HPC.NRW

aiXcelerate 2021: Part I – File I/O

HPC.NRW Competence Network

CC BY SA

Focus of Part I: File I/O

HPC.NRW

- **What is not covered in this workshop?**
  - Deep theoretical background of file I/O
    - Details about file systems not in use on CLAIX
    - Theoretical discussions about hardware

- **What is covered in this workshop?**
  - Focus is operational side on CLAIX
  - Overview of available hardware and software
  - How to configure and analyze applications on the cluster
  - Hands-on examples

CC BY SA

Does I/O Matter to me?

Philipp Martin (RWTH)

aiXcelerate 2021: Part I – File I/O

CC BY SA

Detailed Agenda for Session

HPC.NRW

- What is file I/O?
  - Why do we need files?
  - Metadata and bandwidth
- When does this become Important?
  - Example measurements
  - Application vs cluster
- How to identify and solve Problems
  - Profile and analyze applications
  - Improve code
  - Choose the correct file system
- Outlook

CC BY SA

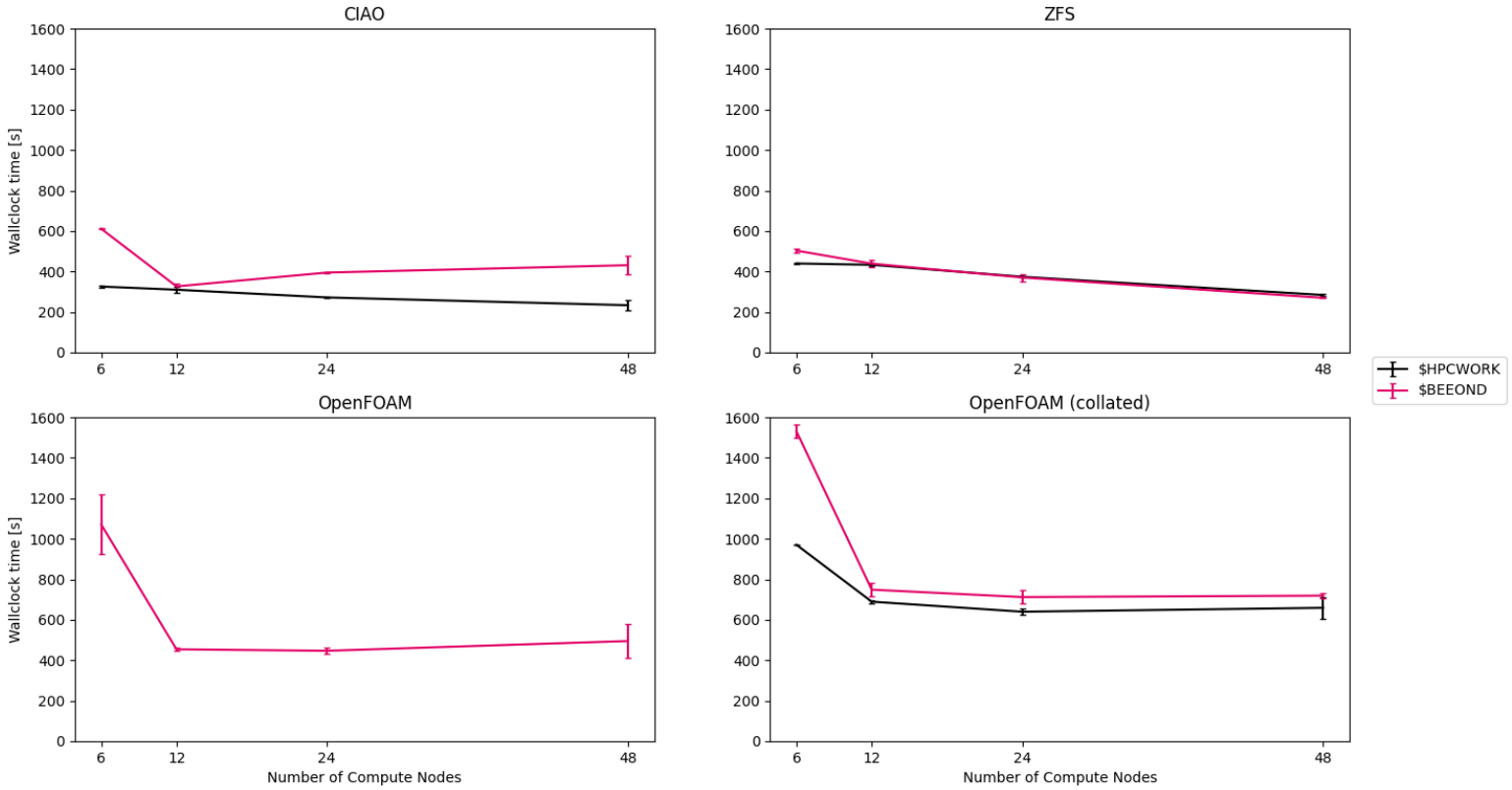
What is file I/O?

HPC.NRW

- Why do we need files?
  - Persistent storage
  - Input data
  - Output data
  - Intermediate data (stop-and-resume)
- Metadata and bandwidth
  - Metadata: anything that is not the content of the file
  - Filename, permissions, creation/modification date, size, location…
  - Metadata operations (IOPS) vs Bandwidth performance (Bytes/s)
  - Many small files vs. few large files

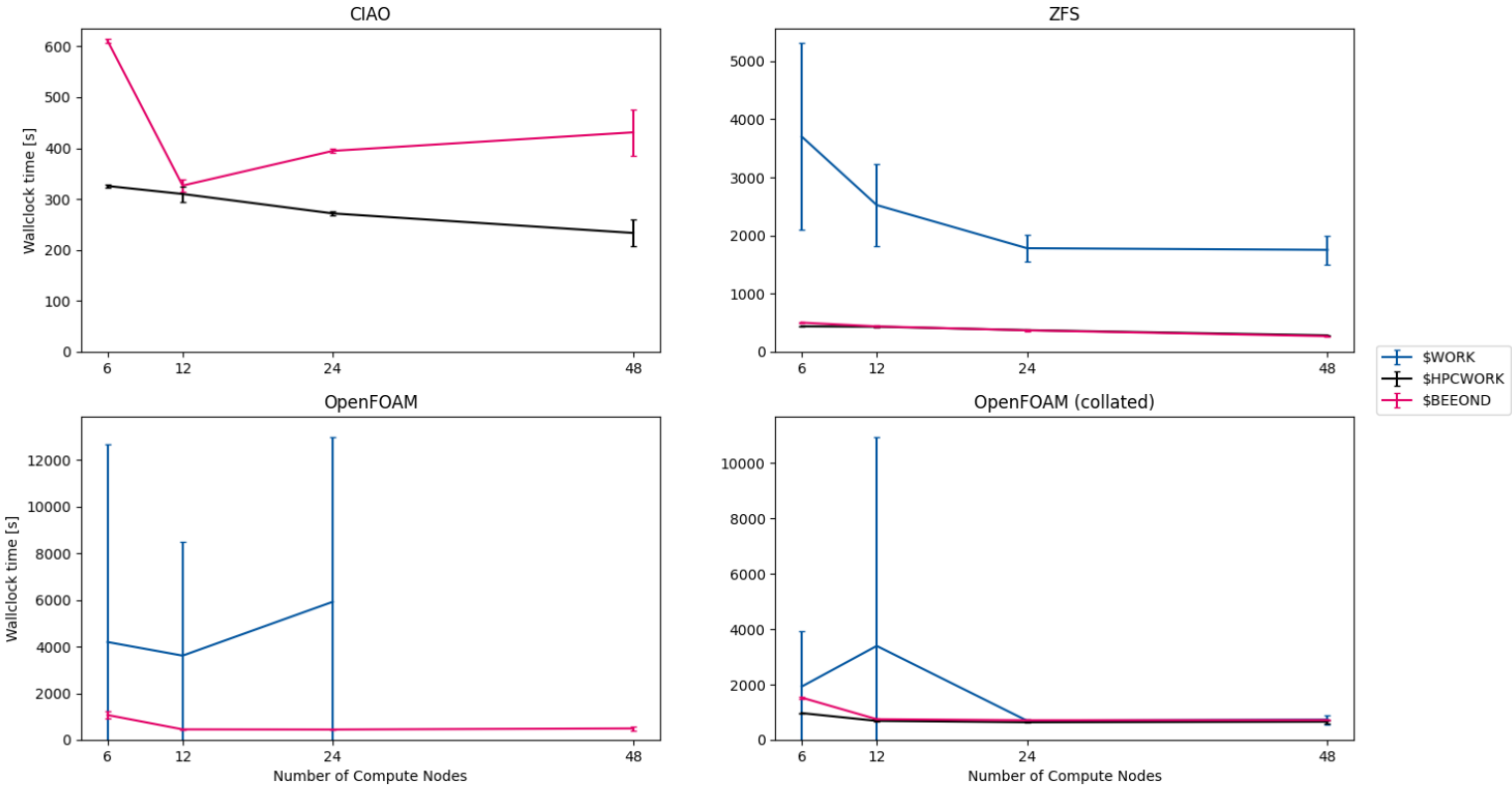
CC BY SA

When does file I/O become important?



CC BY SA

When does file I/O become important?



CC BY SA

When does file I/O become important?

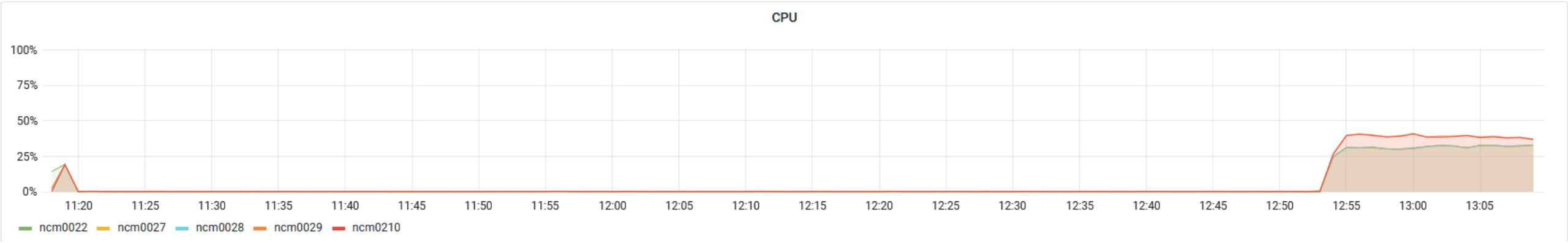
HPC.NRW

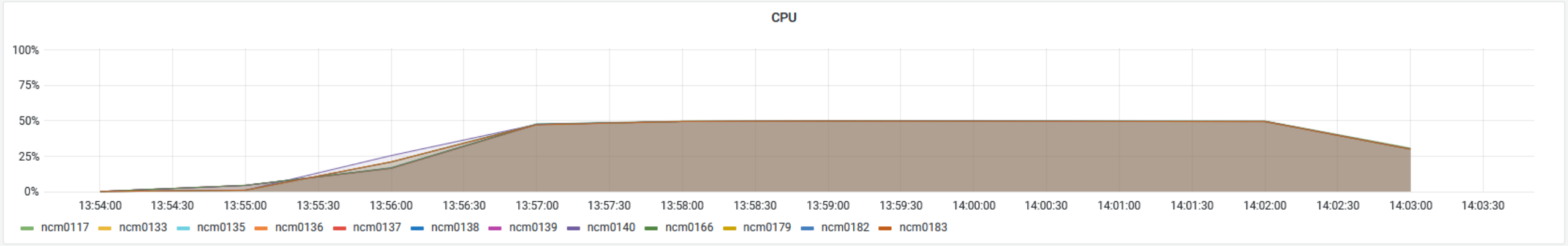
- Application side considerations
  - Am I using a significant number of I/O operations?
  - Am I using a significant amount of data?
  - Simulations: Writing of time-step data
  - Machine learning: Large training sets, large networks
  - Big Data
- Cluster side considerations
  - Is the file system suited for the particular task?
  - Is the file system overloaded?

CC BY SA

When does file I/O become important?

HPC.NRW





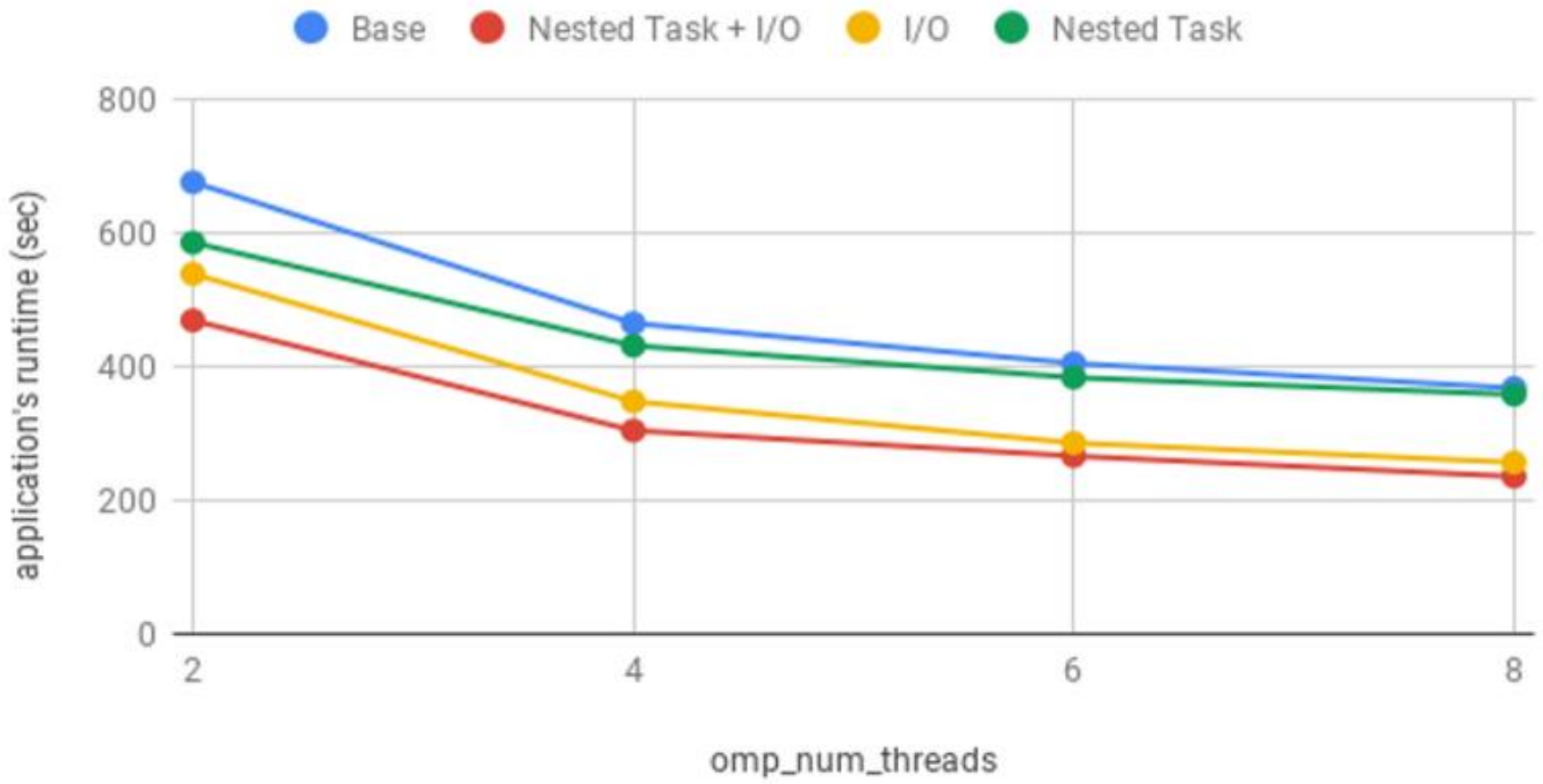
CC BY SA

How to Identify and Solve Problems

- Profile and analyze
- Improve access patterns
- Choose the correct file system

CC BY SA

– **Example: Calculix I/O Buffering (POP2 Project)**



CC BY SA

Schedule – I/O

HPC.NRW

Monday, December 6

Start	End	Topic	Speaker
09:00	10:00	Does I/O matter to me?	RWTH
10:00	10:15	Break	
10:15	11:15	Overview of I/O Technologies on CLAIX	Philipp Martin (RWTH)
11:15	11:30	Break	
11:30	12:30	Using Darshan for I/O Analysis	Radita Liem (RWTH)
12:30	14:00	Lunch Break	
14:00	15:00	Using Score-P & Vampir for I/O Analysis	Marc-André Hermanns (RWTH)
15:00	15:15	Break	
15:15	16:45	BYO: Preparation of benchmarks and job submissions for user codes	

Tuesday, December 7

Start	End	Topic	Speaker
09:00	10:00	I/O Libraries: Overview and MPI-IO	Marc-André Hermanns (RWTH)
10:00	10:15	Break	
10:15	11:00	I/O Libraries: HDF5	Sebastian Lührs (Forschungszentrum Jülich)
11:00	11:15	Break	
11:15	12:00	I/O Patterns Best Practice	Radita Liem (RWTH)
12:00	13:30	Lunch Break	
14:00	15:00	BYO: Review benchmark results	
15:00	15:15	Break	
15:15	16:45	BYO: Lightning talks about take-aways	

CC BY SA