CS554 Project Ideas

FusionFS:CKPT - Efficient Checkpointing with Distributed File Systems

Overview

Checkpointing is the de facto mechanism to achieve fault tolerance on large-scale distributed systems. This project is comprised of three parts: (1) a thorough literature review of state-of-the-art techniques in checkpointing; (2) design and implement a checkpointing mechanism, and integrate it to the FusionFS distributed file system [1]; (3) evaluate the checkpointing system with benchmarks (e.g. BLCR[2]). This work will lead (or partially contribute) to a conference and/or journal publication.

Relevant Systems and Reading Material

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

[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] Paul H. Hargrove and Jason C. Duell. Berkeley Lab Checkpoint/Restart (BLCR) for Linux Clusters, *Technical Report, Lawrence Berkeley National Laboratory*, 2006. Available online: http://crd.lbl.gov/assets/pubs_presos/CDS/FTG/Papers/2006/LBNL-60520.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