

PCJ Java Library as a solution to integrate HPC, Big Data and Artificial Intelligence workloads

Friday, 26 March 2021 14:30 (30 minutes)

With the development of peta- and exascale size computational systems there is growing interest in running Big Data and Artificial Intelligence (AI) applications on them.

Big Data and AI applications are implemented in Java, Scala, Python and other languages that are not widely used in High-Performance Computing (HPC) which is still dominated by C and Fortran. Moreover, they are based on dedicated environments such as Hadoop or Spark which are difficult to integrate with the traditional HPC management systems.

We have developed the PCJ library (Parallel Computing in Java), a tool for scalable high-performance computing and big data processing in Java. In this paper, we present the basic functionality of the PCJ library with examples of highly scalable applications running on the large resources. The performance results are presented for a different classes of applications including traditional computational intensive (HPC) workloads (e.g. stencil), as well as communication intensive algorithms such as Fast Fourier Transform (FFT). We present implementation details and performance results for Big Data type processing running on petascale size systems. The examples of large scale AI workloads parallelized using PCJ are presented.

Primary author: Dr NOWICKI, Marek (N. copernicus University)

Co-authors: Prof. BAŁA, Piotr (University of Warsaw); Dr GÓRSKI, Łukasz (ICM University of Warsaw)

Presenters: Dr NOWICKI, Marek (N. copernicus University); Prof. BAŁA, Piotr (University of Warsaw)

Session Classification: Converging High Performance infrastructures: Supercomputers, clouds, accelerators Session

Track Classification: Virtual Reserach Environment (including Middleware, tools, services, workflow, ... etc.)