Letter from the Special Issue Editor

Scientific computing used to be based on numerical simulations run on mid-range warehouse scale computers. This is no longer the case, due to the combination of strong application pull and technology push.

In order to get realistic models of a phenomenon in natural or engineered systems, scientists must analyze unprecedented volumes of data generated by new generations of instruments and experiments. In addition, they must run simulations at higher spatial resolutions, for longer simulation times and with higher dimension models, possibly combining multiple physical models of a phenomenon, or study multiple simultaneous phenomena. These new computational challenges stemming from scientific applicatins have triggered a convergence of traditional numerical simulation with machine learning and high-performance data analytics. Put differently, data science and eScience are merging.

The technology push is due to the planned transition to Exascale systems. Strictly defined, Exascale computers are capable of 10^{18} floating points operations per second (flops). More interestingly, they are three orders of magnitude faster than the High-Performance Computers deployed a decade ago. The first Exascale systems are expected in the coming year. In the US, three systems are being deployed: Aurora at Argonne National Lab, Frontier at Oak Ridge National Lab and El Capitan at Lawrence Livermore Lab. In China three existing pre-exascale systems are being extended: Sunway at the National Research Center of Parallel Computer Engineering and Technology (NRCPC in Wuxi, Jiangsu), Sugon (installed at the Shanghai Supercomputer Center) and Tianhe at the National Center of Defense Technology (NUDT in Changsha, Hunan). In Japan, Riken and Fujitsu have designed the Fugaku Exascale computer, which has been announced for 2021, 2022. It will be hosted at the RIKEN Center for Computational Science in Kobe. In Europe, three pre-exascale computers are under construction: Mare Nostrum 5 at the Barcelona Supercomputing Center, Leonardo at Bologna's CINECA and LUMI at the CSC Data Center in Kaajani, Finland.

In 2008, Kogge et al. surveyed the technology challenges in achieving Exascale systems. The main roadblock they identified was *transporting data from one site to another: on the same chip, between closely coupled chips in a common package, or between different racks on opposite sides of a large machine room.* Put differently, minimizing data movement is the key challenge on Exascale systems. This is a challenge in terms of architecture, but it is also a challenge for data management.

In this issue, leading researchers from the HPC and database communities present their work on data management at Exascale. The papers will give readers an insight in the nature of the application pull and technology push sketched above. They contain the lessons learnt at the forefront of scientific data management. They are very interesting points of departure for future work.

Mario Lassnig from CERN and his co-authors review their experience with the Rucio system, developed at CERN, to handle data in the ATLAS experiment. They detail the challenges they faced and how Rucio addresses them. They report on recent efforts to adapt Rucio in the context of other large-scale scientific projects.

Jerome Soumagne from HDF Group and his co-authors tackle the issue of performance and resilience for data services at Exascale. They propose Remote Procedure Call as a building block for such data services. The paper describes the design of Mercury, a new form of Remote Procedure Call adapted to large data transfers on low-latency network fabrics.

Jeremy Logan from Oak Ridge National Lab and his co-authors focus on ADIOS, the Adaptable I/O System, that provides a publish/subscribe abstraction for high-performance data services. The paper describe its design and its use in the context of near Exascale use cases. Based on lessons learnt and examples from a range of different projects, the authors discuss challenges and opportunities for future work on data management at Exascale.

Noel Moreno Lemus from LNCC (National Lab for Scientific Computing in Rio de Janeiro, Brazeil) and his co-authors tackle the issue of large-scale spatio-temporal simulations. More specifically, they focus on answering uncertainty quantification queries over such simulation results. This is a great example of the convergence of numerical simulation and query processing.

Finally, Alberto Lerner from the eXascale Infolab at U.Fribourg and his co-authors present their vision for in-network computing and near-storage processing. These techniques are crucial for bringing computation closer to data and thus tackle the issue of data movement. Their vision is based on a thorough analysis of the current generation of platforms and of the computation tasks that could be brought closer to data at rest or in movement.

These papers addresses several aspects of the state of the art in data management at Exascale and they outline a range of challenges. There are many opportunities for the database community to engage with the high-performance computing community to tackle these challenges. As Jim Gray once wrote: *The next decade will be exciting*!