Editorial Board

David Hutchison  
*Lancaster University, Lancaster, UK*

Takeo Kanade  
*Carnegie Mellon University, Pittsburgh, PA, USA*

Josef Kittler  
*University of Surrey, Guildford, UK*

Jon M. Kleinberg  
*Cornell University, Ithaca, NY, USA*

Friedemann Mattern  
*ETH Zurich, Zürich, Switzerland*

John C. Mitchell  
*Stanford University, Stanford, CA, USA*

Moni Naor  
*Weizmann Institute of Science, Rehovot, Israel*

C. Pandu Rangan  
*Indian Institute of Technology, Madras, India*

Bernhard Steffen  
*TU Dortmund University, Dortmund, Germany*

Demetri Terzopoulos  
*University of California, Los Angeles, CA, USA*

Doug Tygar  
*University of California, Berkeley, CA, USA*

Gerhard Weikum  
*Max Planck Institute for Informatics, Saarbrücken, Germany*
More information about this series at http://www.springer.com/series/7409
Preface

This volume contains a selection of the papers presented at the 19th East-European Conference on Advances in Databases and Information Systems (ADBIS 2015), held during September 8–11, 2015, at Futuroscope, Poitiers, France.

The ADBIS series of conferences aims at providing a forum for the presentation and dissemination of research on database theory, development of advanced DBMS technologies, and their advanced applications. ADBIS 2015 in Poitiers continued the series after St. Petersburg (1997), Poznań (1998), Maribor (1999), Prague (2000), Vilnius (2001), Bratislava (2002), Dresden (2003), Budapest (2004), Tallinn (2005), Thessaloniki (2006), Varna (20007), Pori (2008), Riga (2009), Novi Sad (2010), Vienna (2011), Poznań (2012), Genoa (2013), and Ohrid (2014). This edition was special, as it was the first time that ADBIS took place in France. The conferences are initiated and supervised by an International Steering Committee consisting of representatives from Armenia, Austria, Bulgaria, Czech Republic, Estonia, Finland, Germany, Greece, Hungary, Israel, Italy, Latvia, Lithuania, Poland, Russia, Serbia, Slovakia, Slovenia, and the Ukraine.

The program of ADBIS 2015 included keynotes, research papers, two tutorials, and thematic workshops. The conference attracted 135 paper submissions from 39 countries with 330 authors. After rigorous reviewing by the Program Committee (77 reviewers from 22 countries), the 31 papers included in this LNCS proceedings volume were accepted as full contributions, making an acceptance rate of 23%.

Furthermore, the Program Committee selected 18 more papers as short contributions and 30 papers from seven workshops that are published in a companion volume entitled New Trends in Databases and Information Systems in the Springer series Communications in Computer and Information Science. All papers were evaluated by at least three reviewers and most of them by four to five reviewers. The selected papers span a wide spectrum of topics in databases and related technologies, tackling challenging problems and presenting inventive and efficient solutions. In this volume, these papers are organized according to the 15 sessions: (1) Database Theory and Access Methods, (2) User Requirements and Database Evolution, (3) Multidimensional Modeling and OLAP, (4) ETL, (5) Transformation, Extraction and Archiving, (6) Modeling and Ontologies, (7) Time Series Processing, (8) Performance and Tuning, (9) Advanced Query Processing, (10) Approximation and Skyline, (11) Confidentiality and Trust.

For this edition of ADBIS 2015, we had two keynote talks: the first one from Serge Abiteboul from Inria and ENS Cachan, France, on “The Story of Webdatalog” and the second one by Jens Dittrich, from Saarland University, Germany, on “The Case for Small Data Management.” In addition, we had two tutorials: the first by Nicolas Anciaux, Benjamin Nguyen, and Iulian Sandu Popa from Inria Paris-Rocquencourt and INSA Centre-Val de Loire, France, on “Towards an Era of Trust in Personal Data Management” and the second one by Boris Novikov, from St. Petersburg University, Russia, on “Query Processing: Beyond SQL and Relations.”
ADBIS 2015 strived to create conditions for more experienced researchers to share their knowledge and expertises with the young researchers. In addition, the following seven workshops associated with the ADBIS were co-allocated with the main conference:

- Second International Workshop on Big Data Applications and Principles (BigDap 2015), organized by Elena Baralis (Politecnico di Torino, Italy), Tania Cerquitelli (Politecnico di Torino, Italy) and Pietro Michiardi (EURECOM, France).
- Workshop on Data Centered Smart Applications (DCSA 2015), organized by Ajantha Dahanayake (Prince Sultan University, Saudi Arabia) and Bernhard Thalheim (Christian Albrechts University, Germany).
- 4th International Workshop on GPUs in Databases (GID 2015), organized by Witold Andrzejewski (Poznan University of Technology, Poland), Krzysztof Kaczmarski (Warsaw University of Technology, Poland), and Tobias Lauer (Offenburg University of Applied Sciences, Germany).
- Workshop on Managing Evolving Business Intelligence Systems (MEBIS 2015), organized by Selma Khouri (National Engineering School for Mechanics and Aerotechnics (ISAE-ENSMA), France and National High School of Computer Science (ESI, Algeria), and Robert Wrembel (Poznan University of Technology, Poland).
- 4th International Workshop on Ontologies Meet Advanced Information Systems (OAIS 2015), organized by Ladjel Bellatreche (LIAS/ISAE-ENSMA, France), and Yamine Ait Ameur (IRIT-ENSEIHT, France).
- First International Workshop on Semantic Web for Cultural Heritage (SW4CH 2015), organized by Béatrice Bouchou Markhoff (LI, University François Rabelais de Tours, France) and Stéphane Jean (LIAS/ISAE-ENSMA and University of Poitiers, France).
- Workshop on Information Systems for AlaRm Diffusion (WISARD 2015), organized by Rémi Delmas (ONERA, Toulouse, France), Thomas Polacsek (ONERA, Toulouse, France), Florence Sèdes (IRIT, Toulouse, France).

Each workshop has its own international Program Committee. The accepted papers were published by Springer in the series Communications in Computer and Information Science (CCIS).

The best papers of the main conference and workshop were invited to be submitted to special issues of the following journals: Information Systems - Elsevier, Information Systems Frontiers - Springer, and International Journal on Semantic Web and Information Systems - IGI.

We would like to express our gratitude to every individual who contributed to the success of ADBIS 2015. First, we thank all authors for submitting their research papers to the conference. We are also indebted to the members of the community who offered their precious time and expertise in performing various roles ranging from organizational to reviewing - their efforts, energy, and degree of professionalisms deserve the highest commendations. Special thanks to the Program Committee members and the external reviewers for evaluating papers submitted to ADBIS 2015, thereby ensuring the quality of the scientific program. Thanks also to all the colleagues, secretaries, and engineers involved in the conference organization, as well as the workshop organizers. We would like to thank Dr. Mickaël Baron, from LIAS/ISAE-ENSMA, for his endless...
help and support. Special thanks are due to the members of the Steering Committee, in particular, its chair Leonid Kalinichenko and his vice-chair Yannis Manolopoulos for all their help and guidance.

Finally, we thank Springer for publishing the proceedings containing invited and research papers in the LNCS series. The Program Committee work relied on EasyChair, and we thank its development team for creating and maintaining it; it offered a great support throughout the different phases of the reviewing process. The conference would not have been possible without our supporters and sponsors:

- Région Poitou Charentes
- ISAE-ENSMA
- Poitiers University
- INFORSID Association
- CRITT Informatique, Futuroscope
- LIAS laboratory

Last, but not least, we thank the participants of ADBIS 2015 for sharing their works and presenting their achievements, thus providing a lively, fruitful, and constructive forum, and giving us the pleasure of knowing that our work was purposeful.

September 2015

Ladjel Bellatreche
Tadeusz Morzy
Patrick Valduriez
Organization

General Chair
Ladjel Bellatreche LIAS/ISAE-ENSMA, Poitiers, France

Program Committee Co-chairs
Patrick Valduriez Inria of Montpellier, France
Tadeusz Morzy Poznan University, Poland

Workshop Co-chairs
Athena Vakali Aristotle University of Thessaloniki, Greece
Bernhard Thalheim Kiel University, Germany

Doctoral Consortium Co-chairs
Sofian Maabout Labri/Bordeaux, France
Boris Novikov St. Petersburg University, Russia

Publicity Chair
Selma Khouri LIAS/ISAE-ENSMA, France

Website Chair
Mickaël Baron LIAS/ISAE-ENSMA, Poitiers, France

Proceedings Technical Editor
Stéphane Jean LIAS/ISAE-ENSMA, Poitiers, France

Local Organizing Committee Chair
Patrick Girard LIAS/ISAE-ENSMA, France

Local Organizing Committee
Mickaël Baron LIAS/ISAE-ENSMA, Poitiers, France
Frédéric Carreau LIAS/ISAE-ENSMA, Poitiers, France
Brice Chardin LIAS/ISAE-ENSMA, Poitiers, France
Zoé Faget LIAS/ISAE-ENSMA, Poitiers, France
Patrick Girard LIAS/ISAE-ENSMA, Poitiers, France
Laurent Guittet LIAS/ISAE-ENSMA, Poitiers, France
Stéphane Jean LIAS/ISAE-ENSMA, Poitiers, France
Yassine Ouhammou LIAS/ISAE-ENSMA, Poitiers, France
Claudine Rault LIAS/ISAE-ENSMA, Poitiers, France
Okba Barkat LIAS/ISAE-ENSMA, Poitiers, France
Selma Bouarar LIAS/ISAE-ENSMA, Poitiers, France
Ahcène Boukorca LIAS/ISAE-ENSMA, Poitiers, France
Lahcène Brahimi LIAS/ISAE-ENSMA, Poitiers, France
Zouhir Djilani LIAS/ISAE-ENSMA, Poitiers, France
Géraud Fokou LIAS/ISAE-ENSMA, Poitiers, France
Nadir Guetmi LIAS/ISAE-ENSMA, Poitiers, France
Yves Mouafo LIAS/ISAE-ENSMA, Poitiers, France
Guillaume Phavorin LIAS/ISAE-ENSMA, Poitiers, France

Supporters

Région Poitou Charentes
ISAE-ENSMA
Poitiers University
INFORSID Association
CRITT Informatique, Futuroscope
LIAS laboratory

Steering Committee

Paolo Atzeni Italy
Andras Benczur Hungary
Albertas Caplinskas Lithuania
Barbara Catania Italy
Johann Eder Austria
Theo Haerder Germany
Marite Kirikova Latvia
Hele-Mai Haav Estonia
Mirjana Ivanovic Serbia
Hannu Jaakkola Finland
Mikhail Kogalovsky Russia
Yannis Manolopoulos Greece
Rainer Manthey Germany
Manuk Manukyan Armenia
Joris Mihaeli Israel
Tadeusz Morzy Poland
Pavol Navrat Slovakia
Boris Novikov Russia
Program Committee

Reza Akbarinia
Inria, France
Paolo Atzeni
Università Roma Tre, Italy
Andreas Behrend
University of Bonn, Germany
Ladjel Bellatreche
ISAE-ENMSA, France
Omar Boucelma
Aix-Marseille University, France
Mahdi Bohlouli
University of Siegen, Germany
Albertas Caplinskas
Institute of Mathematics and Informatics, Italy
Barbara Catania
DISI-University of Genoa, Italy
Wojciech Cellary
Poznan School of Economy, Poland
Ricardo Rodrigues Ciferri
Federal University of São Carlos, Brazil
Alfredo Cuzzocrea
University of Trieste, Italy
Todd Eavis
Concordia University, Canada
Johann Eder
Alpen-Adria-Universität Klagenfurt, Austria
Markus Endres
University of Augsburg, Germany
Pedro Furtado
University of Coimbra/CISUC, Portugal
Johann Gamper
Free University of Bozen-Bolzano, Italy
Jérôme Gensel
Grenoble University, France
Shahram Ghandeharizadeh
University of Southern California, USA
Matteo Golfarelli
DISI - University of Bologna, Italy
Goetz Graefe
Hewlett-Packard Laboratories, USA
Dawid Gross-amblard
IRISA, Rennes University, France
Jarek Gryz
York University, Canada
Mohand-Said Hacid
University of Claude Bernard Lyon 1 - UCBL, France
Theo Härder
TU Kaiserslautern, Germany
Mirjana Ivanovic
University of Novi Sad, Serbia
Hannu Jaakkola
Tampere University of Technology, Finland
Leonid Kalinichenko
Russian Academy of Science, Russia
Ahto Kalja
Küperneetika Instituut, Estonia
Kalinka Kaloyanova
University of Sofia - FMI, Bulgaria
Mehmed Kantardzic
University of Louisville, USA
Marite Kirikova
Riga Technical University, Latvia
Mikhail Kogalovsky
Market Economy Institute of the Russian Academy of Sciences, Russia
Christian Koncilia
Alpen-Adria University of Klagenfurt, Austria
Margita Kon-popovska  
Ss Cyril and Methodius University, Macedonia

Harald Kosch  
University of Passau, Germany

Georgia Koutrika  
HP Labs, USA

Regine Laleau  
Paris Est Creteil University, France

Wolfgang Lehner  
TU Dresden, Germany

Pericles Loucopoulos  
University of Manchester, UK

Ivan Lukovic  
University of Novi Sad, Serbia

Yannis Manolopoulos  
Aristotle University of Thessaloniki, Greece

Rainer Manthey  
University of Bonn, Germany

Pascal Molli  
Nantes University, France

Tadeusz Morzy  
Poznan University of Technology, Poland

Pavol Navrat  
Slovak University of Technology, Slovakia

Kjetil Nørvåg  
Norwegian University of Science and Technology, Norway

Gultekin Ozsoyoglu  
Case Western Reserve University, USA

M. Tamer Ozsu  
University of Waterloo, Canada

Oscar Pastor  
Valencia University of Technology, Spain

Dana Petcu  
Institute e-Austria Timisoara, Romania

Jean-Marc Petit  
Université de Lyon, INSA Lyon, France

Olivier Pivert  
IRISA, Rennes University, France

Neoklis Polyzotis  
University of California Santa Cruz, USA

Boris Rachev  
Technical University of Varna, Bulgaria

Peter Revesz  
University of Nebraska, USA

Stefano Rizzi  
DEIS - University of Bologna, Italy

Viera Rozinajova  
Slovak University of Technology in Bratislava, Slovakia

Henryk Rybinski  
Warsaw University of Technology, Poland

Gunter Saake  
University of Magdeburg, Germany

Klaus-Dieter Schewe  
Software Competence Center Hagenberg, Germany

Timos Sellis  
RMIT University, Australia

Bela Stantic  
Griffith University, Australia

Manolis Terrovitis  
Institute for the Management of Information Systems, RC Athena, Greece

Martin Theobald  
University of Antwerp, Belgium

Farouk Toumani  
LIMOS, Blaise Pascal University, Clermont-Ferrand, France

Patrick Valduriez  
Inria, France

Panos Vassiliadis  
University of Ioannina, Greece

Jari Veijalainen  
University of Jyvaskyla, Finland

Goran Velinov  
UKIM, Skopje, Macedonia

Krishnamurthy Vidyasankar  
Memorial University, Canada

Stratis Viglas  
University of Edinburgh, UK

Peter Vojtas  
Charles University of Prague, Czech Republic

Gerhard Weikum  
Max Planck Institute for Informatics, Germany

Tatjana Welzer  
University of Maribor, Slovenia
Robert Wrembel  Poznan University of Technology, Institute of Computing Science, Poland

Vladimir Zadorozhny  University of Pittsburgh, USA

Additional Reviewers

Fabian Benduhn  Magdeburg University, Germany
Jevgeni Marenkov  Tallinn University of Technology, Estonia
Sonja Ristic  University of Novi Sad, Serbia
Giorgos Giannopoulos  National Technical University of Athens, Greece
Karoly Bosa  Johannes Kepler University Linz, Austria
Grégory Smits  IRISA, France
Fatma Slaimi  LSIS, Marseille, France
Olga Gkountouna  National Technical University of Athens (NTUA), Athens, Greece

John Liagouris  University of Hong Kong SAR China
Panagiotis Symeonidis  Aristotle University, Thessaloniki, Greece
Konstantinos Theoharidis  IMIS, Research Center Athena, Greece
Mustafa Al-Hajjaji  University of Magdeburg, Germany
Sebastian Dorok  University of Magdeburg, Germany
Loredana Tec  AIT Austrian Institute of Technology GmbH, Vienna, Austria

Anton Dignos  University of Zürich, Switzerland
Felix Kossak  Software Competence Center Hagenberg GmbH, Hagenberg, Austria

Amel Mammar  Telecom/Telecom SudParis, France
Sahar Vahdati  University of Bonn, Germany
Nabil Hameurlain  University of Pau, France
Tarmo Robal  Tallinn University of Technology, Estonia
Hala Skaf-Molli  LINA, Nantes University, France
Zoltan Miklos  Inria, Rennes, France
Farida Semmak  Université Paris-Est, France
Christophe Gnaho  Université Paris Est, France
Lorena Paoletti  Universidad de Santiago de Chile, Chile
Gilles Nachouki  LINA, Nantes University, France
Irina Astрова  Tallinn University of Technology, Estonia
Shuaiqiang Wang  University of Jyvaskyla, Finland
Zoé Faget  LIAS/ISAE-ENSMA, France
Vladimir Ivančević  University of Novi Sad, Serbia
Saulius Gudas  Vilnius University, Lithuania
Dirk Habich  Technische Universität Dresden, Germany
Slavica Kordić  University of Novi Sad, Serbia
Eike Schallehn  Otto von Guericke University of Magdeburg, Germany
Vladimir Dimitrieski  University of Novi Sad, Serbia
Christian Končilia  Alpen-Adria-Universität Klagenfurt, Austria
Ioannis N. Athanasiadis  Hellenic Open University, Kozani, Greece
Keynotes
The Story of Webdamlog

Serge Abiteboul
INRIA Saclay and ENS Cachan

Abstract. We summarize in this paper works about the management of data in a distributed manner based on Webdamlog, a datalog-extension. We point to relevant articles on these works. More references may be found there.

1 The Webdamlog Approach

Information of interest may be found on the Web in a variety of forms, in many systems, and with different access protocols. Today, the control and management of the diversity of data and tasks in this setting are beyond the skills of casual users [1]. Facing similar issues, companies see the cost of managing and integrating information skyrocketing. We are concerned with the management of Web data in place in a distributed manner, with a possibly large number of autonomous, heterogeneous systems collaborating to support certain tasks. We summarize in this paper works in this setting around Webdamlog and point to the relevant articles on it.

The thesis is that managing the richness and diversity of data residing on the Web can be tamed using a holistic approach based on a distributed knowledge base. Our approach is to represent all Web information as logical facts, and Web data management tasks as logical rules. A variety of complex data management tasks that currently require intense work and deep expertise may then greatly benefit from the automatic reasoning provided by inference engines, operating over the distributed Web knowledge base: for instance, information access, access control, knowledge acquisition and dissemination.

We propose to express the peers logic in Webdamlog, a datalog-style rule-based language. In Webdamlog, peers exchange facts (for information) and rules (in place of code). The use of declarative rules provides the following advantages. Peers may perform automatic reasoning using the available knowledge. Because the model is formally defined, it becomes possible to prove (or disprove) desirable properties. Because the model is based on a datalog-style language, query processing can benefit from optimization techniques. Because the model represents provenance and time, the quality of data can be better controlled. Because the model is general, a wide variety of scenarios and protocols may be captured, which is a requirement for today's Web.

This work was realized in the context of the European Research Council grant Webdam [6, 13]. The system is available in opensource at [8]. The work on Webdamlog was inspired by previous works on ActiveXML [3] at INRIA, as well as Bud [7, 12] at Berkeley University. The system has been demonstrated in [2]. An extensive
experimental evaluation of the implementation (showing notably that the computational cost of access control is modest) is presented in [11].

In the remaining of this paper, we briefly mention three main contributions: (i) The Webdamlog language that facilitates the exchange of data and rules between distributed peers; (ii) A collaborative access control mechanism for Webdamlog that enables controlling the dissemination of data in a network; and (iii) A probabilistic semantics for datalog with functional dependencies that can serve as the basis for managing uncertain, noisy, possibly contradicting data.

2 Three Main Contributions

Webdamlog. There is a new trend to use datalog-style rule-based languages to specify modern distributed applications, notably on the Web [9, 10]. The Webdamlog language was first formally described in [4]. It is a version of distributed datalog that allows specifying distributed applications where peers exchange messages (i.e. logical facts) as well as rules (i.e., the analog of code). An example of rule is as follows:

```
[at alice] album@alice($photoId,$photo,$f) :- friend@alice($f),
                   album@$f($photoId,$photo,$source), tags@f($photoId,"Alice")
```

Ignore the details of the syntax. With this rule, Alice deploys, at each peer corresponding to one of her friends, a rule that sends her all photos this friend owns that is tagged by her name. The main originality of the language is the use of delegation that allows delegating rules to other peers. Distributed computing is realized by delegating some rules to perform some tasks to other peers. Knowledge acquisition, i.e., the Webdamlog analog to “downloading apps”, is also performed using rule delegations. The main contribution of [4] is the presentation of the language. A study of the impact on expressiveness of “delegations” is also provided.

Access control. Users wish to share data using these systems, but avoiding the risks of unintended disclosures or unauthorized access by applications has become a major challenge. An important issue for users in a distributed setting is thus the control of the access to their data by others. In [11], we introduce a collaborative access control mechanism for Webdamlog. Using this model, users can specify access control policies providing flexible tuple-level control derived using provenance information.

Inconsistency and imprecision. In [5], we study deduction in the presence of inconsistencies and probabilities for datalog programs. (The results can be extended to Webdamlog in a straightforward manner). Inconsistencies are captured through violations of functional dependencies (FDs). We propose nondeterministic semantics for datalog with FDs. We introduce a PTIME (in the size of the extensional data) algorithm, that given a datalog program, a set of FDs and an input instance, produces a c-table representation of the set of possible resulting worlds.
We then propose to quantify nondeterminism with probabilities, by means of a probabilistic semantics. We consider the problem of capturing possible worlds along with their probabilities via probabilistic c-tables. We then study classical computational problems in this novel context. We consider the problems of computing the probabilities of answers, of identifying most likely supports for answers, and of determining the extensional facts that are most influential for deriving a particular fact. We show that the interplay of recursion and FDs leads to novel technical challenges in the context of these problems.

Acknowledgements. We thank all the researchers who participated in the Webdamlog project and in particular, Meghyn Bienvenu, Pierre Bourhis, Daniel Deutch, Alban Galland, Gerome Miklau, Vera Zaychik Moffitt, Marie-Christine Rousset, Julia Stoyanovich, Jules Testard, and Victor Vianu.

References

The Case for Small Data Management

Jens Dittrich
Saarland University
http://infosys.uni-saarland.de

Abstract. Exabytes of data; several hundred thousand TPC-C transactions per second on a single computing core; scale-up to hundreds of cores and a dozen Terabytes of main memory; scale-out to thousands of nodes with close to Petabyte-sized main memories; and massively parallel query processing are a reality in data management. But, hold on a second: for how many users exactly? How many users do you know that really have to handle these kinds of massive datasets and extreme query workloads? On the other hand: how many users do you know that are fighting to handle relatively small datasets, say in the range of a few thousand to a few million rows per table? How come some of the most popular open source DBMS have hopelessly outdated optimizers producing inefficient query plans? How come people don’t care and love it anyway? Could it be that most of the worlds data management problems are actually quite small? How can we increase the impact of database research in areas when datasets are small? What are the typical problems? What does this mean for database research? We discuss research challenges, directions, and a concrete technical solution coined PDbF: Portable Database Files (open source at https://github.com/uds-datalab/PDBF). See also our VLDB 2015 demo (https://infosys.uni-saarland.de/publications/p2199-dittrich.pdf).

CV. Jens Dittrich is a Full Professor of Computer Science in the area of Databases, Data Management, and Big Data at Saarland University, Germany. Previous affiliations include U Marburg, SAP AG, and ETH Zurich. He is also associated to CISPA (Center for IT-Security, Privacy and Accountability). He received an Outrageous Ideas and Vision Paper Award at CIDR 2011, a BMBF VIP Grant, a best paper award at VLDB 2014, two CS teaching awards in 2011 and 2013, as well as several presentation awards including a qualification for the interdisciplinary German science slam finals in 2012 and three presentation awards at CIDR (2011, 2013, and 2015). His research focuses on fast access to big data including in particular: data analytics on large datasets, Hadoop MapReduce, main-memory databases, and database indexing. He has been a PC member and/or area chair of prestigious international database conferences such as PVLDB, SIGMOD, and ICDE. Since 2013 he has been teaching his classes on data management as flipped classrooms. See http://datenbankenlernen.de or http://youtube.com/jensdit for a list of freely available videos on database technology in German and English (about 80 videos in German and 80 in English so far).
Tutorials
Towards an Era of Trust in Personal Data Management

Nicolas Anciaux\textsuperscript{1}, Benjamin Nguyen\textsuperscript{2}, and Iulian Sandu Popa\textsuperscript{1}

\textsuperscript{1} INRIA Paris-Rocquencourt, Domaine du Voluceau, 78153 Le Chesnay, France

\textsuperscript{2} INSA Centre-Val de Loire, 88 boulevard Lahitolle, 18022 BOURGES, France
Benjamin.Nguyen@insa-cvl.fr

Managing personal data with strong privacy guarantees has become an important topic in an age where your glasses record and share everything you see, your wallet records and shares your financial transactions, and your set-top box records and shares your energy consumption, while several recent affairs have unveiled the severe consequences of the loss of privacy. In this context, more and more alternatives are proposed based on user centric and decentralized solutions, capitalizing on the use of trusted personal devices controlling the data at the edges of the Internet. Decentralized solutions are promising because they do not exhibit the intrinsic limitations of classical centralized solutions, e.g., sudden changes in privacy policies of companies holding the data, data exposures by negligence or because it is regulated by too weak policies, exposure to sophisticated attacks whose benefit/cost ratio is high for centralized databases. Hence, such solutions appear as a sea change for personal data management, where the control over personal data is pushed to the edges of the Internet, within sensors acquiring the data and in a variety of user devices endowed with a form of trust, e.g., tamper-resistant secure hardware-based devices.

This tutorial reviews several existing solutions going in this direction, presents a functional architecture encompassing these alternatives, and exposes the underlying techniques and open issues dealing with user centric and decentralized data management platforms. In a first part, we review the recent initiatives pursuing the objective of reestablishing user control over their data by decentralizing this control in personal secure or trusted devices. We discuss an abstract distributed architecture focusing on secure storing, managing and sharing of personal data, i.e., the asymmetric architecture, and indicate the main challenges inherent to decentralized data management. In a second part, we explore data management techniques exercised within a trusted device at the client side. We review the main attempts proposed in the literature and concentrate on those addressing the specific context of microcontrollers equipping sensors and mobile phones (SIM cards). In a third part, we investigate the problem of performing global processing without any compromise on data privacy. We present the difficulties to overcome to execute privacy preserving computations on populations of personal devices, and illustrate it by focusing on Group By SQL queries and Privacy Preserving Data Publishing. In a fourth part, we conclude the tutorial by presenting existing and future instances of decentralized privacy preserving data management architectures. We mainly focus on attempts and proposals targeting social-medical, smart houses, and rural areas contexts.
Query processing and optimization are essential for any data processing system since introduction of high-level declarative query languages in early 80-ies. During the last decade several new techniques were introduced in order to address requirements of new classes of applications, data models, storage and indexing, and querying paradigms.

Modern query processing and optimization extends far beyond relational queries. Several techniques were revised and a number of new techniques have been introduced to make the query processing efficient. Several systems that were originally designed as low-level storage facilities implementing persistence layer, were augmented with high level declarative features. The declarative scripting languages provide a technique for easy-to-understand specification of complex analytical scenarios that look like sequential but are executed on massively parallel systems.

The main focus of this tutorial is on the query optimization and processing in new environments and for new classes of applications.

Although many of declarative languages are designed as extensions to SQL, the internals of the implementations usually have significant differences with well-known optimization and processing techniques developed for relational systems using row-based storage structures.

Column stores are considered to be the most efficient for analytical processing on modern hardware. The physical algebraic operations for column stores differ from those used in row-based ones, and optimization strategies and heuristics are different.

Distributed data processing systems such as Hadoop weren’t originally intended for declarative query processing. However, several query languages are implemented on top, bringing back the need for optimization. Examples of these languages and systems include ASTERIX, SCOPE, and Apache Hive.

Processing of semi-structured and unstructured data ultimately requires fuzzy (e.g. similarity) queries resulting in several obstacles for relational optimizers that are mostly oriented on re-ordering of join operations. Although some of recently introduced techniques, such as efficient top-down enumeration algorithms might be helpful, many issues are still open.

Parametric and dynamic optimization techniques seem to be especially useful for distributed heterogeneous environments where availability of data statistics is often severely limited and cost estimations are unreliable.

Finally, holistic optimization is an emerging technology that optimizes the database queries and application together with the goal to improve the overall application performance.
## Contents

### Database Theory and Access Methods

Conditional Differential Dependencies (CDDs) .................................................. 3  
Selasi Kwashie, Jixue Liu, Jiuyong Li, and Feiyue Ye

Improving the Pruning Ability of Dynamic Metric Access Methods with Local Additional Pivots and Anticipation of Information. .................. 18  
Paulo H. Oliveira, Caetano Traina Jr., and Daniel S. Kaster

The Structure of Preference Orders ................................................................. 32  
Markus Endres

### User Requirements and Database Evolution

Two Phase User Driven Schema Matching .................................................... 49  
Nick Bozovic and Vasilis Vassalos

CoDEL – A Relationally Complete Language for Database Evolution .......... 63  
Kai Herrmann, Hannes Voigt, Andreas Behrend, and Wolfgang Lehner

### Multidimensional Modeling and OLAP

Implementation of Multidimensional Databases in Column-Oriented NoSQL Systems ................................................................. 79  
Max Chevalier, Mohammed El Malki, Arlind Kopliku, Olivier Teste, and Ronan Tournier

A Framework for Building OLAP Cubes on Graphs .................................. 92  
Amine Ghrab, Oscar Romero, Sabri Skhiri, Alejandro Vaisman, and Esteban Zimányi

A Generic Data Warehouse Architecture for Analyzing Workflow Logs. ... 106  
Christian Koncilia, Horst Pichler, and Robert Wrembel

### ETL

HBelt: Integrating an Incremental ETL Pipeline with a Big Data Store for Real-Time Analytics .................................................. 123  
Weiping Qu, Sahana Shankar, Sandy Ganza, and Stefan Dessloch
Two-ETL Phases for Data Warehouse Creation: Design and Implementation ................................................................. 138
   Ahlem Nabli, Senda Bouaziz, Rania Yangui, and Faiez Gargouri

Direct Transformation Techniques for Compressed Data: General Approach and Application Scenarios ........................................... 151
   Patrick Damme, Dirk Habich, and Wolfgang Lehner

Transformation, Extraction and Archiving

Analysis of the Blocking Behaviour of Schema Transformations in Relational Database Systems .......................................................... 169
   Lesley Wevers, Matthijs Hofstra, Menno Tammens, Marieke Huisman, and Maurice van Keulen

A Benchmark for Relation Extraction Kernels ............................................. 184
   João L.M. Pereira, Helena Galhardas, and Bruno Martins

Web Content Management Systems Archivability .................................. 198
   Vangelis Banos and Yannis Manolopoulos

Modeling and Ontologies

Evidence-Based Languages for Conceptual Data Modelling Profiles ........... 215
   Pablo Rubén Fillottrani and C. Maria Keet

Ontological Commitments, DL-Lite Logics and Reasoning Tractability .......... 230
   Mauricio Minuto Espil, Maria Gabriela Ojea, and Maria Alejandra Ojea

SeeCOnt: A New Seeding-Based Clustering Approach for Ontology Matching ................................................................. 245
   Alsayed Algergawy, Samira Babalou, Mohammad J. Kargar, and S. Hashem Davarpanah

Time Series Processing

ForCE: Is Estimation of Data Completeness Through Time Series Forecasts Feasible? ................................................................. 261
   Gregor Endler, Philipp Baumgärtel, Andreas M. Wahl, and Richard Lenz

Best-Match Time Series Subsequence Search on the Intel Many Integrated Core Architecture ................................................. 275
   Mikhail Zymbler
Feedback Based Continuous Skyline Queries
Over a Distributed Framework ........................................ 287
  Ahmed Khan Leghari, Jianneng Cao, and Yongluan Zhou

Performance and Tuning
Partitioning Templates for RDF ....................................... 305
  Rebeca Schroeder and Carmem S. Hara
Efficient Computation of Parsimonious Temporal Aggregation ............ 320
  Giovanni Mahlknecht, Anton Dignös, and Johann Gamper
TDQMed: Managing Collections of Complex Test Data .................... 334
  Johannes Held and Richard Lenz

Advanced Query Processing
A Self-tuning Framework for Cloud Storage Clusters ...................... 351
  Siba Mohammad, Eike Schallehn, and Gunter Saake
Optimizing Sort in Hadoop Using Replacement Selection .................. 365
  Pedro Martins Dusso, Caetano Sauer, and Theo Härder
Distributed Sequence Pattern Detection Over Multiple Data Streams ...... 380
  Ahmed Khan Leghari, Jianneng Cao, and Yongluan Zhou

Approximation and Skyline
Space-Bounded Query Approximation .................................. 397
  Boris Cule, Floris Geerts, and Reuben Ndindi
Hybrid Web Service Discovery Based on Fuzzy Condorcet Aggregation .... 415
  Hadjila Fethallah, Belabed Amine, and Halfaoui Amel

Confidentiality and Trust
Confidentiality Preserving Evaluation of Open Relational Queries .......... 431
  Joachim Biskup, Martin Bring, and Michael Bulinski
A General Trust Management Framework for Provider Selection in Cloud Environment ........................................ 446
  Fatima Zohra Filali and Belabas Yagoubi
Sybil Tolerance and Probabilistic Databases to Compute Web Services Trust ........................................ 458
  Zohra Saoud, Noura Faci, Zakaria Maamar, and Djamal Benslimane
Erratum to: ForCE: Is Estimation of Data Completeness Through Time Series Forecasts Feasible? .......................... E1

Gregor Endler, Philipp Baumgärtel, Andreas M. Wahl,
and Richard Lenz

Author Index ............................................ 473