

## Financial Services: Boost Confidence in Algorithmic Trading

### Alpha Discovery Benefits

- Accelerate model development with the highest-performance open tools for machine learning, graph and Spark™
- Address massive data requirements with confidence and ease
- Ensure optimal application performance

### Backtesting Benefits

- Increase speed and scope of backtesting
- Handle larger sample sizes and faster model runs — complete jobs in seconds
- Cray addresses the I/O issue:
  - Lowest network latency available
  - Highest I/O bandwidth to reduce data load times
  - I/O acceleration technologies save money and improve performance

### Value for Financial Services Firms

- Reduce cost per simulation
- Run more models without cost increase, for declining TCO as you scale
- Make better decisions sooner
- Improve confidence in your models and trading strategy
- Discover new strategies

### About Cray

Cray provides systems and solutions that help you solve your most difficult computing, storage and data analytics challenges. The company's comprehensive portfolio includes optimized cluster systems, powerful supercomputers, advanced storage systems and high-performance data analytics and discovery platforms.



### Why Financial Firms Must Trade Smarter

High-frequency trading (HFT) now accounts for more than half of the trades on the NYSE and NASDAQ exchanges. HFT provides liquidity in the market and has become firmly entrenched as an approach that's here to stay.

However, as more firms enter this low-latency market (with all trading below the 150 microsecond threshold), profitability from pure HFT is getting squeezed. Visionary firms need to maintain speed but also employ smarter algorithmic strategies. They must also identify new trading signals from multiple big data sources in order to find and employ a sustainable alpha strategy. The opportunity exists for firms to increase profitable trade ratio points and create a sustainable strategy, *if* they can address the related technology challenges.

### Challenges of the Smarter Trading Approach

Developing smarter trading strategies requires addressing two problems: 1) finding new alpha-generating strategies, and 2) mitigating risk by backtesting those strategies with highly granular market data over long time horizons. Doing both requires taking advantage of the masses of data available today.

Smart trading may be the answer to differentiating in this high-pressure marketplace, but how do firms actually execute? They must have the right infrastructure: balanced systems for compute, analytics and storage that can scale efficiently as the data — and the trading opportunities — continue to grow.

### Answering the Alpha Discovery Challenge

As the cost of every execution microsecond increases, firms must use new data sources and apply new analytics techniques such as graph databases and machine learning as rapidly as possible.

Machine learning seeks to model the underlying real world from observed data, and deep learning is one tool for achieving it. Deep learning techniques can be used to spot trading signals from diverse, noisy data such as market data and Twitter feeds. However, training these deep or machine learning models puts high demands on a compute infrastructure. Running models can take hours or days, leaving a researcher under extreme pressure to “go live” with new strategy ideas quickly at the mercy of infrastructure performance. As a result, deep learning tools are gravitating to GPUs for their extreme parallelism.

Bringing new strategies to production quickly depends on giving researchers not only the best software tools, but also the best-performing infrastructure.

## Cray Systems for Alpha Discovery

Cray sells more GPUs than any other vendor and has delivered some of the fastest production GPU systems on the planet. The **Cray® CS-Storm™ system** is a dense, accelerated cluster supercomputer that delivers up to 329 double-precision GPU teraflops in a single rack — a space- and power-efficient solution for users who process massive amounts of data from diverse sources. The **Cray® XC™ supercomputer** is the respected market leader in high-performance, scalable systems with a mix of both CPU and GPUs, and is designed to easily accommodate upgrades and enhancements including adding coprocessor and accelerator components.

Cray systems address the heat issue that limits many GPU offerings. Unless cooled correctly, GPUs will throttle back or fail entirely. If a task is fully parallel and distributed across many GPUs, then the entire task has to wait on the slowest GPU. Even a few poorly cooled GPUs can have a major impact on job latency. And if a GPU fails, it can throw out indeterminate math results — a very serious issue in quantitative analytics. Cray has addressed this through intensive signal analytics work. Additionally, Cray systems are designed to avoid GPU throttling, ensuring maximum productivity and lowest cost per simulation.

## Cray Graph Engine for Graph Analytics

When evaluating potential strategies, traders and asset managers must be able to see how multiple factors (e.g., accounts, sectors, exchanges, market data, reference data, sentiment) compare to historical trends. Then they must determine relationships between various transactions to accurately forecast a stock's value and identify undervalued securities. In addition, traders need to be able to analyze data streams from a plethora of markets, news and social media.

Graph analytics is a discovery tool that helps researchers and analysts uncover relationships between very large, disparate and unstructured data sources. Where relational analytics typically explore relationships by

**“Cray’s Urika-GX platform allows the Smart Data based analytics of NextAngles to scale to handle the workloads of the world’s largest financial organizations. The kind of insights that currently require hundreds of man hours of work can be generated at a fraction of the effort and cost; and within minutes of the events.”**

—Suresh Nair  
Chief Architect & CTO, NextAngles

comparing “one-to-one” or perhaps even “one-to-many,” graph analytics can compare “many-to-many” and support more real-time discovery and insight. For example, an equity researcher attempting to spot undervalued stocks may have traditionally used public data such as market, 10-K or 10-Q data. Now with graph analytics, this same researcher can relate the many alternate data sources online in a graph to discover new insights.

Graph analytics are done in memory against all nodes at the same time. However, server design typically allocates memory to a single server or node for all cores and threads to access. Accessing memory outside that node is possible, but very expensive in terms of latency. Performance falls off dramatically for data that exceeds a typical node (e.g., 256 GB). The **Cray Graph Engine (CGE)** answers this problem. CGE exploits its low-latency interconnect to access memory in other nodes at near local performance. The result is the world's most scalable graph database, available on

Cray® XC™ supercomputers and Cray® Urika®-GX analytics systems.

## Answering the Backtesting Challenge

Backtesting is a critical business process highly dependent not only on dataset quality but also on the integrated analytics, compute and storage infrastructure. To improve the degree of confidence in your strategies, you need larger sample sizes and more simulations run during the same window. Unfortunately, data movement and I/O become bottlenecks, and traditional architectures cannot scale to meet increasing data requirements. In addition, an architecture's ability to rapidly move large quantities of historical data is critical to success, and the frequency of data movement will directly affect job latency. Larger sample sizes and the ability to run more simulations during the same window improves modeling confidence and lowers cost per simulation. Unfortunately, data movement and I/O to compute nodes become bottlenecks, and traditional architectures simply cannot scale to meet the increasing data requirements.

## Cray Systems for Backtesting

Cray supercomputers offer high-performance interconnect and flexible network topology, with low latency and scalable global bandwidth to satisfy the most challenging multi-petaflops applications. The Cray XC supercomputer offers unbeatable performance, with market-leading interconnect and I/O acceleration. In addition, the biggest, fastest GPU-enabled systems on the TOP500 are Cray supercomputers.

## Cray Differentiators

### High-performance supercomputing systems

- Market-leading network for optimized data movement
- Powerful application I/O acceleration to create a persistent data tier for all compute nodes, scaling to 40M IOPS or more and over 1.5 TB/s bandwidth

### Flexible, balanced and scalable clusters

- Densest available GPU system
- Powerful and highly customizable/upgradeable
- Support for CUDA and additional parallel programming models for accelerators

### Agile analytics platform

- Hadoop®/Spark and graph capabilities
- Convenience of an appliance with flexibility of a custom solution

### High-performance scale-out storage

- 38 percent more sustained performance than the competition
- 30 percent reduction in datacenter footprint of petascale systems

### Ease of implementation and reduced time to value

- Pre-integrated and tested solutions implemented in days once onsite



### Cray® DataWarp™ for I/O Acceleration

Cray's approach is to keep the most frequently accessed or persistent datasets closest to the compute nodes. By leveraging the Aries™ network, Cray supercomputing and analytics systems are able to move data in and out of those nodes at the fastest speeds possible — maximizing the number of simulations run per dollar. When running a large set of strategies against the same market data, the problem becomes too large to fit in every node's memory. The answer is to share the data on a file system where every node can access it. However, file systems vary in their I/O bandwidth. The ideal solution is to share SSDs on the network at speed, and Cray has an excellent implementation. The **Cray® DataWarp™ applications I/O accelerator** shares SSDs across all compute nodes directly on the Aries interconnect for XC supercomputer and Urika-GX analytics systems. The DataWarp accelerator delivers pure performance for I/O-hungry applications without modifications, providing 5x the performance of disk-based systems at the same cost, and offloads bursty and I/O-intensive workloads from disk-based parallel file systems.

Market data files can be striped across multiple DataWarp blades, and DataWarp performance excels when many nodes have many threads each streaming data — an example typical of a backtesting workload. In fact, in a simulated backtesting workload job latencies were reduced by 70 percent using the DataWarp accelerator as compared to a Lustre® file system.

### Cray® ClusterStor™ Lustre Storage

For backtesting involving specific slices of data, quants will reduce the amount of data they test against so that simulations complete in a reasonable amount of time. But this approach compromises strategy confidence. The answer to this problem is a storage system with I/O bandwidth high enough for the compute nodes. As a leader in open systems and parallel file systems, Cray builds on community-driven Lustre to unlock the performance of popular x86 Linux® compute clusters and supercomputers using Cray's proven storage system architecture.

The Cray ClusterStor system provides 38 percent more sustained performance than the competition in the same amount of space — achieving 98 GB/s sustained throughput and topping out at a whopping 112 GB/s per rack. The ClusterStor system's compact design also reduces the total datacenter footprint of petascale systems by up to 30 percent over competing solutions. Using less power also provides big savings, reducing the operating costs by 15 to 20 percent over monolithic SAN-based systems.

### Find Out More

#### Cray XC Supercomputing and CS Cluster Systems

[www.cray.com/products/computing](http://www.cray.com/products/computing)

#### DataWarp Applications I/O Accelerator

[www.cray.com/products/storage/datawarp](http://www.cray.com/products/storage/datawarp)

#### Cray Graph Engine

[www.cray.com/products/analytics/cray-graph-engine](http://www.cray.com/products/analytics/cray-graph-engine)

#### Cray ClusterStor Storage Systems

[www.cray.com/products/storage/clusterstor](http://www.cray.com/products/storage/clusterstor)