Desktop as a Service supporting Environmental 'omics

David C.H. Wallom*, Andy Bowery, Ben Collier,
Dawn Field
Oxford e-Research Centre, University of Oxford, UK

Phillip Kershaw
Centre for Environmental Data Analysis, STFC, UK

Timothy Booth*
Centre for Ecology and Hydrology, Wallingford, UK

Anurag Priyam, Yannick Wurm
,School of Biological and Chemical Sciences, Queen
Mary University of London, UK

Abstract— Within the Environmental 'omics community Bio-Linux is a widely used tool. This has the advantage of providing in a single deliverable package all necessary software and tools to support common analyses. With the growth in data volumes within the community and increasing constraints on user access and control over their own desktops an alternative delivery method of Bio-Linux and, in future, the Docker container environment is necessary.

Within the EOS Cloud project we have constructed a Desktop as a Service system to centrally host virtual machines with these tools preconfigured and maintained. To enable efficient use of the resources we have enabled user controlled resource scaling so that users are able to utilise small scale VMs for task configuration and data manipulation and boost to a larger scale to run analysis applications all the while maintaining the user environment in a consistent manner.

Alongside this within the project we have been developed tools to simplify the increasingly popular Docker software usage model. This includes ensure uniformity of behaviour between the host system and the running Docker container. Within the invitation only trial user community we identify two different exemplars groups and explain their usage and how the products and services developed within the project are useful for them. We conclude discussing the useful nature of Desktop as a Service, how it is of great benefit to the bioinformatics community but could also be of great use elsewhere, where the need for a stable user environment with applications already available that do not rely on local ICT support.

Keywords—Cloud Computing; e-infrastructure; Cloud, Bioinformatics, Desktop as a Service, Containers, Scalability

PRESENTATION

The annotated presentation is available from the University of Oxford Research Archive from;

http://ora.ox.ac.uk/objects/uuid:d7ee0184-1971-4eca-be44-ce7a46ac3020

REFERENCES

1. Field, D., Tiwari, B., Booth, T., Houten, S., Swan, D., Bertrand, N., & Thurston, M. (2006). Open software for biologists: from famine to feast. *Nature biotechnology*, *24*(7), 801-803.
2. Möller, S., Krabbenhöft, H. N., Tille, A., Paleino, D., Williams, A., Wolstencroft, K., ... & Plessy, C. (2010). Community-driven computational biology with Debian Linux. *BMC bioinformatics*, *11*(Suppl 12), S5.
3. Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., ... & Nekrutenko, A. (2005). Galaxy: a platform for interactive large-scale genome analysis. *Genome research*, *15*(10), 1451-1455.
4. Open Bioinformatics Foundation (OBF, www.open-bio.org)
5. Krampis, K., Booth, T., Chapman, B., Tiwari, B., Bicak, M., Field, D., & Nelson, K. E. (2012). Cloud BioLinux: pre-configured and on-demand bioinformatics computing for the genomics community. *BMC bioinformatics*, *13*(1), 42.
6. Amazon, E. C. (2010). Amazon elastic compute cloud (Amazon EC2). *Amazon Elastic Compute Cloud (Amazon EC2)*.
7. Watson, J. (2008). Virtualbox: bits and bytes masquerading as machines. *Linux Journal*, *2008*(166), 1.
8. I. Parallels, "An introduction to os virtualization and parallels virtuozzo containers," Parallels, Inc, Tech. Rep., 2010. [Online]. Available: http://www.parallels.com/r/pdf/wp/pvc/Parallels Virtuozzo-Containers-WP-an-introduction-to-os-EN.pdf
9. Rosenblum, M. (1999, August). VMware's virtual platform™. In *Proceedings of hot chips* (Vol. 1999, pp. 185-196).
10. Distributed Management Task Force, Inc., Open Virtualization Format Specification (version 1.0.0), DSP0243, 2009
11. Docker container hosting environment; http://docker.com
12. Wurm, Y. Avoid having to retract your genomics analysis (2015). The Winnower. https://thewinnower.com/papers/avoid-having-to-retract-your-genomics-analysis
13. Lawrence, B. N., V. L. Bennett, J. Churchill, M. Juckes, Philip Kershaw, Stephen Pascoe, Sam Pepler, M. Pritchard, and Adrian Stephens. "Storing and manipulating environmental big data with JASMIN." In *Big Data, 2013 IEEE International Conference on*, pp. 68-75. IEEE, 2013.
14. Puppet, Puppet Labs; https://puppetlabs.com/
15. VMware vCloud, A. P. I. Programming Guide.
16. Gabriel, Mike (August 17, 2013). *X2Go Terminal Server Suite and Debian* (OGV). Debian Conference 2013. Vaumarcus, Switzerland. Retrieved June 14, 2014.