Submission Details: bof139s1
Form first submitted: 2019-02-20 19:07
Form last updated: 2019-02-20 19:07

BoF Title

BoF Title: Defining HDF5 Future

BoF Organizer/Speaker Information

Bot Organizer/Speaker 1:
Salutation: Ms.
Name: Elena Pourmal
Email: epourmal@hdfgroup.org
Company/Institution: The HDF Group
2nd Company/Institution:
Country: United States of America
Job Title: Director of Engineering
Biography:
Ms. Elena Pourmal is one of the founders of The HDF Group, a not-for-profit company with the mission to develop and sustain the HDF technology, and to provide free and open access to data stored in HDF. Ms. Pourmal has been with The HDF Group since 1997 and for 20+ years led HDF software maintenance, quality assurance and user support efforts. Ms. Pourmal currently serves as The HDF Group Engineering Director leading HDF5 engineering effort and is also a member of The HDF Group Board of Directors. Ms. Pourmal received her MS in Mathematics from Moscow State University and MS in Theoretical and Applied Mechanics from University of Illinois at Urbana-Champaign. Her interests include management of software development projects, HPC, and evangelizing HDF5.

Photograph:

Will this person present this BoF at ISC? Yes
Is corresponding author? Yes
Is this person on the Birds of a Feather reviewing committee? No

Bof Organizer/Speaker 2:
Salutation: Mr.
Name: Steven Varga
Email: steven.varga@gmail.com
Company/Institution: The HDF Group
2nd Company/Institution:
Country: United States of America
Job Title: Software Developer
Biography:
Steven Varga is an independent researcher in machine learning and computational finance, providing convex approximations for combinatorial problems, modeling sequential, categorical data, and writing software for high performance computing in C++, Julia, Python and R. Steven works closely with The HDF Group developers in HDF5 C++ and performance areas.

Photograph:

Will this person present this BoF at ISC? Yes
Is corresponding author? No
Is this person on the Birds of a Feather reviewing committee? No
Abstract

Abstract (Maximum 250 words):

HDF5 is a unique, open-source, high-performance technology suite that consists of an abstract data model, library, and file format used for storing and managing extremely large and/or complex data collections. The technology is used worldwide by government, industry, and academia in a wide range of science, engineering, and business disciplines.

We will provide a forum for the HDF5 user community to share ideas, present problems, outline possible solutions and discuss future direction of HDF5 and the HDF5 ecosystem. Elena Pourmal will present HDF5 features under development, the HDF5 roadmap, including upcoming releases and launch a discussion for HDF5 roadmap and community involvement in HDF5 development. Elena’s presentation and discussion will be followed by a discussion led by Steven Varga who is an active member and contributor of the HDF community. He will answer questions from the audience and survey the latest developments and challenges in the HDF5 ecosystem including (but not limited to): ● Storage devices and platforms for HDF5 (cloud, object stores, HDFS, PFS), VFDs and VOLs ● HDF5 language bindings (C++/LLVM, Python, Julia, R) ● HDF5 library development and community code contributions to HDF5 ● HDF5 Performance and productivity (benchmarks, profiling, Tau plugin) ● I/O concurrency (multithreading, MPI-I/O) ● Sparse data and HDF5 ● HDF5 connectors for Apache Spark and Drill ● HDF5 tools, including third-party tools and vendor support Steven and Elena will encourage HDF5 community members to share their experiences and challenges with HDF5. The summary of discussion will be shared with the HDF5 community.

Topic Area

Please select the topic area(s) your BoF belongs to:
Data, Storage & Visualization
HPC Applications
Programming Models & Systems Software
Artificial Intelligence and Machine Learning

Keywords

Keywords: Exascale Systems
HPC workflows
Parallel Applications
Storage Technologies

Fraction of time for presentations

How much of your BoF time will be taken up by "classic" presentations, excluding discussion? 30%

Description of concept for making the BoF interactive

Description of concept for making the BoF interactive (Maximum 150 words):

Two months before the event, we will publish a detailed outline of the areas that will be covered during the BoF on the HDF forum and in other channels. This will give prospective attendees and people who might be unable to attend a chance to propose additional topics and send or prepare detailed questions or problem descriptions. We will also contact and encourage members of the HDF5 community who will attend ISC 2019 and who have experience in the identified areas to work with us on responses and helping us with discussions. During the event, guest access will be available to all attendees to “kick the tires” of some of the new developments such as the HDF Kitla cloud offer and a LLVM Clang-based C++ source transformation productivity tool for HDF5 developers.
Targeted Audience

Targeted Audience (Maximum 100 words):
The intended audience includes anyone interested in data management problems and high-performance and parallel I/O on HPC systems and BigData platforms. While existing HDF5 users and application developers might get the most out of this event, it is also suitable for ● Students and scientists who want to learn about current data management and I/O performance solutions ● Small to medium-size research groups in AI and ML ● Vendors who have integrated HDF5 into their solutions ● Engineers and Data Scientists

Estimated Number of Attendees

How many attendees do you expect for this BoF session? Between 50 and 80

Presentation Slides for Attendees

I agree to the publication of the BoF presentation slides as part of the ISC 2019 proceedings: yes