Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

cloudera®

Accelerate your Big Data Strategy

Execute faster with Capgemini and Cloudera's Enterprise Data Hub Accelerator

Enterprise Data Hub Accelerator enables you to get started rapidly and cost-effectively with an action-focused plan for your first Big Data project

The Big Data debate has inspired organizations to look at their information in new ways - and to see the potential that lies dormant in their existing databases and data warehouses. The big question, however, is how to harness this potential in ways that will deliver genuine value and tangible business benefits.

This leads to a vital and logical question: **how to begin?**

Organizations looking to get started with Big Data need quick wins. They need to start small, experiment and find ways to achieve tangible business benefits. Experimentation and innovation are key but they also need to start out in the right direction, with flexible, cost-effective, and unified solutions that will scale without locking in precious data.

Capgemini and Cloudera have collaborated to build a comprehensive approach to help organizations accelerate and execute on Big Data initiatives. The Enterprise Data Hub Accelerator is an execution framework for Big Data, built around Cloudera's Apache Hadoop-based open-source enterprise data management platform. It helps you to define your first projects, make sure you execute them well, and show you how to grow these to a fully-defined and sustainable Big Data strategy for your organization.

In this white paper, we illustrate how the Enterprise Data Hub Accelerator enables you to get started rapidly and cost-effectively with an action-focused plan for your first Big Data project. We show you how to identify the right business drivers for Big Data, progressively transform your governance processes around data, get your analytics to the next level, and ensure scalability of the platform for all of your data. Addressing all dimensions of your Big Data strategy will ensure that you achieve your goals and make a significant and transformative impact on your business with Big Data.

Data drives the business world

The world we live in has become data driven. Data pours out of wearable electronics, connected devices, and RFID chips. It flows across social media, changing the way we live and the way we do business.

Virtually all of our daily activities leave data "footprints." Even inanimate objects are getting in on the act. Machines, consumer goods, devices, clothes, and raw materials are all becoming connected and integrated to make our lives more comfortable - a development commonly referred to as the "Internet of Things".

Leading organizations are now changing the way they think about data, transforming it from a cost to an asset. They are leveraging new Big Data technologies to get more insight into all of their data:

- Financial services companies improve analytics across virtually all areas of operations - from risk assessment in insurance to fighting fraud, keeping mobile banking operations online, and delivering improved customer insight to marketing teams.
- Telecommunication companies analyze customer behavior, experience, and usage patterns to develop new products and services, optimize networks, and grow market share.
- Energy management companies provide smart metering analytics to customers to help them become "greener".
- Oil & gas companies use sensors in pipelines to deliver data to predictive asset maintenance solutions. By aggregating the sensor data, they can identify potential failures, avoid harming the environment, and save billions on repairs and legal actions.
- Life sciences organizations reduce the time to market of new, effective, and safe drugs by combining data from clinical trials, drug efficacy studies, and markets.
- Travel agencies send discounts on hotel rooms tailored to customers' online activities to increase sales and market share.
- Fraud detection analytics look at open data, social media, or machine data to detect unusual behavior patterns and react in real-time.

The Big Data revolution is underway - don't get left behind

Business models are challenged now more than ever. On the one hand, organizations need to be more relevant to customers and aware of their needs and expectations; on the other hand, they need to be increasingly more efficient, both financially and operationally. All the while, the market competition is becoming fierce.

Organizations are looking at ways to transform themselves, to become customer-centric and services-oriented. Digital transformation, customer experience management, and operational efficiency are key topics that cannot be appropriately tackled without the intelligent use of Big Data.

Knowing how a business or process performed yesterday is not enough anymore. Organizations need to know not only who their customers are and what they are buying, but also why they are buying, what they will buy in the near future and how their networks influence their buying behaviour.

Around information, context is key.

Forward-looking companies need to get in on the game, or they risk being left behind. In this world, data becomes a strategic asset, a way to gain an edge on the competition, making you more efficient and profitable.

But this is a huge challenge. In order to get meaningful and actionable insights, business leaders are demanding integration of all their data sources, including:

- External sources – third party data, competitive data, open data, social media
- New data sets – machine log data, sensor data
- Traditional data sources – online transactional data such as those from CRM and ERP systems, as well as data warehouses and data marts.

These data are currently not well integrated. They are stored in different formats and in different places. They are worlds apart.

Organizations are using Big Data solutions to finally get to a complete view of their business and their market. They use it to increase customer interaction, understand customers better, engage them on a more personal level, or increase operational efficiency and profitability.

Challenges of traditional BI landscapes and processes

Organizations that want to get on this road to transformation are facing challenges that are not new, but are becoming intolerable due to the data deluge:

1. **Data "blindness"** - existing data management systems are expensive to scale, forcing companies to leave data out or archive it offline. This leads to multiple silos of data and a lack of visibility to all data.
2. **Delays** - existing analytical architectures are too rigid and depend on time-consuming data transformations and modeling processes. This prolongs time-to-insight and also compromises the fidelity of the data.
3. **Cost** - existing BI and analytics systems operate at near full capacity in most organizations, making it impossible to accommodate new data without incremental infrastructure spend.
4. Complexity – the combination of multiple data sources, complex transformation processes, and siloed analytical systems make it nearly impossible to provide integrated views of the required data to the appropriate business users.

Figure 1: Challenges of traditional BI landscapes and processes

Big opportunities come with big questions

While Big Data makes big promises, it also raises significant questions. From talking to our clients around the world, we find these are most pressing:

- Where and how do we start?
- How do we create a business case for a pilot?
- What should we expect in terms of ROI?
- What data is relevant?
- How do we acquire that data, how often and how do we manage it?
- Do we start looking at all of it, now?
- Should we move legacy applications and workloads to the new platforms?
- How do we build or acquire a team to do this?
- How do we make the right choices upfront?

Together, Capgemini and Cloudera can address these questions for you and help you take action now.

Cloudera's enterprise data hub at the core of next generation data management platforms

The big question looming around Big Data is simple yet fundamental: how to extract new value from data and deliver tangible business benefits.

Cloudera's enterprise data hub is the core of the solution to this dilemma. Based on the open source technology of Apache Hadoop and delivered via Cloudera Enterprise, an enterprise data hub gives you one centralized platform for processing all your data and managing it in an agile, cost-efficient, and sustainable way.

Instead of moving data around in multiple or point-solutions, it provides a single store for all data, for as long as it is needed. The data is stored "as is" and transformed when necessary for a specific purpose. This means the raw data can be recaptured and reused.

An enterprise data hub is open and flexible, and can be integrated with existing infrastructure and tools (ETLs, reporting and advanced analytics) to run a variety of enterprise workloads — including batch processing, interactive SQL, enterprise search, stream processing, machine learning, and advanced analytics.

In addition to being open and flexible, it is also enterprise-ready with robust security, governance, data protection, and

management capabilities. Organizations such as financial institutions or government agencies have been running this technology for a number of years, with the appropriate security and data privacy measures.

Best of all, it gives you a way to explore your data - all of it - and deliver insights to all kinds of users in the manner that suits them best. It can do this efficiently and cost-effectively, delivering the business benefits promised by the potential of Big Data.

How to put Big Data to work for your organization

Capgemini and Cloudera have partnered to build an execution framework to help you start and deliver your first project and grow it into a fully defined Enterprise Big Data strategy. The framework defines how to:

- Identify business drivers as well as security and governance processes
- Incorporate and leverage machine learning and scientific algorithms
- Identify where to begin, which data sets to start with, how to manage the data and make it secure and governable
- Scale your Big Data Platform over time.

Our execution framework is built around:

1. Five pillars for Big Data execution: Business Drivers, Governance, Analytics, Data, and Platform
2. Four stages of maturity - **Initiate**, **Scale**, **Transform** and **Run**

In this first phase – "Initiate" – we focus on identifying accelerators that will help you deliver your first results quickly and build solid foundations for the future. The next stages will guide your organization on the best way to build on this success and grow to maturity.

Your core technologies are already certified on Cloudera's platform

Cloudera's open platform and rich partner ecosystem - more than 900 partners from the "existing BI" as well as the "new Big Data" worlds - give you maximum flexibility to integrate your existing and new systems, tools, and applications (ETL, analytic, and reporting tools). This openness ensures that your enterprise data hub investment is there for the long run, capable of integrating data from and to legacy technologies as well as emerging solutions.

These partners in this ecosystem include:

The Five Pillars of Information-Driven Business Transformation

Pillar 1 – Identify your **Business Drivers**

You need clarity on why you are launching your Big Data initiatives, the business challenges it will address, and the benefits you expect it to deliver.

Some clients have the aim of simply reducing data storage costs. While Big Data can certainly help to do that, there is potential for so much more. We recommend looking at business areas where greater insight will deliver rapid competitive advantage. Big Data is first and foremost a set of new solutions to help you solve business problems you have been struggling with for a long time.

Sometimes, the right business drivers can be determined and agreed upon by looking at a company's top five corporate challenges, to see how these new solutions could make a difference. Also, it is likely that any organization will have a number of analytical use cases or projects, which currently can't be addressed completely. Typical reasons include:

- Costs around adding data to existing data management systems are prohibitive, either because these systems have reached capacity or the volume of data required is not manageable with the current architecture
- ROI studies are inconclusive or the business need is not 100% mature
- New data is too complex or is coming in too fast, does not fit well in existing structures, and the cost of integration (ETL processes) is too high.

Even if you are not thinking of these use cases as being Big Data problems, the new solutions brought by Big Data could actually help.

As agility is key, business drivers do not have to be tightly defined up-front. The right governance processes will help you refine these along the way.

Pillar 2 - Enhance your data culture with the right **Governance**

Governance around Big Data is focused on agility, exploratory work, and new ways to engage business and IT teams to work together.

It is also about connecting the dots between Big Data initiatives and champions in an organization and creating a community of people and processes that will drive execution forward in business units and IT departments.

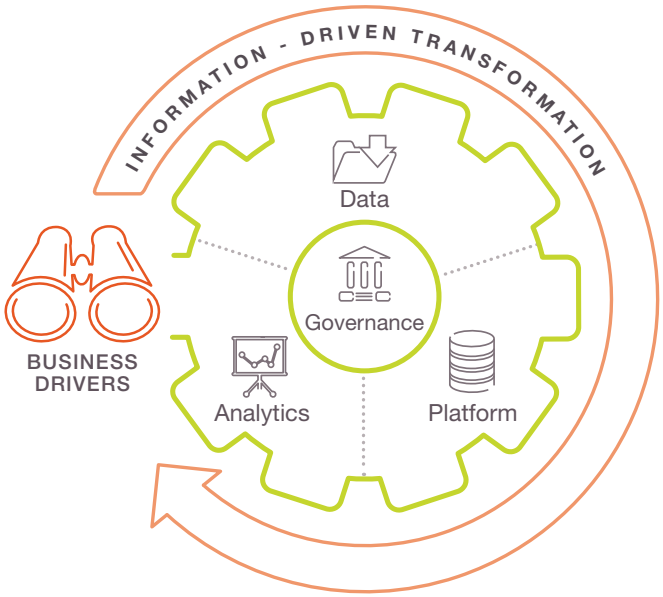
In this execution framework, Capgemini brings best practices used in enterprise-wide analytic environments to help clients define the governance that is right for their organization, and change the experience of their business users around data with agility.

There are two key aspects:

- Establishing the right processes around data management and analytics - how users interact with data, set up processes and management, organize workloads, and handle integration.
- Establishing the right processes for users and teams - how new business needs are integrated, how agile processes are introduced and managed, how collaboration can be improved between teams, and how business drivers evolve over time.

Creating a well-established but agile governance model has traditionally been a major challenge. This is where Big Data solutions like the Cloudera enterprise data hub can help by unleashing the possibilities around data management and analytics.

Figure 2 : The Five Pillars of Information-Driven Business Transformation



Pillar 3 - Do more with your data – try new **Analytics**

Big Data initiatives are about looking at new solutions and techniques to get more insight out of your data. Analytics is central to this process.

The maturity stages of analytics are well known:

- Descriptive analytics and reports depict the right picture of what happened last month, last week, or yesterday
- Predictive and prescriptive analytics are able to predict sales or inventory and recommend a "next best action" to your sales team.

Once again, the key to getting better results from your analytics processes is to keep things simple at the beginning.

Improve efficiency

You can increase the efficiency of your existing analytics by being able to store more data and see a longer history. With larger data sets and longer time spans to examine, your existing predictive algorithms will be able to reveal more accurate trends and enable more precise actions.

Get more agile

Presently, many organizations spend most of their time on data integration, data cleansing, and data transformation activities. With Cloudera Enterprise, instead of moving data every time you need to use a new tool or add new applications, the unified platform lets you analyze data in-place. You can focus on the pure analytics part of the job, as data provisioning is done only once.

Within your enterprise data hub, it is then possible to set up a specific space for experimentation and exploratory work, which will allow you to get your analytics processes to the next level, with new data sets and new types of algorithms.

Pillar 4 – The pieces of the puzzle: get more **data**, get it ready?

If data is the new oil, how do you fill the tank?

In the past, one of the big stumbling blocks on the road to Big Data has been the issue of how to reduce the cost and complexity of storage.

Cloudera Enterprise solves the challenge as it can store all your data in its original format quickly and economically, for as long as you need it. To get there, we will help you define the best approach to select your enterprise and external data, migrate it, and integrate it into your enterprise data hub. Once that is achieved, you will have made a significant step towards solving the issue of data silos. We will also help you set up the right processes for data quality, master data management, data privacy and secure access.

We recommend looking at business areas where greater insight will deliver rapid competitive advantage. Big Data is first and foremost a set of new solutions to help you solve business problems you have been struggling with for a long time

Pillar 5 - A data management **platform** at the service of your business, not the other way around

The aims of deploying an enterprise data hub are simple:

- Acquire and combine any amount or type of data in its original fidelity, in one place, for as long as you need
- Effectively deploy diverse analytic workloads on shared data to reduce complexity and increase business visibility across all your data
- Deliver insights to all kinds of users, as efficiently as possible.

An enterprise data hub solves the problem of data platforms that are too complex or slow, unwieldy, or simply too full to allow business users to launch new projects.

Cloudera is the leading contributor to the Hadoop ecosystem and has created a rich suite of complementary open source projects as part of an enterprise data hub, including security, data management, and search.

The open-source model means your data is never locked in and Cloudera remains fully committed to keeping it this way. It offers a number of key advantages over existing systems:

- Security over who can access which data, through advanced governance and lineage services.
- Transformation and processing ETL workloads run on your enterprise data hub at low cost, in parallel, much faster than before.
- Users can explore data, with full security, using business intelligence tools via SQL and broad-based search (keyword, term, or document) capabilities.
- Multiple computing frameworks enable analytics, search, machine learning and more. This allows you to unlock value in new and old data sources. Simple tabular data can mix with more complex and multi-structured data in ways that were not possible before.

Four Stages of Maturity on your Road to Big Data

Capgemini and Cloudera have come together to help clients act now on the promises of Big Data - helping them to start small, explore the potential, test the waters, and move forward with confidence that the solutions put in place now will scale for the future.

To help our clients, we have clearly defined four key stages on the road towards implementing Big Data solutions across the enterprise.

Initiate: Kick-start success

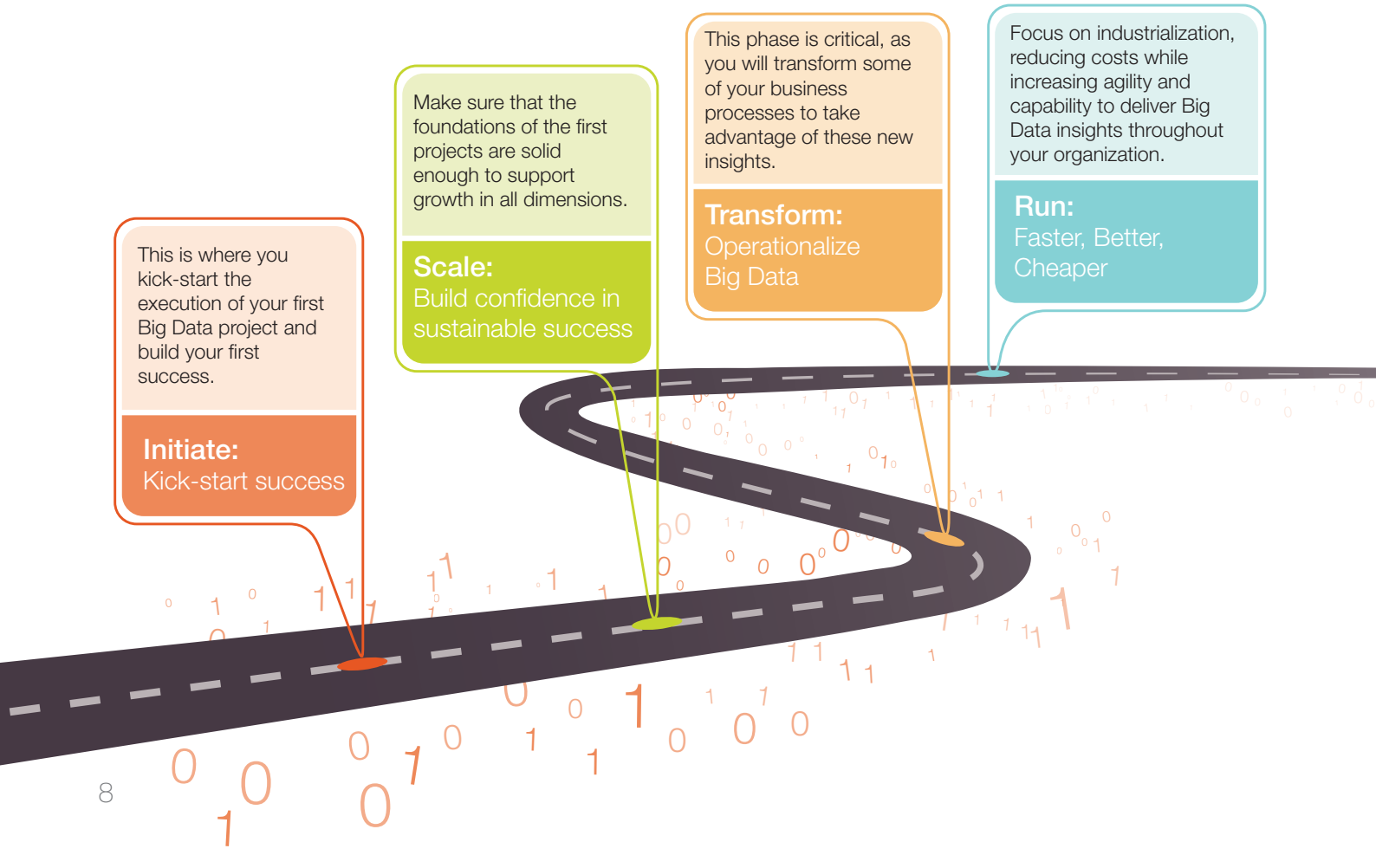
This is where you kick-start the execution of your first Big Data project and build your first success.

Capgemini has created accelerators to help organizations identify the right projects to address and ensure they deliver measurable benefits that help you move forward with confidence. We see three major patterns from our on-going projects

1. **Data optimization**. Use new Big Data architectures next to your existing BI landscape to help you improve storage, modernize your platform and raise the value of the insights you extract from all your data. The ultimate goal is to identify opportunities to optimize cost while improving the business value of analytics.
2. **Data Exploration Lab**. Set up a new environment and the right methodology for identifying new insights and exploring new opportunities from your data. It gives you a way to look at unexplored data sets, test new types of analytics on your data and move fast from one experimentation to another, while ensuring you capitalize on what you learn.
3. **Customer case studies and industry use cases** - Capgemini has developed a series of customer case studies and industry use cases to deliver insight into what other companies and organizations are already doing. These can inspire your early projects.

The key requirement is to create value quickly. To help you achieve this, Capgemini provides its **Elastic Analytic** solution

Figure 3 : Four Stages of Maturity on your Road to Big Data



built on Amazon Web Services. It gives you the flexibility to pay for infrastructure when you need it, not upfront, and adjust it to your business needs.

Scale: Build confidence in sustainable success

As the "Initiate" phase was all about getting first results quickly, the "Scale" phase is about making sure that the foundations of the first projects are solid enough to support growth in all dimensions, for example :

- Governance: New countries or business units added in the scope,
- Data: New data sets,
- Platform:
  - Bigger data volume,
  - Higher number of users,
  - Stronger workload intensity.

This phase will demonstrate that success is repeatable and reliable, and that the foundations are solid for the long run.

You will prepare for the next phase by identifying what business processes could be optimized by the new analytics delivered by the new platform.

Transform: Operationalize Big Data

In this stage, you will move forward to achieving "insights at the point of action". This phase is critical, as you will transform some of your business processes to take advantage of these new insights.

The objective here is to support corporate development through a link to operational process transformation. It will involve the addition of new analytics such as machine learning and discovery algorithms. This will allow the data platform to connect to operational and mission-critical systems.

Effective governance will help you achieve this transformation and get to a whole new level of ROI for your Big Data initiatives.

Capgemini and Cloudera have come together to help clients act now on the promises of Big Data - helping them to start small, explore the potential, test the waters, and move forward with confidence that the solutions put in place now will scale for the future.

Run: Faster, Better, Cheaper

During this stage you will focus on industrialization, reducing costs while increasing agility and capability to deliver Big Data insights throughout your organization.

You get to a level of maturity where you can define and refine iterative processes in order to:

- Constantly fuel your innovation roadmap with, for example, quarterly updates on your business drivers
- Optimize your governance by increasing the agility and time to market for the new business needs you will integrate on your platform
- Constantly optimize your platform from a TCO and performance perspective
- Constantly add new valuable data sets to continue to get richer insights for your business

Capgemini's global capabilities of more than

9,600

professionals

in **Business Information Management** and its specific **Big Data Center of Excellence** (the BIM CUBE), combined with Cloudera's expertise in Big Data platforms.

Capgemini and Cloudera can help you get Value out of Big Data

In this data-driven world, increased volume, variety, and velocity of data have merely started the conversation around Big Data. The real question is how to create value and stay ahead of the competition by transforming your organization around data, and creating the right offers and services for your customers and partners.

Capgemini has collaborated extensively with Cloudera to build an integrated approach to leveraging Big Data. Big Data is not only about technology but also the need to address all dimensions, ranging from business to IT, in order to execute well.

Together we have already helped clients in telecommunications, the public sector, consumer products, retail, manufacturing, and financial services. We have achieved this by utilizing Capgemini's global capabilities of more than 9,600 professionals in Business Information Management and its specific Big Data Center of Excellence (the BIM CUBE), combined with Cloudera's expertise in Big Data platforms.

Capgemini and Cloudera will help you become a truly innovative, information-centric company. Together, we address the full potential for Big Data within your organization, and we make it work for you.

We can get you started right now.

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

cloudera

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, process 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. Only Cloudera offers everything needed on a journey to an enterprise data hub, including software for business critical data challenges such as storage, access, management, analysis, security and search. As the leading educator of Hadoop professionals, Cloudera has trained over 20,000 individuals worldwide. Over 1,000 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.

More information is available at
www.cloudera.com

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

MCOS_GI_MK_20140627