



Contribution ID: 74

Type: Poster

Benchmarking the EGEE grid with Imbench; why it would be useful as a continuous service

Tuesday, 23 September 2008 16:44 (0 minutes)

Describe the activity, tool or service using or enhancing the EGEE infrastructure or results. A high-level description is needed here (Neither a detailed specialist report nor a list of references is required).

The activity we propose involves benchmarking of grid resources, in order to measure real grid characteristics. These serve as a proof that metrics-guided resource selection is nearly imperative. We have come to the conclusion that building performance models will become increasingly more important. We demonstrate that a single flag available under `/proc/cpuinfo` in linux systems, "lm", can account for considerable opportunities in optimizing the current grid infrastructure's job throughput.

Abstracts for online demonstrations must provide a summary of the demo content. Places for demos are limited and this summary will be used as part of the selection procedure. Please include the visual impact of the demo and highlight any specific requirements (e.g. network connection). In general, a successful demo is expected to have some supporting material (poster) and be capable of running on a single screen or projector.

The basic framework for doing benchmarking is already available as a python code package, freely available at sourceforge.net. We wish to demonstrate the usage of the root package (root.cern.ch) as a storage back-end for the benchmarking data, and how useful graphs and conclusions can be extracted from it, without much effort.

Report on the impact of the activity, tool or service. This should include a description of how grid technology enabled or enhanced the result, or how you have enabled or enhanced the infrastructure for other users.

What we want specifically to demonstrate is, that by skipping benchmarking and resource characterization, enormous amount of grid resources can be waisted or sub-optimally exploited. For example, the systems that are best on floating point of a given algorithm, say 64 bit operations, are not the ones that are optimal on memory transfers, and vice versa. The results are conclusive in demonstrating that the current Information System scheme is, at best, incomplete. What is very important to specifically clarify, is that the benefits of applying the benchmarking technique and resource characterization (estimated advantage is at 40%+ for our examples) can greatly outnumber the measurement system's overhead in itself (typically less than 0.5% of grid sites' capacity).

Describe the added value of the grid for your activity, or the value your tool or service adds for other grid users. This should include the scale of the activity and of the potential user community, and the relevance for other scientific or business applications.

We want to document the successful deployment of lmbench within two Virtual Organizations, namely ATLAS and SEE VO, on a temporary basis and for the purpose of data collection. We demonstrate that a metrics service can be built upon them and useful results can be obtained; some major insights are available already.

The metrics service can be deployed on a permanent basis within the context of any VO, in order to benefit the users of that particular realm and in their own discretion or, it could be provided as a service integrated within the WMS mechanism. The latter would provide enormous benefits in resource usage optimization, since it would allow to (a) minimize on a task execution duration (b) minimize on a workflow's CPU*hour product (c) minimize on a workflow's timespan (d) maximize aspects of grid efficiency.

Primary authors: Mr GEORGATOS, Fotis (GRNET); Mr KOURETIS, John (NTUA); Mr KOUVAKIS, John (NTUA)

Presenters: Mr GEORGATOS, Fotis (GRNET); Mr KOURETIS, John (NTUA); Mr KOUVAKIS, John (NTUA)

Session Classification: Demos and Posters

Track Classification: Poster