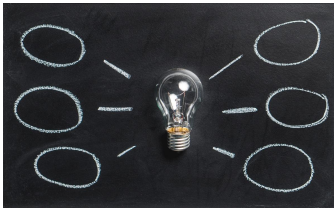
HPS
https://hps.vi4io.org

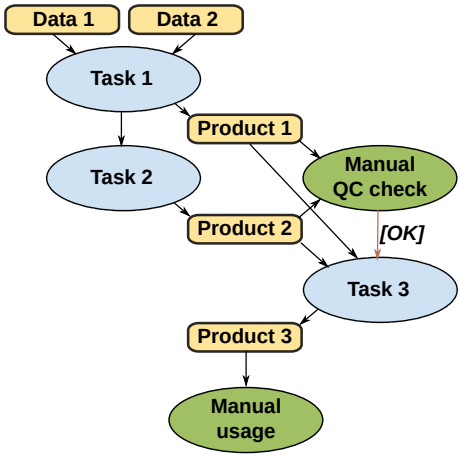
Julian M. Kunkel

Lifting the user I/O abstraction to workflow level a possibility or in vain?



Workflows: My personal definition

- Workflow: Steps from 0 to insight
  i.e. what users are interested in
  - Needs/produces data
  - Uses tasks
    - HPC and big data tools
    - Manual analysis
  - Spans across HPC system, cloud
  - May need months to complete
  - May involve mnual tasks
- Often partially described in scripts
- Would a proper description not support understandability?
  - Could potentially be exploited by (runtime) system?



Is the current abstraction level already good enough?

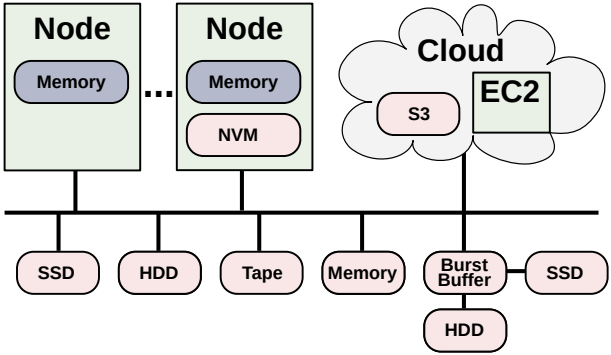
- Why do we still have to analyze I/O access patterns for POSIX?
  - It obfuscates the use-case / rationale behind the low-level I/Os
- Do we have enough insight about what workflows are executed in the DC?
  - Do we exploit this knowledge automatically or manually?
- Is HDF5 or ADIOS good enough to describe I/O in a single application?
- Are current workflow systems good enough to execute 0 to insight?

Can we lift the abstraction level higher?

Exploiting Workflow Knowledge for Planning HPC Resources

- Scientists deliver
  - detailed but abstract workflow orchestration
  - (containers with) all software
  - data management plan with data lifecycle
  - time constraints and budget
- Data centers and vendors
  - Simulate the execution before workflow is executed
  - Determine the best option to run
    Rough estimates for: Costs, runtime, energy consumption
- Systems
  - Utilize the information to orchestrate I/O
  - Make decisions about data location and placement:
    Could trade compute vs. storage and energy/costs vs. runtime

Automatic: Coexistence of Storage/File Systems? Too far away?



- We shall be able to use all compute/storage technologies concurrently
  - Without explicit migration etc. put data where it fits, compute where sensible
  - Across vendor and system boundaries
  - Administrators just add a new technology (e.g., hybrid) and users benefit

A Potential Approach in the Community?

- I believe the community must lift abstraction to enable better analysis

Can we follow the MPI Forum and actually work toward standardization?

- **Standardization** of a high-level *data model & interface* & workflow spec
- Development of a reference implementation of a **smart runtime system**
- **Demonstration** of benefits on socially relevant data-intense apps