Ivo Jimenez

PhD Student, UC Santa Cruz

ivo.jimenez@gmail.com

Summary

I'm a 3rd year PhD student at the UCSC Computer Science Department. I’m a member of the Systems Research Lab, advised by Professor Carlos Maltzahn.

Specialties: Object-oriented design; Java, C++, C; Linux (Unix-like environments); Database Research; Research prototype development; Systems Modeling

Experience

Graduate Student Researcher at University of California, Santa Cruz
September 2010 - Present (2 years 7 months)

Since 2013, I have been a member of the Systems Research Lab, working in the DAMASC project, an enhanced file system where rich data management services for scientific computing are provided as a native part of the file system

From 2010-2012 I was a member of the Database Group of the Computer Science Department, where I worked in the DBTune project, investigating automatic physical design tuning techniques.

Highlights include:

• Implemented a Semi-Automatic Index Tuning Tool. A demo-paper appeared on SIGMOD '12.
• Extended MySQL 5.5 to add what-if optimization capabilities.
• Designed a library for DBMS benchmarking. I have implemented the first components and is currently work-in-progress.
• Designed and implemented a Java API for doing physical design tuning.

Engineering Intern at Aster Data
June 2012 - December 2012 (7 months)

Member of the Data Analytics Platform Development team, researching and developing database components of Aster's highly scalable nCluster database

Technical Consultant at Colegio de Postgraduados
August 2010 - December 2011 (1 year 5 months)

I was a member of the GReNASeR group, where geographic information systems are deployed to compute large-scale environmental models that estimate metrics of interests such as carbon footprints or vegetation indices. I also served as the point of contact of a collaboration with Google.org.

Highlights:

• Redesigned the back end (in C++) of the main image processing system.
• Ran experiments to evaluate the feasibility of alternate database systems (in particular Hadoop and SciDB).
• Incorporated Agile development techniques to the group.

Research Associate at Hewlett-Packard Laboratories
August 2006 - December 2010 (4 years 5 months)

I was the technical lead on several one- and two-year-long projects. In general, my work consisted of interacting with business units to underst and their problems, translate them into formal models, implement them and transfer them back so they could merge these solutions into existing/new products.

Highlights:

• Software architect of the physical design advisor for HP's OLTP (NonStop) and OLAP/Analytical (Neoview) DBMSes. The project was divided in many phases, where the first one, entitled "QuickStart", was incorporated to the set of client tools that pre-sales personnel use. I was in charge of transferring the tool to the development team from the business unit that was responsible of it. We published a demo-paper in ICDE '09.

• Software architect on a project that focused on the obfuscation of relational data. We published an article in ICDE '10 (Industrial Track).

• As part of a collaboration agreement between HPL and the University of Sonora, I was in charge of coordinating the work of two interns.

Research Assistant at Universidad de Sonora
October 2003 - May 2006 (2 years 8 months)

Collaborated with the staff and faculty of the Math Optimization Lab at the Computer Science Department. I worked on the solution-modelling of combinatorial optimization problems, as well as the implementation of probabilistic metaheuristic algorithms to solve them.

One of the highlights of my involvement was the work on the PlaDiet project, which dealt with the planning and personalization of diet regimes. I designed the architecture and implemented the lower layers of it: data-access, back-end. This work won the 3rd place in the Science and Technology Fair of the University of Sonora. We also published an article describing the tool.

Summer Intern at CICESE Research Center
July 2005 - September 2005 (3 months)

Developed a Genetic Algorithm for DNA Sequencing by Hybridization with Isothermic Libraries.

Summer Intern at IIE Research Center

June 2004 - September 2004 (4 months)

Development of a GUI for a Neural-Network-based electrical controller modelling library

Summer Intern at Universidad de Sonora
June 2003 - September 2003 (4 months)

Developed a Search Engine Prototype called POPOTE, targeted towards the indexing of academic content. I designed the architecture and implemented the back-end while other colleagues implemented plug-ins (one for each different type of document [doc,pdf,etc]).

Courses

PhD, Computer Science
University of California, Santa Cruz

Design and Implementation of Database Systems	CMPS278
Computational Models	CMPS210
Machine Learning	CMPS242
Data Mining	TIM245
Principles of Database Systems	CMPS277
Storage Systems	CMPS229
Large-scale Information Extraction and Data Integration	CMPS290H
Big Data Systems	CMPS290S

Publications

Pladiet: Un Sistema de Cómputo para el Diseño de Dietas Individualizas Utilizando Algoritmos Genéticos
Revista Iberoamericana de Sistemas, Cibernética e Informática, International Institute of Informatics and Cybernetics July 1, 2007
Authors: Ivo Jimenez, Pedro Flores, Maria Guadalupe Cota, Desiderio Ramírez Romero, Lluvia Morales, Antonio, Raygoza, Samuel Galaviz, Armida Espinoza, Maria Esther Orozco

En este trabajo se presentan las bases de funcionamiento del sistema de cómputo Pladiet que calcula dietas individualizadas que cumplen con todos los requerimientos nutricionales establecidos por un experto. Para resolver el problema de calcular las dietas, se construye un modelo de Programación Entera no lineal que se resuelve utilizando Algoritmos Genéticos. Aquí se presenta la construcción del modelo, las características especificas del Algoritmo Genético con que se resuelve, y los resultados de las pruebas para ajuste de parámetros correspondientes. Pladiet es el resultado de una investigación conjunta entre los Departamentos de Nutrición y Ciencias de la Computación de la Universidad de Sonora en México y actualmente esta en uso en el Modulo de Orientación Nutricional (MON) de dicha Universidad, en donde se atienden anualmente a aproximadamente 400 pacientes con 2000 citas programadas.

QuickStart: An Upfront Client-Based Design Advisor for Parallel Data Warehouses
2009 IEEE International Conference on Data Engineering April 2, 2009
Authors: Ivo Jimenez, Malu G Castellanos, Umeshwar Dayal, Neal Coddington, Hans Zeller, Steven Euijong Whang

QuickStart is a tool to automate the physical design of data warehouses for HP's Neoview system. It has been researched and prototyped at HP Labs with close interaction from Neoview design experts. It embodies heuristics and best practices of the experts to search for candidate physical features and uses cost calculations to recommend the features that result in good designs. It has some unique characteristics that differentiate it from other physical design advisors. In particular, it is the only advisor that is client-based, does not require a DBMS server installation and can work off a laptop by simply connecting to the customers' flat files. Another unique characteristic of QuickStart is that it provides the rationale for its recommendations.

Data Desensitization of Customer Data for Use in Optimizer Performance Experiments
2010 IEEE 26th International Conference on Data Engineering (ICDE) April 2010
Authors: Ivo Jimenez, bin zhang, Malu G Castellanos, Perla Ruiz, Miguel Durazo, Umeshwar Dayal

Improving the performance and functionality of database system optimizers requires experimentation on real customer data. Often these data are of sensitive nature and the only way to keep them is by applying a non-reversible transformation to obfuscate them. However, in order that the database optimizer generates exactly the same query plans as for the sensitive data, the transformation has to preserve the order and some important properties of the data distribution. Unfortunately, existing data obfuscation techniques do not preserve all of these properties and therefore are not applicable in this context. In this paper we present a Desensitizer tool that we have developed for optimizer performance experiments of HP’s Neoview high availability data warehousing product. The tool is based on novel numeric and string desensitization algorithms which are agnostic to the database system. We explain the core concepts behind the algorithms, how they preserve the required data properties and important implementation considerations that were made. We present the architecture of the Desensitizer tool and results of the extensive validation that we conducted.

Benchmarking Online Index-Tuning Algorithms
IEEE Data Engineering Bulletin November 2011
Authors: Ivo Jimenez, Alkis Polyzotis, Jeff LeFevre, Huascar Sanchez, Karl Schnaitter

The topic of index tuning has received considerable attention in the research literature. However, very few studies provide a comparative evaluation of the proposed index tuning techniques in the same environment and with the same experimental methodology. In this paper, we outline our efforts in this direction with the development of a performance benchmark for the specific problem of online index tuning. We describe the salient features of the benchmark, present some representative results on the evaluation of different index tuning techniques, and conclude with lessons we learned about implementing and running a benchmark for self tuning systems.

Kaizen: A Semi-Automatic Index Advisor
Proceedings of the 2012 international conference on Management of Data May 2012
Authors: Ivo Jimenez, Alkis Polyzotis, Quoc Trung Tran

Index tuning; i.e., selecting indexes that are appropriate for the workload to obtain good system performance, is a crucial task for database administrators. Existing automatic index advisors either work in offline or

online mode. A new recommendation paradigm, termed Semi-automatic Index Tuning, has been proposed recently; it provides a novel feedback mechanism to enable DBAs to interactively refine index recommendations. In this demonstration we introduce Kaizen, an index tuning tool that implements a semi-automatic algorithm that works online to avoid the limitations of commercial tools that require the workload to be known in advance or that ignore the DBA entirely. The feedback and online features of the tool are shown visually in this demonstration.

Patents

Desensitizing Database Information
United States Patent 7797341 Issued April 30, 2007
Inventors: Ivo Jimenez

In a method of preserving characteristics of desensitized database information for use in database management system benchmarking, at least one column of sensitive data from a database is accessed. The at least one column of sensitive data is sorted into a sorted data representation. Desensitized data is generated by sequentially mapping ascending different values of the sorted data representation to ascending desensitized values generated according to a strictly monotone increasing randomly generated function. The mapping results in a plurality of mapped desensitized values which are associated with the sorted data representation. The mapped desensitized values are unsorted into a desensitized database according to sorting information related to the sorted data representation. As a result of the unsorting, cardinalities of the at least one column of sensitive data are maintained within the desensitized database.

Skills & Expertise

C++
Unix
Linux
Ubuntu
Red Hat Linux
Java
Scala
R
Weka
Octave
Mathematica
Databases
Systems Modeling
C
Bash
Object Oriented Design
Haskell
SQL
Relational Databases
Data Warehousing

Database Design
Research
Experimentation
Rapid Prototyping
Agile Methodologies
Subversion
MySQL
Eclipse
Data Mining

Education

University of California, Santa Cruz
PhD, Computer Science, 2010 - 2015

Universidad de Sonora
BS, Computer Science, 2001 - 2006
Activities and Societies: Student Society, Semana de la Computación (Computation Week) organization committee

Honors and Awards

UCMEXUS Doctoral Fellowship 2010

Interests

Databases, music, photography, skateboarding, capoeira

Ivo Jimenez

PhD Student, UC Santa Cruz

ivo.jimenez@gmail.com

Contact Ivo on LinkedIn