Quantum.

A Leading Investment Fund Extends Its Competitive Edge by Deploying ActiveScale™ Cloud

Challenge

- Store over 100 Petabytes of data for long periods, yet be able to readily access that data at high speed for quantitative processing and analysis operations
- Spend less time deploying, managing and maintaining the system, leaving more time to focus on developing new strategies to actually use the data
- Scale the storage system easily to keep up with rapid data growth without busting the IT budget

Solution

Two ActiveScale cloud object storage systems. One for "hot" data under analysis and the other for "cool" archival data that may not be read as often but must be easily accessible. This provided the key benefits of public cloud—simple operation, zero hardware maintenance responsibility and fixed costs—in a low-cost, high-performance on-premises solution.

Key Results

- Deployed the system in under 8 hours—compared to 6-8 weeks with previous approach
- High and consistent performance for both hot and archived data
- Capex costs reduced by over 35 percent
- Slashed operational expenses for data infrastructure by 50 percent
- Data protected today with ability to easily scale capacity in the future
- Less time spent trouble shooting and maintaining equipment
- More time spent developing cutting-edge financial applications and driving maximum value from the company's data

Company Profile

For cutting-edge investment funds, success increasingly depends on data science. The winners in this space have better talent and deeper insights than competitors—and massive amounts of historical and real-time data to fuel their trading models. One such U.S.-based firm followed this blueprint to great success. Starting with just a few hundred employees five years ago, the company has become one of the largest, most profitable funds in the United States. As the firm grew, however, they found that maintaining their competitive edge—and the constantly growing data infrastructure that enabled it—was becoming expensive and unwieldy. To stay on top, they needed a way to continually feed more data into their models without sacrificing performance, and without letting costs get out of control.

Challenges

As the firm continues to develop new trading models, they never know exactly what data they'll require. To that end, the company stores several decades worth of data from more than 10,000 sources—over 100 Petabytes and growing. The data needs to be archived for long periods but also be readily available for analysis. It's a tall order, requiring a serious data infrastructure. The company had originally built an open-source Ceph™ solution with commodity storage and custom software. At the time, this "do-it-yourself" (DIY) approach had seemed like a great way to get high-performance storage at a low cost. However, with such fast data growth, the platform was falling short in both categories.

First, as the firm relied on more complex quantitative processing and analysis for its trading models, the performance just wasn't good enough. The solution was also becoming more difficult to deploy and maintain, requiring constant patching and testing. The firm's infrastructure team spent almost all their time just trying to keep the system running, leaving little to focus on developing new strategies to actually use the data. And, as the pool of data kept growing, scaling was becoming a big, expensive problem.

Finding the Right Solution

The company wanted a low-cost, software-defined storage and archive solution that could maintain a vast and growing amount of data. They needed top performance—from legacy applications as well as new ones, ideally without requiring significant rearchitecting. And, they wanted a platform that would be much easier to manage, freeing up their engineers to focus on strategic value rather than nonstop maintenance.

All of these requirements seemed like a natural fit for a cloud solution. But, due to inconsistent performance and massive "get" costs every time they needed to pull large amounts of data from archive, public cloud was not a viable option.

The Solution

After exploring the options, the company chose an ActiveScale cloud object storage system. The solution gives them all the key benefits of public cloud—simple operation, zero hardware maintenance responsibility, and fixed costs under a service-level agreement (SLA). But it provides them in a low-cost, high-performance on-premises solution.

The company stores data at two locations on separate ActiveScale systems. The architecture bifurcates the data storage into two levels. One supports "hot" data under analysis, while the other stores "cool" archival data that may not be read as often but is easily accessible with predictable performance.

Custom software running on a compute cluster delivers data pulled from a front-end SSD-based cache to ActiveScale using an Amazon S3™-compliant interface. And, the company no longer has to worry about performance lags, with a solution that delivers upwards of 8 GB/s to 20 GB/s throughput.

With ActiveScale, the company can:

- Store two copies of data on a highly durable/highly available pair of ActiveScale systems
- Easily scale each site by multiple Petabytes by adding a single rack at each location
- Stay ahead of explosive growth in their data sets
- Provide always-online access to historical data for use by the high-performance front-end
- Reduce capital costs and dramatically reduce operational expenses compared to the previous "DIY" solution

Results

ActiveScale has proven to be exactly what the firm needed, providing ample support for its constantly growing storage needs, with superior performance at an affordable cost. The infrastructure team was able to deploy the system in under 8 hours—compared to 6-8 weeks with the home-grown Ceph solution. Applications have high, consistent performance for both hot and archived data. And, with ActiveScale, capex costs are 35 percent lower than the DIY platform, and the firm has slashed operational expenses for data infrastructure by 50 percent.

Most important, the firm's expert data engineers are no longer spending their days racing around trying to stay ahead of the data deluge and keep the infrastructure running. With a solution that protects their data and scales easily, they're enjoying a more cloud-like management model. They no longer worry about day-to-day hardware operation, or wrestle with each new software revision. Instead, they can focus on the job they were brought in to do in the first place: Developing cutting-edge financial capabilities. Evolving legacy applications to state-of-the-art. And driving maximum value from the company's data.

Quantum®

www.quantum.com

CS00491A-v02