Julian Kunkel's Papers

Title and Abstract

April 28, 2019

1 HDTrace - A Tracing and Simulation Environment of Application and System Interaction

HDTrace is an environment which allows to trace and simulate the behavior of MPI programs on a cluster. It explicitly includes support to trace internals of MPICH2 and the parallel file system PVFS. With this support it enables to localize inefficiencies, to conduct research on new algorithms and to evaluate future systems. Simulation provides upper bounds of expected performance and helps to assess observed performance as potential performance gains of optimizations can be approximated.

In this paper the environment is introduced and several examples depict how it assists to reveal internal behavior and spot bottlenecks. In an example with PVFS the inefficient write-out of a matrix diagonal could be either identified by inspecting the PVFS server behavior or by simulation. Additionally the simulation showed that in theory the operation should finish 20 times faster on our cluster - by applying correct MPI hints this potential could be exploited.

Comments

2 Towards an Energy-Aware Scientific I/O Interface

Intelligently switching energy saving modes of CPUs, NICs and disks is mandatory to reduce the energy consumption.

Hardware and operating system have a limited perspective of future performance demands, thus automatic control is suboptimal. However, it is tedious for a developer to control the hardware by himself.

In this paper we propose an extension of an existing I/O interface which on the one hand is easy to use and on the other hand could steer energy saving modes more efficiently. Furthermore, the proposed modifications are beneficial for performance analysis and provide even more information to the I/O library to improve performance.

When a user annotates the program with the proposed interface, I/O, communication and computation phases are labeled by the developer. Run-time behavior is then characterized for each phase, this knowledge could be then exploited by the new library.

Comments

- Several best practices are realized within the ADIOS library to increase usability and performance, for instance aggressive write-behind is performed, and MPI collectives transfer file information to decrease the burden on metadata servers.
- Available modules include NetCDF, HDF5, MPI (collective or independent), POSIX and several asynchronous staging modules.

3 Simulating Parallel Programs on Application and System Level

Understanding the measured performance of parallel applications in real systems is difficult - with the aim to utilize the resources available, optimizations deployed in hardware and software layers build up to complex systems. However, in order to identify bottlenecks the performance must be assessed.

This paper introduces PIOsimHD, an event-driven simulator for MPI-IO applications and the underlying (heterogeneous) cluster computers. With the help of the simulator runs of MPI-IO applications can be conducted in-silico; this includes detailed simulation of collective communication patterns as well as simulation of parallel I/O. The simulation estimates upper bounds for expected performance and helps assessing observed performance.

Together with HDTrace, an environment which allows tracing the behavior of MPI programs and internals of MPI and PVFS, PIOsimHD enables us to localize inefficiencies, to conduct research on optimizations for communication algorithms, and to evaluate arbitrary and future systems. In this paper the simulator is introduced and an excerpt of the conducted validation is presented, which demonstrates the accuracy of the models for our cluster.

Comments