

**Your abstract submission has been submitted for the 2014 AGU Fall Meeting. You will receive an email confirmation.**

Click [HERE](#) to print this page now.

**Receipt of this notice does not guarantee that your submission was accepted for the 2014 AGU Fall Meeting. All submissions are subject to review and acceptance by the Program Committee. You may review or edit your abstract submission until the deadline of 7 August 2014 23:59 EDT/03:59 +1 GMT. After this date, no further edits will be made to the submission.**

---

### **Integrating the Apache Big Data Stack with HPC for Big Data**

---

**Geoffrey C Fox**<sup>1</sup>, Judy Qiu<sup>1</sup> and Shantenu Jha<sup>2</sup>, (1)Indiana University Bloomington, School of Informatics and Computing, Bloomington, IN, United States, (2)Rutgers University Newark, Department of Electrical and Computer Engineering, Newark, NJ, United States

#### **Abstract Text:**

There is perhaps a broad consensus as to important issues in practical parallel computing as applied to large scale simulations; this is reflected in supercomputer architectures, algorithms, libraries, languages, compilers and best practice for application development. However, the same is not so true for data intensive computing, even though commercially clouds devote much more resources to data analytics than supercomputers devote to simulations.

We look at a sample of over 50 big data applications to identify characteristics of data intensive applications and to deduce needed runtime and architectures. We suggest a big data version of the famous Berkeley dwarfs and NAS parallel benchmarks and use these to identify a few key classes of hardware/software architectures. Our analysis builds on combining HPC and ABDS the Apache big data software stack that is well used in modern cloud computing. Initial results on clouds and HPC systems are encouraging.

We propose the development of SPIDAL - Scalable Parallel Interoperable Data Analytics Library -- built on system and data abstractions suggested by the HPC-ABDS architecture. We discuss how it can be used in several application areas including Polar Science.

**Session Selection:** Leveraging Enabling Technologies and Architectures to enable Data Intensive Science

**Title:** Integrating the Apache Big Data Stack with HPC for Big Data

**Preferred Presentation Format:** Assigned by Program Committee (Oral or Poster)

**Invited** 1

First Presenting Author

---

***Presenting Author***

---

Geoffrey C Fox

**Primary Email:** gcf@indiana.edu

**Affiliation(s):**

Indiana University Bloomington  
School of Informatics and Computing  
Bloomington IN (United States)

Second Author

---

Judy Qiu

**Primary Email:** xqiu@cs.indiana.edu

**Affiliation(s):**

Indiana University Bloomington  
School of Informatics and Computing  
Bloomington IN (United States)

Third Author

---

Shantenu Jha

**Primary Email:** shantenu.jha@rutgers.edu

**Affiliation(s):**

Rutgers University Newark  
Department of Electrical and Computer Engineering  
Newark NJ (United States)

**If necessary, you can make changes to your abstract submission**

- To access your submission in the future, point your browser to: [Full Menu Options](#).
- Your Abstract ID# is: 5905.
- Any changes that you make will be reflected instantly in what is seen by the reviewers.
- After the abstract proposal is submitted, you are not required to go through all submission steps to make edits. For example, click the "Authors" step in the Abstract Submission Control Panel to edit the Authors and then click save or submit.
- When you have completed your submission, you may close this browser window or submit another abstract <http://fallmeeting.agu.org/2014/Sessionviewer>.

[Tell us what you think of the abstract submission process](#)