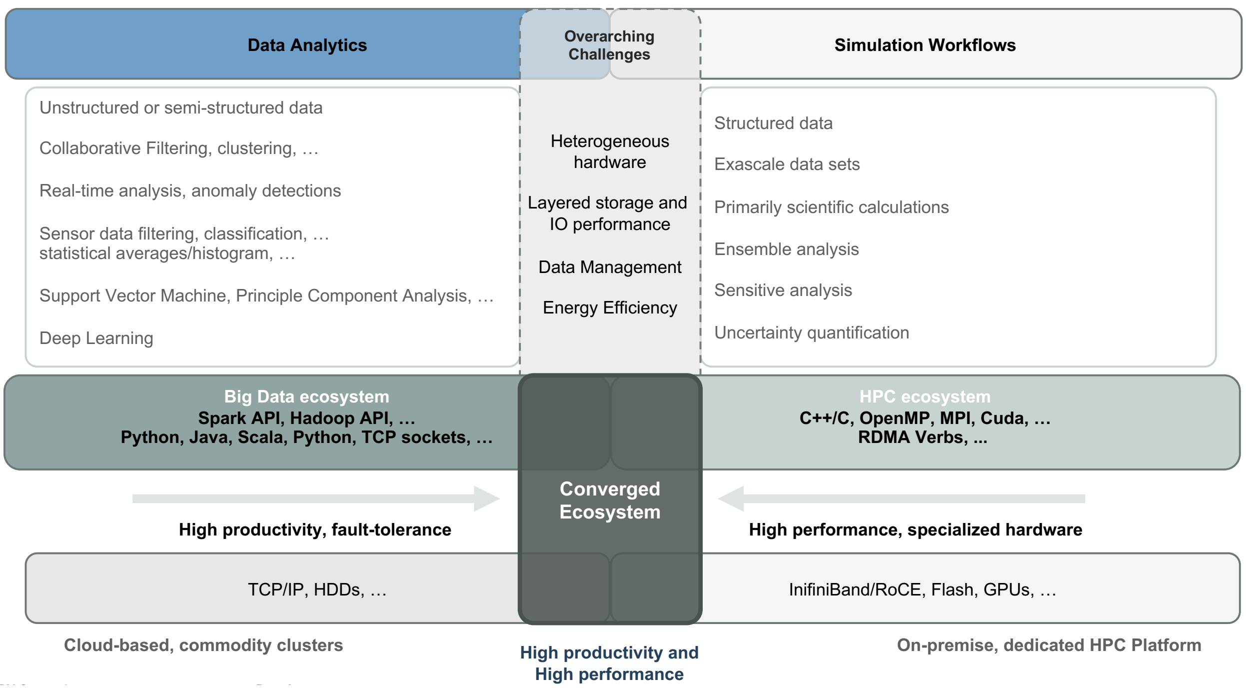
Towards a Converged Software Ecosystem for Data Analytics and Extreme-Scale Computing

Carlos Costa, IBM T.J. Watson Research Center – chcost@us.ibm.com

The explosion of data generated by emerging extreme-scale simulation has been motivating a paradigm shift in scientific discovery: from simulation-centric to data-centric discovery. The unprecedented scale of data sets and the rise of powerful data analytics techniques made a data-driven approach crucial in advancing scientific inquiry in many disciplines in science and engineering.

In these emerging scientific workflows, which often embody the entire inference cycle of discovery, data analytics is a key element in enhancing traditional simulation. Rich and scalable data analytics support for in-situ analysis is expected to enable real-time extraction of actionable information and in-depth post-processing in large amounts of simulation data. A converged software ecosystem for data analytics and extreme-scale simulation would enable existing simulation workflows to be more intelligent, more productive, and more robust, and enable bigger or newer and more accurate science.



A converged ecosystem requires the effective co-deployment and integration of two traditionally disjoint software ecosystems: the Big Data and HPC ecosystems. While these ecosystems share some of the same overarching challenges imposed by trends in system design, distinct design goals and software development cultures make their integration challenging. The HPC community has been traditionally keen to specialization at both the software and hardware levels, often trading productivity for performance at the expense of increased overall system complexity. Conversely, the Big Data community evolved around the goal of enabling cost-effective parallel computing on cloud-based commodity clusters, for which high productivity and fault-tolerance are high priorities.

The challenges in reconciling these two approaches arise at multiple levels. At application and runtime levels, one challenge is how to leverage advantages in both ecosystems without disrupting APIs and breaking existing libraries and frameworks. For Big Data frameworks, that translates into how to leverage specialized hardware and enable effective scaling without disrupting APIs that many libraries depend on (e.g. Apache Spark's MLLib/GrapX, etc…). For traditional simulation code (e.g., MPI-based), the challenge translates into enabling more elastic deployment strategies with cloud technologies (e.g. containers and virtual machines) without sacrificing functionality and performance. At the data management level, one particularly difficult challenge is how to enable a data flow model that allows efficient data exchange across heterogenous processing frameworks (e.g., between an MPI application and a data analytics framework, such as Spark). Complex data-driven workflows typically expose both tightly and loosely coupled data sharing for which traditional low-level approaches, such as traditional message passing, are not efficient. At the resource management and scheduler level, a converged ecosystem would be required to support both batch and stream processing and allow the integration of on-premises and cloud platforms.

In response to these challenges, we envision a converged ecosystem that builds on the strengths of both ecosystems, adding extra services that facilitate interoperability. We have been working towards this vision in multiple fronts. In one front, we have been optimizing Apache Spark, adding support for specialized hardware, such as RDMA-enabled communication and GPU acceleration, and enhancing communication primitives for more effective scaling in large-systems [1][2][3]. These optimizations transparently enhance the performance of Spark libraries, making them more suitable for processing large data sets on extreme-scale platforms. In the data management front, we have been developing the Data Broker, a shared storage framework for data and message exchange. It provides a simple and intuitive API to access persistent or volatile storage through one or more distributed tuple-based namespaces, regardless of programming language and heterogeneity of the data [3][4]. It facilitates the integration of heterogenous workflows composed of simulation and data analytics. In the scheduler front, we have been identifying gaps and adding missing functionally to enable the deployment of simulation code with cloud schedulers, like Kubernetes [3].

While we believe these are still the first steps in the direction of a truly converged ecosystem, we are encouraged by the benefits we have been observing with the application of our efforts to a set of hybrid workflows. In collaboration with Lawrence Livermore National Laboratory (LLNL), we demonstrated the benefits of Spark optimizations for extreme-scale knowledge discovery on the Sierra System [3]. Also with LLNL, we demonstrated the value of the Data Broker for data exchange in a hybrid workflow involving molecular dynamics simulation and machine learning for lipid cell membrane simulation [3]. In collaboration with MIT-Harvard Broad Institute, we demonstrated significantly improved performance and scalability of a reference DNA variant discovery pipeline involving heterogenous steps [5]. These among other efforts have helped us to identify gaps and missing functionally and validate and refine the vision for a converged ecosystem.

References

[1] Leveraging Adaptive I/O to Optimize Collective Data Shuffling Patterns for Big Data Analytics. B Nicolae, CHA Costa, C Misale, K Katrinis, Y Park - IEEE Transactions on Parallel and Distributed Systems (TPDS), 2017. link

[2] Towards Memory-Optimized Data Shuffling Patterns for Big Data Analytics. B Nicolae, CHA Costa, C Misale, K Katrinis, Y Park - Cluster, Cloud and Grid Computing (CCGrid), 2016. link

[3] Converged Ecosystem for Data Analytics and Extreme-Scale Computing, Carlos Costa, Workshop I: Big Data Meets Large-Scale Computing, Part of NSF program: Science at Extreme Scales: Where Big Data Meets Large-Scale Computing, 2018. link

[4] Data Broker - https://github.com/IBM/data-broker

[5] Optimization of Genomics Analysis Pipeline for Scalable Performance in a Cloud Environment, CHA Costa, C Misale, F Liu, M Silva, H Franke, P Crumley, B D'Amora, IEEE International Conference on Bioinformatics and Biomedicine (industry track), 2018 (to appear).

[6] SparkGA: A Spark Framework for Cost Effective, Fast and Accurate DNA Analysis at Scale. H Mushtaq, F Liu, C Costa, G Liu, P Hofstee, Z Al-Ars - ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB), 2017. link