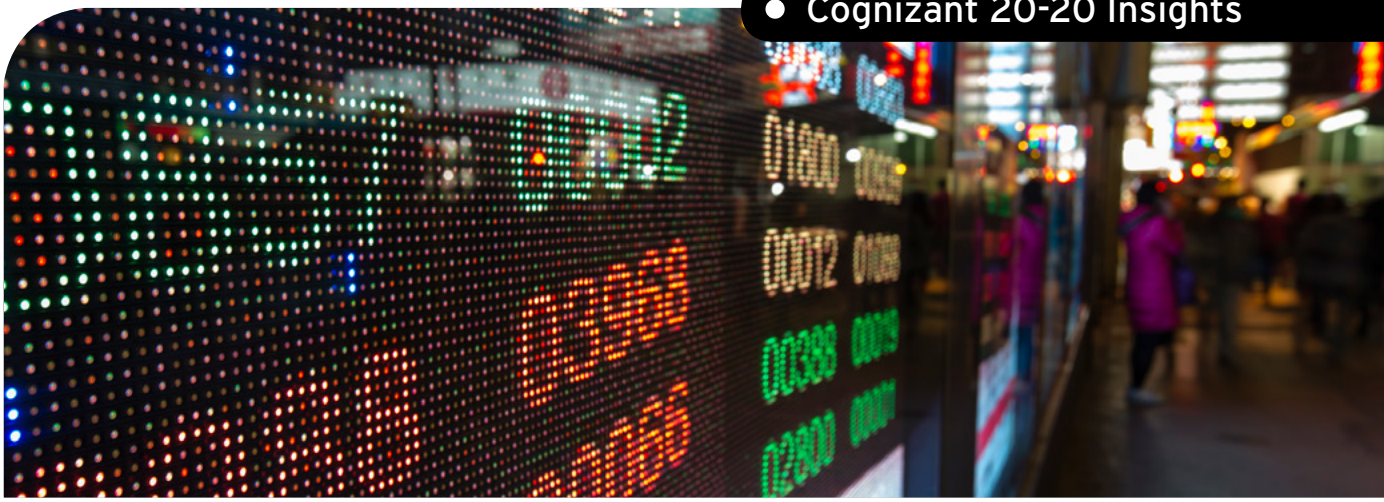


Trade Surveillance with Big Data

The rise of real-time, high-frequency trading has regulatory compliance teams working hard to keep pace with the industry's widening pools of structured and unstructured data. By employing emerging tools and techniques, capital markets firms can improve trade surveillance and spot abuse and irregularities before they can do harm.

Executive Summary

Electronic trading has come a long way since the NASDAQ's debut in 1971. Today's fragmented electronic market venues (the result of non-traditional exchanges competing for trades with traditional exchanges) have created so-called "dark pools of liquidity." Simultaneously, automated and algorithmic trading has become more sophisticated – now enabling individuals and institutions to engage in high-frequency trading (HFT).[1] As a result, the number of trades has increased tenfold in the last decade, from 37 million trades in NYSE listed issues in February 2004 to 358 million in February 2014.[2]

Traders at capital market firms have been at the forefront of these advancements – pushing the envelope along the way. How has this impacted trade surveillance and compliance teams? The rise of algorithmic trading, where split-second execution decisions are made by high-performance computers, plus the explosion of trading venues and the exponential growth of structured and unstructured data, are challenging regulatory and compliance teams to rethink their surveillance techniques. Those that depend on individual alerts can no longer meet most firms' requirements. We believe that capital markets firms require a radically new and holistic surveillance approach.

This paper highlights some of the key issues faced by regulators and compliance teams. We will also describe how new "big data" solutions can help manage them.

Challenges and Changes on the Trading Landscape

While traders march ahead with high-frequency trading and order books that allow them access to liquidity spread across geographies, surveillance and compliance teams cannot seem to catch up. A primary reason is their inability to access and harness huge volumes of data.

Let's look at some of the obstacles faced by the surveillance community today:

1. **Capturing and recalling each event in the lifecycle of a trade.** Trading firms need to maintain and recall the end-to-end lifecycles of individual trades. This helps internal compliance teams perform deep-dive analyses and dig into any issues that may arise. It also enables firms to respond accurately to regulatory investigations when needed. In some instances, this becomes mandatory, as detailed in the following examples:

   - Per Dodd-Frank regulations, a trading firm must be able to provide step-by-step details of the complete lifecycle of any swap

Cognizant

or associated transactions — including information on related deals.

- As stated in the article on Financial Industry Regulatory Authority (FINRA) Rule 5270 (effective from September 3, 2013) no FINRA member broker-dealer shall execute an order to buy or sell a security or a "related financial instrument" when that member has material, non-public market information concerning an imminent block transaction in that security, and that information has not been made public or has not become stale or obsolete.[3] In the event of an inquiry, firms must quickly recall and reconstruct the entire lifecycle of the block trade. This can become very complicated; a single block order sent via an algorithmic engine can be broken into thousands of smaller orders, and may get routed to multiple execution venues over the course of hours, days, or even weeks.

According to financial-services and software vendor Sungard, complete and rapid recalls of the details of a trade require:[4]

- Trade-related data from trade systems and ancillary systems.
- Electronic and telephonic communications data from instant messaging applications, e-mails, phone call logs and transcripts.
- Data from microblogging sites such as Twitter.

Firms must also be able to merge structured and unstructured data. In this way, they can re-create every detail of the trade and gain insight into market conditions and any other information that may have influenced the trader.

2. **Curbing market manipulation in HFT.** Market manipulation techniques such as "quote stuffing," "spoofing" and "pump and dump" are among the top-of-the list items that regulators on both sides of the Atlantic seek to detect and eliminate.

   Firms that actively use HFT have mastered the complexity of the market structure, but institutional investors are far behind HFT. In his new book, "Flash Boys: A Wall Street Revolt,"[5] author Michael Lewis points to some of HFT's major issues:

   - **Electronic front running.** This involves using extremely fast computer programs to detect market orders and jump to the front of that queue. Front-running results in the entity that placed the original order having to buy at a higher price point.
   - **Slow market arbitrage.** When the price of a stock changes on one stock exchange, an HFT algorithm picks up orders available on other exchanges before those exchanges have had a chance to react.

   Today, there are no regulations in place that prevent an HFT algorithm from engaging in techniques such as "electronic front running" or "slow market arbitrage." The reason? Prior to the emergence of HFT, no one considered such scenarios. Clearly, HFTs have exploited gaps in the regulatory frameworks — giving them just enough lead over the public investors to make substantial gains.

   In September 2013, the Commodity Futures Trading Commission (CFTC) announced plans to build a regulatory framework around high-speed and algorithmic futures trading. The CFTC subsequently released a 137-page document[6] that requested public input on various proposed ways to control the associated technology risks while enabling more trades to be made faster, and with less human interaction.

   Firms that actively use HFT have mastered the complexity of the market structure, but institutional investors are far behind HFT.

   Detecting such market manipulation techniques requires real-time surveillance. But given the number of trades executed by human traders and their robotic partners at venues spread across continents, connecting the dots in real time can be daunting. A trader working at multiple venues, for example, could deploy a high-frequency algorithmic system to process trades at exceptional velocities, making those trades impossible to track without equally fast technology.[7]

3. **Assembling the bigger picture for efficient and effective "cross-market and cross-asset surveillance."** At a large firm, traders have access to multiple trading venues; they can view liquidity across all markets. They have a consolidated order book and can access its full depth across different asset classes and venues, including exchanges, ECNs and dark pools of liquidity.

   One question to ask: Do compliance officers have a similar view of the market? A trade surveillance application can scan traders' executions, but it does not take into consideration

the entire order book that was visible to the trader, or trades made by other players in the market across multiple venues worldwide. Trade sureveillance applications do not track a stock's price movement against the volume traded for the stock throughout all venues from which the firm can trade. Nor do they compare a trader's actions with the positions in related future contracts. According to one publication, compliance teams cannot compare the bid placed on an ECN with the best bid/ask. They cannot detect spoofing if the trader is trading using several venues in different time zones and with different currencies.[8]

In a nutshell, compliance officers do not have a complete picture of the market. They cannot perform any cross-market analyses with the information accessible to them. They need to have the same view as the trader – a consolidated order book that spans multiple venues.

Considering the large volumes of data that result from various venues worldwide and the accelerated speed of trading, it is simply not possible for a human being to detect all possible instances of market abuse. Consequently, future surveillance tools must be programmed to detect any suspicious activities, store very large volumes of data, and analyze that data in real time.

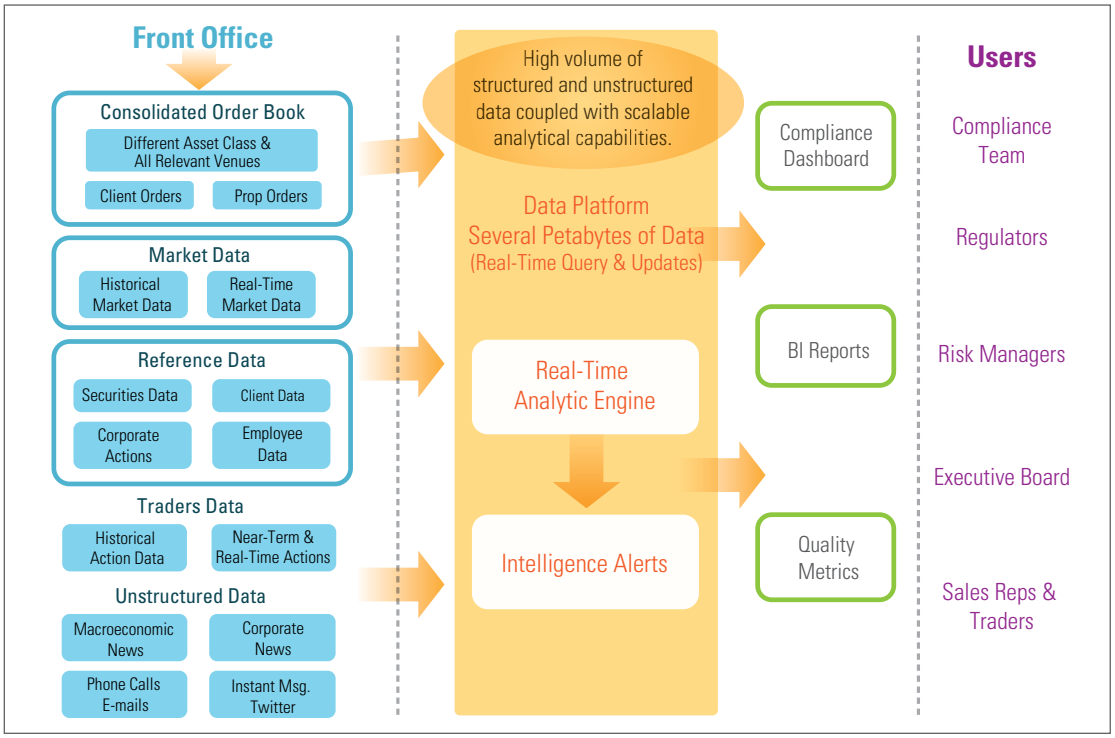
A Regulatory and Industry Roadmap for Overcoming the Hurdles

The explosive growth of data over the last few years is taxing the IT infrastructure of many capital markets firms. Fortunately, there are emerging technologies that can help these companies better manage and leverage ever-bigger data pools. These tools can enable trading firms to end data triage and retain useful historical information. By building a big-data architecture, IT organizations can keep both structured and unstructured data in the same repository, and process substantial bits and bytes within acceptable timeframes. This can help them uncover previously inaccessible "pearls" in today's ever-expanding ocean of data.

Big data analytics involves collecting, classifying and analyzing huge volumes of data to derive useful information, which becomes the platform for making logical business decisions (see figure below).

Relational database techniques have proven to be inadequate for processing large quantities of data, and hence cannot be applied to big data sets.[9] For today's capital markets firms, big data sets can reach multiple petabytes (one petabyte is one quadrillion bits of data).

A Big Data Analytics Reference Architecture



To keep processing times tolerable, many organizations facing big-data challenges are counting on new open-source technologies such as NoSQL (not only SQL) and data stores such as Apache Hadoop, Cassandra and Accumulo.

Considering the large volumes of data that result from multiple venues worldwide and the accelerated speed of trading, it is simply not possible for a human being to detect all possible instances of market abuse.

The figure on the previous page depicts a representative big-data architecture appropriate for modern-day trade surveillance.

A highly scalable in-memory data grid (e.g., SAP's HANA) can be used to store data feeds and events of interest. Real-time surveillance can thus be enabled through exceptionally fast[10] open-source analytic tools such as complex event processing (CEP). CEP technologies like Apache Spark, Shark and Mesos put big data to good use by analyzing it in real time, along with other incidents. Meaningful events can also be recognized and flagged in real time.

There are some key guidelines that firms should bear in mind while formulating big-data strategies for surveillance and compliance. In our experience, superior results can be achieved by keeping note of the following while developing a big data solution:

- **Involvement of senior management.** Regulators and compliance officers are aware of the advantages of putting big data to work. The issue is that a big-data solution cannot be implemented by one or two teams in an organization. Company-wide involvement and a commitment and direction from top management are required to initiate this effort. Senior executives must also share their vision with the entire firm. Open discussions about goals and objectives will help various teams collaborate more effectively.
- **Identify issues and take small steps.** Clearly identify the issues that the organization wishes to address with a big data solution. Pick only a few in the first round of implementation.
- **Planning.** Develop a strategy and a solution roadmap, bearing in mind the strengths and limitations of the business. Also, it is prudent to focus on the three 3 "Vs" of big data applicable to the organization:
  - **Velocity.** The speed at which the data comes in and is stored.
  - **Variety.** The types of structured and unstructured data.
  - **Volume.** The data set size, which can read petabyte proportions.
- **Iterative development.** Develop a solution to prioritize issues for the initial phase. This will help the organization create a mechanism to resolve challenges faced along the way. Solutions can then be added to address all other identified issues as the implementation proceeds.
- **Data quality is essential.** Last but not least, the quality of the data being captured plays a vital role in big-data solutions. IT organizations must focus on allocating the required resources needed to ensure that the highest-quality data is captured for proper and meaningful analysis.

According to industry experts quoted in an article published in TabbFORUM,[11] "next-generation" architectural requirements for surveillance should encompass the following:

- **New types of data sources should be included to gain a complete picture.** Real-time news and data from social media can be helpful in evaluating the circumstances under which a trade was executed.
- **Fuse data across the timeline.** Historical, new and real-time data must come together to provide a 360-degree view of market activity, sentiments and trading behavior.
- **Put activity in context.** Once validated against historical data, a suspicious activity can turn out to be an aberration or a pattern. The NoSQL database aims to supersede and replace legacy RDBMS systems to deal with the rising data tide.
- **Real-time analytics is a must-have.** Market data must be analyzed at the speed of automated trading. CEP technologies can provide the real-time analytics needed for modern surveillance systems.

Looking Ahead

Firms need to equip their regulatory compliance teams with tools and technologies that can keep pace with robot traders. They also need to manage the torrents of data coming at them in various formats from various venues and media in order to discover and apply actionable intelligence.

Using big data solutions, capital markets firms can:

- Have a complete view of historical activities, including highly granular details down to very small time intervals.
- Enable quick recall, review and analysis of large volumes of historical data from algorithmic trading programs. This provides a better view of the trading patterns needed to uncover anomalies.
- Deploy CEP technologies to uncover practices of market abuse that are hard to detect with existing technologies.
- Be seen as a proactive player in the eyes of regulators, the market and customers.

Large, global U.S.-based banks are already raising the bar by applying insights distilled from big data. Implementing big data and CEP together allows them to capture data from active feeds and interpret the signals in real time. In addition, they are creating higher forms of business intelligence by bringing together streams from market news, information from social media such as Twitter, exchange data, market quotes, and insights distilled from their own trade executions in real time.

By tapping into new tools and data streams, surveillance teams can effectively uncover more – and more complex – patterns of abuse, and move beyond the specific alerts that today only spot insider trading.

Footnotes

[1] http://en.wikipedia.org/wiki/High-frequency_trading
"High-frequency trading (HFT) is a type of algorithmic trading, specifically the use of sophisticated technological tools and computer algorithms to rapidly trade securities. HFT uses proprietary trading strategies carried out by computers to move in and out of positions in seconds or fractions of a second."

[2] Consolidated Volume in NYSE Listed Issues, 2014.
http://www.nyxdata.com/nysedata/asp/factbook/viewer_edition.asp?mode=tables&key=306&category=3

[3] Charles S. Gittleman, Russell D. Sacks, Shriram Bhashyam, Michael J. Blankenship & Steven Blau. FINRA Rule 5270 FAQs: Front Running of Block Transactions, 2013.
http://www.shearman.com/~/media/Files/NewsInsights/Publications/2013/01/FINRA%20Rule%205270%20FAQs%20Front%20Running%20of%20Block%20Tran__/Files/View%20full%20memo%20FINRA%20Rule%205270%20FAQs%20Front%20Runnin__/FileAttachment/FINRARule5270FAQsFrontRunningofBlockTransactions__.pdf

[4] Big Data – Challenges and Opportunities for the Energy Industry. Sungard, 2013.
http://financialsystems.sungard.com/~/media/fs/energy/resources/white-papers/Big-Data-Challenges-Opportunities-Energy-Industry.ashx

[5] Flash Boys: A Wall Street Revolt, by Michael Lewis. Published by W. W. Norton & Company; 1st edition. March 31, 2014.

[6] http://www.cftc.gov/ucm/groups/public/@newsroom/documents/file/federalregister090913.pdf

[7] Steve Dew-Jones. Real-Time Surveillance: Mission Impossible? March, 2012.
http://www.waterstechnology.com/waters/feature/2163111/real-surveillance-mission-impossible
(The paper above can be read after subscribing to the Web site).

[8] Michael O'Brien. Cross-Market surveillance is essential In an era of market fragmentation, 2011.
http://www.nasdaqomx.com/digitalAssets/80/80361_75297_cross-marketsurveillancewhitepaper_final.pdf

[9] Big Data – Challenges and Opportunities for the Energy Industry. Sungard, 2013. http://financialsystems.sungard.com/~/media/fs/energy/resources/white-papers/Big-Data-Challenges-Opportunities-Energy-Industry.ashx

[10] Data throughput rates that can range to millions of events or messages per second.

[11] Mark Palmer. Real-Time Big Data and the 11 Principles of Modern Surveillance Systems, 2011. http://tabbforum.com/opinions/real-time-big-data-and-the-11-principles-of-modern-surveillance-systems (Access to the paper requires creation of a free account on the Web site of TabbFORUM).

References

- http://www.bigdata.com/blog/
- http://www.finra.org/Industry/Regulation/Notices/2012/P197389
- http://www.techrepublic.com/topic/big-data/
- http://www.cftc.gov/LawRegulation/index.htm
- http://nosql-database.org/
- http://hadoop.apache.org/

About the Author

Pritesh Bhushan is Senior Manager – Consulting, in Cognizant's Business Consulting, focused on the capital markets domain. Pritesh has worked on several projects for large investment banks, providing technology consulting in the areas of compliance and trade surveillance. He has over 12 years of experience in technology consulting for capital markets firms. Pritesh has an MBA from Indian Institute of Management, Bangalore and a B.Tech from Indian Institute of Technology, Kanpur. He can be reached at Pritesh.Bhushan@cognizant.com.

About Cognizant

Cognizant (NASDAQ: CTSH) is a leading provider of information technology, consulting, and business process outsourcing services, dedicated to helping the world's leading companies build stronger businesses. Headquartered in Teaneck, New Jersey (U.S.), Cognizant combines a passion for client satisfaction, technology innovation, deep industry and business process expertise, and a global, collaborative workforce that embodies the future of work. With over 75 development and delivery centers worldwide and approximately 187,400 employees as ofJune 30, 2014, Cognizant is a member of the NASDAQ-100, the S&P 500, the Forbes Global 2000, and the Fortune 500 and is ranked among the top performing and fastest growing companies in the world. Visit us online at www.cognizant.com or follow us on Twitter: Cognizant.

World Headquarters
500 Frank W. Burr Blvd.
Teaneck, NJ 07666 USA
Phone: +1 201 801 0233
Fax: +1 201 801 0243
Toll Free: +1 888 937 3277
Email: inquiry@cognizant.com

European Headquarters
1 Kingdom Street
Paddington Central
London W2 6BD
Phone: +44 (0) 20 7297 7600
Fax: +44 (0) 20 7121 0102
Email: infouk@cognizant.com

India Operations Headquarters
#5/535, Old Mahabalipuram Road
Okkiyam Pettai, Thoraipakkam
Chennai, 600 096 India
Phone: +91 (0) 44 4209 6000
Fax: +91 (0) 44 4209 6060
Email: inquiryindia@cognizant.com