Data Processing in Modern Hardware

Gustavo Alonso
Systems Group
Department of Computer Science
ETH Zurich
alonso@inf.ethz.ch

ABSTRACT

Data processing is changing in radical ways from how it has developed in the last four to five decades.

On the one hand, data science and big data have brought an unprecedented growth and variety in data sizes, demanding workloads, data types, and applications. From studying social networks on graph data to genomics over string matching algorithms; from low latency key value stores used to retrieve user profiles to large scale data appliances focusing on data warehousing; from real time stream data processing to database engines on cloud platforms, the types, scope, and requirements on data management engines has grown enormously.

On the other hand, hardware is no longer a source of performance as it has been in the last decades. Instead, it has become a complex, fast evolving, highly specialized, and heterogeneous platform that requires considerable tuning and effort to use optimally. Today, hardware is not becoming necessarily faster per se but provides instead a wide range of options for accelerating and tuning applications through new features. Unlike what happened in the past, applications in general and database engines in particular, have to work much harder to extract performance improvements from new hardware as the exploitation of these new features is not automatic and often requires a redesign of the system. In addition, many of the opportunities offered by modern hardware are still without adequate support from high level tools such as compilers or debuggers, placing quite a burden on system designers.

In this talk I will discuss the issues in data processing that arise as a result of modern hardware: the need to deal with parallelism and distribution, the increasing importance of networking, the proliferation of accelerators, and the raise of heterogeneity in the machine. These issues are both a threat and a challenge, demanding a radical redesign of many aspects of data processing and database engines. Using examples from recent work ranging from query scheduling to hardware accelerators, I will present several exciting and radically new directions that are opening up for database research as a result of the advances being made in hardware. An important theme in the talk is the call for database designers and researchers to become proactive and identify the hardware features and characteristics that are needed to better support data processing.