

SC15 BoF Participant Survey

FRESCO: An Open Data Repository for Dependability Research and Practice

Name of the survey participant: _____ Data: Nov. 18, 2015

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

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 of measurement tools		Difficulty in determining what to collect and store, unless a researchers approaches with specific requests	
Cost of deploying measurement tools		Data privacy concerns	
Cost of storing, maintaining and updating such data		Other (please write in)	
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 downloading the data		Visualize the data from search	
Selection and download data in small manageable chunks of a few 100 MBs (e.g. over a short period)		Availability of data for jobs representing applications from diverse domains	
View detailed metadata explaining the data fields next to the data itself		Availability of data for a variety of systems (e.g., accelerators)	
Tools for filtering, extracting and classify error data from various sources		Other desired features (please write in)	

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)	