Wittawat Tantisiriroj

3 Bayard Road, Apt. 56
Pittsburgh, PA 15213
wtantisi@cs.cmu.edu
http://www.cs.cmu.edu/~wtantisi

Gates-Hillman Center 6219
School of Computer Science
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

Research Summary

Research interests: Cloud Computing: storage, databases, and optimization for computing in a massively distributed environment.

I am interested in storage systems for large scale computing, especially data-intensive computing systems such as Hadoop, HDFS, Cassandra, etc. I have been at Carnegie Mellon University for three years working with Professor Garth Gibson in the Parallel Data Lab in 3 projects.

Hadoop Distributed Files System (HDFS) & Parallel Virtual File System (PVFS): Data-Intensive File Systems Comparison

Cloud file systems, such as GFS and HDFS, are emerging as a key component in large scale computing systems that compute on massive amounts of data. In order to run these applications fast, computations are distributed over a large cluster. By exposing data layout, Cloud file systems enable Map-reduce and Hadoop to minimize the transfer of large amounts of data by shipping computation to nodes that store the data. Although it is commonly believed that high performance computing (HPC) systems use specialized infrastructure, that their parallel file systems are designed for vastly different data access patterns, and that they cannot support Internet services workloads efficiently, in fact, many HPC clusters use commodity compute, storage and network infrastructure. Moreover, parallel file systems have mature deployments and data managements, are cost effective, and have high performance. In this project I compared a parallel file system, developed for HPC, and a Cloud file system. Using PVFS as a representative for parallel file systems and HDFS as a representative for Cloud file systems, I configured a parallel file system into a distributed computing system, Hadoop, and tested performance with micro-benchmarks and macro-benchmarks running on a 4,000 core Internet services cluster, Yahoo!s M45. Once a number of configuration issues such as stripe unit sizes and application buffering sizes are dealt with, issues of replication, data layout and data-guided function shipping are found to be different, but supportable in parallel file systems. Performance of Hadoop applications storing data in an appropriately configured PVFS are comparable to those using a purpose built HDFS.

DiskReduce: RAID for Cloud file systems

Cloud file systems, such as GFS and HDFS, provide high reliability and availability by replicating data, typically three copies of each file while high performance computing file systems, such as Lustre, PVFS, and PanFS, achieve tolerance for the same numbers of concurrent disk failures using much lower overhead erasure encoding, or RAID schemes. In this project, I modify HDFS to include RAID6 without change to the HDFS client, reducing storage overhead in HDFS from 200% down to about 25%. My implementation writes three copies initially, using the existing HDFS client code, and asynchronously encodes data into RAID sets. Delaying encoding trades space for the performance optimizations possible when reading can be satisfied by any one of three nodes, and delaying encoding trades space for reducing the amount of work that is done during the encoding. Finally, triplication and RAID6 are both two failure tolerant, that is, no data is lost if only two disks are concurrently failed. But many more than two disks are likely to fail in large data-intensive clusters, so we analyze reliability in these systems more closely to better understand the impact of data loss resulting from lowering capacity overhead. An earlier version on this project has already stimulated Dhruba Borthakur, the Hadoop author at Facebook, to implement and release HDFS-RAID, a vari-

ant on these ideas.

Cloud Database

Cloud distributed databases systems, such as HBase, HyperTable, Cassandra and many others, provide a lightweight database system to manage structured data for cloud applications. Although they typically do not support ACID transactions, they can support a wide range of cloud applications. Given the number of different emerging Cloud database systems and the diverse range of Cloud applications, an apples-to-apples comparison is hard and it is difficult to understand tradeoffs between systems. In this project, I am creating a benchmark suit to represent a diverse range of applications including a real machine learning application code. The goal of this benchmark suits is to highlight a set of important workloads for different types of applications to help developers optimize their systems and help users choose a system that suits their workloads. With such a benchmark suits, we hope to identify areas for improving the Cloud database for future research.

Wittawat Tantisiriroj

3 Bayard Road, Apt. 56
Pittsburgh, PA 15213
412-999-4440
wtantisi@cs.cmu.edu
http://www.cs.cmu.edu/~wtantisi

Gates-Hillman Center 6219
School of Computer Science
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

Education

Ph.D., Computer Science, May 2012 (expected)
Carnegie Mellon University, Pittsburgh, PA
Advisor: Professor Garth Gibson

B.S., Computer Science and Computer Engineering, May 2007
University of Virginia, Charlottesville, VA

Professional Experience

Graduate Student **August 2007 – Present**
Carnegie Mellon University, Parallel Data Lab

- As a member of PDL, I have been conducted research on Data-Intensive Scalable Computing (DISC) and Distributed Storage Systems as explained above.

Technical Intern **June 2010 – August 2010**
Yahoo! – Sunnyvale

- I have implemented a version of DiskReduce RAID 6 for HDFS which lowers capacity overhead significantly. It is built as a tool and a library layered on top of and independent of HDFS. This tool can encode directories into RAID sets and repair corrupted files, and the library can detect and correct missing data while reading. This implementation is available via MAPREDUCE-2036 and is expected to be released with Hadoop 0.22.

Student Researcher **June 2005 – May 2007**
University of Virginia, Multimedia Networks Group

- Design and implement a new protocol for a multicast networking to achieve high performance over the heterogeneity of network devices

Relevant classes

- `18-741`: Advanced computer architecture
- `18-746`: Advanced storage systems
- `15-712`: Advanced operating systems & distributed systems
- `15-744`: Advanced computer networks
- `15-857`: Performance Modeling & Design of Computer Systems (queuing theory)

Special Skills

- Operating systems: Linux and Windows
- Programming languages: C, C++, Java, Perl, Python and Ruby

Publications

On the Duality of Data-intensive File System Design: Reconciling HDFS and PVFS. Wittawat Tantisiriroj, Swapnil Patil, Garth Gibson (CMU), Seung Woo Son, Samuel J. Lang, Robert B. Ross (ANL). Appears in the proceedings of the 24^{th} Supercomputing Conference (SC 2011). November 12-18, 2011. Seattle, Washington, USA.

DiskReduce: Replication as a Prelude to Erasure Coding in Data-Intensive Scalable Computing. Bin Fan, Wittawat Tantisiriroj, Lin Xiao, Garth Gibson. Carnegie Mellon University Parallel Data Lab Technical Report CMU-PDL-11-112. October 2011.

YCSB++: Benchmarking and Performance Debugging Advanced Features in Scalable Table Stores. Swapnil Patil, Milo Polte, Kai Ren, Wittawat Tantisiriroj, Lin Xiao, Julio Lpez, Garth Gibson (CMU), Adam Fuchs, Billie Rinaldi (NSA). Appears in the proceedings of the 2^{rd} Symposium on Cloud Computing (SOCC 11). October 26-28, 2011, Cascais, Portugal.

DiskReduce: RAID for Data-Intensive Scalable Computing. Bin Fan, Wittawat Tantisiriroj, Lin Xiao, Garth Gibson. Appears in the proceedings of the 4^{th} Petascale Data Storage Workshop (PDSW). November 15^{th}, 2009. Portland, Oregon, USA.

Fast Log-based Concurrent Writing of Checkpoints. Milo Polte, Jiri Simsa, Wittawat Tantisiriroj, Garth Gibson, Shobhit Dayal, Mikhail Chainani, Dilip Kumar Uppugandla. Appears in the proceedings of the 3^{rd} Petascale Data Storage Workshop (PDSW). November 17^{th}, 2008. Austin, Texas, USA.

In Search of an API for Scalable File Systems: Under the Table or Above It?. Swapnil Patil, Garth A. Gibson, Gregory R. Ganger, Julio Lopez, Milo Polte, Wittawat Tantisiroj, Lin Xiao. Appears in the proceedings of the 1^{st} Workshop on Hot Topics in Cloud Computing (HotCloud). June 15^{th}, 2009. San Diego, California, USA.

Data-intensive file systems for Internet services: A rose by any other name .... Wittawat Tantisiriroj, Swapnil Patil, Garth Gibson. Carnegie Mellon University Parallel Data Lab Technical Report CMU-PDL-08-114. October 2008.

Talks

On the Duality of Dataintensive File System Design: Reconciling HDFS and PVFS. Presented at the 24^{th} Supercomputing Conference (SC 2011). November 2011.

RAIDTool: A First Step to RAID 6 in HDFS. Presented at the 18^{th} annual Parallel Data Lab Workshop & Retreat. October 2010.

DiskReduce: RAID for Data-Intensive Scalable Computing. Presented at the 4^{th} Petascale Data Storage Workshop (PDSW), Supercomputing '09. November 2009.

DiskReduce: Making Room for More Data on DISCs. Presented at the 17^{th} annual Parallel Data Lab Workshop & Retreat. November 2009.

Crossing the Chasm: Sneaking a Parallel File System into Hadoop. Presented at the 16^{th} annual Parallel Data Lab Workshop & Retreat. November 2008.

References

Dr. Garth Gibson
Professor of CS
Parallel Data Lab
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
garth@cs.cmu.edu

Dr. Gregory R. Ganger
Professor of ECE & CS
Director, Parallel Data Lab
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
ganger@ece.cmu.edu