

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

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 collect, organize and analyze data, gain actionable information, and meet mission, today's leaders in government realize that their agencies 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 sector of government, agencies are on the road to putting data at the center of decision-making, with goals such as improving disaster response, increasing efficiency in the supply chain of our armed forces, or optimizing patient health outcomes.

Over the years, federal agencies 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 documents, and application solutions such as ERP to run major function areas in government. At the same time, in volume, in variety and in velocity, there is more data streaming in than ever before. As the data grows, it becomes increasingly burdensome to move around for each new question of analysis. Compliance regulation and risk management grow more complex and expensive. Worst of all, no single system contains the complete collection of enterprise and mission data that organizations need in order to drive government forward.

Across government, agencies are tackling 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 agency to give them the power and flexibility to be information-driven at a superior return-on-investment 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 workloads—including batch processing, interactive SQL, enterprise search, and advanced analytics—together with the robust security, governance, data protection, and management that government requires. With an enterprise data hub, government is rethinking data, transforming it from a cost center to a crucial mission asset.

### Big Opportunities. Big Struggles.

Big data presents a tremendous opportunity for federal agencies. By tapping into new volumes and varieties of data, agencies can ask questions about their businesses, their missions, and those they serve that were previously inaccessible. For example, agencies are using data to improve their outward-facing services, resulting in improved services and communication between the public sector and its citizens. At the same time, with improved insight into business opportunities, it's possible for agencies to identify areas of inefficiency that, if addressed, can help reduce operating costs— especially important given today's budget cuts facing federal government.

Yet even with so much at stake, most agencies are only just getting started. According to recent research, 70 percent of all agency leaders believe that in the next five years, big data will play a critical role in fulfilling mission objectives. To make sense of this discrepancy, we must first understand what stands in the way.

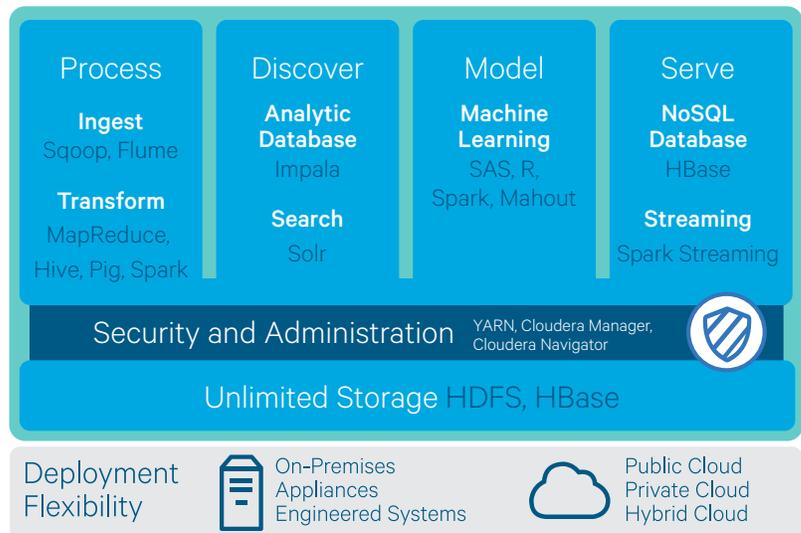
First, the nature of data has changed. In the past, data primarily resided in structured, transactional databases, where schemas were fixed and volumes were limited to the gigabyte and, sometimes, terabyte range. Today's landscape is more complicated given the changes in the volume, variety and velocity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

Second, agencies 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 reliable 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 often span data types, analyst workflows, or even agency departments. These silos continue to proliferate as new data sources arrive and each agency division develops its own requirements fulfilled by its own budget. Without a single view, viability 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 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 agencies be sure that what they 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 this leads to organizational inertia to continue investing in expanding these exiting systems even as they show their limitations.

What government needs 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. Investment in new technology is critical to keeping government at the forefront, ensuring that agencies make mission in an increasingly digital age.



## 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 necessary, and deliver insights to all kinds of users, as fast as possible. And to do so with maximum efficiency of budget 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 requirement 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 service, you can trace access to, and the evolution of, your data over time. This ensures compliance and optimal security practices that help keep sensitive government data safe.
- **Transformation and Processing:** ETL workloads that previously had to run on expensive systems can migrate to the enterprise data hub, where they run at a 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 mission-critical applications.
- **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 workloads they support. An enterprise data hub allows federal users to explore data, with full security, using both traditional interactive business intelligence tools like SQL and keyword search as well as more advanced computing approaches.
- **Advanced analytics:** Multiple computing frameworks that enable capabilities such as analytics, search, and machine learning 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, expanding analytics possibilities.

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

Designed for the demanding requirements of public sector customers, Cloudera is revolutionizing data management with the first unified platform for big data, an enterprise data hub built on Apache Hadoop. Cloudera is the leading contributor to the Hadoop ecosystem, and has created a rich suite of complementary open source projects.

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 since 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 federal 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 security intelligence, fraud detection, or image processing.

## Summary

Unified data management has long been a shared business and IT objective for government, 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 analytics, and storage arrays or archives the only option for keeping diverse data long term.

With Cloudera Enterprise, agencies 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 these existing systems. Powered by Apache Hadoop at the core, agencies using Cloudera Enterprise can extend their 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 that can help agencies manage the data deluge, ultimately gaining better, more comprehensive and actionable information. In the future, most enterprise and mission 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. Federal agencies that deploy an enterprise data hub alongside their existing infrastructure will continue to lead in the world of modern data. That's the future of government.

## About Cloudera

Cloudera is revolutionizing data management with the first unified platform for big data, an enterprise data hub built on Apache Hadoop. Agencies today must be information-driven while managing risk and costs. Cloudera offers agencies a secure and cost-efficient place to store and analyze all their data, empowering them to derive new insights and correlation while extending the value of existing investments.

With Cloudera at the center of an agency's enterprise data hub (EDH), analysis and business users gain unprecedented visibility to contemporary and archival data and releases data sequestered in standalone applications. An EDH offers a wide range of computing capabilities, including interactive SQL, search, and machine learning, and offers mission operators the right mix of flexible analysis and data depth and breadth to find the answers without sacrificing oversight and security to protect and govern the data and its use.

Cloudera helps agencies make the most of their data, their infrastructure, and their most valuable resource -- their people. Cloudera was the first and still is the leading provider and supporter of Hadoop for the public sector and offers software for mission critical data challenges including storage, cloud, security, management, and analysis. Cloudera works with over 1,400 hardware, software, and services partners to meet agencies' data goals.