Geophysical Research Abstracts
Vol. 18, EGU2016-16953, 2016
EGU General Assembly 2016

Massive parallel processing of geospatial data with Python and R on the Intel Xeon Phi platform – Two case studies using Free and Open Source Software

Witold Arndt, Tillmann K. Buttschardt, and André Große-Stoltenberg
Institute of Landscape Ecology, University of Muenster, Germany (witold.arndt@uni-muenster.de)

Geosciences have always been a discipline where very large datasets and complex models were processed. However, many of its fields of research started only recently to employ the possibilities of massively parallel computing. An important factor is the prevalence of unoptimised standard software which is utilising only one CPU in the established workflows although even present desktop computers provide 2-8 CPUs.
Using high-level programming or scripting languages is an established alternative to benefit from the computing power of commercially available hardware.

General-purpose computing on graphics processing units (GPGPU) provided the access to parallel processing without the technical and financial complexities of traditional super computing. However this approach requires the scientist to write program code for specific APIs (CUDA, OpenCL) which were not common in most fields within the geosciences.

Using manycore x86/x64 based designs like the Intel Xeon Phi [1] allows the research to employ already commonly used tools like Python with its NumPy [2] and SciPy [3] extensions and libraries and the R programming language for statistical computing within a Linux environment.

In our first case we use pyMIC [4] to offload processing and analysing RapidEye 3A ortho product [5] remote sensing data of central Portugal. Atmospheric correction and object-based classification is done with RSGISLib [6]. We will compare processing time and optimisation effort for this approach.

Hyperspectral Vegetation Indices can be applied to assess biophysical parameters and to discriminate types of vegetation [7]. In our second case study we will calculate Hyperspectral Vegetation Indices (HVIs) based on airborne hyperspectral data from SW Portugal (EUFAR TA project "DeInVader", EUFAR11-06) using the hsdar R package [8] Intel Xeon Phi Automatic Offload, with Intel Math Kernel Library [9]. We will assess the computation efficiency of our manycore approach to conventional parallelisation.

References

[1]Intel Corporation, Intel Xeon Phi System Software Developer's Guide. Intel Corporation, 2012.
[2]S. van der Walt, S. C. Colbert, and G. Varoquaux, 'The NumPy Array: A Structure for Efficient Numerical Computation', Computing in Science & Engineering, vol. 13, no. 2, pp. 22–30, Mar. 2011.
[3]T. E. Oliphant, 'Python for Scientific Computing', Computing in Science & Engineering, vol. 9, no. 3, pp. 10–20, 2007.
[4]M. Klemm and J. Enkovaara, 'pyMIC: A Python Offload Module for the Intel Xeon Phi Coprocessor', presented at the 4th Workshop on Python for High Performance and Scientific Computing (PyHPC 2014), New Orleans, 2014.
[5]Blackbridge, Satellite Imagery Product Specifications, Version 6.1. 2015.
[6]P. Bunting, D. Clewley, R. M. Lucas, and S. Gillingham, 'The Remote Sensing and GIS Software Library (RSGISLib)', Computers & Geosciences, vol. 62, pp. 216–226, Jan. 2014.
[7]P. S. Thenkabail, I. Mariotto, M. K. Gumma, E. M. Middleton, D. R. Landis, and K. F. Huemmrich, 'Selection of Hyperspectral Narrowbands (HNBs) and Composition of Hyperspectral Twoband Vegetation Indices (HVIs) for Biophysical Characterization and Discrimination of Crop Types Using Field Reflectance and Hyperion/EO-1 Data', IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 6, no. 2, pp. 427–439, Apr. 2013.

[8]L. W. Lehnert, H. Meyer, and J. Bendix, ‘hsdar: Manage, analyse and simulate hyperspectral data in R’, R package version 0.3.1., 2014.
[9]Intel Corporation, Intel Math Kernel Library Reference Manual, Revision 004. 2015.

Acknowledgements
Deutsche Forschungsgemeinschaft DFG (TH 1295/4-1)
European Facility for Airborne Research EUFAR (DeInVader, EUFAR11-06)
German Academic Exchange Service DAAD (A. Große-Stoltenberg)
Rapideye (RESA ID 00169)