

VLDB2015

41st International Conference on Very Large Data Bases, Kohala Coast, Hawaii



Proceedings of the VLDB Endowment

Volume 8, No. 12 – August 2015

**Proceedings of the 41st International Conference on
Very Large Data Bases, Kohala Coast, Hawaii**

Program Chairs and Editors-in-Chief:

Chen Li and Volker Markl

Associate Editors – Research and Innovative Systems Tracks:

**Kevin Chang, Shivnath Babu, Magdalena Balazinska, Felix Naumann, Stefan Manegold, Yi Chen, Fatma Ozcan,
Jignesh Patel**

Associate Editors – Experiments and Analysis Track:

Rainer Gemulla

Proceedings Chairs:

Tyson Condie, Daisy Zhe Wang

PVLDB – Proceedings of the VLDB Endowment

Volume 8, No. 12, August 2015.

The 41st International Conference on Very Large Data Bases, Kahola Coast, Hawaii.

Copyright 2015 VLDB Endowment

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>. Obtain permission prior to any use beyond those covered by the license. Contact copyright holder by emailing info@vldb.org.

Volume 8, Number 12, August 2015: VLDB 2015

Pages ii - xiv and 1334 - 2058

ISSN 2150-8097

Additional copies only online at: portal.acm.org, arxiv.org/corr, and www.vldb.org

TABLE OF CONTENTS

Front Matter

Copyright Notice	ii
Table of Contents	iii
VLDB 2015 Organization and Review Board	x

Letters

Letter from Program Committee Chairs	Chen Li and Volker Markl	xiv
--	--------------------------	-----

Research Papers

Aggregate Estimations over Location Based Services.....	1334
..... <i>Weimo Liu, Md Farhadur Rahman, Saravanan Thirumuruganathan, Nan Zhang, Gautam Das</i>	
Principles of Dataset Versioning: Exploring the Recreation/Storage Tradeoff	1346
..... <i>Souvik Bhattacharjee, Amit Chavan, Silu Huang, Amol Deshpande, Aditya Parameswaran</i>	
SEMA-JOIN: Joining Semantically-Related Tables Using Big Table Corpora.....	1358
..... <i>Yeye He, Kris Ganjam, Xu Chu</i>	
Stale View Cleaning: Getting Fresh Answers from Stale Materialized Views	1370
..... <i>Sanjay Krishnan, Jiannan Wang, Michael J. Franklin, Ken Goldberg, Tim Kraska</i>	
Compressed Spatial Hierarchical Bitmap (cSHB) Indexes for Efficiently Processing Spatial Range Query Workloads	1382
..... <i>Parth Nagarkar, K. Selçuk Candan, Aneesha Bhat</i>	
Selective Provenance for Datalog Programs Using Top-K Queries	1394
..... <i>Daniel Deutch, Amir Gilad, Yuval Moskovitch</i>	
Processing of Probabilistic Skyline Queries Using MapReduce.....	1406
..... <i>Yoonjae Park, Jun-Ki Min, Kyuseok Shim</i>	
Bonding Vertex Sets Over Distributed Graph: A Betweenness Aware Approach	1418
..... <i>Xiaofei Zhang, Hong Cheng, Lei Chen</i>	
A Natural Language Interface for Querying General and Individual Knowledge.....	1430
..... <i>Yael Amsterdamer, Anna Kukliansky, Tova Milo</i>	
Scaling Up Concurrent Main-Memory Column-Store Scans: Towards Adaptive NUMA-aware Data and Task Placement.....	1442
..... <i>Iraklis Psaroudakis, Tobias Scheuer, Norman May, Abdelkader Sellami, Anastasia Ailamaki</i>	
SQLite Optimization with Phase Change Memory for Mobile Applications.....	1454
..... <i>Gihwan Oh, Sangchul Kim, Sang-Won Lee, Bongki Moon</i>	
<i>An Architecture for Compiling UDF-centric Workflows.....</i>	1466
<i>Andrew Crotty, Alex Galakatos, Kayhan Dursun, Tim Kraska, Carsten Binnig, Ugur Cetintemel, Stan Zdonik</i>	

A Scalable Distributed Graph Partitioner	1478
..... <i>Daniel Margo, Margo Seltzer</i>	
Take me to your leader! Online Optimization of Distributed Storage Configurations	1490
..... <i>Artyom Sharov, Alexander Shraer, Arif Merchant, Murray Stokely</i>	
Association Rules with Graph Patterns.....	1502
..... <i>Wenfei Fan, Xin Wang, Yinghui Wu, Jingbo Xu</i>	
Fuzzy Joins in MapReduce: An Experimental Study	1514
..... <i>Ben Kimmatt, Venkatesh Srinivasan, Alex Thoma</i>	
PARADIS: An Efficient Parallel Algorithm for In-place Radix Sort	1518
..... <i>Minsik Cho, Daniel Brand, Rajesh Bordawekar, Ulrich Finkler, Vincent Kulandaisamy, Ruchir Puri</i>	
Join Size Estimation Subject to Filter Conditions.....	1530
..... <i>David Vengerov, Andre Cavalheiro Menck, Mohamed Zait, Sunil P. Chakkappen</i>	
Asynchronous and Fault-Tolerant Recursive Datalog Evaluation in Shared-Nothing Engines	1542
..... <i>Jingjing Wang, Magdalena Balazinska, Daniel Halperin</i>	
Maximum Rank Query.....	1554
..... <i>Kyriakos Mouratidis, Jilian Zhang, HweeHwa Pang</i>	
Performance and Scalability of Indexed Subgraph Query Processing Methods	1566
..... <i>Foteini Katsarou, Nikos Ntarmos, Peter Triantafillou</i>	
Lenses: An On-Demand Approach to ETL	1578
..... <i>Ying Yang, Niccolo Meneghetti, Ronny Fehling, Zhen Hua Liu, Oliver Kennedy</i>	
Keys for Graphs.....	1590
..... <i>Wenfei Fan, Zhe Fan, Chao Tian, Xin Luna Dong</i>	
Spatial Partitioning Techniques in Spatial Hadoop	1602
..... <i>Ahmed Eldawy, Louai Alarabi, Mohamed F. Mokbel</i>	
Extracting Logical Hierarchical Structure of HTML Documents Based on Headings	1606
..... <i>Tomohiro Manabe, Keishi Tajima</i>	
Permutation Search Methods are Efficient, Yet Faster Search is Possible.....	1618
..... <i>ilegsaikhon Naidan, Leonid Boytsov, Eric Nyberg</i>	

Industrial Papers

Distributed Architecture of Oracle Database In-memory	1630
... <i>Niloy Mukherjee, Shasank Chavan, Maria Colgan, Dinesh Das, Mike Gleeson, Sanket Hase, Allison Holloway, Hui Jin, Jesse Kamp, Kartik Kulkarni, Tirthankar Lahiri, Juan Loaiza, Neil Macnaughton, Vineet Marwah, Atrayee Mullick, Andy Witkowski, Jiaqi Yan, Mohamed Zait</i>	
Argonaut: Macrotask Crowdsourcing for Complex Data Processing	1642
..... <i>Daniel Haas, Jason Ansel, Lydia Gu, Adam Marcus</i>	

Building a Replicated Logging System with Apache Kafka	1654
<i>.....Guozhang Wang, Joel Koshy, Sriram Subramanian, Kartik Paramasivam, Mammad Zadeh, Neha Narkhede, Jun Rao, Jay Kreps, Joe Stein</i>	
Indexing and Selecting Hierarchical Business Logic	1656
<i>.....Alessandra Loro, Anja Gruenheid, Donald Kossmann, Damien Profeta, Philippe Beaudequin</i>	
Schema-Agnostic Indexing with Azure DocumentDB	1668
<i>Dharma Shukla, Shireesh Thota, Karthik Raman, Madhan Gajendran, Ankur Shah, Sergii Ziuzin, Krishnan Sundaram, Miguel Gonzalez Guajardo, Anna Wawrzyniak, Samer Boshra, Renato Ferreira, Mohamed Nassar, Michael Koltachev, Ji Huang, Sudipta Sengupta, Justin Levandoski, David Lomet</i>	
JetScope: Reliable and Interactive Analytics at Cloud Scale.....	1680
<i>Eric Boutin, Paul Brett, Xiaoyu Chen, Jaliya Ekanayake, Tao Guan, Anna Korsun, Zhicheng Yin, Nan Zhang, Jingren Zhou</i>	
Differential Privacy in Telco Big Data Platform	1692
<i>Xueyang Hu, Mingxuan Yuan, Jianguo Yao, Yu Deng, Lei Chen, Qiang Yang, Haibing Guan, Jia Zeng</i>	
Optimization of Common Table Expressions in MPP Database Systems.....	1704
<i>Amr El-Helw, Venkatesh Raghavan, Mohamed A. Soliman, George Caragea, Zhongxian Gu, Michalis Petropoulos</i>	
Towards Scalable Real-time Analytics: An Architecture for Scale-out of OLxP Workloads.....	1716
<i>.....Anil K Goel, Jeffrey Pound, Nathan Auch, Peter Bumbulis, Scott MacLean, Franz Farber, Francis Gropengiesser, Christian Mathis, Thomas Bodner, Wolfgang Lehner</i>	
FIT to Monitor Feed Quality.....	1728
<i>.....Tamraparni Dasu, Vladislav Shkapenyuk, Divesh Srivastava, Deborah F. Swayne</i>	
Real-Time Analytical Processing with SQL Server	1740
<i>Per-Åke Larson, Adrian Birka, Eric N. Hanson, Weiyun Huang, Michal Nowakiewicz, Vassilis Papadimos</i>	
Efficient Evaluation of Object-Centric Exploration Queries for Visualization.....	1752
<i>.....You Wu, Boulos Harb, Jun Yang, Gong Yu</i>	
Goblin: Unifying Data Ingestion for Hadoop.....	1764
<i>Lin Qiao, Yanan Li, Sahil Takiar, Ziyang Liu, Narasimha Veeramreddy, Min Tu, Ying Dai, Issac Buenrostro, Kapil Surlaker, Shirshanka Das, Chavdar Botev</i>	
Query Optimization in Oracle 12c Database In-Memory	1770
<i>Dinesh Das, Jiaqi Yan, Mohamed Zait, Satyanarayana R Valluri, Nirav Vyas, Ramarajan Krishnamachari, Prashant Gaharwar, Jesse Kamp, Niloy Mukherjee</i>	
Live Programming in the LogicBlox System: A MetaLogiQL Approach.....	1782
<i>.....Todd J. Green, Dan Olteanu, Geoffrey Washburn</i>	
The Dataflow Model: A Practical Approach to Balancing Correctness, Latency, and Cost in Massive-Scale, Unbounded, Out-of-Order Data Processing	1792
<i>.Tyler Akidau, Robert Bradshaw, Craig Chambers, Slava Chernyak, Rafael J. Fernandez-Moctezuma, Reuven Lax, Sam McVeety, Daniel Mills, Frances Perry, Eric Schmidt, Sam Whittle</i>	

One Trillion Edges: Graph Processing at Facebook-Scale	1804
..... <i>Avery Ching, Sergey Edunov, Maja Kabiljo, Dionysios Logothetis, Sambavi Muthukrishnan</i>	
Gorilla: A Fast, Scalable, In-Memory Time Series Database.....	1816
..... <i>Tuomas Pelkonen, Scott Franklin, Paul Cavallaro, Qi Huang, Justin Meza, Justin Teller, Kaushik Veeraraghavan</i>	
ConfSeer: Leveraging Customer Support Knowledge Bases for Automated Misconfiguration Detection	1828
..... <i>Rahul Potharaju, Joseph Chan, Luhui Hu, Cristina Nita-Rotaru, Mingshi Wang, Liyuan Zhang, Navendu Jain</i>	
Scaling Spark in the Real World: Performance and Usability.....	1840
..... <i>Michael Armbrust, Tathagata Das, Aaron Davidson, Ali Ghodsi, Andrew Or, Josh Rosen, Ion Stoica, Patrick Wendell, Reynold Xin, Matei Zaharia</i>	
Demonstration Papers	
StarDB: A Large-Scale DBMS for Strings.....	1844
..... <i>Majed Sahli, Essam Mansour, Panos Kalnis</i>	
Evaluating SPARQL Queries on Massive RDF Datasets	1848
..... <i>Razen Harbi, Ibrahim Abdelaziz, Panos Kalnis, Nikos Mamoulis</i>	
A Topic-based Reviewer Assignment System	1852
..... <i>Ngai Meng Kou, Leong Hou U, Nikos Mamoulis, Yuhong Li, Ye Li, Zhiguo Gong</i>	
FP-Hadoop: Efficient Execution of Parallel Jobs Over Skewed Data.....	1856
..... <i>Miguel Liroz-Gistau, Reza Akbarinia, Patrick Valduriez</i>	
Data Profiling with Metanome.....	1860
..... <i>Thorsten Papenbrock, Tanja Bergmann, Moritz Finke, Jakob Zwiener, Felix Naumann</i>	
Demonstration of Santoku: Optimizing Machine Learning over Normalized Data.....	1864
..... <i>Arun Kumar, Mona Jalal, Boqun Yan, Jeffrey Naughton, Jignesh M. Patel</i>	
PRISM: Concept-preserving Summarization of Top-K Social Image Search Results	1868
..... <i>Boon Siew Seah, Sourav S Bhowmick, Aixin Sun</i>	
Provenance for SQL through Abstract Interpretation: Value-less, but Worthwhile	1872
..... <i>Tobias Muller, Torsten Grust</i>	
SDB: A Secure Query Processing System with Data Interoperability	1876
..... <i>Zhian He, Wai Kit Wong, Ben Kao, David Wai Lok Cheung, Rongbin Li, Siu Ming Yiu, Eric Lo</i>	
SPARTex: A Vertex-Centric Framework for RDF Data Analytics.....	1880
..... <i>Ibrahim Abdelaziz, Razen Harbi, Semih Salihoglu, Panos Kalnis, Nikos Mamoulis</i>	
I2RS: A Distributed Geo-Textual Image Retrieval and Recommendation System.....	1884
..... <i>Lu Chen, Yunjun Gao, Zhihao Xing, Christian S. Jensen, Gang Chen</i>	
Reformulation-based query answering in RDF: alternatives and performance.....	1888
..... <i>Damian Bursztyn, Francois Goasdoue, Ioana Manolescu</i>	

SAASFEE: Scalable Scientific Workflow Execution Engine	1892
..... <i>Marc Bux, Jorgen Brandt, Carsten Lipka, Kamal Hakimzadeh, Jim Dowling, Ulf Leser</i>	
A Demonstration of HadoopViz: An Extensible MapReduce System for Visualizing Big Spatial Data ..	1896
..... <i>Ahmed Eldawy, Mohamed F. Mokbel, Christopher Jonathan</i>	
QOCO: A Query Oriented Data Cleaning System with Oracles	1900
..... <i>Moria Bergman, Tova Milo, Slava Novgorodov, Wang-Chiew Tan</i>	
TreeScope: Finding Structural Anomalies In Semi-Structured Data	1904
..... <i>Shanshan Ying, Flip Korn, Barna Saha, Divesh Srivastava</i>	
A Demonstration of the BigDAWG Polystore System	1908
<i>A. Elmore, J. Duggan, M. Stonebraker, M. Balazinska, U. Cetintemel, V. Gadepally, J. Heer, B. Howe, J. Kepner, T. Kraska, S. Madden, D. Maier, T. Mattson, S. Papadopoulos, J. Parkhurst, N. Tatbul, M. Vartak, S. Zdonik</i>	
RINSE: Interactive Data Series Exploration with ADS+	1912
..... <i>Kostas Zoumpatianos, Stratos Idreos, Themis Palpanas</i>	
Collaborative Data Analytics with DataHub	1916
<i>Anant Bhardwaj, Amol Deshpande, Aaron J. Elmore, David Karger, Sam Madden, Aditya Parameswaran, Harihar Subramanyam, Eugene Wu, Rebecca Zhang</i>	
Mindtagger: A Demonstration of Data Labeling in Knowledge Base Construction	1920
..... <i>Jaeho Shin, Christopher Ré, Michael Cafarella</i>	
Perseus: An Interactive Large-Scale Graph Mining and Visualization Tool.....	1924
..... <i>Danai Koutra, Di Jin, Yuanshi Ning, Christos Faloutsos</i>	
Smart Drill-Down: A New Data Exploration Operator	1928
..... <i>Manas Joglekar, Hector Garcia-Molina, Aditya Parameswaran</i>	
Virtual eXist-db: Liberating Hierarchical Queries from the Shackles of Access Path Dependence	1932
..... <i>Curtis E. Dyreson, Sourav S Bhowmick, Ryan Grapp</i>	
Annotating Database Schemas to Help Enterprise Search	1936
..... <i>Eli Cortez, Philip A. Bernstein, Yeye He, Lev Novik</i>	
VIIQ: Auto-Suggestion Enabled Visual Interface for Interactive Graph Query Formulation.....	1940
..... <i>Nandish Jayaram, Sidharth Goyal, Chengkai Li</i>	
FLORIN – A System to Support (Near) Real-Time Applications on User Generated Content on Daily News...	1944
..... <i>Qingyuan Liu, Eduard C. Dragut, Arjun Mukherjee, Weiyi Meng</i>	
VINERy: A Visual IDE for Information Extraction	1948
..... <i>Yun Yao Li, Elmer Kim, Marc A. Touchette, Ramiya Venkatachalam, Hao Wang</i>	
KATARA: Reliable Data Cleaning with Knowledge Bases and Crowdsourcing	1952
..... <i>Xu Chu, Mourad Ouzzani, John Morcos, Ihab F. Ilyas, Paolo Papotti, Nan Tang, Yin Ye</i>	

GIS Navigation Boosted by Column Stores.....	1956
..... <i>Foteini Alvanaki, Romulo Goncalves, Milena Ivanovaa, Martin Kersten, Kostis Kyzirakos</i>	
Gain Control over your Integration Evaluations	1960
..... <i>Patricia C. Arocena, Radu Ciucanu, Boris Glavic, Renée J. Miller</i>	
AIDE: An Automatic User Navigation System for Interactive Data Exploration	1964
..... <i>Yanlei Diao, Kyriaki Dimitriadou, Zhan Li, Wenzhao Liu, Olga Papaemmanouil, Kemi Peng, Liping Peng</i>	
A Demonstration of AQWA: Adaptive Query-Workload-Aware Partitioning of Big Spatial Data.....	1968
..... <i>Ahmed M. Aly, Ahmed S. Abdelhamid, Ahmed R. Mahmood, Walid G. Aref, Mohamed S. Hassan, Hazem Elmeleegy, Mourad Ouzzani</i>	
Janiform Intra-Document Analytics for Reproducible Research	1972
..... <i>Jens Dittrich, Patrick Bender</i>	
A Framework for Clustering Uncertain Data	1976
..... <i>Erich Schubert, Alexander Koos, Tobias Emrich, Andreas Zufle, Klaus Arthur Schmid, Arthur Zimek</i>	
EFQ: Why-Not Answer Polynomials in Action	1980
..... <i>Nicole Bidoit, Melanie Herschel, Katerina Tzompanaki</i>	
Error Diagnosis and Data Profiling with Data X-Ray	1984
..... <i>Xiaolan Wang, Mary Feng, Yue Wang, Xin Luna Dong, Alexandra Meliou</i>	
Sharing and Reproducing Database Applications	1988
..... <i>Quan Pham, Severin Thaler, Tanu Malik, Ian Foster, Boris Glavic</i>	
A Demonstration of TripleProv: Tracking and Querying Provenance over Web Data.....	1992
..... <i>Marcin Wylot, Philippe Cudre-Mauroux, Paul Groth</i>	
WADaR: Joint Wrapper and Data Repair.....	1996
..... <i>Stefano Ortona, Giorgio Orsi, Marcello Buoncristiano, Tim Furche</i>	
DATASPREAD: Unifying Databases and Spreadsheets.....	2000
..... <i>Mangesh Bendre, Bofan Sun, Ding Zhang, Xinyan Zhou, Kevin Chen-Chuan Chang, Aditya Parameswaran</i>	
Wisteria: Nurturing Scalable Data Cleaning Infrastructure.....	2004
..... <i>Daniel Haas, Sanjay Krishnan, Jiannan Wang, Michael J. Franklin, Eugene Wu</i>	
CODD: A Dataless Approach to Big Data Testing.....	2008
..... <i>Ashoke S., Jayant R. Haritsa</i>	
Query-Oriented Summarization of RDF Graphs	2012
..... <i>Sejla Cebiric, Francois Goasdoue, Ioana Manolescu</i>	
Universal-DB: Towards Representation Independent Graph Analytics	2016
..... <i>Yodsawalai Chodpathumwan, Amirhossein, Aleyasen, Arash Termehchy, Yizhou Sun</i>	
Tornado: A Distributed Spatio-Textual Stream Processing System	2020
..... <i>Ahmed R. Mahmood, Ahmed M. Aly, Thamir Qadah, El Kindi Rezig, Anas Daghistani, Amgad Madkour, Ahmed S. Abdelhamid, Mohamed S. Hassan, Walid G. Aref, Seleh Basalamah</i>	

Vizdom: Interactive Analytics through Pen and Touch	2024
..... <i>Andrew Crotty, Alex Galakatos, Emanuel Zgraggen, Carsten Binnig, Tim Kraska</i>	
S+EPPs: Construct and Explore Bisimulation Summaries, plus Optimize Navigational Queries; all on Existing SPARQL Systems.....	2028
..... <i>Mariano P. Consens, Valeria Fionda, Shahan Khatchadourian, Giuseppe Pirro</i>	
GraphGen: Exploring Interesting Graphs in Relational Data	2032
..... <i>Konstantinos Xirogiannopoulos, Udayan Khurana, Amol Deshpande</i>	
DBSeer: Pain-free Database Administration through Workload Intelligence	2036
..... <i>Dong Young Yoon, Barzan Mozafari, Douglas P. Brown</i>	

Tutorial Papers

Real Time Analytics: Algorithms and Systems	2040
..... <i>Arun Kejariwal, Sanjeev Kulkarni, Karthik Ramasamy</i>	
On Uncertain Graphs Modeling and Queries.....	2042
..... <i>Arijit Khan, Lei Chen</i>	
A Time Machine for Information: Looking Back to Look Forward	2044
..... <i>Xin Luna Dong, Wang-Chiew Tan</i>	
Structured Analytics in Social Media	2046
..... <i>Mahashweta Das, Gautam Das</i>	
Truth Discovery and Crowdsourcing Aggregation: A Unified Perspective	2048
..... <i>Jing Gao, Qi Li, Bo Zhao, Wei Fan, Jiawei Han</i>	
Tutorial: SQL-on-Hadoop Systems	2050
..... <i>Daniel Abadi, Shivnath Babu, Fatma Özcan, Ippokratis Pandis</i>	

Keynote Papers

Engineering Database Hardware and Software Together	2052
..... <i>Juan Loaiza</i>	
Big Data Research: Will Industry Solve all the Problems?.....	2053
..... <i>Magdalena Balazinska</i>	
Big Plateaus of Big Data on the Big Island.....	2057
..... <i>Todd Walter</i>	
Databases and Hardware: The Beginning and Sequel of a Beautiful Friendship	2058
..... <i>Anastasia Ailamaki</i>	

VLDB 2015 ORGANIZATION AND REVIEW BOARD

General Chairs

Michael J. Carey, University of California, Irvine

Program Chairs and Editors-in-Chief of PVLDB 8

Chen Li, University of California, Irvine

Volker Markl, TU Berlin

Research and Innovative Systems Tracks Associate Editors

Kevin Chang, U Illinois

Shivnath Babu, Duke University

Felix Naumann, Hasso Plattner Institute

Stefan Manegold, CWI Amsterdam

Yi Chen, NJIT

Fatma Ozcan, IBM Research Almaden

Jignesh Patel, University of Wisconsin, Madison

Magdalena Balazinska, University of Washington

Experiments and Analysis Track Associate Editors

Rainer Gemulla, MPI Saarbrücken, Germany

Industrial, Applications and Experience Track Associate Editors

Anhai Doan, University of Wisconsin, Madison

Prasan Roy, Sclera

Gregor Hackenbroich, SAP

Demonstration Chair

Alfons Kemper, TU München

Tutorial Chairs

Tova Milo, Tel Aviv University

Pierre Senellart, Telecom Paris Tech, France

Panel Chair

Joseph M. Hellerstein, University of California, Berkeley

PhD Workshop Chairs

Rachel Pottinger (UBC)

Proceedings Chairs

Daisy Zhe Wang, University of Florida

Tyson Condie, University of California, Los Angeles

Sponsorship Chairs

Michael Franklin, University of California, Berkeley

Edward Change, HTC

Patrick Valduriez, INRIA

Local Organization Chair

Lipyeow Lim, University of Hawaii

Conference and Registration Chairs

Ke Chen, Zhejiang University

Cuiping Li, Renmin University

Publicity and Web Management Chair

Rada Chirkova, North Carolina State

Treasury Chair

Malu Castellanos, HP Labs

VLDB Endowment Liaison

Paul Larson, Microsoft Research

PVLDB Managing Editor

Divesh Srivastava, AT&T Labs

PVLDB Information Director

Gerald Weber, University of Auckland

PVLDB Advisory Committee

H. V. Jagadish, Renée J. Miller, M. Tamer Özsu, Kian-Lee Tan, Michael Böhlen, Susan Davidson, S. Sudarshan, Gerhard Weikum

Research Track Review Board

Daniel Abadi, Yale University
Alberto Abello, UPC Barcelona
Ashraf Aboulnaga, Qatar Computing Research Institute
Foto Afrati, NTU Athens
Sihem Amer Yahia, CNRS LIG
Aijun An, York University
Arvind Arasu, Microsoft Research
Walid Aref, Purdue University
Paolo Atzeni, Roma Tre University
Denilson Barbosa, University of Alberta
Srikanta Bedathur, IBM Research
Philip Bernstein, Microsoft Research
Michael Böhlen, University of Zürich
Peter Boncz, CWI Amsterdam
Angela Bonifati, Lille 1 U and INRIA
Philippe Bonnet, IT U of Copenhagen
Nico Bruno, Google
Alex Buchmann, TU Darmstadt
Mike Cafarella, University of Michigan
K. Selcuk Candan, Arizona State University
Malu Castellanos, HP Labs
Kaushik Chakrabarti, Microsoft Research
Lei Chen, Hong Kong U of Science and Technology
Fei Chiang, McMaster University
Byron Choi, Hong Kong Baptist University
Philippe Cudre Mauroux, Fribourg University
Mahashweta Das, HP Labs
Sudipto Das, Microsoft Research
Amol Desphande, University of Maryland
Stefan Dessloch, TU Kaiserslautern
Jens Dittrich, Saarland University
Alin Dobra, University of Florida
Xin Luna Dong, Google
Jennie Duggan, MIT
Wenfei Fan, University of Edinburgh
Alan Fekete, University of Sydney Australia
Peter Fischer, Universität Freiburg

Avrilia Floratou, IBM
Avigdor Gal, Technion
Minos Garofalakis, Technical U of Crete
Wolfgang Gatterbauer, Carnegie Mellon U
Tingjian Ge, University of Massachusetts Lowell
Floris Geerts, University of Antwerp
Lukasz Golab, University of Waterloo
Torsten Grust, Universität Tübingen
Jarek Gryz, York University
Dimitrios Gunopulos, University of Athens
Hakan Hacigumus, NEC Labs
Wook Shin Han, POSTECH
Seif Haridi, KTH Stockholm
Oktie Hassanzadeh, IBM Research
Bingsheng He, Nanyang Technological University
Jeffrey Heer, University of Washington
Herodotos Herodotou, Microsoft Research
Katja Hose, Aalborg University
Vagelis Hristidis, University of California, Riverside
Jeong-Hyon Hwang, State University of New York at Albany
Stratos Idreos, Harvard University
Yannis Ioannidis, University Of Athens
Zachary Ives, University of Pennsylvania
Christopher Jermaine, Rice University
Ruoming Jin, Kent State University
Alekh Jindal, Massachusetts Institute of Technology
Ryan Johnson, University of Toronto
Eser Kandogan, IBM Research Almaden
Gjergji Kasneci, Hasso Plattner Institute
Asterios Katsifodimos, TU Berlin
Yannis Katsis, University of California, San Diego
Daniel Keim, Universität Konstanz
Bettina Kemme, McGill University
Eamonn Keogh, University of California, Riverside
Martin Kersten, CWI Amsterdam
Daniel Kifer, Penn State University
Hideaki Kimura, Hewlett Packard
George Kollios, Boston University
Donald Kossman, ETH Zurich

Nick Koudas, University of Toronto
 Georgia Koutrika, HP Labs
 Tim Kraska, Brown University
 Harumi Kuno, HP Labs
 Laks Lakshmanan, University of British Columbia
 Paul Larson, Microsoft Research
 Hongrae Lee, Google
 Wolfgang Lehner, TU Dresden
 Alberto Lerner, New York University
 Ulf Leser, Humboldt Universität zu Berlin
 Justin Levandoski, Microsoft Research
 Feifei Li, University of Utah
 Guoliang Li, Tsinghua University
 Jianzhong Li, Harbin Institute of Technology
 Yunyao Li, IBM Research Almaden
 Erietta Liarou, EPF Lausanne
 Xuemin Lin, University of New South Wales
 Ziyang Liu, NEC Labs America
 Eric Lo, Polytechnic University of Hong Kong
 Guy Lohman, IBM Research Almaden
 Jiaheng Lu, Renmin University of China
 Qiong Luo, Hong Kong University of Science and Technology
 Jayant Madhavan, Google
 Ioana Manolescu, INRIA
 Patrick Marcel, University of Tours
 Marta Mattoso, Federal University of Rio de Janeiro
 Alexandra Meliou, University of Massachusetts Amherst
 Sergey Melnik, Google
 Weiyi Meng, Binghamton University
 Sebastian Michel, Saarland University
 Iris Miliaraki, Yahoo Labs Barcelona
 Renee Miller, University of Toronto
 Zhou Minqi, East China Normal University
 Prasenjit Mitra, Penn State University
 Bernhard Mitschang, Universität Stuttgart
 Mohamed Mokbel, Northeastern University
 Barzan Mozafari, University of Michigan
 Hannes Mühleisen, CWI Amsterdam

Arnab Nandi, Ohio State University
 Vivek Narasayya, Microsoft Research
 Jeffrey Naughton, University of Wisconsin Madison
 Rimma Nehme, Microsoft
 Thomas Neumann, TU Munich
 Raymond Ng, University of British Columbia
 Christopher Olston, Google
 Dan Olteanu, Oxford University & LogicBlox
 Beng Chin Ooi, National University of Singapore
 M. Tamer Özsu, University of Waterloo
 Themis Palpanas, Paris Descartes University
 Ippokratis Pandis, IBM Research Almaden
 Dimitris Papadias, Hong Kong U of Science and Technology
 Paolo Papotti, Qatar Computing Research Institute
 Andy Pavlo, Carnegie Mellon University
 Torben Bach Pedersen, Aalborg University
 Jian Pei, Simon Fraser University
 Peter Pietzuch, Imperial College London
 Evaggelia Pitoura, University of Ioannina
 Alkis Polyzotis, Google and University of California, Santa Cruz
 Fabio Porto, LNCC Brazil
 Li Qian, Facebook
 Jorge Arnulfo Quiane Ruiz, Qatar Computing Research Institute
 Tilmann Rabl, University of Toronto
 Erhard Rahm, Universität Leipzig
 Krithi Ramamritham, IIT Bombay
 Ravi Ramamurthy, Microsoft Research
 Vijayshankar Raman, IBM Research Almaden
 Mirek Riedewald, Northeastern University
 Tore Risch, Uppsala Universitet
 Kenneth Ross, Columbia University
 Elke Rundensteiner, Worcester Polytechnic Institute US
 Barna Saha, AT&T Labs Research
 Kenneth Salem, University of Waterloo
 Simonas Saltenis, Aalborg University
 Kai Uwe Sattler, TU Ilmenau

Eric Sedlar, Oracle Labs
Bernhard Seeger, Philipps Universität Marburg
Kyuseok Shim, Seoul National University
Jerome Simeon, IBM Research Watson
Alkis Simitsis, HP Labs
Michael Stonebraker, Massachusetts Institute of Technology
Julia Stoyanovich, Drexel University
Nan Tang, Qatar Computing Research Institute
Yufei Tao, The Chinese University of Hong Kong
Sandeep Tata, Google
Nesime Tatbul, Intel Labs and MIT
Arash Termehchy, Oregon State University
Evimaria Terzi, Boston University
Jens Teubner, TU Dortmund
Martin Theobald, University of Antwerp
Andreas Thor, University of Leipzig
Yuanyuan Tian, IBM Research Almaden
Anthony Tung, National University of Singapore
Vasilis Vassalos, Athens University
Yannis Vassiliou, NTU Athens
Yannis Velegrakis, University of Trento

Rares Vernica, HP Labs
Gottfried Vossen, WWU Münster
Florian Waas, Datometry Inc.
Daisy Zhe Wang, University of Florida
Haixun Wang, Google
Wei Wang, University of New South Wales
Gerhard Weikum, Max Planck Institut für Informatik
Till Westmann, Oracle Labs
Steven Whang, Google
Kyu Young Whang, KAIST
Raymond Chi Wing Wong, Hong Kong University of Science and Technology
Xiaokui Xiao, Nanyang Technological University
Xifeng Yan, University of California, Santa Barbara
Jun Yang, Duke University
Xiaochun Yang, Northeastern University China
Cong Yu, Google Research
Ge Yu, Northeastern University China
Jeffrey Xu Yu, Chinese University of Hong Kong
Xiaofang Zhou, The University of Queensland
Esteban Zimanyi, UL Brussels
Marcin Zukowski, Snowflake Computing

LETTER FROM THE PROGRAM COMMITTEE CHAIRS

As a premier annual international forum for database community, PVLDB 2015 received a large number of submissions to various tracks. The beautiful location in Hawaii makes the conference this year even more attractive to many authors. The research track received 710 high-quality submissions, with 139 papers accepted so far, and 17 papers still being revised, some of which will appear at VLDB 2016. The industrial track received 68 submissions, of which 21 were accepted. The demonstration track received 148 submissions, of which 49 were accepted. We are glad to see that our community is continuously developing new technologies in the space of data management to make an impact on businesses, the sciences, and society as a whole. We see a variety of topics ranging from traditional indexing, query processing and optimization, search, graph data management, to crowdsourcing and big data management. In addition to those papers above, we also have 21 roll-over papers from VLDB 2014, 6 tutorials, 4 keynote speeches, and 2 panels as well as a variety of exciting workshops. We are expecting an exciting technical program with many interesting presentations at VLDB 2015, including poster sessions, which will feature all research papers presented at the conference as well as papers from the VLDB Journal.

It has been a busy year since we officially became the program chairs in March 2014. When saying “Yes” to Mike Carey’s invitation to play this role, we knew it’s going to be quite a journey. The experience tells us that the job requires a lot of commitment, attention, organization, patience, and sometimes creativity. Each month we needed to manage several sessions in parallel. We had to do good bookkeeping to track that all reviews were submitted on time, ensure that a thorough discussion happened for each paper, and arrive at a timely conclusion based on high quality reviews. At the same time, it has been a very rewarding journey, working with an excellent team of co-chairs, associate editors, and reviewers, who have been instrumental in creating the VLDB 2015 research program. The co-chairs of the industrial, demo, panel, workshop, PhD workshop, and tutorial track have been running their respective tracks very smoothly. The nine associate editors did a fantastic job handling their piles of research papers each month and have driven each paper to a timely conclusion. The reviewers deserve a lot of credit for carrying out most of the evaluation work, not to mention their tolerance of our frequent “ping messages.” And of course we thank all authors for their high-quality contributions, which make VLDB the premier forum for scientific exchange in data management. The proceedings chairs, Tyson Condie and Daisy Zhe Wang, handled the proceedings process efficiently. The CMT team at Microsoft deserves our gratitude for support and modifications to the tool, patiently and quickly handling our frequent requests. We also want to take this chance to thank Divesh Srivastava and H.V. Jagadish for providing great advice whenever we had questions.

We look forward to seeing you in Hawaii! Aloha!

Chen Li and Volker Markl
PVLDB Volume 8 Editors in Chief
VLDB Program Committee Chairs