Title: *Data Series Management: Fulfilling the Need for Big Sequence Analytics*

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 times not analyzed in their full detail due to their sheer size. However, no existing data management solution (such as relational databases, column stores, array databases, and time series management systems) can offer native support for sequences and the corresponding operators necessary for complex analytics.

In this talk, we argue for the need to study the theory and foundations for sequence management of big data sequences, and to build corresponding systems that will enable scalable management and analysis of very large sequence collections. 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 discuss novel techniques that adaptively create data series indexes, allowing users to correctly answer queries before the indexing task is finished. Finally, we present our vision for the future in big sequence management research, including the promising directions in terms of storage, distributed processing, and query benchmarks.

Short Bio:

Themis Palpanas is Senior Member of the Institut Universitaire de France (IUF), a distinction that recognizes excellence across all academic disciplines, and professor of computer science at the Paris Descartes University (France), where he is director of diNo, the data management 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 at IBM T.J. Watson Research Center, and visited Microsoft Research, and the IBM Almaden Research Center.

His interests include problems related to data science (big data analytics and machine learning applications). He is the author of nine US patents, three of which have been implemented in world-leading commercial data management products. He is the recipient of three Best Paper awards, and the IBM Shared University Research (SUR) Award.

He is currently serving on the VLDB Endowment Board of Trustees, as an Editor in Chief for the BDR Journal, Associate Editor for VLDB 2019, Associate Editor in the TKDE, and IDA journals, as well as on the Editorial Advisory Board of the IS journal, and the Editorial Board of the TLDKS Journal. He has served as General Chair for VLDB 2013, Associate Editor for VLDB 2017, and Workshop Chair for EDBT 2016, ADBIS 2013, and ADBIS 2014, General Chair for the PDA@IOT International Workshop (in conjunction with VLDB 2014), and General Chair for the Event Processing Symposium 2009. April 13th, 2018.