Graham Cormode Ke Yi
Antonios Deligiannakis Minos Garofalakis(Eds.)

First International Workshop on Big Dynamic Distributed Data (BD3)

Workshop at VLDB 2013
Riva Del Garda, Italy, August 30, 2013

Proceedings

©2013 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. Re-publication of material from this volume requires permission by the copyright owners.

Editors' contacts:
G.Cormode@warwick.ac.uk, yike@cse.ust.hk, adeli@softnet.tuc.gr, minos@softnet.tuc.gr

Preface

As the amount of streaming data produced by large-scale systems such as environmental monitoring, scientific experiments and communication networks grows rapidly, new approaches are needed to effectively process and analyze such data. There are several promising directions in the area of large-scale distributed computation, that is, where multiple computing entities work together over partitions of the massive, streaming data to perform complex computations. Two important paradigms in this realm are continuous distributed monitoring (i.e., continually maintaining an accurate estimate of a complex query), and distributed and cluster-based systems that allow the processing of big, streaming data (e.g., IBM System S, Apache S4, and Twitter Storm).

The aim of the BD3 workshop is to bring together computer scientists with interests in this field to present recent innovations, find topics of common interest and to stimulate further development of new approaches to deal with massive dynamic and distributed data.

August 2013

Graham Cormode, Antonios Deligiannakis,
Minos Garofalakis, Ke Yi

Organizing Committee

General Chairs:
Minos Garofalakis
Technical University of Crete
minos@softnet.tuc.gr

Antonios Deligiannakis
Technical University of Crete
adeli@softnet.tuc.gr

Program Chairs:
Graham Cormode
University of Warwick
G.Cormode@warwick.ac.uk

Ke Yi
Hong Kong University of Science and Technology
yike@cse.ust.hk

Publicity Chair:
Odysseas Papapetrou
Technical University of Crete
papapetrou@softnet.tuc.gr

Program Committee

Alin Dobra	U. Florida
Pascal Felber	Universite de Neuchatel
Christof Fetzer	TU Dresden
Ling Huang	Intel Research
Daniel Keren	Haifa
Andrew McGregor	UMass-Amherst
Stavros Papadopoulos	HKUST
Odysseas Papapetrou	Technical University of Crete
Jeff Phillips	Utah
Peter Pietzuch	Imperial College London
Neoklis Polyzotis	UC Santa Cruz
Assaf Schuster	Technion
Izchak Sharfman	Technion
Nesime Tatbul	Intel Labs / MIT
Srikanta Tirthapura	Iowa State
Suresh Venkatasubramanian	Utah
Milan Vojnovic	Microsoft Research
Qin Zhang	IBM Research

Contents

Safe-Zones for Monitoring Distributed Streams
Daniel Keren, Guy Sagy, Amir Abboud, David Ben-David, Izchak Sharfman, and Assaf Schuster **7**

Communication-Efficient Distributed Online Prediction using Dynamic Model Synchronizations
Mario Boley, Michael Kamp, Daniel Keren, Assaf Schuster and Izchak Sharfman **13**

Communication-efficient Outlier Detection for Scale-out Systems
Moshe Gabel, Daniel Keren and Assaf Schuster **19**

Elastic Complex Event Processing under Varying Query Load
Thomas Heinze, Yuanzhen Ji, Yinying Pan, Franz Josef Grueneberger, Zbigniew Jerzak, and Christof Fetzer **25**

Adaptive Selective Replication for Complex Event Processing Systems
Franz Josef Grünberger, Thomas Heinze and Pascal Felber **31**

Dynamic Partitioning of Big Hierarchical Graphs
Vasilis Spyropoulos and Yannis Kotidis **37**

Scalable and Robust Management of Dynamic Graph Data
Alan G. Labouseur, Paul W. Olsen Jr. and Jeong-Hyon Hwang **43**

Towards Elastic Stream Processing: Patterns and Infrastructure
Kai-Uwe Sattler and Felix Beier **49**

Task Graphs of Stream Mining Algorithms
Sayaka Akioka **55**

Large-scale Online Mobility Monitoring with Exponential Histograms
Christine Kopp, Michael Mock, Odysseas Papapetrou and Michael May **61**

Multi-Stage Malicious Click Detection on Large Scale Web Advertising Data
Leyi Song, Xueqing Gong, Xiaofeng He, Rong Zhang and Aoying Zhou **67**