

# Department of Computer Science and Software Engineering

## Master Thesis Defense

Speaker:	Ruhan Sayeed
Supervisor:	Dr. T. Eavis
Examining Committee:	Drs. Chalin, Rilling and Dr. Constantinides (Chair)
Title:	High Performance Analytics with the R3-cache
Date:	Monday April 6, 2009
Time:	1:00 pm
Place:	EV3.309

## ABSTRACT

Contemporary data warehouses now represent some of the world's largest databases. As these systems grow in size and complexity, however, it becomes increasingly difficult for brute force query processing approaches to meet the performance demands of end users. Certainly, improved indexing and more selective view materialization are helpful in this regard. Nevertheless, with warehouses moving into the multi-terabyte range, it is clear that the minimization of external memory accesses must be a primary performance objective. In this paper, we describe the R3-cache, a natively multi-dimensional caching framework designed specifically to support sophisticated warehouse/OLAP environments. R3-cache is based upon an in-memory version of the R-tree that has been extended to support buffer pages rather than disk blocks. A key strength of the R3-cache is that it is able to utilize multi-dimensional fragments of previous query results so as to significantly minimize the frequency and scale of disk accesses. Moreover, the new caching model directly accommodates the standard relational storage model and provides mechanisms for pro-active updates that exploit the existence of query "hot spots". The current prototype has been evaluated as a component of the Sidera DBMS, a "shared nothing" parallel OLAP server designed for multi-terabyte analytics. Experimental results demonstrate significant performance improvements relative to simpler alternatives.