BioBlend – automating bioinformatics with Galaxy and CloudMan

Clare Sloggett1, Nuwan Goonasekera1,2,4, Enis Afgan1,3,4
1 Victorian Life Sciences Computation Initiative (VLSCI), University of Melbourne
2 Victorian eResearch Strategic Initiative (VeRSI), University of Melbourne, Melbourne, Australia
3 Center for Informatics and Computing (CIR), Ruder Bošković Institute (RBI)
4 Galaxy Project (http://galaxyproject.org)

BioBlend is a python library which wraps and documents the functionality of both the Galaxy and the CloudMan REST-based APIs. The Galaxy API gives users access to a rapidly expanding set of Galaxy functionality, while the CloudMan API allows users to manipulate the CloudMan cloud-based job runtime platform. These APIs can be used individually, or can be used together to automate the end-to-end provision and operation of a Galaxy cluster on the cloud.

The library is easily installable via PyPi and comes with detailed documentation and example scripts. BioBlend is released under the MIT license. Documentation and installation instructions can be found at http://bioblend.readthedocs.org/.

BioBlend wraps the REST-based APIs of Galaxy and CloudMan in a high-level language (python). It can be used to automate the provision of cloud-based infrastructure and/or operations on a Galaxy server. We plan to include the Galaxy Toolshed API into BioBlend as it matures. See below for an illustration of sample BioBlend calls.

```python
from bioblend.cloudman import CloudManConfig, CloudManInstance

# CloudMan

# Configuration:
cfg = CloudManConfig(access_key, secret_key, cluster_name, ami_id, instance_type, password, block_till_ready=True)

cmi = CloudManInstance.launch_instance(cfg)

cmi.active

cmi.enable_autoscaling(min_nodes=0, max_nodes=10)

# Galaxy API

gi = GalaxyInstance(url='http://galaxy_url', key=galaxy_api_key)

my_workflow = gw.workflows.import_workflow_from_json(my_string)

my_dataset = gi.libraries.upload_file_from_local_path(library_id, local_path)

datamap = {input_id: my_dataset["id"]}

result = gw.workflows.run_workflows(my_workflow["id"], datamap, history_name="Example output")
```

CloudMan allows users to launch an SGE compute cluster in the cloud, with a pre-configured Galaxy server installed tools, and reference genomes. The virtual cluster can be scaled dynamically to match demand, reconfigured, saved and relaunched, and shared for re-instantiation by other users. CloudMan runs on cloud providers such as Amazon (AWS) and the Australian Research Cloud (OpenStack).

CloudMan has its own REST-based API which is wrapped by BioBlend.


Use, contribute!

BioBlend wraps a subset of the still-growing Galaxy REST API. Code contributions, feedback and suggestions are welcome.

Documentation: http://bioblend.readthedocs.org/

Source: https://github.com/afgane/bioblend

Galaxy REST API source: https://bitbucket.org/galaxy/galaxy-central/src and explore lib/galaxy/webapps/galaxy/api/