HARNESSING BIG DATA VOLUMES

PhD Candidate Bogdan DINU[1], PhD Stefan IOVAN[2,3]
[1]ORACLE SA, Bucharest, ask_rules_req@yahoo.com
[2]West University of Timisoara, Computer Science Department, stefan.iovan@infofer.ro
[3]Railway Informatics SA, Bucharest, stefan.iovan@infofer.ro

Abstract: Big Data can revolutionize humanity. Hidden within the huge amounts and variety of the data we are creating we may find information, facts, social insights and benchmarks that were once virtually impossible to find or were simply inexistent. Large volumes of data allow organizations to tap in real time the full potential of all the internal or external information they possess. Big data calls for quick decisions and innovative ways to assist customers and the society as a whole. Big data platforms and product portfolio will help customers harness to the full the value of big data volumes. This paper deals with technical and technological issues related to handling big data volumes in the Big Data environment.

Key words: big data, business intelligence, analytics, cloud computing, skills

1. FOREWORD

Each day we create 2.5 quintillion bytes of data generated by a variety of sources – from climate information, to posts on social media sites, and from purchase transaction records to healthcare medical images. We believe that *big data* and *analytics* are a catalyst to help clients become more competitive and drive growth. Companies are helping clients harness this big data to uncover valuable insights and transform their business. A comprehensive portfolio of big data and analytics technologies and solutions, spanning services, software, research and hardware [1, 2], is now available on the market.

The analysis [3] was conducted across nearly 1,000 business and IT executives from 70 countries. Survey topics included executive activities, business processes, data management practices, human resource management, and competency measures software usage and hardware implementation. The company conducted this survey using a global team of business experts, consultants, data scientists and statisticians.

While more than 75% of the highest performing organizations cite growth or innovation as the top value of *business analytics*, almost two-thirds of them say that political or executive resistance is the primary barrier to realizing the full value of their investments.

The fifth global examination of the adoption and usage of analytics has found that a constantly growing number of respondents (more than two-thirds) are applying business analytics to support revenue-generation strategies in the detriment of cost containment.

Additionally, the study [3] has found that almost 40 percent of companies see a rapid return on investment (ROI) within the first six months from adoption of analytics systems.

The study has also revealed three essential factors supporting the implementation of business analytics within organizations - sponsorship, trust and skills – as well as some significant gaps in each of these areas. By addressing these gaps, analytics leaders will have an opportunity to increase executive advocacy, strengthen trust across the organization and build sounder knowledge and skills.

250

Fiabilitate si Durabilitate - Fiability & Durability Supplement No 1/ 2014
Editura "Academica Brâncuşi" , Târgu Jiu, ISSN 1844 – 640X

In order to harness the insights generated by data, organizations need to identify different successful leaders and get them to fully sustain the use of analytics. Emerging roles like Chief Data Officer (CDO) and Chief Analytics Officer (CAO) are helping companies build a data strategy across the entire enterprise and gain competitive advantage. The right alignment of strategy, standards, technology and organizational structure is required in order to turn to good account the full potential offered by business analytics.

1.1. Executive Advocates

Today only a small percentage of leaders around the world are strongly advocating the use of Big Data and analytics to speed-up decision-making and institutional change. The study [3] reveals that one-quarter of all Chief Executive Officers (CEO) and Chief Operating Officers (COO) acts as the lead advocates for the use of analytic insights (24 percent). While this represents, nevertheless, a 10 percent increase versus 2012, we are still far from the global leadership required to trigger change. Besides the CEOs and COOs, other executives such as the CIO and CMO have the power to make decision as to the implementation and use of big data and analytics across the enterprise.

Executive sponsorship and involvement in big data and analytics are crucial to value creation. Analytics implementations in organizations with low levels of executive support are hampered by lack of funding, resources and follow-through. As organizations wish to grow, there is a need to identify advocates inside the organization, who are willing to help with implementation of projects [4].

This draws attention to the fact that executive sponsors play a critical role in building a data-driven culture. In the future, more companies will think of appointing their CDO or their CAO to help the company analyze and drive results, build an analytics-driven corporate culture and ultimately drive its business success.

1.2. Lack of Trust

Most organizations today prioritize their investments and their decisions regarding data at the business-unit level. A lack of trust within organizations proves to be one of the most significant hurdles to deriving value. While about 60 percent of top professionals enjoy high levels of trust among individuals within their organization, the level of trust drops significantly (47 percent) when it is anonymized between business units and IT departments in general.

A fragmented approach can result in a lack of trust among different groups of people who may be accessing, interpreting and using data in different ways. This trust gap springs from an underlying distrust as to who is actually qualified and competent to analyze and act upon the data. A lack of trust among executives, analysts and data managers can significantly deter people from sharing data, relying on insights and working together to generate value.

The study [3] has found that a trust gap among individuals leads to a lack of trust in the veracity of data. When this happens, the overall costs to the business are high. For example, a major global banking institution decided to duplicate its data collection efforts from two different sources, because banking executives did not trust the work of another team from a different geographic area, which led them to doubt the authenticity of the data.

Nationwide Insurance in Columbus, Ohio, is an example of a company that relies on enhanced communication and education to bridge the trust gap. By empowering the organization to build reliable relationships, the enterprise has accelerated its big data and analytics objectives and has broken down trust barriers by an increased focus on education and personal interaction, so that everyone across the company understands and interprets data in a consistent manner. Trust is a must-have – trusting both the data and individuals – in any organization that is keen to act quicker on the insights generated from their data.

1.3. Growing Skill Gap
Another factor creating a barrier to the success of big data and analytics deals with skills. The gap between the global demand for Big Data and analytics talent and the local supply of talent is one of the major obstacles to analytics implementations across all organizations. One-third of respondents cited the lack of skills in analyzing and interpreting data and converting it into relevant business actions to be the greatest business challenge hampering better value analytics within their organizations.

The widest skills gap is the ability to combine analytics skills with business knowledge in order to draw meaningful insights from the data. The analyst who is capable to understand the business and also to perform higher mathematic tasks is the most coveted in the market. More than one-third of all respondents (36 percent) cited this as the main cause of the skills gap inside their organization, followed by analytic skills (24 percent), data management skills (21 percent) and business skills (19 percent).

2. LEVERAGING BIG DATA
HP has announced the expansion of its portfolio of solutions for big data volumes, designed to help companies gain valuable information from their own data and deliver results in real time. According to a research commissioned on behalf of HP, in 2013 nearly 60% of the companies surveyed will spend at least 10 per cent of their IT innovation budget on big data. The research also reveals that big data initiatives have failed in over one third of the companies surveyed.

Expanded company portfolios are now delivering services and solutions designed to facilitate the successful implementation of such initiatives and to enable organizations to effectively manage the ever growing volume, variety, velocity and vulnerability of data, which can cause failure of big data initiatives.

HP launched the *HAVEn* solution. HAVEn is a big data analytics platform, which leverages HP's analytics software, hardware and services to create the next generation of big data-ready analytics applications and solutions.

2.1. Unifying Analytic Insights for Big Data
HAVEn combines established technologies from *HP Autonomy* and *HP Vertica* with key industry initiatives such as Hadoop, enabling clients and partners to:
- Avoid vendor lock-in by providing an open architecture that supports a broad range of analytics tools.

- Protect investments with support for multiple virtualization technologies.
- Speed time-to-value with highly optimized hardware solutions.
- Gain value from 100 percent of information, including structured, semistructured and unstructured data, via HP's portfolio of more than 700 connectors into HAVEn.

HP has also expanded its big data portfolio with new capabilities that help clients build and operate their big data solutions.

HP Operations Analytics is the first integrated big data analytics solution built on HAVEn, which delivers insight into all aspects of IT operations. The solution allows organizations to effectively ingest, manage, and analyze massive streams of IT operational data from a variety of HP products as well as from third-party sources.

These solutions enable clients to implement analytics and extract the insights hidden within big data, as well as to streamline their key organizational processes such as customer offers, procurement, and supply chain and inventory operations [4]. HP has also expanded its solution to help companies implement successfully big data solutions and ensure optimal IT infrastructure performance as well as support for increasing big data demands.

3. BIG DATA SECURITY CHALLENGES

Security challenges associated with big data volumes are involving both mobile devices and cloud computing platforms. In an attempt to reduce the biggest security inhibitors organizations are facing in implementing cloud, mobile and big data initiatives, a broad set of security software has been released, to help them to holistically secure data and identities.

The new software capabilities provided by these solutions help clients to better maintain security control over mobile devices, mitigate internal and external threats, reduce security risks in cloud environments, and extend database security to gain real-time insights into big data environments and to automate compliance and data security management. This set of scalable capabilities supports a holistic, proactive approach to security threats encompassing people, data, applications and infrastructure.

A major shift is now taking place in the way organizations are protecting data. Today, data resides everywhere – in mobile devices, in the cloud and on social media platforms. This creates massive amounts of data, forcing organizations to move beyond a traditional siloed perimeter to a multi-perimeter approach in which security intelligence is applied in a more target-dedicated manner.

According to the 2012 survey [5], experienced executives have identified some IT system risks that are a major concern: from data thieves to the use of emerging technologies including cloud systems, mobile devices and social media platforms.

Today, there are tens of new products and enhancements available on the market, conceived to help organizations deliver real time security for big data, mobile and cloud computing.

3.1. Real Time Security for Big Data

As information is steadily growing in volume, variety and velocity, organizations are looking beyond relational data sources in their search for relevant insights and their attempt to make businesses more agile and find answers to questions they were once incapable to grasp. Today, state-of-the-art technologies including Hadoop based environments have opened the door to a world of possibilities.

As organizations are ingesting more and more data, they are faced with a significant risk across a complex threat environment, while they are at the same time required to abide by a great number of compliance regulations. Traditional approaches to data protection are often unable to meet these requirements.

IBM is amongst the first companies to offer data security solutions for Hadoop and other big data storage environments. These solutions are now providing real time monitoring and automated compliance reporting for Hadoop based systems. With controls across data sources, clients can now understand data and application access patterns and hence become savvier in preventing data leakage and enforce data change controls.

Built-in audit reporting can be used to generate compliance reports on a scheduled basis, to distribute them to oversight teams for electronic sign-offs and to document the results of remedial actions.

Current data security solutions provide organizations with tools that allow them to automate their vulnerability detection systems and suggest priority remedial actions in heterogeneous IT infrastructures. In addition, they offer data masking to de-identify sensitive data as it moves into and out of big data systems [6].

In addition to securing data in multi-perimeter environments, IBM is announcing improvements to data encryption management that allows organizations to automate key recovery and support the latest version of the *Key Management Interoperability Protocol* (KMIP).

3.2. Relevant Data Security Solutions

Mobile security framework improves access and threat protection. Some companies have announced risk-based authentication control for mobile users, integration of access management into mobile application development and deployment, as well as an enhanced mobile device control.

With the launch of new access management capabilities, a greater context-aware access control is now offered to mobile users, coupled with improved mobile threat protection and enhanced mobile device control. With a broader portfolio of mobile security and management solutions – including solutions for mobile application security and mobile security intelligence – security companies are able to help with protection against security breaches, whether malicious or unintentional – anytime, anywhere and from any device.

Furthermore, due to the simplicity of these mobile devices, which makes them pervasively and seamlessly integrated into everyday life, new threats are stemming from mobile-based activities such as retail purchases, bank accounts management and social networks updating.

The ubiquitous nature of mobility across both businesses and consumers requires that securing the smartphone cover the device, the network and the applications on the device, so that employees, consumers and partners may know that their transactions are executed across a secure environment.

3.3. Mobile Device Security Solutions

Companies are transforming cloud security from an inhibitor to an enabler. While the cloud can increase productivity by providing access to information anywhere, anytime, it can nevertheless present additional challenges for enterprise security.

To realize the value of cloud computing, organizations are looking for integrated security solutions to help address risks.

Security portfolio enhancements are announced, designed to address the new challenges and provide improved visibility and enhanced levels of automation and patch management, and to help demonstrate compliance, prevent unauthorized access and safeguard against the latest threats using advanced security intelligence.

With the new patch management solution, patches are now managed automatically at any location and remediation cycles are diminished from weeks to hours, thereby reducing security risks. For example, IBM is announcing enhancements to its *QRadar Security Intelligence Platform* that provides a unified architecture for collecting, storing, analyzing and querying data related to logs, threats, vulnerability and security from distributed locations, using cloud technology to obtain a wider insight into enterprise activity and enable informed business decisions-making.

The new system is designed to proactively address the growing threat concerns and to help prove compliance across the organization. The system facilitates sign-on to cloud applications and is now available with improved integration with adopted SaaS (*Software as a Service*) applications and services.

4. CONCLUSIONS

For IT professionals with Big Data skills the number of job opportunities has increased. There are now more Big Data management and analysis job opportunities than a year ago and many IT professionals are ready to provide training in this area, as the survey [5, 7] reveals. Forty-eight per cent of respondents said they saw more job opportunities in Big Data environment, and sixty-three percent of them are willing to invest time and money to acquire big data and analysis skills. According to the survey [5], the market of Big Data technologies and services is set to increase from $3.2 billion in 2010 to $16.9 billion in 2015.

With data volumes constantly on the rise, it is vital for countries like UK, for example, to have the skilled workforce to manage the huge data volumes. To meet the demand for Big Data expertise, recruiters should make sure that the professionals they are hunting for are qualified in managing big data volumes. Given the increase in the market demand for professions qualified in data management and data analysis and current the shortage of knowledgeable and skilled professionals in this field, it is vital for recruitment companies to understand the needs of their clients, so that they may be able to supply them the candidates they are looking for [7, 8].

Skills deficit risk has increased in the context of an increasing demand for Big Data expertise. The new technologies are calling for implementation of Big Data solutions and are generating a higher demand for professionals trained in managing this type of information. Despite predictions that Big Data will create about 4.4 million jobs by the year 2015, chances are that only a third of these jobs openings will eventually be filled.

Big quantities of data come in many forms. New technologies are now available to download, store, process and visualize this data in real time. According to a report [8] drawn up by software provider SAS – job openings in the Big Data market are expected to increase by 92 % by the year 2017.

5. BIBLIOGRAPHY
[1] **Iovan, St. and Ionescu, P-V.** (2011) *Cloud Computing: A Short Introduction*, Proc. of 12th European Conference (***E_COMM_LINE 2011***), Bucuresti, Romania;
[2] **Iovan, St. and Ionescu, P-V.** (2012) *Security Issues in Cloud Computing Technology,* Proc. of 13th European Conference (***E_COMM_LINE 2012***), Bucuresti, Romania;
[3] **IBM Report**, Analytics: A Blueprint for Value, *IBM Institute for Business Value,* [accessed in November 2013];
[4] **Iovan, St. and Litra, M.** (2012) *A New Challenge: Large Volumes of Unstructured Data*, Proc. of 13th European Conference (***E_COMM_LINE 2012***), Bucuresti, Romania;
[5] **IBM Report**, *IBM Global Reputational Risk and IT,* [accessed in November 2013];
[6] **Iovan, St. and Daian, Gh. I.** (2012) *New Challenges: "Big Data" and "Consumer Intelligence" / Noi provocari: "Big Data" si "Consumer Intelligence"*, Annals of the "Constantin Brancusi" University of Targu Jiu, Engineering Series, Issue **4**/2012, (***CONFERENG 2012***), pag. 318-329;
[7] **Iovan, St. and Litra, M.** (2013) *Decision Making in Big Data Era*, Proc. of 14th European Conference (***E_COMM_LINE 2013***), Bucuresti, Romania;
[8] **SAS Report**, *Big Data Analytics: An Assessment of demand for labour and skills, 2012-2017,* [accessed in December 2013];