MARKET COMMENTARY

Higher Performance and More Collaboration for Quants

Regulation is transforming the way banks operate. On one hand, new requirements are tremendously increasing both real-time and overnight calculations, generating more IT costs. On the other, banks are required to have more capital reserve, so scrutiny on spending. Such a problem is not easy to solve.

Recently, I had an interesting conversation with an executive at a top tier European bank. Our chat was about their ongoing Fundamental Review of the Trading Book (FRTB) project. They expect the impact of FRTB, which comes into force on 1 January 2020, to require 32 times more calculations than what they currently have to calculate their overnight Value-at-Risk (VAR).

Front-to-risk consistency, brought by the FRTB Internal Model Approach (IMA), is forcing banks to reconsider how their current quant platforms are designed and managed as they try to bring down silos and ensure consistency across departments. This means banks are now considering disruptive technology to achieve the agility required by the front office, the performance required by their risk department, and, of course, temper their overall IT costs.

On a recent trip to Asia, I met another executive from a major Asian bank. He told me his bank needs to perform Credit Valuation Adjustment (CVA) calculations on their book within six hours. We crunched the numbers and quickly realized we were talking about billions of Present Value (PV) calculations to be generated, meaning dozens of millions of PVs to be computed per second.

The number of calculations needed is exploding. Meanwhile, the timeframe during which they need to be executed remains the same as before: either overnight or intraday. Often the only way to cope with this new additional load seems to be throwing more hardware at the problem. The total bill runs high and is talked about in millions of dollars. In the context of reduced margins and higher capital charges, this makes for difficult spending decisions, especially when Return on Investment (ROI) remains to be proven.
Front-to-risk consistency, brought by the FRTB Internal Model Approach (IMA), is forcing banks to reconsider how their current quant platforms are designed and managed."

Since 2012, Finastra has invested in solutions to cope with real-time computational performance. We have seen the emergence of multi-cores and Single Instruction Multiple Data (SIMD) platforms. We made the decision to build a framework that can leverage vectorization while hiding the technical complexity. The result is FusionFabric Parallel Processing. It allows Quants to focus on the algorithm, written in standard scripts, letting the platform taking care of the conversion of those scripts on the fly into vectorized code.

Encouraged by positive feedback from our customers, we are now pushing the concept further. We are moving this platform to the cloud to foster more collaboration with Finastra’s Platform-as-a-Service (PaaS) solution, FusionFabric.cloud.

With FusionFabric.cloud Valuation, Finastra offers banks, academics, and Fintechs a secured and scalable cloud platform designed from the ground up for the finance industry. With a complete Web Integrated Development Environment (Web IDE or WIDE), Quants can write codes sitting next to traders. They can immediately test and validate the results. Building on a collaborative approach, Quants can now work with banks’ various teams and, if they want to, work in association with Quants from other banks or academics.

The generated code is automatically parallelized, freeing up Quants from the burden of IT constraints and letting them focus instead on the algorithm. The framework not only converts the code into OpenCL or CUDA to leverage SIMD but it also takes care of data management to ensure coalescence of data (memory alignment to ensure maximum vector effect). This adds to create a seamless and transparent experience for Quants.

Finastra is working with schools and universities to develop courses supported by FusionFabric.cloud. The platform is open to students and academics and give them an entry point in to the financial services ecosystem. Testing can be done using WIDE, or using the Microsoft Excel plug-in that allows the management of scenarios, market data, and trade data with the convenience of a spreadsheet.

If we look at performance, which is the key driver underpinning FusionFabric.cloud, figures speak for themselves: about 50 million prices per second for swaps and 60,000 prices per second for callable bonds. These pricing figures exclude post-processing times.

Such performance has been achieved on regular instances from cloud providers such as Microsoft Azure or Amazon Web Services. Here we used the Amazon EC2 P2 instances, using eight of the 16 NVIDIA GPUs available.
FusionFabric.cloud uses CUDA and is able to fully harvest the full benefits of NVIDIA GPUs. In our benchmark, we used NVIDIA Tesla K80, and we are now looking forward to testing the new Pascal and Volta architectures for even higher results. “With the strong performance advantage that our GPU computing platform offers, financial institutions around the world are turning to NVIDIA to accelerate their most important applications,” said Andy Steinbach, Head of Financial Services, NVIDIA. “This new cloud offering from Finastra is a great example of how a company is using our platform to offer a potentially game-changing advantage for the financial industry.” FusionFabric.cloud Valuation delivers performance and collaboration “as-a-service”. It delivers multiple benefits to banks to achieve regulatory compliance and to lower operating costs. It also gives financial institutions a competitive advantage to release new products quickly, reducing time to market and allowing banks to give their full attention to agility and innovation. Combined with the full suite of tools offered by FusionFabric.cloud, both students and developers can create powerful and innovative apps whether they are in academia, are employed by financial institutions, or work for Fintechs.