Letter from the Special Issue Editor

Globalization has significantly altered the way in which the world operates. More than ever before are businesses required to operate at a global scale with customers expecting fast and seamless access to applications independent of where they are located. Similar, also the availability requirements for applications have changed over the last years. Whereas a decade ago, it was still sufficient to be able to sustain a single machine failure, nowadays mission-critical applications are expected to tolerate an entire data-center outage. Thus, it is no longer sufficient to think about data management at the scale of a few servers in some basement. Rather, we need to think about data management at the global-scale; from how is data replicated across data centers, over how bandwidth and latency of slow wide-area replication can be avoided or hidden to not negatively impact the application latency, up to how local regularization, e.g., privacy regulations in Europe, impacting what data can be stored at which location.

In this issue, we brought together an exciting collection of recent work in the space of global-scale data management addressing issues from reducing latency to how to execute machine-learning algorithms across data-center boundaries. The first paper, "A System Infrastructure for Strongly Consistent Transactions on Globally-Replicated Data" by Faisal Nawab and others, describes a first system infrastructure for global-scale transaction processing with serializability guarantees. The authors analyze the fundamental limits of executing transactions across data-center boundaries and develop a theoretical framework of the optimality of strongly-consistent transaction latency. Based on those findings, the authors present a transaction-processing protocol which takes advantage of the previously determined lower bounds on the transaction latency.

The second paper, "Writes: the dirty secret of causal consistency" by Lorenzo Alvisi and others, discusses in detail an alternative geo-replication strategy, which is based on causal consistency. Causal consistency has several advantages over other (stronger) consistency guarantees as it is able to provide high availability even in the presence of network partitioning. At the same time, causal consistency is still not widely adopted. The authors argue that one of the main reasons for this is, that current implementations do not handle write operations, especially write-skew, efficiently. Finally, by proposing two system, TARDiS and Occult, the authors aim to overcome this issue and make causal consistency more broadly applicable.

The third paper, "Cost-Effective Geo-Distributed Storage for Low-Latency Web Services" by Zhe Wu and Harsha V. Madhyastha, addresses the issue of the cost-latency trade-off by spreading the data across data centers of multiple cloud providers. Furthermore, the authors propose additional techniques, such as augmenting every GET/PUT request with a set of redundant requests to mitigate the effect of isolated latency spikes, among other techniques to reduce the high cost of geo-replication.

The final paper, "Towards Geo-Distributed Machine Learning" by Ignacio Cano and others, makes the case for Geo-Distributed Machine Learning (GDML). Regulatory requirements often prohibit to move data out of a country, but at the same time, many Machine-Learning applications require a global view of the geo-distributed data in order to achieve the best results. Thus, the authors propose a system architecture for geo-distributed training and show by means of an empirical evaluation on three real datasets that GDML is not only possible but also advisable in many scenarios.

I would like to thank all authors for their insightful contributions to this special issue. Happy reading!

Tim Kraska
Brown University
Rhode Island, US