Running with Scissors: Fast Queries on Just-in-time Databases

Anastasia Ailamaki
Professor, Computer and Communication Sciences,
Ecole Polytechnique Federale de Lausanne (EPFL)

Abstract

The amount of data collected in the last two years is higher than the amount of data collected since the dawn of time. Businesses are drowning in data, and need several months of ETL processing to barely prepare them for querying. Domain scientists collect data much faster than they can be transformed into valuable information and are often forced into hasty decisions on which parts to discard, potentially throwing away valuable data before it has been exploited fully. The reason is that query processing, which is the mechanism to squeeze information out of data, becomes slower as datasets grow larger. At the same time, the continuously increased number of hardware contexts ends up slowing processing down further, as keeping all cores busy with doing useful computation is difficult. Today’s query engines cannot harness but a fraction of the potential of new hardware platforms.

Is it possible to decouple query processing efficiency from the data growth curve? As data grows exponentially, which new techniques can we invent to process today’s data with the same efficiency as yesterday’s data (although the latter was half the size)? How can we remain hardware-aware without creating systems that are too specialized to today’s microarchitectures (and useless tomorrow)?

This talk advocates a departure from the traditional “create a database, then run queries” paradigm. Instead, data analysts should run queries on raw data, while a database is built on the side. In fact the database should become an implementation detail, imperceptible by the user. To achieve this paradigm shift, query processing should be decoupled from specific data storage formats. Ad-hoc primitives and dynamically synthesized operators are key for just-in-time query optimization and processing. Finally, exploitation of compute and memory resources should be seamless and based on hardware hints; extreme vertical integration is an enemy to forward compatibility.

Speaker Biography:

Anastasia Ailamaki is a Professor of Computer and Communication Sciences at the Ecole Polytechnique Federale de Lausanne (EPFL) in Switzerland. Her research interests are in database systems and applications, and in particular (a) in strengthening the interaction between the database software and emerging hardware and I/O devices, and (b) in automating data management to support computationally-demanding and data-intensive scientific applications. She has received an ERC Consolidator Award (2013), a Finmeccanica endowed chair from the Computer Science Department at Carnegie Mellon (2007), a European Young Investigator Award from the European Science Foundation (2007), an Alfred P. Sloan Research Fellowship (2005), eight best-paper awards in database, storage, and computer architecture conferences (2001-2012), and an NSF CAREER award (2002). She holds a Ph.D. in Computer Science from the University of Wisconsin-Madison in 2000. She is a senior member of the IEEE and a member of the ACM, serves as the ACM SIGMOD vice chair, and has served as a CRA-W mentor.