

The Business Value of an Enterprise Data Hub



Table of Contents

Overview	3
Business Meets Data	3
Business Opportunity: Expand Intelligence of Customers and Channels	4
Business Opportunity: Products and Services	7
Business Opportunity: Managing Security, Risk, and Compliance	9
Summary	11
About Cloudera	11

Overview

Big data is no longer just hype. As the Apache Hadoop ecosystem has evolved and matured, enterprises are graduating from evaluation and prototype into production, deploying Hadoop-based platforms as enterprise data hubs to deliver transformative business value.

Hadoop brings the benefits of high performance and predictable scalability without schema restrictions to complex data with minimal cost, making it very attractive to budget-conscious IT organizations. But its relevance to broader business objectives requires a platform approach that ensures superior performance and operational efficiency aligned to a wide variety of use cases and additional functionality.

To achieve full value from all your data, an enterprise data hub architecture builds on Hadoop's principal benefits by delivering an array of powerful processing, exploratory, analytic, and real-time capabilities certified to work seamlessly with existing infrastructure and tools. Cloudera's enterprise data hub complements the flexibility and extensibility of the open source tools at its core with the systems management, data governance, and robust security that enterprises demand of all their production systems. With this powerful new architecture, you can streamline your path to a data strategy and drive action from the rich insights your business requires.

Business Meets Data

The most significant challenge facing today's enterprise is how to maximize the business benefit of data. Big data has become a global watchword—for the past few years, tech-savvy companies have been using an abundance of information to challenge traditional models and drive new business opportunities. Now, leaders across all industries are focused on building new capabilities and becoming more agile, competitive, and market-driven using data.

However, data size and its practical value have long maintained an inverse relationship. Historically, larger and more complex data sets required custom management and analytics solutions, which placed a premium on the tools and their specialized capabilities but treated the data itself as a commodity. The rich insights about customers, operations, and markets that drive business growth and efficiency were limited by data silos and the prohibitively high cost of scaling legacy solutions or deploying new ones.

Today, the introduction of an enterprise data hub at the core of your information architecture promotes the centralization of all data, in all formats, available to all business users, with full fidelity and security at up to 99% lower capital expenditure per terabyte compared to a traditional data warehouse.

An enterprise data hub serves as a flexible repository to collect and keep unlimited data, whether for compliance purposes or for sophisticated applications such as real-time anomaly detection. It speeds up business intelligence reporting and analytics to deliver markedly better throughput on key service-level agreements. And it increases the availability and accessibility of data for the activities that support business growth and can provide a full picture of your operations to enable process innovation.

Most importantly, centralizing and enabling analytics across all your data unlocks new data-driven business opportunities that were previously too expensive or complex for most enterprises. An enterprise data hub delivers advanced capabilities—such as synchronous customer models based on social networks and offline behaviors, truly real-time analysis of streaming data-in-motion, proactive security against fraud and cyber-attacks—with a unified, flexible, scalable platform that's quick to implement and easy to grow with your business.

Business Opportunity: Expand Intelligence of Customers and Channels

In order to win and retain customers, today's leaders realize their organizations must become information-driven. This means developing a culture where individuals have access to the right information at the right time, and are empowered to make decisions with the facts at hand. Only by building a comprehensive view of their customers and channels can organizations hope remain ahead of the competition.

Build a 360-Degree View

Perhaps no industry has been more affected by the data deluge than retail. E-commerce now compels even brick-and-mortar stores to develop an integrated online sales and marketing strategy. Supply channels are more complex, and the web has removed barriers that previously kept manufacturers from vertically integrating into the retail space. Retailers now have the opportunity to capture and analyze real-time interaction and behavioral data, combining it with transactional data to get more complete perspectives into the customer journey across in-store, online, and mobile channels. Unfortunately, these repositories can also be an attractive target for cyber-attackers that threaten privacy, denial of service, and theft.

Service providers are among the world's biggest aggregators of consumer data and work under the most uncertain regulatory conditions. Securely storing billions of records and providing transparent, real-time customer access has historically required multiple expensive systems to handle the huge size, complexity, and variety of data.

At the same time, healthcare organizations have a unique opportunity to map the patient experience across care centers, online and at home. With the move towards the digitization of patient data, together with the adoption of wearables, fitness trackers, and other connected health devices, care providers are able to build a more complete picture of how treatments and preventative measures impact patient health.

With an enterprise data hub, these organizations can map diverse human interaction, transaction, and device data across multiple platforms in a single place at considerably lower cost. Given rapid systems consolidation and the competition to acquire and retain business, you have the tools to research, develop, and bring new products to market in a fraction of the time. A deeper and more synchronous view of behavior - while maintaining security and privacy - is a prime benefit of a powerful, centralized data hub.

Magnify Analytic Solutions delivers a self-serve, 360-degree customer view to Fortune 100 clients like Chrysler, DuPont and Ford. Magnify has built broad expertise processing large datasets for customers to support things like business-to-consumer (B2C) online marketing contests and product giveaways. Magnify recently managed an automotive client's customer relationship management (CRM) system and evolved it into a centralized data hub delivering a 360-degree view of customers, encompassing broad data types from different sources, including vehicle information from local motor vehicle departments, dealer distribution statistics, and parts data from services organizations. Magnify now offers clients a web-based solution through which they interact directly with Hadoop.



Increase Conversion Rates

One of the most obvious areas of focus for online retailers is in improving their conversion rates. Why do shoppers abandon their carts? What would help them drive towards a purchase decision faster, even increase the scope of each transaction? If they leave, how can we get them back? Data holds the answer, but only by looking beyond the transactional record, to clickstream, social, and location data. Capturing this diverse and voluminous data is enough of a challenge using traditional infrastructure. Analysts then need the freedom to explore and model customer behavior, and the ultimate benefits can only be realized if those models can be embedded into the online experience to impact that behavior at the point of interaction.

Many leading retailers now rely on an enterprise data hub to complement their existing analytical approaches, giving their product, marketing, and data science teams the power and freedom to test offers and recommendations in real-time, analyze the results, and iterate quickly. Experimentation is key to a modern online product strategy, and an enterprise data hub helps make it possible.

Quaero helps clients such as ESPN, MSNBC, Keurig Green Mountain, and Bravo gain insight into individual consumer behavior. Its AdVantage platform uses digital data to improve acquisition, increase engagement across channels, and deliver revenue. A Cloudera-powered enterprise data hub provides AdVantage the flexibility, scalability, and performance to process behavioral data that traditional MPP systems weren't built for. As the platform manages both anonymous and identified data, Quaero can stitch together disparate pieces of information. This helped one media and entertainment client turn "casual clickers" into "engaged audiences" with double-digit growth. The client was able to specifically measure a 250% increase in its active registered user base and a 300% increase in campaign response rates.



Reduce Customer Churn

Voluntary churn is the service provider's enemy; unfortunately it's increasingly difficult to understand what keeps customers happy or why they leave. As the number of touchpoints increase - at home or at work; online, mobile, or physical - the opportunities for a brand to thrill or disappoint, even anger us, increase in kind. Predicting customer churn risk and judiciously applying tactics to retain key customers - such as those with high social media influencer scores - has never been more critical.

Leading organizations know how to connect the dots of behavior and desire to assess the right channels to delight their customers. Nothing beats engaging directly with loyal users to increase their likelihood to recommend, or with dissatisfied users at the critical moment to avoid attrition. An enterprise data hub combines the analytical power necessary to build comprehensive models of churn behavior with the real-time capabilities to take action on those models.

SFR, France's second largest telecommunications operator, sought a way to improve the experience for its 21 million mobile clients. To deliver on this goal, SFR leverages a Cloudera-powered enterprise data hub to provide employees across the organization with a 360-degree customer view that is accessible for real-time search, reporting, and analysis. To truly understand the customer journey and minimize customer churn, SFR recognized it needed to bring in multi-structured data from new sources, such as customer behavior across SFR's many channels.



Grow Share of Wallet

These days, loyalty isn't enough. New channels create new opportunities for customers to interact with a brand's products and services, or opportunities for competitors to encroach. Media companies know this all too well, as online streaming services battle for market share in a space that hardly existed even a few years ago, and designing the right bundling strategies becomes imperative. Financial institutions compete across a broad array of offerings, where the best integrated customer experience can be the difference between a committed customer and one who will seek more compelling offers elsewhere.

By developing a more complete view of consumer behavior through analysis and modeling of the diverse data available in an enterprise data hub, organizations drive strategies to grow share of wallet and identify new revenue channels, solidifying or expanding their competitive positions in the market.

A leader in the casino-entertainment industry, **Caesars** has used big data for years to improve the customer experience. As the industry has evolved, a greater share of customer spending is happening in the non-gaming aspects of its properties. By implementing a Cloudera Hadoop environment, Caesars can now integrate and analyze the diverse data sources required to study its customers across these properties. With this data, they can deliver promotions and offers that reflect the overall customer value and experience, not just for a single channel or property.



Business Opportunity: Products and Services

In the modern business environment of globalization, facing increased competition, leading organizations have learned to leverage data to build and improve winning products and services.

Improve Product Quality

Every company is concerned with the quality of their products and services. With the increased online availability of consumer reviews and recommendations, coupled with the persistent risk of a bad customer experience gone viral, it's never been more important to use data to improve product quality. Assessing quality, however, is extremely challenging. Unfortunately for the analyst, root causes are buried in unstructured text data and not easily analyzed. For example, in the automotive space, the answer might lie in the mechanic's notes after fixing a car, in an online forum for upset customers, in a government test report, in an insurance claim. Any one of these might hold the key to finding the cause of a current problem, or uncovering a potential future issue and avoiding it in the first place.

With an enterprise data hub, it is now possible to aggregate quality data from the entire manufacturing process and supply chain, from customer service channels, and from the online and social sphere, giving organizations the opportunity to comprehensively assess quality, and align investments where they are most needed.

Omneo has deployed a multi-tenant enterprise data hub behind its supply chain cloud solution, which ingests machine data and existing system data from throughout the manufacturing process, including from clients' factory data, supplier data, field services, after-market repairs, and re-manufacturing data.

The solution delivers a 360-degree view of the supply chain process in seconds, allowing manufacturers to access their data in different ways, on the fly. If something happens at any supplier that drives a sudden increase in quality issues, they can figure out where the issue stems from and why in minutes or hours. In traditional environments, these investigations would take weeks or months.



Accelerate New Product Introduction

Successfully delivering innovative new products to market is challenging for many reasons. The risk of a dud product can be measured in millions of dollars worth of marketing expense and surplus inventory, not to mention brand damage. Accurate data on customer needs and behaviors informs product-market fit, but often these data are trapped in silos or simply unavailable due to the difficulty or expense of capturing it all. Every product team would love to have the ability to gather all relevant data and get real-time feedback from customers, but traditional data management and analytic approaches make this all but impossible.

By centralizing and enabling immediate access to critical data, an enterprise data hub helps break down business silos that impede effective collaboration during product development, testing, and launch. Teams have access to more data with which to quickly assess the impact of their product and, ultimately, its success in the market. The powerful combination of mobile devices, embedded sensors, clickstream aggregation, social media, and more enable organizations to gather real-time intelligence and feedback from customers, so they can get a faster read on successful launches or, on the other hand, avoid inventory buildup of an unsuccessful product, potentially saving millions.

It can take **Monsanto** from 5-10 years to bring one new product to market because of the intensive research, testing and evaluation that needs to be done during the R&D process for its seeds and genetic traits. Monsanto's data from sources such as labs, the field, and literature were all stored separately, and their researchers were working in special purpose analytical systems that made it difficult to share their results and combine information. The biotech company deployed Cloudera Enterprise with Cloudera Search and Cloudera Impala to knock down data silos and help researchers share their data, automate many data-driven decisions in the R&D pipeline, and answer questions like what traits to integrate into a germ plasm. Bringing data such as indexed images of plants at various lifecycle stages together with other data sources, optimized the production processes, ultimately helping them reduce the time to market of new products.



Monetize Data Assets

Many companies now realize they possess a vast untapped asset: their data. While needing to balance with privacy regulations and concerns, the opportunity to monetize corporate databases - customer profiles, buying patterns, location data, and more - is tremendous.

An enterprise data hub combines unlimited storage and real-time data access with the security and governance necessary to build an infrastructure for an ethical and compliant data monetization strategy. By blending internal and external data and selectively exposing it through Hadoop's various interactive and online interfaces, it is now possible to deliver data as a product to grow revenues while managing the risk of doing so.

Tubular Labs uses a Cloudera enterprise data hub to track and synthesize billions of YouTube video views, comments, and related "events" as well as social engagements including tweets, comments, and likes. This data becomes Tubular's business as it helps content marketers and content publishers leverage the data to gain competitive intelligence—and deeper insights—that will help them identify, grow, and engage passionate communities of followers on YouTube. Tubular uses APIs to get injections of billions of social actions from YouTube, as well as from Twitter, web crawls, and dozens of other publicly available data sources. Tubular customers, in turn, are growing their YouTube subscriber base 56% faster than non- Tubular users, and are growing their views 52% faster.



Business Opportunity: Manage Security, Risk, and Compliance

Increased regulatory pressures and well-publicized corporate data breaches have forced organizations to become more risk averse. With the volume of transactional, interactional, and observational data available today, it's increasingly difficult to detect complex and coordinated fraud or security attacks. Illicit activities cost organizations millions in revenue, not to mention lost customer trust, which enterprises simply cannot and will not tolerate.

As oversight increases, organizations need to ensure they are able to comply with modern regulations that mandate data retention and business transparency in a timely manner. There is less tolerance for non-compliance, driving the need to maintain strict data retention and governance practices to ensure that information can be traced throughout its lifecycle and across the organization.

Secure the Network

Cybersecurity has never been a higher priority for across every industry and in the public sector. With the average cost per breach reaching \$12.7 million in 2014, organizations are turning to new technologies in order to avoid massive reputational and monetary losses. Unfortunately, achieving network awareness and protection from advanced persistent threat (APT) attacks has never been more challenging. The data growth from the proliferation of network connected devices and the modern technology available to hackers and insider agents alike create a complex environment that traditional security, information, and event management (SIEM) tools are ill-equipped to handle.

By enabling security professionals to cost-effectively and scalably capture a deep and broad history of network activity, analyze it interactively using the tools they rely on, and operationalize their threat models to respond in real-time, modern organizations have a chance to remain ahead of attackers and protect the business from an expensive breach.

CounterTack's security solution analyzes system-level information collected from devices throughout an enterprise to identify patterns or anomalies that might be associated with malicious behavior. The data this requires quickly outgrew the restrictions of their original MySQL database which allowed coverage of 250 endpoints (and thus customers with 100 employees). With Cloudera Enterprise, they now support deployments over 100,000 endpoints. In just a few minutes, CounterTack can detect, analyze and automatically respond to an embedded piece of malicious code, or purpose-built malware, compared to the industry average of almost 300 days.



Assess Risk and Detect Fraud

Every sector of the global financial industry faces tremendous risk and regulatory need on a daily basis. From commercial banking and credit card processing to insurance companies and stock brokerages, a strategy that enables all business decision-makers to become information-driven can save tens of millions of dollars every year. Telecommunications, healthcare, retail, and other sectors face similar challenges.

When data is freed from silos, secured, and made available to analysts that answer key questions about the markets, customers, or patients — as they need it, in its original form, and accessed via familiar tools — everyone in the C-suite can rest assured that they have a complete view of the business, perhaps for the first time.

UK insurer, **Markerstudy**, implemented a Cloudera-powered enterprise data hub (EDH) to support both its policy underwriting and pricing as well as its claims and fraud management. It uses real-time and existing customer data to provide the most appropriate price and product to every customer. By storing, governing, and analyzing diverse data sets in its enterprise data hub, Markerstudy has identified an approximately £5 million reduction in claim costs through better fraud detection and prevention at point-of-quote.



Retain Data for Compliance

You need to retain data to meet compliance regulations, but you're already at capacity with existing data. Should you archive it to physical media, risking data loss and losing the ability to query that data when required? Or pay for an expensive expansion of existing systems? How would you know what's worth keeping online?

The answer is simple: stop archiving data offline – keep it all accessible online. Offload unlimited raw, historical, or low value data to an enterprise data hub for similar cost to storage or physical media, while retaining query access through existing BI and analytic tools. Avoid expanding existing systems, while freeing them to focus on high-value data. Because it is secure, you control who sees what; because it delivers governance and lineage services, you can trace access to, and the evolution of, your data over time. Finally, a cost-efficient way to address compliance requirements and deliver data on demand at the same time.

MasterCard brought its own enterprise data hub to PCI compliance and is now taking advantage of other PCI-compliant data sets by bringing them together for greater analytics capabilities. Further, MasterCard has entered into a Systems Integrator partnership with Cloudera where they're offering to help other financial services institutions achieve PCI certification.



Summary

Without the rich data that drives opportunity in the modern business context, decision-makers across every industry will continue to struggle with information paralysis and its high cost to the enterprise. Increasingly, CFOs will battle fraud and theft; CIOs will contend with regulatory non-compliance and overspending; CMOs will deal with a woefully incomplete view of the customer; R&D will be limited by sample myopia; and CEOs will fail to compete with more informed challengers as markets transform.

It's imperative that enterprises drive meaningful change in their infrastructure. Traditional data storage, management, and staging solutions have kept data limited, fulfilling specific objectives, usually for a few restricted user groups. Specialized systems continue to support the business but a new approach is required to support the new realities of customer service, product development, and corporate risk management.

With Cloudera, customers can now apply advanced analytics to unlimited data, turning that data into a strategic asset. By deploying an enterprise data hub, multiple users and applications can simultaneously access necessary information in real-time with full fidelity and governance based on role and profile. No other data platform offers such a powerful combination of flexibility and security to enable the sophisticated recommendation systems, security information and event management (SIEM), graph analytics, and machine learning capabilities that monetize data without the costs typically associated with traditional approaches.

Cloudera delivers the additional performance and support that enterprises require, supercharging open-source Hadoop with multiple processing engines. Complemented by state-of-the-art operational, governance, and security capabilities and certified on the industry's broad ecosystem of partner tools, Cloudera's enterprise data hub is aligned to the modern realities that drive business growth, operational efficiency, and innovation. It is the platform that powers the data-driven enterprise.

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, access, process, secure, 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. Cloudera's open source Big Data platform is the most widely adopted in the world, and Cloudera is the most prolific contributor to the open source Hadoop ecosystem. As the leading educator of Hadoop professionals, Cloudera has trained over 30,000 individuals worldwide. Over 1,450 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.

cloudera.com

1-888-789-1488 or 1-650-362-0488

Cloudera, Inc. 1001 Page Mill Road, Palo Alto, CA 94304, USA

© 2015 Cloudera, Inc. All rights reserved. Cloudera and the Cloudera logo are trademarks or registered trademarks of Cloudera Inc. in the USA and other countries. All other trademarks are the property of their respective companies. Information is subject to change without notice.