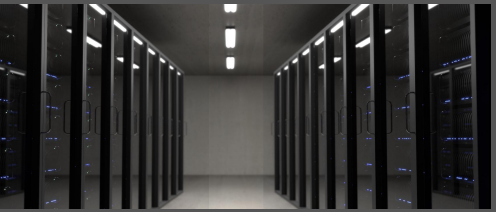
Department of Computer Science

HPS

University of Reading

Smarter Management using Metadata and Workflow Expertise



ESD

Limitless Storage
Limitless Possibilities

https://hps.vi4io.org

Julian M. Kunkel

BoF: Knowledge Is Power: Unleashing the Potential of Your Archives Through Metadata

2019-11-21

LIMITLESS **POTENTIAL** | LIMITLESS **OPPORTUNITIES** | LIMITLESS **IMPACT**

Outline

University of Reading

Data Organization

University of Reading

What scientists care about

- Organization of the (hierarchical) namespace
- Choosing the storage location/s
  - Often (if not policy driven): when to archive/how?
- Management of the applications that generate data
- Loads of scripts to run applications, manage data
- The reproduction of experiments after their data is archived
- Sometimes: High-level databases to manage data
  - The allow to ingest metadata and organize data accordingly

Data Organization

University of Reading

What scientists should care about

- Setup of the workflow all processing steps
- Providing accurate metadata describing experiments and data
- Documentation of experiment: Ensuring reproducibility
  - Supported by tools that record the data lineage
- Information lifecycle management
  - How long to keep data/workflow/reproducible data
  - Metadadata and rules guiding the life of data

How to Search and Address Data?

University of Reading

High-Level questions relevant to scientists

- What experiments did I run yesterday?
- Show me the data of experiment X, with parameters Z...
- Cleanup unneeded temporary stuff from experiment X
- Compare the mean temperature of one model for one experiment across model versions

A Semantic Namespace might help!

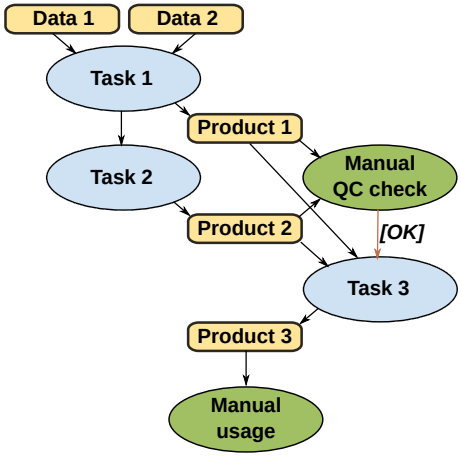
- Allow to explore data based on user metadata
- User-defined properties but provide means to validate schemas
- Similar to an MP3 library (search by Genre/Year/Artist/...)

Outline

Workflows

University of Reading

- Consider workflow from 0 to insight
  - Needs/produces data
  - Uses tasks
    - Parallel apps?
    - Big data tools?
    - Manual analysis
  - May need month to complete
  - Manual tasks are unpredictable
  - What are users interested in?
- Not well described in HPC
  - Mostly hardcoded in scripts
- Can we exploit workflows?
  - Can we use **archives** more?
  - Enforce ILM as needed by users



Scenario: Large Simulation

University of Reading

- Assume large scale simulation, timeseries (e.g., 1000 y climate)
- Assume manual data analysis needed (but time consuming)
- We need all 1000 y for detailed analysis!

A typical workflow execution

- Run simulation for 1000 year simulation time
  - Store various data on (online) storage
  - Keep checkpoints to allow reruns
  - Maybe backup data in **archive**
- Explore data to identify how to analyze data
- At some point: Run the analysis on all data
- Problem: Occupied storage capacity

Alternative Workflows Done by Scientists

University of Reading

Recomputation

- Run climate simulation
  - Store checkpoints
  - Store only selected data (wrt. resolution, section, time)
- Explore data
  - Run recomputation to create needed data (e.g., last year)
- At some point: run analysis across all data needed
- This is a manual process, must consider
  - Runtime parameters
  - System configuration/available resources
  - We are trading compute cycles vs. storage
  - It would be great if a system considers costs and does this automatically

Another Alternative Workflows

University of Reading

Provided by more intelligent storage and better workflows

- Run simulation
  - Store checkpoints on node-local storage
    - Redundancy: from time to time restart from another node
  - Store selected data on online storage (e.g., 1% of volume)
    - Also store high-resolution data sample (e.g., 1% of volume)
  - Store high-resolution data directly in a cold **archive**
- Explore data on snapshot
- Month later: schedule analysis of data needed
  - The system retrieves data from the **archive**
  - Performs the scheduled operations on **streams** while data is pulled in
  - Informs user about analysis progress
- Some people do this manually or use some tools to achieve similarly
  - Aim for domain & platform independence and heterogenous HPC landscapes

Outline

University of Reading

ESiWACE: http://esiwace.eu

University of Reading

The Centre of Excellence in Simulation of Weather and Climate in Europe

- Prepare the European weather and climate community
  - Make use of future exascale systems
- Goals in respect to HPC environments
  - Improve efficiency and productivity
  - Supporting the end-to-end workflow of global Earth system modelling
  - Establish demonstrator simulations; run at the highest affordable resolution
- Funding via the European Union's Horizon 2020 program

esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

Earth-System Data Middleware

University of Reading

A transitional approach towards a vision for I/O addressing

- Scalable data management practice
- The inhomogeneous storage stack
- Suboptimal performance and performance portability
- Data conversion/merging

Design goals of the Earth-System Data Middleware

1. Relaxed access semantics, tailored to scientific data generation
2. Site-specific (optimized) data layout schemes
3. Ease of use and deploy a particular configuration
4. Enable a configurable namespace based on scientific metadata

ESiWACE2 Plans for ESDM

University of Reading

- FUSE prototype to dynamically build a hierarchical semantic namespace
  - E.g., `<model>/<date>/<variable>`
- Supporting post-processing, analytics and (in-situ) visualization
  - Support of computation offloading within ESDM
  - Integration with analysis tools, e.g., Ophidia, CDO
  - Direct data exchange between processes

Long-term goals

- Cost-modelling for optimized workflows utilizing heterogenous storage
- Performing operations while streaming data from **tape**
  - Or any storage/compute opportunity $\Rightarrow$ **Liquid Computing**

Outline

University of Reading

Summary

University of Reading

- Manual data placement leads to suboptimal results
- Semantic namespace is an opportunity to abstraction
  - Necesary to manage the data deluge
- With ESDM, we explore some aspects as part of the ESiWACE project
- Workflows must be lifted to a higher level
  - Utilization of heterogenous storage/compute infrastructure
  - See NGI initiative at VI4IO
  - https://ngi.vi4io.org