ABSTRACT
In more and more industries, competitive advantage hinges on exploiting the largest quantity of data in the shortest possible time - and doing so cost-effectively. Data volumes are growing exponentially, while businesses are striving to deploy sophisticated and computationally intensive predictive analytics. Often, massive data is stored in a data warehouse running on dedicated parallel hardware, but advanced analytics is performed on a separate compute platform. Moving data from the data warehouse to the compute environment can constitute a significant bottleneck. Organizations resort to considering only a fraction of their data or refreshing their analyses infrequently. To address the data movement bottleneck and take full advantage of parallel data warehouse platforms, vendors are offering new in-database analytics capabilities. They are opening up their platforms, allowing users to run their own user-defined functions and statistical models as well as vendor- and partner-supplied advanced analytics on the database platform, close to the data, in parallel, without transporting the data through a host node or corporate network. In this talk, we will present the need for in-database analytics and discuss a number of the new solutions available, highlighting case studies where solution times have been reduced from hours to minutes or seconds.

Categories and Subject Descriptors
H.2.7 [Database Administration]: Data warehouse and repository, H.2.8 [Database Management]: Data Mining.


Keywords: Large-scale data warehousing, in-database analytics, predictive analytics.

Bio
Dr. Mario E. Inchiosa is U.S. Chief Scientist at Netezza, an IBM Company, where he develops data-intensive high performance computing appliances. His work focuses in particular on the juncture of data warehousing and parallelized advanced analytics and optimization. Dr. Inchiosa received an A.B. in Physics from Harvard College and an A.M. and Ph.D. in Physics from Harvard University. At Harvard, he combined his dual interests in Physics and Computer Science by applying statistical physics to the study of neural network associative memories. He moved on to study the dynamics of neural network associative memories as a post-doc at the Technical University of Munich. Next, he joined SPAWAR Systems Center San Diego, specializing in stochastic non-linear dynamics, signal detection, Monte Carlo simulation, and high performance computing. He was awarded four patents as a result of his research, and he has published over 30 papers, earning Publication of the Year and Technical Publication Excellence awards. In 2001 Dr. Inchiosa joined BiosGroup, a Santa Fe Institute complexity science spin-off (subsequently NuTech Solutions), applying evolutionary algorithms and swarm-like agent based modeling to problems in business and government. He developed pipeline simulation and optimization engines, served as Principal Investigator researching general and geospatial reasoning under uncertainty, and used agent based models to study global market dynamics and co-evolutionary business strategy optimization. As NuTech’s Chief Science Officer, Dr. Inchiosa was involved with Netezza’s acquisition of NuTech as part of Netezza’s strategy to bring advanced analytics capabilities to data warehouse appliances.