

SOLVING TICK DATA RETENTION CHALLENGES WITH POWERFUL TECHNOLOGY DELIVERING STORAGE-AS-A-SERVICE



By Terry Smith, Senior Director, Advanced Solutions Group
M.S. Computer Science

Overview

One of the largest banks in the United States had outgrown its storage capacity for tick data and needed to overhaul their system. Given that tick data is not valuable without analysis, the solution had to address rigorous access needs as well as meet organizational and regulatory retention requirements. At the same time, the bank had business challenges related to capital and operating expenses that had to be taken into account, including a desire to limit additional headcount required to support any new technology.

After exploring various options, the bank selected Penguin Computing for a software-defined storage (SDS) data center solution based on the FrostByte™ storage system with Red Hat® Gluster Storage, which was developed in close cooperation between Penguin Computing and Red Hat to provide optimal performance.



To meet the client's business needs, we suggested a multi-site design with Penguin Computing's expert system administrators running the data center as virtual employees. To reduce cost and allow maximum flexibility, though, we proposed that the off-site replica of the bank's production system be hosted by Penguin Computing, further decreasing the impact to the client's bottom line.

Problem

A significant portion of modern financial services data analytics is performed by sifting through large volumes of tick data, which is the measurement of upward or downward movement in the price of a security. This results in a massive volume of securities data, especially because tick data for stocks over \$1 has been measured down to the 1 cent level since 2001.

Data about the change in securities prices over a period of time is extremely valuable, as it allows financial institutions, researchers, and traders to run analytics, build historical simulations, develop trading and market-making strategies, and build transaction-cost models. Financial institutions know they can make more insightful decisions by analyzing their tick data. And, regulatory standards in the financial organization may mandate specific data retention policies.

It is not a simple matter for organizations to retain the huge quantity of desired data due to the cost and capacity constraints of their storage systems. This is because tick data contains multiple parameters, including timestamped bid or ask prices, bid or ask sizes, quote time, trade time, and exchange information for each specific security that is traded. As a result, tick data for a typical organization can use hundreds of gigabytes to terabytes or even petabytes of storage. Retention of tick data for long-term trend analysis for one of the largest banks in the United States would understandably be an even larger challenge.

A technically sound solution that met the functional and regulatory needs of the bank would normally be a very expensive proposition, well beyond the bank's planned spend. The bank needed a solution provider that went beyond robust product offerings and could, instead, address the total client need. This is when they found Penguin Computing.



Solution

The bank selected the FrostByte storage system from Penguin Computing, opting for a version of this turn-key solution that included – and was optimized for – the Red Hat Gluster file storage system.

The FrostByte line was built on the premise that flexible SDS, deployed on industry-standard hardware, and supported by expert services can deliver the performance and scale that large data centers – such as those in financial institutions with huge quantities of tick data – need to address their storage challenges.

Predicting the future is difficult, so, if you can build flexibility into your data center, you can prepare for likely scenarios and make better use of the infrastructure as situations change. With SDS policy-based provisioning and management, data centers get dynamic optimization of infrastructure capabilities to application service-level requirements, yielding both storage and staff efficiencies. This is a significant benefit when dealing with the huge volumes of data that accompanies most financial transactions.

This flexibility gave the bank both cost-effective storage and the ability to scale as needed in the future. And, given the exorbitant cost of storage appliances and how rigidity in capacity expansion can further stress financial institutions, the design ensured that expansion could be managed cost-effectively by simply adding more FrostByte building block nodes.

Why Red Hat Gluster? Gluster is a very flexible, scale-out, distributed storage system, and Red Hat Gluster Storage is rigorously tested to ensure enterprise-class performance. Having an optimized and vetted hardware/software solution is critical in financial services, where sub-par performance can lead to unacceptable latencies in making buy/sell decisions.

With this design philosophy, FrostByte lifts the limitations imposed by traditional storage appliances while giving organizations open, agile, and cost-effective SDS. In addition, customers switching from traditional, monolithic storage appliances to a solution based on Red Hat Gluster Storage gain tunable performance and flexible capacity scaling. Customers switching from storage appliances like EMC Isilon to Red Hat Gluster Storage for tick data also report cost savings on solution acquisition as well as ongoing maintenance and support.

If you profile a similar workload on both an EMC Isilon system and a Red Hat Gluster Storage solution, just four nodes of Red Hat Gluster Storage outperforms a 16 node EMC Isilon cluster. The Red Hat solution can sustain the rigorous workload pattern of a large number of directory

creations because tick data is stored in distributed time series, historical database partitioned format (HDPF) files across many directories.

To reduce cost and allow maximum flexibility, Penguin Computing proposed creation of an on-site FrostByte-based production system and an off-site replica. This recommendation for a multi-site design made use of another of Gluster's capabilities, geo-replication, in which the data are asynchronously replicated off-site. The key, however, was that the off-site replica would be hosted by Penguin Computing in a unique Storage-as-a-Service capability, further decreasing the impact to the client's bottom line. (We've been running HPC-as-a-Service via our Penguin Computing On-Demand™ (POD) platform since 2009 so supporting client data center needs with remote access as well only made sense.)


Inexperience with a large-scale storage solution represents a real operational risk that many data center managers are – rightly – unwilling to take. That's why the final component of this comprehensive, customer need-focused solution featured Penguin Computing's expert system administrators running the data center as virtual employees, part of our SysAdmin-as-a-Service offering.

With managed services, data center managers can avoid the immediate cost of additional headcount while delivering critical and expanded services. In this case, the level of expertise of Penguin Computing's Managed Services team – particularly with an embedded administrator ensuring close communication and alignment – became the key to solving the bank's problems. Our close relationship with Red Hat ensures that our Professional Services team, including our Managed Services staff, are up to date on the latest Red Hat developments and able to bring those to our clients.

However, Penguin Computing also has deep experience in tick data from having jointly built a record-setting high-frequency trading solution, nicknamed the Ultimate Trading Machine, announced on October 17, 2017 at the STAC Summit in Chicago, IL. The solution set the bar for tick-to-trade performance by achieving a 98-nanosecond tick-to-trade latency, making this solution 18% faster than the previous world record.

As a result of this effort, the bank received a cost-effective Storage-as-a-Service solution, optimized to meet the its unique workload, regulatory, and cost challenges. Moreover, this model allowed the client to shift what could have been large capital expenses to smaller, more manageable operational expenses. This has enabled the bank to grow the managed storage capacity as needed and at significant cost savings over a traditional approach.



SOLUTION COMPONENTS
Penguin Computing FrostByte with Red Hat Gluster Storage
Penguin Computing Managed Services


Learn More

See related solutions at www.penguincomputing.com/frostbyte

For pricing on your specific design needs, contact a representative by email at sales@penguincomputing.com or call **1-888-PENGUIN (736-4846)**.

Purchase with Financing

Finance products, services, even soft costs with Penguin Computing Capital. Choose from options such as no money down, flexible billing choices, extended repayment timelines, and a variety of end-of-term alternatives.

About Penguin Computing

Penguin Computing, Inc. is a 20-year-old, U.S.-based global provider of high-performance computing (HPC), artificial intelligence (AI), and data center solutions with more than 2,500 customers in 40 countries, across eight major vertical markets. Penguin Computing offers a comprehensive portfolio of hardware, software, and services, including solutions based on the Open Compute Project (OCP), as well as financing and top-rated customer support. Penguin Computing products include Linux-based servers, software, integrated, turn-key clusters, enterprise-grade storage, and bare metal HPC on cloud via Penguin Computing® On-Demand™ (POD).