Capgemini's Data Optimization for the Enterprise with Cloudera

An advanced service to rationalize and optimize your data with enterprise data management powered by Apache Hadoop™

In this age of big data, new technologies and methods have evolved and matured to the point where organizations are able to deal with the increasing volume, variety, and velocity of data at a fraction of the cost of traditional Data Warehouse technologies.

In order to derive meaningful and actionable insights, business leaders are now increasingly demanding integration of external sources (such as third party data, open data, social media, and the Internet of Things) with traditional data sources (such as CRM and ERP systems).

Big data technologies are allowing organizations to harness this unstructured and semi-structured data, much of which was previously inaccessible owing to its sheer volume and to the structured bias of traditional Enterprise Data Warehouse (EDW) systems.

Realizing Value from Big Data

Many industries today are highly regulated and required to archive large volumes of historical data, or need to have near real-time access to all relevant data to provide a sustainable competitive advantage. In either case, the ability to **obtain greater value from all your data assets can be the difference between success and failure**. Gartner has stated

that the three Vs of big data are volume, velocity, and variety. However, **the V that matters most to any organization is value** – otherwise you are collecting data for its own sake.

Organizations must reconsider how they think about value. They must use data to answer questions like:

- What is happening now?
- Why is it happening?
- What are the potential causal factors?
- What will happen next?
- What decision(s) can I make right now to improve results?
- How can I rapidly inform the business and push the results to the PC, tablet, and mobile platforms?

Although they now have access to more mature technology and processes for measuring and creating efficiencies and improving decision-making capabilities, organizations have encountered challenges in adopting, and evolving to, the next generation of data management architectures. Before a business can gain deeper insight and competitive edge through information, it must implement tools and techniques for collection and storage of historic data on the one hand, and innovative methods and technologies for deriving value from the data on the other.

Solution Overview

Data Optimization for the Enterprise is a service offering that helps organizations achieve significant additional value from their data in return for relatively low capital investment. Clients can take full advantage of big data technologies to optimize the ratio between the value of data and its storage costs, while also gaining extended capabilities to handle complex data and providing their users with a richer analytical experience.

Data Optimization for the Enterprise Powered by Cloudera

Capgemini has collaborated extensively with Cloudera to build an integrated Data Optimization solution leveraging the Cloudera platform. Cloudera has worked with Fortune 500 companies to help them **enhance their capabilities for complex data handling and advanced analytics**.

Capgemini and Cloudera now bring their big data technology and advanced analytic capabilities together. Cloudera's expertise in optimizing existing Data Warehouse architectures using Hadoop complements Capgemini's broader view of these analytical landscapes, from the variety of source systems (internal or external) to the interfaces (business applications) that expose the data in the form of practical insights for your business users.

Strategic Value Assessment

Capgemini Data Optimization uses a specially-designed Strategic Value Assessment (SVA) methodology. By reviewing your current technology and usage landscape through the lens of articulated business use cases and key business challenges, we rapidly determine the potential for data optimization.

An SVA enables your organization to understand how enterprise data collection, storage, and information dissemination can be optimized to address key information gaps and cost-effectively align the information with analytical needs. You will see how you can **manage all data as a strategic asset**, improve storage utilization, take advantage of platform modernization, and **raise the value of the insights** you extract from all of this. The ultimate goal is to identify opportunities to take out cost while improving the business value of analytics.

Reviewing Your End-to-End Data Architecture

Big data initiatives have a direct impact on storage costs and infrastructure. Capgemini's approach exploits innovative big data solutions that are designed to run on industry-standard hardware. Cloudera Enterprise includes capabilities for interactive SQL querying (with Cloudera Impala) and full-text, interactive search (Cloudera Search) to support advanced analytics on a wide variety of data. This platform complements existing technology investments and data assets and will, in many cases, help extend the value, reach, and capability of these existing investments.

Figure 1: Analytical architecture complemented by Cloudera
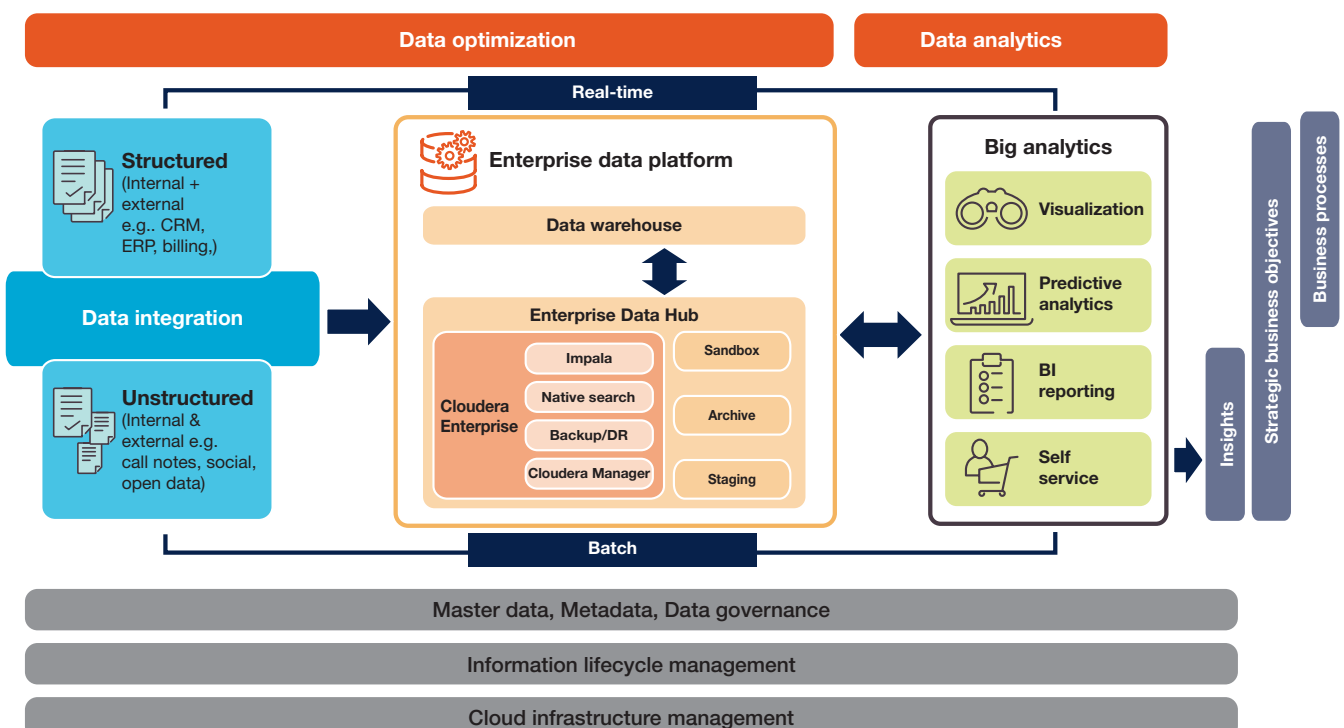
Drill down into Your Existing Data Warehouse

One important piece of your analytical landscape will certainly be an existing Data Warehouse, and the logical, physical, or hybrid Data Marts built to answer the needs of your business users. Capgemini leverages Cloudera's experience in complementing such architectures with Hadoop to help you determine the data sets and/or the workloads that you could optimize by moving them to Hadoop. To do this, we:

- Identify migration candidates by **scanning the less optimized workloads, analyze your data usage** to determine how the data sets are accessed, and classify these current usages by workload type (tactical or strategic), complexity (volume and algorithms), and frequency of access.
- **Define the target reference architecture that will best fit your context** – taking into account workload future positioning, and impact on infrastructure capacity planning.

The value of such optimization approaches, integrated with our broader Data Optimization SVA, offer multiple opportunities:

- As Cloudera Enterprise provides scalable storage, offloading some of your data sets will allow you to realize **significant savings on upcoming infrastructure costs**, and allocate your more valuable resources to deeper and more complex analytics.
- Cloudera Enterprise allows you to keep all your data, rather than archiving or taking data offline to reduce data volume. **Analytics can be performed on all the data, rather than sampled or incomplete data**, increasing the value of the insights gained.
- Thanks to the capabilities of Cloudera and a new approach to data modeling, you will no longer be forced to define the way you want to look at your data when you integrate it into your Data Warehouse ("schema on write"). Instead, you can leverage Hadoop's **"schema on read"** capability. This allows you to be much **more agile** when a new product or feature is released, or when you want to change the angle of your analysis. You drastically increase your agility and reduce time-to-market as a result.
- With Cloudera as the enterprise data hub plus the ability to **query data with Cloudera Impala** and perform full-text **search with Cloudera Search**, your business users dramatically improve their ability to test out analytical hypotheses. This can now be done across the total amount of historical data, and across both structured and unstructured data. It becomes possible to work on large-scale exploratory tasks such as private or department-specific studies, ad-hoc and self-service analytics, and benchmarks of new KPIs or analytical models. All this increases the adaptive power to be predictive, precise and responsive to market changes.
- Ultimately, your business teams no longer need to rely only on IT processes and resources when they need to obtain data and insights for just-in-time business analytics.

Capgemini and Cloudera can help you exploit all these optimization opportunities. Our Data Optimization approach creates a **transformation roadmap and an investment case** that allows you to leverage Cloudera to reallocate time and resources to activities that add real value, such as advanced analytics, instead of on tactical activities, such as data integration and data crunching.

As a result of our Data Optimization approach, you will gain an end-to-end view of the positive impact of these new technologies on your entire data value chain, for immediate business value, efficiencies, and cost optimization.

Many companies are realizing immediate benefits by combining their existing Data Warehouse architecture with Cloudera Enterprise. They have been able to optimize their IT spending profiles, maximize ETL/ELT load performance, reduce data latency, increase agility, and improve business-critical workload SLAs. They have obtained a richer set of analytics, building users' confidence in the data they are accessing.

Outputs from the SVA

On completion of the assessment, clients will have a clear understanding of any gaps in their data and analytical capabilities, recommendations as to the most valuable improvements they could make, and a **roadmap that takes into consideration their unique business and technology requirements**. Other recommendations cover:

- Strategic SWOT alternatives on available options
- Practical solution scope and requirements
- Information future state architecture and roadmap
- Data asset architecture and roadmap
- Global standards and process recommendations
- Resource and project plans
- Business case for data optimization and investment
- Sizing, configuration, scope, and licensing suggestions (as applicable)

Figure 2: Data optimization assessment and harnessing value

Data discovery workshop	Define business value and craft scenarios	Create tangible results	Tune the initial scenarios for productive use
Where you are & where you could be	**See through the lens of the Business**	**Made real in a test-drive environment**	**Delivered to realize business value**
We underpin the company's vision and strategy with the potential of Capgemini's data optimization solution	The idea created during the discovery will be matched against business requirements, elaborated upon and assessed for feasibility	Vision comes down to Earth: identified scenarios can be brought to life at feasibility level, or with a PoC in our CUBE laboratory	Finally the scenario will be implemented and made available for the broad user group within the company

The Big Data Journey

Your journey starts with a Data Discovery Workshop where Capgemini helps the business and IT community agree on how big data and analytics can help achieve the company's vision and priorities. We then drill down into key objectives and discuss their feasibility.

You can then "test-drive" your ideas in a sandbox or with a Proof of Concept (PoC) to help demonstrate tangible value in a limited environment. PoCs can be constructed in Capgemini's CUBE lab environment, where big data and analytic technologies are available for clients to try out.

> CUBE is our Customer BIM Experience showcase, where you can experience innovative BIM solutions. You can interact with subject matter experts, examine solutions created to address specific client issues, and review PoCs, technology innovations, and productivity tools. Everything in CUBE focuses on helping clients realize the value of information management and become Intelligent Enterprises.

Once we have identified the value you want to achieve and agreed on the path forward, Capgemini can help you deliver that value through a mature, managed approach to implementation. We will partner with you to achieve the right project prioritization, funding, and Rightshore® resourcing to deliver identified initiatives.

Capgemini Guides You from Data to Insight

Data Optimization for the Enterprise is the fast, cost-effective way to start getting full value from data. Through the process outlined above, we will help you design and implement a robust reporting and analytics infrastructure that will transform big data into intelligence and intelligence into action.

We have already helped many clients evolve their legacy information architectures to exploit massive data volumes and new data types, taking advantage of the rapid evolution of technology. **Contact us today to find out how we can do the same for you**.

About Capgemini

With more than 130,000 people in over 40 countries, Capgemini is one of the world's foremost providers of consulting, technology and outsourcing services. The Group reported 2013 global revenues of EUR 10.1 billion.

Together with its clients, Capgemini creates and delivers business and technology solutions that fit their needs and drive the results they want.

A deeply multicultural organization, Capgemini has developed its own way of working, the Collaborative Business Experience™, and draws on Rightshore®, its worldwide delivery model.

For further information visit
www.capgemini.com/bim

or contact us at
bim@capgemini.com