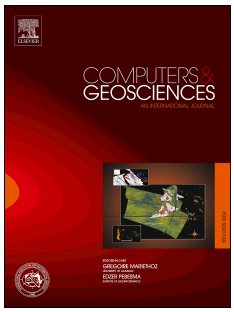
Accepted Manuscript



Arc4nix: A cross-platform geospatial analytical library for cluster and cloud computing

Jingyin Tang, Corene J. Matyas

PII: S0098-3004(17)30107-3

DOI: 10.1016/j.cageo.2017.11.011

Reference: CAGEO 4052

To appear in: *Computers and Geosciences*

Received Date: 29 January 2017

Revised Date: 7 November 2017

Accepted Date: 11 November 2017

Please cite this article as: Tang, J., Matyas, C.J., Arc4nix: A cross-platform geospatial analytical library for cluster and cloud computing, *Computers and Geosciences* (2017), doi: 10.1016/j.cageo.2017.11.011.

Arc4nix: A Cross-platform Geospatial Analytical Library for Cluster and Cloud Computing

Jingyin Tang[a,*], Corene J. Matyas[a]

[a] *3141 Turlington Hall, Gainesville, Florida, USA*

Abstract

Big Data in geospatial technology is a grand challenge for processing capacity. The ability to use a GIS for geospatial analysis on Cloud Computing and High Performance Computing (HPC) clusters has emerged as a new approach to provide feasible solutions. However, users lack the ability to migrate existing research tools to a Cloud Computing or HPC-based environment because of the incompatibility of the market-dominating ArcGIS software stack and Linux operating system. This manuscript details a cross-platform geospatial library "arc4nix" to bridge this gap. Arc4nix provides an application programming interface compatible with ArcGIS and its Python library "arcpy". Arc4nix uses a decoupled client-server architecture that permits geospatial analytical functions to run on the remote server and other functions to run on the native Python environment. It uses functional programming and meta-programming language to dynamically construct Python codes containing actual geospatial calculations, send them to a server and retrieve results. Arc4nix allows users to employ their arcpy-based script in a Cloud Computing and HPC environment with minimal or no modification. It also supports parallelizing tasks using multiple CPU cores and nodes for large-scale analyses. A case study of geospatial processing of a numerical weather model's output shows that arcpy scales linearly in a distributed environment. Arc4nix is open-source software.

Keywords: Spatial Analytics, Geographic Information System, Cloud

*Corresponding author. Email Address: jtang8756@ufl.edu