Researchers create first-of-its-kind composable storage platform for high-performance computing

November 12 2018, by Amy Loeffler

Ali Anwar is first author on a paper announcing the advent of BESPOKV, a novel framework for HPC systems. He is a former student at Virginia Tech and is currently employed at IBM Research. Credit: Virginia Tech

Large-scale, advanced high-performance computing, often called

supercomputing, is essential to solving both complex and large questions.

Everything from answering metaphysical queries about the origins of the universe to discovering cancer-fighting drugs to supporting high-speed streaming services, requires processing huge amounts of data.

But storage platforms essential for these advanced computer systems have been stuck in a rigid framework that required users to either choose between customization of features or high availability.

Now, Virginia Tech researchers have found a way to give high-performance computing (HPC) data systems the flexibility to thrive with a first-of-its-kind framework called BespoKV, perhaps helping to one day achieve the HPC goal of performing at the exascale, or 1 billion billion calculations per second.

The researchers will present their findings at the Association of Computing Machinery/IEEE Supercomputing Conference in Dallas, Texas, on Nov. 13.

The main ingredient to the functioning of the new platform is key value (KV) systems. KV systems store and retrieve important data from very fast memory-based storage instead of slower disks. These systems are increasingly used in today's high-performance applications that use distributed systems, which are made up of many computers to solve a problem. High-performance computing relies on having computers intake, process, and analyze huge amounts of data at unprecedent speeds. Currently, the best systems operate at a quadrillion calculations per second, or a petaflop.

The research is relevant to industries that process large amounts of data, whether it be the space-hogging, intense visual graphics of movie streaming sites; millions of financial transactions at large credit card

companies; or user-generated content at social media outlets. Think large media sites like Facebook where content is everchanging and continually accessed. When users upload content to their profile pages, that information resides on multiple servers.

Professor of Computer Science Ali Butt is a co-author on the paper being presented at Supercomputing 2018 in Dallas. Credit: Virginia Tech

But if you have to continually access certain content, KV systems can be far more efficient as a storage medium because content loads from the faster in-memory store nearby, not the far-away storage server. This allows the system to provide very high performance in completing tasks or requests.

"I got interested in key value systems because this very fundamental and simple storage platform has not been exploited in high-performance computing systems where it can provide a lot of benefits," said Ali Anwar, first author on the paper being presented and a recent Virginia Tech graduate who is currently employed at IBM Research. "BespoKV is a novel framework that can enable HPC systems to provide a lot of flexibility and performance and not be chained to rigid storage design."

The main innovation of BespoKV is that it supports composing a range of KV stores with desirable features. It works by taking a single-server KV store called a datalet and enables immediate and ready-to-use distributed KV stores. Now, instead of redesigning a system from scratch to accomplish a specific task, a developer can drop a datalet into BespoKV and offload the "messy plumbing" of distributed systems to the framework. BespoKV decouples the KV store design into the control plane for distributed management and the data plane for local data storage.

The framework also enables new HPC services for workloads that businesses and institutions have yet to anticipate.

One of the major limiting effects of current state-of-the-art KV stores is that they are designed with pre-existing distributed services in mind and are often specialized for one specific setting. Another limiting factor is the inflexible monolithic design where distributed features are deeply baked into a system with backend data stores that do things like manage inventory, orders, and supply. The rigid design of these KV stores is not adaptive to everchanging user demands for myriad backend, topology, consistency, and a host of other services.

"Developers from large companies can really sink their teeth into designing innovative HPC storage systems with BespoKV," said Ali Butt, professor of computer science. "Data-access performance is a major

limitation in HPC storage systems and generally employs a mix of solutions to provide flexibility along with performance, which is cumbersome. We have created a way to significantly accelerate the system behavior to comply with desired performance, consistency, and reliability levels."

BespoKV can be nimble because it allows an arbitrary mapping between desired services and available components while supporting distributed management services to realize and enable the distributed KV stores associated with the datalet.

"Now that we have proven that we can make the efficient and simple action of using KV systems in powerful HPC systems, customers won't have to choose between scalability and flexibility," said Butt.

Provided by Virginia Tech