User-Defined Functions in Relational Databases: Challenges and Promising Solutions based on YeSQL

Yannis Ioannidis[1]

[1]*University of Athens, Greece*

Abstract
On 3rd July 2023, Yannis Ioannidis delivered a keynote talk at the 31st Symposium on Advanced Database Systems in Galzignano Terme (Padua, Italy). The following is the abstract of his talk and a short biography.

Keywords
User-Defined Functions, YeSQL, SQL, SEBD 2023 Keynote

Abstract of the Keynote

The diversity and complexity of modern data management applications have led to the extension of the relational paradigm with syntactic and semantic support for User-Defined Functions (UDFs). Although well-established in traditional DBMS settings, UDFs have become central in many application contexts, such as data science, data analytics, and edge computing. Still, a critical limitation of UDFs is the impedance mismatch between their evaluation and relational processing.

In this talk, I first overviewed the area and the technical challenges that UDF processing brings. I then presented YeSQL, an SQL extension with rich UDF support, and a pluggable architecture to easily integrate it with server-based or embedded database engines. YeSQL supports Python UDFs fully integrated with relational queries such as scalar, aggregator, or table functions. Key novel characteristics of YeSQL include easy implementation of complex algorithms and several performance enhancements, including tracing JIT compilation of Python UDFs, parallelism and fusion of UDFs, stateful UDFs, and seamless integration with a database engine. Experimental analysis showcases the usability and expressiveness of YeSQL. It demonstrates that its techniques of minimizing context switching between the relational engine and the Python VM are very effective and achieve significant speedups of up to 68x in common, practical use cases compared to earlier approaches and alternative implementation choices.

Short Biography

Yannis Ioannidis (Ph.D., Computer Sciences, UC Berkeley; MSc, Applied Mathematics, Harvard University; Diploma, Electrical Engineering, National Technical University of Athens) is the

President of the Association of Computing Machinery (ACM). He is a Professor at the Department of Informatics and Telecommunications of the National and Kapodistrian University of Athens as well as an Associated Faculty at the "Athena" Research and Innovation Center, where he also served as the President and General Director for 10 years. His research interests include Database and Information Systems, Data Science, Data and Text Analytics, Data Infrastructures and Digital Repositories, Recommender Systems and Personalization, and Human-Computer Interaction, topics on which he has published over 170 articles in leading journals and conferences and also holds four patents. His work is often inspired by and applied to data management and analysis problems that arise in industrial environments or in the context of other scientific fields (Social Sciences and Humanities, Life Sciences, Physical Sciences) and the Arts. He has been the coordinator and legal entity head of OpenAIRE, which implements the European policies on open access to research publications and data. He is the software director of the European Human Brain Project flagship initiative, the coordinator of the EOSC Future strategic project, which implements the core elements of the European Open Science Cloud, and a coordinator or partner in tens of other European and national research and innovation projects. He has also led or is currently leading the creation of new international or spin-off companies. He is an ACM and IEEE Fellow, a member of Academia Europaea, and a recipient of several research and teaching awards. He is currently the Greek delegate to the European Strategy Forum on Research Infrastructures (ESFRI) and a co-chair of the Global Climate Hub of the UN Sustainable Development Solutions Network.