

## [A high performance parallel computing framework to support big climate data analysis](#) [1]

Submitted by Fei Hu on Mon, 2015-12-21 12:51 **Event:** [Winter Meeting 2016](#) [2]

### **Abstract:**

Big earth science data (e.g., over 100 PB climate data) are being created by climate observation and model simulations to support Earth sciences. Ideally, such big data can be provided to scientists with on-demand analytical and simulation capabilities to relieve them from time-consuming computational tasks. However, it is challenging to realize it because processing such big data requires efficient big data management strategies, complex parallel computing algorithms and scalable computing resources. Based on the spatiotemporal index for array-based data, big climate data analytics, Spark, and HDFS, we develop a high performance computing framework to efficiently utilize big climate data to analyze climate models. This paper will be a pilot study on how to support on-demand big Earth data analytics in real time for climatologists.

**Collaboration Area:** [Cloud Computing](#) [3]

**Reference:** Li Z., Hu F., Schnase J., Duffy D., Lee T., Yang C., Bowen M. (2016), A Spatiotemporal Indexing Approach for Efficient Process of Big Array-based Climate Data with MapReduce, International Journal of Geographic Information Science (In press)

Buck, J. B., Watkins, N., LeFevre, J., et al., 2011. SciHadoop: Array-based query processing in hadoop. In Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis, ACM, p. 66

### **Author(s):**

**Name:** [Fei Hu](#) [4]

**Organization(s):** [George Mason University](#) [5]

**Email:** [fhu@gmu.edu](mailto:fhu@gmu.edu) [6]

**Name:** [chaowei yang](#) [7]

**Organization(s):** [GMU](#) [8]

**Email:** [cyang3@gmu.edu](mailto:cyang3@gmu.edu) [9]

**Name:** [John Schnase](#) [10]

**Organization(s):** [Goddard Space Flight Center](#) [11], [NASA](#) [12]

**Email:** [John.L.Schnase@nasa.gov](mailto:John.L.Schnase@nasa.gov) [13]

**Name:** [Daniel O. Duffy](#) [14]

**Organization(s):** [Goddard Space Flight Center](#), [NASA](#) [15]

**Email:** [daniel.q.duffy@nasa.gov](mailto:daniel.q.duffy@nasa.gov) [16]

**Keywords:** [Big Data](#) [17]

[high performance distributed computing](#) [18]

[Climate](#) [19]

**Source URL:** <http://commons.esipfed.org/node/8922>

### **Links:**

[1] <http://commons.esipfed.org/node/8922>

[2] <http://commons.esipfed.org/2016WinterMeeting>

[3] <http://commons.esipfed.org/collaboration-area/cloud-computing>

[4] <http://commons.esipfed.org/node/8920>

[5] <http://commons.esipfed.org/taxonomy/term/836>

[6] <mailto:fhu@gmu.edu>

[7] <http://commons.esipfed.org/node/1145>

- [8] <http://commons.esipfed.org/taxonomy/term/213>
- [9] <mailto:cyang3@gmu.edu>
- [10] <http://commons.esipfed.org/node/1352>
- [11] <http://commons.esipfed.org/taxonomy/term/727>
- [12] <http://commons.esipfed.org/taxonomy/term/228>
- [13] <mailto:John.L.Schnase@nasa.gov>
- [14] <http://commons.esipfed.org/node/8921>
- [15] <http://commons.esipfed.org/taxonomy/term/726>
- [16] <mailto:daniel.q.duffy@nasa.gov>
- [17] <http://commons.esipfed.org/taxonomy/term/348>
- [18] <http://commons.esipfed.org/taxonomy/term/2065>
- [19] <http://commons.esipfed.org/taxonomy/term/448>