

#### SOLUTION BRIEF

# Power Quant Trading with Pure Storage and Code Willing

Accelerating time to insight and alpha generation.

In the landscape of data science and quantitative trading, serving analysts consistently and with agility is paramount. As financial firms strive to gain a competitive edge for alpha generation, a high-performance, scalable, and secure data solution is a key element for success. The strategic partnership between Pure Storage<sup>®</sup> and Code Willing introduces a significant enhancement, integrating industry leading, multi-cloud data solutions from Pure Storage with Code Willing's CWIQ FS file system. This integration not only addresses traditional data handling challenges but also introduces a new paradigm in managing and serving on-premises and cloud-based data with unprecedented agility and efficiency. This partnership yields a solution that enables models to be developed and tested efficiently, yielding greater outcomes versus teams being hampered by infrastructure.

## The Conundrum of Internal Data Privacy

Data strategy owners are often met with the challenge of providing permission to various pods with access to specific data while not allowing others to be able to see what it or other pods are using or accessing. Legacy file systems can't easily perform this task so many firms turn to replicating data and sending it to individual pods separately—which is time-consuming and costly—requiring them to store multiple copies of the same data. This challenge permeates the firm, keeping the compliance team under the constant pressure of multiple data copies being updated and delivered to multiple pods or individual users on a daily basis. Factor in growth from both a compute and data volume standpoint and individual P&L/budget owners are negatively impacted from exponentially increasing costs of doing business.

What if there was a solution capable of delivering a single global file system capable of spanning hundreds of machines, managing billions of files and petabytes of data, all while delivering the full aggregate bandwidth of managed object stores, such as S3?

What if metadata was updated in real-time across all machines so the cluster can see a consistent view of the world? What if data hiding and date-range restrictions, not only on the pod level, but also on the user level were capable? What if data feed subscription usage was able to be audited for maximum ROI? What if data was able to be strategically tiered, taking both performance improvements and cost savings into consideration?





#### **Scalable Efficiency**

Tailored to handle highly parallel workloads, Pure solutions provide the seamless scalability necessary for improved analytic runtimes and development of complex financial models.



## Advanced Data Management

With seamless tiering and independent scaling of performance and capacity, FlashBlade empowers financial institutions to handle massive volumes of tick and historical data cost-effectively, for precise and efficient compliance, backtesting, and analysis.



## Complete Data Access Control

CWIQ FS is built to support massive data pipelines with full control over what teams and individual data consumers are able to see and a rich feature set enabling deduplication, multi cloud deployment and full audit logging capabilities.



## The Ultimate Data Management Tool for Model Development

Pure Storage and Code Willing recognize that the teams supporting quantitative trading and data scientists face a myriad of challenges, including the management of increasing data volumes, centralizing and serving data to authorized teams, the need for rapid processing capabilities, and stringent compliance requirements. Being able to serve teams with the data they need at scale means they can act on signals faster build, test, and tune models quickly, and execute strategies efficiently. Underpinning data with a highly performant, consistent, and fully observable platform simultaneously yields dramatic decrease in model testing runtimes and management overhead. FlashBlade//S<sup>®</sup> from Pure Storage delivers the low latency, maximum throughput, and continuous uptime needed to meet the demands of data scientists and PMs. FlashBlade// E<sup>®</sup> empowers firms to build strategic data tiers to manage and grow their historical datasets with unrivaled efficiency. Code Willing's CWIQ FS delivers content-based hashing, auto-deduplication, straightforward distribution and caching, control of user access to and visibility of data based on date range and dataset, and full file system audit capabilities so teams can see what data is being used, by whom, and how often. The alliance of Pure Storage and CWIQ FS creates a game changing toolset, delivering a scalable, searchable, and secure data repository with enhanced performance for data science and alpha generation.

## **The FlashBlade Platform**

FlashBlade<sup>®</sup> is a distributed, scale-out unstructured data platform that provides the flexibility needed to exploit alpha opportunities by effectively balancing loads and processing data at exabyte scale. The unique modular architecture of FlashBlade allows capacity and performance to be scaled independently, enabling the platform to meet evolving workload capacity and I/O needs. Native NFS, SMB, and S3 protocols, along with fully-featured management REST APIs, make deployments, management, and upgrades simple and non-disruptive.

A single unified platform combining compute, storage and networking simplifies deployment, management and upgrades. And Pure Storage DirectFlash® modules eliminate the need for SSDs or HDs, ensuring peak performance, security and reliability.

The FlashBlade platform includes:

- FlashBlade//S<sup>\*</sup>: The Ideal data platform for the most demanding unstructured data workloads. It offers multidimensional performance for billions of small and large files or objects, delivering both the high throughput and low latency that are critical to speed-sensitive trading strategies. Cutting-edge capabilities without complexity.
- FlashBlade//E<sup>\*</sup>: Empowers firms to manage and grow their historical data sets with unrivaled efficiency. The //E series from Pure Storage is an all-flash capacity-optimized data platform, designed to tackle both structured and unstructured data repository workloads that until now have largely remained on disk-based storage systems. You gain simplicity and sustainably with better economics than disk.

#### The CWIQ FS File System

CWIQ FS is a shared file system that scales automatically, with unlimited storage and high performance. This clustered file system provides a single global file system that is capable of spanning hundreds of machines, managing billions of files and petabytes of data. Delivering the full aggregate bandwidth of managed object stores like S3, metadata is updated in real-time across all machines to provide a consistent view of a firm's data.



## **Use Case: Complex Quant Research**

A quantitative analyst needed to run a complex simulation against two years of a CME's most detailed order-by order-raw feed (about 150GB of compressed data per day) for a new client. The analyst needed to be able to run all 500+ days in parallel as quickly as possible and have the ability to rerun the analysis whenever their research was ready.

### Solution

The Code Willing and the Pure Storage team provisioned a semi-private cloud with 10 servers and 500TB on a Pure Storage FlashBlade<sup>®</sup>. Cloud tools from Code Willing made the client's private cloud appear as Amazon 'spot' compute on-demand resources. The analyst was able to reserve capacity dynamically to run their jobs using the cloud tools with no cloud specific code changes on their side.

#### Outcome

The solution resulted in some impressive results, including:

- Time to complete a 500+ day simulation run in parallel was 30 minutes of elapsed time.
- CPU utilization was between 90-99%.
- The max network Transfer rate was 19GB/sec, floating between 12-14.9 GB.
- Nodes were brought from idle to active and then back to idle in matter of minutes.

Most importantly, the analyst was able to complete all their research and deploy their strategy for live trading in two months incredibly valuable given the opportunities offered by the recent elevated volatility.

## Your Partner in Quant Trading and Analytics

Firms within the financial sector require experienced partners who can offer valuable insights and specialized knowledge, enabling them to streamline their model training and trading platforms, as well as foster continuous innovation to maintain a competitive edge. As a globally recognized leader in providing near real-time and historical data solutions, Pure Storage stands as the trusted collaborator for banks, investment funds, asset managers, and trading firms worldwide. Pure Storage is dedicated to supporting our clients and partners in accelerating their alpha generation strategies to market and optimizing their return on investment.

#### **Additional Resources**

- Learn more about our <u>quant trading solutions</u>.
- Learn more about our partners at Code Willing.

#### purestorage.com

800.379.PURE





©2024 Pure Storage, the Pure Storage P Logo, DirectFlash, FlashBlade, FlashBlade//E, FlashBlade//S, and the marks in the Pure Storage Trademark List are trademarks or registered trademarks of Pure Storage Inc. in the U.S. and/or other countries. The Trademark List can be found at <u>purestorage.com/trademarks</u>. Other names may be trademarks of their respective owners.