Hash-Based Algorithms For Operator Load-Balancing In Database Middleware Systems

By: Angel Villalain, MS Student
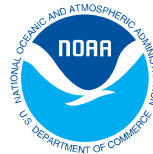
Advisor:

Prof. Manuel Rodriguez
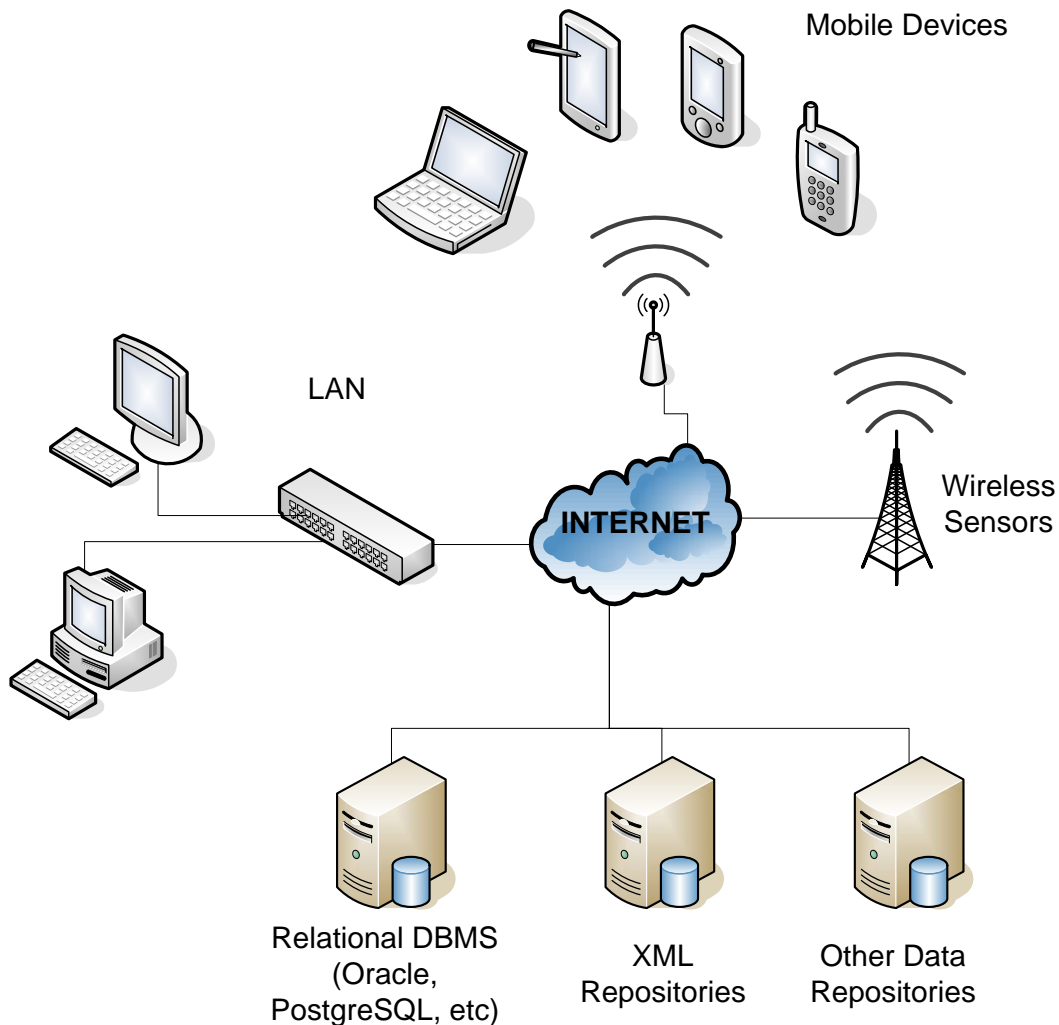
Advanced Data Management Laboratory
University of Puerto Rico at Mayaguez (UPRM)
March 14, 2007

Problem Formulation

Mobile Devices

LAN

INTERNET

Wireless Sensors

Relational DBMS
(Oracle,
PostgreSQL, etc)

XML
Repositories

Other Data
Repositories

- Wide Area Networks (WAN) are becoming common ground for the development of advanced tools.

- WAN environments provides several benefits:
  - Reliability
  - Large computational resources
  - Access to vast quantities of data

- Applications that successfully exploits WAN environments need to take into consideration the several limitations as a result of the heterogeneous characteristics of WAN.

Problem Formulation
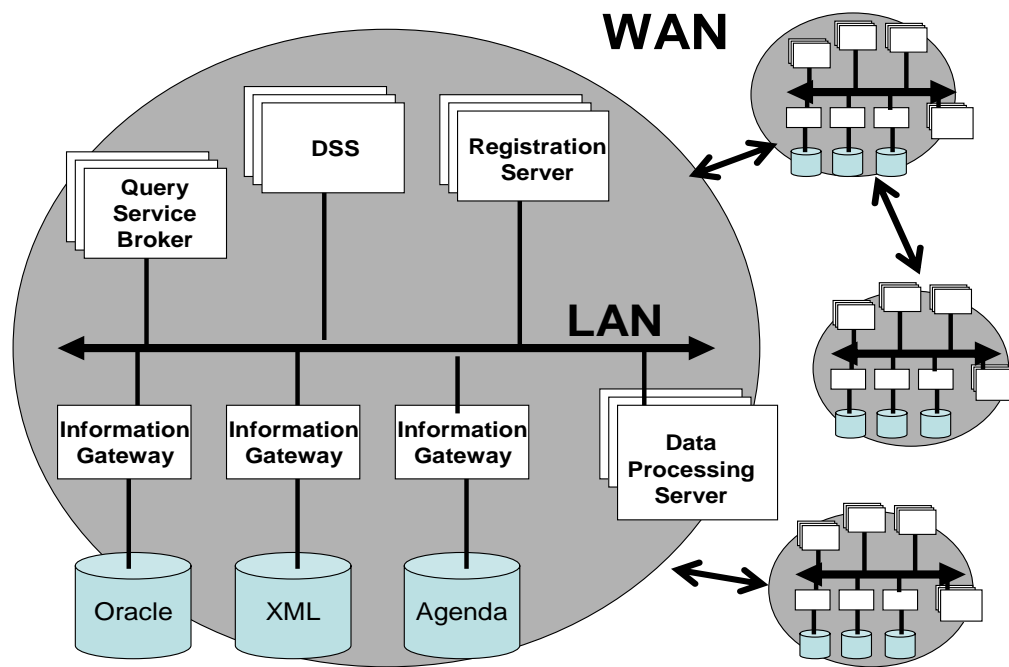
- Our solution: NetTraveler. The idea behind NetTraveler is to build a highly scalable Database Middleware Systems (DMS) that could efficiently exploit the capabilities of WAN environments.

- NetTraveler will facilitate and optimize the access of data across WANs.
  - Transparent data access
  - Uniform access interface

- NetTraveler is a server application implemented as Java Web Service (Axis toolkit).

Proposed Solution
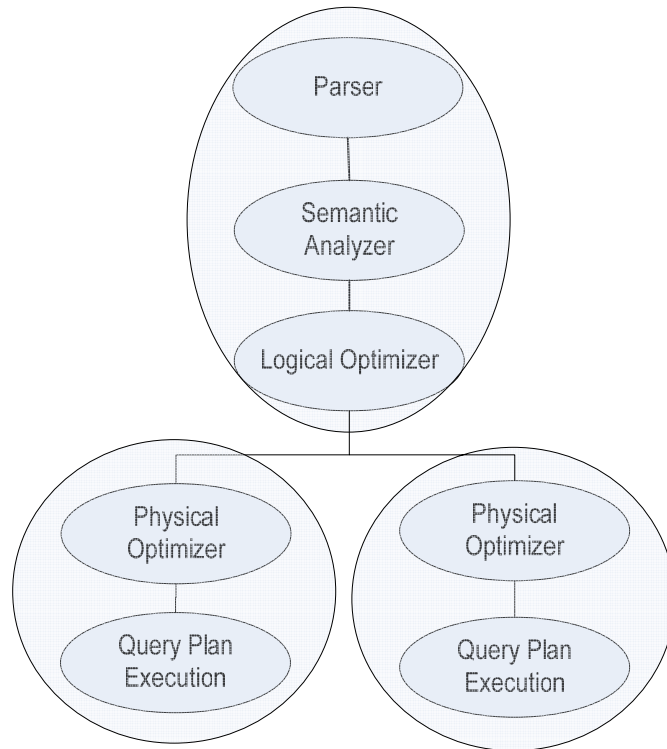
Decentralized Query Optimizer

- DMS usually try to implement a similar architecture as those implemented on traditional DBMS.

- One important area of research concerns the development of Query Optimizer for WAN.
  - Centralized Query Optimizers fails to scale to WAN.

- We proposed the development of a Decentralized Query Optimizer to efficiently exploit WAN environment.

- Another aspect comes from the possibility of exploiting replication and parallel execution.
  - Replication can be effectively used to solve distributed queries in a cooperative fashion by partition the query execution load among the replicas of a targeted data set

Proposed Solution
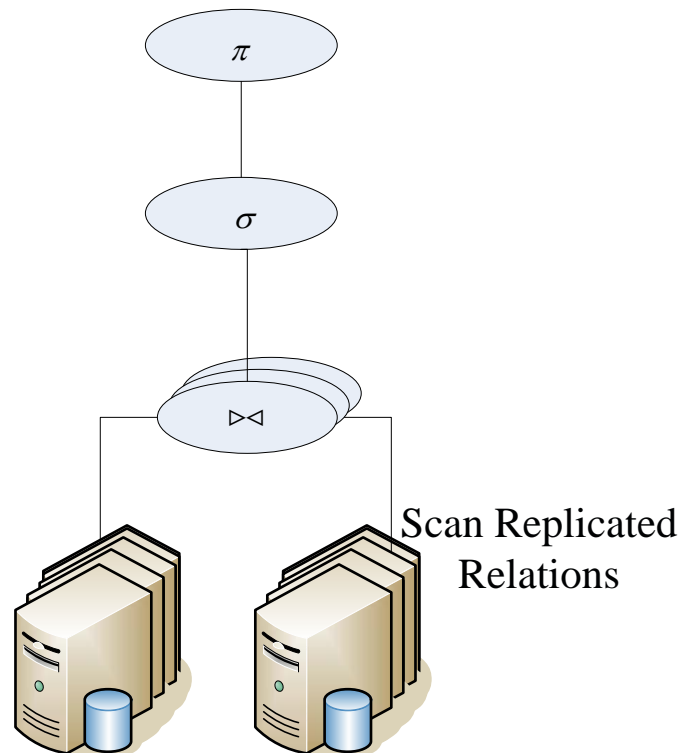
Distributed and Parallel
DBMS Plan

- To address this last issue we proposed the development of:
  - Monitoring and Scheduling mechanism for distributing the load across participating sites

  - Mechanism to support parallel and distributed query execution
    - Partitioned access to data sets

    - Parallel aware operators and coordinating operators