

**DEPARTMENT OF COMPUTER SCIENCE**  
University of Houston

**FACULTY SEMINAR SERIES - SPRING 2013**

**WHEN:** FRIDAY, JANUARY 25, 2013  
**WHERE:** PGH 232  
**TIME:** 10:30 AM

**SPEAKER:** Dr. Carlos Ordonez, University of Houston

Host: Dr. Christoph Eick

**TITLE:** Numerical Linear Algebra in a Database System for Big Data Analytics

**ABSTRACT:**

Linear algebra is behind most scientific analytic computations, including numerical methods, statistics, physics and data mining. On the other hand, database systems provide extensive features to manage and query data, but not to analyze it. The integration of linear algebra algorithms into a database system is difficult due to the relational model discrete set foundation, which in consequence hinders efficient array processing. In this talk, we discuss our progress towards solving two prominent linear algebra problems on large data sets: singular value decomposition and least squares regression. We present a common matrix foundation to summarize large data sets with an outer product and exploiting such summarization we discuss how to efficiently solve linear algebra problems with iterative methods. We explain optimization of data movement going from secondary storage to primary memory storing the data set matrix with several layouts, presenting a tradeoff between high dimensionality and efficiency on row stores, the most common in modern database systems. We then compare different programming solutions to solve linear algebra problems going from completely expressing computations as queries to calling LAPACK routines, where processing of matrix operators and iterative methods is pushed to primary storage. We conclude with a brief experimental comparison showing database systems represent a promising platform for big data analytics. We emphasize speed and scalability, but we also consider numeric stability of our optimized algorithms.

**Short Bio:**

Carlos Ordonez studied at UNAM University (Universidad Nacional Autonoma de Mexico), where he got degrees in applied mathematics and computer science. He continued PhD studies at the Georgia Institute of Technology under Edward Omiecinski, conducting research on database systems and data mining. Dr Ordonez worked for Teradata (formerly part of NCR), from 1998 to 2006 as intern and researcher, developing DBMS software in SQL, C++ and Java, as well as providing consulting to perform data mining on massive relational databases to more than 40 companies. Dr Ordonez cooperated in the optimization of several cube, statistical and machine learning algorithms that work inside the Teradata parallel DBMS to analyze large data sets. In 2006 Carlos Ordonez joined the Department of Computer Science at the University of Houston, where he leads the active DBMS group. His research spans database systems, data mining, big data analytics and high performance computing.