

May 2014, HAPPIEST MINDS TECHNOLOGIES

Big Data: Why should enterprises adopt it

Author

Manish Kumar



SHARING. MINDFUL. INTEGRITY. LEARNING. EXCELLENCE. SOCIAL RESPONSIBILITY.

Copyright Information

This document is an exclusive property of Happiest Minds Technologies Pvt. Ltd. It is intended for limited circulation.

Table of Contents

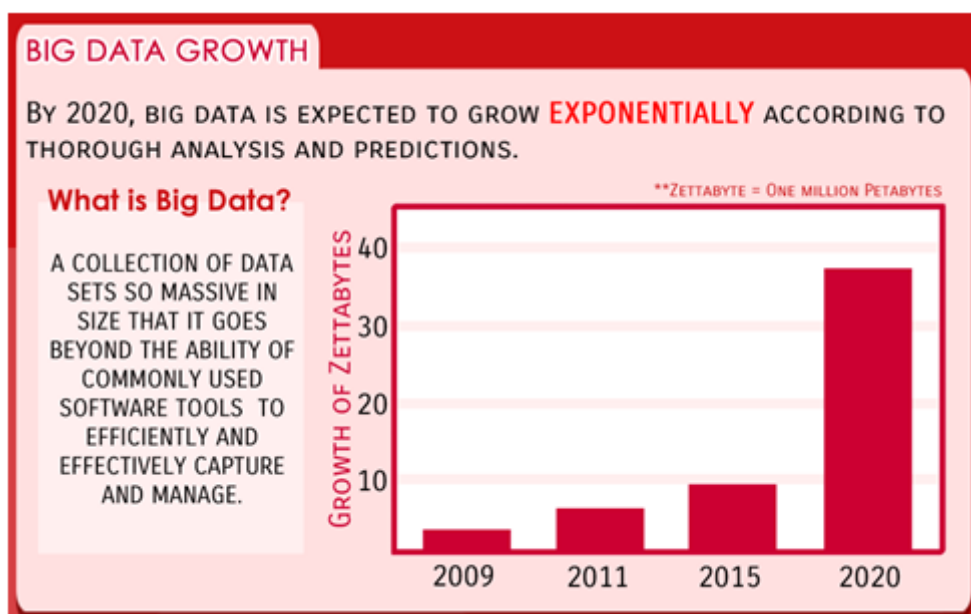
1. Introduction	4
2. Mega Trends: A HappiestMinds Point of View	4
3. Changing Business Models as the Mega Trends Gain Ground	8
4. Staying Relevant with Big Data.....	10
5. Tools for Analysing Business.....	11
6. Conclusion	13

Introduction

Today enterprises are in the midst of a transformation that compels them to be analytics-driven to remain competitive in the coming years. Analysing their data will not only enable them to get a thorough insight about their business, but also give significant competitive edge in the market place. Example of analytics-driven activities includes service innovation, customer experience improvements, detection and remediation of anomalies, and reduction in time to market for products and services. To meet the need for analytics-driven activities, businesses needs to collect, analyze, and store more and more data, often from their diverse sources. They should also adopt such tools and workflows that allow them to quickly and continuously analyse results from stored and transient, high velocity and voluminous data and change their business functions accordingly. This means the systems will be more agile than their predecessors.

This white paper illustrates the key factors that play a vital role in assisting an enterprise to decide why big data is the way forward, and the tools that can help achieve it.

Mega Trends: A HappiestMinds Point of View



Big data growth trend - Source: Syndacast

The above picture depicts how exponentially data is expected to grow in the next few years. A huge volume of valuable unstructured data is being generated on a daily basis from sources like web, social interactions to shopping transaction.

Here are some facts: -

- The New York Stock Exchange generates about one terabyte of new trade data per day
- Facebook hosts approximately 10 billion photos, taking up one petabyte of storage
- Ancestry.com, the genealogy site, stores around 2.5 petabytes of data
- The internet archive stores around 2 petabytes of data, and is growing at a rate of 20 terabytes per month
- The Large Hadron Collider near Geneva, Switzerland, will produce about 15 petabytes of data per year

According to Gartner report, while big data has certainly got people's attention, 2014 will be the year that is catapulted to the forefront. Most of the enterprises will start to focusing on processes and hit their stride for investment in big data.

Here are some mega trends in different sector of enterprises which indicates a growing need of big data in enterprises: -

Mega Trend 1: In healthcare sector

Traditionally, the healthcare industry has trailed behind other industries in the use of big data. Part of the problem stems from resistance to change. Providers are accustomed to making treatment decisions independently, using their own clinical judgment, rather than relying on protocols based on big data. Other obstacles are more structural in nature. Many healthcare stakeholders have underinvested in information technology because of uncertain returns. Although their older systems are functional, they have limited ability to standardize and consolidate data. The nature of the healthcare industry itself also creates challenges: while there are many players, there is no possibility share data easily among different providers or facilities, partly because of privacy concerns. And even within a single hospital, payor or pharmaceutical company, important information often remains siloed within one group or department because organizations lack procedures for integrating data and communicating findings.

But as the market is becoming competitive and a core demand for more is everywhere, stakeholders have now started to access promising new threads of knowledge. Pharmaceutical-industry experts, payors and providers are now beginning to analyze big data to obtain insights. Although these efforts still remain in their early stages, they could collectively help the industry address problems related to variability in healthcare quality and escalating healthcare spend. For instance, researchers can mine the data to see what treatments are most effective for particular conditions, identify patterns related to drug side effects or hospital re-admissions, and gain other important information that can help

patients and reduce costs. Many innovative companies in the private sector, both established players and new entrants, are now focusing and building applications and analytical tools that help patients, physicians and other healthcare stakeholders identify value and opportunities. Fortunately, as their technological capabilities and understanding advance, we expect that innovators will develop solutions which will transform healthcare in a big way, some of which could help substantially reduce the soaring cost of healthcare by delivering consistent, quick and accurate diagnosis or prognosis.

Mega Trend 2: In educational sector

New technologies allow schools, colleges and universities to analyse absolutely everything that happens - from student behaviour, testing results, careers developments of students and educational needs based on changing societies. A lot of these data has already been stored and is used for statistical analysis by government agencies. With the increasing value from high quality online education and development, all data get a completely new meaning. Big data allows changes in the educational field that will revolutionize the way students learn and teachers teach.

Improve student results: The overall goal of big data within the educational system should be to improve student results. Better students are good for society, organisations and educational institutions. Currently, the answers to assignments and exams are the only measures to assess the performance of students. However, during his/her student life, every student generates a unique data trail. This data trail can be analysed in real-time to deliver an optimal learning environment for the student and to gain a better understanding in his/her individual behaviour. It is possible to monitor every action of the students - how long they take to answer a question, which sources they use, which questions they skipped, how much research was done, what is the relation to other questions answered, which tips works best for which student etc. Answers to questions can be checked instantly and can automatically (except for essays) give instant feedback to students.

Improve the learning experience in real-time: When students start working on their own, in their customized learning program, the vast amount of teaching, which most of the time is covered by general topics that have to appeal to all students from different levels, can be done online and by themselves. The professor can monitor all students in real-time and start a much more interesting and deeper conversation on the topic of choice. This will give students the possibility to gain a better understanding of the topics.

Create mass-customized programs: All this data will help to create a customized program for each individual student. Big data allows customization at colleges and universities, even if they have 10.000s of students. This will be created with blended learning; a combination of online and offline learning. It will give students the opportunity to develop their own personalized program, following those classes that they are interested in, working at their

own pace, while having the possibility for (offline) guidance by professors. Providing mass customization in education is a challenge, but algorithms make it possible to track and assess each individual student.

Reduce dropouts, increase results: All these analyses will improve the students' results and perhaps also reduce dropout rates at universities or colleges. Dropouts are expensive for educational institutes and for the society. When students are closely monitored, receive instant feedback and coached based on their personal needs, it can help reduce dropout rates, as mentioned in post one of the technical site.

Mega Trend 3: In social sector

We all know that social media generates huge amounts of data. The explosive growth of social media is one of the reasons that 90% of all the data in the world has been generated in the last two years alone. There are over a billion active users of social media network worldwide, many of whom are frequently active and can be connected by means of their smartphones and tablets. Social media indeed has become a main communication network in the daily lives of people around the world. Did you know that any activity in social media is generating insurmountable data? These big data are waiting to be explored. In fact, social media now embodies the leading and biggest source of consumer data.

Mega Trend 4: In government sector

Although the business sector is leading big data application development, the public/government sector has begun to derive insights to help support decision making in real time from fast-growing in-motion data from multiple sources, including web, biological and industrial sensors, video, email, and social communications. Many white papers, journal articles and business reports have proposed ways governments can use big data to help them serve their citizens and overcome national challenges (such as rising healthcare costs, job creation, natural disasters and terrorism). There is also some skepticism as to whether it can actually improve government operations, as governments must develop new capabilities and adopt new technologies to transform it into information through data organization and analytics.

Mega Trend 5: In insurance sector

The insurance industry is one of the most data driven industries on the planet, so it is amazing how few of them use crowd sourcing solutions to enhance their business. Joe Martinez, an Insurance Adjuster from Austin, TX outlines the problem. "Most insurance companies are content to stand on the shoulders of experts, adjusters like me, accountants and actuaries. They miss out on a big opportunity every year. Big data would allow the

insurance companies to determine variables that they have not been looking at, like the link between mobile phone usage and accidents. “There are data points other than demographics that can more accurately assess risk than demographic data points like age or sex.

Mega Trend 6: In sales sector

Sales department has a good service that can only be acquired through financial transaction, and most buyers are already motivated. It becomes a process of facilitating the transaction and this is where data comes in. According to a top automobile salesman from Redmond, WA says that analytics play a huge part in how he pursues his leads. “I was able to use data on how long it has been since a customer bought a vehicle and how often they had to get it fixed, to calculate a percentage of them that were likely to be unhappy with their vehicle. By focusing direct marketing on this segment I actually saw a huge jump in sales from those who were in the market for a new car without even realizing it. This data gave me jump on the competition.”

Big Data can also help sales teams determine what to concentrate on. The COO of a company that uses data to help companies engage, retain and develop the workforce, talks about a study that asked employees to rank the importance of sales skills. After comparing employee data with the real data, the company found that focus on ‘value creation’ was the most important aspect to the sales process and not ‘sales presentation.’ “In this case, the analysis showed that investing time and money on emphasizing value creation would likely to have the biggest impact on sales revenue,” explains the COO.

Apart from the above explained sectors there are many more other sectors e.g. agriculture, non profits, securities, life sciences, surveillance, defence, intelligence, cyber etc, which are also focusing on adopting big data as per trend shows.

Changing Business Models as the Mega Trends Gain Ground

All the major industries (healthcare, government, education, social, insurance and sales) tracked as part of this research have given clear indication of shifting business models to respond better to their customers journey and be in par with the competitors.

Our researchers worked to identify how mega trends in enterprises are impacting the traditional business models:

1. How to improve efficiency and reduced costs of all machines operated while farming

2. How to improve productivity and efficiency of crops as well as animals
3. How to mitigate weather conditions and optimize pricing for products

On the other hand, for any non-profitable/profitable enterprise, it is really important for a non profit enterprise to turn profitable while any profitable enterprise to continue as a profitable organization.

Additionally, in defence/security/intelligence, today's warfighter have access to an ever-increasing number of sensors, imagers, internet artifacts, open source and other sophisticated collection devices and mechanisms. This has created a challenge in sifting through massive amount of information to find the most critical and actionable items of intelligence. Increasingly, this must be accomplished in near-real time and the information must be packaged in a format capable of being shared with all other pertinent parties.




	<p>Healthcare researchers can mine the data to see what treatments are most effective for different conditions, identify patterns related to drug, and gain other important information that can help patients and reduce costs</p>		<p>Security warfighter have access to an ever-increasing number of sophisticated collection devices that a major challenge has become how to sift through this massive amount of information to find the most critical items of intelligence.</p>
	<p>With the rise of more and more online education and the development, all the data gets a completely new meaning. Big data allow for very exciting changes in the educational field that will revolutionize the way students learn and teachers teach</p>		<p>The Agriculture industry, which has always focused on traditionally farming are transforming to the era of digital enhanced agriculture where analysis needs to be done prior to seeding to reap the best result.</p>

Exhibit 1 – The Changing Models in Big Data

Staying Relevant with Big Data



IT providers will need to play an industry-defining role by transforming into a 'master of big data game' to stay relevant in the emerging industry context. Providers are not dealing with just discrete technologies or one monolithic industry.

Enterprises today are seeking business transformation capabilities from IT providers by adopting big data for:

- **Risk Management** - Inaccurate risk assessment can lead to poor decision making, high costs and scrutiny from regulators. With so much data to analyze, companies need a systemic approach to effectively identify and assess all of the risk exposures, known and unknown, that their business faces. Organizations must identify every risk, from threats to the company's reputation to data breaches and risk of non-compliance with regulations, and weigh them against business opportunities. By analyzing with big data, they can strike a balance between risk and opportunity.
- **Driving R&D** - Big Data for R&D is less about velocity and more about variety, viability and sometimes volume. The key analytics capability for data is the ability to visualize relationships and patterns. By combining real-world outcomes data with clinical data and through the mining of genetic data and a broader understanding of regional and population data, analytically savvy organizations can begin to recognize research failures faster, design more efficient trials and speed the discovery and approval of new innovation while lowering costs along the way.
- **Analyzing Consumer Behavior** - The emergence of internet, e-commerce and social media has radically altered the landscape of available consumer behavior data. Cash registers and Point-of-Sale (POS) systems are being replaced by e-commerce sites that record every move consumers make, even when they don't buy something. Casual

telephone conversations about recent purchases are being replaced by tweets that can be scanned and analyzed by anyone who follows those Twitter feeds. All of this data on actual consumer behavior and experiences is there to be measured and analyzed.

In a nutshell, the ability to capture, translate and leverage increasing amounts of information, from consumer and shopper data, customer or retailer data, social media and the real-time visibility into the demand and supply chains is critical. Investing in the necessary tools and analyzing business-critical information can help develop a deeper level of understanding of customers and products, enhance brand loyalty, increase sales and drive competitive advantage. As such, organizations must equip themselves with the ability to store and quickly process these massive volumes of data with the appropriate technology to develop actionable insights.

Tools for Analysing Business

Businesses struggle with huge quantities and varieties of data and ever-faster expectations for analysis. Here are some of the data analytics tools which could help enterprises analyze their businesses based on their use cases: -

- 1. Apache Hadoop:** A nine-year-old most popular open-source data-processing platform first used by internet giants including Yahoo and Facebook, leads the big data revolution. Cloudera introduced commercial support for enterprises in 2008, and MapR and Hortonworks piled on in 2009 and 2011, respectively. Most of the companies have their own flavor of Hadoop. Hadoop is a popular tool for organizing the racks of servers, and NoSQL databases are popular tools for storing data on these racks. These mechanism can be much more powerful than the old single machine, but far from being as polished as the old database servers. Hadoop jobs are written in Java which requires another level of sophistication. The tools for tackling big data are just beginning to package this distributed computing power in a way that's easier to use.
- 2. Jaspersoft BI Suite** - The Jaspersoft package is one of the open source leaders for producing reports from database columns. The software is well-polished and already installed in many businesses turning SQL tables into PDFs that everyone can scrutinize at meetings. The company is leveraging big data, that is, adding a software layer to connect its report generating software to the places where big data gets stored. The JasperReports Server now offers software to suck up data from many of the major storage platforms, including MongoDB, Cassandra, Redis, Riak, CouchDB and Neo4j. Hadoop is also well-represented with JasperReports providing a Hive connector to reach

the inside of HBase. JasperSoft is expanding by making it easier to use these sophisticated reports with newer sources of data. It isn't offering particularly new ways to look at the data, just more sophisticated ways to access data stored in new locations.

3. **Pentaho Business Analytics** - Pentaho is another software platform that began as a report generating engine; it is, like JasperSoft, branching into big data by making it easier to absorb information from the new sources. We can hook up Pentaho's tool to many of the most popular NoSQL databases such as MongoDB and Cassandra. Once the databases are connected, we can drag and drop the columns into views and reports as if the information came from SQL databases.
4. **Karmasphere Studio and Analyst** - Many of the big data tools did not begin as reporting tools. Karmasphere Studio, for instance, is a set of plug-ins built on top of Eclipse. It 'i's a specialized IDE that makes it easier to create and run Hadoop jobs. There are a number of stages in the life of a Hadoop job, and Karmasphere's tools walk us through each step, showing partial results along the way. As we set up the workflow, the tools display the state of test data at each step. We see what the temporary data will look like as it is cut apart, analyzed, then reduced.
5. **Talend Open Studio** - Talend also offers an Eclipse-based IDE for stringing together data processing jobs with Hadoop. Its tools are designed to help with data integration, data quality and data management, all with subroutines tuned into these jobs. Talend Studio allows us to build up jobs by dragging and dropping little icons onto a canvas. If we want to get an RSS feed, Talend's component will fetch the RSS and add proxying if necessary. There are dozens of components for gathering information and dozens more for doing things like a "fuzzy match." Then we can output the results.
6. **Skytree Server** - Not all tools are designed to make it easier to string together code with visual mechanisms. Skytree offers a bundle that performs many sophisticated machine-learning algorithms. You just need to type the right command into the command line. Skytree is more focused on the guts than the shiny GUI. Skytree Server is optimized to run a number of classic machine-learning algorithms on data, using an implementation the company claims can be 10,000 times faster than other packages. It can search through data looking for clusters of mathematically similar items, then invert this to identify outliers that may be problems, opportunities or both. Algorithms can be more precise than humans and can search through vast quantities of data looking for the entries that are a bit out of the ordinary.

7. **Tableau Desktop and Server** - Tableau Desktop is a visualization tool that makes it easy to look at data in new ways, then slice it up and look at it in a different way. While many of the other reporting tools are built on a tradition of generating reports offline, Tableau wants to offer an interactive mechanism so that we can slice and dice data again and again. Caching helps deal with some of the latency issues of a Hadoop cluster.

8. **Splunk** - Splunk is different from the other options. It is not exactly a report-generating tool or a collection of AI routines, although it accomplishes much of that along the way. It creates an index of our data as if our data were a book or a block of text. Databases also build indices, but Splunk's approach is much closer to a text search process. This indexing is surprisingly flexible. Splunk comes already tuned to particular application, making sense of log files. It is also sold in a number of different solution packages, including one for monitoring a Microsoft Exchange server and another for detecting Web attacks. The index helps correlate the data in these and several other common server-side scenarios.

Conclusion

We believe enterprises will dive into their big data and the approaches would be very different from the kind we see today. You can expect to see a much more real time analytics-focused and agile set of organizations, as they leverage big data in business and information technology. Traditional businesses will not change, but the convergence of innovative technologies to change business process and models will bring in greater agility. The new business model based on the mega trends will be the key differentiator between competing firms. Therefore, firms must prepare themselves today to remain relevant tomorrow.