Big Data Technology Predictions and Trends

Md. Rafeeq  
M.Tech (PhD), Assoc. Prof., 
CMRTC Kandlaikoya-Hyderabad.

Rajeshwar Rao Kodipaka  
Asst. Prof., Dept. of CSE, 
MREC Secunderabad.

Dr. C. Sunil Kumar  
Professor, Dept. of ECM, 
SNIST Ghatkesar, Hyderabad

Abstract – In 2014 saw big data initiatives inside the enterprise increasingly move from test to production. In 2015, big data will push further into the enterprise with even more use cases specifically real-time use cases. We live in a world of data (both at a personal and professional level) and people express themselves through the work they do with it. Inside each organization, users want to be actively engaged with their data; however, they haven’t yet had the technology to do so. In 2015, by having access to Business Intelligence (BI) solutions that allow true self-service, users will move from passively consuming data to actively using it to glean important information. Data is not only big but it is also fast. Data is rapidly being generated from mobile devices, factory sensors, wearables, retail transactions, on-line advertisements and many other sources. The insights from that fast, Big Data disappear as rapidly as they appear so taking action in “business time” is critical. Business time is the speed at which a business must service its customers, and it ranges from sub-seconds (real-time) to minutes and hours.

Keywords – Data Agility, Big Data Analytics, Hadoop-As-a-Service, The Ubiquity of YARN.

I. INTRODUCTION

Data Agility Emerges as a Top Focus, Data agility has been one of the big drivers behind the development of big data technologies, as the processes around legacy databases and data warehouses have proven too slow and inflexible for many business needs. In 2015, Schroeder says data agility will become even more central as organization shift their focus from simply capturing and managing data to actively using it.[2] “The reality is that the tools are still emerging, and the promise of the [Hadoop] platform is not at the level it needs to be for business to rely on it,” says Loconzolo. But the disciplines of big data and analytics are evolving so quickly that businesses need to wade in or risk being left behind. “In the past, emerging technologies might have taken years to mature,” he says. “Now people iterate and drive solutions in a matter of months — or weeks.” So what are the top emerging technologies and trends that should be on your watch list — or in your test lab? Computer world asked IT leaders, consultants and industry analysts to weigh in. Here’s their list. With big data, analysts have not only more data to work with, but also the processing power to handle large numbers of records with many attributes, Hopkins says[1].

Traditional machine learning uses statistical analysis based on a sample of a total data set. “You now have the ability to do very large numbers of records and very large numbers of attributes per record” and that increases predictability, he says. The combination of big data and compute power also lets analysts explore new behavioral data throughout the day, such as websites visited or location[3]. Hopkins calls that “sparse data,” because to find something of interest you must wade through a lot of data that doesn’t matter. “Trying to use traditional machine-learning algorithms against this type of data was computationally impossible. Now we can bring cheap computational power to the problem,” he says. “You formulate problems completely differently when speed and memory cease being critical issues,” Abbott says. “Now you can find which variables are best analytically by thrusting huge computing resources at the problem. It really is a game changer.”[4].

“To enable real-time analysis and predictive modeling out of the same Hadoop core, that’s where the interest is for us,” says Loconzolo. The problem has been speed, with Hadoop taking up to 20 times longer to get questions answered than did more established technologies. So Intuit is testing Apache Spark, a large-scale data processing engine, and its associated SQL query tool, Spark SQL. “Spark has this fast interactive query as well as graph services and streaming capabilities.[4].”

II. ORGANIZATIONS MOVE FROM DATA LAKES TO PROCESSING DATA PLATFORMS

In some ways, 2014 was the year of the data lake (or data hub), an object-based storage repository that stores raw data in its native format - whether structured, unstructured or semi-structured - until it's ready for use. Data lakes have a strong value proposition in that they represent a scalable infrastructure that's economically attractive (with a reduced per- terabyte cost) and extremely agile[6].

Schroeder says that the data lake will continue to evolve in 2015 with the capability to bring multiple compute and execution engines to the data lake to process the data in-place. That's not only more efficient, it creates a single point of governance and a single point of security[7].

C-Level Executives Seeing Big Results From Big Data, A study by Accenture finds that many organizations are putting big data initiatives into production and the vast majority of those that are do are pleased with the results. Between February and early April 2014, Accenture screened more than 4,300 CIOs, COOs, CMOs, CFOs, chief data officers (CDOs), chief analytics officers (CAOs) and other senior technology, data and analytics leaders from companies in 19 countries and across seven industries. Of that number, 36 percent said their company...
Figure 1: Big Data Lake

had not completed and weren't currently pursuing a big data installation and four percent said their organizations were in the process of implementing their first big data project. Of the remaining 2,600 respondents that had applied big data to their businesses, Accenture focused on a sample of 1,007 executives.

"Businesses are at a transition point where instead of just talking about the potential results that can be achieved from big data, they are realizing actual benefits including increasing revenues, a growing base of loyal customers and more efficient operations," adds Narendra Mulani, senior managing director of Accenture Analytics. "They're recognizing that big data is one of the cornerstones of digital transformation." "Today, even the most basic items like water pipes can generate and provide data," adds Mulani. "While the Internet of Things is giving rise to massive sources and quantities of data, new big data technologies are emerging that help uncover crucial business insights from the data. Companies not implementing big data solutions are missing an opportunity to turn their data into an asset that drives business and a competitive advantage."

"In 2015, data lakes will evolve as organizations move from batch to real-time processing and integrate file-based, Hadoop and database engines into their large-scale processing platforms," he says. "In other words, it's not about large-scale storage in a data lake to support bigger queries and reports; the big trend in 2015 will be around the continuous access and processing of events and data in real time to gain constant awareness and take immediate action"[6].

III. HADOOP VENDOR CONSOLIDATION: NEW BUSINESS MODELS EVOLVE.

In early 2013, Intel made a splash with the Introduction of its own Hadoop distribution, saying that it would differentiate itself by taking a ground-up approach in which Hadoop was baked directly into its silicon. But just a year later, Intel ditched its distribution and threw its weight behind Hadoop distribution vendor Cloud era instead. At the time, Intel noted that customers were sitting on the sidelines to see how the Hadoop market would shake out. The number of Hadoop options were muddying the waters. Schroeder believes Hadoop vendor consolidation will continue in 2015 as the also-rans discontinue their distributions and focus elsewhere in the stack[5].

"Hadoop is early in the technology maturity lifecycle with only 10 years passing since the seminal MapReduce white papers were published by Google," he adds. "Hadoop adoption globally and at scale is far beyond any other data platform just 10 years after initial concept. Hadoop is in the innovation phase, so vendors mistakenly adopting "Red Hat for Hadoop" strategies are already exiting the market, most notably Intel and soon EMC Pivotal"[6].

Schroeder believes 2015 will see the evolution of a new, more nuanced model of OSS that combines deep innovation with community development. "The open source community is paramount for establishing standards and consensus," he says. "Competition is the accelerant transforming Hadoop from what started as a batch analytics processor to a full-featured data platform."

"The lines between IT and marketing are blurring in the age of digital marketing. For that reason, it's more important than ever that CIOs and CMO communicate consistently and effectively. Yet a strong partnership between IT and marketing is crucial if big data and analytics are to succeed. CMOs may increasingly take the lead on big data projects, but CIOs are the key to implementing, maintaining and scaling these solutions. "Marketing is the driver of the big data car," says Todd Merry, CMO of global hospitality and food service company Delaware North, “but it doesn't go anywhere without IT."[4]" Mainstream.

Previously, IT would be required to establish centralized data structure “Hadoop has made the enterprise comfortable with structure-on-read for some use cases. Advanced organizations will move to data bindings on execution and away from a central structure to fulfill
ongoing requirements. This self-service speeds organizations in their ability to leverage new data sources and respond to opportunities and threats.”

More Distributions, More Competition:
The large enterprise vendors such as HP, Oracle, SAP, Software AG, and Tibco will create their own Hadoop distribution, as IBM and Pivotal already have. Specialized public and private cloud distributions also will emerged, generating plenty of competition and customer choice[7].

As the presence of the Internet of Things (IoT) — such as connected devices, sensors and smart machines — grows, the ability of things to generate new types of real-time information and to actively participate in an industry’s value stream will also grow. Essentially, things become agents for themselves, for people and for businesses. Think of the car that alerts emergency services and an insurance company or the smart thermostat that schedules service. The added connectivity, communications and intelligence of things will make many of them agents for services that are currently requested and delivered via human intervention [5].

**IV. ENTERPRISE ARCHITECTS SEPARATE THE BIG HYPE FROM BIG DATA**

2015 will see enterprise architects take center stage as their improving understanding of the Hadoop technology stack leads to a better defined and more sophisticated statement of requirements for big data applications, including elements like high availability and business continuity. "As organizations move quickly beyond experimentation to serious adoption in the data center, enterprise architects move front and center into the big data adoption path," Schroeder says[8]. "IT leaders will be vital in determining the underlying architectures required to meet SLAs, deliver high availability, business continuity and meet mission-critical needs."

In 2014 the booming ecosystem around Hadoop was celebrated with a proliferation of applications, tools and components. In 2015 the market will concentrate on the differences across platforms and the architecture required to integrate Hadoop into the data center and deliver business results.

Security is a major concern with big data. To make more sense from the big data, organizations would need to start integrating parts of their sensitive data into the bigger data. To do this, companies would need to start establishing security policies which are self-configurable: these policies must leverage existing trust relationships, and promote data and resource sharing within the organizations, while ensuring that data analytics are optimized and not limited because of such policies.

What good is a prediction if its not based on data? When I look at our own internal data and the larger big data market in 2014, there are a few interesting trends that emerged that I believe are strong indicators for what will come in 2015. For an in-depth explanation of 2014 trends and my 2015 predictions. Big Data will be a business initiative. Historically, what we saw was that IT tried to service business, and built the infrastructure for them, but its clear now that business users themselves want to get their hands on big data. It’s no longer that business sends IT off to find analytical results, the line of business wants to do that themselves.

There is a really strong demand for the democratization of data. Hadoop deployments will shift from being centralized in IT toward departmental deployments. We also saw that central, IT-driven deployments are shifting to department specific deployments[5].

**V. SECURITY AND MY DATA-DRIVEN PREDICTIONS**

Big data technologies have evolved at a torrid pace that shows every sign of continuing in 2015. MapR CEO and co-founder John Schroeder predicts five major developments will dominate big data technology in the new Year[4].

**VI. BIG DATA DEEP DIVE CONCLUSION: THE FUTURE OF ANALYTICS**

As data volumes have grown, and as the complexity of data that is collected and analyzed has increased, new novel software architectures have emerged. What I’m most excited is the fact that software is open-source, and...
we’re playing a key role in driving where that software is going. And what I’m most excited about on top of that is the commoditization of that software – I’m tired of talking about the container in which you put your data,” Hammerbacher said. “No one should be required to invest tremendous amounts of money in their container anymore. They should be identifying novel data sources, algorithms to manipulate that data and the smartest people for using that data. Big Data technologies are changing all that: mathematical analysis makes it possible to weed out bad info in a highly efficient manner.

ACKNOWLEDGEMENT

I would like to extend thanks to the Director Dr. A. Raji Reddy, Chairman Mr. C. Gopal Reddy, CMR Technical Campus for encouragement of research work. and I thank to Dr. C. Sunil Kumar Professor, for his valuable suggestions and supervision.

REFERENCES


AUTHOR'S PROFILE

Md. Rafeeq
Doing external PhD in JNTUH- Hyd, Telangana, India. This work will be supported in doing research work as well as understanding the various predictions for big data. I would like thank to my guide Dr. Sunil Kumar, and co-guide Dr. Subhash Chandra for constant support for Understanding the subject.

Rajeshwarrao. Kodipaka
(ISTE, CSI Life Member), working as Asst.Prof in CSE, MREC since 1 year, Having 7 years teaching experience and interested domain is Data mining, cloud computing, Computer Networks.