

Taller de Ingeniería de Software y Bases de Datos

Maestría en Ciencia e Ingeniería de la Computación, UNAM

Viernes 9 de noviembre de 2018

Auditorio del IIMAS

Introducción al área

El Posgrado en Ciencia e Ingeniería en Computación de la UNAM, cuenta con varios tutores y profesores en las áreas de Ingeniería de Software y de Bases de Datos que en este taller presentarán a los alumnos sus temas de interés. Adicionalmente, un grupo de exalumnos platicarán sus experiencias después del posgrado.

9:00 – 9:15

Presentación del área de Ingeniería de Software
Guadalupe Ibargüengoitia G.

9:15 – 9:30

Aplicación de procesos de Ingeniería de Software
M. en C. Guadalupe Ibargüengoitia G.

9:30 – 9:45

Los estándares no tienen la culpa.
Dra. Hanna Oktaba

9:45 – 10:00

Desarrollo de aplicaciones en internet.
M. en C. Gustavo A. Márquez F.

10:00 – 10:15

Análisis del usuario como base para el desarrollo de software que transforme.
Dr. Fernando Gamboa

10:15 – 11:00

Mesa de profesores del área
M. en I. Ana Yuri Ramírez
Dr. Francisco Valdés Souto

11:00 – 11:15

Café

11:15 – 11:30

Presentación del área de Bases de Datos del PCIC

Ma. del Pilar Ángeles, Javier García García

11:30 – 11:45

De los sistemas de archivos a hadoop, evolución o retroceso.

Dra. Pilar Ángeles

11:45 – 12:00

En qué consiste la investigación en temas de bases de datos

Dr. Javier García García

12:00 – 13:00

Can Parallel Database Systems Help Big Data Analytics?

Dr. Carlos Ordoñez*

13:00 – 13:20

Café

13:20 – 13:45

Innovaciones en RDBMS comerciales

Dr. Víctor González Castro

13:45 – 14:45

Mesa de egresados

Sandra Ramírez

Israel Sandoval

Judith Jaramillo

Ma. Dolores Mendoza

14:45 – 16:00

Bocadillos

* Conferencia invitada:

Abstract:

Big data analytics remains an active research area. We argue parallel database systems can indeed solve important, but not all, big data problems. We present a review of parallel processing alternatives, storage mechanisms, scalable algorithms and data structures. We discuss three major tasks: data preprocessing (integrating, cleaning, aggregating, transforming data), computing common machine learning models (clustering, dimensionality reduction, regression, classification) and solving graph problems (paths, clique detection, page rank). On the mathematical side, we contrast relational tuples/tables

and vectors/matrices as fundamental abstractions. On the systems side, we compare features of parallel DBMSs (evaluating SQL queries, UDFs, C/C++ code) and the Hadoop stack (HDFS, MapReduce/Spark, Java/Scala code) as the two major competing technologies. We conclude with an outline of solved problems and open research issues. This talk should be interesting to both practitioners and researchers.

Bio:

Carlos Ordonez studied at UNAM, Mexico, getting a [B.Sc.](#) in applied math and an M.S. in computer science. He continued PhD studies at the Georgia Institute of Technology focusing on extending database systems with data mining algorithms. Thereafter Carlos worked at NCR Corp. for 8 years, collaborating in the optimization of machine learning and cube algorithms to work on the Teradata parallel DBMS. In 2006 Carlos joined the University of Houston, where he currently leads the Parallel Database Systems lab. From 2013 to 2015 Carlos collaborated with Michael Stonebraker, regularly visiting the Database Group at MIT. From July 2014 to July 2015 Carlos was a visiting researcher at ATT Labs (formerly ATT Bell Labs), where he worked on stream analytics and big data warehousing with Divesh Srivastava. Carlos has served as Editor on the IEEE TKDE, Intelligent Data Analysis and Information Systems journals and has been on the PC of venues like SIGMOD, KDD, DOLAP, ICDE, CIKM, DaWaK, and AMW. His research has been funded by NSF.