Summer | 14

Risk and Reward: The Effect of Big Data on Financial Services

Jose Gutierrez, Thomas Anzelde, Galliane Gobenceaux

Big Data will minimize risk in fraud detection, compliance and portfolio management. This risk reduction, in combination with improving trading strategies, has the potential to give financial service companies a substantial competitive advantage.

INTRODUCTION

The advent of big data has changed the way financial service companies do business. Big Data will minimize risk in fraud detection, compliance and portfolio management. This risk reduction, in combination with profit strategy optimization, has the potential to give financial service companies a substantial competitive advantage.

For firms that interact with the public markets, big data has enabled new strategies that go beyond simple enhancements. As financial products and systems become more complex, they create new fields of opportunity for fraudsters. Financial firms are now turning to Big Data to quickly detect and prevent evolving and complicated fraud schemes. Regarding compliance and audit, both regulators and firms are leveraging the ability to implement policies that govern workflows, complex business logic, and other large data sets. Government departments are using Big Data to assess systemic risk in prominent financial markets to implement safeguards against threats such as bubbles and recessions. At the same time, firms are adopting preventive measures to avoid punishments that could threaten firms' viability and core business. These pervasive changes have forced financial firms to evolve or perish.

1. THE IMPACT OF HIGH FREQUENCY TRADING

Trading technology and the large amounts of data required have drastically changed the way many financial firms interact with the public markets. Advances in both have led to a rise in algorithmic trading, where programs make the trading decisions. High frequency trading, or HFT firms leverage a combination of the latest technology, algorithms, and data to influence and profit from the markets. HFT firms trade billions shares a day, and account for about half of all stock trades in the U.S. (http://knowledge.wharton.upenn.edu/article/high-frequency-trading-profiting-news/) Profits from HFT average a twentieth of a penny per share, yet the cumulative effect on the markets can be in the hundreds of billions. (http://www.businessweek.com/articles/2013-06-06/how-the-robots-lost-high-frequency-tradings-rise-and-fall#p1). The effect of HFT is so widespread that many large financial firm buying or selling securities must either adapt or sacrifice millions of dollars in profits.

1.1 Technological Factors

Multiple technologies contribute to the efficacy of HFT. Fiber optic transmissions are the standard for performing arbitrage between markets in different locations. For some applications, microwaves are being used to transmit data faster than the fiber optic cables. This is possible because microwaves can transmit data in a straight line through the atmosphere, whereas fiber optic cables are not laid in a straight line. Many companies use Hadoop in order to make sense of large streams of market data. Hadoop distributes the processing responsibilities across clusters of computers. (http://hadoop.apache.org/) It allows for parallel processing without requiring complex code. Hadoop MapReduce is used by some firms to organize the data. (http://www.pcmag.com/article2/0,2817,2424495,00.asp) The Map function performs sorting capabilities that can be used to identify which stocks have prices with the most potential for profit. The Reduce function can then group them into smaller, more usable sets.

Machine learning algorithms are run on trillions of historical observations of market data. These algorithms identify new patterns to be exploited with a variety of strategies. (http://heartland.org/sites/default/files/ssrn-id2034858.pdf) Trading algorithms are generally compact and conceptually simple, because speed is critical. They are generally designed by highly educated quantitative analysts. Because of this, they are smart, small in size, and extremely valuable. In 2010 a programmer at Goldman Sachs was arrested by the FBI on allegations that he stole their proprietary trading code. The size of the code was only 32 MB of data. (http://www.cnbc.com/id/39041598#) As a result of the speed and processing required for algorithmic trading, winning an arms race is often a factor of success. The winner-take-all nature of some strategies increases the stakes even more. If HFT firms do not continuously innovate and invest in cutting edge technology, they risk becoming obsolete.

1.2 Regulatory Factors

Many regulatory factors played a part in the rise in HFT. Pre 1998, the NYSE and NASDAQ exchanges possessed a majority of the trading activity in the US. In order to spur competition, the Securities Exchange Commission (SEC) passed its Regulation on Alternative Trading Systems in 1998. This regulation authorized the existence of electronic trading systems, also known as Electronic Communications Networks, or ECNs. ECNs functioned as a hybrid of an exchange, a broker, and a market maker, and could be created and run by independent firms. This allowed individuals and firms to trade around the clock without many of the restrictions of the major exchanges. The growth of ECNs led to increased operational efficiencies as well as more opportunities to profit from HFT.

In 2001 another major factor, decimalization, came into effect. This changed the minimum possible change of a stock's price from 1/8 of a dollar to 1 cent. (http://online.wsj.com/public/resources/documents/HFT0324.pdf) It allowed buyers and sellers to come to a closer agreement on prices, which led to significantly more trading. It also evaporated the profits that were made by human floor traders. These traders, known as market makers, would perform arbitrage on small price fluctuations and keep the difference. It had the opposite effect on algorithmic traders, which were able make an increased number of trades to offset the smaller profits per trade. Non-HFT traders, such as large institutions, had to adopt algorithmic trading in order to protect against algorithms of predatory HFT firms.

The final major factor was the Regulation on National Market System, published by the SEC in 2005. Before this, brokers had discretion over which exchange they would use to fill an order. They would often fill the order at one exchange. The regulation included the requirement that market orders be posted electronically and executed at the best price nationally. This broke up large orders so that they would be executed as smaller orders to get the best prices at different exchanges. When the first of these smaller orders arrived at an exchange, HFT algorithms could detect it. This simple awareness allowed HFT firms to use strategies in order to exploit this knowledge. (http://www.washingtonpost.com/blogs/wonkblog/wp/2014/04/04/a-veteran-programmer-explains-how-the-stock-market-became-rigged/)

1.3 Strategies

There are a variety of strategies that HFT firms use to profit. Some are improvements on old techniques. The first strategy that falls into this category is profiting from the spread, also known as market making. The spread is the difference between the amounts that investors are willing to buy and pay. In order to make money off of the spread, firms place orders to buy below or sell above the current market price. Previously, markets and trading firms hired individuals known as market makers to perform this function on the floor of the stock exchange. Now, HFT firms leverage Big Data to analyze vast quantities of stock prices along with the relative favorability of their

spreads. In conjunction with greater speed, these advantages increase profits and reduce the risk of adverse price changes occurring while entering into positions. As with many advances made by HFT firms, the efficiencies are so great that the human competition has been decimated.

The second strategy that improves on previous techniques is low-latency arbitrage between different locations. This strategy creates profits by exploiting price differences in two locations for the same stock. Before computers existed, there were instances of traders using outrunning the competition by identifying faster methods to communicate. In 1865, American financier Jim Fisk chartered fast boats to outrun mail boats with news of the outcome of the Civil War. He was then able to short Confederate bonds. (http://www.forbes.com/forbes/2010/0927/outfront-netscape-jim-barksdale-daniel-spivey-wall-street-speed-war.html). A common modern-day application of this strategy is trading a stock listed at different prices on different exchanges. Some HFT algorithms can detect the first trade of a large, multi-exchange purchase of a stock on one exchange and then race to buy the same stock elsewhere. If the projected orders materialize, a large profit is made on the spike in demand. If they do not, then the position can be unwound for a small loss. Another typical application involves futures and the securities they represent. For example, an S&P 500 futures contract on the Chicago Mercantile Exchange will sometimes sell at a higher price than the individual S&P 500 stocks on the NYSE. An algorithm can simultaneously sell the futures contract and buy the stocks. Because the two securities are tightly linked in value, they will theoretically return to price parity, and the trader will profit. Because of the hedged nature of this trade, there is very low risk of the prices fluctuating and causing a loss. The financial markets research and advisory firm TABB Group has estimates that "annual aggregate profits of low-latency arbitrage strategies exceed $21 billion, an amount which is spread out among the few hundred firms that deploy them." (http://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=1211&context=dltr)

The third strategy in the category is statistical arbitrage. Statistical arbitrage relies on analysis of market data. HFT firms mine market data to determine correlations between different stocks. For example, the stock price of an oil company will increase with the price of crude oil while the price of airline stocks decrease. When these correlations are not maintained by the market, HFT firms will trade the stock in order to buy at a discount or sell at a premium. This creates a profit when the stocks return to their expected relative value.

Other strategies employed by HFT firms are not improvements on old techniques, but rather ways of bypassing the standard mechanisms involved in trading. One common strategy that does this is called pinging. Pinging involves placing small limit orders inside the spread to detect hidden orders. Limit orders have a limit on the price that can be paid. Hidden orders are limit orders that are unobservable to other participants in the market. They are used by large funds wanting to avoid detection for the large trades they need to make. Traders can exploit the potential supply-demand imbalance at the fund's expense. (http://scholar.princeton.edu/sites/default/files/JiangminXu_JobMarketPaper_Revised_0.

pdf) If the small limit orders created by the pinging algorithm are not executed, they are cancelled immediately. This possible result is essentially risk-free, since regular market participants would not be able to react fast enough to take advantage of those small orders. On the other hand, if the small limit orders used for pinging are executed immediately, then there are likely hidden orders inside the spread. The HFT firm will then buy or sell shares to prevent the hidden order from getting filled. Eventually, the large fund will be forced to adjust its price in order to complete its order. Then, the HFT firm sells its accumulated shares to the large fund firm at a premium.

Another HFT technique that bypasses traditional means is directional trading. Directional trading leverages sources of data outside of a company's official financial reports. HFT firms use algorithms to parse through news releases for information that implies a profitable trade. An algorithm can search for words such as "increase" or "higher" next to "earnings" in news releases. Upon detection, the future direction of the stock can be assumed, and the HFT firm can buy the shares before other investors. Other algorithms search for specific events that are indirectly tied to the stock market. Some HFT firms leverage Twitter's data feed, which includes 400 million tweets a day, and use Big Data to process the information. On April 23, 2013, false news was posted by a hacker on the Associated Press' Twitter feed. The tweet described an explosion at the White House injuring president Obama. Within two minutes, HFT algorithms caused the S&P 500 to lose 1% of its value and the US stock market to drop in value an estimated $200 Billion. (http://www.businessweek.com/articles/2013-04-24/how-many-hft-firms-actually-use-twitter-to-trade#p1) Although this was a case where the strategy did not necessarily make money, it also implies the profits available to be taken when significant events really occur.

Illegal strategies have also been used by HFT firms. They are difficult to detect and malicious intent is difficult to prove. This is largely due to the amount of trades placed and the complex interaction of the market and the traders. Momentum ignition strategies are a modern version of the pump-and-dump scheme. These involve creating multiple fake orders along with a few real ones that get executed. These are intended to trick competitor algorithms into assuming that a trend has started to develop and the stock price will continue in the same direction. The competitor algorithm will then try to buy or sell in order to profit from the trend. The algorithm originating the strategy will trade in advance of the competitor and accommodate the competitor's demand at unfair prices.

1.4 Risk Mitigation

There is little room for error when operating in the market at the speed of HFT. An HFT firm is in some ways more at risk from human error than one that relies solely on human trading. This is because by the time a mistake is detected, a computer may have executed millions of erroneous trades. On August 1, 2012, a HFT firm called Knight Capital traded 397 million shares (http://dealbook.nytimes.com/2013/10/16/knight-capital-to-pay-12-million-fine-on-trading-violations/?_php=true&_type=blogs&_r=0)

using a program that had a glitch. The company lost $10 million per minute until the problem was detected – resulting in a $440 million loss. This exceeded their $365 million cash on hand and the firm's stock dropped 63%. (http://online.wsj.com/news/articles/SB10000872396390443866404577564772083961412) In order to decrease this risk, companies will test their algorithms on large amounts of historical market data.

Any firm attempting to buy and sell stocks with non-HFT methods or inferior technology risks trading at suboptimal prices. Competitors use Big Data to search for telltale patterns in the markets and can often detect when an institution is buying and selling large amounts shares. In order to reduce the risk of having predatory HFT firms exploit their trades, many financial firms have to mask their behavior in order to avoid tipping off others. One of the ways in which a firm can mask its behavior is to mine large amounts of the market data to identify the normal trading patterns of the stock to be traded. Then, trade within that pattern in order to hide in the crowd. Another way is to use round lots for trades. Round lots are orders for round numbers of shares. Using these will help to mislead the HFT firms into concluding that retail investors are making the trades, and not a large institution. Creating misdirection is also a way that an institution can confuse HFT firms. In 2013, a Vice President at T. Rowe Price described what happened after a trade was sent to one of their brokers. He discovered that "To hide the purchase from fast-moving traders, the broker placed and canceled many smaller orders all across the stock market, creating a dense smoke screen of phantom interest in the security. In total, the broker offered to buy 750 million shares of the stock while actually purchasing just 2.5 million." (http://online.wsj.com/news/articles/SB10001424052702303281504579219962494432336)

1.5 Broader Effects of HFT

The major benefit to the broader market created by HFT is the narrowing of spreads. Spreads have "fallen by an order of magnitude since 2004, from around 0.023 to 0.002 percentage points." (http://www.bankofengland.co.uk/publications/Documents/speeches/2011/speech509.pdf) Decreasing spreads lead to less frictional costs for investors, who are no longer paying large commissions to compensate for human slowness. They also allow the market to give a fairer price to both buyers and sellers, since the narrower spread more closely resembles an implied correct value for the stock.

A danger to the overall market posed by HFT firms is the potential reduction of liquidity during a market downturn. Generally, algorithmic trading increases liquidity since it creates many more opportunities to buy and sell than humans can. But when the market drops sharply, many algorithms sell all of their stock and effectively shut down in order not to lose money. This can drastically reduce liquidity when the markets need it most. Human traders have traditionally been able to change strategy immediately and profit from a downturn by buying shares at depressed prices. Now, in a sharp downturn, prices can change at much more rapid pace. In that situation, human traders are unable

to know the current state of the market. They have no choice but to stop trading, just like the algorithms. When both algorithmic and human traders exit a falling market, it makes the situation worse. This happened during the "Flash Crash" of 2010. Accenture shares fell over 99%, all the way down from $40 to $0.01. (http://www.bankofengland.co.uk/publications/Documents/speeches/2011/speech509.pdf)

2. BIG DATA IN THE FIGHT AGAINST FINANCIAL CRIME

As financial products and systems are becoming more and more complex, financial fraudsters are becoming more adaptive, which gives them more opportunity to commit crime. Every year, dozens of billion of dollars are being lost by financial institutions due to fraudulent practices. The response to this trend have been relatively low-tech in the past, but there has been a clear shift in this paradigm for the past few years, as Big Data technologies are getting more sophisticated and affordable. Big Data IT solutions targeted to financial firms are getting popular on the market, and this technology will surely become more powerful the next few years, as research interest in this field is flourishing.

2.1 Financial Crime Management (FCM): Current Situation and Challenges

2.1.1 The Billion-Dollar Opportunity behind Financial Crime

According to the ACFE's 2012 Report to the Nations (http://www.acfe.com/rttn-highlights.aspx), on average, 5% of a company's total revenue is lost due to financial fraud; for financial companies, this loss is even more significant, estimated at $80 billion annually for the industry (http://www.research.ibm.com/foiling-financial-fraud.shtml). Financial crime for these institutions affects companies in two different ways: the effect can be either direct, on the income statement bottom line, or indirect, through unexpected expenses resultant from the fraud. Bottom line losses "vary by type of transaction, such as ATM, credit card, or wire fraud", according to IOVATION (https://www.iovation.com/industries/financial-services), and are generally the most expensive. Indirect losses include tangible (cost of case investigation, customer phone support), and intangible costs (clients loss, compromised trade secrets, affected image and reputation).

Fraudsters are fast, intelligent and adapt quickly to any change in the environment. Different studies, elaborated by Deloitte (http://www.deloitte.com/view/en_US/us/Services/additional-services/deloitte-analytics-service/2d5678046eb3e310VgnVCM2000003356f70aRCRD.htm) and Ernst & Young (http://www.ey.com/Publication/vwLUAssets/Annual_Financial_Crime_Trends_Survey_2012/$FILE/EY_Annual_Financial_Crime_Trends_Survey_2012.pdf) among other firms, have pointed out that that the complexity of financial products and systems, the increase of connectivity and data availability, as well as higher rates of digitalization, increase companies and institutions vulnerabilities and provide criminals with more opportunities to commit financial crimes.

2.1.2 Traditional FCM Solutions

Traditionally, financial institutions have invested numerous resources into fraud prevention by focusing their effort on "new account" fraud, which "occurs on an account within the first 90 days that it is open" (http://www.acfe.com/uploadedFiles/Shared_Content/Products/Self-Study_CPE/Financial%20Institution%20Fraud%202013_Chapter%20Excerpt.pdf). Since 2001, systematic background checks have been performed to verify the identity of every entity seeking to open an account, in compliance with the PATRIOT Act. Individual account representatives realize those systematic checks, generally through the labor-intensive comparison of information provided by the customer with information obtained from consumer-reporting agencies, public databases, among others. In the case of businesses, the ACFE recommends that employees visually visit the company's installations. In any case, most of fraud detection is done according to personal intuition and experience.

Financial institutions also have focused their efforts in training and development as a way to detect and prevent financial crime. As stated in Ernst & Young's Financial Compliance Report, though more than 90% of financial professionals have received training in fraud detection, only 44% of them think the cost/benefit ratio of such programs is high. Training courses generally include the study of broad fundamental fraud schemes such as counterfeit check activity, new account fraud, forged instruments, identity theft and kiting.

2.1.3 IT Systems

Even though technology has had a growing influence on the development of financial products and systems since the Internet boom, financial institutions have had issues in the past to take advantage of its full potential, according to DeSantis (http://deloitte.wsj.com/riskandcompliance/2014/03/31/managing-the-risks-of-financial-crime-and-the-role-of-big-data/):

Over the last 10 to 15 years, there has been a leap in the use of data analytics to mine big data to identify the patterns, trends and anomalies that are often indicators of fraud or other types of financial crime. In the past, these have been siloed efforts in response to particular incidents or investigations.

To prevent new account fraud, financial firms have traditionally used internal consumer reporting systems that verify the applicant's information (name, social security number, date of birth, and address of a specific customer). (http://www.acfe.com/uploadedFiles/Shared_Content/Products/Self-Study_CPE/Financial%20Institution%20Fraud%202013_Chapter%20Excerpt.pdf) Consumer reporting agencies (NCRAs) also provide useful insights on a customer credit history; in the US, the three major agencies are Equifax Information Services LLC (Equifax), TransUnion LLC (TransUnion), and Experian Information Solutions Inc. (Experian). (http://files.consumerfinance.gov/f/201212_cfpb_credit-reporting-white-paper.pdf)

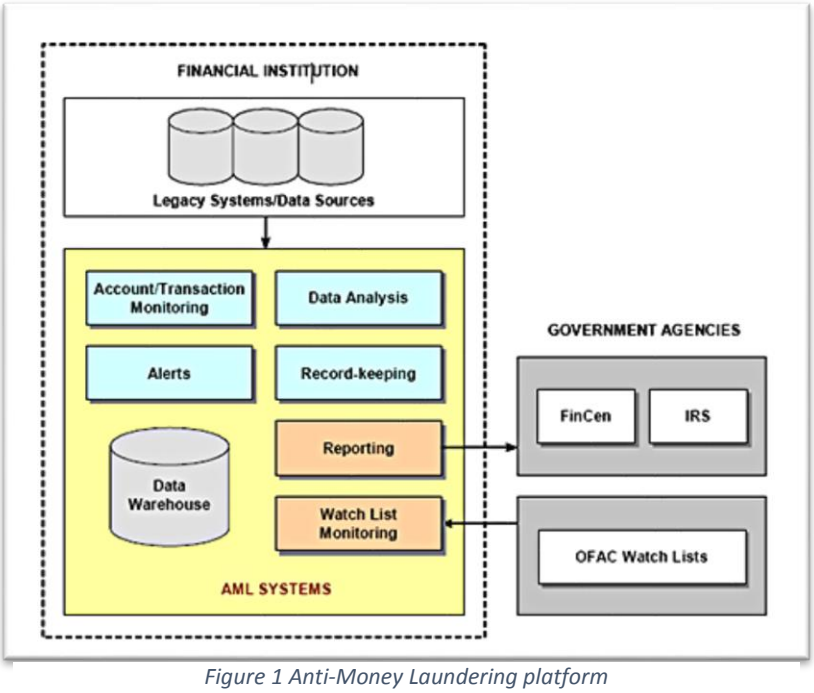

Figure 1 Anti-Money Laundering platform

For detecting suspicious transactions, one of the most popular computer programs is the anti-money laundering system. The AML filters incoming transaction data, classifies it according to its suspicion level, and inspects it for anomalies. Any transaction higher than $10,000 is screened (by law), but smaller suspicious transactions are also analyzed (deposits of large sum as multiple smaller sums in a short period of time). An employee must investigate each flagged transaction individually.

Credit card fraud is also an important source of loss for financial institutions. Traditional IT solutions perform anomaly screening by detecting suspicious activities such as unusual IP addresses or login times (outlier detection). (http://hortonworks.com/blog/how-big-data-is-revolutionizing-fraud-detection-in-financial-services/) This results in numerous "false-positives", which affect the customer's opinion, as well as the company's reputation.

2.2 Improving FCM with Big Data

2.2.1 Using Big Data to Prevent and Detect Financial Crime

67% of financial professionals believe technology is the most important factor in the management of financial crime risk, according to E&Y. The Big Data boom has been seen as a great opportunity for fraud detection and prevention in finance. According to DeSantis (http://deloitte.wsj.com/riskandcompliance/2014/03/31/managing-the-risks-of-financial-crime-and-the-role-of-big-data/), Big Data is the ideal tool to "proactively detect, deter and prevent financial crimes", by identifying trends and patterns indicative of risks that are not otherwise easily discernable.
In addition, 43% of financial professionals agree that "the use of more sophisticated technology is going to become the most important factor in the way organizations manage financial crime risk". Big Data analytics systems are becoming a necessity for financial firms for two main reasons (http://www.ey.com/Publication/vwLUAssets/Annual_Financial_Crime_Trends_Survey_2012/$FILE/EY_Annual_Financial_Crime_Trends_Survey_2012.pdf):

> *"The first [reason is] efficiency: increasingly large volumes of electronic transactions, when combined with limited access to anti- money laundering specialists, present a significant challenge for firms. Technology is perceived to be the most cost effective tool available to mitigate and detect financial crime. The second reason [is] insight: financial products and processing systems are becoming more complex."*

2.2.2 Big Data IT Solutions on the Market

As a response to an increasingly complex and fast changing environment, several IT firms have developed Big Data IT solutions for the financial market. Recently, Hadoop has started to emerge as the favorite software framework for several commercial distributors of anti-fraud solutions. Hadoop-based Big Data solutions for financial firms are starting to be commercialized, and allow detecting fraud without disrupting service to customers. They are founded on a Real Time Big Data Architecture (RTBDA) proposed by David Smith, and are divided into four tiers (http://cdn.oreillystatic.com/oreilly/data/big-data-finance-collection.pdf):

- In RTBDA, the data tier include structured data (in RDBMS, NoSQL, Hbase, or Impala), unstructured data (in Hadoop MapReduce), streaming data (from the web, social media, sensors and operational systems), basic descriptive analytics, and tools such as Hive, HBase, Storm and Spark.

- The analytics layer includes a production environment (for deploying real-time scoring and dynamic analytics), a development environment (for building models), and a periodically updated local data mart (for improved performance).

- The integration tier is the link between the end-user applications and analytics tools.

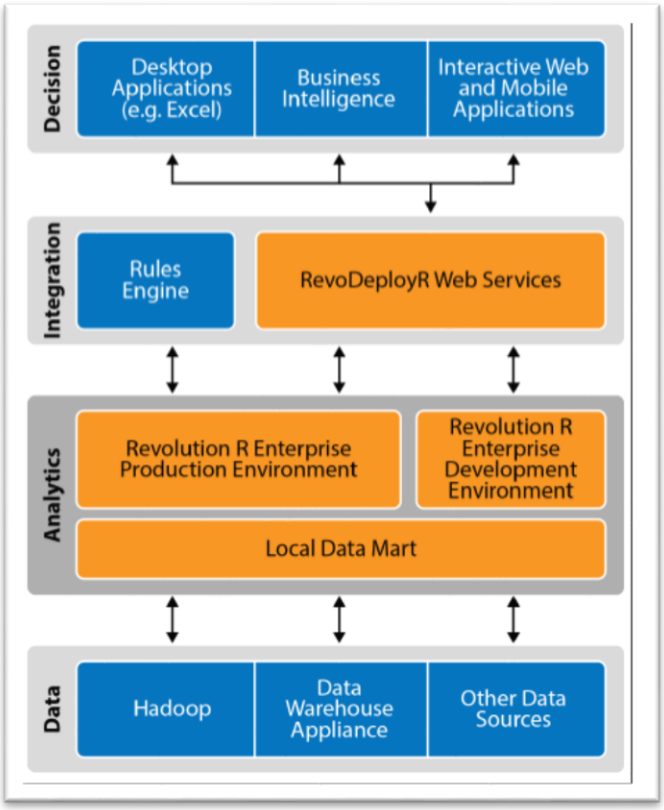

Figure 2. RTDBA stack

- The decision layer includes end-user applications (desktop, mobile, interactive web apps), as well as business intelligence software. It's where the interaction between users and real-time big data analytics system takes place.

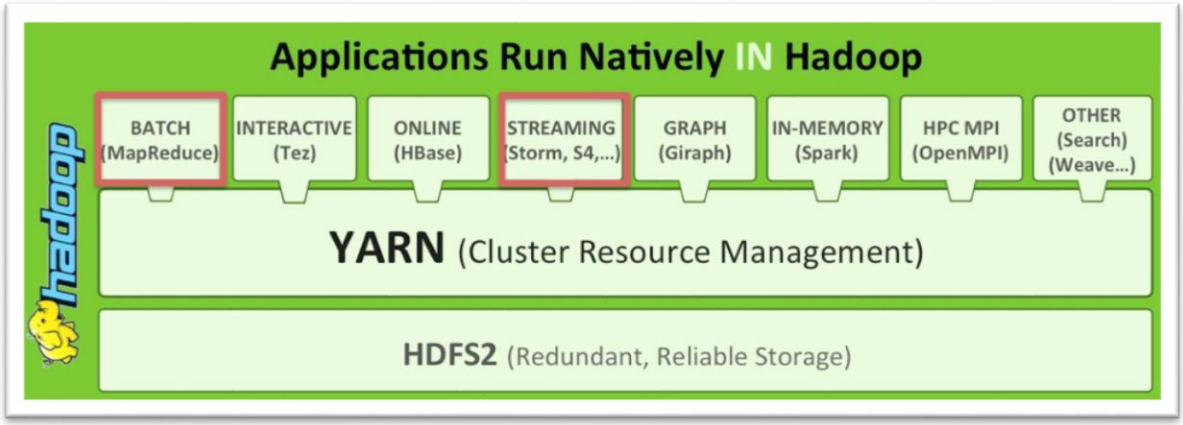

Figure 3. Hadoop's YARN system

HortonWorks and Pactera recently partnered up to develop a system based on RTBDA (http://hortonworks.com/blog/how-big-data-is-revolutionizing-fraud-detection-in-financial-services/). It implements both batch processing (through MapReduce) and stream processing (using Apache Storm) on the data layer. This solution empowers clients by taking advantage of the distributed platform of Storm, and enables them to read, analyze and react to streamed data of fraudulent transactions in less than a second. Hadoop's data operating system, YARN, allows HortonWorks solution to support multiple workloads, which results in sub-second decision making for intelligent financial decisions that can potentially save banks and consumers millions of dollars. In other words, as Wooledge (http://www.mapr.com/blog/combating-financial-fraud-with-big-data-and-hadoop#.U9RUFFbltjA) mentions, using this kind of analytics allow firms to detect fraud more accurately (less false-positive), identify fraud sooner (real time activities), and predict future attacks.

Oracle also targeted the financial market by offering a set of solutions against financial crime in the Financial Crime and Compliance Management for Financial Services suite, for which it won the Fraud and Financial Crime Software Provider of the Year award in 2014 (http://www.oracle.com/us/industries/financial-services/operation-risk-reg-oracle-2203424.pdf). The suite aims to tackle fraud and compliance risks by offering 7 sets of solutions: Oracle Financial Services Operational Risk, Anti Money Laundering, Know Your Customer, Currency Transaction Reporting, Trading Compliance, Broker Compliance and Fraud. All of them run on the Oracle Financial Services Analytical Applications platform, which gives Oracle the advantage of leveraging the overlap of all analytics approaches across the whole business.

The Oracle Financial Services Fraud in particular provides an integrated fraud management platform that is able to identify and correlate events within the company ("including intelligent aggregation and correlation of alerts and exceptions from point fraud solutions", http://jmrinfotech.com/Financial-Crime-and-Compliance-Management-for-Financial-Services), to perform real time monitoring and interdiction capabilities ("including online access and authentication"), to simulate different fraud scenarios

(including behavior detection, profiling techniques, and advanced risk scoring), and to provide case management tools ("including robust workflows, documentation, loss and recovery data, and audit trail"). This solution provides a holistic understanding of suspicious events by detecting new fraud schemes and entry points in different products and channels. Not only does it detect outsider fraudsters (identity theft or account takeover), but also insider fraud from employees, and online fraud.

As far as Anti-Money Laundering is concerned, the Oracle solutions "provides automated, comprehensive, and consistent surveillance of all accounts, customers, correspondents, and third parties in transactions across all business lines." (http://jmrinfotech.com/Financial-Crime-and-Compliance-Management-for-Financial-Services) It basically monitors every transaction, matching it to historical and accounts profiles to detect any fraudulent activity, thus reducing the false positive alerts. The platform also focuses on changes in national and international regulation to manage the risk.

The Know Your Customer Platform, apart from helping financial institutions with compliance issues, allows customers to identify and prevent financial crimes by quickly assessing the risk associated with each customer, and leveraging the level of customer behavior monitoring that has to be performed. Key benefits of the KYC platform for fraud management include more agility in the collection of identity information (documentation and verification are automated from public and private databases), and a better allocation of resources by leveraging collected data to assess customer risks, allowing firms to direct their effort on relevant and accurate information.

IBM also targeted the fraud prevention and detection market for financial institutions through its WebSphere Solutions and the Financial Transactions Manager (FTM); IBM offers a common platform that addresses general fraud threats in organizations. Even though they don't provide case management tools, as Oracle does, IBM customers can integrate WebSphere Solutions with their own case management software in order to "improve detection rates, productivity and customer satisfaction." (http://www-01.ibm.com/software/websphere/industry/banking/risk-fraud-management/)

IBM developed the FTM as a framework for "integrating, orchestrating, and monitoring financial transactions." (http://books.google.com/books?id=r7AkAwAAQBAJ&dq=WebSphere+Business+Integration+for+Financial+Networks+fraud&source=gbs_navlinks_s) This framework's main goal is to manage the interaction and integration of the complex systems and data of financial institutions, which result in fraud detection capabilities, but does not aim to prevent and detect fraud per se:

> *"In this context, [the financial transaction manager] provides the process orchestration, business insight in processing state, and integration and message management capabilities (such as fraud detection, syntax validation, and message creation and repair) to implement business process."*

(http://books.google.com/books?id=r7AkAwAAQBAJ&dq=WebSphere+Business+Integration+for+Financial+Networks+fraud&source=gbs_navlinks_s)

Some key benefits of IBM's FTM are that transactions are monitored in real-time across multiple channels, suspicious transactions are identified and assessed by their level of threat in relation to customer impact, and that different types of transactions (ATM, Remote Banking, Wire/ACH Transfers, Deposits) are monitored, no matter if they are done by internal or external sources.

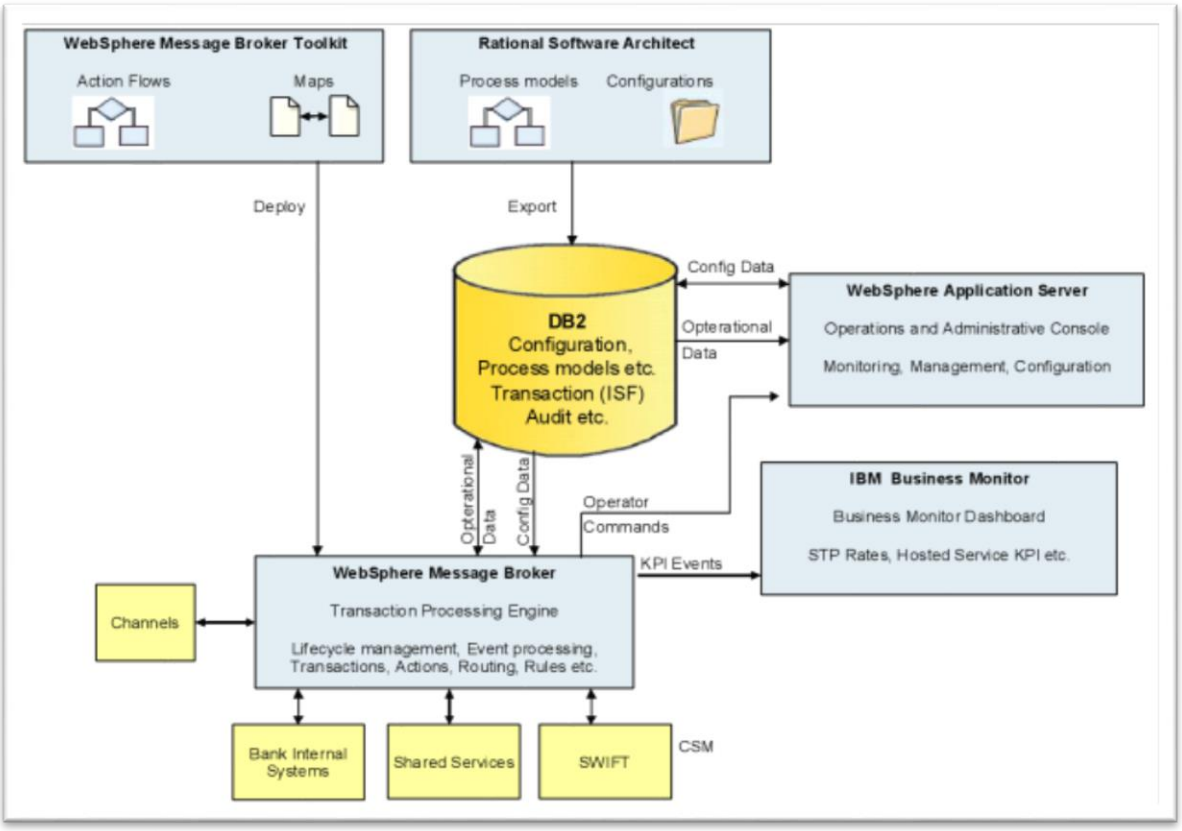

Figure 4. IBM FTM Architecture

2.3 Big data in the financial crime fight: future opportunities and challenges

2.3.1 Limitations and Challenges of Current Solutions

Even though there is a recent effort from big and medium private companies to offer more sophisticated Big Data solutions against financial crime, there is still a feeling from financial professionals that IT systems lack flexibility in order to keep up with the pace of change. According to E&Y recent study, most systems are still "relatively

unsophisticated, producing a lot of false alerts that are time-consuming to manage and distract from managing the real risks."

According to Verma and Mani (http://www.infosys.com/industries/financial-services/white-papers/Documents/big-data-analytics.pdf), improving Big Data systems' velocity of response is still one of the major challenges for IT solutions providers. The velocity of data, as well as improved accuracy of analytics results will be the key to a more precise fraud detection, which will result in better customer satisfaction and corporate reputation. It is also important for IT solutions providers to keep on improving the quality and pertinence of data collected, both structured and unstructured.

2.3.2 New Models for Internal and External Fraud Detection and Prevention

Machine learning has emerged as a new challenge and opportunity for companies to increase fraud detection rates. This technology allows different systems to interact and "learn" from one another by discovering patterns buried in data (http://www.banktech.com/leveraging-big-data-to-revolutionize-fraud-detection/a/d-id/1296473). IBM just started to research about a way to integrate machine learning and stream computing to detect financial fraud (http://www.research.ibm.com/foiling-financial-fraud.shtml):

Rather than singling out specific types of transactions, the solution analyzes historical transaction data to build a model that can detect fraudulent patterns. This model is then used to process and analyze a large amount of financial transactions as they happen in real time, also known as stream computing. (http://www-01.ibm.com/software/websphere/industry/banking/risk-fraud-management/)

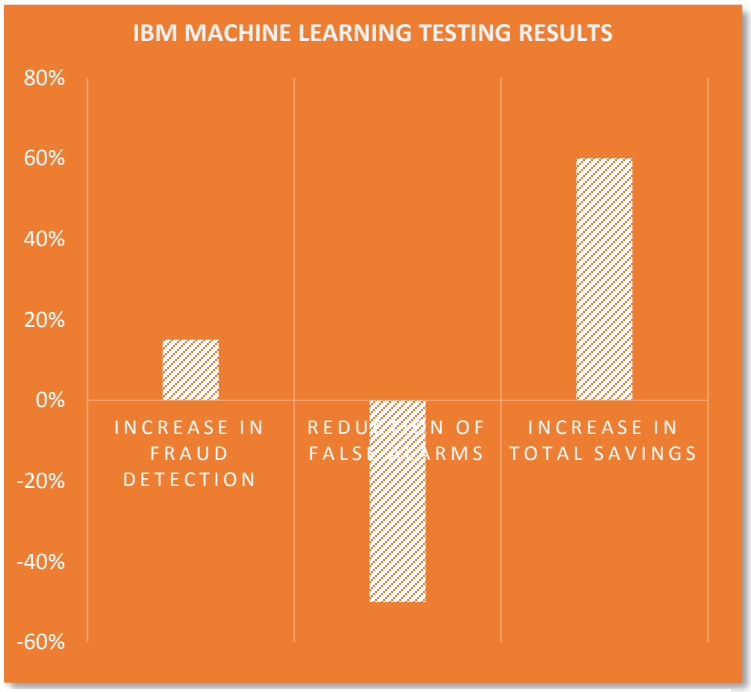

Figure 5. IBM's Machine Learning significantly improves fraud detection at a lower cost

The model is first customized with each customer's information. For every transaction, the model detects its probability of being fraudulent, and then scores the event accordingly. Machine learning methods, along with a constant flow of structured and unstructured data through stream computing, enable

IBM to find complex fraud patterns that could not be detected by humans. The first test on the market have shown that, using this technology, fraud detection rates are higher, there are less false alarms and an increase in total savings. There are still great opportunities for improving machine learning models performance, and further research is needed in order to provide more accurate business solutions.

Ideal future IT models for fraud detection should also allow cross-border B2B co-operation by integrating international private data flows into internal fraud detection systems. In the Internet era, criminal enterprises can move operations in a borderless digital world; in order to keep up with the fraudsters pace of change, companies must create policies in order to join efforts as far as data collection and analysis is concerned. There is a real need of a global, cross-border IT model for financial crime detection and prevention.

It is clear that technology has the power to both enhance human capability to detect and correct patterns related to financial crime, but also make financial firms more vulnerable to digital attacks. This is why, along with elaborated fraud-detection systems, companies must invest in efficient training and development of employees. Financial institutions are facing a growing need for competent "digital forensics" (http://www.ey.com/Publication/vwLUAssets/Annual_Financial_Crime_Trends_Survey_2012/$FILE/EY_Annual_Financial_Crime_Trends_Survey_2012.pdf) that have comprehensive knowledge of systems functioning and the capabilities to extract important insights for the decision making process.

Best practices must also be implemented at the corporate level for managing Big Data systems: companies must ensure the use of high quality data ("separate the signal from the noise", http://www.banktech.com/leveraging-big-data-to-revolutionize-fraud-detection/a/d-id/1296473), they must know their regulatory environment to avoid important penalties, and ensure IT and business units are collaborating, so everybody is working toward the same goal.

3. IMPROVING COMPLIANCE STANDARDS WITH BIG DATA

Despite already being a major concern for financial institutions, compliance and regulation have become even more important in light of the 2009 recession. The major changes have come in the form of the Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank), a law passed in 2010 intended to create regulatory safeguards to prevent another recession. This law has changed the landscape of government regulation in financial services by requiring firms to share significantly more data with the government more frequently in order to facilitate data-driven policymaking. These additional rules make adoption challenging because financial firms must adapt their operational frameworks, IT infrastructure, and workflows quickly to avoid the possibility of adverse legal action and sanctions by regulators. To complicate matters further, the law changes constantly: in fact, according to corporate law firm Davis Polk, only 52% of the rules required by Dodd-Frank have actually been implemented since it was passed into law in 2010 (http://money.cnn.com/2014/07/20/investing/dodd-frank-progress/). Despite this lack of statutory clarity, firms are still required to comply, and those that are found in violation of the law face stiff penalties including civil and criminal prosecution.

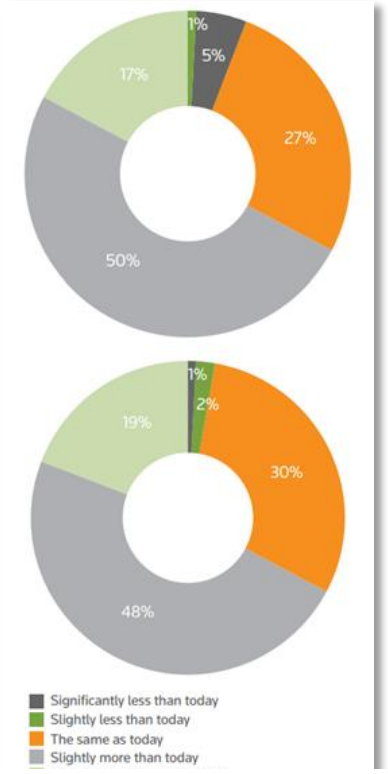

Figure 6. A 2013 study by Thomson Reuters found that compliance

3.1 The Cost of (Non-)Compliance

Despite the fact that compliance and audit divisions are not client-facing units in a firm's core business, they are responsible for one of the most essential operations: keeping firms on the right side of the law. Since Dodd-Frank was enacted, compliance teams have seen increased work related to updating policies and procedures. The increased workload translates to higher costs for an entire typical compliance division. A 2013 survey conducted by Thomson Reuters found that among 800 compliance practitioners at financial services firms, 50% foresaw a slight increase and 17% predicted a significant increase in compliance team budgets over the upcoming fiscal year http://accelus.thomsonreuters.com/sites/default/files/GRC00186.pdf). Additionally, the survey discovered that 67% of practitioners expected a higher staffing costs. When asked the reasons behind these increases, respondents attributed the "implementation of regulatory change" brought forth by Dodd-Frank, US and UK regulatory reform, keeping track of the regulatory changes, and data protection as some of the major compliance challenges driving up costs in the immediate future.

Ironically, a greater source of costs is non-compliance, which often results in severe penalties for firms that are caught breaking rules. In the post-recession era, regulators have made enforcement stricter, leading to multi-billion dollar fines and criminal prosecution. For example, in 2013, JP Morgan Chase reached a record-breaking $13 billion settlement with the US Department of Justice to avoid a lawsuit regarding the alleged sale of fraudulent mortgage-backed securities (http://www.forbes.com/sites/danielfisher/ 2014/02/11/jpmorgan-settlement-challenge-right-on-philosophy-weak-on-the-law/). Considering that JP Morgan Chase made $21 billion in all of 2012, we can see that the potential costs of compliance are growing beyond small back-office expenses and are becoming increasingly hard to predict. Another noteworthy case occurred in May 2014, when Credit Suisse was convicted of criminal conspiracy to aid US taxpayers in committing tax evasion by hiding offshore accounts. As part of the lawsuit, the firm agreed to pay fines of $1.8 billion to the Department of Justice, $100 million to the Federal Reserve, and $715 million to the New York State Department of Financial Services—a $2.6 billion total. In addition, eight of the firm's executives were indicted during the course of the investigation, which started in 2011 (http://www.justice.gov/opa/pr/2014/May/14-ag-531.html). Whether dealing with compliance or noncompliance, banks must ensure that they take proper steps to adhere to the law. Failure to do so can lead to catastrophic monetary and reputational losses.

3.2 Dodd-Frank Data Reporting Requirements

3.2.1 Swap Data

Dodd-Frank's monumental impact on the landscape of regulation in financial services has forced institutions to invest more resources in technology in order to provide the necessary data required by law. As a result, many of the largest Wall Street banks have turned to big data in hopes of garnering deep analysis about their business activities that will help them satisfy reporting requirements and help themselves become more aware of their own operations.

An example of new reporting requirements is Dodd-Frank's new rule regarding swap data reporting, one of the most significant features of the new law's reform on derivatives trading. This section of the law took effect between 2012 and 2013 over a 90-day period and stipulated strict swap data recordkeeping guidelines and reporting requirements for all parties involved in a swap. One of the novel features of this legislation was the creation of an entity called a swap data repository (SDR) that would be designed as a central data warehouse from which data could be prepared for different use scenarios such as real-time public dissemination and confidential regulatory use (http://www.pwc.com/en_US/us/financial-services/regulatory-services/publications/assets/pwc-swap-data-reporting-ready-to-deliver.pdf). Despite providing some degree of direction about the organization of swap data infrastructure, the Dodd-Frank Act failed to specify exactly what kinds of data needed to be gathered. According to Title VII of the act, which authorized the Commodity Futures Trading

Commission to develop the specific requirements of the law, parties involved in swap trading

> *"… must keep full, complete, and systematic records, together with all pertinent data and memoranda, of all activities relating to the business of such entities or persons with respect to swaps, including, without limitations, records or all data required to be reported in connection with any swap… throughout the existence of the swap and for five years following [its] final termination …"*

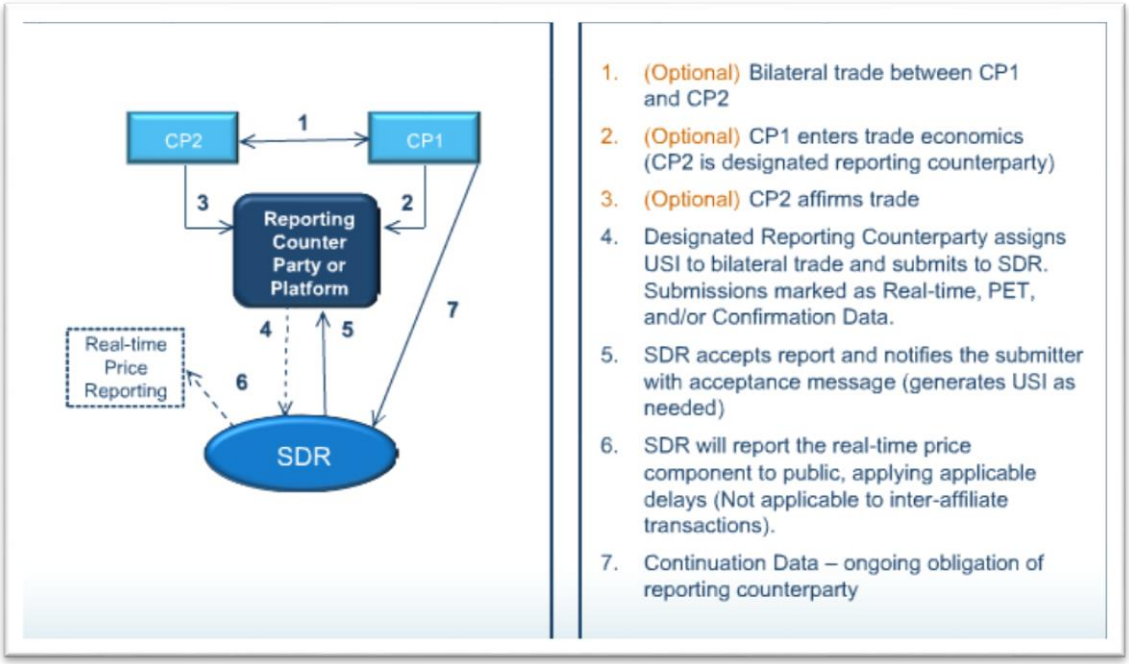

Figure 7. A typical SDR interaction between two reporting counterparties involved in a swap [http://thomas.loc.gov/cgi-bin/query/z?c111:H.R.4173]

Under such a broad definition, financial institutions take a conservative stance on interpreting what data is required to be kept. In typical cases, swap data can include unstructured information such as voice conversations, emails, phone calls, and logs that are often siloed away in application-specific databases (http://www.ndm.net/archiving/pdf/Dodd-Frank%20recordkeeping%20compliance%20requirements.pdf). To make matters worse, given the fact that banks conduct millions of swaps on a frequent, regular basis, the sheer amount of data becomes nearly impossible to manage or regulate with traditional databases. Thus, the only way to manage data and ensure compliance is to adopt a centralized approach that can keep track of interactions between related financial data.

In response to the swap data regulation, a market around Dodd-Frank-compliant SDRs has emerged in recent years. One of the more popular SDRs is the CME Group's SDR, which automatically integrates with their existing financial software CME Clearing, a product that checks the financial integrity of transactions, and CME ClearPort, another product that verifies the validity of contracts. CME's SDR was officially approved by the CFTC in November 2012, and since then has become one of the more widely used SDR systems (http://investor.cmegroup.com/investor-relations/releasedetail.cfm?ReleaseID=722857).

3.2.2 Legal Entity Identifier (LEI)

3.3 Dodd-Frank Restructuring Leads to Big Data Adoption among Regulators

Given the enormous scope of Dodd-Frank, lawmakers realized that a new organization of government agencies would be required in order to implement the desired reform in the industry effectively. As a result, in addition to creating new rules for the financial industry, the Dodd-Frank Act authorized the creation of new government agencies to oversee financial institutions and enforce regulations after all rules are finalized. Among these new agencies are the Financial Stability Oversight Council (FSOC), the Office of Financial Research (OFR), and Consumer Financial Protection Bureau (CFPB), which were created to assess firms' stability and identify potential systemic risks, perform research on financial data, and protect consumers in the financial industry. Under the act, these agencies are given broad powers to create the necessary reporting infrastructure to keep financial markets stable. For example, the CFPB is given the broad authority to "administer, enforce, and otherwise implement federal consumer financial laws, which includes the power to make rules, issue orders, and issue guidance" (http://www.law.cornell.edu/wex/dodd-frank_title_X). Likewise, the FSOC is given the task of conducting research on financial stability without specific details on

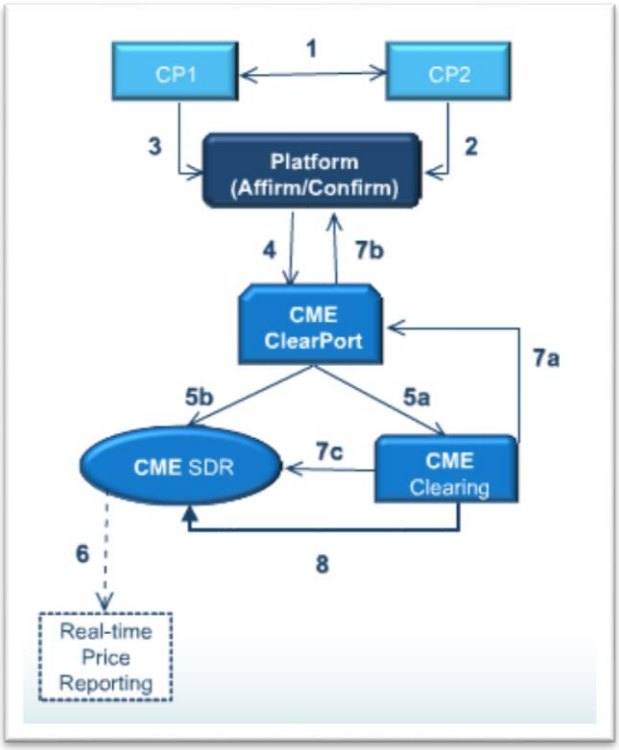

Figure 8. The CME Group's swap data reporting solution

implementation methods. The OFR is also given the responsibility to create a "Data Center" that will collect data from financial institutions for analysis, and a "Research and Analysis Center" to develop tools to gain quantitative insight into market risk and financial positions (http://www.law.cornell.edu/wex/dodd-frank_title_I).

Dealing with these regulatory challenges would be tremendously difficult using existing bureaucratic avenues. Given the freedom to choose how they implement their duties, regulatory agencies have chosen to explore nontraditional methods such big data technologies in order to take on their new responsibilities. One of the most interesting cases has been the CFPB's partnership with Socrata, a Seattle-based cloud software company. In order to satisfy the transparency aspect of its mandate, it decided to open its credit card complaint database to the public. By using Socrata's open data platform, the CFPB has managed to create a highly interactive way to visualize complaint data and hold financial firms accountable to consumers.

Big Data in the service of trading generates value by allowing companies to leverage vast amounts of information in order to exploit opportunities provided by markets. It mitigates the risk of these strategies by restricting the potential damage that this increased ability can wreak. Enhanced Big Data systems for fraud detection provide firms an important competitive advantage for two main reasons. First, by reducing risks, it limits future losses, thus increasing the company's present value. Secondly, appropriate and timely fraud detection methods increase customers' confidence, boosting overall corporate reputation.