Inside Market Data

May 2010 www.insidemarketdata.com

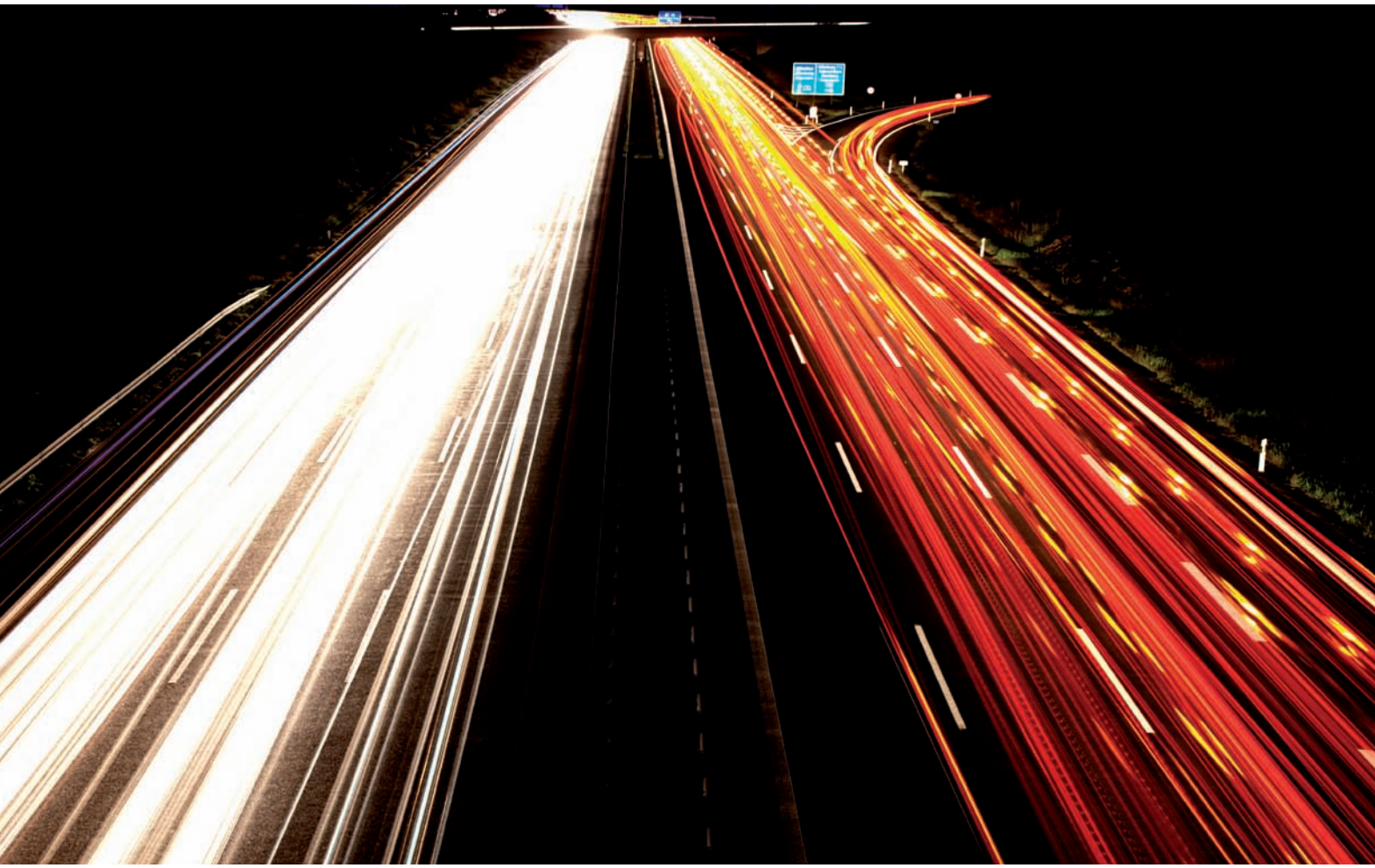
LATENCY

SPECIAL REPORT

Sponsored by:

Interactive Data

ROUNDTABLE



The Long Road to Zero Latency

Despite the financial crisis, trading firms have continued to invest heavily in low-latency data solutions to support algorithmic and high-frequency trading, reducing latency levels even further, but requiring firms to invest in new technologies to get the most out of the data. But how are firms handling low-latency feeds of other types of content, and what impact will regulators' investigation of market structure have on latency?

IMD: How have the levels of latency changed over the past year, and what impact have the financial crisis and volatile markets had on firms' ability—and desire—to invest in low-latency solutions?

Lawrence Hansen, director of product management for market data, Lime Brokerage: Overall latency for many market participants has continued to decrease due to technological enhancements across the various segments of the client-venue communication path—for both data and trading. The combination of the financial crisis and volatile markets have brought with them an even greater need for speed across all types of strategies, not just in traditional high-frequency trading strategies such as virtual market making. In today's environment, statistical arbitrage, exchange-traded fund arbitrage, algorithmic trading and options trading among others, require faster market signaling, more robust processing speed and near-immediate dispatch to venues to stay ahead of the queue and ensure the profitability of these strategies. Volatility and increased competition bring with them ever-more fleeting opportunities, so strategies with longer time horizons now require the same low latency as microsecond-level strategy horizons. Every firm that trades in the market owes it to themselves and their investors to reevaluate the mechanics and latency of their trading activity. The financial crisis has had the direct impact of compelling firms to more closely evaluate where both macro and micro levels of latency exist in their infrastructure, and the impact on how they wish to trade, rather

"The combination of the financial crisis and volatile markets have brought with them an even greater need for speed across all types of strategies, not just in traditional high-frequency trading strategies such as virtual market making."

Lawrence Hansen, director of product management for market data, Lime Brokerage

than simply rushing into an expensive infrastructure overhaul. If building the low latency to compete is too costly, especially in light of the current fiscal constraints, there are providers in the marketplace such as Lime Brokerage who have made the necessary investments in infrastructure which these firms can quickly and easily tap without adding too much to the bottom line.

Donal Byrne, chief executive, Corvil: In general, we have seen a significant overall reduction in latency levels within the trading environment. One to two years ago, everyone was talking about milliseconds. Today it is microseconds. Tomorrow, it will be nanoseconds. Investment in low-latency technologies seems to be one of the few areas in banks where growth was experienced [over the past year]. In many respects, the existence of highly volatile markets was positive for many of the trading strategies used in high-frequency trading.

Mark Skalabrin, chief executive, Redline Trading Solutions: Only a few years ago, low-latency market data systems were considered state of the art if they were processing data in less than a few hundred microseconds. Trading technology has rapidly improved over the past couple of years to where consistent single-digit microsecond performance is required for latency sensitive applications such as smart order-routing and high-frequency trading. In addition, the number of other applications where latency improvements increase profitability has expanded as market infrastructure has improved. These infrastructure changes have come from exchanges and technology providers aggressively competing to deliver better latency performance. What we have seen is that after the shock of the financial crisis wore off, many firms aggressively moved to improving latency with a clear objective to look outside their firms for leading-edge solutions.

Adam Honoré, senior analyst, Aite Group: Latency has been an arms race that has not really slowed down through the crisis. Aite Group has seen consistent technology spend for firms that require the lowest latency, despite budget cuts in other areas. Actually, because of the financial crisis and market conditions, I would argue that latency has become a two-tier battle. First, to get faster. Second, to do more processing without getting slower. That second tier is driven largely by increasing sophistication of trading styles. Lastly, now that markets have returned, firms are very focused on expanding the battle beyond local markets and traditional asset classes. Expect Asia as a geography and foreign as an asset class to grow in the next year.

Adam Honoré
Aite Group

Sinan Baskan, senior director, global financial services industry solutions, Sybase: In the wake of the financial crisis, financial firms' love affair with low-latency technology solutions continues to grow even stronger. The economic turmoil has given firms reason to reflect, fine-tune and accelerate plans to deliver low-latency solutions to the algorithmic/high-frequency trading community. Firms rely on technology for capabilities ranging from high-frequency trading to risk management to derivatives pricing. If anything, that dependence will only grow deeper in the years to come as companies and regulators seek technological answers to portfolio and systemic risk.

As complex event processing technology continues to advance, levels of latency continue to decrease. Some algorithmic trading infrastructures boast 400 or 500 microsecond response times, resulting in near-zero latency. That's about as close to "real-time" data as you can get.

John Heflin, senior vice president, global operations, Interactive Data 7Ticks: We have noticed that latency levels have dropped in almost every part of the technology stack: the telecom circuits, hardware, and software layers. The financial crisis hasn't necessarily led to a desire to invest in low-latency solutions, but we have noticed that it has fueled a desire to control costs, and we've seen an uptick in demand as a result of that. We also believe that volatility equates to trading opportunity. If there's a lot of opportunity, firms want to get from price to order quickly, and that can drive the need for low-latency solutions.

Steven Sadoff, chief information officer, Knight Capital: Market data volumes have continued to follow a Moore's Law-type curve, and most marketplaces are continuing to invest in

the arms race to lower latencies in a similar fashion. However, I don't believe the financial crisis had any significant impact with respect to this race to zero.

IMD: How can firms take advantage of the current environment to gain a latency advantage over their rivals?

Heflin: Firms can take advantage of the current environment across the stack. At the hardware layer, network hardware vendors seem to always be coming out with new or improved products that can improve performance. In addition, hardware accelerators such as FPGAs and kernel bypass cards are continuing to reduce latency. At the software layer, there have been improvements in areas such as messaging and ISVs that seem to be constantly improving the latency profiles of their core software processing. For firms not using co-location, the telecommunications industry also seems to be constantly improving the latency of its circuit routes. Regardless of where you are in the technology stack, everyone looks to be racing to get as close as possible to zero latency.

Interactive Data

John Heflin
SVP, global operations
Interactive Data 7Ticks
john.heflin@interactivedata.com
www.interactivedata7ticks.com

Hansen: The current environment has given rise to a whole host of vendor utility solutions that focus on specific pieces or segments in addressing the drive towards lower latency among buy-side and sell-side participants, execution venues and related service providers. Many of these vendors are new entrants taking advantage of current demand but without applying the appropriate scientific rigor or having an established track record to address these issues effectively. Others have been focusing on the issue for a number of years preceding the recent buzz on the topic. These firms typically offer monitoring solutions or other critical functionality to only a portion of the end-to-end low-latency stream for both market data and trading. At Lime Brokerage, we constantly evaluate the market versus what we have engineered internally, but as an agency broker and high-throughput, low-latency technology provider we have implemented an end-to-end solution which we provide to our clients. In recognizing the current financial and regulatory environment, we have brought to market products such as our Citrius market data solutions, and our LimeInside co-located sponsored access solution to drive our own and thus our clients' latency to the lowest levels attainable.

> **"The challenge today is that many competing rivals are all doing the same thing, and therefore the latency advantage that one may have had 12 months ago has eroded significantly.... As costs excalate, we believe that information about speed is where the next level of competitive advantage lies."**
>
> Donal Byrne, CEO, Corvil

Skalabrin: Today the fastest solutions exploit co-location, hardware acceleration, low-latency switching, and dedicated network connectivity. For most firms, it is not possible to make evolutionary changes to their existing solutions to get the desired benefits from these technologies. Achieving competitive advantage from latency requires addressing each stage of the trading pipeline with solutions that are designed from the ground up to perform at market-leading levels. Fortunately, there is a growing base of leading-edge solution vendors such as Redline that enable firms to quickly replace legacy approaches to achieve best in class performance.

Byrne: The challenge today is that many competing rivals are all doing the same thing, and therefore the latency advantage that one may have had 12 months ago has eroded significantly. Most serious players employ co-location, direct market feeds, and new generation algorithmic trading platforms to maximize their execution speed. However, as costs escalate, we believe that information about speed is where that next level of competitive advantage lies. Those with the best quality and most accurate information on latency will be able to use it to optimize trading decisions while avoiding the infinite cost trap inherent in the race to zero.

Honoré: Technology focused on speed only gets you so far. You also have to build good models, have a deep understanding of the market microstructure, and keep your eye on the future so you don't box yourself in with an inflexible platform. For instance, if you want to add unstructured data to your model or add new asset classes, will what you build to support today's trading opportunities be able to provide for tomorrow's opportunities?

Sadoff: I'm not sure if the current environment changes the classic formula for success. It comes down to getting the best and brightest people, investing in the latest and greatest hardware, and creating a productive team environment.

Baskan: In today's markets, there is an exponential cost associated with every microsecond saved. We may be approaching

the point where costs outweigh return. This has led market leaders to look in other directions. One direction is to be smarter, and so we see more very high-performance pre-trade analytics. Another direction is to news analytics and analysis of a wider range of data, some structured and some unstructured. It is in these areas that the next latency wars may take place—how much deep analysis can you do to identify trading opportunities that today's ultra-low latency applications can't detect?

IMD: Where do the highest levels of latency still exist, and what will the industry focus on this year to reduce latency in these and other areas?

Byrne: In our experience there is no single element within the trading loop, which is the standout latency bottleneck point for all environments. Each customer environment has its own challenges. In one deployment it might be a problem with a feed handler, and in another it might be the FIX gateway or any other myriad of problems that might arise. The focus on latency reduction is a function of the type of environment and business at play. If one is co-located, the network will hopefully be less of a problem. However, as people continue to expand their trading strategies to multiple locations then latency between co-location centers becomes a key issue to consider.

Heflin: It varies greatly, depending on the overall trading solution. In some instances, the trading application is a good place to look. For a firm running a third-party ISV in a co-location situation, the highest latency would more likely exist within the trading application, since the firm's server is sitting right next to the exchange's matching engine. However, firms that have optimized their application may look toward hardware acceleration. In addition, the telecommunications circuit is the area of highest opportunity in many instances where firms are not using co-location services.

Skalabrin: The highest latency events for most co-lo deployed systems are the direct result of traffic bursts. Most trading applications do alright on average but fall far behind when one or more exchanges saturate the network with market data. While these bursts are typically short, they can result in delays of tens to hundreds of milliseconds if not handled properly. Redline has attacked this issue through hardware-accelerated solutions that deliver the bandwidth to consume bursts without delay. Better end-to-end and point-to-point latency measurements with time-accurate traffic patterns are a key industry focus, and are needed to understand and improve this and other similar sources of latency.

Honoré: From what we see, improperly configured networks and bad code seem to be the latency frontrunners. That said, expect to see more hardware acceleration, more solid-state disk drives, and more in-memory database usage. Also look for commodity hardware upgrades based on progress made by Intel and AMD.

Lawrence Hansen
Lime Brokerage

Hansen: The war on latency is being fought on two fronts—internally and externally. On the internal front, firms are realizing that their own network infrastructure—along with the actual physical location of their trading and market data infrastructure—is typically a significant handicap in lowering their latency. Advanced hardware, messaging infrastructure and co-location will continue to be a primary focus for the industry this year. Firms that optimize these will put themselves in pole position to use and connect to advanced low-latency market data and trading applications offered by the market. On the external front, there is a significant latency disparity among and between the execution venues, the market data disseminated from each, as well as the ways that data is consolidated. Trading venues are making a competitive effort in their quest to demonstrate their value with respect to execution and latency by providing step-by-step latency measurements on flows into and out of the venues, and providing them as a value-added service. This is a clear sign that firms are becoming more focused on latency across all fronts. Specifically you cannot simply pawn off the issue of latency on the exchanges, as they are all better prepared to show their numbers to you and to your clients.

Sadoff: Typically, the longer an asset class has been trading electronically, the lower the latency. That ends up being a function of the volume, the number of participants, and the cumulative investments applied toward the ecosystem. For instance, when we acquired our foreign exchange ECN in 2006, unlike most equity or futures marketplaces, it didn't have an event-based feed. So in January 2008 we introduced Hotspot FX ITCH, and now all of our latency-sensitive customers use this feed. We did the same for our fixed income ECN, Knight BondPoint. At the end of last year we finished

the re-architecture of our market data feeds in a similar fashion. Although all of our fixed income clients are on the new feed, at the request of a few clients, we have turned on throttling for their specific feeds because their infrastructure can't handle the transmission rates of our new ticker plant.

IMD: How do the challenges around sourcing, delivering and processing low-latency, machine-readable news differ from other kinds of data?

Honoré: First, the consistency of sentiment is not nearly as high as the consistency of something like an economic indicator from a trading behavior perspective. Second, some high-performance databases used for testing trading strategies don't support unstructured data. Third, you have to be careful about trusted sources. One bad piece of information can cost quite a bit of money or affect the market—ask Apple and United Airlines.

Heflin: Although the content is different than real-time market price and trade data, the core challenges generally remain the same—getting information from point A to point B. However, an important difference is that there are a lot fewer off-the-shelf solutions that support machine-readable news, compared to traditional electronic trading and market data.

"There is an incremental cost associated with every microsecond saved. We may be approaching the point where costs outweigh return. This has led market leaders to look in other directions."

Sinan Baskan, senior director, global financial services industry solutions, Sybase

Sadoff: Although similar in many regards on the delivery front, news still lags in terms of models to accurately interpret the data. Natural language processing was one of the first goals when the field of artificial intelligence research was founded in the late 1950s, and although it has advanced, the consensus is still that a computer won't be able to pass the Turing Test anytime in this decade. Hence, only a small percentage of firms seem to be using this data in sophisticated fashions above and beyond simply throttling back models when news is published that is relevant to the securities being traded.

Byrne: One of the biggest challenges is that machine-readable news has to be distributed over large geographies and therefore must solve the problem of precision time management. Measuring and understanding precisely the time at which the news is sourced and when it is received is critically important. Millisecond precision on a global scale is required today, and this is likely to become sub-millisecond in the not-so-distant future.

Baskan: With traditional relational database technology, queries run against static data to find information or to summarize or analyze data. While those queries may seem fast in some environments, it doesn't provide the continuous and immediate insight needed within capital markets. Complex event processing sends incoming data through a set of pre-defined continuous queries to immediately process it in real time.

SYBASE®

Sinan Baskan
Senior Director, Global Financial Services Industry Solutions
Sybase
Tel: +1 212 596 1150
www.sybase.com/capitalmarkets

The process of sentiment analysis—and other analytics applied to machine-readable news—is an order of magnitude more complicated that dealing with streaming market data. Sentiment engines, such as RavenPack, Need to Know News, and others, can provide low-latency results, but unfortunately, these are available to all subscribers, so it is incumbent on any trading firm that wants to develop a strategic edge to take in multiple streams of raw and analyzed data, apply proprietary logic across a large set of structured and unstructured data, and make decisions. In our opinion, a combination of complex event processing technology (data cleansing, aggregation, keyword identification, etc.)—coupled with a large in-memory analytics capability—is key.

Hansen: Machine-readable news has not yet gained major inroads within the high-frequency trading segment in a significant way, but is gaining traction with certain types of tailored strategies. Recent products are now enabling real-time analysis of the relevance of news events while providing a variety of related metrics and heatmap-style displays in consistent formats. As with traditional data, quality, consistency and speed are critical, but given the subjective nature of machine-readable news, additional refinements will be necessary to make this a viable source for high-speed trading strategies. Vendors apply their own specific techniques to filter and ensure a consistent quality of what comes through, but it is still just subjective evidence of a trend. While Lime can accommodate its rise as firms tailor their algorithms to use these new sources in the search for alpha, any trading decision from machine-readable news requires the ability to execute as expected, and this will continue to be a function of high-quality, low-latency market data and market access.

IMD: What impact will any potential regulation of high-frequency trading and co-location have on the data industry?

Sadoff: It depends on how onerous the regulation. For co-location specifically, I haven't heard anything on the regulatory front that will have a significant impact. On the other hand, if

Steven Sadoff
Knight Capital

they outlaw high-frequency trading, impose a transaction tax or take some similar approach, it could be a significant setback. It has never been a better time to be an investor than right now, especially in markets with significant technology advances like equities. Turning the clock back will not just hurt the high-frequency community, but it will also harm other segments of the marketplace, including the retail investor.

Baskan: There are different possible regulation scenarios, and more than one scenario may apply. Firstly, there may be limits on speed of execution, to slow the markets down so regulators can monitor them in real-time. We think this type of regulation is unworkable and unlikely—trading will simply move to another jurisdiction where low-latency trading is allowed. Another option is to make trading strategies undergo approval before they are implemented. Again, it seems laudable to want to understand the "thinking," but it is probably not feasible. In the end, we think the issue is more about the viability of the products than the buying and selling mechanism—the crisis of 2008 was due to faulty products and ignorant buyers rather than electronic trading. Even the latest 1,000 point drop in the markets was quickly corrected, and (as of writing) was caused initially by human error.

Skalabrin: It is a little difficult to predict, given the spotlight on high-frequency trading and the wide-ranging potential regulations being discussed. What seems most likely at this point is that there will be increased requirements on risk management, whether accomplished through sponsored access or otherwise. While these types of changes will have some affect on the industry's structure, they should not materially change how trading applications perform.

Byrne: While it is still very early days with respect to regulation, it does seem that improved latency transparency is an area gaining increased traction with regulatory authorities. The publication and reporting of latency performance by market centers in a transparent and universally accessible manner seems to be preferred over and above ideas that might look to handicap the speed of the underlying trading systems. This will likely create further debate and discussion concerning the area of latency compliance reporting.

Heflin: It's inevitable that technology will play a key role in helping firms comply with regulation and support increasing market data volumes regardless of what regulations are created. As long as electronic trading exists, it is logical to conclude that technology will be involved in supporting and extending firms' capabilities.

Hansen: Regardless of the specific aspects of new regulations forthcoming on trading and co-location, the overall impact on the market data industry has been very positive thus far. The current attention in the industry to disparities in the mechanics of market data, its consistency and latency measurement standards, have brought forward an effort among the best providers to self-regulate prior to any forthcoming mandates to do so. Industry efforts such as STAC have been well embraced and have caught the attention of providers and consumers in regards to the standardization of latency measurements. We welcome the industry-wide education that such efforts have produced in helping us promote the perspective of the entire trading cycle value chain, and not just the value of a utility or appliance in a specific segment.

"[Regulators] turning the clock back will not just hurt the high-frequency community, but it will also harm other segments of the marketplace, including the retail investor."

Steven Sadoff, chief information officer, Knight Capital

Honoré: We hear a lot of noise from regulators and politicians about being concerned for the long-term investor. If they were really concerned about long-term investors, they probably wouldn't worry too much about low-latency trading. Unfortunately, low-latency trading and high-frequency trading have blended together in our vocabulary. High-frequency traders are generally proprietary firms or market makers. Low-latency trading can be done by anybody looking for better execution with access to technology. Both leverage the same basic technology stack. Senator Ted Kaufman issued some guidelines in March urging regulators to take a more active stance when it comes to technical advantages like co-location.

I think the net impact of the execution technology arms race to the average retail investor is positive, even if they lose best execution to an electronic trading firm every time they trade. According to the Investment Company Institute, the median amount invested by investing households in mutual funds is $100,000. My research into online brokerage habits indicates the average retail investor trades equities once per month. The average number of shares traded per execution across the market is about 240 shares. Instead of costing investors, innovation in low-latency trading saves them $471 per year if the retail investor loses a penny per share. Even if investors lose 10 cents per share on every trade, they still make $212 more per year than they would in a theoretical level playing field. We don't need to legislate a cap-and-trade-style competitive disadvantage in our local markets based largely on misunderstanding and little on fact. Personally, I hope my fund manager beats me in the market every time I trade. If I ever find out they are not winning best execution against me, I'll start investing in different funds. ■