

Data Series Management: The Road to Big Sequence Analytics

Themis Palpanas

Paris Descartes University (France)

Friday, January 22, 2016

11:00 am - 12:30 pm

Room: APB 3105, Nöthnitzer Straße 46

Abstract: There is an increasingly pressing need, by several applications in diverse domains, for developing techniques able to index and mine very large collections of sequences, or data series. Examples of such applications come from social media analytics and internet service providers, as well as from a multitude of scientific domains. It is not unusual for these applications to involve numbers of data series in the order of hundreds of millions to billions, which are often not analyzed in their full detail due to their sheer size.

In this talk, we describe recent efforts in designing techniques for indexing and mining truly massive collections of data series that will enable scientists to easily analyze their data. We show that the main bottleneck in mining such massive datasets is the time taken to build the index, and we thus introduce solutions to this problem. Furthermore, we discuss novel techniques that adaptively create data series indexes, allowing users to correctly answer queries before the indexing task is finished. We also show how our methods allow mining on datasets that would otherwise be completely untenable, including the first published experiments using one billion data series.

Finally, we present our vision for the future in big sequence management research, including the promising directions in terms of distributed processing.

Bio: Themis Palpanas is a professor of computer science at the Paris Descartes University (France), here he is a director of the Data Intensive and Knowledge Oriented Systems (diNo) group. He received the BS degree from the National Technical University of Athens, Greece, and the MSc and PhD degrees from the University of Toronto, Canada. He has previously held positions at the University of Trento and the IBM T.J. Watson Research Center. He has also worked for the University of California, Riverside, and visited Microsoft Research and the IBM Almaden Research Center. His research solutions have been implemented in world-leading commercial data management products and he is the author of nine US patents. He is the recipient of three Best Paper awards (including ICDE and PERCOM), and the IBM Shared University Research (SUR) Award in 2012, which represents a recognition of research excellence at worldwide level. He has been a member of the IBM Academy of Technology Study on Event Processing, and is a founding member of the Event Processing Technical Society. He has served as General Chair for VLDB 2013, the top international conference on databases.

