cloudera®

WHITE PAPER

Lowering Business Risk with Data

Version: 101

At its core, risk is the exposure to loss. When looking at risk through the lens of business, there are many issues that organizations must navigate in order to avoid monetary and reputational losses. A majority of these business risks manifest themselves in the form of cyber threats, fraud, and ever-looming compliance regulations. And despite these risk areas advancing over the years, they simply do not seem to be getting the attention they deserve. Why else would we as an industry continue to lose $190B[1] to credit card fraud annually? Why did we have half a billion[2] personal records stolen in 2015? And why do we expose ourselves to potentially losing 2 percent of revenue to a new compliance regulation?[3]

While external risks are changing every day, so do the complexities that enterprises face behind the scenes to keep up with these changes. Applications continue to be bolted on as band-aid fixes, emerging technologies open up new threat surfaces, and data that can help combat risk continues to sit stagnant as a cost to the organizations. But it doesn't have to be that way. There is a way to future-proof your risk strategy by making fundamental shifts in the way you leverage data and analytics.

We will spend some time discussing how open source technology is helping organizations leverage data and analytics in new ways to reduce business risk as it relates to cybersecurity, fraud, and compliance.

Cybersecurity

Cybersecurity has become the topic of conversation for organizations across every industry. With the average breach costing $200 per lost customer record,[4] and even more for lost intellectual property, organizations are looking for new ways to detect and investigate cyber threats. As attackers have become more sophisticated, attack surfaces have expanded, and the number of attacks has increased, organizations find themselves exposed to an onslaught of novel and previously unseen attacks. Combined with the threat of inside rogue users, it's clear organizations face an enormous challenge. The tools available to the Security Operations Center (SOC) are not built for the hyperconnected world they now operate within. Cloudera's Enterprise Data Hub, built on Hadoop and the latest open source project, brings rise to a new class of cybersecurity solutions designed to detect previously unseen threats early in the kill chain—helping organizations avoid financial and reputational damage.

The threat landscape is changing rapidly. The number of touch points is exploding and so is the number of entry points for malicious activity. Hackers are getting more sophisticated. With traditional cybersecurity systems, such as a security information events management (SIEM), organizations face data and analytic constraints that are causing threats to go unnoticed and data breaches to happen. SIEM cannot monitor every corner of the enterprise because of technology and economic constraints; they can't discover known threats until it is too late, and they only hold a subset of data that makes it difficult to use historic data for investigation and remediation. With 71 percent[5] of organizations saying it is impossible to leverage advanced analytics on traditional systems to discover advanced threats, this is forcing organizations to rethink their cybersecurity strategy.

A new approach is needed to address the changing threat landscape. That is where the open source ecosystem and Apache Hadoop come into play. Hadoop modernizes an organization's cybersecurity architecture to detect advanced threats 2.25 times faster[6] and accelerate threat mitigation leveraging big data and advanced analytics (machine learning, predictive analytics, etc.). Unlike traditional solutions that provide signature and correlation analysis across subsets of security data, a Hadoop-based cybersecurity solution can ingest, store, process, and analyze any volume of data with any analytic type. Having access to all the raw data in one place can help uncover new insights and patterns. This allows for behavior-driven analytics that can detect

[1] "Solving the $190 Billion Annual Fraud Problem: More on Jumio"—Forbes. 12 Oct. 2016
http://www.forbes.com/sites/haydnshaughnessy/2011/03/24/solving-the-190-billion-annual-fraud-scam-more-on-jumio/1

[2] "Over Half a Billion Personal Information Records Stolen"—Symantec." 2016. 12 Oct. 2016
https://www.symantec.com/content/dam/symantec/docs/infographics/istr-reporting-breaches-or-not-en.pdf

[3] "General Data Protection Regulation"—Wikipedia, the free encyclopedia. 2013. 12 Oct. 2016
https://en.wikipedia.org/wiki/General_Data_Protection_Regulation

[4] Data Breach Costs Top $200 Per Customer Record"—CIO. 2016. 12 Oct. 2016
https://www.cio.com/article/2421114/security0/data-breach-costs-top--200-per-customer-record.html

[5] "Cybersecurity Analytics with Apache Hadoop"—Cloudera. 2016. 13 Oct. 2016
http://www.cloudera.com/resources/infographic/cybersecurity-analytics-with-apache-hadoop.html

[6] "Cybersecurity Analytics with Apache Hadoop"—Cloudera. 2016. 13 Oct. 2016
http://www.cloudera.com/resources/infographic/cybersecurity-analytics-with-apache-hadoop.htm

the smallest changes in user or system behavior—traditionally the most reliable indicators of compromise. Integrating existing cyber defenses, Hadoop allows organizations to quickly deploy and improve their security posture with no disruption.

Fraud

Challenges in fraud detection have increased dramatically with the introduction of new access points to service offerings and increased sophistication of perpetrators. Furthermore, as companies expand into new markets, they face new fraud risks that must be modeled.

With each new access point, firms are more susceptible to new methodologies and ever more complex cross-channel fraud. Let's take money laundering as an example. Unlike other forms of fraud that are identified with machine learning algorithms that detect anomalies and outliers, money laundering schemes are designed to closely mimic typical banking behaviors—hiding clandestine or illegal activities behind a taxable, auditable mainstream business front—and are, therefore, characteristically less anomalous. The thresholds mandated by reporting policies like Bank Secrecy Act (BSA) and utilized by first- and second-generation AML systems are well known, so criminals have little difficulty modeling the source of their above-board trade and transaction behaviors to be largely imperceptible, even to specialized software. As a result, these systems must be enriched with much larger and more diverse data sets to isolate signals of possible money laundering. Apache Hadoop is the ideal platform for AML because it augments all of the core functions of a specialized system to better handle big data: data collection, data preparation, automated evaluation, model building, and investigation. The modern AML architecture is fully integrated with an enterprise data hub, with Hadoop initially staging massive complex data for legacy solutions to provide run times for the predictive models and perform the actual fraud detection. Beyond the introductory use case of more expansive and affordable storage, Hadoop's natural fit for back testing against long-term descriptive data is gaining popularity for more advanced AML workloads, as is the use of other components in the Hadoop stack for exploration, discovery, investigation, and forensics.

More data-driven risk modeling is needed to address the types of fraud in this changing landscape. It is no longer sufficient to rely on brilliant quant staff with complex algorithms driven by data samples, which leaves firms susceptible to hidden deficiencies and irreconcilable predictions. The most valuable tools available help firms combat fraud by providing detailed traces across all operational systems. Since perpetrators work hard to exploit gaps in financial or billing systems, firms must be vigilant and self-aware of every place where they are exposed. By leveraging Hadoop to collect and analyze detailed behaviors from online channels and automated systems, fraud detection teams can combine logically linked accounts by looking for common patterns of money movement and related transactions. Similar to the way social networking companies find relationship links that are complicated to identify, antifraud teams search for connections that are implied by 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. Fraud detection needs advanced analytics, including data mining techniques. Hadoop brings the processing power to analyze massive amounts of data to quickly identity and prevent fraud.

Compliance

The cost and complexity of compliance for organizations have escalated significantly in recent years. Stringent regulatory compliance laws have been put in place to improve operational transparency, digital privacy, and customer protection. 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 a minimum of seven years. These records must be available on demand, or in some cases must be normalized and sent to regulators proactively. Increasingly, organizations are setting up internal audits to identify and prevent rogue employee behavior or loopholes in internal systems or processes.

Frequent changes in regulations have tested the ability of legacy compliance systems to respond in a timely manner. Partly because of these pressures, leading 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.

Cloudera helps organizations create a secure, auditable, and searchable unified repository for all the data at a fraction of the cost of traditional systems. This platform can be used for multiple risk and compliance use cases: PCI, HIPAA, VAR, Monte Carlo simulations, CCAR, Basel III, Solvency II, BCBS239, MiFID II, FRTB, etc. New data sources can be added easily and changes can be implemented more quickly. By eliminating data silos and time-consuming processes such as ETL/ELT, the time it takes to prepare the data for analysis can be drastically reduced.

How would I get started?

We are here to help. The first step is to arrange for an initial meeting with our team to understand what you're trying to accomplish, and identify business-impacting use cases. After that, most of our customers prefer to do some form of proof-of-concept to get a sense of the power of our solutions.

In some cases, we'll recommend one of our systems integrator partners, or perhaps you'd rather engage directly with Cloudera's professional services team—either way is fine with us. And, whether you want to deploy on-premise or in the cloud, we will support you 100 percent.

Lastly, to help ensure the likelihood of success, we strongly recommend you provide your team with training. In fact, Cloudera provides comprehensive and in-depth training for all your Hadoop needs, and for various roles within the organization. In addition to making sure you get the most out of your Cloudera deployment, our training will help you with recruiting and retention of the DevOps people essential to Hadoop.

Summary

The risk landscape is changing rapidly. Future-proofing your risk strategy by using open source technologies to leverage data and analytic more effectively is key. Unifying data, opening up access, and applying advanced analytics against it will allow you to stay one step ahead of the looming risk threats. Cloudera is here to help you reduce your overall risk exposure by more effectively leveraging your data.

About Cloudera

Cloudera delivers the modern platform for data management and analytics. The world's leading organizations trust Cloudera to help solve their most challenging business problems with Cloudera Enterprise—the fastest, easiest, and most secure data platform built on Apache Hadoop. Our customers can efficiently capture, store, process, and analyze vast amounts of data—empowering them to use advanced analytics to drive business decisions quickly, flexibly, and at lower cost than has been possible before. To ensure our customers are successful, we offer comprehensive support, training, and professional services. Learn more at cloudera.com.

cloudera.com

1-888-789-1488 or 1-650-362-0488
Cloudera, Inc. 1001 Page Mill Road, Palo Alto, CA 94304, USA