

Panel: Querying Networked Databases

Nick Koudas
AT&T Labs-Research
koudas@research.att.com

Divesh Srivastava
AT&T Labs-Research
divesh@research.att.com

1. MOTIVATION

A large number of useful databases are currently accessible over the Web and within corporate networks. In addition to being frequently updated, this collection of databases tends to be highly dynamic: new databases appear often, and databases (just like Web sites) also disappear. In this environment, the goal of providing flexible, timely and declarative query access over all these databases remains elusive.

- Traditional data integration technology is useful in providing an integrated view over a collection of small, typically domain-specific, databases.

While query access is usually declarative and timely, it is not obvious how this technology can be used in a flexible way over a heterogeneous, dynamic collection of databases.

- For conventional Web search engine technology to be applicable, the contents of these databases would need to be published as Web documents, which is typically infeasible.

Even under this assumption, while this approach would be applicable to a dynamic collection of databases, timeliness issues (search engines indices tend to be out of date) make this approach less than ideal for querying frequently updated databases.

- Peer-to-peer and datagrid technologies have proven useful in locating replicas of a large collection of files and datasets in a timely fashion, using centralized and distributed metadata catalogs.

However, it is less clear how this technology could be used to provide declarative query access over a large set of databases.

The panel will focus on the challenging problem of querying networked databases, and identify methodologies and techniques from different disciplines of Computer Science, including Databases, Information Retrieval, and Networking, that are needed for its solution.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGMOD 2003, June 9-12, 2003, San Diego, CA
Copyright 2003 ACM 1-58113-634-X/03/06 ...\$5.00.

2. PANELISTS

FCRC'03 provides a valuable opportunity to engage panelists with database expertise from different communities of interest in addressing this problem. The following panelists will serve on the panel.

- Michael Carey, BEA Systems Inc.
- Hector Garcia-Molina, Stanford U.
- Tim Griffin, AT&T Labs-Research
- Alon Halevy, U. of Washington
- Tomasz Imielinski, Rutgers U.
- Prabhakar Raghavan, Verity Inc.

The panelists include a mix of academics and industry people, from different disciplines of relevance to the topic of the panel. They bring expertise in database systems, data integration technology, information retrieval systems, networking systems, and peer-to-peer technology.

3. MAIN QUESTIONS FOR PANELISTS

The panelists will address the following main questions:

- What would the architecture of a system, meeting the requirements of providing flexible and timely access to networked databases using a declarative query language, look like?
- Is there a need to revisit issues related to distributed databases and data integration technologies, based on this architecture?
- What role can modern networking techniques and paradigms play in such an architecture?
- How can information retrieval technologies be used in this context?
- What are the specific research questions arising in this context?