WHITE PAPER

Optimizing Electronic FX Trading Performance

Summary

Greater use of electronic execution has brought substantial benefits to the FX trading industry, but has also created new challenges. Trading venues and participants alike operate in a more diverse and competitive marketplace, in which both the volume and velocity of trading have increased. Successfully navigating these waters requires a differentiated strategy, utilizing carefully managed technology investments to ensure competitiveness while avoiding excessive costs. In this paper we outline how best-in-class streaming analytics leveraging network data can help firms achieve these objectives. The key is to deploy a data-driven approach to position and optimize platforms, manage customer experience, and interact with key trading partners and stakeholders. We describe key requirements that firms should look for when establishing their IT Operations Framework, and streaming analytics best practices that implementers can employ to successfully deliver on these requirements.

Introduction

Electronic trading now accounts for a substantial percentage of the FX markets and is widely recognized to have brought significant benefits to the industry. The market place is more diverse and competitive than it was in the days when it relied solely on voice trading and was dominated by a relatively small number of inter- dealer brokers. Liquidity is easier to find and transaction costs are lower. As a consequence, the market now supports a broader range of business models and trading strategies, and offers greater price and execution transparency than in the past.

The changes engendered by electronic trading have also created new challenges for FX trading firms. Traditional barriers to entry in many sectors of the industry have been lowered or eliminated, implying a greater need to differentiate offerings in a more competitive marketplace. The shift to e-trading requires continual investment in computing and communications technology capable of handling ever-greater trading speeds and volumes. The emergence of new high-frequency FX trading strategies has led to concerns about speed requirements and latency arbitrage. At the same time, high profile 'glitch' incidents in other markets have highlighted the need to manage advanced technology safely. While these changes have been taking place, government regulators have also become more active in seeking detailed oversight of trade execution processes.

These challenges and opportunities have important implications for how electronic FX trading firms manage the health and performance of their high-performance technology platforms. Traditionally, the approach has been to use simple system monitoring to assist internal operations, often using coarse measures of system load, performance, and availability, with partial records of processing activity. Greater scrutiny of speed, safety, technology cost and compliance means that firms are now turning to more real-time and authoritative sources of data to characterize what's happening in the trading infrastructure. Today's forward-looking companies are leveraging streaming analytics platforms that go far beyond networking monitoring to

address today's IT Operations challenges on several different fronts. By arming you with the accurate data you need to demonstrate the superior speed and performance of your platform, today's streaming data analytics platforms help build client confidence and establish your firm's position as a technology leader. Real-time streaming analytics also generates detailed records of system activity that you can use to address current and upcoming market regulations. It assists client support by supplying the data needed to rapidly troubleshoot customer problems at both the technical and trading levels, and it improves client retention by proactively identifying poor customer experience.

Steaming Analytics from network data that can also be correlated with other data sources is also a vital discipline for mitigating technology risks, especially in complex environments where multiple systems distributed across different firms interact with each other at high speed. It helps identify 'latent' or hidden problems that could contribute to a major incident if left uncorrected; it acts as the last line of defense for detecting and intervening

Diagram 1: 5 use cases for monitoring

against risks that escape from automated controls; and it provides data for post-incident reconstruction and analysis efforts needed for future prevention.

Given the high costs often associated with low- latency technologies, firms need to take a data- driven approach to determine how fast they need to be in order to achieve their business objectives, and to identify those areas where investment will produce meaningful returns. Streaming analytics should be helping firms to determine, in real time, where performance has an impact on trading and where it does not, and to select the right service providers and technology partners to achieve their commercial goals.

These objectives constitute an ambitious program for technology to do more to help FX trading firms address the strategic challenges of modern electronic execution. In this paper we outline some of the key requirements that today's modern streaming analytics platforms should meet, and best practices that they can employ, to deliver on this program successfully.

Key Requirements for Monitoring Electronic FX Trading

Detailed accurate information

The breakneck speed of electronic trading and the fact that trading activity typically occurs in bursts lasting much less than a second imply that network analytics over short time increments is necessary to keep track of what's going on. For example, to understand system load and capacity, it's vital to have visibility into short bursts of messages that must be processed in milliseconds or less. It's also often necessary in trading environments to be able to reconstruct what happened to particular transactions or orders, for example to respond to a client query about why an order traded in the way that it did. Regulatory requirements are also becoming steadily more stringent in terms of the level of detail specified for captured data, with regulators generally seeking to be able to reconstruct in full the timing and sequencing of trade activity (see e.g. the CFTC's proposed rule 309 for SEFs). Thus streaming analytics from network data must provide real-time, detailed, precise and accurate information via high frequency measurements and accurately time-stamped per-message data capture.

Results available in real-time

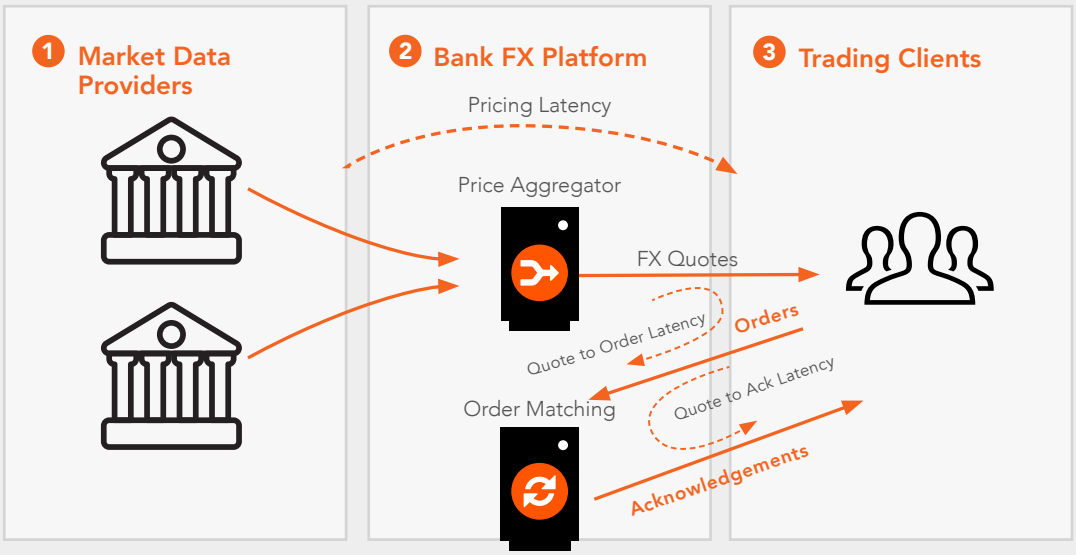
The speed of trading also means that large financial sums can potentially be placed at risk very quickly when things go wrong. Streaming analytics platforms need to react instantly to any detected faults or anomalous conditions, which means that results must be available in real time rather than being computed post-facto from stored information. A store-first-compute-later approach risks falling far behind real time during periods when systems are especially busy, which is often when up-to-date visibility is most required. Note that yesterday's 'real-time' monitoring has often meant alerts and dashboards that update anywhere from every 30 seconds to every 5 minutes. These time-scales are not sufficient for electronic trading, where participants should aim for time-scales of seconds or less.

Non-intrusive with minimal system impact

The impact of collecting this information on trading system performance must, however, be kept to an absolute minimum, both to maintain a competitive latency profile and to avoid the possible introduction of new system failure modes. Where avoidable, monitoring should not rely on trading system components to either generate or forward monitoring information. Sticking to this principle also helps avoid the problem of 'opt-in self reporting', i.e. relying on system components to self-diagnose and report faults. (It is desirable that components should report faults when they can, but potentially dangerous to rely solely on a common standard of reporting from components created by diverse teams possibly spread across different organizations).

A multi-layered network/application/ trading view of system actvity

Trading is a distributed activity that relies on applications and networks working together to achieve a business outcome. Results can be impacted by problems arising at either the network or application layer, and therefore streaming analytics platforms must provide visibility into both. This means keeping track of transaction-handling and message-processing within software components, to ensure correct and fast results. It also means watching out for any congestion, delays or data loss that may occur within network stacks, in a firm's internal networks, or over links connecting



1. Incoming market-data feeds are monitored for any gaps and for upstream latency based on feed timestamps, and to detect which copy of A/B feeds are faster.
2. Wire-to-Wire latency across the pricing engine is measured, providing a direct measurement of the quoting speed.
3. Client order flow is monitored, both the business layer (dollar value traded across currency pairs) and the performance (quote/response and order/ack) per client.

remote sites and external parties. Streaming analytics platforms are most effective when business results (such as fill-rates and trading volumes) can be visualized side-by-side with infrastructure metrics and KPIs. This makes sure that business-impacting problems are prioritized and diagnosed quickly. The result is a multi-layered network/application/trading view of system activity.

Comprehensive latency and performance metrics

Traders worry about latency impact and indeed there are numerous latency factors that can affect trading performance, ranging from poor order-execution times resulting in slippage, to slow pricing engine updates leaving traders open to arbitrage. If your sources of market data are slow, then your systems may already be behind when the starting pistol fires for an attractive trading opportunity. And even if internal systems are lightning fast, your clients may still suffer if their network connectivity is poor. Streaming analytics platforms should provide comprehensive metrics, in real time, that reveal the different latency factors present in your environment and help you to identify and eliminate the important bottlenecks in both internal and external systems. Achieving this requires an ability to analyze both quote and order message flow, to track the causal relationships between different stages of transaction processing, and to provide hop-by-hop latency breakdowns across applications and networks.

Transparent KPIs that facilitate communication with clients and partners

The data and insights generated by streaming analytics platforms can help build client confidence in your trading platform, and assist partners and service providers seeking to troubleshoot problems or improve their offerings. To be useful for these purposes the data must be clear and unambiguous, and presented in a format that is commonly understood across industry participants. Avoid debates about what key metrics mean, and whether they are affected by internal or external factors, by being transparent about where and how measurements are made. Be clear about how the raw data is processed to generate summary statistics where applicable. Ideally, streaming analytics should maintain traceability from KPI values back to the details of the transactions from which they were generated.

Streaming Analytics: Best Practices

Analyze the perimeter

The perimeter of a trading system – where it sends and receives quotes and orders to/from the outside world – is a good place to get an unambiguous view of system performance. By capturing data here you can create an authoritative record of interactions with clients and partners. Performance measurements made at the perimeter – ideally at the network links that provide external connectivity - enable insight into 'firm-level' latency, incorporating all potential sources of internal latency including network equipment, operating system stack, and application processing. Latency across execution venues, and across client systems and network connections, can also be assessed unambiguously by perimeter monitoring. So can the quality

and integrity of inbound market data feeds. These measurements are invaluable for troubleshooting and characterizing system performance to clients and other stakeholders. Compliance and client support use cases also benefit from access to definitive records of orders and messages entering and leaving the system that allow the timing and sequencing of trading activity to be accurately reconstructed.

Implement a purpose-built streaming data analytics platform and capability

The need to collect and analyze large amounts of detailed data in real-time, without disturbing trading system performance (and without necessarily depending on system implementers to provide all of the required inputs) points to a stand-alone streaming data analytics platform with its own compute and storage resources as the preferred solution. Not only does this ensure minimal trading impact, it also guarantees that the platform provides an independent analysis of system activity and doesn't share common assumptions or points of failure with system components.

Consider the accuracy of time-stamping

The accurate time-stamping of events within high speed trading systems is a topic that deserves careful consideration. Timestamps are needed to accurately and with precision measure the latency between related events, such as receiving and processing a client order. They are also needed when reconstructing the sequence of events involved in a particular transaction or pattern of trading activity. The time differences between events of interest in modern trading can range from just a few microseconds (transmission across a single system component such as a network switch, for example), to tens or hundreds of microseconds (end-to-end processing of a client order), or milliseconds (communication between remote systems across a WAN).

A streaming analytics platform for low-latency high-volume trading will require sub-microsecond time-stamping precision

The required accuracy of time-stamping is therefore governed by latency performance management goals, and by the pace at which orders are received, processed, and transmitted. A streaming analytics platform for low-latency high-volume trading will require sub-microsecond time-stamping precision to support measurement of very short latencies and correct sequencing of hundreds of thousands of events per second. This level of precision depends on hardware time-stamping support because software timestamp accuracy is limited to tens of microseconds at best, due to operating system effects such as scheduling and buffering.

When timestamps are made using different clocks, perhaps located in different places, the issue of clock synchronization must also be considered. Accurate latency measurements can be made using clocks that are not externally synchronized by tracking their offset relative to each other using carefully designed network signals. However, reconstructing the sequence of events relative to wall-clock time requires

some level of synchronization to an external time source. Several technology options are available for this purpose, differing in scalability, complexity of implementation, and achievable accuracy. Network Time Protocol (NTP) can synchronize systems across a plain IP network and achieves millisecond level accuracy at best – that is, when running across an uncongested LAN. Precision Time Protocol (PTP) can achieve sub-microsecond accuracy but requires special support in network switches and end devices.

Finally time-stamping systems can be synchronized directly to a co-located GPS master clock using low-level Pulse Per Second (PPS) signals running over dedicated RF cables. This approach has limited scalability, but provides superb accuracy in the tens of nanoseconds range. The correct choice of technology for your environment will depend on how distributed your system is and the speed and volume of events that you intend to monitor.

Use network data to power application and trading layer analytics and performance management

Perimeter analytics, accurate time-stamping, and the deployment of dedicated streaming data analytics resources are all facilitated when data from the network is used as a direct primary input to the real-time analytics platform. Network data on external connections can be time-stamped in hardware, either by direct tapping or using switches from vendors such as Arista, Gigamon, IXIA, Cisco and VSS Monitoring forwarded to a dedicated system for real-time analysis and subsequent storage. Data from the network is very detailed and foundational in nature, but can be reassembled into message flows between trading system components and then decoded to reveal its business level content. This approach, now made available through modern streaming analytics platforms, uniquely supports analysis and detection at all three levels encompassing network communications, application-layer behavior, and trading results.

Make streaming analytics a source of data, not just a consumer

As envisaged in this document, modern, real-time streaming analytics platforms support multiple use cases ranging from latency and performance optimization, to client support, risk mitigation and compliance verification. The set of data and analytics generated to drive these use cases is a rich and detailed reflection of activity within the trading platform and its interactions with external parties. As such, it has the potential to power further applications such as Transaction Cost Analysis and Customer Experience Management. In our experience, the results of a dedicated streaming data analytics platform are often used for additional purposes that weren't initially foreseen. Plan to maximize the value by making its results accessible to other systems and users within your organization via appropriate easy-to-use APIs and interfaces.

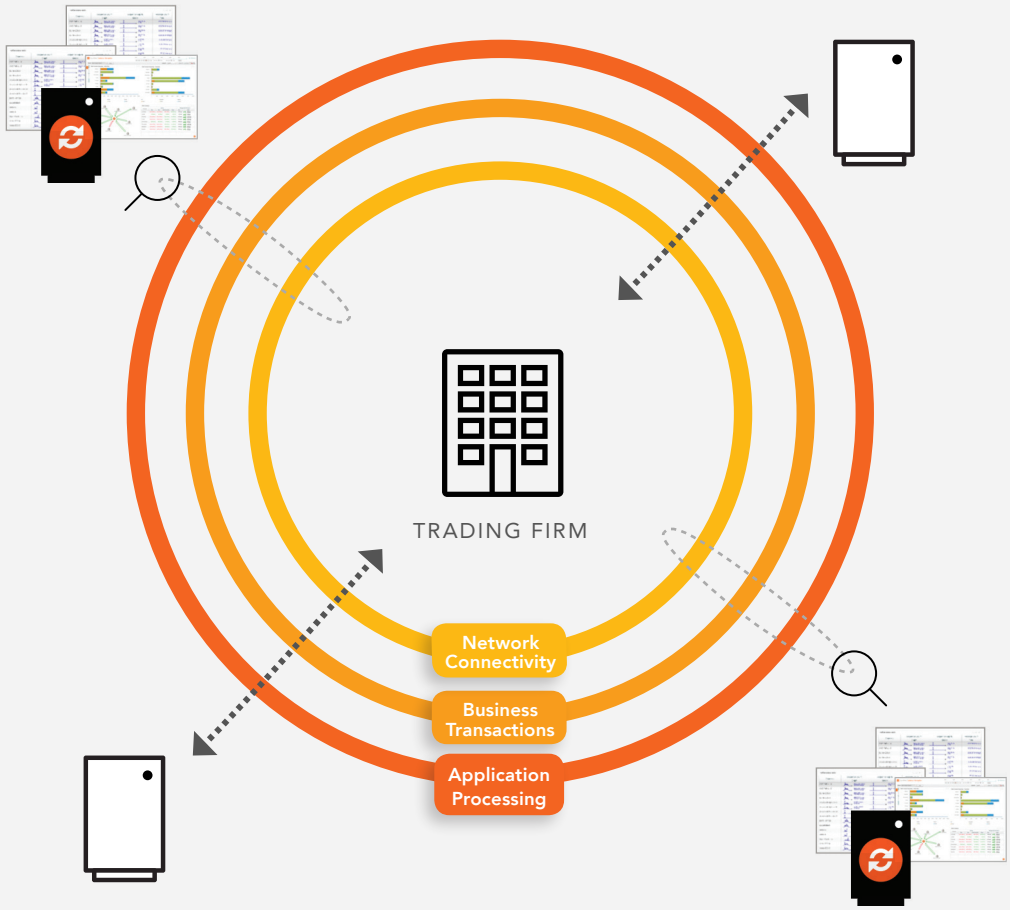


Diagram 3: Streaming Analytics Best Practices: Analyze the perimeter, gain visibility into the network, application and trading layers, and capture detailed accurate records of trading interactions.

For more information or to contact us, visit **www.corvil.com**

CWP-080814