



23rd International Conference on Scientific and Statistical Database Management

2011 Conference Program

July 20 - July 22, 2011
Portland, Oregon

Welcome from the General Chair

Welcome to the 23rd International Conference on Scientific and Statistical Database Management in Portland, Oregon. This year marks the return of SSDBM to the Pacific Northwest, having previously been held in Olympia, Washington (1997). The shift of the conference time from winter to summer may deprive you of enjoying the famous Portland rain, but I hope that lack is compensated by the other attractions of the city: its brewpubs and baristas; the urban parks and gardens; the wineries and farmers markets; and the nearby natural wonders of the Columbia River Gorge, the Oregon Coast, the High Dessert near Bend and Mount St. Helens. And don't miss trying one of the city's 300+ food carts while you're here.

SSDBM is an independent conference, and succeeds by the efforts of its all-volunteer organizing committee. On the technical side, I thank Judy Cushing and Jim French for their recruitment of the program committee and oversight of the review process. Shawn Bowers has excelled as proceedings editor and EasyChair wrangler. Thanks as well to the PC members and ancillary reviewers. In the financial realm, Len Shapiro and Michael Grossniklaus set up SSDBM 2011 as an Oregon Corporation, and Michael oversees our budget and banking. Bill Howe ran the registration site and was instrumental in arranging our sponsorships. Locally, Laura Bright and Kristin Tufte have managed our arrangements within the hotel, as well as doing tireless research on a suitable restaurant for our dinner. Dave Hansen is overseeing all the student volunteers you see around you. Thanks as well to Stephanie Lewis and the staff of University Place. For conference communications, Pete Tucker has maintained our web site and David Chiu has coordinated announcements and information for SSDBM participants. I also thank the SSDBM Steering Committee, particularly Arie Shoshani for his corporate memory, and Michael Gertz for information on the Heidelberg conference (and surplus funds!).

I thank our Gold Sponsor, Microsoft Research, and Silver Sponsors, the eScience Institute at the University of Washington, The Gordon and Betty Moore Foundation and Paradigm4 for their generous support of this year's conference. Their generous contributions have helped support discounted registration for students, the keynote speaker, and social events for student volunteers. Also, I express our gratitude to Springer-Verlag, for their continuing role as our proceedings publisher.

Finally, I thank all those who submitted papers and proposed panels. Your interest and participation is what keeps the quality of SSDBM high and the topics timely. Anyone for SSDBM 2012 in Crete?

David Maier
General Chair

Message from the Program Co-Chairs

We are pleased to offer this year's high quality research and thank all those who are participating and attending the conference – as well as those who submitted papers. We also express our sincere appreciation to the 65 members of the program committee and 32 additional reviewers (recruited for specialized expertise) for their hard work and dedication.

In addition to our keynote speaker, Prof. Michael Stonebraker who will inaugurate the conference with *The Architecture of SciDB*, we have two excellent panels: *Data Intensive Science: Moving Towards Solutions* chaired by Dr. Terence Critchlow and *Data Scientists, Data Management and Data Policy* chaired by Dr. Sylvia Spengler. These “hot topics” were selected to promote fruitful discussion on directions of the scientific and statistical data management field.

As always, the “main course” of the conference is the research presented: 23 long and 12 short papers spanning the field from the more practical (*Applications and Models, and Architectures and Privacy, Workflows and Provenance*) to more theoretical (*Clustering and Data Mining, Ranked Search, Temporal Data and Queries, and Graph Querying*). In addition to research presentations, the poster and demo session, to be held during an extended Thursday afternoon break from presentations, is to increase active participation in the present conference. Fifteen posters will be presented, as well as three demos (chosen from the Friday *Applications and Models Session*).

Please enjoy the conference, and feel free to call on either of us during the conference!

Judith Bayard Cushing and James French

Conference Schedule

Tuesday, July 19, 2011

18:00 – 20:00 | **Registration**
Location: University Place Lobby

Wednesday, July 20, 2011

8:00 – 17:00 | **Registration**
Location: Outside Columbia Falls Ballroom

8:30 | **Opening**

8:45 | **Keynote Presentation:**
The Architecture of SciDB, Michael Stonebraker

9:45 | **Coffee Break**
Location: Wahkeena Falls Room

10:15 – 12:15 | **Session 1: Ranked Search**
Session Chair: James Frew

- **Location-Based Instant Search**
Shengyue Ji and Chen Li
- **Continuous Inverse Ranking Queries in Uncertain Streams**
Thomas Bernecker, Hans-Peter Kriegel, Nikos Mamoulis, Matthias Renz and Andreas Zuefle
- **Finding Haystacks with Needles: Ranked Search for Data Using Geospatial and Temporal Characteristics**
V.M. Megler and David Maier
- **Using Medians to Generate Consensus Rankings for Biological Data**
Sarah Cohen-Boulakia, Alain Denise and Sylvie Hamel
- **A Truly Dynamic Data Structure for Top-k Queries on Uncertain Data**
Manish Patil, Rahul Shah and Sharma V. Thankachan

12:30 | **Lunch Break**
Location: Willamette Falls Room

Conference Schedule

13:30 – 15:00

Session 2: Temporal Data and Queries

Session Chair: Magdalena Balazinska

- **Efficient Storage and Temporal Query Evaluation in Hierarchical Data**
Hui Wang, Ruilin Liu, Dimitri Theodoratos and Xiaoying Wu
- **Update Propagation in a Streaming Warehouse**
Theodore Johnson and Shkapenyuk Vladislav
- **Efficient Processing of Multiple DTW Queries in Time Series**
Hardy Kremer, Stephan Günemann, Anca-Maria Ivanescu, Ira Assent and Thomas Seidl
- **Probabilistic Time Consistent Queries over Moving Objects**
Xiang Lian and Lei Chen

15:00

Coffee Break

Location: Wahkeena Falls Room

16:00 – 17:20

Session 3: Workflows and Provenance

Session Chair: Tamas Budavari

- **Knowledge Annotations in Scientific Workflows: An Implementation in Kepler**
Aida Gandara, George Chin, Paulo Pinheiro Da Silva, Chandrika Sivaramakrishnan, Signe White and Terence Critchlow
- **Improving Workflow Fault Tolerance through Provenance-based Recovery**
Sven Koehler, Sean Riddle, Daniel Zinn, Timothy McPhillips and Bertram Ludaescher
- **ProPub: A Declarative Approach for Publishing Customized, Policy-Aware Provenance**
Saumen C Dey, Daniel Zinn and Bertram Ludaescher
- **Provenance-Enabled Automatic Data Publishing**
James Frew, Greg Janée and Peter Slaughter

17:30 – 19:00

Reception

Location: Poolside, University Place

Conference Schedule

Thursday, July 21, 2011

8:00	Poster and Demo authors initial setup <i>Location: Wahkeena Falls Room</i>
8:30 – 10:00	Panel - Data Intensive Science: Moving Towards Solutions Panel Chair: Terence Critchlow <ul style="list-style-type: none">• Stefan Heinzl (Rechenzentrum Garching der Max-Plank-Gesellschaft)• Tony Hey (Microsoft Research)• Bill Howe (University of Washington)• Kerstin Kleese Van Dam (Pacific Northwest National Laboratory)
10:00	Coffee Break <i>Location: Wahkeena Falls Room</i>
10:30 – 12:00	Session 4: Querying Graphs Session Chair: Alfredo Cuzzocrea <ul style="list-style-type: none">• Querying Shortest Path Distance with Bounded Errors in Large Graphs Miao Qiao, Hong Cheng and Jeffrey Xu Yu• PG-Join: Proximity Graph Based String Similarity Joins Michail Kazimianec and Nikolaus Augsten• A Flexible Graph Pattern Matching Framework via Indexing Wei Jin and Jiong Yang• Subgraph Search over Massive Disk Resident Graphs Peng Peng, Lei Zou, Dongyan Zhao, Lei Chen and Xueming Lin• BR-Index: An Indexing Structure for Subgraph Matching in Very Large Dynamic Graphs Jiong Yang and Wei Jin
12:00	Lunch Break <i>Location: Willamette Falls Room</i>

Conference Schedule

13:00 – 14:40

Session 5: Clustering and Data Mining

Session Chair: Tony Hey

- **CloudVista: Visual Cluster Exploration for Extreme Scale Data in the Cloud**
Keke Chen, Huiqi Xu, Fengguang Tian and Shumin Guo
- **Efficient Selectivity Estimation by Histogram Construction based on Subspace Clustering**
Andranik Khachatryan, Emmanuel Müller, Klemens Böhm and Jonida Kopper
- **Finding Closed MEMOs**
Htoo Htet Aung and Kian-Lee Tan
- **Density Based Subspace Clustering Over Dynamic Data**
Hans-Peter Kriegel, Peer Kröger, Irene Ntoutsi and Arthur Zimek
- **Hierarchical Clustering for Real-Time Stream Data with Noise**
Philipp Kranen, Felix Reidl, Fernando Sanchez Villaamil and Thomas Seidl

14:45

Demo and Poster Session

Location: Wahkeena Falls Room

16:30 – 18:00

Session 6: Architectures and Privacy

Session Chair: Terence Critchlow

- **Energy Proportionality and Performance in Data Parallel Computing Clusters**
Jinoh Kim, Jerry Chou and Doron Rotem
- **Privacy-Preserving Group Linkage**
Fengjun Li, Yuxin Chen, Bo Luo, Dongwon Lee and Peng Liu
- **Dynamic Anonymization for Marginal Publication**
Xianmang He, Yanghua Xiao, Yujia Li, Qing Wang, Wei Wang and Baile Shi
- **Pantheon: Exascale File System Search for Scientific Computing**
Joseph L. Naps, Mohamed F. Mokbel and David H. C. Du
- **Massive-scale RDF Query Processing Using Compressed Bitmap Indexes**
Kamesh Madduri and Kesheng Wu
- **SQLShare: Database-as-a-Service for Long Tail Science**
Bill Howe, Garret Cole, Alicia Key, Nodira Khoussainova, Leilani Battle, Emad Soroush and Paraschos Koutris

19:00 – 21:00

Conference Banquet at *Thirst*, Portland Waterfront. Student volunteers and organizers to meet attendees at hotel lobby starting **18:30** to walk over.

Conference Schedule

Friday, July 22, 2011

9:00 – 10:00	Panel - Data Scientists, Data Management, and Data Policy Panel Chair: Sylvia Spengler
10:00	Coffee Break <i>Location: Wahkeena Falls Room</i>
10:30 – 12:05	Session 7: Application and Models Session Chair: Arie Shoshani <ul style="list-style-type: none">• Context-Aware Parameter Estimation for Forecast Models in the Energy Domain Lars Dannecker, Robert Schulze, Matthias Böhm, Wolfgang Lehner and Gregor Hackenbroich• Implementing a General Spatial Indexing Library for Relational Databases of Large Numerical Simulations Gerard Lemson, Tamas Budavari and Alexander Szalay• Histogram and Other Aggregate Queries in Wireless Sensor Networks Khaled Ammar and Mario Nascimento• Efficient In-Database Maintenance of ARIMA Models Frank Rosenthal and Wolfgang Lehner• Recipes for Baking Black Forest Databases Julio López, Colin Degraf, Tiziana Di Matteo, Bin Fu, Eugene Fink and Garth Gibson• CrowdLabs: Social Analysis and Visualization for the Sciences Phillip Mates, Juliana Freire, Claudio Silva and Emanuele Santos
12:05	Closing Session
12:30	Adjourn

Posters and Demos

Demos

Papers corresponding to these demos can be read in *Session 7: Applications and Models*.

- **Context-Aware Parameter Estimation for Forecast Models in the Energy Domain**
Lars Dannecker, Robert Schulze, Matthias Böhm, Wolfgang Lehner and Gregor Hackenbroich
- **Implementing a General Spatial Indexing Library for Relational Databases of Large Numerical Simulations**
Gerard Lemson, Tamas Budavari and Alexander Szalay
- **CrowdLabs: Social Analysis and Visualization for the Sciences**
Phillip Mates, Juliana Freire, Claudio Silva and Emanuele Santos

Posters

- **Heidi Visualization of R-tree Structures over High Dimensional Data**
Shraddha Agrawal, Soujanya Vadapalli and Kamalakar Karlapalem
- **Towards Efficient and Precise Queries Over Ten Million Asteroid Trajectory Models**
Yusra Alsayyad, K. Simon Krughoff, Bill Howe, Andrew J. Connolly, Magdalena Balazinska and Lynne Jones
- **Keyword Search Support for Automating Scientific Workflow Composition**
David Chiu, Travis Hall, Farhana Kabir and Gagan Agrawal
- **FastQuery: A General Indexing and Querying System for Scientific Data**
Jerry Chou, John Wu and Prabhat
- **Retrieving Accurate Estimates to OLAP Queries over Uncertain and Imprecise Multidimensional Data Streams**
Alfredo Cuzzocrea

Posters and Demos

Posters (cont.)

- **Hybrid Data-Flow Graphs for Procedural Domain-Specific Query-Languages**
Bernhard Jaecksch, Franz Faerber, Wolfgang Lehner and Frank Rosenthal
- **Scalable and Automated Workflow in Mining Large-Scale Severe-Storm Simulations**
Lei Jiang, Gabrielle Allen and Qin Chen
- **Accurate Cost Estimation Using Distribution-based Cardinality Estimates for Multi-Dimensional Queries**
Andranik Khachatryan and Klemens Böhm
- **Session-based Browsing for Better Query Reuse**
Nodira Khoussainova, Yongchul Kwon, Wei-Ting Liao, Magdalena Balazinska, Wolfgang Gatterbauer and Dan Suciu
- **The ETLMR MapReduce-based ETL Framework**
Xiufeng Liu, Christian Thomsen and Torben Bach Pedersen
- **Top-k Similarity Search on Uncertain Trajectories**
Chunyang Ma, Hua Lu, Lidan Shou, Gang Chen and Shujie Chen
- **Fast and Accurate Trajectory Streams Clustering**
Elio Masciari
- **Data-Driven Multidimensional Design for OLAP**
Oscar Romero and Alberto Abelló
- **An Adaptive Outlier Detection Technique for Data Streams**
Shiblee Sadik and Le Gruenwald
- **Power-Aware DBMS: Potential and Challenges**
Zichen Xu, Yicheng Tu and Xiaorui Wang

Conference Organization

Organizing Committee

General Chair:

David Maier, Portland State University

Assistant Chair:

Leonard Shapiro, Portland State University

PC Co-Chair:

Judith Cushing, The Evergreen State College

PC Co-Chair:

James French, CNRI and University of Virginia

Proceedings Editor:

Shawn Bowers, Gonzaga University

Information Officer:

David Chiu, Washington State University, Vancouver

Web and Publicity:

Peter Tucker, Whitworth University

Registration:

Bill Howe, University of Washington

Treasurer:

Michael Grossniklaus, Portland State University

Local Arrangements (food, events, scheduling):

Laura Bright, McAfee

Local Arrangements (AV, computing):

Kristin Tufte, Portland State University

Student Volunteers:

David Hansen, George Fox University

Steering Committee

Michael Gertz, Univ. of Heidelberg

Bertram Ludäscher, UC Davis

Nikos Mamoulis, Univ. of Hong Kong

Arie Shoshani, Lawrence Berkeley National Laboratory (chair)

Marianne Winslett, Univ. of Illinois

Conference Organization

Program Committee

Ken Barker, University of Calgary
Randal Burns, Johns Hopkins University
Sarah Cohen-Boulakia, University of Paris-Sud 11
Isabel Cruz, University of Illinois at Chicago
Alfredo Cuzzocrea, University of Calabria
Nilesh Dalvi, USC Information Sciences Institute
Ewa Deelman, Information Sciences Institute
Dejing Dou, University of Oregon
Amr El Abbadi, University of California, Santa Barbara
Juliana Freire, University of Utah
James Frew, University of California, Santa Barbara
Johann Gamper, Free University of Bozen-Bolzano
Michael Gertz, University of Heidelberg
Carole Goble, University of Manchester
Michael Goodchild, University of California, Santa Barbara
Wilfried Grossmann, University of Vienna
Dimitrios Gunopulos, University of Athens
Amarnath Gupta, San Diego Supercomputer Center
Theo Härder, University of Kaiserslautern
Bill Howe, University of Washington
Ray Idaszak, RENCI, UNC
H. V. Jagadish, University of Michigan
Matthias Jarke, RWTH Aachen
Chris Jermaine, Rice University
Matthew Jones, University of California, Santa Barbara
Conny Junghans, University of Heidelberg
Jessie Kennedy, Napier University
Larry Kerschberg, George Mason University
Martin Kersten, Centrum Wiskunde & Informatica
Hans-Joachim Klein, University of Kiel
Peer Kröger, Ludwig-Maximilians University of Munich

Conference Organization

Program Committee (cont.)

Zoe Lacroix, Arizona State University

Ulf Leser, Humboldt University of Berlin

Feifei Li, Florida State University

Bertram Ludäscher, University of California, Davis

Yannis Manolopoulos, Aristotle University of Thessaloniki

Claudia Bauzer Medeiros, University of Campinas

Kyriakos Mouratidis, Singapore Management University

Wolfgang Mueller, HITS gGmbH

Silvia Nittel, University of Maine

Frank Olken

Beng Chin Ooi, National University of Singapore

Gultekin Ozsoyoglu, Case Western Reserve University

Andreas Reuter, HITS gGmbH

Philippe Rigaux, Université Paris-Dauphine

Kenneth Ross, Columbia University

Doron Rotem, Lawrence Berkeley National Laboratory

Nagiza Samatova, North Carolina State University

Linda Shapiro, University of Washington

Sylvia Spengler, National Science Foundation

Jianwen Su, U C Santa Barbara

Kian-Lee Tan, National University of Singapore

Yufei Tao, Chinese University of Hong Kong

Dimitri Theodoratos, New Jersey Institute of Technology

Shengru Tu, University of New Orleans

Can Türker, Functional Genomics Center Zürich

Andrew Westlake, Survey and Statistical Computing

Kesheng Wu, Lawrence Berkeley National Laboratory

Yan Xu, Microsoft Research

Jeffrey Yu, Chinese University of Hong Kong

Xiangliang Zhang, King Abdullah University of Science and Technology (KAUST)

Daniel Zinn, University of California, Davis

Conference Organization

Student Volunteers

Scott Britell, Portland State University

Fredton Doan, Washington State University, Vancouver

Brent Dombrowski, Portland State University

Hema Kumar, Portland State University

Patrick Leyshock, Portland State University

Jesse Lopez, CMOP-OHSU

Sharmadha Moorthy, Portland State University

Mandar Patil, Portland State University

Helena Scheuble, Washington State University, Vancouver

Phillip Jacques Sitbon, Portland State University

Jeremy Steinhauer, Portland State University

James Whiteneck, Portland State University

Mark Wong, Portland State University

More Conference Info

Social Events

Wednesday Reception

The reception Wednesday will take place on the patio by the pool, from 17:30 - 19:00. It will feature Oregon microbrew beers and light appetizers. Please wear your name badge and bring your mug.

Thursday Banquet

The banquet Thursday will be at *Thirst* restaurant on the Willamette riverfront, from 19:00 - 21:00. It will feature Oregon wines (including a Pinot Noir tasting), heavy appetizers and dessert. We plan for participants to walk down (see the map in your packet). For anyone wanting guidance, meet in the hotel lobby starting at 18:30, and we will have conference organizers and student volunteers escort groups down. (If you have mobility issues, contact one of the conference staff, and we can arrange a ride.) We plan on having shuttle service back to the hotel available afterward, for those who want it.

SSDBM 2011 Is Eco-Friendly

Portland is hailed as the “Greenest City in the US.” In 2008 *Popular Science* ranked it #1 in this category, and according to *Grist Magazine*, it is second only to Reykjavik, Iceland globally. Half the city’s power comes from renewable energy, and it has 100+ buildings certified by the US Green Building Council. Portland is the first city to enact an extensive plan to reduce CO2 emissions, and as part of this initiative, its transportation system is designed to keep cars off the road. As a consequence, a quarter of the workforce commutes by bike, carpool, or public transportation. The Portland area boasts 92,000 acres of green space and more than 74 miles of hiking, running, and biking trails. The spirit of fostering sustainability has also guided several decisions made by the SSDBM 2011 organizing committee.

- Your registration kit does not include a conference bag, which is often produced from non-recyclable materials. Also, the majority of these bags are tossed during or shortly after the conference, which has an adverse effect on the environment.
- As a conference souvenir we wanted to have something that serves a purpose. We have provided you with a branded mug and encourage you to carry it with you during the conference. At the reception, you can use it to taste some of the revered local beers, and during the session, it replaces water bottles. Additionally, the lanyard on your badge is made from recycled materials.
- Finally, we encourage you to make frequent use of Portland’s extensive public transport system. The streetcar is only a few blocks from the hotel and is ideal for exploring this great city. Downtown Portland is a “Free Fare Zone” for all rail transport. Refer to the city map in your registration kit to reveal the boundaries of the “Free Fare Zone.”

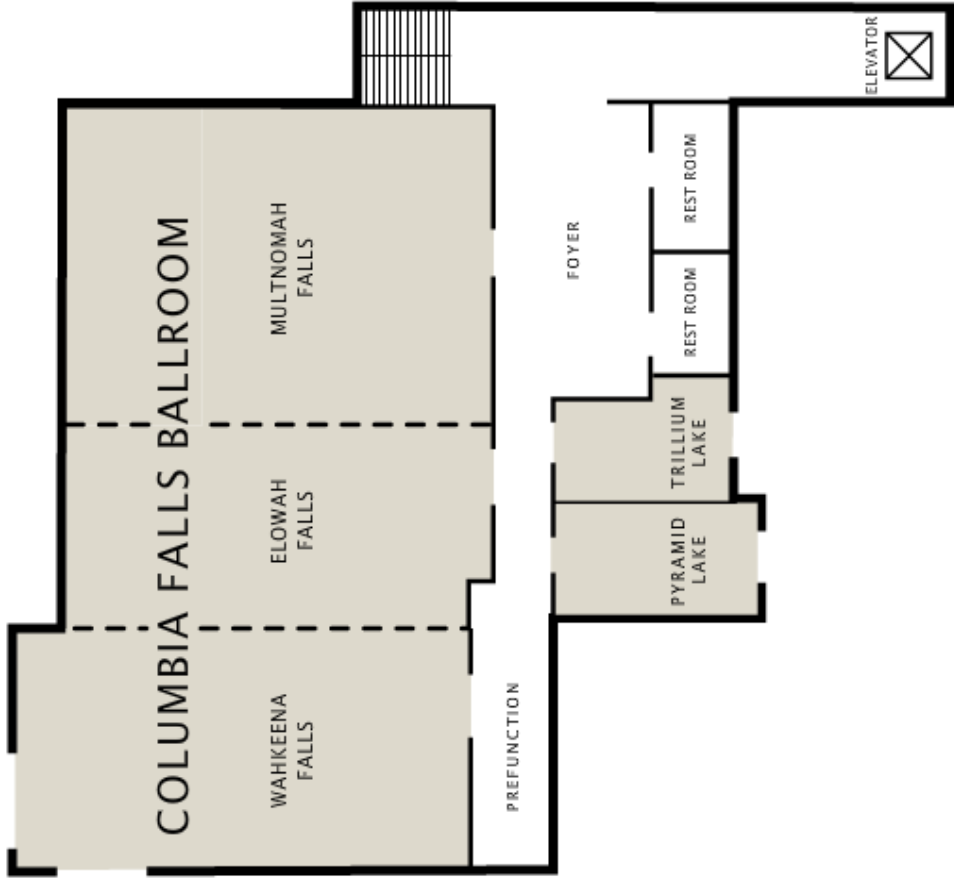
Conference Venue Map: Level 1

LEVEL ONE



Conference Venue Map: Level 2

LEVEL TWO



This international conference brings together scientific domain experts, database researchers, practitioners and developers for the presentation and exchange of current research on concepts, tools and techniques for scientific and statistical database applications. SSDBM provides a forum for original research contributions and practical system design, implementation and evaluation. Individual themes differ year to year with the main focus remaining on database theory and application in the scientific and statistical fields. Past themes have included bioinformatics (genomics, biodiversity informatics including biological databases), geospatial and sensor databases, geological databases, data mining, metadata management, conceptual models, data integration and visualization and system architectures. The priority areas are supplemented with invited talks and panel sessions and illustrated with demonstrations of research prototypes and industrial systems. The conference takes place in a single plenary session to facilitate the viewing of all presentations by all attendees. The conference is hosted by institutions alternating every other year between North America and abroad.



Gold Sponsor



Silver Sponsors

