The CODES Project

Enabling Co-Design of Multi-Layer Exascale Storage Architectures

Sam Lang (Technical Lead), Bill Allcock, Rob Ross, Argonne National Laboratory
Chris Carothers, Rensselaer Polytechnic Institute

Overview

The data demands of science and the limited rates of data access place a daunting challenge on the designers of exascale storage architectures. *Co-design* of these systems will be necessary to find the best possible design points for exascale systems. Designers must consider performance, reliability, and power consumption in the context of the I/O patterns and requirements of applications and analysis tools at exascale. Meeting these constraints will require the development of a multi-layer hardware and software architecture incorporating devices that do not yet exist. The most promising approach for co-design of such systems is simulation.

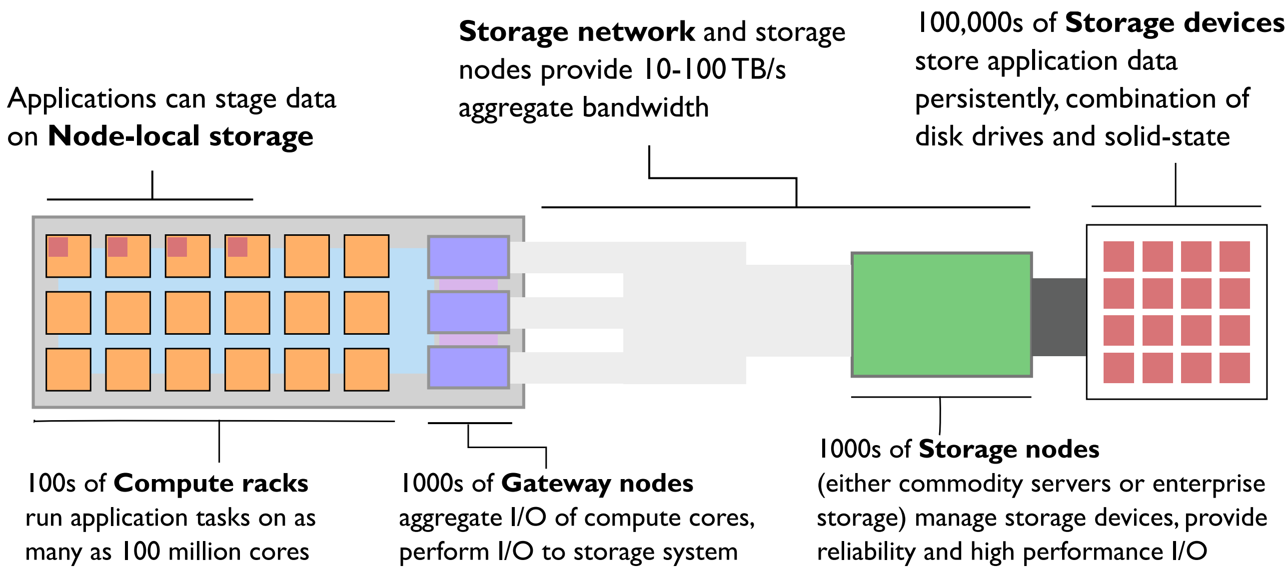
The goal of this project is to enable the exploration and co-design of exascale storage systems by providing a detailed, accurate, and highly parallel simulation toolkit for exascale storage. We will develop models to realistically represent application checkpoint and analysis workloads. These models will be joined together using the Rensselaer Optimistic Simulation System (ROSS), a discrete-event simulation framework that allows simulations to be run in parallel, decreasing the simulation run time of massive simulations to hours. Building on our prior work in highly parallel simulation and using our new high-resolution models, our system will capture the complexity, scale, and multi-layer nature of exascale storage hardware and software, and it will execute in a time frame that enables "what if" exploration of design concepts.

With this new toolkit we will investigate design options and trade-offs related to improving the reliability, performance at scale, and power consumption of potential exascale storage architectures. We will work with industry, DOE computing facilities, and the computer and computational science communities to refine our models and to encourage the use of this powerful tool in the design of future extreme-scale storage systems.

Impact

The project will advance the goals of the ASCR mission through development of an exascale storage simulation framework, deployment of that framework to storage system researchers, vendors, and experts in the community, and most importantly, through discovery of novel storage architectures, storage software algorithms, scalable application interfaces to storage. Through new storage system designs that improve energy efficiency, this research will help to lower the cost of Exascale computing for the ASCR. Through better I/O concurrency and system resilience, this research will enable the scalability of scientific applications of interest to the Department of Energy and ultimately, to advance scientific discovery.

The CODES Project



In the near term, storage architectures are likely to evolve to include storage in the compute system, but otherwise will look similar to those seen on systems at Oak Ridge and Argonne today. Accurate simulations of these systems will help us better understand potential bottlenecks and allow us to compare to revolutionary alternatives.

Current Activities

Network Modeling - We are developing a highly detailed, flexible, large-scale torus network model to understand the performance implications of different exascale architecture components with a focus towards scalable, reliable storage. Our initial results demonstrate strong-scaling of the torus network model using up to 32,000 cores on a 16 million node torus model. Simulations of the torus are run using the massively parallel optimistic simulator system (MPOSS) developed at Rensselaer Polytechnic Institute.

Failure Tolerance Simulations - We are evaluating the scalability, load and failure tolerance properties of a set of distributed protocols for detecting failures and disseminating information throughout a large storage system. Our initial results show that epidemic algorithms with probabilistic guarantees provide scalable alternatives to more static protocols that provide strong guarantees about state. We have integrated our core networking infrastructure into ROSS (the Rensselaer Optimistic Simulation System) to allow for realistic simulation of our storage algorithms and protocols, and have begun simulating epidemic and traditional failure-detection methods on a set of projected exascale storage designs.

Contacts

For more information on the project, please contact Sam Lang <slang@mcs.anl.gov>. For additional information on the ROSS.Net simulation system and possible uses, please contact Chris Carothers <chrisc@cs.rpi.edu>