

CS554 Project Ideas

FusionFS:DAS - Data-Aware Scheduling for Distributed File Systems

Overview

FusionFS [1] is a new distributed file system designed for exascale systems. In the current implementation, data movement is not optimized to increase data locality. For example, a job's data is written into the local storage and will be transferred to the requested node for reading, on the fly. While this strategy is optimal for some specific workload (e.g. checkpointing), a more general and intelligent mechanism is desired for achieving higher end-to-end data throughput and shorter wall time. In this project we will design and implement a scheduling algorithm, implement it in the Matrix [2] framework, and create a communication interface between Matrix and FusionFS in order to allow tasks to be scheduled close to the data. Work from 2009 [3] has insight into the centralized data-aware task scheduling system Falkon. This work will be part of the next major release of FusionFS.

Relevant Systems and Reading Material

Please refer to the following papers for the current status of this project:

[1] Dongfang Zhao, Chen Shou, Zhao Zhang, Iman Sadooghi, Xiaobing Zhou, Tonglin Li and Ioan Raicu. FusionFS: a distributed file system for large scale data-intensive computing, *2nd Greater Chicago Area System Research Workshop*, 2013. Available online: http://datasys.cs.iit.edu/~dongfang/download/poster_fusionfs_revised.pdf

[2] Anupam Rajendran. MATRIX: Many-task Computing Execution Fabric, *Master thesis, Illinois Institute of Technology*, 2013. Available online: http://datasys.cs.iit.edu/publications/2013_MS-IIT_MATRIX.pdf

[3] Ioan Raicu, Ian Foster, Yong Zhao, Philip Little, Christopher Moretti, Amitabh Chaudhary, Douglas Thain. "The Quest for Scalable Support of Data Intensive Workloads in Distributed Systems", *ACM HPDC* 2009. Available online: http://www.cs.iit.edu/~iraicu/research/publications/2009_HPDC09_data-diffusion.pdf

Preferred/Required Skills

Principles: operating system, distributed systems, computer network

Programming: Shell Script, Perl/Python, C, C++, PThread, sockets, FUSE

Operating systems: Linux

Project Mentor

Dongfang Zhao

Email: dzhao8@hawk.iit.edu