

EXPLORING HIGH PERFORMANCE STORAGE WITH DAOS

Kenneth Cain
Mohamad Chaarawi
Johann Lombardi



Hewlett Packard
Enterprise

Adrian Jackson
(a.jackson@epcc.ed.ac.uk)
Nicolau Manubens



Aims

- Understand object storage hardware and software
- Learn about DAOS and filesystems on DAOS
- Learn about DAOS lower-level APIs and using them for your applications
- Get hands on with DAOS and high performance storage hardware
- Learn how to program, and design, for object storage systems

Aims cont.

- Understand/think about your application data requirements
 - Both storage and discovery
- Thinking about different ways you undertake I/O or storing data
- Move beyond bulk, block-based, I/O paradigms

Format

- Lectures and practicals
- Slides and exercise material available online:
 - <https://github.com/adrianjhpc/ObjectStoreTutorial>
 - Exercises will be done on remote system (NEXTGenIO)
 - We will provide access for these

Timetable

- 08.30 Introduction
- 08.40 Object storage and storage approaches
- 09.20 Practical: Benchmarking different storage approaches
- 09.30 DAOS programming APIs
- 10.00 Break
- 10.30 DAOS programming APIs, filesystems, and examples
- 11.30 Practical: Using DAOS for applications
- 11.50 Performance, design, and summary
- 12.00 Finish

Object Storage

- Design and performance considerations are the challenge
 - Programming against the interfaces is (relatively) easy(ish)
 - Direct use often straight forward (i.e. filesystem interfaces)
 - More intelligent functionality takes more work/more specialised
- Design for functionality
 - When to store, when transactions should happen, what granularity I/O operations should be, what failures can you tolerate, etc..
- Design for performance
 - Memory size, I/O, data access costs, etc...

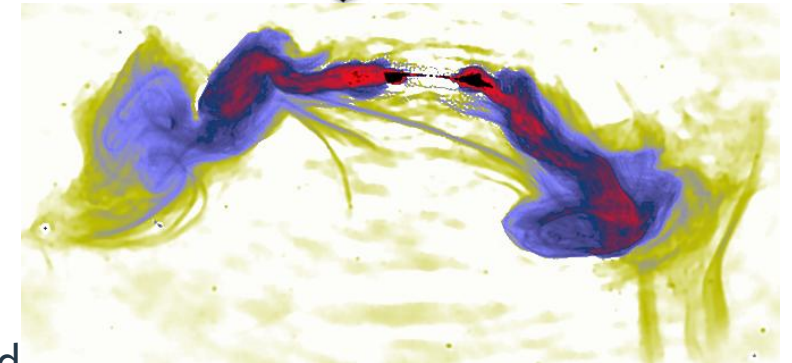
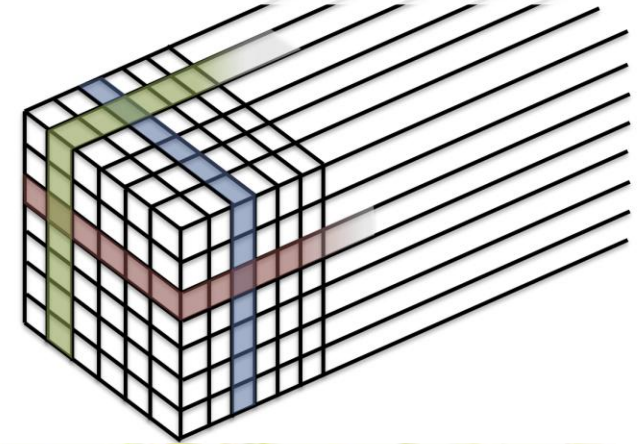
Object storage

- Data stored in unstructured objects
 - Data has identifier
 - Size and shape can vary
 - Metadata can also vary
- Originally designed for unstructured data sets
 - Bunch of data with no specific hierarchy required
- Can also enable efficient/fast access to data in different structures
 - Supports different creation, querying, analysis, and use patterns

Object storage

- Can enable efficient/fast access to data in different structures
 - Supports different creation, querying, analysis, and use patterns
- Granular storage with rich metadata
 - Data retrieval leverages metadata
 - **Build structure on the fly**
- Weather/climate
 - Pursuing optimal I/O for applications
 - Weather forecasting workflows
 - End-to-end workflow performance important
 - Simulation (data generation) only one part
 - Consumption workloads different layout/pattern from production
- Radio astronomy
 - Data collected and stored by antenna (frequency and location) and capture time
 - Reconstruction of images done in time order
 - Evaluation of transients or other phenomenon undertaken across frequency and location

Clients want to do **different** analytics across **multiple** axis



Summary

- Please don't hesitate to ask questions!
- We have a practical sessions
 - Login account will remain active for you to try out using DAOS after the tutorial (not really enough time to finish them during the tutorial today)
 - Email a.jackson@epcc.ed.ac.uk to get an account on the system we will use for practical/try out sessions (if you use the subject "Object store account" that will help us respond to the emails)