A Scientific Paper Reproducible Environment with **Overlay Cloud Architecture**

Shigetoshi Yokoyama*, Yoshinobu Masatani*, Tazro Ohta[†], Osamu Ogasawara[‡], Nobukazu Yoshioka^{*}. Kento Aida^{*}

*National Institute of Informatics, †Database Center for Life Science, **±National Institute of Genetics**

Back ground

Reproducibility of Scientific Activities

Data and Infrastructure are ready to Reproduce

Data Centric Science \Rightarrow Reproducibility of IT environments in Scientific Activities



Issues in Reproducibility of Scientific Activities

Data don't move and Processes don't, either.

- 1) Huge Scientific Data (Distributed)
- 2) Huge Scientific Computation
- 3) Complex Scientific Computation Software



Cloud Federation Approaches

Cloud Standardization is not the only way to go.



On-demand Cloud: Academic Inter-Cloud

(academic community cloud hub)



On-demand cloud approach works but ...

Need to Federate beyond Community Clouds.

 On-demand cloud approach works fine in our academic community cloud environment with a limitation.



Public clouds do not always support bare metals.

Overlay Cloud

Container Revolutions in Two Industries

1950s -









?

Transport Industries	Contents	Containers	Container Management System	International Container Management System
IT Industries	Applications	Virtual Macines	Cloud	Inter-cloud

2000s -







Separation of Concerns in Transport Industries



Separation of Concerns



Separation of Concerns in Transport Industries



Separation of Concerns



Separation of Concerns







Container Revolution in Transportation Industries





Container Revolution in IT Industries

Containers wrap software Applications Inside Container Separation of Concern Outside Container Cluster Cluster Cluster Flynn **Overlay Cloud** L2Tunnels L2Tunnels L2Tunnels • **Cloud User RM** BM VM VM VM VM VM VM **Separation of Concern** Real Real Real Real **Cloud Provider Cloud Provider Cloud Provider Cloud Provider Cloud Provider** Gunnii DigitalOcean amazon webservices Container Engine, BM: Bare-metal Machine, VM: Virtual Machine

"Container" = Container

Overlay Cloud Approach is the way to go.



Virtual Cloud Provider

Virtual Cloud Provider (VCP)

VCP is the Middleware for Overlay Cloud.

- Virtual Cloud Provider
 - Functions (Deploy/Manage/Monitor)
 - Registry (Container images/Cluster Templates)



Base Containers and Application Containers

VCP manages Two level Containers.

- Base container contains virtual cloud platform
- Application container contains application software



VCP Prototype

Quick Prototype with Open Source Products.





Case Study of Bioinformatics Application

Genome Sequencing

 application of the inter-cloud system to genomic sequence annotation and analysis workflows



Genome Sequencing On Overlay Cloud



NII National Institute of Informatics

Workflow Example (RNA-Seq)



Conclusion

Overlay Cloud Approach Works.

- Proposed overlay cloud architecture.
- Developed middleware,
 VCP (<u>Virtual Cloud Provider</u>).
- Case study shows VCP enables the user to deploy computing environment that produces <u>reproducible</u> results.

History: The Road to Cloud Computing



Loosely Connected Independently Governed Domains Work Together

Recent: New Wave is Coming

- The 3rd Wave is coming ...
 (1st Wave) Public Cloud (AWS, ...)
 (2nd Wave) Open Cloud (OpenStack, ...)
 (3rd Wave) Containerization (Docker, ...)
- They work together and make inter-cloud real.



Vision

Inter-Cloud Everywhere ...

- Third Wave of Cloud Revolution is coming…
 - (1st Wave) Public Clouds (AWS, \cdots)
 - (2nd Wave) Open Clouds (OpenStack, …)
 - (3rd Wave) Containers on Clouds (Docker, …)
- They co-exist and realize inter-cloud everywhere



