grid-control: The Swiss Army knife of job submission tools

Fred Stober, H. Stadie, C. Garbers, N. Kovalchuk, P. Schleper, M. Fischer

What is grid-control?

- grid-control is an open source job submission tool
- it supports all available grid interfaces
- it provides a consistent interface to different components
- it is a powerful tool for submitting and managing jobs

Parameterized jobs

Parameterized jobs are a feature of grid-control that allows users to specify job parameters in a concise and easy-to-use way. This feature is particularly useful for running jobs multiple times with different parameters, as it saves time and effort. With grid-control, users can define parameters and their ranges, and the system will automatically submit jobs with different parameter values. This makes it easy to perform sensitivity analyses or to explore the impact of different parameter settings on the job's outcome.

Requirements / Installation

- Requirements on the submission host:
  - Python 3.6 - 3.10
  - flake8 - black
  - pyyaml

- Requirements on the worker nodes:
  - basic, GNU/Make, ssh, web

Installation:

```
$ pip install -user grid-control
```

Performance

- Memory usage: on standard setup, the submission box uses approximately 2gb memory, while the worker nodes use 1gb memory.
- Time to submit jobs: on standard setup, the submission box takes approximately 1 minute to submit 100 jobs, while the worker nodes take approximately 30 seconds.

Design and Architecture

grid-control is using a highly modular design, where all functionality is provided by specialized plugins. The core package provides around 300 plugins, most of which belong to one or more categories. The OMS experiment software integration (OMS) supports CMM (CMM v1.6, DAS/DASH/EDGES) is done with over 2000 small plugins.

This is very similar to the plugin system used in CMM, which is a modular system that is adapted to the particular needs of the user and new functionality. The interplay between the plugins is the key to the framework.

Dataset resynchronization

Changes to the dataset file (additions or removals) as well as on the working level (resubmission or shrinking files) are propagated through the parameterized and transparent job requests to the grid framework.

For grid-control to work properly, it is important to run a "freeze" analysis on the dataset file using a single processing job that records changes on the dataset file before new jobs are started.

Example

- The information from the dataset file "example.dat" will yield 3 jobs (with filesize=90G, 90G, 90G).
- The environment reads GEN_PROCESS will be set to "center" for the first 2 jobs. Every 7th file of the dataset will be re-read and the specified dataset will be scanned and jobs created/destroyed if new files are found (files are missing).

Documentation

The documentation is currently in progress and will be made available in a user-friendly format. The documentation includes examples and instructions for setting up the grid-control environment, as well as best practices for utilizing the system.

https://github.com/grid-control/grid-control

Licensed under the Apache License 2.0
Copyright © 2007-2016 Karlsruhe Institute of Technology