A case study of advancing remotely sensed image processing

Roberto Giachetta and István Fekete

Today's information systems present virtually unlimited opportunities for data analysis due to advances in algorithmic capabilities, evolution of parallel and distributed processing, leading to the increasing popularity of cloud computing. Due to data sets becoming increasingly large and complex (usually noted as *Big data*), these new paradigms cannot be ignored. In recent years many systems have been developed for data processing in the cloud, most notably the *MapReduce* model [1], and its open-source implementation, the *Apache Hadoop* framework [2]. However to take advantage of today's state of the art computing, previous data processing methodologies and workflows have to revisited and redesigned.

In case of geographical information systems (GIS) and remotely sensed image analysis, the new paradigms have already been successfully applied to several problems [3], and systems have been developed to support processing of geographical or remotely sensed data in the cloud (e.g. *Hadoop-GIS* [4]). However in the case of complex processing operations the paradigm shift involves several considerations affecting both design and implementation. In contrary, most spatial data analysis processes performed at organizations such as the *Institute of Geodesy, Cartography and Remote Sensing (FÖMI)* have their evolved workflows using multiple (proprietary and open-source) software and GIS expertise.

For example, in the case of *waterlogging and flood detection*, the current practice involves a workflow with multiple steps with differrent software and including several parameters usually specified manually by a remote sensing expert. This is primarily due to the nature of waterlogging phenomena depending on the environment. For this problem, some specific algorithms have been introduced with environmental circumstances in mind [5, 6]. The process also involves multiple datasets including high-resolution remotely sensed imagery and digital elevation model (DEM). Thus the size of input is usually too large to be handled by a single machine, making the process worthwhile to be moved into cloud architecture. For this purpose, not only the distributed execution, but also complete automation is required. Hence, the advancement of the workflow to the state of the art requires both architectural, and algorithmic considerations.

In our paper, waterlogging and flood detection serves as a case study of shifting remotely sensed image processing from its current, semi-automatic, multi software environment to an integrated and fully automated distributed system based on the Hadoop framework.

The algorithm itself is enhanced by including segmentation based image analysis techniques based on reference data with auto tuning [7]. This approach enables the exclusion of intervention by the expert during the process, as the parameters are automatically tuned towards best result (with respect to reference data). The application of segmentation is not only an option, but a necessity in the processing of very high resolution images, as their pixels usually cannot be interpreted individually. Using reference data and automatic parameter calibration, the manual steps of the process can be eliminated, leading to a fully automated workflow.

To enable the distributed execution of the process, we rely on the proven Hadoop implementation, but extend its functionalities to enable a more flexible and controllable execution of the workflow and a more advanced management of resources. Instead of the direct approach to implement a MapReduce versions of the used algorithms, the distribution is accomplished by providing an environment, in which the process is automatically transformed to the MapReduce paradigm. For this purpose, we extend the capabilities of the *AEGIS spatio-temporal framework* [8], as it contains a generally defined, flexible processing system enabling operations to be executed in a variety of environments without any modification. This approach enables other operations to be executed in the Hadoop environment, not just our case study. However this technique also comes with the challenge of enhancing the data management possibilities of

Hadoop distributed file system, as the basic data splitting methodology does not allow execution of complex operations (such as segmentation and clustering) on the distributed data in a uniform manner.

In conclusion our approach advances the process of waterlogging and flood detection in both automation and efficiency using state of the art technology. It enables the replacement of multiple software to a single, generic framework, and it also paves the way for applying the paradigm shift to other GIS workflows.

References

[1] Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. *Communications of the ACM*, 51 (1), 107-113, 2008.

[2] Agrawal, D., Das, S., Abbadi, A. E.: Big data and cloud computing: current state and future opportunities. *Proceedings of the 14th International Conference on Extending Database Technology (EDBT/ICDT '11)*, 530-533, 2011.

[3] Almeer, M. H.: Hadoop Mapreduce for Remote Sensing Image Analysis. *International Journal of Emerging Technology and Advanced Engineering*, 2 (4), 443-451, 2012.

[4] Aji, A. et. al.: HadoopGIS: A High Performance Spatial Data Warehousing System over MapReduce. *Proceedings of the VLDB Endowment*, 6 (11), 1009-1020, 2013.

[5] Sreenivas, K., Dwivedi, R. S., Singh, A. N., Raviprakash, S.: Detection of sub-surface waterlogging using Terra-1 MODIS data. *Journal of the Indian Society of Remote Sensing*, 38 (1), 119-132, 2010.

[6] El Bastawesy, M., Ali, R. R., Deocampo, D. M., Al Baroudi, M. S.: Detection and Assessment of the Waterlogging in the Dryland Drainage Basins Using Remote Sensing and GIS Techniques. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, 5 (5), 1564-1571, 2012.

[7] Dezső, B., Fekete, I., Gera, D., Giachetta, R., László, I.: Object-based image analysis in remote sensing applications using various segmentation techniques. *Annales Universitatis Scientitarum Budapest, Sectio Computatorica*, 37 (1), 103-120, 2012.

[8] Giachetta, R.: AEGIS - A state-of-the-art spatio-temporal framework for education and research. *OSGeo Journal*, 13, 68-77, 2014.