desventajas.
- Teorema CAP. Consistencia.

Caso 3 – ¿Qué guardar que valga la pena? ¿Qué recomendar a otros?

- Agrupando y encontrando elementos comunes, clases o conceptos (Clustering, LSH, reducción de dimensiones)
- Otras técnicas, SVD, CUR decomposition
- ¿Cómo saber si un mail es SPAM? ¿Cómo saber si un mensaje que me mandan es de una persona real? Aplicaciones.

Caso 4 – ¿Cómo capturar a un asesino serial?

- Limitaciones en el análisis de datos: generalizaciones/extrapolaciones (cisne negro, drunken walk). Principio de Bonferroni.
- Limitaciones y problemas típicos en los proyectos de data mining y big data.
- Calidad de datos.

12.- BIBLIOGRAFÍA:

Mining of Massive Datasets, Jure Leskovec Stanford Univ. Anand

Rajaraman Milliway Labs and Jeffrey D. Ullman Stanford Univ., Palo

Alto, 2010

A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Weiner, "Graph structure in the web," Computer Networks 33:1–6, pp. 309–320, 2000.

H. Garcia-Molina, J.D. Ullman, and J. Widom, Database Systems: The Complete Book Second Edition, Prentice-Hall, Upper Saddle River, NJ, 2009.

J. Dean and S. Ghemawat, "Mapreduce: simplified data processing on large clusters," Comm. ACM 51:1, pp. 107–113, 2008.

Dr. ESTEBAN FEUERSTEIN
DIRECTOR
Depto. COMPUTACION
FCE y N - UBA

S. Ghemawat, H. Gobioff, and S.-T. Leung, "The Google file system," 19th ACM Symposium on Operating Systems Principles, 2003.

A.Z. Broder, "On the resemblance and containment of documents," Proc. Compression and Complexity of Sequences, pp. 21–29, Positano Italy, 1997.

M. Datar, N. Immorlica, P. Indyk, and V.S. Mirrokni, "Locality-sensitive hashing scheme based on p-stable distributions," Symposium on Computational Geometry pp. 253–262, 2004.

M. Theobald, J. Siddharth, and A. Paepcke, "SpotSigs: robust and efficient near duplicate detection in large web collections," 31st Annual ACM SIGIR Conference, July, 2008, Singapore.

The MADlib Analytics Library or MAD Skills, the SQL Joseph M.

Hellerstein Christopher Ré Florian Schoppmann Zhe Daisy Wang Eugene

Fratkin Aleksander Gorajek Kee Siong Ng Caleb Welton Xixuan Feng Kun

Li Arun Kumar Electrical Engineering and Computer Sciences University

of California at Berkeley Technical Report No. UCB/EECS-2012-38

http://www.eecs.berkeley.edu/Pubs/TechRpts/2012/EECS-2012-38.html

April 3, 2012

The Byzantine Generals Problem

LESLIE LAMPORT, ROBERT SHOSTAK, and MARSHALL PEASE SRI International; ACM Transactions on Programming Languages and Systems, Vol. 4, No. 3, July 1982, Pages 382-401.

Bad Data Handbook, Ethan McCallum, 2013. Published by O'Reilly Media, Inc.

Data Analysis with Open Source Tools, Philipp K. Janert, 2013. Published by O'Reilly Media, Inc.

Harness Oil and Gas Big Data with Analytics: Optimize Exploration and

Production with Data-Driven Models, Keith R. Holdaway, Ed: Wiley, 2014.

Dr. ESTEBAN FEUERSTEIN
DIRECTOR
Depto. COMPUTACION
FCE y N - UBA

Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales

Referencia Expte. N° 505.243/15

VISTO: Buenos Aires, 07 SEP 2015

la nota presentada por el Dr. Esteban Feuerstein, Director del Departamento de Computación, mediante la cual eleva la información y el programa del curso de posgrado **Big Data: una perspectiva desde la práctica,** que será dictado durante 2015 por el Dr. Daniel Yankelevich,

CONSIDERANDO:

lo actuado por la Comisión de Doctorado,

lo actuado por la Comisión de Postgrado,

lo actuado por este Cuerpo en la sesión realizada en el día de la fecha,

en uso de las atribuciones que le confiere el Artículo 113° del Estatuto Universitario,

EL CONSEJO DIRECTIVO DE LA FACULTAD DE CIENCIAS EXACTAS Y NATURALES
R E S U E L V E:

Artículo 1°: Autorizar el dictado del curso de posgrado **Big Data: una perspectiva desde la práctica** de 20 hs. de duración.

Artículo 2°: Aprobar el programa del curso de posgrado **Big Data: una perspectiva desde la práctica**, obrante a fs 3 y 4 del expediente de la referencia.

Artículo 3°: Aprobar un puntaje máximo de un (1) punto para la Carrera del Doctorado.

Artículo 4°: Comuníquese a la Dirección del Departamento de Computación y a la Biblioteca de la FCEyN (con fotocopia del programa incluido). Comuníquese a la Dirección de Alumnos y a la Secretaría de Postgrado (sin fotocopia del programa). Cumplido Archívese.

2181

RESOLUCION CD N°_____
SP/ga 02/09/15

Dr. JOSÉ OLABE IPARRAGUIRRE
SECRETARIO DE POSGRADO
FCEN - UBA

Dr. JUAN CARLOS REBOREDA
DECANO