CS Seminar

Explaining Outputs in Modern Data Analytics

John Liagouris
ETH Zurich

Date:
November 2, 2016
Wednesday
2:30 pm

Venue:
Room 308
Chow Yei Ching Building
The University of Hong Kong

Abstract:
In this talk I will present our recent work on the design and implementation of a general framework for interactively explaining the outputs of modern data-parallel computations, including iterative data analytics. To produce explanations, existing works adopt a naive backward tracing approach which runs into known issues; naive backward tracing may identify: (i) too much information that is difficult to process, and (ii) not enough information to reproduce the output, which hinders the logical debugging of the program. The contribution of our work is twofold. First, we provide methods to effectively reduce the size of explanations based on the first occurrence of a record in an iterative computation. Second, we provide a general method for identifying explanations that are sufficient to reproduce the target output in arbitrary computations – a problem for which no viable solution existed until now. We implemented our approach on differential dataflow, a modern high-throughput, low-latency dataflow platform. We added a small (but extensible) set of rules to explain each of its data-parallel operators, and we implemented these rules as differential dataflow operators themselves. This choice allows our implementation to inherit the performance characteristics of differential dataflow, and results in a system that efficiently computes and updates explanatory inputs even as the inputs of the reference computation change. We evaluated our system with various analytic tasks on real datasets, and we show that it produces concise explanations in tens of milliseconds, while remaining faster – up to two orders of magnitude – than even the best implementations that do not support explanations.

About the Speaker:
John Liagouris is a post-doctoral researcher at ETH Zurich, and a member of the Systems Group. Before joining ETHZ, he was a visiting research fellow at the University of Hong Kong (2013-2014), and a research assistant at the Institute for the Management of Information Systems (IMIS) of the Research and Innovation Center 'Athena', Greece (2009-2015). Dr. Liagouris obtained a diploma in Electrical and Computer Engineering in 2008, and a PhD in 2015, both from NTU Athens, Greece. His research interests lie in the areas of datacenter monitoring, modelling, and simulation, real-time data analytics, distributed system profiling, and software defined networks.

All are welcome!
For enquiries, please call 2859 2180 or email
enquiry@cs.hku.hk
Department of Computer Science
The University of Hong Kong