

CS554 Project Ideas

FusionFS:FFSZ – Exploring Data Compression in Distributed File Systems

Overview

Modern applications are generating large volume of data at an unprecedented rate. While state-of-the-art distributed file systems alleviate this pressure by splitting big data into chunks stored in multiple nodes, it remains unclear how to efficiently improve the data throughput for the chunk itself. Data compression has been actively studied to explore the feasibility of reducing the data size in the context of parallel processing systems [1]. In this project, you will design and implement a transparent data compression and decompression layer in the FusionFS [2] distributed filesystem, so that applications would read and write a smaller chunk of data, which will in turn deliver a higher throughput as long as the computation overhead introduced by (de)compression is relatively small. You will compare FusionFS I/O performance on various workloads with and without compression, and exploring a variety of compression algorithms.

Relevant Systems and Reading Material

Please read the following papers (and their references) before submitting your proposal:

[1] T. Bicer, J. Yin, D. Chiu, G. Agrawal and K. Schuchardt. Integrating Online Compression to Accelerate Large-Scale Data Analytics Applications, *IEEE International Parallel and Distributed Processing Symposium*, 2013. Available online: <http://www.cse.ohio-state.edu/~bicer/research/papers/ipdps2013.pdf>

[2] 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

Preferred/Required Skills

Principles: operating system, distributed systems, computer network, compression algorithms

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

Operating systems: Linux

Project Mentor

Dongfang Zhao

Email: dzhao8@hawk.iit.edu