



# LOW LATENCY DATA DISTRIBUTION IN CAPITAL MARKETS: **GETTING IT RIGHT**

---

**PATRICK KUSTER**

Head of Business Development, Enterprise Capabilities, Thomson Reuters  
+358 (40) 840 7788; [patrick.kuster@thomsonreuters.com](mailto:patrick.kuster@thomsonreuters.com)



THOMSON REUTERS™

---

## LOW LATENCY DATA DISTRIBUTION IN CAPITAL MARKETS: GETTING IT RIGHT

---

In seeking the most effective solution for low latency access to data from multiple venues, financial firms are challenged by the fact that technology has moved on, but the way they accumulate data has not. The platforms built in response to regulatory requirements to deliver best execution are now nearly a decade old and firms are looking for a faster way to execute on their clients' behalf.

---

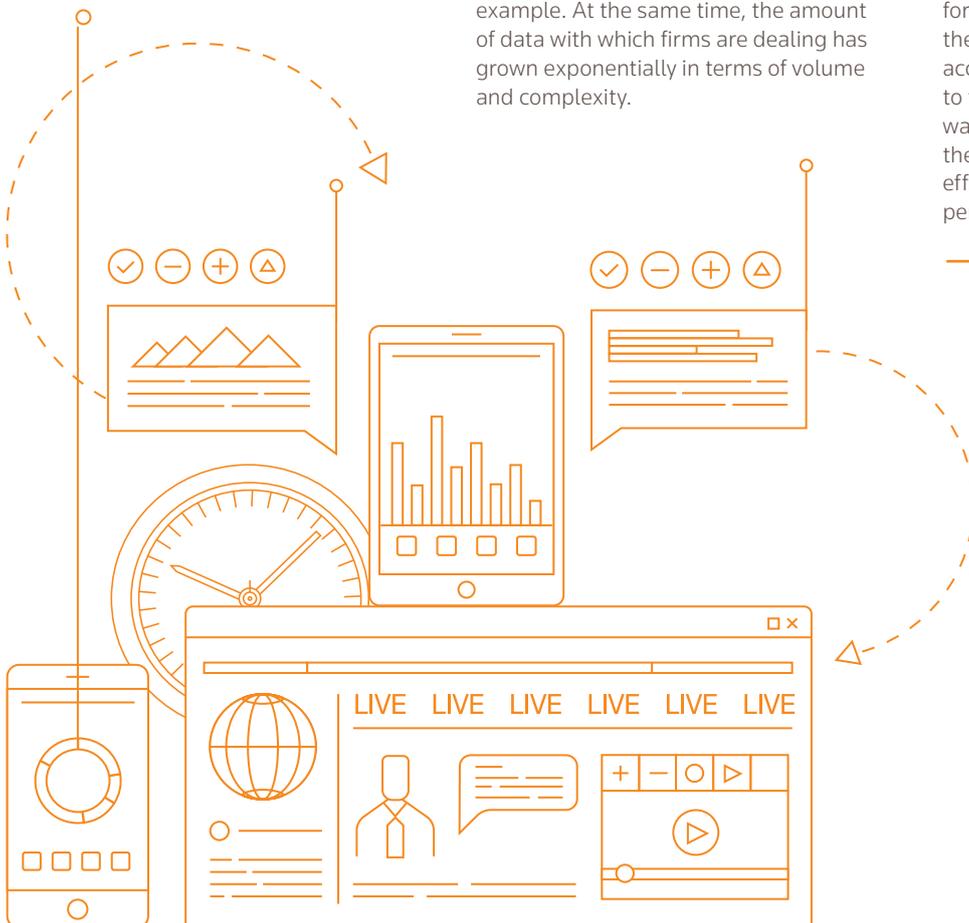
### A THIRD WAY

---

During the past ten years, technology options have become broader and less costly –single core processors have been replaced with multi-core processors, for example. At the same time, the amount of data with which firms are dealing has grown exponentially in terms of volume and complexity.

Today, firms require high-performance access to real-time market data sourced directly from individual trading venues in a standardised, reliable and flexible format. Recently, a debate has arisen about the virtues of embracing either hardware acceleration or software as a solution to these requirements. However, a third way – a hybrid approach that harnesses the value of both – is in our view the most effective way of delivering the required performance.

---

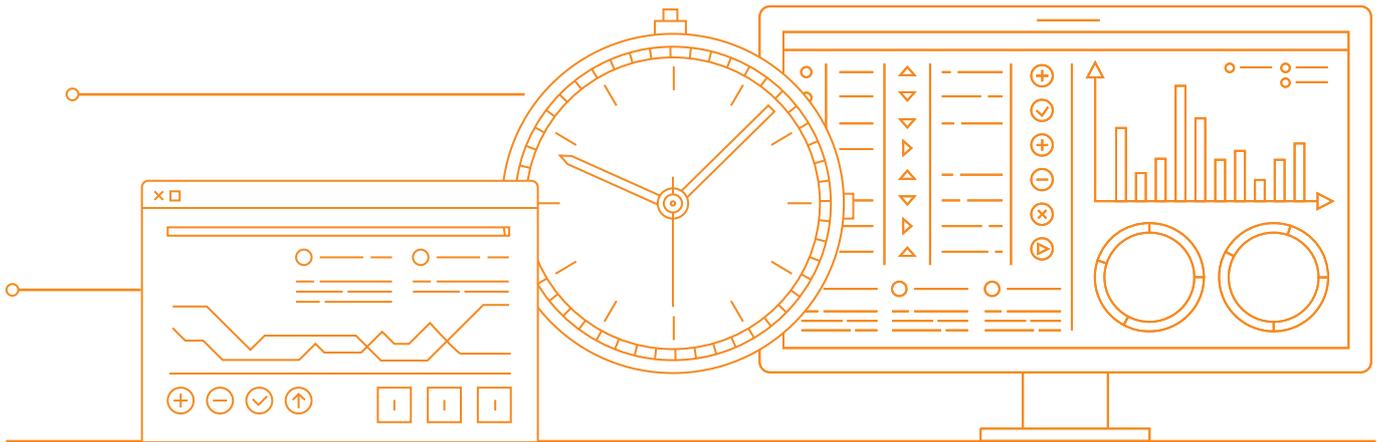


---

## LOW LATENCY DATA DISTRIBUTION IN CAPITAL MARKETS: GETTING IT RIGHT

---

### HARDWARE ACCELERATION VS SOFTWARE: FALSE DICHOTOMY?



---

### FPGA HAS GAINED MUCH ATTENTION IN FINANCIAL SERVICES

---

Hardware acceleration – the use of computer hardware to perform functions faster than is possible in software running on a general purpose CPU – delivers low latency performance. Of the hardware acceleration techniques, field-programmable gate array (FPGA) technology has gained much attention in financial services. FPGA is an integrated circuit designed to be configured by a user or a designer after manufacturing. The circuits can undertake complex digital computations rapidly. FPGA is best at processing streams of data quickly and deterministically – meaning users can accurately predict how long a process will take. The technology delivers parallel and concurrent resources so that users can deploy it, for example, for five or 500 tasks and the latency will be the same.

The FPGA technology ensures that the latency from the venue to the receiving application is deterministic. You can be sure that the application is always a fixed amount of time “behind” real time, which is important since it affects the probability that a particular order/quote will have been filled by the time the app sees it. This allows trading applications to accurately predict the impact that a trade will have had on market conditions. This is important for the performance of an algorithm, as well as for risk management. For example Market Makers have quotes that can be exposed when the market moves, so they need to know how current their market data is to take appropriate action and set their spread appropriately.

Both software and hardware approach the problem of increase in data rates through parallel processing, however software does allow a much more dynamic and flexible environment to work in. The flexibility and ease of programming of software are the main benefits. A challenge with the software approach is that the underlying server is designed for general purpose computing and therefore the software tool is competing with everything else that runs on that server. When the market is very busy and has high liquidity, latency tends to grow with a software solution.

# LOW LATENCY DATA DISTRIBUTION IN CAPITAL MARKETS: GETTING IT RIGHT

## IMPROVEMENTS IN FPGA

One of the most important factors that is making FPGA a key technology in financial services is the growth in size of the FPGA chips. The FPGA is made up of a fixed set of logic cells, which the developer will configure into a certain arrangement to carry out the required task. The more logic cells that are available on an FPGA, the more complex operations are possible on a single chip.

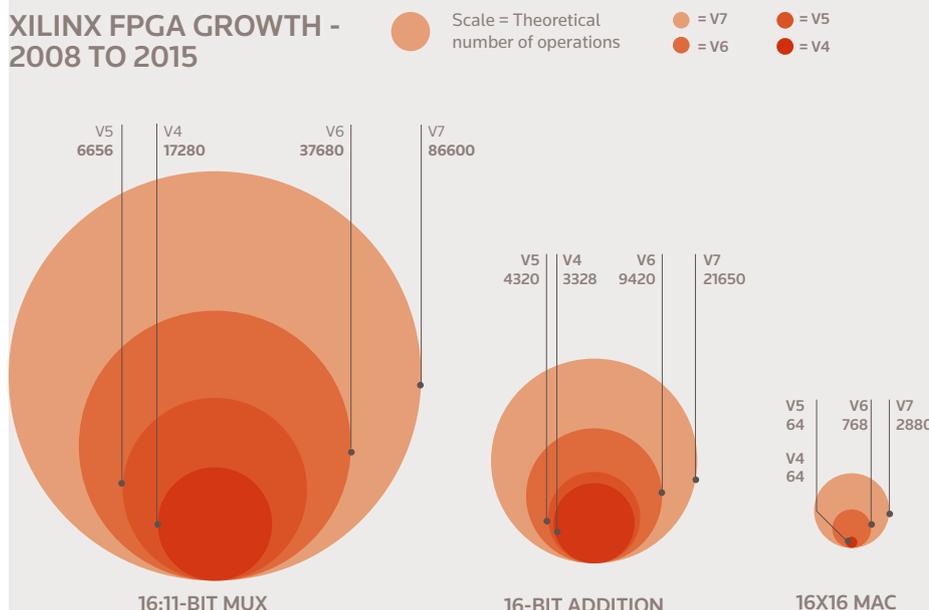
If we look at the past seven years, we can see a huge increase in the size of these chips available to the market. For example, the Xilinx FPGA cards on which the Thomson Reuters Elektron Direct Feed is based demonstrate huge increases in both computation power and storage capabilities.

THEORETICAL MAXIMUM TIMES OPERATION CAN BE CARRIED OUT ON THE FPGA

FAMILY	DEVICE	LOGIC CELLS (THOUSANDS)	BRAM (KB)	16X16 MAC	16-BIT ADDITION	16:11-BIT MUX
V4	XC4VLX60	60	360	64	3328	6656
V5	XC5VLX110T	110	666	64	4320	17280
V6	XC6VLX340T	240	1872	768	9420	37680
V7	XC7VX690T	690	5310	2880	21650	86600

- The available Logic Cells on each generation of card. This is an approximate “equivalent gate count” on the FPGA itself.
- The available BRAM on each generation – this is the directly accessible memory available to the FPGA device.
- The 16x16 MAC is a 16 bit x 16 bit multiply and accumulate function, implemented in special DSP blocks rather than in the generic FPGA fabric.
- 16 bit addition is a simple binary addition of two 16 bit numbers.
- 16:1 1-bit mux is a multiplexer that selects one bit from 16 values (this is a basic hardware function).

### XILINX FPGA GROWTH - 2008 TO 2015



For basic operations, the available space on the FPGA has increased by more than tenfold during the past seven years. This will likely continue as the chips continue to reduce in size. In practice this means that there is a huge opportunity to use these cards for more complex operations and therefore bring the deterministic latency benefits of the hardware to more complex data sets.

---

## LOW LATENCY DATA DISTRIBUTION IN CAPITAL MARKETS: GETTING IT RIGHT

---

### THE BENEFITS OF A HYBRID SYSTEM

Hardware acceleration is good for some tasks but not others – the same can be said of software tools. Therefore a hybrid solution that combines hardware acceleration to achieve low latency and software optimisation for recovery and static data mapping will help firms to achieve higher standards of best execution. This approach ensures that customers receive fast and robust real-time data feeds, which can be installed and integrated into their applications quickly and easily.



**HARDWARE FOR  
SIMPLE TASKS,  
SOFTWARE FOR  
MORE COMPLEX  
TASKS**

The determinism of FPGA is most useful for data feed handlers when providing real-time updates from data feeds. For other activities such as recovering logic, handling static data etc software is more appropriate.

However, once a firm wants to run complex algorithms for trading strategies software programs come into their own. Simple, repetitive tasks can be handled by hardware accelerators, with software performing the more analytical, complex activities.

As a general rule, the simpler a task is, the more appropriate it is to use hardware acceleration. Along a scale of complexity, the indicator will move towards software the more complex, dynamic or changeable a task becomes. If the market data feed is unbundled from other tasks and handled via hardware acceleration, firms will achieve deterministic latency.

---

# LOW LATENCY DATA DISTRIBUTION IN CAPITAL MARKETS: GETTING IT RIGHT

---

## HARDWARE AND SOFTWARE USE CASES

---

**Ed Rainer, Group Product Manager, Real-Time Information Services, London Stock Exchange Group, says that FPGA technology coupled with a high performance, scalable software chassis was a natural progression for LSEG's big data technology solutions.**

"Innovation is at the core of everything we do at LSEG and through the research and development teams at MillenniumIT we've got all of the tools to bring an exciting technology offering to market.

The behaviour of the Group's varied customer base - Member Firms, Information Vendors and Network Service Providers - continues to be influenced by our product's performance and in particular the latency of our trading and information product suite. Feedback consistently indicates that constant performance and latency is what's important to our customers. This is a key benefit of an FPGA implementation - deterministic, predictable performance and a scalable product build.

Our Group Ticker Plant, a product designed to standardise real-time data publication across Group trading venues, irrespective of the trading technology employed, leverages the Group's FPGA research and promotes it to production environments for the very first time. The platform FPGAs disseminate latency-sensitive data from our matching engines such as order book changes, executions and trading sessions, directly to customers in IP multicast. This happens in parallel with more complex processing such as statistics, reference data management and TCP-based data recovery being powered by a traditional software chassis."

Phase 2 Group Ticker Plant services, including full-depth order books powered by FPGA, are available across Group test environments now. Further detail on the programme can be found at: [www.londonstockexchange.com/oneproductinfiniteness](http://www.londonstockexchange.com/oneproductinfiniteness).

---

## A KEY BENEFIT OF AN FPGA IMPLEMENTATION - DETERMINISTIC, PREDICTABLE PERFORMANCE...

Software is best deployed in scenarios where large amounts of stored data are required, and the analysis of this data needs to change on a regular basis. For example, a firm may be using a solution that analyses the past 30 years of news headlines for specific pieces of information, and based on those results will adjust the program to look for new information. Due to the large amount of data needed, and the need for quick evolution of the search and analysis criteria, software would make a natural fit.

---

On the other hand, FPGA and other hardware accelerators are best used for constant, predictable data. An example of this may be an equities or derivatives data feed from an exchange or other venue. The feed may change only twice or three times a year, so there is some degree of flexibility required to cope with these changes. Hardware accelerators can parse the information very efficiently, particularly for venues that provide a binary, fixed-length message. The predictability dovetails very well with the determinism that FPGA offers.

Another scenario is that of an internal processing system, such as a large data set, database or capture system, the format of which is controlled by the trading firm. The requirement here is to move a large amount of data from one location to the other, using a standard translation.

---

# LOW LATENCY DATA DISTRIBUTION IN CAPITAL MARKETS: GETTING IT RIGHT

---

## DRIVERS OF FPGA

---

### A MASSIVE INCREASE IN DATA RATES

---

One of the key drivers of FPGA use in recent years is the massive increase in data rates. This has been combined with an increase in the cost of data centre space (due largely to the cost of cooling). At a time when exchanges have moved to co-location strategies and trading firms pay a premium for being as close to the exchange servers as possible, space is costly. When you add up the rack space itself, the increased costs of power and cooling needed for modern data centres, there is a large cost of ownership for every server deployed. Putting a large number of servers running software in a location is no longer efficient. Hardware accelerators enable firms to do more with less kit. Configurable, multicast FPGA technology for real-time data processing means that multiple direct feeds can be supported on a single server, reducing the footprint of what is needed to process data. As a result, a firm's infrastructure and operating costs can be reduced.

FPGA also efficiently handles microbursts – huge spikes of data delivered by exchanges in very short periods of time. Determinism enables hardware accelerators to process data as it comes in, at the rate it comes in. Software cannot deal with microbursts as efficiently and will typically smooth the data out, buffering the information and thus delaying its delivery.

When running software systems with multiple cores each one can be dedicated to a single process. A handoff is then required to the network interface, then on to a system that publishes the data out to multiple users on the network that want to consume the data. If distribution is done out of FPGA, these handoffs are removed as all steps are handled by the FPGA card.

As already stated, FPGA is not always appropriate for all data feed activities. If an activity requires the storage of large volumes of data, software and storage on a server will deliver better results than FPGA. Also, low-volume data feeds with low throughput of data are more efficiently processed via software than via hardware accelerators. Software is cheaper to develop than FPGA therefore feeds that are low-volume or are not contingent on having very low latency are better handled via software.

---

Benjamin Stephens, Head of Latency Sensitive Electronic Trading at Instinet Europe, says there were two main reasons for choosing hardware acceleration: deterministic latency and a reduction in the firm's hardware footprint. "Deterministic latency allows trading algorithms to perform more consistently and also reduces performance penalties during times of heavy trading and high data volumes. Additionally, the cost of co-location space is high and hardware accelerators enabled us to significantly reduce the hardware footprint we needed to run a large number of venues in co-location sites. For example, we could run a US equity smart order router on one server with hardware acceleration. This delivered large savings and a good performance benefit."

As explained, hardware accelerators are not always the best fit. Stephens says for advanced calculations or complex data or when the logic required is constantly changing, software is more appropriate.

---

---

### HARDWARE ACCELERATION - DETERMINISTIC LATENCY AND REDUCTION IN HARDWARE FOOTPRINT

---

---

## LOW LATENCY DATA DISTRIBUTION IN CAPITAL MARKETS: GETTING IT RIGHT

---

### CONCLUSION

The deployment of a hybrid approach to low latency data access can be complex; a thorough understanding of when FPGA can be used versus when software should be considered is key.

---



In today's markets, electronic market makers and brokers are no longer achieving the price realisation they were ten years ago. In looking for faster ways of executing on their clients' behalf, they realise that they must balance this need for speed with cost.

Using hardware acceleration to strengthen software-based feed handlers is becoming more common, even among market vendors. The technology underpinning

hardware accelerators is complex and firms that are not on the bleeding edge of latency or that don't have very specialised business models should seek out partners to help them develop hybrid approaches.

Partnering with companies that are developing combined software and hardware direct feed solutions for multiple clients on a scale basis, will enable financial firms to cost effectively and reliably gain access to ever-increasing volumes of market data.

**USING HARDWARE  
ACCELERATION  
TO STRENGTHEN  
SOFTWARE-BASED  
FEED HANDLERS  
IS BECOMING  
MORE COMMON**

---

#### **PATRICK KUSTER**

Head of Business Development, Enterprise Capabilities, Thomson Reuters  
+358 (40) 840 7788; [patrick.kuster@thomsonreuters.com](mailto:patrick.kuster@thomsonreuters.com)  
Visit [thomsonreuters.com](http://thomsonreuters.com)