Geoprocessing in Hybrid Clouds

Theodor Foerster, Bastian Schaeffer, Bastian Baranski & Kristof Lange
{theodor.foerster;schaeffer;bastian.baranski;kristof.lange}@uni-muenster.de
Institute for Geoinformatics, University of Muenster, Germany

Geoprocessing on the web is considered to be the next evolution step in Spatial Data Infrastructures (Kiehle et al., 2006). Increasing network bandwidth and processing capacities along with common communication protocols for Web Services such as established by ISO, OASIS and OGC have paved the way (Kradilis, 2007). On this basis comprehensive architectures for distributed geoprocessing has been demonstrated as i.e. by Schaeffer et al. (2009). However, following the Model as a Service paradigm (Dumitru et al., 2009) by exposing complex model on the web via standardized interfaces implies two general problems in production environments:

- the complexity of geoprocessing models requires large computing capacities
- the number of user cause high computing loads even for simple geoprocessing models.

Cloud Computing (Foster et al., 2008) addresses these two problems. Different critical aspects concerning the performance are addressed such as the availability of sufficient computational resources to solve given computational problems in a satisfying manner. Performing remote functionality in a cloud infrastructure means that the service (cloud) consumer is able to allocate as much resources as required (e.g. sufficient disk storage, network bandwidth or amount of CPUs) and therefore can expect a process to be performed according to specific Quality of Service (QoS) parameters (e.g. to be finished in a specific time period). The allocation of sufficient hardware resources in advance can either be done manually by the cloud consumer or realized through defined set of rules according to which the cloud infrastructure has to allocate automatically additional computational resources (e.g. in case of high request rates of a web service).

Such a Cloud infrastructure can be hosted either by the organization itself (Private Cloud) or by an external provider (Public Cloud) (Armbrust et al., 2009). The Private Cloud approach allows organizations to use their existing hardware infrastructure more efficiently through a combination of well established mechanisms like Virtualization and Grid Computing. In case of a Public Cloud, operational functionality is outsourced and realized on the hardware infrastructure of the external provider for additional usage costs. From an organizational perspective using a Private Cloud is applicable to scale the process effort on the organization's own infrastructure. This however might not always meet the required QoS parameters due to a lack of sufficient computational resources in the organization's own infrastructure. Therefore, to meet specific QoS parameters in peak times (e.g. the availability of a web service as required by INSPIRE) and to generally optimize operational costs (e.g. the hardware utilization rate), sometimes organizations must dynamically outsource (parts of) their business processes and operational functionality to Public Clouds (the so-called Hybrid Cloud approach).

In the proposed paper we will describe an abstract architecture of a Hybrid Cloud for geoprocessing based on the Web Processing Service (WPS) interface specification (OGC, 2007). Furthermore we will present a prototypical implementation of the proposed architecture (Figure 1). The implementation is based on the 52°North Web Processing Service (52n WPS) that utilizes the Open Nebula software package for building a Cloud Computing environment (Private Cloud). The implementation is demonstrated in a risk assessment use case in which critical processing steps during the analyzation of satellite data are dynamically outsourced to the Amazon EC2 service platform (Public Cloud). Finally, we will summarize our experiences in implementing the Hybrid Cloud approach and evaluate the existing gaps and bottlenecks in utilizing Cloud Computing in the geospatial domain.

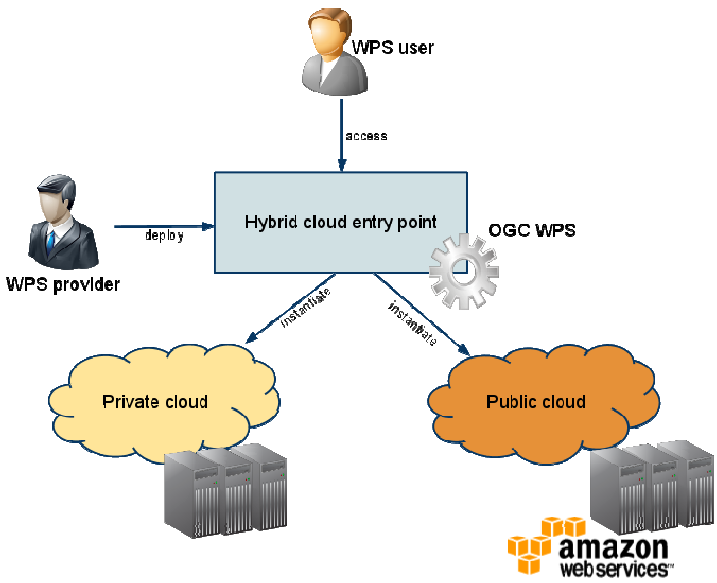

Figure 1: Architecture of the Hybrid Cloud using a service providers own infrastructure (Private Cloud) and the infrastructure of external providers such as Amazon (Public Cloud).

REFERENCES

Armbrust, M., Fox, A., Griffith, R., Joseph, Anthony D., Katz, Randy H., Konwinski, A., Lee, G., Patterson, D. A., Rabkin, A., Stoica, I. & Zaharia, M. (2009). Above the Clouds: A Berkeley View of Cloud Computing. Technical Report. EECS Department, University of California, Berkeley.

Foster, I., Zhao, Y., Raicu, I. & Lu, S. (2008). Cloud computing and grid computing 360-degree compared. [Online]. Available: http://arxiv.org/abs/0901.0131.

Kiehle, C., Greve, K., & Heier, C. (2006). Standardized Geoprocessing - taking spatial data infrastructures one step further. In *9th AGILE International Conference on Geographic Information Science* (pp. 273-282). Visegrad, Hungary.

Kralidis, A. T. (2007). Geospatial Web Services: The Evolution of Geospatial Data Infrastructure. In A. Scharl & K. Tochtermann (Eds.), *The Geospatial Web*, Advanced Infromation and Knowledge Processing Series (pp. 223-228). London, UK: Springer.

OGC. (2007). *OpenGIS Web Processing Service*. OGC implementation specification, Open Geospatial Consortium. Retrieved from http://www.opengeospatial.org/standards/wps.

Roman, D., Schade, S., Berre, A. J. , Bodsberg, N. R. & Langlois, J. (2009). Model as a Service (MaaS).Proceedings of Grid Technologies for Geospatial Applications - pre-conference workshop AGILE 2009. Hannover, Germany.

Schaeffer, B., Baranski, B., Foerster, T., & Brauner, J. (2009). A Service-Oriented Framework for Real-time and Distributed Geoprocessing. In *International Opensource Geospatial Research Symposium*. Presented at the OGRS 2009, Nantes, France: Springer Verlag.