

## Enterprise Data Hub: The Key to the Information-Driven Enterprise

Big data investments in 2013 continue to rise, with 64% of organizations investing or planning to invest in big data technology compared with 58% last year.

### Overview

In order to win and retain customers, control costs, and remain ahead of the competition, today's leaders realize their organizations must become information-driven. This means developing a culture where individuals are empowered to make decisions with facts, rather than relying on opinion. Across virtually every industry, organizations are on the road to putting data at the center of business transformation, with goals such as better understanding their customer's behaviors, detecting fraud, or improving patient health outcomes.

Over the years, organizations have built solutions aimed at specific business problems: relational databases for transaction processing, data warehouse systems for analysis and exploration, document management systems for storing and searching business documents, and application solutions such as ERP and CRM to run major functional areas in the company. At the same time, in volume, in variety and in velocity, there is more data streaming in than ever before. As data grows, it becomes increasingly burdensome to move around for each new business question or analysis. Compliance regulations and risk management grow more complex and expensive. Worst of all, no single system contains the complete collection of enterprise data that organizations need in order to drive their business forward.

Many organizations are attacking these challenges head-on by building out a new capability in their data centers. They deploy a new platform, an enterprise data hub, that puts data at the center of the enterprise to give them the power and flexibility to be information-driven at superior price relative to traditional data management offerings.

An enterprise data hub (EDH) is one place to store all data, for as long as desired or required, in its original fidelity; integrated with existing infrastructure and tools; with the flexibility to run a variety of enterprise workloads—including batch processing, interactive SQL, enterprise search, and advanced analytics—together with the robust security, governance, data protection, and management that enterprises require. With an enterprise data hub, leading organizations are changing the way they think about data, transforming it from a cost to an asset.

### Big Opportunities. Big Struggles.

Big data presents a tremendous opportunity for enterprises across industries. By tapping into new volumes and varieties of data, organizations can ask questions about their customers and their business like never before. For example, organizations are using data to deliver a better customer experience, thus resulting in a more loyal customer base from which they can derive greater value. At the same time, with improved insight into business operations, it's possible to identify areas of inefficiency that, if addressed, can help reduce operating costs.

Yet even with so much at stake, most organizations are only just getting started. Gartner research shows that big data investments in 2013 continue to rise with 64% of organizations investing or planning to invest in big data technology compared with 58% last year\*. To make sense of this discrepancy, we must first understand what stands in the way.

First, the nature of data has changed. In the past, data primarily resided in structured transactional databases where schemas were fixed and volumes were limited to the gigabyte and, sometimes, terabyte range. Today's landscape is more complicated given the changes in the volume, variety and velocity of data.

“Deploying Cloudera allows us to process orders of magnitude more information through our systems, and that technological capability in combination with Experian’s expertise in bringing together data assets is driving new, real insights into tomorrow’s marketing environments.”

Jeff Hassemer,  
VP of Product Strategy, Experian

Second, organizations currently employ a variety of systems to support their diverse data driven goals, and so lack a unified view of information. They have enterprise data warehouses for operational reporting; storage systems to keep data available and safe; specialized massively-parallel databases for large-scale analytics; data archiving systems for cost-effective backup; and enterprise search systems for finding and exploring information with the ease of a web search engine. They are silos, each unable to share data easily among themselves for analyses that span data types or lines of business. These silos continue to proliferate as new data sources arrive and each business unit develops its own requirements fulfilled by its own budget. Without a single view, business visibility and insight are severely constrained, making it difficult to draw complete or consistent conclusions, and requiring expensive additional technology investments for each new data source or workload.

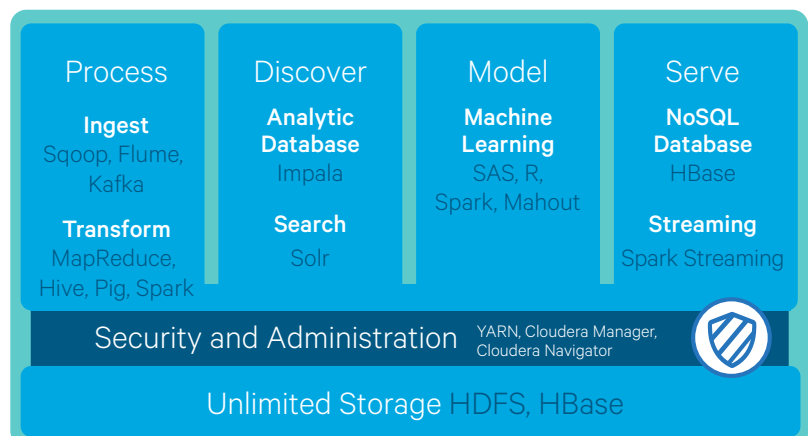
Third, moving large amounts of data is hard. Once IT is working with terabytes and petabytes of data, it’s often prohibitively expensive - or at best extremely inefficient - to transform and relocate data to specialized systems simply to answer new business questions. And, by definition, every transformation alters the original data such that information can be irretrievably lost to analysis. Going back to IT to change data models, to incorporate new data sources, or even to add new columns can take weeks or months. How can we be sure that what we leave behind doesn’t contain the next critical insight?

Fourth, despite these limitations, existing investments in people, business process, and technology remain valuable. Developers and administrators are certified on relational databases; business and IT understand how to use the existing BI and data warehouse platforms with proven workflows and governance processes; all of which leads to organizational inertia to continue investing in expanding these existing systems even as they show their limitations.

What’s needed is not a replacement for these systems but a new capability in the data center that extends and complements these investments with the scalability, flexibility, and price/performance point necessary to conquer the new world of data.

### The Enterprise Data Hub: The Key to Success in Big Data

The objective is simple: acquire and combine any amount or type of data in its original fidelity, in one place, for as long as is necessary, and deliver insights to all kinds of users, as fast as possible. And do so with maximum efficiency of capital and resources. Enter the Cloudera Enterprise Data Hub Edition.



Cloudera Enterprise Data Hub Edition has key advantages over existing systems:

- **Active Archive:** One place to store all your data, in any format, at any volume, for as long as you like, allowing you to address compliance requirements and deliver data on demand to satisfy internal and external regulatory demands. 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.
- **Transformation and Processing:** ETL workloads that previously had to run on expensive systems can migrate to the enterprise data hub, where they run at very low cost, in parallel, much faster than before. Optimizing the placement of these workloads and the data on which they operate frees capacity on high-end analytic and data warehouse systems, making them more valuable by allowing them to concentrate on the business-critical OLAP and other applications that they run.
- **Self-Service Exploratory BI:** Users frequently want access to enterprise data for reporting, exploration, and analysis. Production enterprise data warehouse systems must often be protected from casual use so they can run the mission-critical financial and operational workloads they support. An enterprise data hub allows users to explore data, with full security, using traditional interactive business intelligence tools via SQL and keyword search.
- **Advanced Analytics:** Multiple computing frameworks that enable analytics, search, machine learning, and more unlock value in new and old data sources. Rather than examining samples of data, or snapshots from short time periods, all historical data, in full fidelity, can be combined in comprehensive analyses. Simple tabular data can mix with more complex and multi-structured data in ways that were never before possible.

Of course, an enterprise data hub must also provide key capabilities to allow operation of the system with confidence, including:

<b>Secure and Compliant</b>	<ul style="list-style-type: none"> <li>• Robust access controls</li> <li>• Built-in data encryption and enterprise key management</li> <li>• Shared security policies</li> </ul>
<b>Enterprise Data Governance</b>	<ul style="list-style-type: none"> <li>• Metadata management</li> <li>• Data lineage and tethering</li> <li>• Audit logging and reporting</li> </ul>
<b>Unified and Manageable</b>	<ul style="list-style-type: none"> <li>• A single shared storage system for all data, across all analytic and processing tools</li> <li>• On-premise, cloud and managed service deployment</li> <li>• Highly available with support for backup and disaster recovery</li> </ul>
<b>Open Architecture</b>	<ul style="list-style-type: none"> <li>• Open source platform guards against vendor lock-in</li> <li>• Native APIs and engines for multiple workloads</li> <li>• Extensible for third parties</li> </ul>

## Powered by Cloudera

Built on the transformative Apache Hadoop open source software project, Cloudera Enterprise is designed for the demanding requirements of enterprise customers. Cloudera is the leading contributor to the Hadoop ecosystem, and has created a rich suite of complementary open source projects that are included in Cloudera Enterprise.

Hadoop has evolved into a stable, scalable, flexible core for next-generation data management. Yet, alone it lacks several critical capabilities necessary for deploying it as the center of an enterprise data hub. It lacks a comprehensive security model across the entire ecosystem of projects. It was built for batch-mode data processing workloads, which limits Hadoop to an ancillary position in the data center – a central enterprise data hub must be real-time. And Hadoop doesn't support the range of industry-standard interfaces for query and search applications, among others, that business users require. Cloudera has addressed all of these challenges and more with its solution.

Cloudera offers a single platform from which organizations tackle diverse critical business problems:

- Automatically archiving the complete set of enterprise data to meet compliance requirements with immediate online access;
- Complementing existing enterprise data warehouses to offload data and workloads to improve performance while managing costs, and enabling the delivery of high value data sets for operational reporting;
- Supporting self-service business intelligence, through familiar tools, on more data and more kinds of data than ever before possible;
- Enabling and consolidating enterprise search on data and documents in-place within the single environment; and
- Accelerating advanced analytics solutions, such as recommendation engines, fraud detection, or image processing.

## Summary

Unified data management has long been a shared business and IT objective, unattainable until now due to the limitations and costs of traditional approaches. For decades, data warehouses and marts based on relational database technology were the only option for deploying enterprise analytics, and storage arrays or archives the only options for keeping diverse data alive long term.

With Cloudera Enterprise, customers can now easily handle the rapidly increasing data volume and variety they face, addressing a growing share of data and workloads from legacy infrastructure while optimizing the efficiency of those existing systems. Powered by Apache Hadoop at the core, Cloudera Enterprise extends your existing investments with a central scalable, flexible, secure environment for gaining insights from all data, without limits.

An enterprise data hub is a powerful new platform. In the future, most enterprise data will land first in an enterprise data hub, and increasingly it will stay there. In the near term, an enterprise data hub delivers unprecedented flexibility to comprehensively and economically analyze and process data in new ways. Organizations that deploy an enterprise data hub alongside their existing infrastructure will continue to lead in the world of modern data.

## About Cloudera

Cloudera delivers the modern platform for machine learning and advanced analytics built on the latest open source technologies. The world's leading organizations trust Cloudera to help solve their most challenging business problems by efficiently capturing, storing, processing and analyzing vast amounts of data.

Learn more at [cloudera.com](https://cloudera.com)