TERASCALA

High Throughput Storage at a Major Renewable Energy Laboratory

Terascala, Inc.
145 Bodwell St
Avon, MA 02322
www.terascala.com

Summary

A major U.S. Government laboratory researching renewable energy and energy efficiency recently selected the Terascala™ RTS 1000 storage appliance to provide high throughput storage for its research initiatives.

Background

As the demand for renewable energy and energy efficiency research skyrockets worldwide, this major government laboratory has been in the spotlight recently. The scientific computing staff at the Laboratory is responsible for keeping up with the agency's computing demands.

"One of the things we're seeing as we move forward is an enormous growth in the amount of data we collect," said the laboratory's director. "The ability to store and retrieve data is just the start of it all. We're constantly looking for ways to increase our performance in terms of storing and retrieving data because it enables us to effectively search, analyze and mine our data, not just to store it and access it periodically. That's where we get real value out of it."

As an example of just one area the laboratory is involved with, within one photovoltaic center, a team of people synthesizes new materials and another team characterizes those materials under different environmental conditions. Data is collected each step of the way, including what material is being analyzed, what the conditions are, and how the materials are performing.

"We are helping keep track of all of that data, what changes were made and what impact they have," said the director. "The goal is to find trends that help us discover the right sort of controls that can improve the efficiencies of the resulting devices based on how it was synthesized."

"It doesn't matter if we're working in materials or biology or any other type of experiment. They all have the same sort of scientific process involved and there's an informatics component that we're trying to exploit so they can do it better. And that's data. Data is the modern currency of the scientific process. You absolutely have to keep track of your data to analyze and improve your results," said the director.

Lab Evaluated Terascala vs. "Homegrown" Lustre implementation

The scientific computing team had been using NFS for some of the laboratory's's storage needs but found that its implementation wasn't scalable and didn't provide the performance needed to keep up with demand. The team decided to evaluate the Lustre™ file system and built its own implementation on an existing storage platform in the lab. At the same time, the team decided to and to evaluate the Terascala RTS 1000 Storage Appliance.

"We performed numerous tests on our internal Lustre environment and compared the results with the Terascala RTS 1000," said the director. "We found that the Terascala solution was approximately 20 percent faster on InfiniBand."

"The Terascala RTS 1000 is optimized to run Luster and it provides us with the high levels of performance we need," said the director. "It allows us to perform data analysis quickly with the compute nodes and to access that data rapidly. For us, that means that we can analyze data in a day instead of the three or four days it used to take. The improved performance will enable us to do more with our data in the future. We collect data from all sorts of experiments, and our ability to quickly store and retrieve information enables us to more effectively mine our data."

The laboratory purchased the Terascala RTS 1000 for use within their high performance computing environment. They also leverage the appliance as centralized storage for its Postgres database and for storing application run time results.

Cost-effective Terascala RTS 1000 Provides Scalability, Superior Support

"We looked at other solutions, but the Terascala solution gave us the scalability we needed to meet our future requirements," said the director. "We also found the Terascala team very easy to work with. They have a great deal of Lustre knowledge and are both proactive and responsive to our inquiries. And in terms of price, the RTS 1000 was very cost effective."

Terascala's RTS 1000 architecture provides the ability to expand throughput or capacity by adding additional nodes without impacting the existing environment. Additionally, the RTS 1000 incorporates "plug and play" replaceable units, hot swappable components for reliability, redundant power supplies and high density packaging along with the choice of GbE, 10GbE or InfiniBand.

Single Name Space Simplifies Overall Storage Environment

"We used to have multiple storage environments and were moving data around to process it. Now, because our data is within one file system and one namespace, it's easy to manage," said the director. "It's a very cost-effective solution from a total cost of ownership perspective."

"Some of our experimental data was being stored in flat files and in some databases, but what we really needed was the ability to look at all of our data as a whole. In the past, much of the data was stored on individual PC's and sometimes further information about the experimental design sits in a notebook. It may take weeks to go back and get the data you need to publish one paper," said the director. "Now, all of that is captured online. It's uploaded into the file system and people have the ability to search it and to work with it. Scientists are now able to readily track synthesis data and correlate it with characterization data, and it's truly helping to advance our research."

"Simply put, our ability to collect and analyze data is going to be a gating factor on our ability to advance our research and developments efforts," said the director.. "We feel that the ability to manage data is an increasingly important part of our research and development efforts and Terascala's products are playing a critical role in that for us," he said.

Terascala's Storage Appliance

The Terascala RTS 1000 Storage Appliance is a Lustre parallel file system-based system that offers high throughput and high capacity. The RTS 1000 is designed to deliver the maximum throughput to enable applications to run at peak efficiency. It can deliver over 2GB/sec from a single enclosure and up to 10 GB/sec for a full rack solution. Designed to plug directly into the compute client network environment, the RTS 1000 has an optimized data path from the client network through to the disks within the storage device.

Simplification of deployment and ongoing management is a key aspect of the RTS 1000. It is delivered as an appliance with all of the software installed and tuned to deliver performance. With its built-in management system, it is easy to add additional capacity and throughput or to fine tune key aspects of the environment. The RTS 1000 is designed to be operated and managed by system administrators without extensive storage experience.

Terascala leverages Lustre because it is an open source, high performance clustered software initially developed for applications needing very high throughput, scalability and capacity. It offers high levels of reliability, scalability and performance, having been deployed in some of the largest compute installations in the world. Lustre leverages a simple metadata/stored object architecture, where the metadata server stores location information about data and the object store servers act as the repository for the actual data. This approach allows throughput to be scaled by simply adding additional object store servers. Terascala has an optimized architecture and has tuned Lustre for the specific capabilities of the RTS 1000, so the appliance delivers a simple, easy to use and expandable solution.

With its simplified architecture and the use of commodity technologies where possible, the RTS 1000 is able to deliver tremendous price/performance while simplifying the overall deployment and management of a high throughput storage solution.

Contact Terascala at www.terascala.com or info@terascala.com to learn more.

About Terascala

Terascala develops high throughput, high capacity and cost-effective storage solutions. Its unique storage appliance approach is changing the dynamics of the performance driven computing market, enabling existing users to do more for less while enabling new users to maximize the capabilities of their processing infrastructure. Founded in 2005, Terascala is based in Avon, MA. Contact Terascala at 508-588-1501 to learn more about how we can solve your storage throughput challenges, or visit our website at www.terascala.com

TERASCALA

Terascala, Inc.
145 Bodwell St.
Avon, MA 02322
Tel: 508-588-1501
Email: sales@terascala.com
www.terascala.com