

Making the case for big data and Hadoop in the enterprise



1

Introduction

2

Healthcare

3

Financial
services

4

Energy and
utilities

5

Telecom

6

Automotive
and
manufacturing

7

How
BigInsights
helps

8

The big
data payoff

Introduction

Big data—data from many sources, of varying formats, both structured and unstructured—means different things in different industries. But as different as their needs and usage of big data may be, there is one commonality among all industries: the opportunity to plumb big data for better, more informed perspectives on their customers, products, partners, competitors and strategies.

For example, in the loosely coupled world of healthcare payers, patients and providers, big data and analytics can provide insight into patterns that help improve care, patient care and organizational efficiency. In financial services, big data technologies can analyze data from multiple sources, including

streaming data-in-motion, to identify potential fraud faster. Manufacturers can dig deep into everything from product quality to supply chain optimization by analyzing big data from machine sensors as well as customer records.

As organizations begin to explore the possibilities enabled by big data and analytics, they need new ways to store and access data—fast. Apache Hadoop provides an answer to that challenge. Hadoop is able to fulfill several key aspects of data analysis that were traditionally too complex or expensive. For instance, it can access unstructured and semi-structured complex data in both traditional applications—such as documents, notes, contacts, spreadsheets



and emails—and in innovative applications such as news media, web feeds, video, graphics and sensor data.

This e-book highlights the benefits of Hadoop across several industries and explores how IBM® BigInsights for Apache™ Hadoop® combines open source Hadoop with enterprise-grade management and analytic capabilities.

Critical success factors in deploying Hadoop

Organizations in any industry can derive consistent benefits from deploying Hadoop, including better system performance leading to faster analytics results; improved risk management and operational efficiency; and faster innovation of new products and services.

But these advantages are gained only by paying attention to specific best practices for deploying a Hadoop-based project. Here are four to keep in mind.

1

Organizations should choose projects with a high potential return on investment, for which data sources are readily accessible and already in electronic form, and establish clear goals and quantifiable metrics. There should be a strong business need for making the resulting data easily accessible to broad user communities.

2

The data architecture should be extensible to allow addition of other data sources, including streaming data, as needed.

3

As the project continues, create a feedback loop to inform other departments of insights derived about products, marketing and sales. This helps promote the value of analytics, builds a culture that focuses on deriving even better information from analytics, and instills a high level of trust in the data's veracity and completeness.

4

Surround Hadoop with a strong ecosystem of big data tools and analytics capabilities. The richer the portfolio of capabilities in the selected Hadoop solution, the more freedom teams have to solve problems and advance the organization's insights.

Healthcare

Data is growing and moving faster than healthcare organizations can consume it, yet getting insights from that data into the hands of practitioners can quite literally make the difference between life and death.

Real-time data analysis and distribution is an ongoing hurdle for healthcare organizations. Patient records is just one of many crucial information resources; others include claims, finance records, customer relationship management (CRM)

systems, business and care partner organization references, research logs and physician correspondence. Much of this data is unstructured and changes constantly. In addition, it is usually spread across multiple sources and departments. **Getting access to this valuable data and factoring it into clinical and advanced analytics is critical to improving care and outcomes, incentivizing patient behavior and driving efficiencies.**

Successfully harnessing big data with Hadoop and streaming technology unleashes the potential to achieve several critical objectives for healthcare transformation, including:

- Building sustainable healthcare systems and health information exchanges
- Improving clinical treatment effectiveness and reducing readmission rates
- Reducing medical errors and supporting collaboration
- Detecting claims fraud and other attempts to misuse medical resources

Case in point: Asian health bureau

A health bureau in Asia wanted to address the booming area of telemedicine while helping radiologists reduce eye fatigue, so it developed a medical imaging analytics system to automatically detect anomalies. However, it was a challenge to accommodate both disease-specific algorithms and the huge data volume process, so it chose Hadoop in conjunction with IBM technology to mine these large imaging data sets, which can range anywhere from 10 GB to terabytes or even petabytes.

The medical imaging diagnostics platform is expected to significantly improve patient healthcare by allowing physicians to exploit the experience of other physicians in treating similar cases, and inferring the prognosis and the outcome of treatments. The platform brings together all relevant data, allowing physicians to see consensus opinions as well as differing alternatives, and helps reduce diagnosis uncertainty.



Financial services

Big data is a familiar management challenge for the financial services industry, which has long been characterized by the tremendous volumes, varieties and velocities of highly detailed data generated by a wide array of sources, customers, partners and electronic feeds. However, today's wealth of unstructured data (including social media posts) doesn't always fit into traditional transaction-focused analysis systems, and transaction and customer data streams represent a tempting target for hackers and fraudsters.

Financial services firms that can harness and protect big data, in the form of transactions, real-time market feeds, customer-service records, correspondence and social media posts, can derive more insight about their business than ever before. Successfully harnessing and protecting big data can help financial services firms achieve three critical growth objectives:

- **Improve customer insight** by developing a 360-degree view of their customers, enabling better-targeted marketing campaigns, customized offers and cross-selling opportunities and faster responses through all contact channels.

- **Better manage risk and detect fraud** through real-time analysis of streaming data and combining multiple data sources to determine relationships and patterns that may indicate fraud attempts.
- **Increase flexibility and streamline operations** by using big data technologies to support processes that enhance efficiency, from speedier and better-informed customer service responses to enterprise payment hubs.

Case in point: Teikoku Databank

Tokyo-based Teikoku Databank serves financial services customers in areas such as corporate credit research, credit risk management services, database services, marketing services and e-commerce support services. It maintains corporate credit report files for 1.6 million companies, financial statements from 680,000 companies, a corporate profile database of 1.42 million companies and other corporate data for 4.1 million companies.

To provide even deeper analysis, it needed to integrate public information from other sources, so it set up a system using Hadoop and IBM technology. The company can now process billions of items of textual data in 30 minutes, a process that used to take several days. It can also analyze almost five times as much data for customers, enabling faster response to customer requests. With these advanced capabilities, Teikoku now has a key competitive differentiation in a fast-moving market.



Energy and utilities

Streaming data is a huge challenge—and opportunity—for energy and utility companies. Smart meters, sensors on oil rigs and power stations, plant efficiency systems, thermometers, water flow monitors: all of these resources and more take and send measurements multiple times per day, sometimes multiple times per minute. **Big data technologies can turn the information from smart meter and smart grid projects into meaningful operational insights and insights about company assets and customer behavior.**

Energy and utility companies are relying on Hadoop to help curb energy consumption by analyzing user data and making those insights available to consumers. The goal: show the customer how behavior changes and appliance or other upgrades reduce energy consumption, and enhance direct customer communication.

Operational efficiency also benefits from big data and analytics. Data from the smart grid can help utilities get more from existing resources and plan for proactive maintenance and capacity management.

Analysis of weather and geographic data helps companies identify locations for new green energy deployments (wind farms, solar systems and so on) as well as optimize the transportation of work crews and timely delivery of raw materials or building components to job sites and processing facilities.

Case in point: Vestas

Aarhus, Denmark-based firm Vestas develops, manufactures, sells and maintains wind power systems, and has installed more than 43,000 land-based and offshore wind turbines in 66 countries on six continents. It needs to precisely place wind turbines to maximize their performance and useful life, and so wanted to expand its wind library to incorporate a larger range of weather data over a longer period of time.

This required a more powerful computing platform to run global forecasts much faster, so it deployed IBM

enterprise-grade Hadoop and other IBM technology, ultimately reducing the response time for wind forecasting information by approximately 97 percent—from weeks to hours—to help cut systems development time. The Hadoop platform also improves accuracy of turbine placement with capabilities for analyzing a greater breadth and depth of data. Ultimately, this lowers the cost to customers per kilowatt hour produced and increases customers' return on investment, as well as reducing their IT footprint and costs.



Telecom

Telecommunications is a necessarily data-driven and capital-intensive business. Mobile network rollouts and the increasing use of mobile devices and social media generate huge amounts of customer and market data. Quick responses to changing market conditions are imperative to remaining competitive.

Adopting Hadoop can help service providers reduce infrastructure costs while enabling speedy go-to-market moves. Networks must run at peak performance to both grow and retain the customer base. Hadoop's extreme scalability can accommodate large data volumes at reasonable cost, and its machine learning capabilities can help prevent service disruptions. Effective capture of phone and device data supports innovative

product development and enables companies to respond to product failures in a timely manner.

Hadoop can also handle billions of call data records, allowing that data to be stored in a diminutive, cost-effective fashion. With greater and deeper visibility into their customers, telcom organizations can provide services and content in line with current interests and reduce customer churn.

Case in point: Celcom

The Axiata Group, of which Kuala Lumpur-based Celcom is a member company, serves more than 190 million mobile subscribers. While its core business remains prepaid and postpaid mobile voice services, it has rapidly growing segments in mobile broadband, mobile commerce, enterprise solutions and bulk wholesale business. Because it operates in a saturated and highly competitive market, Celcom needed to dramatically improve its understanding of its customers to help retain them and increase its market share. It wanted to use existing customer information and analytics to manage a targeted marketing campaign.

To achieve this, the company deployed Hadoop and IBM analysis tools to target individual customers with marketing messages most appropriate to their usage. By providing customers with more suitable plans, Celcom could help build customer satisfaction and loyalty, in addition to raising average revenue per user (ARPU) rates and reducing churn. With these analysis tools, it was able to reduce new campaign launch time by more than 80 percent and improve campaign performance by more than 70 percent.



Automotive and manufacturing

Growing use of telematics and other sensor technologies is changing the way automotive and manufacturing companies approach their customers and their business strategies. This streaming data joins the massive volume of data from internal business operations, R&D activities, supply chain management and enterprise resource planning (ERP) systems, and customer data to present organizations with huge opportunities to gain a competitive edge.

Deploying Hadoop and other big data and analytics technologies allows auto and manufacturing companies to manage all of that data effectively. These three imperatives are especially critical in a global marketplace:

- **Increase customer satisfaction and retention:** Hadoop enables organizations to integrate data from call center operations, social media, service records, product surveys and more to gauge customer sentiment and respond appropriately to feedback. Marketing and sales teams can better identify cross-sell opportunities, and design teams can collect opinions on new and proposed design concepts.

- **Enhance product quality:** From proactive service alerts to analyzing the manufacturing process and testing component quality, big data holds a wealth of insight that can inform manufacturing efforts for companies that can process and analyze it accurately.
- **Optimize the supply chain:** Internal suppliers, external suppliers, transportation providers and inventories generate massive amounts of complex data. Hadoop can collect and analyze the structured and unstructured data collected by supply chain managers, RFID tags and manufacturing floor sensors to generate lead-time analysis that helps prevent stock-outs, breakdowns and delays in the process.

Case in point: American auto manufacturer

A Detroit-based automobile manufacturer wanted to systematically capture and analyze massive amounts of data, improve system efficiency, streamline data analysis and accelerate decision making. It had a functionally oriented, application-driven IT and data management environment, with data siloed in largely outsourced applications. The applications also belonged to separate, large functional areas, such as engineering or marketing, with minimal visibility and data sharing among the different areas.

The IT department deployed Hadoop and IBM technology to consolidate data in a single enterprise data warehouse, making data available across business lines

so the company could use analysis to answer cross-functional questions and improve its business performance and competitiveness. This opened up new analysis points, allowing the company to develop a project that collects data from dealers at the point of repair, including sales and marketing information as well as services information and more. Consolidating this data, the system analyzes it against manufacturing data and product design information to help the manufacturer identify service problems better, including warranty issues, issues with specific suppliers or assemblers, or problems in product design and engineering.



How IBM BigInsights helps

IBM provides a standard implementation of open source Hadoop so customers can leverage its rich functionality, while taking advantage of additional enterprise-class data capabilities. These capabilities include:

- **IBM Big SQL:** SQL-on-Hadoop with massive parallel processing query optimization and execution that faithfully implements the ANSI SQL standard and operates on native Hadoop data types
- **IBM BigSheets:** A spreadsheet-style data manipulation and visualization tool for business users
- **Big R:** An integrated implementation of the open source R language that automatically parallelizes statistical calculations across a Hadoop cluster

- **Adaptive MapReduce:** An alternative low-latency framework compatible with Apache MapReduce that delivers better performance and workload reliability
- **IBM GPFS™ FPO:** An alternative POSIX-compliant Hadoop Distributed File System (HDFS)

IBM also offers a number of development tools, including AQL for accelerated development of text analytic applications; JAQL, a development and query language (based on JSON); development toolkits for rapidly building and deploying particular types of applications, such as social media or machine data analytics; and an integrated Eclipse-based integrated development environment (IDE) that streamlines the process of developing and maintaining big data applications.

IBM BigInsights makes it simpler to use Hadoop to get value out of big data and build big data applications. It bolsters the open source technology to withstand enterprise demands by adding:

- Optimized workloads for performance
- Best-in-class analytical capabilities
- Data discovery, development and provisioning features
- Security and data management enhancements
- Administrative and support capabilities

The result is a more user-friendly solution for complex, large-scale big data and analytics projects.

The big data payoff

No matter what industry you're in, your organization can derive advantages from using BigInsights for Hadoop. Because IBM has integrated its relational database management system tools with Hadoop, IT can use familiar technology and processes to take advantage of Hadoop (such as SQL, spreadsheets and R). IBM also makes it easy to accommodate data regardless of source, from structured CRM and supplier relationship management systems to unstructured e-mail, social media and other industry-specific applications. In addition, IBM has created links that make it easier to pull data out of unstructured data sources, helping to bridge the skills gap that may appear when companies deploy new technology.

Hadoop's ability to bring together previously disparate data sources means bigger, better insights for companies. BigInsights makes the benefits of Hadoop more accessible. By combining the two, organizations can leverage both internal and third-party external data and address state-of-the-art data analysis. These cutting-edge capabilities span a variety of industries and requirements, including real-time analysis of streaming data, analysis of purchasing data to identify for potential fraud, analysis of social media channels for incipient support issues, and analysis of healthcare records to potentially identify disease breakouts.



By combining Hadoop processing speed with IBM analytics expertise and innovation, organizations can now realize the full value of all enterprise data resources—and get answers to industry-specific questions that were previously unfeasible.

Resources

You can learn more about Hadoop and big data, and how to get started with big data in your organization, by visiting the following sites:

- www.ibmbigdatahub.com
- www.bigdatauniversity.com
- <https://developer.ibm.com/Hadoop>
- <https://developer.ibm.com/streamsdev>

To learn more about Hadoop, big data analytics and IBM BigInsights for Apache Hadoop, please contact your IBM representative or IBM Business Partner, or visit: ibm.com/hadoop

You can also download IBM BigInsights QuickStart Edition, free for non-production use, to start learning Hadoop today at: ibm.com/software/data/infosphere/biginsights/quick-start

Additionally, IBM Global Financing can help you acquire the software capabilities that your business needs in the most cost-effective and strategic way possible. We'll partner with credit-qualified clients to customize a financing solution to suit your business and development goals, enable effective cash management, and improve your total cost of ownership. Fund your critical IT investment and propel your business forward with IBM Global Financing. For more information, visit: ibm.com/financing



© Copyright IBM Corporation 2015

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
March 2015

IBM, the IBM logo, ibm.com, BigInsights, GPFS, and InfoSphere are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at ibm.com/legal/copytrade.shtml

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions. THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.

Actual available storage capacity may be reported for both uncompressed and compressed data and will vary and may be less than stated.



Please Recycle