OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

Parallel Computing and I/O

Lecture Parallel Storage Systems

Parallel programming is becoming increasingly important since even phones and laptops contain multiple processor cores nowadays. Supercomputers can contain up to several million cores and have become a useful and important tool for a wide rang of scientific domains. The analyses and simulations enabled by them have accelerated the process of gaining scientific insight considerably.

The amount of collected and produced data is growing exponentially; it has to be stored, analyzed and processed efficiently sinc I/O significantly affects overall performance. Vastly different rates of performance development for processors and storage hardware result in a performance imbalance, which makes it even more important to take a close look at storage systems in ord to be able to meet future demands.

The lecture will teach the fundamentals of parallel storage systems and I/O; the exercises will allow transferring and applying the acquired skills with a system programming language such as C, C++ or Rust.

As part of the lecture, we will cover the complete storage stack: Storage devices and networks (hard disk drives, solid-state disk storage area networks etc.), local and distributed file systems (in kernel and user space, novel concepts like snapshots and deduplication) as well as the I/O interfaces layered on top (POSIX, MPI-IO, NetCDF and ADIOS). Moreover, we will discuss reasons and solutions for performance problems as well as alternative approaches for I/O (such as cloud interfaces). Problems and examples will be motivated using real-world scientific applications.

- Lecture: Thursday 11:15-12:45 (G29-307)
- Exercises: Monday 11:15-12:45 (G22A-111)
- Contact: Michael Kuhn and Michael Blesel

Participants will learn how parallel applications perform I/O using different programming concepts and how I/O can be optimized. Additionally, they will gain insight into and practical experience with the internals of storage and file systems.

Required skills:

- Practical knowledge of a programming language and the ability to create simple applications

Recommended skills:

- Basic knowledge about operating systems
- Basic knowledge about parallel programming

- 2022-04-07: Introduction
- 2022-04-14: Storage Devices
- 2022-04-21: File Systems
- 2022-04-28: Modern File Systems
- 2022-05-05: Parallel Distributed File Systems
- 2022-05-12: MPI-IO
- 2022-05-19: Libraries
- 2022-06-02: Optimizations
- 2022-06-09: Performance Analysis
- 2022-06-16: Data Reduction
- 2022-06-23: Grid, Cloud and Peer-to-Peer
- 2022-06-30: Current and Future Developments

- 2022-07-07: Research Talks

- 2022-04-07: Introduction
  - Deadline: 2022-04-21, 23:59
- 2022-04-14: Debugging and Checkpoints
  - Deadline: 2022-04-28, 23:59
- 2022-04-28: I/O Tools
  - Deadline: 2022-05-05, 23:59
- 2022-05-05: Dummy File System
  - Deadline: 2022-05-19, 23:59
- 2022-05-19: Memory File System
  - Deadlines: 2022-05-26 and 2022-06-09, 23:59
- 2022-06-09: Persistent File System
  - Deadlines: 2022-06-16 and 2022-06-30, 23:59

- High Performance Parallel I/O (Prabhat und Quincey Koziol)