

HPC strategies for financial services: The playing field widens



VERNE GLOBAL

LEADERS IN HIGH PERFORMANCE COMPUTING

produced by The Realization Group in association with Verne Global

HPC strategies for financial services: The playing field widens



Introduction

The financial marketplace today is dominated by growing regulatory requirements, constantly changing liquidity conditions and increasingly complex asset-class strategies by firms across the spectrum. In such an environment, it's no longer only top-tier banks that need vast amounts of high performance computing (HPC) capabilities to handle the huge number of calculations needed for risk management, trade idea generation, compliance or a host of other functions that sell-side and buy-side organisations carry out.

These HPC capabilities are typically delivered via a grid computing environment – where clusters of computers are configured together, each with one or more multi-core processors, in order to generate large amounts of compute power.

The financial sector is known for its compute-intensive needs, although those needs do vary significantly depending on the type of firm and its market profile. Banks and brokers will typically use HPC and grid computing for tasks such as real-time risk management, as they handle thousands of transactions for clients around the world at any given moment. Proprietary trading outfits, meanwhile, often need HPC to run sophisticated models. For instance, so-called Monte Carlo simulations may involve hundreds of thousands of what-if scenarios whose numbers need to be crunched to work out the probabilities of different trade strategies paying off.

As new vendors and operational models emerge in the HPC space, banks and financial firms of all sizes are finding that HPC is essential to their strategies. Some remain focused on on-premise environments that feature “bare metal” approaches (where physical servers are not shared between tenants via virtualisation). Others are shifting a greater share of operations to the cloud, either on a bare metal or virtualised basis. Still others may be opting for hybrid approaches encompassing a combination of cloud, colocation and on-premise.

But deciding on an HPC and grid compute strategy must take account of a range of issues linked to a firm's customer base, market focus and geographical footprint. All of these factors will have a direct impact on what scope there is for different infrastructural approaches. In other words, each firm is likely to have a unique set of HPC challenges and opportunities.

The upshot is that HPC considerations now must form an essential component of any forward-thinking financial firm's strategy.

A new landscape

The ability to perform high-intensity quantitative analysis has long been recognised as a key factor for banks, brokers, hedge funds and a range of financial market participants. But as compute costs have come down and cloud-based vendor models have emerged, firms have come to realise that they can increasingly use HPC to speed up their compute processes.

Pre-trade analytics, quant analytics and risk management essentially involve stress testing against rapidly updating portfolios as well as against large amounts of historical data. What HPC and grid does, in part thanks to advancements in multi-core technology, is allow firms to take a given set of inputs and apply exponentially more conditions. Over the years the industry has moved from single-core CPUs to quad-core CPUs to cases now where 40 cores are available within a single box of traditional CPU performance.

For some proprietary trading firms, whose entire business models may be predicated on their ability to find signals amid vast amounts of noise, the importance of having large amounts of compute power cannot be overestimated. In the crudest terms, extra compute power lets firms run bigger, more complex models faster.

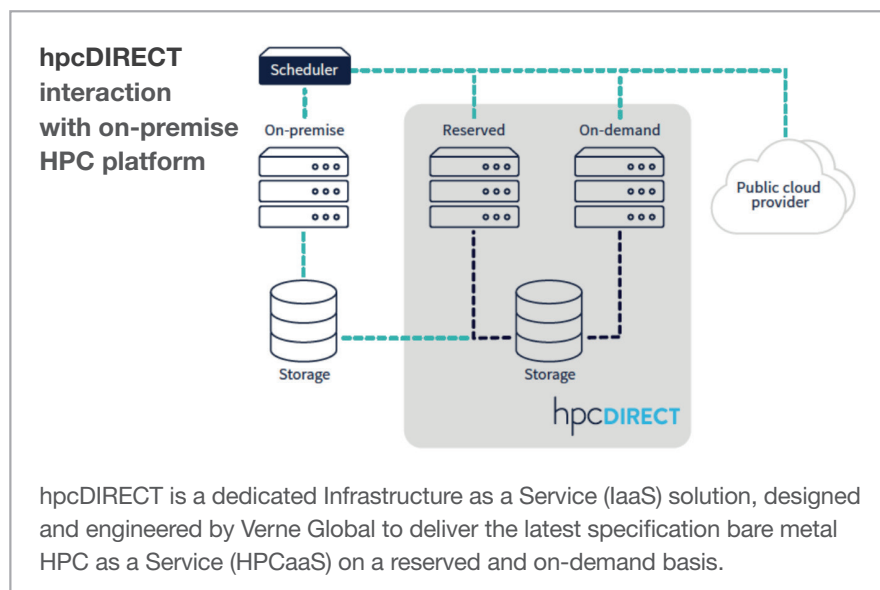
“Some of these prop trading shops have HPC environments that rival the national labs for scale. They’re just huge”

Peter Lankford, founder and director at US-based Securities Technology Analysis Center (STAC).

HPC options

HPC environments fall into five broad categories, each of which will feature different cost/benefit calculations depending on the ways firms require and use compute power.

- 1 Firms build their own on-premise HPC infrastructure:** This will be common for bigger firms that cater to large numbers of clients. Compute needs are characterised by constant high demand or corporate security requirements.
- 2 Firms co-locate infrastructure to external specialist providers:** This is popular for hedge funds and alternative investment firms, which often require long-term, low costs and high levels of scalability.
- 3 Cloud (virtualised):** This may be most appropriate for firms that sporadically need ultra-high levels of compute power for running models, and which are less sensitive to latency concerns. The virtualised environment may keep costs lower.
- 4 Bare-metal Cloud (non virtualised):** If fully optimised this allows for the full performance of the servers to be realised – what some analysts call *TrueHPC*, while taking advantage of the cost, flexibility and scalability benefit of cloud automation. An example of this is Verne Global’s hpcDIRECT solution (see diagram on page 4), which is gaining interest in the financial services sector.
- 5 Hybrid:** A combination of some or all of the above depending on applications, workloads, locations and other factors. Firms that operate in multiple markets with multiple functions may be attracted to hybrid solutions.



By many accounts, hybrid approaches are gaining in popularity. For instance, a firm may have a local on-premises facility but will have some data and workload sitting in a hosted facility elsewhere.

Also, cloud-based vendors aren't restricted to selling compute power on an as-and-when basis. Firms that know when, on a regular basis, they will need bursts of compute power, can opt for dedicated resources at set times and intervals.

“With the advent of essentially limitless capacity on public clouds and heavy workloads from regulatory reporting and complex sets of calculations in trading algorithms, it is simply uneconomic to do much of that in your own server farms”

Michael Davison, Industry Principal, Financial Services at Atos, a digital transformation specialist firm.

Factors to consider

Those firms that depend on ultra-low latency are restricted in how much they can take advantage of the cloud since most cloud environments don't offer the high-performance, low-latency networks needed to compete in the high-frequency trading space.

Virtualised environments will generally mean a loss of some performance in latency terms; but the prospect of being able to perform more jobs due to the endless capabilities of the cloud, or the attraction of lower capex costs, may outweigh any performance downside. That is especially the case for proprietary firms that are competing based more on trade strategies rather than latency. For these companies, hybrid solutions can allow them to keep their proprietary algorithms under wraps in private domains, while encrypted datasets can pass through the cloud to derive results.

Meanwhile, for those firms that operate their own HPC infrastructure, there are several physical factors that need to be considered. For instance, during the past few decades, Moore's Law has driven IT strategies as compute power essentially doubled every 18 months.

But what is different now is that physical limits are beginning to be reached. As a result, machines are increasingly parallel. That creates its own issues because there is a skills shortage for large-scale parallel machine programming.

Another factor for firms that operate their own facilities is the heat that is generated by server farms. The power and cooling costs can become prohibitive, particularly in those countries where energy is more expensive. That is pushing companies to explore the possibility of quantum computing, although that option is not yet able to be practically deployed.

... *“The approach of just putting additional compute right next to the trading engines is not financially sustainable”*

... **Tate Cantrell, Chief Technology Officer at Verne Global.**

Whether the setup is on-premise bare metal, cloud-based or hybrid, firms will still place a high premium on security. For some, that may tip the balance towards on-premises solutions because their security teams insist on it. But that view is becoming increasingly outdated as vendors that allow participants to hyperscale offer ultra-secure networks. On the other hand, if a firm's security teams require numerous mechanisms to enhance the security of cloud solutions, it could affect the price-performance proposition to such a degree as to cancel the benefits of the cloud.

Another common issue is regulation. All firms are subject to some form of regulation, although measures that are currently in the pipeline will affect some parts of the market more than others. The biggest firms are those that face the greatest amount of regulatory scrutiny. For instance, under the Fundamental Review of the Trading Book (FRTB), as called for by the Basel Committee, banks and brokers will need to make widespread changes in the way they treat securitised and non-securitised products in terms of capital requirements. They also will need to change their risk management approaches to capture more “tail risk”. All of this means extra number crunching.

Opportunities and challenges

While scalability is one of the chief attractions of cloud-based HPC solutions, time to market and flexibility are also major draws. Some vendors are now billing down to one-second intervals.

Access to the cloud for high bursts of compute power also opens up new possibilities for companies that might have had to run large simulations-based programmes, such as Monte Carlo-type models. Historically, these firms might have run their models overnight to generate a set of results that will determine the next day's trading strategy. Now, the ability to draw on as much compute power as desired means companies can run their models multiple times a day, allowing them to develop more intraday trading strategies.

One challenge for companies which outsource their compute needs is the lack of predictability of costs. Yet some providers, such as Verne Global, recognise this and are basing their offerings on stable, low-cost power sources, allowing them to offer their clients a high degree of price predictability, such as 12-year (or longer) energy pricing.

For smaller, proprietary-based firms, the ability to experiment with new compute-intensive models creates attractive business opportunities. Firms can be as nimble as they desire in terms of algorithmic innovation, time horizons and asset classes.

... *“The right HPC solution can allow firms to deploy something in a couple of weeks rather than a couple of months”*

... **Marko Djukic, Managing Director of cloud solutions provider Hentsü**

And for most of the marketplace, regulatory demands, particularly those involving massive amounts of computing to satisfy pre- and post-trade risk management, are making HPC strategy more important than ever before. The enormous grid computing networks that feature hundreds of thousands of cores may be too expensive for many firms to contemplate. But with the cloud, there are now plenty of alternatives for getting the compute muscle companies need.

Six key recommendations

- 1 Consider the regulatory pipeline and its impact on your operations.** As measures such as FRTB come on stream, firms' HPC needs could increase sharply.
- 2 Factor in the IT horizon.** GPUs, parallel computing and eventually quantum computing offer enormous potential as firms ramp up their HPC capabilities. That also means that considerations such as cooling costs need to be taken into account.
- 3 Security is a priority for all participants, regardless of size or market focus.** Consider how security needs will be met by your HPC approach and whether costly additional measures could come back to haunt your firm.
- 4 Be prepared to make trade-offs.** Cost, flexibility and scalability are all variable factors that will shift depending on the approach taken. Knowing what a firm's priorities are in this respect can make the difference in developing an effective HPC and grid compute strategy.
- 5 Widen your trading horizons.** For proprietary trading firms in particular, HPC solutions could mean faster signal generation as they can run more demanding models more frequently. That can increase the numbers and types of markets that firms participate in.
- 6 Cloud is not a silver bullet.** Hyperscale cloud providers offering HPC have their place, but are not a panacea. When considering them, be aware of data transfer costs, custom application incompatibilities with platforms, and a lack of specialist HPC infrastructure or support.

The Realization Group would like to thank the following for contributing their insights to this article: Tate Cantrell of Verne Global, Michael Davison and Andy Grant of Atos, Marko Djukic of Hentsü, Peter Lankford of STAC and Matt Barrett of Adaptive.

For more information on Verne Global visit **www.verneglobal.com**

For more information, specification and pricing on hpcDIRECT for Financial Services and Grid Compute: **www.verneglobal.com/solutions/hpcdirect/financial-services-and-grid-compute**