

Report on PODS 2015*

Diego Calvanese
PODS 2015 PC Chair
Free University of Bozen-Bolzano, Italy
`calvanese@inf.unibz.it`

October 31, 2015

The 34th edition of the ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (PODS) took place from May 31 to June 4, 2015, in Melbourne, Victoria, Australia. The symposium was organized jointly with the ACM SIGMOD International Conference on Management of Data (SIGMOD). PODS focuses on theoretical aspects of data management systems and techniques, and the co-location with SIGMOD stimulates interaction between theory-oriented and system-oriented research.

The proceedings of PODS 2015 are published by ACM Press, and can also be found on the SIGMOD website (<http://www.sigmod.org/>), as well as the ACM Digital Library (<http://www.acm.org/dl/>). The conference program included a keynote talk by Michael I. Jordan, two invited tutorials (by Todd J. Green and Graham Cormode), and 25 contributed papers selected by the Program Committee from 80 submissions. Similarly to past PODS conferences, most of the contributed papers are preliminary reports on work in progress, and many are likely to appear in more polished form in scientific journals.

In his keynote talk, Michael Jordan argued that the classical approaches in computer science on the one hand, and statistical science on the other, are no longer adequate for addressing the challenges of Big Data. Therefore, he advocates a novel foundational perspective on data analysis that blends research in computational thinking, focusing on scalable algorithms and abstractions, with inferential thinking, focusing on underlying populations, sampling patterns, and predictions. This collaboration between researchers working in databases and those working in statistical machine learning will benefit both disciplines.

The tutorial by Todd J. Green gave an overview of *LogiQL*, a declarative language that relies on Datalog to provide solid foundations for enterprise level data management and analytics. The language is practical, and allows building complex applications involving a mixture of tasks including analytics, transactions, machine learning, and optimization problems. The tutorial also presented techniques for efficient evaluation of

*Editor of the Database Theory Column: Victor Vianu, UC San Diego, vianu@cs.ucsd.edu

LogiQL programs, emphasising the theoretical foundations. The accompanying paper *LogiQL: a Declarative Language for Enterprise Applications* by Todd J. Green provides details on these techniques, including purely functional data structures used at all levels of the system, the leapfrog trijoin algorithm and its extension supporting incremental maintenance, and transaction repair, a novel concurrency control mechanism supporting concurrent write transactions.

The tutorial by Graham Cormode addressed a central challenge in any system dealing with Big Data, namely how to compute and represent complex features of the data in a way that can be processed by many single machines in parallel. It introduced the concept of *compact summaries*, which are data structures that capture key features of the data, and that can be created efficiently over distributed data sets. It discussed prominent examples of summaries for various operations, including variants of counts (approximate, distinct), set operations, median and other statistics, matrix and linear algebra operations. It also discussed problems for which no compact summary can exist. The accompanying short paper *Compact Summaries over Large Datasets* by Graham Cormode discusses the need for compact summaries in the context of Big Data, outlines the structure of the tutorial, and provides a list of useful pointers to the literature.

Among the contributed papers, the Program Committee selected the paper *Parallel-Correctness and Transferability for Conjunctive Queries* by Tom J. Ameloot, Gaetano Geck, Bas Ketsman, Frank Neven, and Thomas Schwentick for the PODS 2015 Best Paper Award. The paper provides a general framework for reasoning about one-round evaluation algorithms under arbitrary distribution policies in the Massively Parallel Communication model of computation. It is a generalization of earlier work in that it considers the problem for families of distribution policies and not for specific ones. In this framework, two main technical problems are investigated, namely parallel-correctness and transferability, and the complexities of the decision versions of these problems are established.

Similarly to preceding years, the contributed papers covered a wide range of subjects, including data structures and algorithms, model theory and logics, foundations for Big Data, query languages, incompleteness and uncertainty, streaming data, distributed and parallel data management, tree and graph-structured data, data integration and exchange, and metadata management.

The PODS 2015 conference also presented the ACM PODS Alberto O. Mendelzon Test-of-Time Award. The Award is given each year to one or a small number of papers published in the PODS proceedings ten years prior that had the most impact over the intervening decade. The 2015 Award Committee, consisting of Dan Suciu (chair), Foto Afrati, and Frank Neven, selected two papers for this award: *XPath Satisfiability in the Presence of DTDs*, by Michael Benedikt, Wenfei Fan, and Floris Geerts; and *Views and Queries: Determinacy and Rewriting*, by Luc Segoufin and Victor Vianu.

Overall, the conference was extremely interesting and stimulating. The co-occurrence with the SIGMOD conference remained beneficial for both conferences. Despite the distance from Europe and North-America, and an earlier start than in previous years, the 2015 SIGMOD/PODS conference enjoyed very good attendance.

As in past editions, the significant number of young authors and attendees bodes well for theoretical research in databases.

While the number of submissions remained lower than in most previous years, for the first time since 2010 there has been again an increase in submissions with respect to the preceding year (from 68 for PODS 2014 to 80 for PODS 2015). This may be due at least in part to two novelties that the PODS Executive Committee introduced for PODS 2015: *(i)* PODS provided two submission cycles, where the first cycle allowed for papers to be revised and resubmitted; *(ii)* PODS broadened its scope, inviting explicitly for submission not only papers featuring a deep theoretical exploration of topical areas central to data management, as done traditionally, but also papers presenting new formal frameworks that aim to provide the basis for deeper theoretical investigation of important emerging issues in data management, and papers providing a validation of theoretical approaches from the lens of practical applicability to data management. To assess the effectiveness of these measures in increasing the number of submissions without compromising their quality, these changes will be maintained at least for two additional editions of PODS.

PODS 2016 will take place in San Francisco, California, USA, from June 26 to July 1, 2016. Tova Milo will be the PODS General Chair and Wang-Chiew Tan the PC Chair. More Information can be found at <http://www.sigmod2016.org/>.