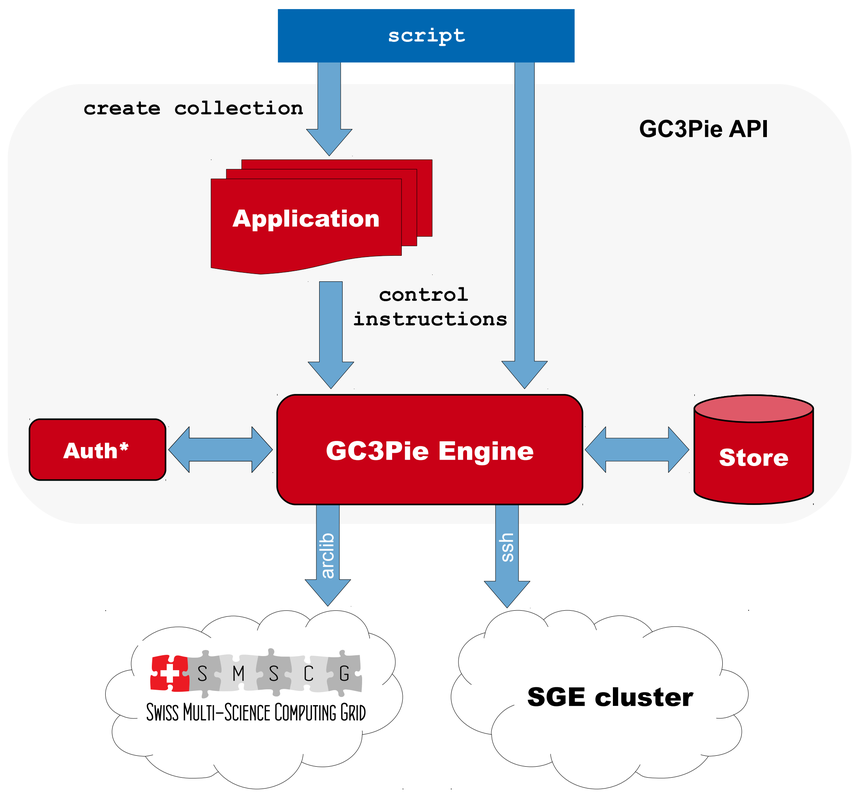
EGI-InSPIRE

Enabling High-Throughput Computational Chemistry on the Grid

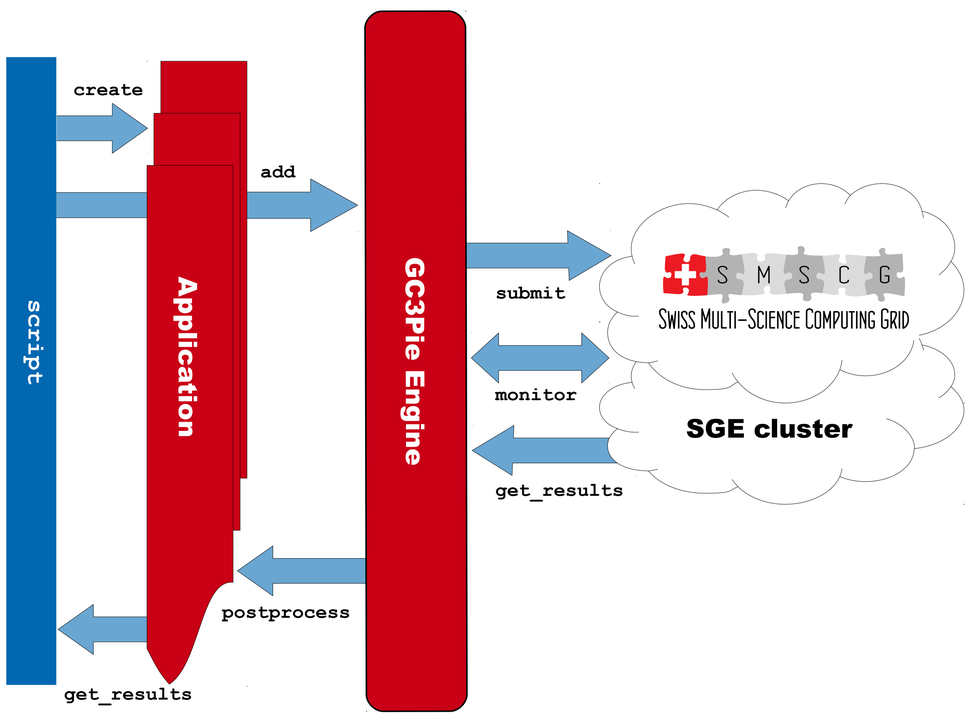
http://gc3pie.googlecode.com

GC3Pie is a suite of Python classes (and command-line tools built upon them) to aid in submitting and controlling batch jobs to clusters and grid resources seamlessly. GC3Pie aims at providing the building blocks by which Python scripts that combine several applications in a dynamic workflow can be quickly developed.



GC3Pie features

- Interact with underlying grid infrastructure controlling the execution of large collections of application objects.
  - Automated grid credentials management
- Resilient to network interruptions
- Runs user-level; no admin required.
- Tuned for the SMSCG national grid infrastructure (based on ARC) and for local SGE clusters. Extension for support glite on request
- Python & LGPL-licensed: Easy to understand & modify.



GC3Pie python APIs

- ***Engine****:* controls execution of large application collections in both synchronous and asynchronous mode
- ***Authentication***: credential management (slcs, grid/voms proxy, mproxy, ssh-keys); does not store user password
- ***Application***: User-specific, i.e. executables, input/output, data. Control mechanisms to determine failure.
- ***Task***: A meta-*Application*; can contain other Applications or Tasks to build workflows.
- ***Persistence***: Store and retrieve *Task* information for long runs & to tolerate problems.

High-throughput GAMESS analysis: ggamess

ggamess mainly has been designed for a method to predict reaction pathways for molecules without usage of chemical knowledge (hypersphere method[1]):

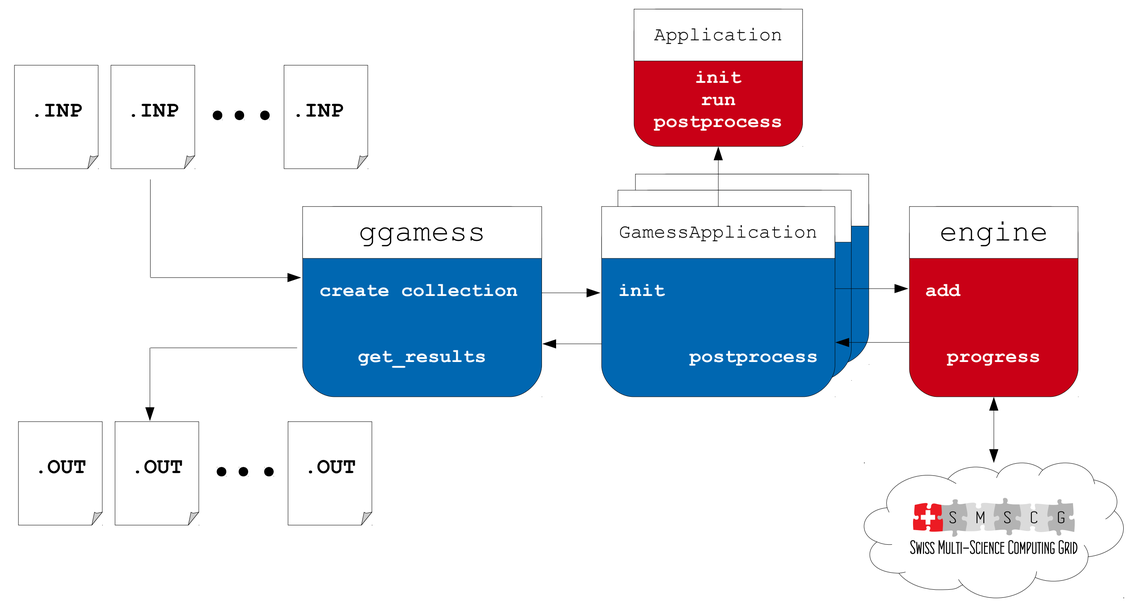
- Initial ground state geometry (GS) with its vibrational modes (normal coordinates)
- n normal coordinates constituting n-dimensional search space with centre representing GS
- Spherical depiction of points around the centre representing deformations of GS
- Submission of deformed geometries of each sphere to GAMESS by ggamess gives DFT-energies
- Plotting DFT-energies of points on surface of n-dim. sphere to figure out directions to transition states (TS)
- Calculation of reaction pathways from TS's leading to new GS's
- Repetition of 1. - 6. for each new GS

This breadth -first-search finally produces a complete network of reaction pathways for a given molecule (or set of molecules).

ggamess scans the specified INPUT directories recursively for '.inp' files, and submit a GAMESS job for each input file found; keeps a record of jobs (submitted, executed and pending) in a session file; job progress is monitored and, when a job is done, its '.out' and '.dat' file are retrieved back either to the submission host or to a gridFTP compliant storage service.

For this example, a single ggamess session has generated DFT-energy calculation for **17713** molecules on the ARC-based SMSCG infrastructure (failure rate: 0.05%)

[1] Maeda, S.; Ohno, K.; Chemical Physics Letters, 2003, 381, 177



On-line quantum chemistry databases: grundb

grundb is an interface for starting GAMESS analyses of molecules from the online GAMESS.UZH database (http://ocikbgtw.uzh.ch/gamess.uzh) on the Grid resources from the Swiss National Infrastructure SMSCG and local compute clusters.

- grundb launches a GAMESS job for each molecule of the chosen subset(s) of the GAMESS.UZH database.
- manages the job lifecycle, and
- finally print out a comparison table of stoichiometry reference data (from the database) and the same quantities as computed by GAMESS.
- A single run of grundb on the full test on the GAMESS.UZH database generates **1168** jobs.

grundb allows to easily test new QM methods (currently focused on GAMESS) against available on-line chemistry databases. It also provides utilities for building a local benchmark database.

People
Sergio Maffioletti, Riccardo Murri, Mike Packard, Timm Reumann

Acknowledges
Prof. Kim Baldridge (UZH)
Swiss Multi Science Computing Grid project (http://www.smscg.ch)
NGI-CH, SwiNG (http://swing-grid.ch)

Info: info@gc3.lists.uzh.ch