Executive Summary

Financial institutions have always been closely aligned with high-performance computing or HPC. And with a challenging economy that has made it increasingly demanding and complex for financial firms to endure and thrive, this relationship has grown even stronger.

Capital markets firms require high-frequency, low-risk, compute-intensive and reliable solutions to remain competitive. It is important to mention that HPC helps the industry predict performance and extend competitive reach – both to retain and expand client base.

HPC systems are used to solve the most complex financial computations, such as spotting market trends and calculating risk. These clusters of computers are easier to deploy, more scalable and more cost-effective than standard computers.

Firms typically set up a computational cluster that is split into grids for specific tasks, such as processing transactions at specific times of the day or overnight. The grid computing approach has proven cost-effective because it drives up hardware usage.
The Situation

Capital markets firms deploy HPC systems for one purpose: to gain a competitive edge. These systems provide a range of advantages to financial services organizations, from speeding up trade processing and providing real-time risk analysis to supporting applications critical to financial business.

HPC systems help firms quickly analyze huge volumes of financial data and accelerate trade speed and execution. With the advent of high-frequency trading, a firm’s systems must be able to buy and sell financial instruments after owning them for only a fraction of a second, with the firm skimming a profit from tiny price fluctuations. HFT systems operate over high-speed networks at microsecond speeds with latencies in the nanosecond range.

The TABB Group, a Westborough, Mass. consultancy, suggests that profits made from high-frequency trading may have peaked at $7.2 billion in the United States in 2009, the biggest year for HFT on record. Then again, HFT represented 65 percent of all U.S. stock trading during the first two weeks of August 2011, according to TABB.

Indeed, high-frequency traders made record profits of about $60 million in the United States on one day: Aug. 8, 2011. What’s clear is that, for most capital markets firms, HFT drives not only market volatility but also transaction volume and data — raw processing that requires HPC systems.

These systems also provide real-time market risk analysis, risk analytics and portfolio calculations, giving firms a handle on their financial exposure and credit risks. HPC systems perform all of these calculations at sub-millisecond speeds so that firms can run simulations of different economic scenarios and gain a better understanding of their exposure to defaults and other risks before they finalize transactions.

When capital markets firms buy HPC systems, they care about more than just performance. Ease of management, ease of deployment and support of existing third-party applications are also critical. Chief among the benefits of purchasing HPC clusters — especially systems built using off-the-shelf multi-core processors — is the ease with which firms can scale such systems to run third-party cluster management software.

J.P. Morgan deployed a special-purpose HPC system last year that performs risk calculations in about four minutes — a task that took eight hours back in 2008. The system’s speed lets the company run its risk calculations continuously, instead of as an overnight process, and to consider many more potential risk scenarios than was possible previously.

Simply put, HPC systems let capital markets firms make better decisions faster, and at a lower cost. They provide multitasking traders with real-time data and analysis about market news and price fluctuations so they can make informed decisions and remain within their credit limits. In the financial industry today, all lines of business — including equities, derivatives and algorithmic trading groups — now rely on HPC clusters.

Ultralow Latency

For HPC environments to achieve ultralow latency, they need speedy processors, networks, messaging platforms and applications.

Ultrafast processors: Current mainstream HPC systems have six-, eight- or 12-core processors, speeding their ability to work through complex calculations and create accurate, detailed forecasts, simulations and data visualizations. The powerful processors at the core of these systems must be balanced with the right amount of memory bandwidth for optimal performance.

Capital markets firms often harness these multi-core servers into clusters that can automatically reconfigure themselves based on time of day or processing load. Such clusters make it easier to scale HPC configurations, so firms can add systems as needed to handle peaks more efficiently. Clusters are cost-effective because they increase hardware utilization, allowing expansion on an as-needed basis.

Another benefit of today’s ultrafast processors is that they reduce latency from microseconds to nanoseconds, by reducing input/output latency and improving I/O bandwidth. With increased floating-point performance, multi-core servers can host more trading algorithms per server.

Ultrafast processors not only speed up trading applications, but they also have the throughput to handle data-intensive apps such as real-time risk management. Some vendors market additional hardware features to enhance data security and protect against hackers, while offering improved management capabilities.

Chip companies continue to push the envelope on the development of multi-core processors, pointing to a future of even faster processors. In April, Intel announced new quad-core Ivy Bridge processors. However, the firm has been shipping experimental 48-core processors to research institutions since 2010.

Last fall, AMD shipped 16-core processors based on its new Bulldozer architecture, which involves fitting as many cores as possible on a chip while minimizing power requirements. And startups are selling 16-, 34-, 64- and 100-core processors for embedded applications that require massively scalable performance as well as enhanced power efficiency.

Low-latency networks: HPC systems need low-latency networks to provide millisecond transaction processing. Organizations typically use traceroute or ping tests to determine how long it takes for a packet to make a round trip from a server to an end user’s device to measure network latency.
Latency is the result of bottlenecks that prevent systems from taking full advantage of available bandwidth, the maximum data rate — measured in bits per second — that a network connection can support. A low-latency network connection, therefore, experiences little delay, while a high-latency network suffers from delays.

Distance, network congestion and the number of boxes between an end user's device and the server can all affect latency. One cause of high latency is a propagation delay as a message travels a long distance across a network.

A transmission delay refers to the time it takes to push all of a packet's bits onto the network and typically results from a packet's length. Processing delays refer to the time it takes routers to process packets. Packets that take many network hops, or router functions performed on packets (such as network address translation), usually account for these types of delays.

To achieve low-latency networks, system designers will integrate hardware, systems and protocols specifically designed to minimize the time it takes to move packets from one point on a network to another. Such networks also will apply queuing and congestion management to reduce latency.

Obviously, low-latency networks aren't all about hardware. It's also critical to fine-tune operating systems to handle packets as quickly as possible, and to assure that applications have speedy access to system resources such as distributed memory.

Capital markets firms, obviously, desire ultralow-latency networks for their trading infrastructures. They need direct and redundant connectivity to exchanges and institutional investors for real-time data feeds. Collocation facilities in major financial centers around the globe provide the proximity needed to reduce propagation delays.

Carrier-grade network infrastructure ensures that packets travel across the fastest available paths on long-distance connections. Typically, real-time network monitoring systems track latency along with other key metrics.

**Hardware-accelerated messaging:** The fastest trading platforms can handle millions of messages per second and scale to support hundreds of servers and thousands of message recipients.

Without middleware servers between the message senders and recipients, these platforms provide sub-millisecond messaging and can scale easily as message volume rises. For increased flexibility, HPC messaging systems support multiple transport protocols, including Transmission Control Protocol (TCP), multicast, shared memory and Remote Direct Memory Access (RDMA). Some also support low-bandwidth, machine-to-machine messaging and still others can handle real-time data streaming.

Instead of best-effort delivery, as the Internet provides, these high-performance messaging platforms offer guaranteed delivery and automatic failover to improve reliability and performance.

While some high-performance messaging vendors are delivering innovative software architectures, others offer messaging appliances that feature tightly integrated hardware and software systems to accelerate messaging. Both types of systems provide ultralow latency and strong message throughput.

Special-purpose message switches can be interconnected to deliver more than 10 million messages per second with latency that averages in the microseconds. Even better, these systems provide constant levels of performance even during peak periods of volatility.

With closely coupled hardware and software, these message switches provide deterministic flows through traffic management. These systems also have the smarts to offload message management overhead for speedier delivery of messages.

Some messaging platforms feature shared memory messaging, which enables applications to communicate using interprocess communication (IPC). This technique eliminates latency and accelerates communications between applications. In trading environments, IPC allows data feeds, trading algorithms and risk engines to share information faster than traditional messaging technology. For example, one IPC solution boasts average latency of less than 400 nanoseconds.

These appliances support unified messaging, meaning that they can provide ultralow-latency messaging across a variety of delivery channels, including enterprise, web and mobile platforms. Banks, hedge funds, currency trading advisers, foreign exchange traders, exchanges, investment managers and asset managers are all common users of hardware-accelerated messaging systems.

**Precise time synchronization:** Capital markets firms also need the most accurate systems available for tracking and synchronizing times associated with transactions. Cutting-edge appliances provide time accuracy to the microsecond — some, even to the nanosecond — so that firms can assure precise time-stamping of market data and transactions.

The latest time synchronization appliances rely on global positioning system data to synchronize servers and applications. They support industry standards such as the Network Time Protocol (NTP) and Precision Time Protocol (PTP). They also feature precise clock management software that works in conjunction with synchronization hardware to provide superior clock and time-stamp accuracy. These appliances use time-stamping network interface cards and
Linux, while supporting standard network interfaces such as 10-Gigabit Ethernet and high-speed InfiniBand.

Accurate system clocks and time-stamping are increasingly important as trading firms migrate to high-frequency trading and real-time risk management. Current hardware-software combos synchronize time across networks and HPC application clusters, minimizing the overhead involved with pushing time data to an application after it is requested.

This is an especially tricky feat when apps span multiple processors, which is why time synchronization appliances push time data all the way into the apps to eliminate delays. Appliances on the market solve this so-called last-mile problem without requiring custom application programming interfaces or code changes.

As traffic enters one of these appliances, it receives a time stamp at line rates, creating a permanent time record that’s precise to the microsecond — or nanosecond — depending on the level of accuracy required by the particular application. That time remains permanently attached to the Ethernet packet as it traverses the network.

Capital markets firms are buying time synchronization appliances not just to comply with regulations but also to improve competitiveness. These systems help trading firms reduce application latency. Apps can use the accurate time-stamped data to more precisely execute trades, manage liquidity and minimize risk.

**Efficient applications:** Capital markets firms not only need the fastest, lowest-latency HPC systems, but they also need efficient apps to run on them. In the wake of the global financial crisis, Wall Street firms are less willing to pay whatever it takes to increase performance or lower latency. Instead, they are focused on cost-effectiveness, which extends to the return on investment (ROI) their apps provide.

Trading software makers, therefore, strive to design increasingly efficient transaction processing platforms that lower transaction costs while providing high-quality customer service. For example, the newest portals give traders the ability to monitor and respond to messages in real time across multiple apps. Other new apps streamline trade execution, reducing the time between when a trader receives real-time market data and when he responds to it.

The IT teams at trading firms apply the same techniques — particularly virtualization and application acceleration — as a way to reduce the servers they operate, drive down data center costs and improve operating efficiency.

Virtualization lets capital markets firms dial back the acquisition of dedicated servers and storage capacity. They no longer purchase excess capacity to handle peak traffic volume, a practice that leaves hardware underutilized for much of the time.

With virtualization, organizations can partition a single server to run multiple OSs and apps simultaneously. Virtualization means firms can significantly reduce the servers they operate, which reduces related management costs and power and cooling needs.

---

**Typical High-performance Computing (HPC) Workflow**

Source: EMC Corporation

---

TWEET THIS!
Application optimization adds another layer of efficiency by letting companies deploy automated provisioning of compute resources in real-time to respond to demand. This not only improves a company’s ability to respond to sudden increases in trading volume, but also reduces the total cost of operations.

Winterflood Securities, a London-based liquidity provider, recently upgraded its data center architecture to a new platform specifically built to accelerate the virtualization process. The firm adopted a three-phase methodology that involves consolidation, virtualization, and automation. The result is tighter integration of servers, networks and storage systems, which has driven performance up and costs down. That’s significant, given that the firm simultaneously has seen its traffic grow, from four million messages per day in 2007 to 250 million messages per day in 2012.

**High-performance Trading Fabric**

Driven by the need to interact with clients using mobile devices such as smartphones and tablets, financial services firms have begun turning to new network-based computing environments known as high-performance trading fabrics. These emerging environments combine in-house computing resources with cloud-based applications to provide on-demand services to on-the-go traders and customers.

Proponents of the high-performance trading fabric model say it will reduce IT expenditures for capital markets firms. For example, traders will use a trading app in the cloud on an as-needed basis and pay for it accordingly. Data centers run by individual exchanges would be integrated to create a giant mesh network accessible via the cloud. The high-performance trading fabric would unify compute, switching and storage resources across data centers into a single environment capable of delivering services to users via the cloud.

In 2011, the NASDAQ Stock Market purchased a dozen high-end switches and two dozen appliances called fabric extenders to create a scalable Fibre Channel over Ethernet mesh for its options trading exchange. The aim was to reduce latency, maintain high performance and create an environment that could scale as capacity grows.

NASDAQ’s options exchange trades more than 2,600 equity options and represents 27 percent of total U.S. equity options trading. NASDAQ’s IT executives say the new networking and data center architecture helps the exchange meet its need for reliability, predictability and ultralow latency.

Vendors creating high-performance trading fabrics claim they can help capital markets firms deal with high-frequency trading, which hinges on processing speed and low latency to drive up trade profitability. Milliseconds and microseconds give trading firms a competitive edge. Furthermore, traders must execute these deals during periods of market volatility, such as when markets open or close.

---

**Technology Enablers of High-performance Trading Fabrics**

- **Ultralow latency**: Migrates delays in data transmissions
- **High throughput**: Handles extreme volatility, including sustained periods of peak traffic and microbursts, without data loss
- **Fast recoverability**: Assures built-in redundancy
- **Scalability**: Allows trading networks to grow as business or demand grows
- **Deterministic behavior**: Provides predictable system performance that adjusts to meet business needs
- **Granular visibility**: Identifies performance bottlenecks and other threats

**HPC in the Cloud**

Cloud applications represent a new computing paradigm for how financial services firms purchase and deliver IT services to their employees and customers. Cloud providers offer hosted applications in an on-demand fashion in shared, multitenant environments accessible on any device, from any location. Cloud apps lower the time to market, compared with custom programs, and they reduce total cost of ownership (TCO) because firms don’t need to maintain or upgrade them.

Financial institutions began the shift to cloud computing with private clouds inside their own data centers, which offered a way to reap the benefits of reduced operating costs while retaining security controls. Bank of America, Citibank, Merrill Lynch, Morgan Stanley and State Street Bank were in the vanguard of firms implementing private clouds.

Additionally, GE Capital and Wells Fargo Bank have begun to migrate some back-office applications to public cloud providers. So far, capital markets firms seem most confident migrating to the public cloud for non-mission-critical applications, essentially administrative services that are competition differentiators.

Until recently, the cloud seemed a less-than-ideal fit for trading applications, which require customization, advanced functionality, high performance and low latency. Instead, these apps tended to be enterprise class deployments, even if they sometimes were hosted at third-party facilities. But new high-end trading apps offer enterprise-level performance, full functionality and security with the advantages of cloud-based delivery. These software as a service (SaaS) offerings let financial firms run the trading apps without the necessity of maintaining the infrastructure.

Specialty financial offerings are also emerging, such as NYSE Technologies’ Capital Markets Community Platform, which provides a cloud trading platform with rapid provisioning, high scalability, lower operating costs and top-notch security.
Early customers include GF Futures, Millennium Management and Pico Quantitative Trading, all of which turned to the cloud platform for the flexibility and efficiency benefits.

“NYSE’s cloud customers should also be able to leverage the NYSE cloud to their own advantage, creating and expanding business–critical stickiness through their own value-added services built on top of the NYSE cloud services,” says Bruce Guptill, senior vice president and head of research at Saugatuck Technology, a Westport, Conn., research firm.

“While an initial attraction of the NYSE cloud will be cost reduction, its real power will be in generating strategic value for traders.”

Underpinning these successful cloud implementations is virtualization of servers, networks and, increasingly, desktop clients. With client virtualization, IT departments are able to deliver a user’s customized desktop, applications and data to any device — PC, tablet or smartphone — at any location, in a secure fashion.

Another related technology is cloud storage, which allows enterprises and cloud service providers to store, manage and protect unstructured content and provide it as storage as a service. When provided as a private cloud storage infrastructure, these platforms help reduce storage costs, simplify management and enable policy–based retention for regulatory compliance.

One financial services firm that’s reaping the rewards of cloud–based storage is WSFS Bank in Wilmington, Del., which is using 76 percent less storage capacity and 63 percent less data center space, and is achieving rapid failover from its cloud storage platform. Bank executives say cloud–based storage provides improved agility, scalability and cost–effectiveness. It also reduces failure at every level, more efficient data protection, high availability and robust disaster recovery. WSFS Bank reports that it’s exceeding its operational risk targets for recoverability with its new cloud–based storage platform.

Whether tapping public or private cloud implementations, financial services firms are finding that the approach allows for consolidation of scattered cluster resources, rapid service delivery and service integration. Not only does cloud computing provide faster time to market, but it also enables greater analytic insights.

State Street Bank expects to save $600 million by migrating to a private cloud service delivery model over the next two years. The bank’s CIO says savings will come both from virtualization of its client, server and storage infrastructures, as well as from increased utilization rates and automated provisioning.

### Benefits of Cloud–based Applications

One financial services firm reported key performance improvements:

- Reduced time to market for new business services by 50%
- Reduced total cost of ownership (TCO) by 60%
- Improved transaction performance by 50%
- Reduced time to deploy and change applications by 70%

Source: Next Generation Datacenters in Financial Services, by Tony Bishop (Elsevier Science, 2009)

### 3 Questions to Ask Before Embarking on a Cloud Strategy

1. **What is the right deployment model: private cloud, public cloud or industry–specific cloud?**

   Capital markets firms are adopting a mix of all three approaches. The biggest firms — think State Street and Morgan Stanley — have built private cloud infrastructures to deliver services to their employees.

   Smaller firms are choosing industry–specific clouds such as NYSE Technologies’ Capital Markets Community Platform.

   Companies of all sizes are choosing public cloud offerings for applications such as e-mail and customer relationship management (CRM), where secure, compliant software as a service (SaaS) offerings are available at a fraction of the cost of running in–house applications.

2. **What quantifiable benefits can we achieve through adopting cloud computing, and what will be the source of those benefits?**

   Firms that deploy cloud–based infrastructure and SaaS offerings are reporting measurable benefits in several areas. Cost savings result from fewer hardware purchases, reduced data center floor space, smaller electric bills and fewer IT staff required to provision and maintain systems and applications.

   Gains in agility also can be measured, as firms reduce time to market for deploying new functionality and eliminate internal software development projects. Some companies report higher availability metrics after turning to cloud–based approaches for backup and recovery.

3. **How much change will this mean for our IT organization, and can we handle it?**

   Financial services CIOs will have to rethink their approach to delivering IT services to their employees. They need a service–oriented architecture (SOA) as an overarching plan for cloud deployment. They also need a strategy for keeping information secure and for complying with relevant regulations. Finally, they need tools for managing a combination of in–house and third–party IT services, as well as skilled IT staff to oversee cloud service providers.
But State Street also projects that it will be able to lower what it spends on software development because a standardized IT infrastructure means more code can be reused. The bank aims to reduce the amount of code it has to write by 30 percent and its testing time by 30 percent. With a standardized, simplified IT architecture, State Street expects to collect more data about its customers and be able to mine it better by incorporating analytics throughout its apps.

In this way, the new private cloud infrastructure will improve the quality of service that State Street provides customers while also reducing overhead IT costs. Overall, cloud computing is expected to create significant savings for the bank, whose annual IT budget is estimated at $1 billion.

**Data Capture and Parallel Storage**

As large financial services firms turn to private clouds for their IT infrastructures, they often struggle with how best to handle the massive amounts of data that they gather from customer transactions and other sources. To address this data management challenge, firms increasingly are turning to parallel storage.

Parallel storage systems create a single pool of storage that can be accessed by the most demanding apps at extreme speeds.

Vendors create parallel storage appliances specifically for HPC environments and private clouds, where data is aggregated, IT resources are shared and capacity is provided on demand. The appliances, built on industry standards, are modular and, therefore, easy to scale. They offer easier management by eliminating islands of storage, enhanced data protection and automatic load balancing.

By combining parallel file system software with high-performance storage arrays, these appliances are capable of serving up petabytes of data at faster speeds than older network-attached storage (NAS) technology.

For capital markets firms, a major concern arises from the fact that pricing and trading data are increasing exponentially. For example, the global foreign exchange (FX) market is growing dramatically, reaching $4.7 trillion in average daily trade volume at the end of 2011, according to Aite Group, a Boston consultancy. This represents a doubling of the average daily FX trading volume since 2004, with the growth coming primarily from electronic and algorithmic trading.

“The FX asset class represents one of the largest global capital market opportunities for banks, investors and the vendor community,” Aite Group reported in May 2012. As firms jump on opportunities such as FX trading, they must struggle with the data deluge and seek the most efficient storage strategies.

One European bank is using parallel storage to power its credit risk analysis application. The bank’s goal was to provide traders with faster and more accurate risk data throughout the trading day. Manipulating that credit risk data, based on complex simulations, had outstripped the processing power of the bank’s existing storage area network (SAN).

By upgrading to a parallel storage appliance, the bank was able to increase the performance of its storage systems by several gigabytes per second, which enabled it to increase the risk scenarios it could produce a hundredfold — but in the same amount of processing time. The bank also reduced its storage costs because it no longer needed to maintain an additional database system.

**Architecting the Best HPC Solution**

In their quest for speed, capital markets firms are turning to HPC solutions because of their low latency and their ability to drive down transaction times to milliseconds. They are willing to try any emerging technology — from ultrafast processors to special-purpose appliances and network-centric fabrics — to gain a competitive edge. They are at the cutting-edge of IT industry trends, from client virtualization and parallel storage to private and public cloud deployments.

Even so, capital markets firms face challenges architecting the best HPC system possible because of their need for speed. They also seek to drive quicker ROI in terms of reduced capital expenditures and operating costs, as well as faster time to market for new applications. Indeed, they are creating a new model for global market trading that is both faster and more efficient than ever before.

State Street Bank, for example, has dubbed its private cloud project “an overall IT transformation.” While the bank estimates that it will reap savings of $600 million over the next two years, the new environment also reduces the time it takes to deploy a new app from several weeks to sometimes as little as five minutes.

The new system enhances information security by featuring federated identity management and role-based access. The bank expects that the project will impact two-thirds of its 1,000 home-grown apps, creating opportunities for more efficient business processes. With standardized components, open-source code and reusable code modules, State Street anticipates that the new platform will ease maintenance chores and be easier to upgrade in the future.

The trick, therefore, lies in mapping out existing systems, identifying the options available through the introduction of HPC products and services, and prioritizing spending based on ROI.
CDW: An HPC Partner that Gets IT

CDW understands high-performance computing and the related technologies needed by capital markets firms: low-latency trading systems, network-based trading fabrics, cloud applications and high-availability storage.

Our account managers are dedicated to capital markets and securities organizations. They solve similar issues for industry players on a daily basis. Our solution architects offer expertise in designing customized solutions, while CDW advanced technology engineers assist with implementations and long-term management of those solutions.

CDW’s high-performance computing offerings for the financial services market include technologies such as InfiniBand architecture, 10-Gigabit Ethernet switching, solid-state memory, high availability/high-volume storage, high-performance servers, power management systems and more.

To learn more about CDW’s high-performance computing solutions, contact your CDW securities and investments account manager, call 888.706.4239 or visit CDW.com/securities

In addition, offering a complete portfolio of products and services allows us to work on the entire financial industry stack. Our areas of focus include:

- **Data Storage**: Let us work with you to devise a scalable solution to meet ever-increasing network demands.
- **Servers**: We’ve formed alliances with strong technology brands to help your data center operate at peak performance – even while using limited resources.
- **Network Security**: Our brand partnerships help create the airtight, multilayer network defense you and your clients expect.
- **Telecommunications**: Discover how the latest telephony and unified communications technologies can greatly improve the efficiency and success of your firm.

---

HP’s outstanding product portfolio spans printing, computing, software, data storage, networking hardware, services and IT infrastructure. HP ProLiant servers continue to deliver on their heritage of engineering excellence with increased flexibility and performance, enterprise-class uptime and HP Insight Control manageability.

Arista Networks was founded to deliver software-defined cloud networking solutions for large data center and high-performance computing environments. Arista delivers a portfolio of 1/10/40 and 100GbE-capable products that redefine network architectures, bring extensibility to networking and dramatically change the price/performance of data center networks.

Fusion’s ioMemory platform has been built, from the ground-up, to minimize latency to make use of idle CPU time. This translates into faster processing with less scale-out to meet performance needs.