Contribution ID: **26** Type: **Oral Presentation**

A Scientific Paper Reproducible Environment with Overlay Cloud Architecture

Thursday, 17 March 2016 14:00 (30 minutes)

The Inter-Cloud is a promising approach for the distributed application demands in some of HPC applications, like Next-Generation-Sequencing Data analytic. However, building the Inter-Cloud environments requires IT expert knowledge. This paper introduces an architecture called Overlay Cloud and Virtual Cloud Provider (VCP), which is a middle-ware to automatically build a set of virtual resources on the Inter-Cloud and ease the knowledge requirements. That aims to help to realize ubiquitous scientific paper reproducible environments.

In a data-centric science fields such as bioinformatics, demands for reproducibility to ensure experiments in papers is strong. For example, we share genomic analysis programs as open source software and have the public databases of the DNA sequences in the bioinformatics field. However, the following issues related to program execution environments still exist.
1. Data processing software is complex and diverse.
2. Massive data from many data sources, such as the next generation sequencers.
3. Amount of data analysis processing is increasing.

In my study, I utilize Overlay Cloud architecture in order to solve these problems. Overlay Cloud is an architecture that is Overlay a container environment over existing cloud environments (private / community / public) and overlay clouds do not depend on cloud environments in order to separate the interface between the application user and the cloud infrastructure. Therefore, the user can freely select the container execution platform. Paper readers can simply press a button which is linked to the paper in order to obtain the paper reproduction verification environment on their selected cloud.

As a prototype, I constructed a bioinformatics workflow reproduction environment on inter-cloud in Overlay Cloud architecture. This is aiming to show that the paper reproduction environment has following properties.
1. Data analysis software portability across the clouds by two level containerization.
2. Network delay reduction between the data and the data analysis program by container distribution.
3. Processing performance improvement of data analysis by distributed processing infrastructure to deploy containers.
In this paper, we report in particular a result of our efforts to ensure portability, which is the first problem.

Summary

The Inter-Cloud is a promising approach for the distributed application demands in some of HPC applications, like Next-Generation-Sequencing Data analytics. However, building the Inter-Cloud environments requires IT expert knowledge. This paper presents an architecture called Overlay Cloud and Virtual Cloud Provider (VCP), which is a middle-ware to automatically build a set of virtual resources on the Inter-Cloud and ease the knowledge requirements. That aims to help to realize ubiquitous scientific paper reproducible environments.

Primary author: Prof. YOKOYAMA, Shigetoshi (National Institute of Inofrmatics)

Presenter: Prof. YOKOYAMA, Shigetoshi (National Institute of Inofrmatics)

Session Classification: VRE Session

Track Classification: Virtual Research Environment (including Middleware, tools, services, workflow, … etc.)