HPC in the cloud

Is this you?

Do you have a sprawling legacy grid platform in need of modernisation? Are you under-using expensive infrastructure in mutually incompatible silos? Do you face an ever-increasing need for higher HPC workloads that are hard to predict and costly to deliver to the business? If so, read on!

Working with GFT to deploy high performance computing (HPC) workloads to the cloud lets you deliver computing power to the business at the exact time when it's needed and gives you the ability to react flexibly to uncertain markets and increasing regulation.

Many financial institutions are seeing ever-increasing demands for more compute power to deal with incoming regulation such as FRTB, increased business volumes, and high volatility market events such as Brexit, COVID-19 and financial instability. Laser-focused on cost, our banking and insurance clients are looking to streamline or retire ageing server fleets, recover expensive data centre space, remove on-premise operational complexity whilst bringing their internal grid platforms right up to date.

Outside of financial services, HPC is used in diverse fields such as engineering, life sciences, weather prediction, etc. The use cases may be different from finance, but the technological challenges in deploying HPC to the cloud are very similar.

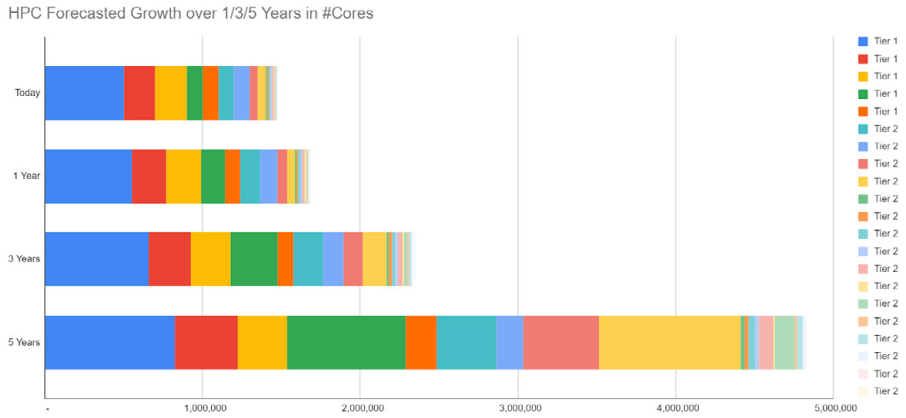
THE ‘WHY?’

Riding the HPC escalator

More than ever, accurate timely risk calculations for the desk and the regulator have a direct impact on the bottom line and are a major business enabler. The competitive advantage of precise risk and better capital management drives compute growth. With compute demands projected to exceed hundreds of thousands of cores, it is increasingly untenable to try to maintain traditional on-premise grid environments. Moving grid computing to the cloud can reduce operational complexity, dramatically improve scalability, resilience and agility, whilst offering bottom-line savings through the use of heavily discounted bulk and “spot/ pre-emptible” cloud compute pricing.

Over the past decade, the importance of risk management within financial institutions has risen dramatically, as have the prevalence and sophistication of the tools that service it. High performance computing (HPC) grids in the cloud have enabled financial institutions to dramatically increase the efficiency of their front office and risk calculations, e.g., reducing capital requirements and facilitating intraday risk measures. Cloud computing now offers scalability and flexibility to HPC grids, and the ability to move away from inflexible and restrictive on-premise architectures.

HPC forecasted growth over one, three and five years amongst tier 1 and 2 banks

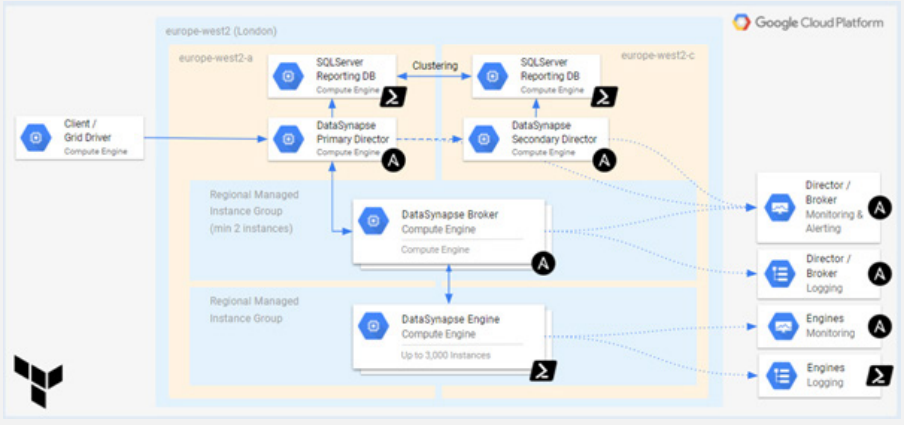


THE ‘HOW?’

GFT HPC accelerator

GFT has extensive experience with HPC deployment, modernisation and operations in tier 1 banks, both on-premise and in the cloud. We can help optimise and enhance on-premise environments, build new automated environments in the cloud or architect hybrid solutions with both on-premise and cloud components. Partnering with the major public cloud vendors GCP, AWS and Azure and with grid middleware vendors such as TIBCO, GFT’s engineers fuse cutting-edge DevOps tooling and methodologies with solid traditional infrastructure management skills.

Example of a highly resilient scalable GridServer implementation



Getting started

GFT has made substantial investments in tools and people and has a strong HPC practice. Our HPC acceleration tool can create TIBCO GridServer environments in minutes, run compute at scale, then tear them down when the workloads are complete. Whatever your cloud ambitions and the sale of you application requirements, we can work with you to **assess, plan** and **implement** your HPC cloud migration every step of the way.

Evaluation workshops

We would conduct an interactive workshop with the right stakeholders in your organisation to develop an understanding of your requirements and challenges, explore the marketplace for HPC on the cloud and demonstrate GFT's HPC accelerators.

Statement of work

Based on the results of the evaluation workshop, we would work with your teams to develop, prepare and provide a statement of work to help you progress to achieve your objectives. This could include:

- Functional POC cloud-only or hybrid HPC deployment to benchmark against your on-premise infrastructure to show proof of value
- Pilot HPC environment to migrate and test selected workloads
- Full migration planning for production HPC workloads
- Training, consultation and support.

Our credentials

GFT has a deep pool of expertise across banking and financial services in general and HPC in particular. Our team contains architects with hands-on operational experience of some of the industry's largest grid platforms, developers with experience of many risk platforms and a strong DevOps function. Our business-focused consultants have delivered key digital and cloud transformation projects to many major financial institutions around the globe.

Our experts

Scott Holman
Head of Google Cloud, GFT
scott.holman@gft.com

Vimmy Patel
UK Partnership Manager, GFT
vimmy.patel@gft.com

About GFT – Shaping the future of digital business

blog.gft.com
twitter.com/gft_en
linkedin.com/company/gft-group

GFT Technologies SE is a global technology partner focused on digital transformation. Founded in 1987, we have over 10,000 specialists in 15 countries.

gft.com

GFT › SUCCESS STORY

HSBC Equities/FI/FX migration to cloud

One of Google Cloud Platform's tier 1 banking clients wanted to migrate its on-premise Equities VaR risk processing workload using TIBCO GridServer onto GCP. This was a particularly urgent need due to a combination of a pressing hardware and software refresh requirement, as well as strong demand for increased compute capacity due to impending new financial regulations. Like many such on-premise environments, the Equities division was facing an increasingly high cost for its hardware estate and inefficient usage due to fragmentation across application silos.

Aside from Equities, HSBC also used GridServer for running HPC workloads for Fixed Income, FX Options, Market Risk and Credit Risk. GFT was engaged by Google to develop a deployable reference architecture for running GridServer on GCP to be used in the Equities migration, which was then to be reused by the other asset classes.

GFT deployed a full team consisting of an HPC cloud architect/project manager, three DevOps cloud engineers, and four cloud application developers to the project for nine months. We designed a hybrid cloud model for the deployable reference architecture and created a backlog in order to deliver the project in an agile fashion, with weekly sprints and 'show and tells'.

The backlog included 'user stories' to create a basic GridServer GCP deployment; automated using Terraform, Ansible and Powershell. We incorporated networking and security best practices, implemented autoscaling functionality using both CPU metrics and a custom metric reading the performance data from GridServer. Other features of the deployment included full encryption using Customer Managed Encryption Keys (CMEK), integration of logging and monitoring activity with Stackdriver, and implementation of DevOps practices to deploy code payloads using a Jenkins CI/CD pipeline.

The reference architecture is now available and has been used to successfully migrate Equities workloads to GCP. Subsequently, the Fixed Income (FI) and Foreign Exchange (FX) businesses have also used the same reference architecture to migrate their workloads to GCP.

There is further interest from other areas in the bank to use the reference architecture in the future.

The main benefits delivered by the reference architecture are the ability to 'stand up' and 'tear down' a full GridServer environment at the touch of a button, parameterisable engine types and initial engine pool sizes, enabling cost-optimisation, quicker time-to-market for new HPC projects, ability to economically run a full-scale UAT and to easily support exceptional workloads such as VaR re-runs. The reference architecture can be easily modified to support AWS, IBM Spectrum Symphony, etc.

From a business perspective, the client has made substantial in-year savings on internal infrastructure costs and has been able to scale their grid environment as needed to meet extraordinary demands for additional risk calculations due to market uncertainty around COVID-19 and Brexit. Their ongoing running costs are now around 60% of their equivalent for on-premise infrastructure with greater flexibility and direct ownership of their grids by the application owners.

GFT › SUCCESS STORY

Cloud-native transformation of Credit Risk workloads

GFT's tier 1 banking client needed to update its overnight credit risk processing architecture to meet deadlines associated with the incoming FRTB / Basel III regulation and to facilitate intraday credit risk calculations. The client's existing credit risk HPC infrastructure was running on-premise and was reliant on vendor supplied risk valuation models (MarkIT Quic), commercially licensed HPC software running Microsoft Windows Server as the base operating system for the grid compute nodes, and another set of commercially licensed software for the "What If" simulation functionality (IBM WebSphere).

In a truly ground-breaking project, GFT and the client designed a completely new cloud-native GCP architecture from the ground up, with the vendor risk valuation models being replaced by in-house models built on open source QuantLib runnable on open source Linux (replacing both the expensive Quic models and the expensive fleet of Windows Server operating system licenses). In the new architecture, Google Cloud Dataflow takes on the role of scheduling the HPC workloads (running a PaaS model), running Apache Beam open-source pipelines at scale. This enabled the client to decommission a significant number of expensive HPC grid scheduler licenses, which would have required more hands-on maintenance and support if deployed to the cloud in an IaaS model. The interactive "What If" functionality was also migrated from a monolithic application architecture running on licensed IBM WebSphere software to a microservices-based application architecture deployed to Google Kubernetes Engine (GKE) on GCP.

The entire project was fully delivered in a time span of less than 12 months, starting off with a 2 month proof of concept phase, followed by a full development, UAT, and production delivery for one European and South East Asia jurisdiction by calendar year end. After the initial production go-live, the client initiated a global rollout programme for the rest of its global jurisdictions (more than 30 in total). As this included jurisdictions that don't allow data to be stored in cloud, an on-premise Flink version of the cloud-native GCP architecture had to be designed and implemented for a limited number of the credit risk workloads.

The project saved the bank several million USD through a more flexible and scalable technology stack, which was also cheaper to run, and by achieving a reduction in capital requirements due to faster, better and more accurate delivery of credit risk metrics.