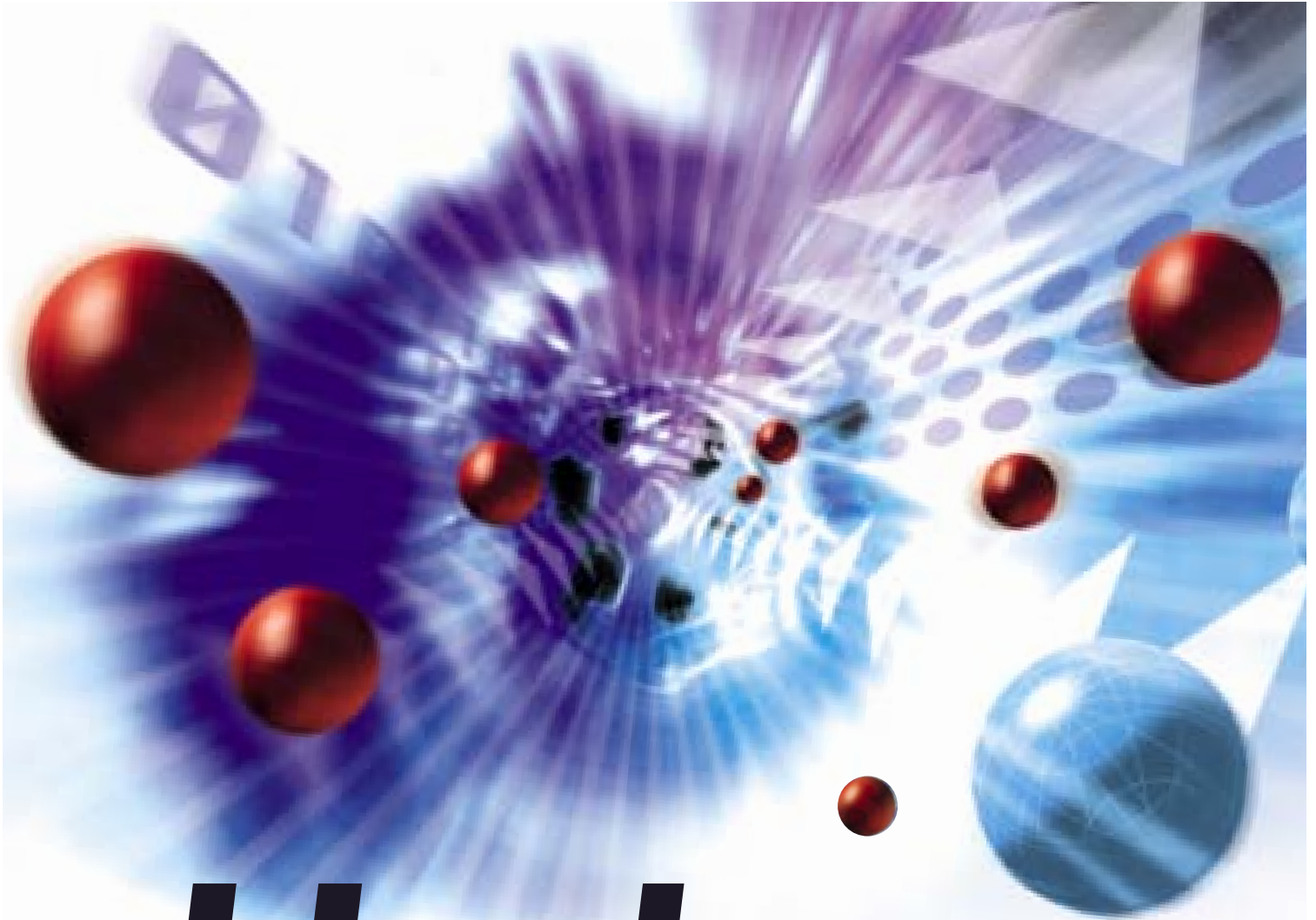


IT *in* Europe

DATA MANAGEMENT / BI

In association with SearchDataManagement.co.UK



Hadoop

THE FUTURE OF ENTERPRISE DATA WAREHOUSING?

Solving the problems posed by "big data" has led data architects beyond traditional database and business intelligence technologies to open source Hadoop.



'Big Data' and Mobile BI Open Up New Fields to Till

HOME

EDITOR'S NOTE

IS HADOOP
THE FUTURE OF
ENTERPRISE DATA
WAREHOUSING?

MOBILE BI TAKES
BABY STEPS INTO
THE ENTERPRISE

HOW TO MANAGE
GLOBAL DATA
QUALITY

THE WORLD OF data is pregnant with new business possibilities. Such is the underlying theme of the three articles in this issue of *IT in Europe: Data Management & BI Edition*.

Hadoop, once a toy elephant, is now [a software framework that supports big-data analytics](#), harnessing up to thousands of nodes and petabytes of data. Groupon, the daily deal site, uses a distribution of Hadoop to analyse transaction data generated by over 70 million users. NYSE Euronext, which operates equities and derivatives markets in Europe and New York, uses similar technology to manage transaction data that is growing by up to 2 TB per day.

Big data is beginning to reshape enterprise architecture, as it reaches beyond the bounds of traditional database and business intelligence tools. It bears a cost advantage, but there is a skills deficit and other technical impediments, such as a lack of support for analysing Hadoop-resident data by BI tools vendors. That is, however, beginning to change.

[Rendering BI more mobile](#) could also open up new business possibilities. HTI, a UK-based toy maker, is benefiting from real-time stock reports delivered to the BlackBerrys of senior management. Dutch mortgage broker De Hypotheekshop is sending real-time market information to its advisors on iPads; the company reports significant cost savings and other benefits.

In theory, the Internet has opened global markets to even the smallest business. But you need to [get the management of data quality right internationally](#), across a sea of maddening differences. The rewards can be considerable if you do. ■

Brian McKenna

UK Bureau Chief, SearchDataManagement.co.UK



Is Hadoop the Future of Enterprise Data Warehousing?

Solving the problems posed by “big data” has led data architects beyond traditional database and business intelligence technologies to open source Hadoop.

BY JESSICA TWENTYMAN

HOME

EDITOR'S NOTE

IS HADOOP
THE FUTURE OF
ENTERPRISE DATA
WAREHOUSING?

MOBILE BI TAKES
BABY STEPS INTO
THE ENTERPRISE

HOW TO MANAGE
GLOBAL DATA
QUALITY

PROBLEMS DON'T CARE how you solve them,” said James Kobielus, an analyst with Forrester Research, in a recent blog post on the topic of ‘big data.’ “The only thing that matters is that you do indeed solve them, using any tools or approaches at your disposal.”

In recent years, that imperative—to solve the problems posed by big data—has led data architects at many organisations on a journey of discovery. Put simply, the traditional database and business intelligence (BI) tools that they routinely use to slice and dice corporate data are just not up to the task of handling big data.

To put the challenge into perspective, think back a decade: few enterprise data warehouses grew as large as a terabyte. By 2009, Forrester analysts were reporting that more than two-thirds of enterprise data warehouses (EDWs) deployed were in the 1- to 10-TB range. By 2015, they claim, a majority of EDWs in large organizations will be 100 TB or larger, “with petabyte-scale EDWs becoming well-entrenched in sectors such as telecoms, financial services and Web commerce.”

What’s needed to analyse these big-data stores are special new tools and approaches that deliver “extremely scalable analytics,” said Kobielus. By “extremely scalable,” he means analytics that can deal with data that stands out in four key ways: for its volume (from hundreds of terabytes to petabytes and beyond); its velocity (up to and including real-time, sub-second delivery); its variety (diverse structured, unstructured and semi-structured formats); and its volatility (wherein scores to hundreds of data sources come and go from new applications, online services, social networks and so on).

HADOOP FOR BIG DATA ON THE RISE

One of the principal approaches to emerge in recent years has been Apache Hadoop, an open source software framework that supports data-intensive distributed analytics potentially involving thousands of nodes and petabytes of data.

The underlying technology was originally invented at search engine giant Google by in-house developers looking for a way to usefully index textual data and other “rich” information and present it back to users

in meaningful ways. They called this technology MapReduce—and today’s Hadoop is an open source version, used by organizations for analytics that are deep and computationally intensive. In addition to Hadoop MapReduce, the Apache Hadoop framework includes a distributed file system and a set of utilities.

In a recent survey of 163 enterprise users conducted by Ventana Research on behalf of Hadoop specialists Cloudera and Karmasphere as well as Pervasive Software, 94% of the respondents using Hadoop reported that they can now perform analyses on large volumes of data that weren’t possible before; 88% said that they analyse data in greater detail; and 82% can now retain more of their data.

It’s still a niche technology, but Hadoop’s profile received a serious boost over the past year, thanks in part to startup companies such as Cloudera and MapR that offer commercially licensed and supported distributions of Hadoop. Its growing popularity is also the result of serious interest shown by large vendors like EMC, IBM and Teradata. EMC bought Hadoop specialist Greenplum in June 2010; Teradata announced its acquisition of Aster Data in March 2011; and IBM announced its own Hadoop offering, InfoSphere BigInsights, in May 2011.

What stood out about Greenplum for EMC was its x86-based, scale-out massively parallel processing (MPP) system with a shared-

“The cost-performance curve of commodity servers and storage is putting seemingly intractable complex analysis on extremely large volumes of data within the budget of an increasingly large number of enterprises.”

—MARCUS COLLINS, *Gartner*

HOME

EDITOR'S NOTE

IS HADOOP THE FUTURE OF ENTERPRISE DATA WAREHOUSING?

MOBILE BI TAKES BABY STEPS INTO THE ENTERPRISE

HOW TO MANAGE GLOBAL DATA QUALITY

nothing design, said Mark Sears, architect at the new organization EMC Greenplum. "Not everyone requires a setup like that, but more and more customers do," he said.

"In many cases," he continued, "these are companies that are already adept at analysis—at mining customer data for buying trends, for example. In the past, however, they may have had to query a random sample of data relating to 500,000 customers from an overall pool of data of five million customers. That opens up the risk that they might miss something important. With a big data approach, it's perfectly

possible and relatively cost-effective to query data relating to all five million."

"With a big data approach, it's perfectly possible and relatively cost-effective to query data relating to all five million customers."

—MARK SEARS,
EMC Greenplum

HADOOP ILL UNDERSTOOD BY THE BUSINESS

While there's a real buzz around Hadoop among technologists, it's still not widely understood in business circles. In the simplest terms, Hadoop is engineered to run across a large

number of low-cost commodity servers and distribute data volumes across that cluster, keeping track of where individual pieces of data reside using the Hadoop Distributed File System (HDFS). Analytic workloads are performed across the cluster, in an MPP model, using related open source tools such as Pig and Hive. Results, meanwhile, are delivered as a unified whole.

Earlier this year, Gartner analyst Marcus Collins mapped out some of the ways he's seeing Hadoop used today: by financial services companies, to discover fraud patterns in years of credit-card transaction records; by mobile telecoms providers, to identify customer churn patterns; by academic researchers, to identify celestial objects from telescope imagery.

"The cost-performance curve of commodity servers and storage is putting seemingly intractable complex analysis on extremely large volumes of data within the budget of an increasingly large number of enterprises," he concluded. "This technology promises to be a significant competitive advantage to early adopter organisations."

Early adopters include daily deal site Groupon, which uses the Cloud-

HOME

EDITOR'S NOTE

IS HADOOP THE FUTURE OF ENTERPRISE DATA WAREHOUSING?

MOBILE BI TAKES BABY STEPS INTO THE ENTERPRISE

HOW TO MANAGE GLOBAL DATA QUALITY

era distribution of Hadoop to analyse transaction data generated by over 70 million registered users worldwide, and NYSE Euronext, which operates the New York Stock Exchange as well as other equities and derivatives markets across Europe and the US. NYSE Euronext uses EMC Greenplum to manage transaction data that is growing, on average, by as much as 2 TB each day. US bookseller Barnes & Noble, meanwhile, is using Aster Data's nCluster technology (now owned by Teradata) to better understand customer preferences and buying patterns across three sales channels: its retail outlets, its online store and e-reader downloads.

HOME

EDITOR'S NOTE

IS HADOOP
THE FUTURE OF
ENTERPRISE DATA
WAREHOUSING?

MOBILE BI TAKES
BABY STEPS INTO
THE ENTERPRISE

HOW TO MANAGE
GLOBAL DATA
QUALITY

SKILLS DEFICIT IN BIG-DATA ANALYTICS

While Hadoop's cost advantage might give it widespread appeal, the skills challenge it presents is more daunting, according to Collins. "Big-data analytics is an emerging field, and insufficiently skilled people exist within many organisations or the [wider] job market," he said.

Skills-related issues that stand in the way of Hadoop adoption include the technical nature of the MapReduce framework, which requires users to develop Java code, and the lack of institutional knowledge on how to architect the Hadoop infrastructure. Another impediment to adoption is the lack of support for analysis tools used to examine data residing within the Hadoop architecture.

That, however, is starting to change. For a start, established BI tools vendors are adding support

for Apache Hadoop: one example is Pentaho, which introduced support in May 2010 and, subsequently, specific support for the EMC Greenplum distribution.

Another sign of Hadoop's creep towards the mainstream is support from data integration suppliers, such as Informatica. One of the most common uses of Hadoop to date has been as an engine for the transformation stage of extract, transform and load (ETL) processes: the MapReduce technology lends itself well to the task of preparing data for loading and further analysis in both conventional RDBMS-

"Big-data analytics is an emerging field, and insufficiently skilled people exist within many organisations or the [wider] job market."

—MARCUS COLLINS, *Gartner*

based data warehouses or those based on HDFS/Hive. In response, Informatica announced that integration was underway between EMC Greenplum and its own data integration platform in May.

Additionally, there's the emergence of Hadoop appliances, as evidenced by EMC's Greenplum HD (a bundle that combines MapR's version of Hadoop, the Greenplum database and a standard x86-based server), announced in May, and more recently, in August, the

Dell/Cloudera Solution for Apache Hadoop (which combines Dell PowerEdge C2100 servers and PowerConnect switches with Cloudera's Hadoop distribution and its Cloudera Enterprise management tools).

Finally, Hadoop is proving well suited to deployment in cloud environments, a fact which gives IT teams the chance to experiment with pilot projects on cloud infrastructures. Amazon, for example, offers Amazon Elastic MapReduce as a hosted service running on its Amazon Elastic Compute Cloud (EC2) and Amazon Simple Storage Service (S3). The

Apache distribution of Hadoop, moreover, now comes with a set of tools designed to make it easier to deploy and run on EC2.

"Running Hadoop on EC2 is particularly appropriate where the data already resides within Amazon S3," said Gartner's Collins. If the data does not reside on Amazon S3, he said, then it must be transferred: the Amazon Web Services (AWS) pricing model includes charges for network costs and Amazon Elastic MapReduce pricing is in addition to normal Amazon EC3 and S3 pricing. "Given the volume of data required to run big-data analytics, the cost of transferring data and running the analysis should be carefully considered," he cautioned.

The bottom line is that "Hadoop is the future of the EDW and its footprint in companies' core EDW architectures is likely to keep growing throughout this decade," said Kobielus. ■

"Hadoop is the future of the EDW and its footprint in companies' core EDW architectures is likely to keep growing throughout this decade."

—JAMES KOBIELUS,
Forrester

HOME

EDITOR'S NOTE

IS HADOOP THE FUTURE OF ENTERPRISE DATA WAREHOUSING?

MOBILE BI TAKES BABY STEPS INTO THE ENTERPRISE

HOW TO MANAGE GLOBAL DATA QUALITY

Mobile BI Takes Baby Steps into the Enterprise

Mobile BI needs to go beyond pushing static reports to smartphones to bring significant new value. European corporate IT adoption is nascent.

BY JIM MORTLEMAN

HOME

EDITOR'S NOTE

IS HADOOP THE FUTURE OF ENTERPRISE DATA WAREHOUSING?

MOBILE BI TAKES BABY STEPS INTO THE ENTERPRISE

HOW TO MANAGE GLOBAL DATA QUALITY

S SMARTPHONES AND tablet PCs continue their incursion into the enterprise, IT departments are increasingly being asked to provide access to business data on such devices. The IT industry has cottoned on to the trend, and one of the notable areas being hyped at the moment is mobile business intelligence (BI).

But while the early adopters of the technology tended towards pushing static reports to devices such as BlackBerrys through email, it is the enhanced visualisation and interactivity possible on bigger-screen tablet devices—predominantly Apple's iPad—that seem to offer the most potential for the future.

"The BI vendors were very quick to develop apps," Gartner analyst Andreas Bitterer said. "MicroStrategy brought out a mobile application just two or three months after the iPad came out, for example, quickly followed by others like IBM, Information Builders, Oracle and SAP."

But Bitterer added that some of the more interesting offerings emanate from smaller, newer players like PushBI and MeLLmo (with Roambi), although he anticipates these newcomers will be acquired by major players within two years. "The established BI companies come out with middle-lane user interfaces even though they have the resources to develop something compelling, but it is the creativity of the smaller companies that shines out at the moment," he said.

Early adopters such as the organizations profiled below are showing that even basic mobile BI can offer significant benefits. But Ovum analyst Mike Davis thinks we've a long way to go before the market reaches anything like maturity. "At the moment, the effort is largely

on replicating desktop BI on smaller mobile devices," he said. "But as organisations' awareness develops, they will start to put pressure on vendors to provide richer, real-time, interactive BI on these devices. But maturity? That's probably three to five years away."

PLAYING AROUND WITH MOBILE BI

HTI is a UK-based toy manufacturer and distributor that supplies major UK retailers such as Argos, Toys R Us, Tesco and Asda with brand-name toys, including the Toy Story, Peppa Pig and Hello Kitty merchandise lines. Ensuring that the company has the right stock for the busy Christmas and summer shopping seasons is no mean feat. With a three- to four-week lead time to obtain orders from manufacturers in the Far East, understanding market trends in real time is critical to meeting peaks in demand for particular products and maximising profit margins.

Back in 2007, all transactions, orders and stock were handled by the company's ERP system, which was unable to produce the meaningful real-time reports that business users needed. For the next two years, IT director John Lord worked with UK-based IBM business partner Enterprise BI to implement a Cognos data warehousing and reporting system that would give the business the data it needed, when it needed it. The system uses the Cognos Go Mobile service to produce customised reports for key management personnel, such as HTI's sales director, managing director and staff members; the reports are emailed weekly in a BlackBerry-friendly format.

"Organisations will start to put pressure on vendors to provide richer, real-time, interactive BI on these devices."

—ANDREAS BITTERER, *Gartner*

HOME

EDITOR'S NOTE

IS HADOOP THE FUTURE OF ENTERPRISE DATA WAREHOUSING?

MOBILE BI TAKES BABY STEPS INTO THE ENTERPRISE

HOW TO MANAGE GLOBAL DATA QUALITY



Lord said, "We're sending out targeted reports which only contain information critical to individual buyers or product managers in particular areas or sectors, as well as summary reports for the board."

Although the email reports are static and don't allow for rich visualisation or drilling down into information, Lord said managers at HTI generally don't need those functions. "The point is they no longer have to wait for month-end reports and have the information at their fingertips to act quickly when they spot an issue," he said. "The system can also produce further customised reports fairly easily should these be needed."

Lord added that he would be interested in moving to a tablet-based user interface in future. "Because of its size, there's a limit to what you can do on a BlackBerry," he said. "Using iPads to see that information in visual form would be great. When there's an app that replicates what Cognos can do on the desktop, with dashboards and the like, that will be fantastic—but we're not there yet."

DE HYPOTHEEKSHOP: AT HOME ON iPADS

Independent Dutch mortgage adviser De Hypotheekshop has 180 franchise branches across the Netherlands and deals with more than 15,000 clients a year. As a leading player in the volatile and heavily regulated financial market, anything the company can do to ease access to critical information will be key to its ongoing business success and its regulatory compliance efforts.

As well as providing real-time data on

"When there's an app that replicates what Cognos can do on the desktop, that will be fantastic—but we're not there yet."

—JOHN LORD, *IT director, HTI*



HOME

EDITOR'S NOTE

IS HADOOP THE FUTURE OF ENTERPRISE DATA WAREHOUSING?

MOBILE BI TAKES BABY STEPS INTO THE ENTERPRISE

HOW TO MANAGE GLOBAL DATA QUALITY

costs, revenues, margins and distribution channels, De Hypotheekshop wanted to give franchisees real-time access to market information and the ability to more easily meet demands for compliance-related information from regulators.

“The financial slump and increasing market demand for transparency has been game-changing for the financial sector,” said Patrick Nelissen, the company’s franchise and formula manager. “To ensure

a return on investment from our activities as a financial intermediary, we have to run things on the basis of up-to-date information rather than gut feeling. In addition, it is paramount [that] franchisees receive management information quickly, in a user-friendly way.”

In 2009, the company opted for a QlikView-based BI application implemented by Netherlands-based specialist financial systems provider Hippo Line. QlikView had the advantage of being able to integrate data from De Hypotheekshop’s Kompas CRM system, SQL Server database and .Net budgeting module, as well as up-to-date market information from the Dutch land reg-

istry database and central bureau of statistics. It also enabled the mortgage adviser to meet demands for transparency by using a compliance module.

But the critical step in terms of getting the system adopted, and embraced, quickly was the company’s decision to deploy the software on iPads across its branch network. “The iPad has a wow factor that makes users want to try it out. Companies normally have to invest in user adoption programs, but with the iPad we didn’t need to do this,” Nelissen said.

Since the deployment of the mobile BI application, De Hypotheekshop says it has seen significant business benefits, including 10% average cost savings across branches, with proportion of sales leads or customer approaches that are converted into actual sales up 30%, 15% savings on marketing and the ability to answer 90% of inves-

“To ensure a return on investment from our activities as a financial intermediary, we have to run things on the basis of up-to-date information rather than gut feeling.”

—PATRICK NELISSEN

*Franchise and formula manager,
De Hypotheekshop*

HOME

EDITOR'S NOTE

IS HADOOP THE FUTURE OF ENTERPRISE DATA WAREHOUSING?

MOBILE BI TAKES BABY STEPS INTO THE ENTERPRISE

HOW TO MANAGE GLOBAL DATA QUALITY

tigative questions—from regulators about compliance issues, from auditors and from people making e-discovery or Freedom of Information requests that must legally be fulfilled within a day.

HOME

OASIS MEDICAL SOLUTIONS: A HEALTHY MOBILE FUTURE?

Around 30 UK hospitals use the Web-based Oasis Patient Administration System (PAS), which among other functionality includes the ability to access patient test results. Developer Oasis Medical Solutions wanted to extend the capabilities of the system to give both

EDITOR'S NOTE

IS HADOOP
THE FUTURE OF
ENTERPRISE DATA
WAREHOUSING?

“[There is clinician demand for the applications ... but the] stumbling block tends to be the powers that be in the hospital. They see a mobile phone; they see a toy; they see a security risk”.

—ROBERT CAMPBELL

Managing director, Ecommnet

clinical and nonclinical staff rapid access to test results, such as pathology and radiology lab reports, through iPhones. The company knew any increase in the speed of access to such results could help hospital staff decide on treatments and actions more quickly, potentially improving care and saving more lives.

Because the data needed to be moved between iPhones and the back-end Oracle database used by the PAS in hospitals, where 3G and Wi-Fi coverage is often patchy, there was also a need for reliable mobile data management and synchronisation capabilities. Working with UK-based development partner Ecommnet, Oasis opted for SQL Anywhere from

SAP's Sybase subsidiary, which enabled it to develop the application quickly without the need to build a dedicated system to manage the storage and transmission of data. Development speed was key, since Oasis wanted to demonstrate the system at an upcoming trade show.

Another consideration was privacy, so the application also employed the Sybase Afaria security solution to protect sensitive patient information on the iPhone. Oasis began piloting the application at three hospitals last year. The company is now looking to extend the application to work with other mobile devices and potentially add mobile support to other components of the PAS.

Robert Campbell, managing director of Ecommnet, said he can see

MOBILE BI TAKES
BABY STEPS INTO
THE ENTERPRISE

HOW TO MANAGE
GLOBAL DATA
QUALITY

potential for similar applications on mobile devices oriented towards patient discharge, which should free up resources, such as beds, to increase throughput. There is clinician demand for the applications, he said, but the “stumbling block tends to be the powers that be in

the hospital. They see a mobile phone; they see a toy; they see a security risk.

“But security is sandboxed within the application itself, and user authentication is tied back to the Oracle database. The Sybase tools encrypt the database and the underlying communications are also encrypted.”

As for the iPhone, Campbell said: “To scale out to nursing staff, the costs might be prohibitive, so Android might [if cheaper] then be used. You could do the devel-

opment within a few days because the underlying technology would be the same.”

He maintained the implementation is “very unusual in that other applications we are seeing in health care are not as direct as this. And integration with a PAS is unusual in the UK.

“Making rapid decisions on the move is the point of this. Getting to a terminal in a hospital is inconvenient, as is carrying around lots of paper.” ■

*Additional reporting by **Brian McKenna**.*

“Making rapid decisions on the move is the point of this. Getting to a terminal in a hospital is inconvenient, as is carrying around lots of paper.”

—ROBERT CAMPBELL

Managing director, Ecommnet

HOME

EDITOR'S NOTE

IS HADOOP THE FUTURE OF ENTERPRISE DATA WAREHOUSING?

MOBILE BI TAKES BABY STEPS INTO THE ENTERPRISE

HOW TO MANAGE GLOBAL DATA QUALITY

How to Manage Global Data Quality

Global data quality management is no cinch: many countries, country-specific values, differing languages and special characters.

Learn how to do it well.

BY KATHY HUNTER

HOME

EDITOR'S NOTE

IS HADOOP
THE FUTURE OF
ENTERPRISE DATA
WAREHOUSING?

MOBILE BI TAKES
BABY STEPS INTO
THE ENTERPRISE

HOW TO MANAGE
GLOBAL DATA
QUALITY

HANKS TO THE Internet, even the smallest business can be global. Customers from all over the world can see what a business offers, and if the company is prepared to operate beyond its own country's borders, these global customers can provide sales opportunities that didn't exist previously.

Larger businesses that have offices worldwide are also moving to operate on an international basis, centralizing databases and business operations to gain efficiencies and save valuable funds. However, many of those who have ventured into the global arena have found that managing the international data that enables truly global business operations is not an easy task.

Global data introduces variation, as it originates from multiple countries, contains varying local formats and country-specific values and is stored in differing languages or special characters.

When dealing with global data, there is much to take into consideration. For example, the differences in international addresses are vast. Not only are there nearly 10,000 languages around the world, but there are also dozens of different addressing formats—some fairly similar, others quite diverse.

Naming conventions for both individuals and businesses vary by country. Many languages have alphabets that include diacritical marks, and some are iconic—using graphical images rather than individual letters. Other data elements, like dates, times and phone numbers, also differ.

For effective global data quality management, those and other

challenges need to be understood and properly managed.

Address formats. Currently, there are 131 different address formats in the world. At the most extreme, some start with the street information and end with the postal code, while others begin with the postal code and end with the street information.

Some postal systems have preferences about the casing of address elements and also the use of abbreviations and punctuation. Likewise, some countries prescribe the use of symbols in a precise manner in order to denote elements of a complex address.

There are 131 different address formats in the world. ... Some start with the street information and end with the postal code, while others begin with the postal code and end with the street information.

Postal address maturity. Postal systems vary around the world. Many countries are just beginning to develop formal processes. In many countries, mail is not delivered to premises. Post office boxes fill the gap, and people and companies travel to the post office to pick up their mail.

This varying maturity means that postal address files used for data validation are often limited and may even be ambiguous, as many entries might

only contain a city name and country.

Address formats may also evolve over time. Old, outdated formats are replaced with newer ones, but the use of new formats takes time to be accepted by the entire population. While this happens, both formats must be anticipated as part of a global data management strategy.

Personal names. The components and order of personal names can differ by country and also by culture. These diverse possibilities highlight the problems associated with using the common form of "First Name" and "Last Name." In the majority of countries, the expectation is that names are shown as given name followed by family name. But that isn't the case in many places. In fact, some countries have multiple options based mainly on culture or religion.

HOME

EDITOR'S NOTE

IS HADOOP
THE FUTURE OF
ENTERPRISE DATA
WAREHOUSING?

MOBILE BI TAKES
BABY STEPS INTO
THE ENTERPRISE

HOW TO MANAGE
GLOBAL DATA
QUALITY

Regional variations. Names and addresses are just the start. Many aspects of customer information have variations based on location:

- › **Phone numbers:** There is a standard format for international phone numbers. However, each section of a number can have a different length, including the country code. The overall length of a phone number changes by country as well. In order to make sense of numbers, people often add characters like hyphens, parentheses and even slashes to them, potentially causing confusion within systems.
- › **Business naming conventions:** Types of business—public companies, limited companies, partnerships, sole traders and so on—are generally country-specific. In order to build a business database, it's important to understand the different types in order to standardize business details and avoid duplicate entries.
- › **Post office boxes:** As mentioned above, post office boxes have great importance in some countries because of the lack of premises-level postal deliveries. It is often the case that a P.O. box will be used for a mailing address and should be stored alongside the physical address of a company, if one exists.

Cultural influences. It is often easier to deal with language barriers than to overcome cultural issues. Unfortunately, getting this wrong can have a dire impact on people's perception of an organization. Care must be taken to handle the strict protocols that some cultures have for names and addresses.

A company's reputation can be damaged when personal details are presented in a culturally disrespectful way. A combination of access to local knowledge and excellent language skills is the best way to overcome these issues.

Diacritics and other special characters. A diacritic is a mark attached to a letter to change its pronunciation or stress. The use of diacritics is central to most alphabets. As a result, when dealing with global data, there likely will be quite a number of diacritical marks that need to be maintained. For example, languages like German and French present issues because of the presence of enhanced characters, such as the umlaut (ü) in German and the accent aigu (é) in French.

Diacritical marks are just the beginning of issues related to differ-

HOME

EDITOR'S NOTE

IS HADOOP
THE FUTURE OF
ENTERPRISE DATA
WAREHOUSING?

MOBILE BI TAKES
BABY STEPS INTO
THE ENTERPRISE

HOW TO MANAGE
GLOBAL DATA
QUALITY

ent character sets. Around the globe, there is a large variety of different character sets used by individual languages. Many of the world's emerging markets are in areas that use the most "unusual" of these. For example, some character sets are based on pictograms, as are some Asian languages. Other places that use different forms of char-

acter sets include Russia, Bulgaria, Greece and countries in the Middle East.

When data is managed locally, there is generally no need to put anything special in place to deal with a country's character sets because the existing local IT infrastructure will have been set up to handle that process. However, when data is managed globally or even regionally, multiple character sets can become quite a problem.

Character sets require encoding so they can be supported within systems. These encodings are commonly known as code pages—tables of mappings that manage the relationships between the codes and the characters they represent.

A set of code pages designated as Unicode brings all known character sets together in a way that is language-independent. The use of Unicode will greatly simplify the management of multiple character sets within IT systems and will safeguard against data corruption. To build a single, centralised database for global data management, it is essential to use Unicode to provide the most effective results; currently, UTF-8 is the dominant character encoding format for Unicode implementations.

IS GOING GLOBAL WORTH IT?

With all the issues that need to be handled, it would be easy to think going global on data management isn't worth the effort. However, the challenges are not insurmountable, and the rewards can be great. An excellent reference source for managing global data is *The Global Sourcebook for Name and Address Data Management*, written and published by Graham Rhind. You can find it at www.grcdi.nl; I swear by it.

With a clear understanding of the challenges and in-depth knowledge of the multinational data that needs to be managed, you can successfully unlock your organization's global data potential. ■

The use of Unicode will greatly simplify the use of multiple character sets and will safeguard against corruption.

HOME

EDITOR'S NOTE

IS HADOOP THE FUTURE OF ENTERPRISE DATA WAREHOUSING?

MOBILE BI TAKES BABY STEPS INTO THE ENTERPRISE

HOW TO MANAGE GLOBAL DATA QUALITY

ABOUT THE AUTHORS



JESSICA TWENTYMAN is a business and technology journalist. Over the last 15 years, she has been a regular contributor to some

of the UK's most respected national newspapers and trade magazines, including the *Financial Times*, *The Wall Street Journal's Europe Edition*, *Director* magazine and *Computer Weekly*. The bulk of her work focuses on the technology products and services that smart companies use for real competitive advantage.



JIM MORTLEMAN is an independent writer and commentator with more than 20 years' experience covering new technologies and their implications for business and society.



KATHY HUNTER is an information management consultant at Kynetika. She has more than 20 years of information

systems experience and started her focus on information quality improvement 13 years ago.



Brian McKenna
Site and News Editor
bmckenna@techtarget.com

Hannah Smalltree
Editorial Director
hsmalltree@techtarget.com

Craig Stedman
Executive Editor
cstedman@techtarget.com

Barney Beal
News Director
bbeal@techtarget.com

Kerry Flood
Associate Editor
kflood@techtarget.com

Jason Sparapani
Copy Editor
jsparapani@techtarget.com

Linda Koury
Director of Online Design
lkoury@techtarget.com

FOR SALES INQUIRIES, PLEASE CONTACT:

Brent Boswell
Director of Sales, EMEA
bboswell@techtarget.com
+44 (0) 7584 311 889

HOME

EDITOR'S NOTE

IS HADOOP
THE FUTURE OF
ENTERPRISE DATA
WAREHOUSING?

MOBILE BI TAKES
BABY STEPS INTO
THE ENTERPRISE

HOW TO MANAGE
GLOBAL DATA
QUALITY