J. Parallel Distrib. Comput. 74 (2014) 2967–2982

Contents lists available at ScienceDirect

J. Parallel Distrib. Comput.

journal homepage: www.elsevier.com/locate/jpdc

Experience with using the Parallel Workloads Archive

Dror G. Feitelson [a,*], Dan Tsafrir [b], David Krakov [a]

[a] Department of Computer Science, The Hebrew University, 91904 Jerusalem, Israel
[b] Computer Science Department, Technion – Israel Institute of Technology, 32000 Haifa, Israel

HIGHLIGHTS

- Reliable performance evaluations require reliable data about workloads.
- Workload logs may have data quality problems despite being automatically generated.
- Finding data quality problems is hard and findings must be reported.
- In some cases data quality can be improved e.g. by filtering dubious data.
- Such data analysis and cleaning is an important component of the scientific method.

ARTICLE INFO

Article history:
Received 14 October 2012
Received in revised form
29 May 2014
Accepted 25 June 2014
Available online 5 July 2014

Keywords:
Workload log
Data quality
Parallel job scheduling

ABSTRACT

Science is based upon observation. The scientific study of complex computer systems should therefore be based on observation of how they are used in practice, as opposed to how they are assumed to be used or how they were designed to be used. In particular, detailed workload logs from real computer systems are invaluable for research on performance evaluation and for designing new systems.

Regrettably, workload data may suffer from quality issues that might distort the study results, just as scientific observations in other fields may suffer from measurement errors. The cumulative experience with the Parallel Workloads Archive, a repository of job-level usage data from large-scale parallel supercomputers, clusters, and grids, has exposed many such issues. Importantly, these issues were not anticipated when the data was collected, and uncovering them was not trivial. As the data in this archive is used in hundreds of studies, it is necessary to describe and debate procedures that may be used to improve its data quality. Specifically, we consider issues like missing data, inconsistent data, erroneous data, system configuration changes during the logging period, and unrepresentative user behavior. Some of these may be countered by filtering out the problematic data items. In other cases, being cognizant of the problems may affect the decision of which datasets to use. While grounded in the specific domain of parallel jobs, our findings and suggested procedures can also inform similar situations in other domains.

1. Introduction

The study and design of computer systems requires good data regarding the workload to which these systems are subjected, because the workload has a decisive effect on the observed performance [1,15,38]. As an example, consider the question of scheduling parallel jobs on a large-scale cluster or supercomputer. As each job may require a different number of processors, this is akin to bin packing [7,25,36,48]. Hence the best scheduling algorithm may depend on the distribution of job sizes, or on the possible correlation between job size and runtime [27].

But how can we know what the distribution is going to be? The common approach is to collect data logs from existing systems and to assume that future workloads will be similar. The Parallel Workloads Archive, whose data is the focus of this paper, is a repository of such logs; it is accessible at URL www.cs.huji.ac.il/labs/parallel/workload/. The archived logs (see Table 1) contain accounting data about the jobs that executed on parallel supercomputers, clusters, and grids, which is necessary in order to evaluate schedulers for such systems. These logs have been used in many hundreds of research papers since the archive was started in 1999. Fig. 1 shows the accumulated number of hits that the Parallel Workload Archive gets when searching for it in Google Scholar (supplemented by the number of hits associated with the Grid Workloads Archive [21], which serves a similar purpose). The high citation count bears witness to the need for such data in the

* Corresponding author.
E-mail address: feit@cs.huji.ac.il (D.G. Feitelson).

http://dx.doi.org/10.1016/j.jpdc.2014.06.013

Table 1
Main logs in the Parallel Workloads Archive. (Some additional logs with mainly serial jobs or low utilizations are not listed.)

Log	*Period*	*Months*	*PEs*	*Users*	*Jobs*	*Util.*	*File*	*Cleaned*
NASA iPSC	10/93–12/93	3	128	69	42,264	0.47	NASA-iPSC-1993-3.swf	yes
LANL CM5	10/94–09/96	24	1,024	213	201,387	0.75	LANL-CM5-1994-4.swf	yes
SDSC Par95	12/94–12/95	12	400	98	76,872	0.72	SDSC-Par-1995-3.swf	yes
SDSC Par96	12/95–12/96	12	400	60	38,719	0.76	SDSC-Par-1996-3.swf	yes
CTC SP2	06/96–05/97	11	338	679	79,302	0.85	CTC-SP2-1996-3.swf	yes
KTH SP2	09/96–08/97	11	100	214	28,489	0.70	KTH-SP2-1996-2.swf	
SDSC SP2	04/98–04/00	24	128	437	73,496	0.84	SDSC-SP2-1998-4.swf	yes
LANL O2K	11/99–04/00	5	2,048	337	122,233	0.70	LANL-O2K-1999-2.swf	
OSC cluster	01/00–11/01	22	178	254	80,714	0.14	OSC-Clust-2000-3.swf	yes
SDSC Blue	04/00–01/03	32	1,152	468	250,440	0.77	SDSC-BLUE-2000-4.swf	yes
Sandia Ross	11/01–01/05	37	1,524	204	85,355	0.50	Sandia-Ross-2001-1.swf	
HPC2N	07/02–01/06	42	240	258	527,371	0.70	HPC2N-2002-2.swf	yes
SDSC Datastar	03/04–04/05	13	1,664	460	96,089	0.63	SDSC-DS-2004-2.swf	
SHARCNET	12/05–01/07	13	6,828	412	1,195,242	n/a	SHARCNET-2005-2.swf	
LLNL uBGL	11/06–06/07	7	2,048	62	112,611	0.56	LLNL-uBGL-2006-2.swf	
LLNL Atlas	11/06–06/07	8	9,216	132	60,332	0.64	LLNL-Atlas-2006-2.swf	yes
LLNL Thunder	01/07–06/07	5	4,008	283	128,662	0.88	LLNL-Thunder-2007-1.swf	yes
MetaCentrum	12/08–06/09	7	806	147	103,656	0.36	METACENTRUM-2009-2.swf	
ANL Intrepid	01/09–09/09	8	163,840	236	68,936	0.60	ANL-Intrepid-2009-1.swf	
PIK IPLEX	04/09–07/12	40	2,560	225	742,965	0.38	PIK-IPLEX-2009-1.swf	
RICC	05/10–09/10	5	8,192	176	447,794	0.87	RICC-2010-2.swf	
CEA Curie	02/11–10/12	20	93,312	722	773,138	0.29	CEA-Curie-2011-2.swf	yes

"PEs" was nodes or CPUs in old logs, today it typically represents cores.
"util" is the system utilization, i.e. the fraction of the resources that were allocated to jobs. It is not computed for SHARCNET because this is a grid system, and the constituent clusters became available at different times.
File names include a version number, as most logs were re-converted to swf when errors were found or new considerations were introduced.
"cleaned" specifies whether a cleaned version exists, where problematic data has been filtered out.

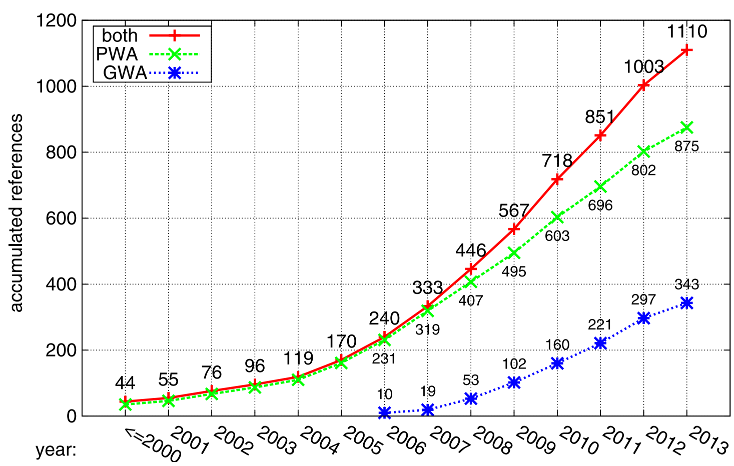

Fig. 1. Accumulated yearly number of hits received when searching for the Parallel Workloads Archive (PWA) and the Grid Workloads Archive (GWA) in Google Scholar as of 28 October 2013. GWA contains those logs from PWA that pertain to grid systems, as well as a few other grid logs. The query used was "Parallel Workload(s) Archive" (both singular and plural) and the archive's URL, and likewise for the grid archive. Papers that cite both archives are only counted once in "both".

research community and highlights the importance of using the data judiciously.

At first blush it seems that accounting logs should provide reliable and consistent data. After all, this is just a mechanistic and straightforward recording of events that happened on a computer system (as opposed to, say, genome data, which is obtained via complex experimental procedures that lead to intrinsic errors [30]). But upon inspection, we find that the available logs are often deficient. This is not a specific problem with the data that is available to us. All such logs have data quality problems, and in fact the logs available in the Parallel Workloads Archive actually represent relatively good data. We have additional logs that were never made public in the archive because an initial investigation found the data contained in them to be so lacking.

The issue of data quality has a long history (the International Conference on Information Quality has been held annually since 1996). The most general definition of data quality is "fitness for use", implying that it is not an objective but rather a context-sensitive attribute [44]. Indeed, work on data quality has identified no less than 20 dimensions of data quality, the top five of which are accuracy, consistency, security, timeliness, and completeness [23]. In the context of computer systems, practically all discussions have been about the quality of data *handled* by the system, e.g. the data contained in enterprise databases [6,28]. Low quality data has been blamed for bad business decisions, lost revenue, and even implicated in catastrophes leading to the loss of human life [16,17,31]. The quality of data in scientific repositories, such as biological genome data, has also been studied, both to assess the quality of existing repositories and to suggest ways to improve data quality [19,26,30]. Likewise, there have been problems with repositories used for empirical software engineering research; for example, massive repetitions of records taint evaluations of learning schemes that attempt to identify defective modules, by causing overlaps between the training and test datasets [18,34].

At the same time, there has been little if any work on the quality of data *describing* computer systems, such as workload data. In this paper we report on our experience with the data available in the Parallel Workloads Archive. We start the discussion by considering