We have made advances in the following areas:

* Data cubes: these recent data-warehouse products need a way to optimize the use of space by selecting some views to maintain permanently. We have identified the "monotonicity" property --- choosing one view cannot increase the value of materializing another view --- as guaranteeing the existence of a polynomial-time, competitive (guaranteed to come within a constant fraction of optimum) solution. In one important nonmonotone case, data cubes with indexes on views, we showed how to find a polynomial, competitive algorithm.

* Self-Maintenance of views: We have techniques for deciding whether or not a view that is defined by a conjunctive query can be maintained in the face of an update to a base relation, without issuing queries to one or more base relations. For a variety of situations, we showed how to express this "self-maintainability" condition as an SQL query. We also can take advantage of functional dependencies to simplify the test.

---

<table>
<thead>
<tr>
<th>Title and Subtitle</th>
<th>Efficient Management of Active Databases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors</td>
<td>Jeffrey D. Ullman and Jennifer Widom</td>
</tr>
<tr>
<td>Organization</td>
<td>Stanford University</td>
</tr>
<tr>
<td></td>
<td>Dept of Computer Science</td>
</tr>
<tr>
<td></td>
<td>Stanford, CA 94305</td>
</tr>
<tr>
<td>Funding Numbers</td>
<td>DAAH04-95-1-0192</td>
</tr>
<tr>
<td>Organization</td>
<td>U.S. Army Research Office</td>
</tr>
<tr>
<td></td>
<td>P.O. Box 12211</td>
</tr>
<tr>
<td></td>
<td>Research Triangle Park, NC 27709-2211</td>
</tr>
<tr>
<td>Sponsor</td>
<td>U.S. Army Research Office</td>
</tr>
<tr>
<td></td>
<td>P.O. Box 12211</td>
</tr>
<tr>
<td></td>
<td>Research Triangle Park, NC 27709-2211</td>
</tr>
<tr>
<td>Distribution</td>
<td>Approved for public release; unlimited</td>
</tr>
<tr>
<td>Subject</td>
<td>Active Databases</td>
</tr>
</tbody>
</table>
EFFICIENT MANAGEMENT OF ACTIVE DATABASES

DAAH04-95-1-0192

JEFFREY D. ULLMAN, JENNIFER WIDOM, PI's

Stanford University
Department of Computer Science
Gates Hall, 4A Wing
Stanford CA 94305
Ullman Phone: (415) 725-4802
Widom Phone: (415) 723-7690
FAX: (415) 725-2588
email: {ullman, widom}@db.stanford.edu

Research Achievements

Active elements in databases are becoming progressively more important commercially. Rules and
constraints give databases intelligent capabilities and are an essential part of the emerging SQL3
standard as well as being present to an extent in SQL2. Materialized views are attracting a great deal of
new interest. For example, "data cubes" or other forms of "data warehouses" that support on-line
analytic processing (OLAP) for applications such as mining data for unexpected patterns are popular in
marketing and could just as well be used to analyze strategic threats and opportunities. As another
example, intelligence services materialize views into warehouses, even for unclassified information, so
that they can be queried without exposing the questions that are being asked.

Below is a summary of the accomplishments of the project, which unfortunately was cut short because
of the inability of ARO to provide the third year's funding.

Data Cubes

We developed techniques for improving the efficiency of OLAP queries on a data cube. The primary
idea is to materialize some views that are aggregations of the raw data in the data cube. These
materialized views are in a sense "little warehouses" helping to answer queries that would take too long
if asked on the "big warehouse" --- the entire data cube.

In Harinarayan, Rajaraman, and Ullman [1996] (winner, of SIGMOD best-paper award), we gave the
basic idea of how to design data cubes by optimizing the choice of subcubes to materialize. The key
theorem is that the simple greedy algorithm is guaranteed never to be worse than 63% of optimum (in
experiments, it is much better than that).

Gupta [1997] addresses the more general problem of selecting views to materialize in any warehouse,
and identifies the "monotonicity" property (picking a view must not make another view more valuable)
as the key to guaranteeing competitive performance, i.e., a polynomial algorithm that gets within some
constant fraction, often the 63% alluded to above, of the optimum.
In Gupta, Harinarayan, Rajaraman, and Ullman [1997] we address the problem of what to do if the application does not have the monotonicity property. In particular, we address the important design problem of data cubes with indexes on some of the materialized subcubes. If we think of an index as another "view" to materialize, it does not make sense to materialize an index $I$ until its underlying view $V$ is materialized. Thus, materializing $V$ causes a jump in the value of $I$, violating monotonicity. In this paper we show that a variation of the greedy algorithm is competitive, with a ratio of 46%.

Ullman [1996a] presented a summary of these results to the KDD (knowledge discovery and data-mining) conference last summer.

**View Self-Maintenance**

The goal is to keep a materialized view (i.e., a warehouse or part of a warehouse) up to date as the underlying source data changes. Ideally, we would like to make the changes at the warehouse after being notified of the underlying change without having to examine any other source data, i.e., using only the update and the contents of the warehouse. Deciding whether one can do so for a particular update, and deciding with a tractable query to the warehouse, are very hard problems in general.

Pierre Huyn has attacked a number of problems in this general area. In Huyn [1996a] he gives the general idea and solves the problem for views that are defined by conjunctive queries without self-joins, i.e., no two subgoals have the same predicate. Huyn [1996c] handles conjunctive queries with self-joins, and In Huyn [1996b] includes functional dependencies into the framework. Huyn [1997a] shows that consistency-preserving updates can be efficiently detected when the relations under constraints are not completely given. The tests are given in the form of nonrecursive Datalog queries with negation. Huyn [1997b] solves the problem of view self-maintenance in the presence of multiple views and under arbitrary base updates. For a subclass of conjunctive-query views, the paper shows how to solve the problem in polynomial time and in particular how to generate queries for maintenance and for testing self-maintainability.

**Distributed Constraint Maintenance**

Huyn also shows how to maintain constraints that involve data at several different sites, without looking at anything but local data unless absolutely necessary. In Huyn [1998] extends earlier results to include negated subgoals in the query form.

**Mediation**

The keynote ICDT paper Ullman [1997] synthesizes two different approaches to mediation, the ATT Labs "Information Manifold" approach, and Stanford's "Tsimmis" approach.

Levy, Rajaraman, and Ullman [1996] contributes to the automatic generation of mediators in the following way. This paper shows how to describe the capability of a source to answer queries in a grammar-like notation and then to find whether a given query has one of the (possibly infinite) forms that are described by this grammar.

Ullman [1996b] is a survey of mediation theory for the AI audience.

**Representative Objects**
Nestorov, Chawathe, Ullman, and Wiener [1997] is a contribution to the LORE object-oriented warehouse system being developed at Stanford. The idea is to guide the user of semistructured data by showing the local structure of the hierarchy as it is explored top-down. The paper shows how to use some classical algorithms from finite-automaton theory to construct a concise representation of the structure efficiently.

Publications


Awards and Honors

1. V. Harinarayan, A Rajaraman, and J. D. Ullman: Best Paper Award, SIGMOD 1996.

2. J. D. Ullman: SIGMOD Contributions Award, 1996.


Books Published


Patents Filed

None.

Number of Graduate Students Supported, by Gender and Minority Group

Pierre (Nam) Huyn, male, Asian-American.

PhD's Awarded to Students Who Worked on the Grant


2. Pierre (Nam) Huyn, PhD, 1997, returned to Hewlett-Packard, where he worked before, and has now taken a position at Hitachi-America in Santa Clara CA.

Nonexpendable Instrumentation Purchased

None.