

# **Integrating Database Systems and Data Mining Algorithms**

Carlos Ordonez

Department of Computer Science. University of Houston , Texas. USA

Data mining remains an important research area in database systems and a major challenge in computer science. We present a review of processing alternatives, storage mechanisms, algorithms, data structures and optimizations that enable data mining on large data sets. We focus on the computation of well-known multidimensional statistical and machine learning models. We pay particular attention to SQL (together with UDFs) and MapReduce as two competing technologies for large scale “in-database” analytics, especially with parallel computing. We conclude with a summary of solved major problems and open research issues.