Building a Holistic Picture of the US Securities Market From 50 Billion Daily Events

Overview

The Financial Industry Regulatory Authority, or FINRA, is an independent regulator overseen by the Securities and Exchange Commission that is obligated to monitor the markets it operates. FINRA also conducts market surveillance of US equity and options exchanges pursuant to regulatory service agreements and presently conducts cross market surveillance for 99 percent of the listed equity volume and 80 percent of the listed options volume in the U.S. FINRA oversees every brokerage firm and broker doing business with the U.S. public and monitors trading on the U.S. stock markets.

FINRA finds evidence of market manipulation by assembling 50 billion market events into a holistic picture of the U.S. securities market every day. By rewriting this process using Cloudera and other big data technologies, FINRA is creating a regulatory platform that will handle future market growth while containing escalating costs and increasing operational efficiency.

The Challenge

Every day, FINRA must integrate comprehensive data feeds from securities firms and exchanges into a cohesive representation of U.S. securities markets. The result is a huge, and constantly evolving, graph database of market events that supports analytics of various sorts, including ad hoc access by analysts, as well as analysis based on an extensive library of surveillance patterns.

Handling such volumes and complexity would be difficult enough in a static environment, but FINRA must do so while addressing the challenges of a dynamic, evolving market:

- Market volumes are volatile and steadily increasing.
- Exchanges are dynamically evolving.
- Regulatory rules are continuously being created and enhanced.
- New securities products are regularly introduced.
- Market manipulators are constantly innovating.
FINRA has a long history of meeting these challenges through technological innovation, and in 2014, FINRA embarked on a comprehensive journey to shift all market monitoring applications to modern big data platforms in the cloud.

**Solution**

The decision to redevelop FINRA’s market regulation technology came as the result of an exploratory process that began when it became clear that open source technologies such as Apache Hadoop had reached a level of maturity that allowed them to compete with established big data appliances such as Pivotal Greenplum and IBM Netezza. With this in mind, FINRA initiated a series of proof-of-concept projects of gradually increasing scope and personnel commitment designed to identify the appropriate technology stack and vendor relationships, and to provide data to support a business case.

This business case led to the approval, in early 2014, of a program with the following objectives:

- Support future market growth and a dynamic regulatory environment.
- Enable fast, reliable analytics on comprehensive data for end user applications.
- Develop a platform to contain future technology expenditures.

Apache Hadoop surfaced early as the foundation of FINRA’s new platform – it was a good technological fit for the regulatory purpose and had achieved traction both within FINRA and across the data management community. FINRA also valued its open source nature, which promised cost savings and rapid innovation cycles.

A Hadoop architecture allows FINRA to use a public cloud infrastructure, in this case provided by Amazon Web Services. This cloud-based implementation delivers the elasticity, cost-savings, and enterprise-grade infrastructure support that FINRA needs.

Cloudera was selected as the Hadoop distribution for the critical process of building the market event graph database described above, and for providing rapid access to that data for regulatory analysts, based on a few key factors:

- Cloudera’s software platform delivers a pervasive, enterprise-ready Apache Hadoop distribution.
- Cloudera Manager provides a sophisticated and user-friendly management console, simplifying and streamlining Hadoop cluster administration.
- Cloudera’s deep expertise, available through Professional Services and Support, would be critical to solving technical challenges inherent in a Hadoop-based architecture of this scale.

Using Hadoop to build the market event graph database was the first major challenge of the redeployment program. Every day, FINRA receives data feeds containing orders, quotes, and trades from securities firms and various exchanges. Since transactions may span several data providers over days or weeks, understanding order lifecycles requires that huge volumes of data be linked across these different feeds and updated as new information is received. This requires very rapid data access, which was provided by implementing Apache HBase as part of FINRA’s CDH deployment.

Providing access to this data for regulatory analysts was in some ways even more daunting. Rather than just looking at currently active data, analysts need access to at least two years of historical data, preferably interactively. Again, using Cloudera with HBase, order lifecycle graph data output from the process above is repartitioned and distributed over a cluster of servers specifically optimized for rapid data access. This is
an area where Cloudera’s expertise strongly contributed to FINRA’s success. In addition to assistance in a complex configuration, Cloudera designed a snapshot backup for HBase clusters that works at unprecedented volumes and will be contributed back to the Apache community.

**Impact: Keeping Pace with Market Growth**

The new system must be able to handle increases in market volumes. The Cloudera Hadoop platform will assist FINRA in keeping pace with this growth in the following ways:

- By enabling horizontal rather than vertical scaling, FINRA is no longer obliged to buy ever larger, more expensive data appliances to support increasing volumes.
- Similarly, the Hadoop architecture doesn’t impose “largest available appliance” limitations.
- The use of industry standard hardware allows for more flexibility in selecting suppliers and architectures to cope with future growth.

**Impact: Supporting the Mission with Increased Performance & Faster Innovation**

FINRA’s big data platform supports its mission of investor protection and market integrity by helping the organization increase performance and accelerate innovations in the following ways:

- The use of HBase to access the order lifecycle graph database has reduced response times for certain complex queries by orders of magnitude. Queries that took hours now take seconds. One complex query in particular that took ninety minutes to run was reduced to ten seconds. This creates a far more interactive experience, allowing analysts to rapidly iterate and quickly converge on answers that would have been prohibitive in the prior system.
- With its open source core for data processing, FINRA is able to leverage the rapid innovation cycles in the Hadoop and Big Data community.

**Impact: Better Regulatory Systems**

FINRA’s program to implement market regulation applications using Big Data and cloud technologies including Cloudera is containing escalating platform costs and improving operational efficiency. This frees resources to focus on improving systems to provide greater flexibility and effectiveness for regulatory analysts in several key ways.

- The specialized nature of its previous generation data appliances resulted in significant hardware costs and ongoing support requirements. FINRA’s Hadoop-based platform allows the use of less expensive, industry standard hardware.
- Unlike proprietary data appliances, a Hadoop-based architecture lends itself to deployment on a public cloud such as Amazon Web Services. This allows FINRA to benefit from the operational economies of scale of such a vender, as well as providing the elasticity to avoid overprovisioning to handle peak loads.
- The legacy environment was also challenged with capacity limitations, causing FINRA to spend much effort and analysis determining where and when available capacity could be used for various types of analytics. By removing the capacity limitations, the Hadoop/cloud combination virtually eliminates this effort.

“Queries that took hours now take seconds... This creates a far more interactive experience, allowing analysts to rapidly iterate and quickly converge on answers that would have been prohibitive in the prior system.”
• Cloudera Manager automatically manages extremely complex operational tasks within Hadoop, allowing FINRA to focus resources and expertise on areas of strategic regulatory importance rather than fundamental IT tasks. This saves costs while increasing FINRA’s ability to accomplish core objectives.

About Cloudera

Cloudera is revolutionizing enterprise data management by offering the first unified Platform for Big Data, an enterprise data hub built on Apache Hadoop. Cloudera offers enterprises one place to store, process and analyze all their data, empowering them to extend the value of existing investments while enabling fundamental new ways to derive value from their data. Only Cloudera offers everything needed on a journey to an enterprise data hub, including software for business critical data challenges such as storage, access, management, analysis, security and search. As the leading educator of Hadoop professionals, Cloudera has trained over 40,000 individuals worldwide. Over 1200 partners and a seasoned professional services team help deliver greater time to value. Finally, only Cloudera provides proactive and predictive support to run an enterprise data hub with confidence. Leading organizations in every industry plus top public sector organizations globally run Cloudera in production. www.cloudera.com.