

# Report on the 8th International Workshop on Business Intelligence for the Real-Time Enterprise (BIRTE'14)

Malu Castellanos  
HP Labs, USA  
malu.castellanos@hp.com

Umeshwar Dayal  
Hitachi Labs, USA  
umeshwar.dayal@hal.hitachi.com

Nesime Tatbul  
Intel Labs and MIT, USA  
tatbul@csail.mit.edu

Damianos Chatziantoniou \*  
AUEB, Greece  
damianos@aueb.gr

Qiming Chen \*  
HP Labs, USA  
qiming.chen@hp.com

## 1. OVERVIEW

The 8th International Workshop on Business Intelligence for the Real-Time Enterprise (BIRTE) was held on September 1, 2014 in conjunction with the VLDB 2014 Conference in Hangzhou, China. Like in previous years, the workshop was well attended by an engaged audience from both academia and industry, breaking an attendance record in the history of the BIRTE workshop series with more than 80 participants during the keynote session. In addition to the keynote speech, the workshop included two invited industrial talks, and four presentations of peer-reviewed papers covering a wide range of real-time BI topics with an overarching emphasis on big data analytics. The workshop developed as follows: After the official opening of the workshop, two papers were presented - one on BI analytics on graph-structured data and another on contextual analysis in temporal databases. The invited industrial talks session that followed included a talk from Microsoft Research about their data processing solutions for complex big data analytics and a talk from HP Labs about a new fault tolerance technique that they developed for distributed stream processing. The afternoon sessions opened with the highlight of the workshop: Dr. C. Mohan's keynote speech providing a critical survey of the current big data landscape. The workshop closed with a session that consisted of the two remaining papers - one on exploratory OLAP over linked data and another on data stream partitioning.

## 2. KEYNOTE SPEECH

This year's keynote speaker was Dr. C. Mohan from the IBM Almaden Research Center. Dr. Mohan has been a well-known pioneer in database sys-

\*This author served as a session chair at the workshop.

tems and has made numerous contributions to relational database research and technology in various different roles at IBM for more than 30 years. In his talk "Big Data: Hype and Reality", Mohan delighted the audience with a concrete and detailed picture of the current landscape of big data systems. According to Mohan, as users and developers gain a deeper understanding of the needs of real use cases (including real-time BI applications), the initial hype around big data systems (including NoSQL, NewSQL, and others) has been fading away. It is now becoming clearer that most of the so-called distinctive features of big data systems have in fact been well-known principles of relational database systems for decades. Mohan's comprehensive and critical survey of this popular field attracted much attention from a big audience and was very well received.

## 3. INDUSTRIAL TALKS

Inviting industrial speakers to present their perspective on real-world BI problems, solutions, and applications has been a tradition of BIRTE since its inception in 2006. This year's workshop featured two industrial talks. First, in his talk entitled "Building Analytics Engines for the Big Data Age", Dr. Badrish Chandramouli of Microsoft Research presented the challenges of a temporal streaming engine called Trill. Trill has been architected as a library to support embedded execution within cloud applications and distributed fabrics. Second, Dr. Qiming Chen of HP Labs gave a talk about "Optimistic Failure Recovery in Distributed Stream Processing". More specifically, he presented the back-track-based and the window-oriented recovery mechanisms implemented as part of the Fontainebleau distributed stream analytics system built on top of

the Storm platform. Both of these talks, covering industry-scale stream processing engines, intrigued an audience of around 50 participants.

#### 4. PAPER PRESENTATIONS

The BIRTE'14 program committee selected a total of 4 paper submissions presented at the workshop in two research sessions, one being the opening and the other being the closing session of the workshop. The first of these sessions consisted of a position paper by co-authors from TU Dresden in Germany and the SAP Labs in USA, and a research paper by co-authors from University at Buffalo, SUNY and the Oracle Corporation in USA. First, Michael Rudolf presented a flexible approach for multi-dimensional graph data analysis in their paper entitled "SynopSys: Foundations for Multidimensional Graph Analytics". The key feature that distinguishes SynopSys from existing technologies, which require upfront modeling of analytical scenarios and are difficult to adapt to changes, is the ability to express ad-hoc analytical queries over graph data. The second paper, entitled "Detecting the Temporal Context of Queries" and presented by Ying Yang, focuses on the concept of contextual dependency - a term used by the authors to explain and attribute mistaken assumptions made by end users of BI applications. A formal definition for contextual dependence is given, followed by several strategies to efficiently detect and quantify the effects of contextual dependence on query outputs.

The second paper session consisted of an application paper jointly written by co-authors from Aalborg University in Denmark and Universite Libre de Bruxelles in Belgium, and a research paper from University of Southern Denmark. First, Dilshod Ibragimov explained during his talk, entitled "Towards Exploratory OLAP over Linked Open Data - A Case Study", how to integrate real-time data from web sources described in RDF into the analysis process in BI environments. To achieve this, a system that uses a multi-dimensional schema of the OLAP cube expressed in RDF vocabularies is proposed. Finally, the last presentation of the workshop was on "Efficient Pattern Detection over a Distributed Framework" by Ahmed Khan Leghari. In this talk, Leghari described an event stream partitioning scheme that partitions streams over time windows without considering any key attributes.

Though not as popular as the invited speaker sessions, there were 30+ attendees during both of these paper sessions.

#### 5. CONCLUSIONS

Overall, we are proud to report that the 8th edition of the VLDB-colocated BIRTE workshop series has been a great success. We have once again witnessed that real-time business intelligence continues being a topic of great relevance for both database researchers and practitioners. Talks included a diverse set of real-world BI use cases, and indicated strong collaborations between academic and industrial community in this field. This year, we have seen that big data analytics has been a common theme for all BIRTE presentations, with a striking emphasis on the analysis of streaming and graph-structured data. BIRTE'14 presentation slides and a preliminary version of the corresponding papers can be found on our workshop webpage at <http://db.csail.mit.edu/birte2014/>. The final version of the papers (including papers from our invited speakers) have also been published as part of a joint BIRTE 2013-2014 post-proceedings in Springer LNBIP Volume 206.

#### 6. ACKNOWLEDGEMENTS

We would like to thank all the authors of submitted papers for their interest in the workshop and their high-quality submissions, as well as the distinguished PC members for their dedicated work in reviewing those submissions. This workshop would not have been as successful as it was without the keynote given by Dr. C. Mohan, as well as the invited talks given by Dr. Badrish Chandramouli and Dr. Qiming Chen, to whom we are deeply grateful. We would also like to thank the VLDB 2014 organizers for supporting BIRTE, especially the general chairs (Chun Chen and Sharad Mehrotra), the workshop chairs (Anastasia Ailamaki and Kaushik Chakrabarti), and the local organization chair (Lidan Shou). Finally, we would like to acknowledge our proceedings chair Jennie Duggan, who did an excellent job in putting together the workshop post-proceedings.