SC15 BoF Participant Survey FRESCO: An Open Data Repository for Dependability Research and Practice

Name of the survey participant:	Data:	Nov. 18, 2015
realise of the salvey participants	Bata.	1101. 10, 2013

Part I: Utilization of data

1a. Indicate the usefulness of the following types of data in an open systems and workload data repository (enter a number 1-3, with 3=Important, 2=Neutral, 1=Not important):

Type of data	Answer	Type of data	Answer
Job-level activity and performance data		Syslog messages	
(libraries, executables and environment user			
accessed, performance measurement of IB,			
CPU, memory, filesystem during job runtime)			
Hardware performance counter		Type of application executed (eg.	
measurements		Genomics, Weather Forecast,	
		Structural analysis, Image	
		processing, etc)	
Measurements from system monitoring		Expert level of the user (e.g.,	
tools like Nagios or Ganglia		experienced, intermediate or	
		new/casual)	
Accounting logs for job submission (e.g., how		Other (please write in)	
long did a job run, did it terminate			
successfully or not)			

1b. What are the challenges in collecting such datasets from a cluster? (enter a number 1-3, with 3=Important, 2=Neutral, 1=Not important)

Challenges	Answer	Challenges	Answer
Degradation of job performance by the use		Difficulty in determining what to	
of measurement tools		collect and store, unless a	
		researchers approaches with specific	
		requests	
Cost of deploying measurement tools		Data privacy concerns	
Cost of storing, maintaining and updating		Other (please write in)	
such data			
Cost of documenting failure events			

1c. What would be useful usability features for the data repository? (enter a number 1-3, with 3=Important, 2=Neutral, 1=Not important)

Useful features	Answer	Useful features	Answer
Run analysis scripts on the server without		Visualize the data from search	
downloading the data			
Selection and download data in small		Availability of data for jobs	
manageable chunks of a few 100 MBs (e.g.		representing applications from	
over a short period)		diverse domains	
View detailed metadata explaining the data		Availability of data for a variety of	
fields next to the data itself		systems (e.g., accelerators)	
Tools for filtering, extracting and classify		Other desired features (please write in	1)
error data from various sources			

SC15 BoF Participant Survey FRESCO: An Open Data Repository for Dependability Research and Practice

Part II: Data sharing

2a. What issues are important to you when you consider sharing data through a repository like this? (enter a number 1-3, with 3=Important, 2=Neutral, 1=Not important):

Type of data	Answer	Type of data	Answer
Complete anonymization of the data(the		Prominent public recognition of the	
data sets will be non-identifiable to the		PIs and institutions contributing data	
actual source/person who contributed)			
Partial anonymization of the data (sensitive		A large consumer base for the data	
fields in the dataset like user name and		in the research community	
application name will be removed but the			
institution and machine names will be			
available)			
Data uploaded should be easy to cite and the		A large consumer base for the data	
contributor credited for the dataset		in the commercial community	
Other issues? (please write in)	•		•

2b. What type of data can you not share at all?

Ex: Application names, Library names, any framework used like mpi, hadoop, etc.

<<To be answered by Participants>>

Data that cannot be shared		

SC15 BoF Participant Survey FRESCO: An Open Data Repository for Dependability Research and Practice

Part III: Your role in the computational environment

3a. What kind of computational infrastructure do you have access to? (check all that applies)

Computational infrastructure	Answer (Y or N)
Desktop, lab servers	
Campus clusters	
XSEDE systems	
Open Science Grid (OSG)	
BlueWaters	
Commercial cloud services	
Other (please specify)	

3b. What is your role in it? (Ex: Cluster administrator, Researcher from academia, Practical user of the system, etc.)

Computational infrastructure	Answer (Y or N)	
System (cluster) administrator		
Researcher in academia		
Computational end-user of HPC systems		
System vendor		
Other (please specify)		