RyuGoo-Saba
a CWL-based execution engine utilizing cloud resources for big data analysis in life science

Hirotaka Suetake¹ and Ryo Nakaki² ¹The University of Tokyo ²Rhelixa, Inc.

Project Website: https://github/Rhelixa-inc/RyuGoo-Saba/wiki
Source Code: https://github.com/Rhelixa-inc/RyuGoo-Saba

Introduction
We have created a command line tool and a cluster environment for running CWL on Azure.

How to use
1. Input Azure user ID, Password and SubscriptionID
2. Input Data URL and CWL file

Differences from existing ones
Similar things can be container orchestrations such as Kubernetes and Mesos, and other cloud enabled CWL runners such as Toil and Cromwell.

As a feature of this engine,
1. Automatically build the cluster environment itself using the user's ID and Subscription
2. Since the VM is assigned to each job, there is no need to be conscious of the resource pool

Method
RyuGoo-Saba is divided into three components.
1. Saba-CLI
   - Command Line Tool
2. Saba-Master
   - Job Handler
3. Saba-Worker
   - Job Executor

Saba-CLI
- Command line tools running on docker-compose
- Receive Azure user ID and Subscription ID, launch Saba-Master with Ansible on Azure
- Use Saba-Master's Rest API to provide commands to users

Saba-Master
- Provide REST API endpoint
- Receive CWL job from Saba-CLI, start VM as Saba-Worker with Ansible on Azure
- Manage Job status, log of Job, URL of result file, etc.
- When CWL job finishes, delete the Saba-Master VM

Saba-Worker
- Management of input data and output data using network storage
- Execute CWL job using sedLoad

- Each component is all running on docker-compose
- Communication with the REST API endpoint is encrypted by the JSON / TLS protocol.

Discussion and Future Work

Limitation
• Dependence on Azure
• Preparation time of VM
We think that these limitations are mainly caused by creating the server image data with Azure. Therefore, we are planning to keep VM image itself somewhere and to use this.

This engine’s Future Work
• Support for AWS and GCP
• Use distribution of multiple machines during analysis

Our Future Work
We are thinking about developing a web platform with this RyuGoo-Saba as a component. By collecting and analyzing job metrics run on a standardized cloud server, we would like to create this platform which enables users to receive recommendations when creating workflows in the GUI. In addition, we would like to create an environment that makes it easy to share workflows created by users.