Lakshmikant Shrinivas
lshrinivas@wisc.edu
Ph: 608-628-2871

Interests: Database Systems, Query Processing and Optimization, Distributed Systems

Skills:

Programming Languages:	C, C++, Java, PERL, C#, 8085/8086 Assembly Language, Pascal, Scheme, BASIC
Operating Systems:	Linux, Unix, DOS, Windows 98/2000/XP, Mac OS X
Databases:	PostgreSQL, MySQL, DB2, MS SQL Server, Oracle, JDBC, J2EE (EJB)
Miscellaneous:	HTML, XML, XQuery, Xpath

Publications:

1. Issues in Applying Data Mining to Grid Job Failure Detection and Diagnosis – Poster appeared in ACM/IEEE International Symposium on High Performance Distributed Computing, Boston, June 2008.

Experience:

Ongoing: Data mining approaches to co-scheduling jobs and queries on a shared grid system – Parallel database systems are becoming popular to handle large scale OLAP applications. These systems are typically deployed on a cluster composed of commodity hardware, which could also be used for grid computing applications. Economic concerns naturally raise questions about whether sharing resources among the database and grid computation system is a good idea. I am currently investigating scenarios where grid jobs and parallel database queries can be co-scheduled to improve throughput, and the application of data mining to better guide the scheduling, based on past query and job workload. (submitted to ICDE 2009)

Using data mining for grid job failure diagnosis – A running grid computation system is awash in data; indeed there is so much data that it is difficult for users or administrators to get a big picture of the system. Data mining approaches have been shown to be useful in finding commonalities among failures, and thus aid users to find out what went wrong with their jobs, as well as predict when certain jobs are likely to fail. However, there are many ways of applying data mining to this problem, and it is not always clear which techniques are useful or feasible. In this project, I looked at the tradeoffs involved in applying data mining to the problem of grid job failure detection. (Poster in HPDC 2008)

Summer ’07: Internship at Microsoft – The aim of my project was to provide a back-end storage solution for small businesses desiring to use a storage cloud for their database needs. The goals were to support multiple tenants in the same database, provide easy creation and evolution of client side schemas, and to allow standard ADO.NET interfaces for data access. The solution revolved around storing data in XML format and translating client requests (which were in terms of the client side business objects) into SQL and XQuery over the underlying database and XML schemas.

Summer ’05: Internship at IBM Almaden Research Center – Businesses generate large volumes of semi structured and unstructured data, in the form of emails, spreadsheets, and office and XML documents. Putting together information from all these sources is a nightmare, given the large number of sources and formats. In this project, we explored techniques to provide query facilities over a best effort information warehouse. Our system correlated information from various sources, and allowed users to search and browse through the data, based on keywords and structural information present in the data. The results were ranked according to some notion of relevance, so that useful information is separated from noise.

Summer ’04: Internship at IBM Almaden Research Center – Modern business applications involve a lot of distributed data processing and inter-site communication, for which they rely on middleware products. These products provide the data access and communication framework for the business applications. During my internship I worked on a project that integrates messaging operations into a DBMS, so as to provide a single API

for data processing and communication through messages. We implemented a prototype system based on the IBM DB2 Universal Database and IBM Websphere middleware platform.

Sep '03 – Apr '04: Querying external scientific data using a database – People working with scientific data sets are reluctant to use relational DBMS's to manage their data for a variety of reasons, such as presence of legacy applications and scripts which work with the data. Our goal in this project is to enable the benefits of declarative relational query processing over scientific data that resides in the file system, external to the DBMS. The nature of scientific data presents some special challenges, which do not arise in commercial data. The work involved modifying an open source DBMS (PostgreSQL) to include support for querying external scientific data.

Aug '02 – Apr '03: Integrating Semi Joins in a Volcano based Distributed Query Optimizer – Enumerating join plans, with semi joins, for queries involving large relations, seems to be a promising technique for optimizing their execution. In the project I studied a Distributed Query Optimizer based on the *Volcano* algorithm, and modified it to incorporate semi joins. The work involved developing suitable *query transformation rules* and algorithms to apply them in a *cost based fashion*. This project was my Senior Undergraduate Thesis.

Jan '02 – Apr '02: Phoebe: An Artificial Creativity Engine for Generating Music - We developed a new model for creativity, based on Minsky's *Society of the Mind* model. Specifically for the domain of music, we developed *Phoebe*, an engine that can be trained with music samples and then asked to generate music on its own.

Aug '01 – Dec '01: An Online Stock Exchange - We designed and implemented a full size stock market system that distributes services over the internet and allows a huge number of concurrent biddings. The system then breaks deals between the users and updates its massive database of equity ownership.

Jan '01 – Apr '01: A Distributed Peer to Peer File Sharing System – We developed a Java based system for decentralized sharing of files on a Peer to Peer basis. The system identifies new hosts wanting to join the community. It also supports an intelligent file search and a smart download management approach to optimize multiple downloads.

Jan '02 – Apr '02: Compiler for a Pascal like language – We designed and implemented a compiler for a Pascal like language using the Java based lexer-generator *JFlex*, and the LALR parser-generator *CUP*.

Education:

Present: **Candidate for Ph.D. in Computer Science**
University of Wisconsin-Madison

May 2005: **Master of Science (M.S.) in Computer Science**
University of Wisconsin-Madison

Aug 2003: **Bachelor of Technology** (**B. Tech.) In Computer Science and Engineering**
Indian Institute of Technology, Bombay

Academic Honors:

- Awarded the *"Aditya Birla Scholarship (1999-2000)"*. The Scholarship, instituted by the *Aditya Birla Group of Companies,* is awarded each year to ten students from various IITs after a rigorous selection procedure, based on the theme *"Tomorrow's Leaders Today"*
- Ranked among the top 0.01% students appearing for *IIT-JEE* 99
- Awarded the *National Talent Search Scholarship*, 1997 (*N.C.E.R.T., New Delhi*)
- Actively participated in the *Regional and National Mathematics Olympiad* from Std. X to Std. XII. Stood 2nd in my state in the *Regional Math Olympiad (1997)*. Attended the *International Math Olympiad Training Camp* held at HBCSE in Std. XI and XII.

Extra Curricular Activities

- President of the *Indian Graduate Students Association* and *Wisconsin International Students Association* (student organizations at the *University of Wisconsin-Madison*) (2004-05)

- Organized various social and cultural activities as part of student organizations at the *University of Wisconsin-Madison* (2003-04).
- Completed a 7.0 km stretch of continuous swimming in 7.15 hours in the *Swimmathon* organized by the *Swimming Club, IIT Bombay* (2000)
- Won 1st prize in *Model Building Competition* in *TechFest*, a technological festival organized by *IIT Bombay* (2000)