

BigStorage

Storage-Based Convergence Between HPC and Cloud
to Handle Big Data

Michael Kuhn

Research Group Scientific Computing
Department of Informatics
Universität Hamburg

2016-10-06



informatik
die zukunft

About us: Scientific Computing



- Analysis of parallel I/O
- I/O & energy tracing tools
- Middleware optimization
- Alternative I/O interfaces
- Data reduction techniques
- Cost & energy efficiency

We are an Intel Parallel Computing Center for Lustre
("Enhanced Adaptive Compression in Lustre")

1 BigStorage

2 DKRZ Contributions

3 Summary

BigStorage

- BigStorage is a European Training Network (ETN)
 - Main goal is to train future data scientists
 - Estimated demand in the 100,000s over the next few years
- Focus on performance and energy consumption
 - Backed by use cases
- Consortium with ten partners
 - Seven research centers/universities
 - Three large companies
- Three associated partners from industry
- 15 Early Stage Researchers (ESRs)

BigStorage...

- Problems are addressed on different layers of the I/O stack
 - Applications that can exploit the data
 - Middleware for HPC and cloud environments
 - Infrastructure that provides compute and storage capabilities
- Appropriate techniques and algorithms have to be available
 - Areas include statistics, machine learning, visualization, databases and high performance computing
 - ESRs should be educated on different facets of data science

BigStorage...

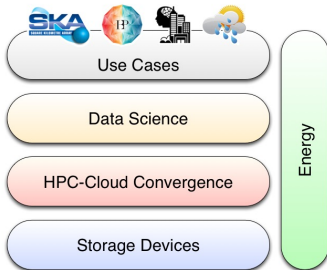
Project overview

Data Science

- Modelling Big Data processing
- Energy-efficient analysis
- Data-driven decision making for Big Data applications

HPC-Cloud Convergence

- Applications
- Middleware, operating in the cloud and HPC environments
- Infrastructure for Storage and Computing



Storage Devices

- Storage acceleration
- Storage convergence
- Storage isolation

Energy

- Compression or de-duplication for storage footprint reduction
- Hints from application to storage system, enabling energy consumption reduction

- Split into four work packages and use cases
- Three work packages cover the whole I/O stack
 - Energy is a cross-cutting issue
 - Solutions are evaluated using use cases

BigStorage...

- Four use cases to evaluate developed solutions
 - Human Brain Project
 - Square Kilometre Array
 - Smart Cities
 - Climate Science
- Each use case is handled by a working group
 - Working groups consist of 3–4 ESRs and 1–2 advisors
 - Gathering requirements and extracting benchmarks

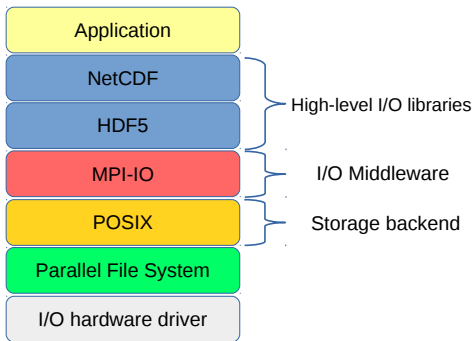
BigStorage...

- Activities to support training
- Training schools and workshops
 - Topics: big data, storage solutions, HPC, cloud, data science etc.
- Seminars
 - Topics: intellectual property, commercial exploitation of results, entrepreneurship etc.
- Additional language courses, scientific writing etc.
- Secondments and internships
 - Typically two secondments at other partners
 - Duration of 3–6 months

DKRZ

- Involved in HPC-cloud convergence and energy
 - Make use of cloud technologies for HPC workloads
 - Provide interface extensions to improve cost efficiency
- Leading the climate science working group
 - Analysis of I/O requirements for climate applications
 - Benchmark extraction for other interested parties
- Workloads: checkpointing/output, post-processing
 - Many climate applications still use serial I/O
 - Ensemble runs (tens to hundreds of model runs)
- Challenges
 - High volumes and exchangeability of data
 - Result analysis, interpretation, visualization
 - Data life cycle management

I/O stack



- Many climate science applications use CDI or NetCDF
 - Easy to change the underlying storage technology
- NetCDF is the de facto standard format
 - Export to NetCDF is a requirement
- Existing technologies have problems regarding scalability

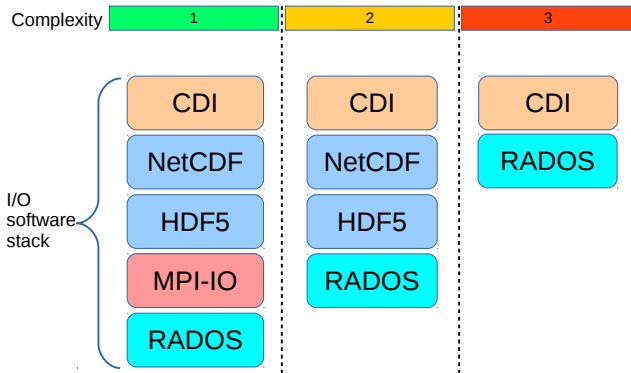
Data storage

- Data is typically stored in a parallel file system
 - HDF uses MPI-IO, which stores data in Lustre
 - Access to shared files has performance issues due to POSIX
 - Object stores provide sufficient functionality for MPI-IO
- Ceph is a storage platform
 - Provides block, file and object storage
 - No single point of failure
 - Resilient through replication
 - Scalable into the exabyte range

Data storage...

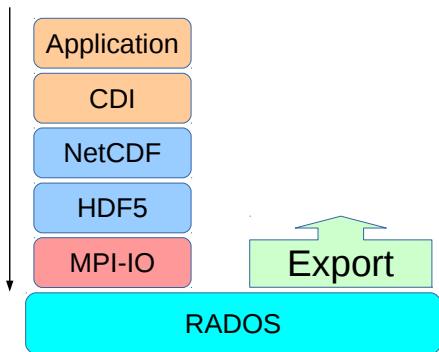
- Ceph consists of several components
 - RADOS (reliable, autonomous, distributed object store)
 - librados (direct access to RADOS)
 - radosgw (REST gateway with S3 and Swift compatibility)
 - RBD (reliable, distributed block devices)
 - CephFS (POSIX-compliant distributed file system)
- RADOS provides the basis with additional features built on top
 - Direct access if additional features are not required
 - Allows avoiding POSIX overhead

Data storage...



- RADOS can be integrated at different levels of the stack
 - Increasing difficulty from left to right
- Proof of concept: RADOS backend for MPI-IO

Data storage...



- Have to be able to export normal NetCDF files
 - Important for archiving and exchanging data
 - Might be as easy as h5copy

Summary

- BigStorage aims to train data scientists
 - Future demands are high
- Different facets of data science are taught
 - Applications, HPC and cloud middleware, storage infrastructure
 - Additional seminars and courses
- Climate applications use NetCDF
 - MPI-IO functionality can be provided by an object store
 - RADOS backend for MPI-IO allows avoiding POSIX overhead
- More information at
 - <http://bigstorage-project.eu/>
 - <https://wr.informatik.uni-hamburg.de/research/projects/bigstorage/start>