cloudera®
Ask Bigger Questions

WHY ARE FINANCIAL SERVICES FIRMS ADOPTING CLOUDERA'S BIG DATA SOLUTIONS?

Table of Contents

cloudera
Ask Bigger Questions

Why Are Financial Services Firms Adopting Cloudera's Big Data Solutions?

Thanks in large part to the availability of data and the movement from in-person to online banking, today's financial firms look very different from those of yesterday. Because the scale of data is overwhelming traditional systems, firms must adapt to new technologies to unlock the power of their data. Cloudera helps companies implement and manage Apache Hadoop so they can derive competitive advantage from Big Data.

Instead of establishing relationships with a local service provider, consumers often choose the most convenient and inexpensive online offering available.

On the Brink. Too Much Data.

The digital revolution has dramatically changed the financial services industry. Four main factors have driven the need for financial services companies to collect, store and analyze massive volumes of data:

1. **Commoditization and digitization of financial products and services.** Consumers no longer need to visit their local bank to make deposits, make investment decisions or complete their banking transactions. Buyers and sellers execute trades online instead of relying on floor traders and brokers. Individuals file their taxes using online tools versus meeting with a tax accountant to prepare and file. As the industry has increasingly moved online, it's become faster, easier, and more affordable for consumers to self-sufficiently handle their own banking and finance transactions.

   The result: financial services and products have become commoditized. Instead of establishing relationships with a local service provider, consumers often choose the most convenient and inexpensive online offering available. Personal connections and customer loyalty have become nearly obsolete. However, every digital action made by a consumer can be captured and analyzed by organizations that seek to understand their customers' behaviors and preferences as they would have traditionally done through face-to-face interactions.

2. **Increased activity.** The ease and affordability of executing financial transactions via online vehicles has led to ever-increasing activity and expansion into new markets. Individuals can make more trades, more often, across more types of accounts, because they can do so with the click of a button in the comfort of their own homes, or on the go from a mobile device. Individuals in South Africa can make trades in the US stock market over the internet. Increased access and ease of use translates into increased activity, which in turn translates into rapidly growing data volumes.

   It is important for banks, investment firms, and other financial services organizations to be able to collect and analyze this information in order to accurately assess risk and market trends. This became painfully evident during the massive market crash of 2007-2008, when banks and brokerage houses scrambled to understand the implications of massive capital leverage and their ability to model and refine liquidity management. Just one bank captures internal transactions exceeding two billion per month, in addition to collecting public data of over a billion monthly transactions. These tremendous transaction volumes have made it nearly impossible to create models that take into account multi-year data sets using detailed data. Financial firms manage anywhere from 10PB to thousands of petabytes of data total, yet most systems they use today build models using only samples — as little as 100GB at a time. Even with large memory technologies that can accommodate one terabyte (TB) or more, that's still a fraction of a percent of the multi-petabytes (and growing) of data that firms have access to. Relying on data samples requires aggregations and assumptions, resulting in large inaccuracies in projections, limited visibility into actual risk exposure, instances of undetected fraud, and poorer performance in the market.

cloudera®
Ask Bigger Questions

The ease and affordability of executing financial transactions via online vehicles has led to ever-increasing activity and expansion into new markets.

3. New data sources. The digital revolution has led to new sources of data that are complex to ingest, such as data from derivative trading platforms, social media, blogs and other news feeds. This information, if combined with individual financial transactions and history, can help to paint a holistic picture of individuals, families, organizations, and markets.

Bringing together large volumes of data from many sources and in a variety of formats — including both structured and unstructured — is impossible using traditional relational database and data warehousing technologies. The pains and costs associated with capturing, combining, and making this data usable are overwhelming for organizations to accomplish successfully.

4. Increased regulations. In recent years, federal stress tests have increased the demand for predictability and integrated solutions for capital asset management. Stringent regulatory compliance laws have been put in place to improve operational transparency. Financial services organizations are held much more accountable for their actions, and are required to be able to access years of historical data in response to regulators' requests for information at any given time. For example, the Dodd-Frank Act requires firms to maintain records for at least five years; Basel guidelines mandate retention of risk and transaction data for three to five years; and Sarbanes-Oxley requires firms to maintain audit work papers and required information for at least seven years. These records must be available on demand, or in some cases must be normalized and sent to regulators proactively.

Partly because of these pressures, leading financial services companies have realized that the key to optimizing their business operations is in maintaining an efficient and large-scale data management infrastructure. This is very expensive and complex to accommodate using traditional systems.

Due to these four main factors, the scale of data that financial services companies need to manage today is overwhelming traditional systems. It's truly a Big Data problem: six years of publicly available market data amounts to about 200TB, and the proprietary data collected by individual firms today adds up to tens of petabytes (PBs) altogether. Firms must adapt to new technologies in order to unlock the power of their data. Financial services companies' ability to access this data and turn it into action is driving competition; those that put all of their data to use have a significant competitive advantage over those who don't.

Ask Bigger Questions

The Hadoop Opportunity

The world of finance set out to find new technologies that would allow them to manage and take advantage of larger data sets. They discovered the technologies deployed by massive web and online advertising companies like eBay and Facebook. In short, they found Hadoop: open source software that enables distributed parallel processing of huge amounts of data across inexpensive, industry standard servers. With Hadoop, no data is too big or complex.

These technologies are now available for enterprises in the form of Cloudera Enterprise, a combination of three things: Cloudera's open source Hadoop stack (CDH), a powerful management platform (Cloudera Manager), and Cloudera's expert technical support.

Cloudera Enterprise is a stable and reliable platform that financial organizations can trust for their Big Data needs. Many of today's top financial services firms have already deployed Cloudera Enterprise to create central data hubs combining large volumes of diverse and detailed data, creating competitive advantage in several key business applications:

> **Consumer risk models:** giving greater insight into liquidity and capital availability.

> **Personalization and recommendations:** opening doors to sophisticated cross-sell and up-sell techniques so firms can offer new financial products to their existing high value customers.

> **Fraud detection and anti-money laundering capabilities:** allowing firms to find critical breaches faster than ever, saving money that directly affects the bottom line.

> **Market risk models:** using detailed data to give firms better insight into the real behavior of complex instruments and creating more predictable results from their investments.

> **Mortgage portfolio valuations:** more accurately identifying which mortgages are candidates for prepayment or default.

The rest of this paper will explore each of the aforementioned big data application areas in greater detail.

cloudera®
Ask Bigger Questions

Using Big Data solutions driven by Cloudera allows financial services organizations to optimize capital leverage while maintaining the reserves required by regulators.

Consumer Risk Modeling

> **Better understand customers through automated consumer facing services**

> **Unmask dependencies on groups with related risk profiles for liquid capital**

> **Optimize capital leverage while maintaining adequate reserves**

The variety and volume of digital interactions — e.g. credit and debit purchases, direct deposits, and detailed web browsing trails — is increasing the abundance of automated consumer-facing service opportunities while generating massive data volumes. But even though firms have more access to data about their customers than ever before, they know less about those customers today than they did years ago. Where banks were once able to make local decisions based on personal relationships, they now have global relationships with less personal knowledge of their customers.

The wide reaching effects of the economic crisis in 2008 were largely due to the lack of visibility into consumer actions and groups with related risk profiles, combined with the increased flow of consumer funds into firms. Companies (and the underlying economy) were more susceptible than statistical models had predicted. Since realizing this, financial firms have recognized the need to alter how they look at the data "exhaust" produced by online interactive systems. The Cloudera Difference: Hadoop enables the large-scale data ingestion that is necessary to keep pace with today's digital world, capturing and analyzing all interactions that allow companies to holistically understand their clients. Billions of digital, detailed interactions can now be used to redefine customer groups and to tailor information for customers' future interactions. Using Cloudera Enterprise and advanced analytics, firms can continuously model and refine consumer groups while maintaining higher quality risk profiles. And by linking these risk profiles to market events, firms can make more precise decisions about capital investment. Using Big Data solutions driven by Cloudera allows financial services organizations to optimize capital leverage while maintaining the reserves required by regulators.

By tracking user interactions across systems and channels in real time, firms can categorize customers dynamically and experiment with uniquely targeted offers, investment strategies, and savings advice to help customers improve their financial discipline, thus resulting in a better and more personalized client experience.

Personalization and Recommendation

> **Differentiate offerings based on a firm's collective intelligence**

> **Cross-sell and up-sell to a well-known consumer base**

> **Retain loyal customers with personalized service**

With the move from in-person to online banking and finacial services, firms have been able to offer more convenient, functional services to customers digitally while sacrificing the personal touch. If companies can harness the data from today's digital fire hose, they are better equipped to tailor services to their clients at a personal level. By integrating and analyzing data across multiple channels, from website click stream data to account profile information to social media activity, financial organizations can target customers with product offerings that better match their financial needs. The same methodologies used to group consumers for risk assessment are employed to feed recommendation engines for cross-sell and up-sell opportunities. Access to detailed data from a variety of sources is providing firms with a consistent and comprehensive view of their customers across all operational dimensions. This type of view has not been possible since the introduction of electronic banking.

cloudera®
Ask Bigger Questions

The Cloudera Difference: Cloudera Enterprise opens the door to a more personalized online experience. By tracking user interactions across systems and channels in real time, firms can categorize customers dynamically and experiment with uniquely targeted offers, investment strategies, and savings advice to help customers improve their financial discipline, thus resulting in a better and more personalized client experience. This also gives firms the ability to build brand loyalty and differentiate themselves in what has become an increasingly commoditized financial services market.

Fraud Detection and Anti-money Laundering

As financial firms roll out new systems and features to customers, the perpetrators of fraud are constantly adapting to fraud prevention techniques.

> **Respond to new threats from increased access points and complex transactions**
> **Identify the signals left by perpetrators hidden in the daily activity noise**
> **Respond to coordinated attacks with consolidated monitoring**

In the race to combat fraud and money laundering, financial firms are challenged to create a balance between increasing service offerings to customers while at the same time reducing exposure to perpetrators. Challenges in fraud detection have increased dramatically with the introduction of new access points to financial services offerings and increased sophistication of perpetrators. Furthermore, as companies expand into new markets, they face new risks that must be modeled.

With each new access point, firms are more susceptible to new methodologies and ever more complex cross-channel fraud. For example, customers commonly create separate accounts for different service offerings without linking them; they may create a checking account for a relative without also giving the relative access to their personal savings. With multiple accounts sharing access to the same financial products, a perpetrator has the opportunity to execute complex trades and transfers between independent accounts to mask fraudulent behavior.

The demand for more data-driven market risk modeling is increasing at a feverish pace. It's no longer enough to rely on brilliant quants with complex algorithms, which leave firms susceptible to hidden deficiencies and irreconcilable predictions. The most valuable tools available to help firms combat fraud are detailed traces from all operational systems. Since perpetrators work hard to exploit gaps in financial systems, firms must be vigilant and self-aware of where they are exposed. By collecting detailed behaviors from online channels and automated systems, fraud detection teams can recombine logically linked accounts by looking for common patterns of money movement. Much in the same way that social networking companies identify unlinked relationships, anti-fraud teams search for connections that are implied by the detailed trace data. Collecting detailed information on both customer and internal interactions leads to new models that help identify patterns of normal and suspect behavior.

The Cloudera Difference: The process of instrumentation, collection, research and analysis is an ongoing cycle for firms. As financial firms roll out new systems and features to customers, the perpetrators of fraud are constantly adapting to fraud prevention techniques. Sophisticated firms are using Cloudera Enterprise to perfect their fraud detection models. Using Hadoop's ability to capture and process detailed data, these companies are able to compare experimental results with simulations that leverage actual recorded data. They deploy large testing sandboxes for research and development and real-time fraud detection engines to catch bad behavior in flight. This tiered system combining development, simulated testing and real time detection has proven highly effective in improving the quality and reducing the time to detect new cases of fraud.

Market Risk Modeling

Using Cloudera Enterprise, firms can build and assess models in hours instead of weeks, leading to significant competitive advantage.

> **Introduce new detailed data sources into models**

> **Use simpler algorithmic building blocks**

> **Build multiple models in parallel**

Access to ever increasing sources of data presents an opportunity and a challenge as firms expand the sophistication of their market risk modeling. Traditionally, organizations have focused on the market measures from simple daily open/close prices all the way to detailed options quotes. By adding information from internal trading systems, logs from trade execution systems and external sources such as social media data and news feeds, firms now have access to larger volumes and richer data than ever before. This robust collection system creates a new challenge: how to make sense of all this new data?

For decades, financial firms have focused primarily on hiring smarter scientists to build better algorithms. The increased availability of data allows firms to rethink how they approach problems such as market risk modeling. Instead of sophisticated market models based on sampled data sets, analysts can build simpler models that take into account all of the detailed data. For example, while it is possible to run a static analysis based on fundamentals and related market performance, these models make gross assumptions about what's driving the market and distribution of trading activities. Instead, by collecting all possible quote and trade data, combined with historical events, firms can build out more accurate models based on actual distributions.

The Cloudera Difference: Cloudera Enterprise breaks down the barriers in existing market data modeling systems. By placing computation next to the data and scaling out with cost effective industry standard servers, firms are able to build multiple models in parallel. This rich capability reduces the time to experiment and identify possible risks in open positions. Using Cloudera Enterprise, firms can build and assess models in hours instead of weeks, leading to significant competitive advantage.

Mortgage Portfolio Valuation

Cloudera Enterprise expands support for unstructured data sources offering greater visibility into mortgage models, such as property valuation sites, local consumer offers such as shops, crime report data and public sentiment data.

> **Access external influences, the largest variable in mortgage payment modeling**

> **Remove compute and storage limitations on market conditions and payment modeling**

> **Evaluate models using higher quality multi-variate comparisons**

When determining mortgage portfolio valuations, collecting data is the first step toward gaining visibility. In order to forecast, firms build out predictive models, changing various inputs such as rates for interest and inflation, direction of the market or expansion or contraction of local consumer services. These models produce enormous data volumes that must then be cross-referenced with actual market behavior.

By ingesting unlimited data volumes and types, companies' predictive models become more accurate, allowing firms to map out mortgages that are going underwater or are likely candidates for prepayment.

The Cloudera Difference. Cloudera Enterprise enables firms to store longer consumer history based on payment data and structured external credit data sources. Cloudera Enterprise also expands support for unstructured sources such as property valuation sites, local consumer offers such as shops, crime report data and public sentiment data. By combining this data geographically and aligning it chronologically, firms are able to enrich the attributes related to particular mortgages and apply more data to their models.

cloudera
Ask Bigger Questions

Conclusion

Using Cloudera Enterprise, firms have access to a powerful platform providing both highly scalable and low cost data storage tightly integrated with scalable processing. Financial firms are now able to tackle increasingly complex problems by unlocking the power of their data. The capability to understand and act upon their data opens the door to a richer and more robust financial ecosystem.

The Result: firms can better serve their customers, understand their risk exposure and reduce incidents of fraud. Cloudera is your Hadoop implementation and management partner, helping you derive competitive advantage through the use of your Big Data.

About Cloudera

Cloudera, the leader in Apache Hadoop-based software and services, enables data driven enterprises to easily derive business value from all their structured and unstructured data. As the top contributor to the Apache open source community and with tens of thousands of nodes under management across customers in financial services, government, telecommunications, media, web, advertising, retail, energy, bioinformatics, pharma/healthcare, university research, oil and gas and gaming, Cloudera's depth of experience and commitment to sharing expertise are unrivaled.