

Opportunities for Integrating I/O Capabilities with Cylc



Limitless Storage
Limitless Possibilities

<https://hps.vi4io.org>

Julian M. Kunkel

Cylc Weather User Group; Supercomputing 2018

2018-11-15

Opportunities of Integrating I/O* with Cylc from 10k feet



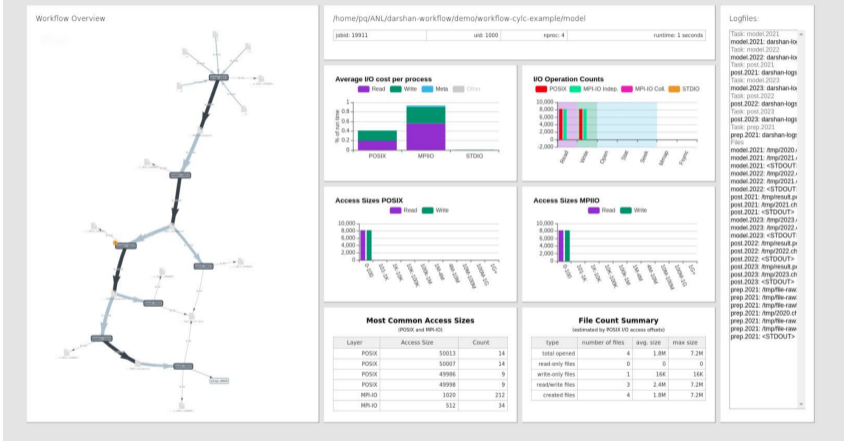
- Diagnosis: increasing understanding
- Prescriptive analysis: optimizing performance
- Predictive and prescriptive analysis: for users, data centers, vendors

Monitoring of I/O behavior through workflows

- Understanding and assessing I/O behavior
- See performance regressions
- Tracking of other utilization metrics is possible (e.g.; Grafana; DKRZ)
- Detecting abnormal I/O behavior

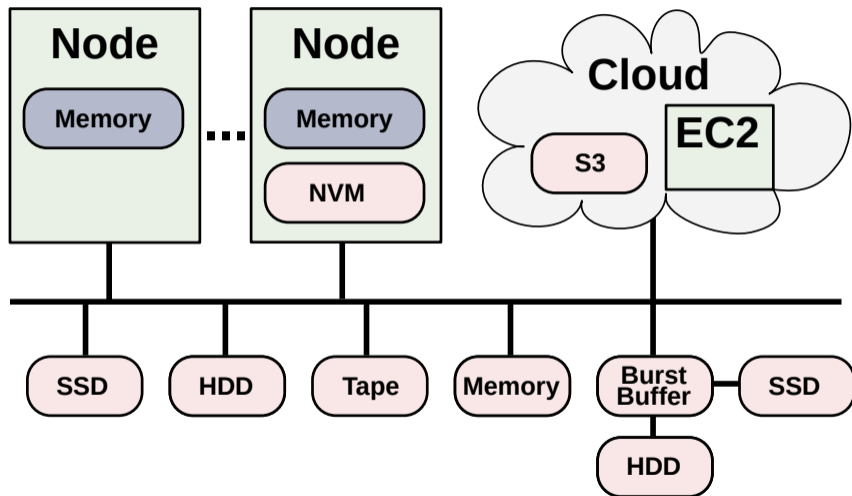
I/O Monitoring for Workflows

http://my.datacenter/workflow-io?workflow_id=314159



Paper: *Toward Understanding I/O Behavior in HPC Workflows*, 2018, PDSW-DISC Workshop, J. Luttgau, S. Snyder, P. Carns, J. Wozniak, J. Kunkel, T. Ludwig

Optimizing Heterogeneous Storage Landscape



Opportunities of Integrating I/O* with Cylc from 10k feet



■ Improved I/O management

- ▶ Including Information Lifecycle Management infos (data life)
 - How long to keep data, type of data...
 - Also: include interactive usage
- ▶ Potentially: Avoid using filenames at all
 - Explicit input/output definition (dependencies) instead of implicit

```
$ export CYLC_TASK_CYCLE_POINT=2012-08
```

```
$ cylc cycle-point --offset-years=2 --template=foo-CCYY-MM.nc  
foo-2014-08.nc
```

- ▶ Specify scientific metadata (like Mars)
 - Instead of a filename; API switches possible...
- ▶ Record data lineage (individual steps of data provenance)

■ Automatic scheduling of storage resources

- ▶ E.g. reusing local storage where possible
- ▶ Better resilience to data loss (restart and automatic recomputation?)
- ▶ Schedulers to create optimal mappings to diverse hardware

Opportunity Knowing Workflows Long-Term Perspective

- Performance modelling (simulation or via. recorded behavior)
 - ▶ How long will the workflow run?
 - ▶ How to change the system / storage to improve performance?
 - ▶ What are the costs to run a workflow?
 - ▶ Imagine to include compute model, too!
- Submission of workflow descriptions for project proposal
 - ▶ Data center could predict benefit, costs
 - ▶ May hand over to vendors, explore usage of alternative systems