TITLE OF HIGHLIGHT: **Efficient View-Design Algorithms to Achieve Near-Optimal Performance of Sets of Relational Queries**

PRIMARY STRATEGIC OUTCOME GOAL: Discovery
CATEGORY: Disciplinary/Interdisciplinary Research

1) Why is this research outcome notable and/or important, and how does it address the strategic outcome goal(s) of the NSF strategic plan?

As data warehouses keep growing in size, evaluating many common queries – such as aggregate queries - in online analytical processing (OLAP) may require significant transformations of large volumes of stored data. Aggregate queries are widely used in data warehouses and decision support. Optimization based on the reuse of query answers is particularly promising for aggregate queries, as often an enormous amount of data is scanned to produce a single aggregate value. As a result, the requirement of good overall performance of frequent and important aggregate queries necessitates optimal choices in choosing and executing query plans.

A significant aspect of query performance is the choice of auxiliary data. In our previous work we considered *materialized views*, relations that were computed by answering certain queries on the (original) stored data in the database, as auxiliary data to accelerate answering queries. In our current research, we are considering not only views, but also indexes over views to reduce the response time of OLAP queries. While it can be relatively easy to improve to some degree the query evaluation costs by using, for instance, greedy strategies for choosing indexes or views, it is highly nontrivial to arrive at a globally optimum solution, one that reduces the processing costs of typical OLAP queries as much as is theoretically possible.

In our current work we introduce an integer programming (IP) model for the view and index selection problem to select the set of "optimal" views and indexes. To be able to solve larger instances of the problem using our IP model, we propose methods for reducing the sizes of the pools of views and indexesOur major contribution so far for this aim is reducing the number of potential indexes for each view from n! to 2^n, where n is the number of attributes in the view. We did this by considering the structural properties of indexes in reducing the cost of answering queries. This reduction in the sizes of the search spaces of views and indexes enabled us to solve much larger instances of the view and index selection problem optimally. We also propose several heuristic methods to further reduce the sizes of the pools of views and indexes to be able to solve even larger instances of the problem.

2) Does this highlight represent transformative research?

No

3) Does this highlight represent Broadening Participation?

No

4) Are there any existing or potential societal benefits, including benefits to the U.S. economy, of this research of which you are aware?

Yes: Improving the response time of OLAP queries is beneficial for business intelligence and can help scientific, governmental, and commercial decision-support systems in a variety of application areas.