

TITLE OF RESEARCH WORK:


Research Description:

A traditional Database Management System (DBMS) is divided on several components. One of the most important components is the Query Optimizer (QO). From a given query, the QO is in charge of selecting an optimal plan, based on a set of performance metrics, for the execution of a given query. A query plan is a representation of operators that are needed for the execution of a given query. This plan contains all the information needed by other components on the DBMS to execute a given query, like relational operations, relations, attributes, information on how each relation is stored on disk, and so on.

Traditional centralized QO fail to operate on a wide area network environment for several reasons: network traffic, extensive consideration of multiple parameters, large size of search space for query plans, etc.

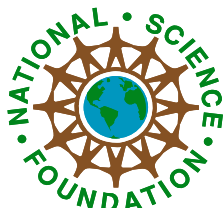
NetTraveler is a distributed database middleware system that takes into consideration the dynamic characteristics of wide area networks to efficiently run queries on it. Our approach consists on dividing the QO into two separate processes. The first process gives a logical plan that represents an optimal plan based on local (and perhaps imprecise) information. This plan is then refined via further optimizing on a second phase carried out by target components across the system. This break with the traditional idea of a centralized architecture for query optimization.

The focus of this research is towards the development of this second phase of the optimization process and the necessary operators and mechanism to develop a successful environment for distributed and parallel query execution on NetTraveler. Our goal is to provide the means to improve response time, load balancing and a robust environment for query execution.

<p>WALSAIP GROUP ASSOCIATION: Advanced Data Management Group</p>	
<p>THESIS TITLE: Distributed and Parallel Query Execution environment on NetTraveler</p>	
<p>THESIS ADVISOR: Prof. Manuel Rodriguez</p>	
<p>INSTITUTION: Department of Electrical and Computer Engineering University of Puerto Rico at Mayaguez</p>	
<p>PERSONAL WEBSITE: http://amadeus.uprm.edu/~s08413/</p>	<p>NAME OF RESEACH ASSISTANT: Angel L. Villalaín-García</p>

RESEARCH PROJECT OUTCOMES:

Publications:



Tools and Applications:

NetTraveler is implemented as a collection of Java Web Services, thus enabling interoperable integration of software applications running on several different platforms and/or frameworks. Our implementation of web service relies on the Apache Axis toolkit which is an implementation of the Simple Object Access Protocol (Soap).

MySQL, PostgreSQL, are being used to represent our database servers. We plan to use other technologies to represent, as close as we can, the heterogeneity of this kind of system in the real world. Data access relies on the Java Database Connectivity (JDBC) API, providing NetTraveler with a call-level API for accessing SQL-based databases. In addition, depending on which other technologies we used for representing our data sources (e.g. XML files, text files), we shall need other Java APIs for accessing these data sources.

Other secondary technologies that are being used on NetTraveler are:

- Java API for XML Binding (JAXB)
- Apache Tomcat server

RELATION OF RESEARCH WORK TO WALSAIP PROJECT:

The WALSAIP project, as well as numerous scientific applications, requires access to a vast volume of imagery, and measurements from multiple remote sensing instruments. These data resides on geographically distributed data sources, with heterogeneous schemas, heterogeneous server systems, and different usage policies and security restrictions.

Real time access to this huge amount of distributed information needs an efficient environment that could harness distributed computation resources to efficiently distribute the load required for query processing in order to have a suitable response time.

In order to achieve this goal, the development of a decentralized query optimizer and new mechanisms and operators for making a suitable query distributed and parallel query execution environment is needed. This is the main focus of the proposed research.

This would provide WALSAIP not only with a mechanism for extracting information but also with a robust and asynchronous environment that minimizes response time.

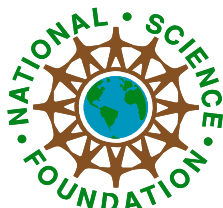


IMAGE REPRESENTATIVE OF RESEARCH WORK:

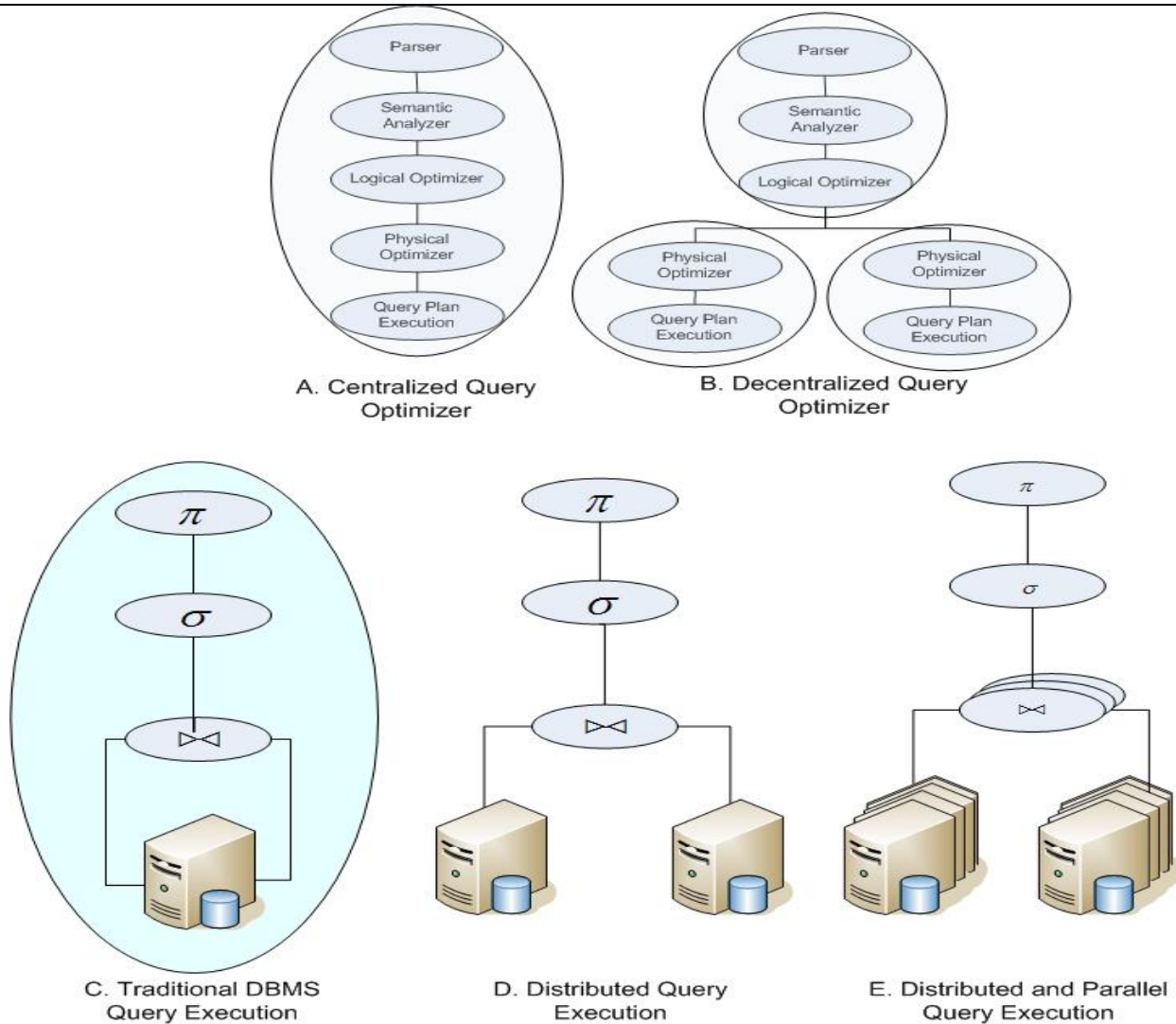
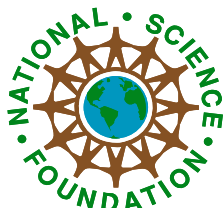


Figure 1 Title needed

Figure 1.A represents the different process to execute a query on a traditional DBMS. As explained above this approach fails to succeed on a distributed execution environment due to the excessive information needed at the optimization steps. Figure 1.B represents what we proposed as a solution. This figure shows that the different steps are organized on two different phases. The idea here is to distribute the QO phase between the different peers that can participate in solving the query. The next three figures (Figure 1.C, 1.D and 1.E) present different plans for solving a query. The first one is the traditional query execution on a single database server; in fact each step is executed as part of the DBMS. This approach is not suitable because fails to integrate remote data sources. The next two are the approaches that we are considering for our project, in which each step could be executed by different peers. Depending on some metrics the selected plan could either present one of those behaviors or even integrate both in order to improve the response time of the system.

RESEARCH DEMONSTRATION:



Although we have a current available running version, with a custom console for testing purpose that displays our results, it is not a final version, and components are still under development. So we decided not to provide it as a publicly available demo, because it also requires of several steps of configuration and installation that are not suitable to any individual that only wants to test the system and see its performance. Also its use interface is not suitable for public release.

For this reason we are proposing a web based demo that provides the user with a friendly user-interface that would be accessible without the need of any installation and configuration steps. This demo would also provide the user with mechanism allowing customizing some of the system parameters instead of being completely managed by our prototype allowing the user to test the performance behavior of the system, and a feature that will display statistical information of the deployed services on NetTraveler. The following figure shows how this demo will look like.



Figure 2 Proposed Demo

