Fine-Grained Profiling for Data-Intensive Workflows†

Nan Dun, Kenjiro Taura, Akinori Yonezawa
Graduate School of Information Science and Technology, The University of Tokyo

Introduction

ParaTrac is a user-level profiler using file system and process tracing techniques for data-intensive workflow applications. ParaTrac enables users to quickly understand the detailed I/O characteristics from entire application to specific processes or files, it also automatically exploits fine-grained data-processes interactions in workflow to help users intuitively and quantitatively investigate realistic execution of data-intensive workflows.

- Effortless user-level profiling unmodified distributed applications
- Scalability of tracing thousands of concurrent processes up to 16% overhead
- Automatic generation of informative I/O-specific and workflow-specific profiles
- Fine-grained investigation of complex applications

Profiling Approaches

- Application Tracing
  - File system call tracing by FUSE (Filesystem in Userspace)
  - Process tracing by /proc file system, ptrace, and taskstats
  - Trace log persistence by SQL database (optional)
- Profile Generation
  - I/O profiles generation by statistical analysis
  - Workflow profiles generation by causal analysis
- Profile Analysis
  - Using standard statistical analysis
  - Applying graph manipulation and graph-theoretic algorithms on workflow DAGs
- Workflow Optimization
  - Tuning of underlying I/O subsystems, e.g., distributed file system
  - Optimize scheduling of workflow according to real data-job interactions

Open Source and Availability

- ParaTrac: http://paratrac.googlecode.com/
- GXP: http://gxp.sourceforge.net/

Conclusion and Future Work

- Effective assistance for data-intensive applications study and optimization
- Reuse profiles as a macro benchmark for workflow management systems by consistent replaying of profiles.

References

[1] I. Taylor, E. Deelman, D. B. Gannon, and M. S. (Eds), *Workflows for e-Science*. Springer-Verlag, 2006.

[2] S. Pandey and R. Buyaa, "Scheduling and management techniques for data-intensive application workflows," in *Data In- tensive Distributed Computing: Challenges and Solutions for Large-Scale Information Management*, T. Kosar, Ed. IGI Global, 2009.

[3] S. Bharathi, A. Chervenak, E. Deelman, G. Mehta, M.-H. Su, and K. Vahi, "Characterization of scientific workflows," in *Proc. The 3rd Workshop on Workflows in Support of Large- Scale Science*, 2008.

[4] L. Ramakrishnan and D. Ganno, "A survey of distributed workflow characteristics and resource requirements," Indiana University.

[5] Workflow generator. [Online]. Available: `http://vtcpc.isi.edu/pegasus/index.php/WorkflowGenerator/`

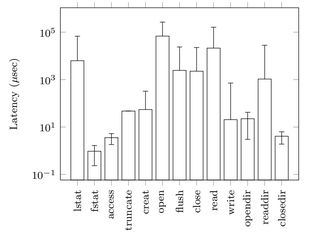
Profiling Real-World Workflows: Montage Astronomy Applications

I/O Characteristics

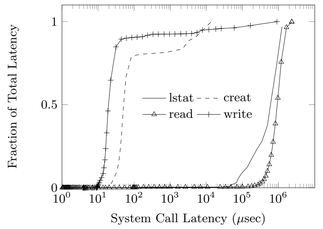
- I/O behaviors in different phases
- Dominant metadata or I/O operations
- Potential bottlenecks, e.g., high-latency operations
- Access characteristics of specific file/process

Phase	Jobs	I/O Size (MB) read	I/O Size (MB) write	Number of Files ro	Number of Files wo	Number of Files rw	Number of Files none	Number of Processes ro	Number of Processes wo	Number of Processes rw	Number of Processes none
Data A											
mProjectPP	28	57.77	221.07	30	0	56	0	1	0	28	0
mDiffFit	63	1023.87	25.32	58	126	63	0	64	63	63	0
mConcatFit	1	0.533	0.0127	61	1	0	0	1	0	1	0
mBgModel	1	0.0672	0.0144	3	1	0	0	1	0	1	0
mBackground	28	223.63	221.07	59	0	56	0	1	0	28	0
mImgtbl	1	0.644	0.0156	30	0	1	0	1	0	1	0
mAdd	1	133.93	75.81	59	2	0	0	1	0	1	0
mShrink	1	37.95	4.22	2	1	0	0	1	0	1	0
mJPEG	1	4.28	0.23	2	1	0	0	1	0	1	0
Total	125	1481.93	547.75	33	65	243	0	64	63	125	0
Data B											
mProjectPP	308	637.87	2472.25	310	2	614	0	1	0	308	0
mDiffFit	913	15114.33	675.64	618	1804	927	0	914	913	909	0
mConcatFit	1	0.34	0.18	915	1	0	0	1	0	1	0
mBgModel	1	0.314	0.015	3	1	0	0	1	0	1	0
mBackground	308	2524.48	2472.25	619	2	614	0	1	0	308	0
mAdd*	4	248.14	1587.56	613	8	0	0	1	0	4	0
mShrink	4	793.85	7.96	5	4	0	0	1	0	4	0
mImgtbl	1	0.085	0.0315	6	0	1	0	1	0	1	0
mAdd*	1	8.02	12.27	7	2	0	0	1	0	1	0
mJPEG	1	7.67	0.357	2	1	0	0	1	0	1	0
Total	1542	21568.07	7231.48	322	897	3084	0	914	913	1538	0

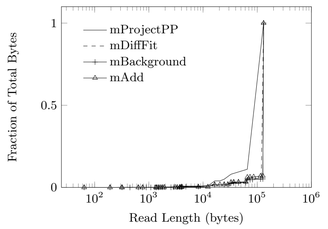
*mAdd is invoked twice in different phases.
ro: read only; *wo*: write only; *rw*: read+write; *none*: create/open only



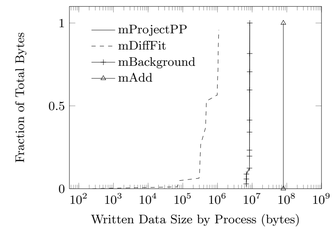
(a) System call latency



(b) CDF of system call latency



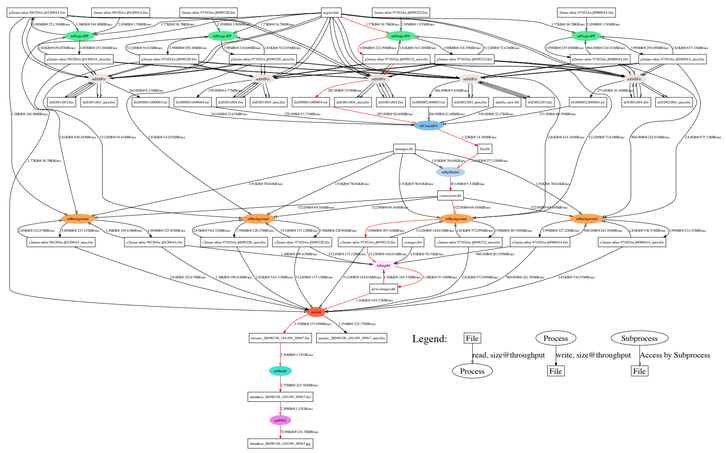
(c) CDF of overall read length



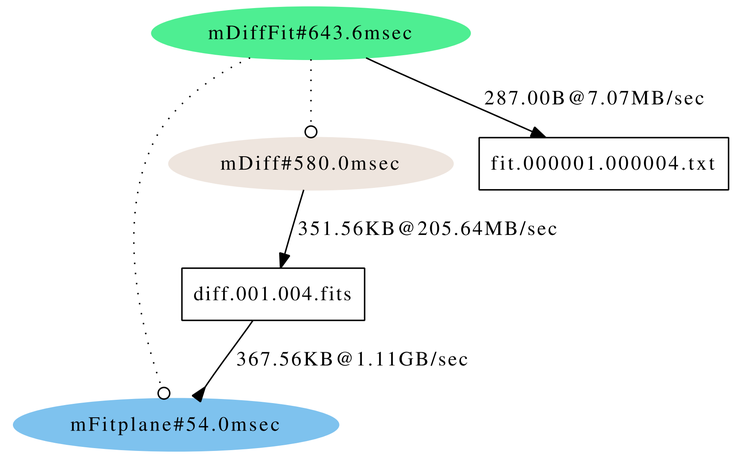
(d) CDF of written size by process

Fine-Grained Workflow Analysis

- Intuitive workflow DAGs annotated with realistic execution informations: Process command line, data path, data transfer volume, data transfer rate, I/O access type, etc.
- Applying graph-theoretic algorithms to find essential workflow information: critical path, critical nodes, hot spots (jobs/data with high degree), etc.
- Workflow debugging or generation of workflow description file by exploring data-job dependencies
- Detection of inefficient scheduling (e.g, data-job locality) if scheduling information from workflow management systems is given
- Suggestion of using data replicating, prefetching, or throttling strategies for optimal execution



(e) Full Workflow



(f) Sub Workflow

†*This research is supported in part by the MEXT Grant-in-Aid for Scientific Research on Priority Areas project "New IT Infrastructure for the Information-explosion Era" and Grant-in-Aid for Specially Promoted Research.*