ROLE OF BIG DATA AND PREDICTIVE ANALYTICS FOR BUSINESS AND IT

Boris Zibitsker, PhD
President and CEO BEZNext,

BEZNext, Chicago, USA, bzibitsker@beznext.com

Introduction

Today's fast-paced businesses have to make effective business decisions in real-time. That creates pressure on IT leaders to develop Big Data applications capable to processing large volumes of structured and non-structured data and applying advanced analytics to provide business recommendations in near-real time. In this presentation we will review role of Big Data Predictive Applications, challenges of developing Big Data applications and examples of incorporating performance assurance methodology focusing on satisfying constantly changing business needs.

Low cost of Big Data infrastructure and open source software, ability to process large volumes of data and provide real-time advice attracts businesses and investors.

The biggest drivers of Big Data Analytics are operational technology (OT), the Internet of Things (IoT), Mobil, Social Media and Cloud Computing.

According to industry analysts:

- By 2017 more than $2 billion in online shopping will be performed exclusively by mobile digital assistants and 60% of big data projects will fail to go beyond piloting and experimentation and will be abandoned

- By 2018 data discovery and predictive analytics will converge and 80% of business processes will be reinvented, digitalized or eliminated

- By 2020 a quarter billion connected vehicles will on the road with new in-vehicle services and automated driving capabilities and IoT will unleash inexpensive surveillance $50 billion market. The connected kitchen will contribute at least 15% savings in the food and beverage industry, while leveraging big data analytics.

Gartner's Group famous hype cycle for Big Data shows that applications incorporating Predictive and Prescriptive Analytics did not reach a peak of expectations yet and will mature in 2- 5 and 5-10 years respectfully. The interest in Big Data is growing, because the hardware is cheap and software is free, so why the rate of deployment of Big Data applications is slower than expected? The major causes are

shortage of people having experience with Big Data, difficulties of managing Big Data environment, security and data privacy challenges.

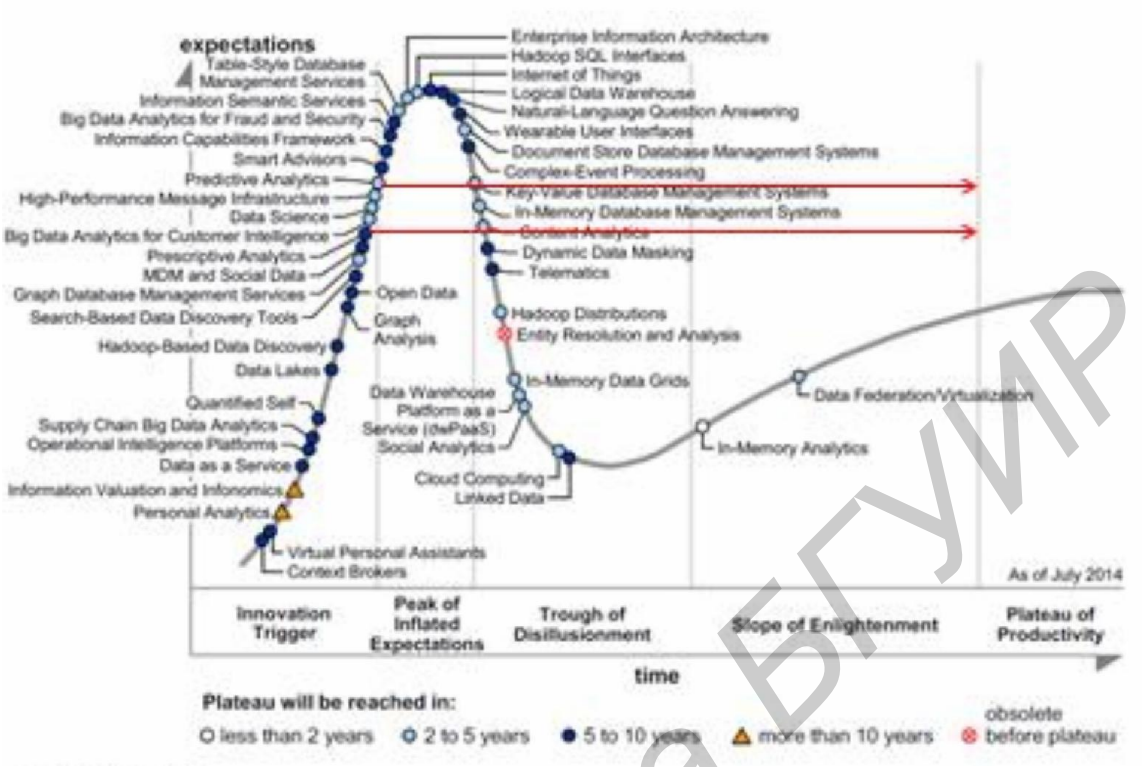

Fig. 1.

What should be done to develop successful Big Data Predictive Analytics applications capable continuously meeting business requirements and service level goals?

Selection of algorithms, infrastructure, and many other decisions made during application life cycle, including design, development, testing, tuning, workload management and capacity planning affects cost, performance and scalability of Big Data applications.

Selection of the Hadoop infrastructure for Batch processing, including the number of Hadoop Mappers and Reducers per workload and Tez parameters require analysis of options to reduce batch elapsed time.

Selection of multi-tier, distributed, virtualized and parallel processing architecture to support real-time Predictive Analytics also require analysis of the tradeoffs and options, workload management and performance optimization.

We will analyze tradeoffs during building infrastructure to support typical Big Data application based on Kafka, Spark or Storm and Casandra Apache projects.

Big Data Analytics Functions

Big Data analytic applications support several functions, including: Descriptive Analytics, Diagnostic Analytics, Predictive Analytics, Prescriptive Analytics and Control Analytics. These functions are implemented using Machine Learning, Artificial Neural Networks, and other Advanced Analytics algorithms.

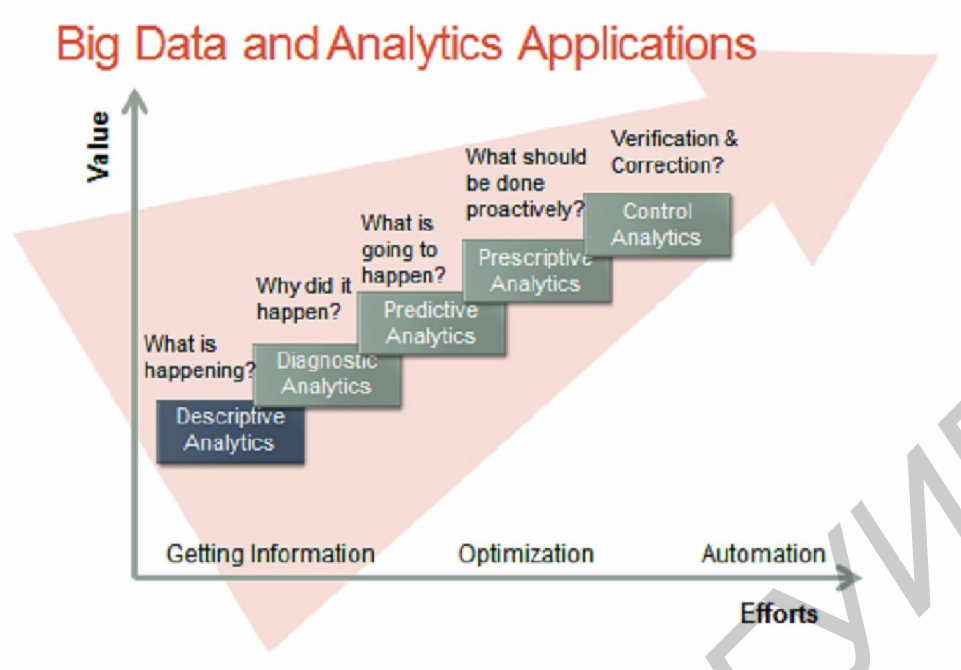

Fig. 2.

Descriptive Analytics

Descriptive Analytics is widely used by businesses and IT to obtain data and information characterizing what is happening now and what happened in the past. ML algorithms are used to find representative clusters of workloads, typical performance profiles and profiles of usage resources and data. Discovery of the seasonal peaks for the individual workloads is used to select representative intervals for further analysis and performance prediction. Information gathered by descriptive analytics is visualized on Dashboards.

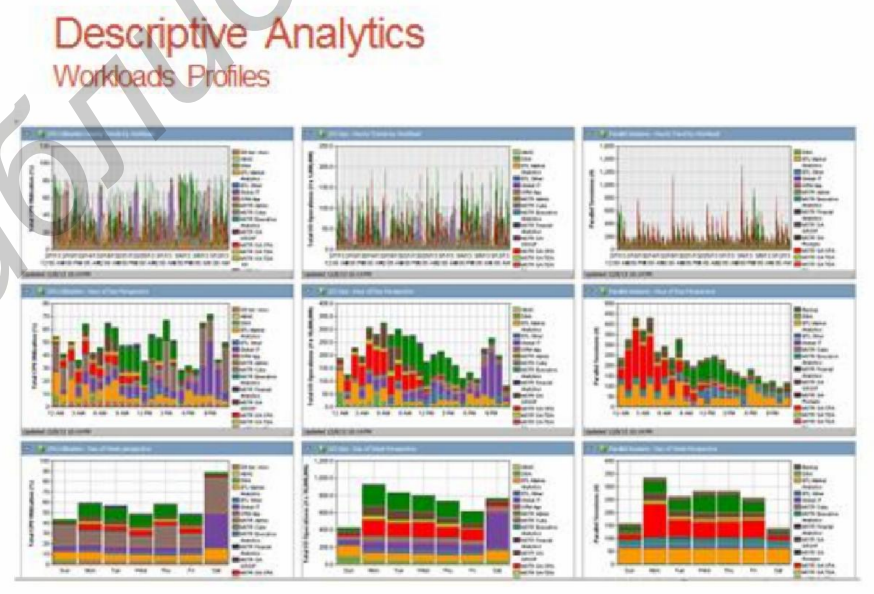

Fig. 3.

Diagnostic Analytics

Diagnostic Analytics: is used to discover unusual changes or increase in usage of resources, change of the pattern accessing data or performance degradation for each of the workload Diagnostic Analytics incorporates functions of Root Cause analysis.

Determining the root cause of the problems narrows down the scope of the tuning efforts, improve the quality and reduce the cost of the management and maintenance. Application of the Predictive Analytics in this case can be used to predict future problems and recommend proactive actions to avoid the problem

Different ML algorithms including Decision Trees and Logistic Regression Analysis are used to identify causes of performance degradation. Predictive analytics is used to identify future causes of performance degradation and generate proactive recommendations.

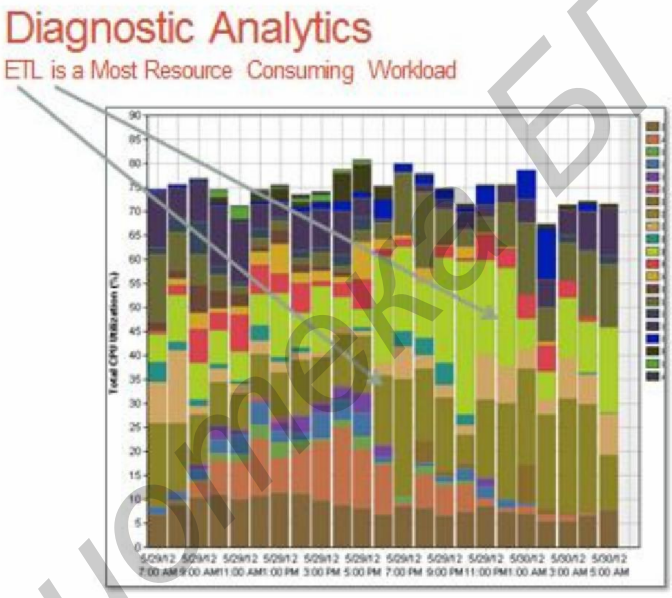

Fig. 4.

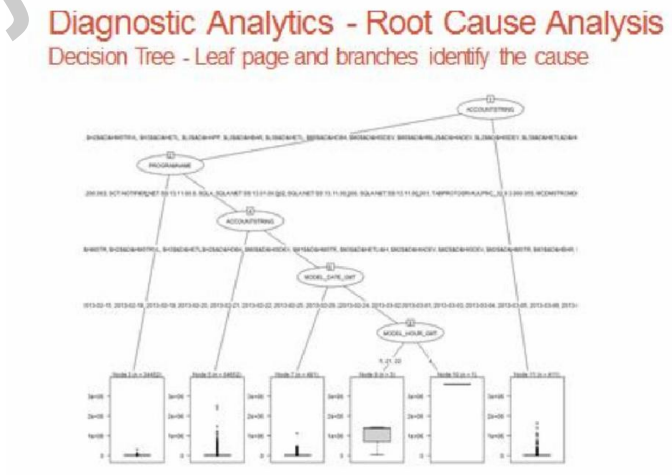

Fig. 5.

Predictive Analytics

Predictive analytics answer what if questions and predict what will happen in a future. It really differentiates Big Data from Business Intelligence, which is focusing on the past and present.

Predictive analytics help to evaluate the impact of expected changes, like workload growth, implementation of new applications, change of infrastructure and other what if questions.

Fig. 6.

Prescriptive Analytics

The objective of Prescriptive Analytics is to evaluate options and recommend what should be done proactively. This is what business people and IT leader's really need. They do not have time to do an analysis. They need actionable recommendations, including what, when and why it should be done. Example below shows the prescriptive set of performance management, workload management and capacity planning actions.

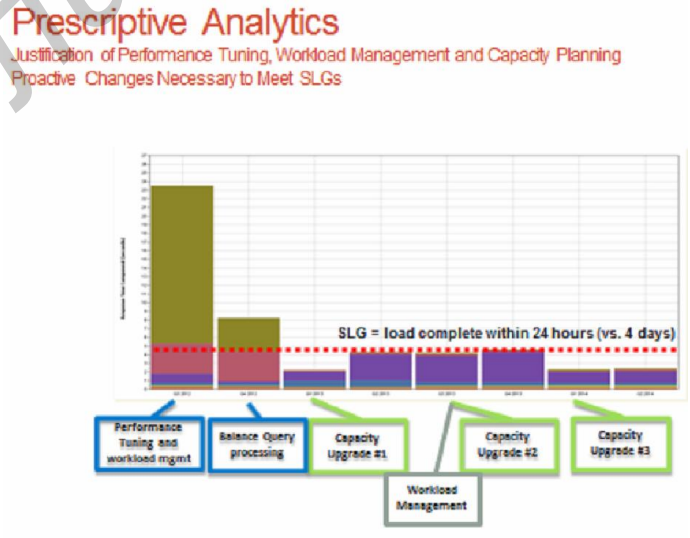

Fig. 7.

Control Analytics

Business people and IT Leaders need expectations prior to making strategic, tactical and operational decisions in order to verify results and justify follow up actions. Comparison of the actual results with expected shows the accuracy of prediction, improves confidence and enables automated control and organization of the continuous management process.

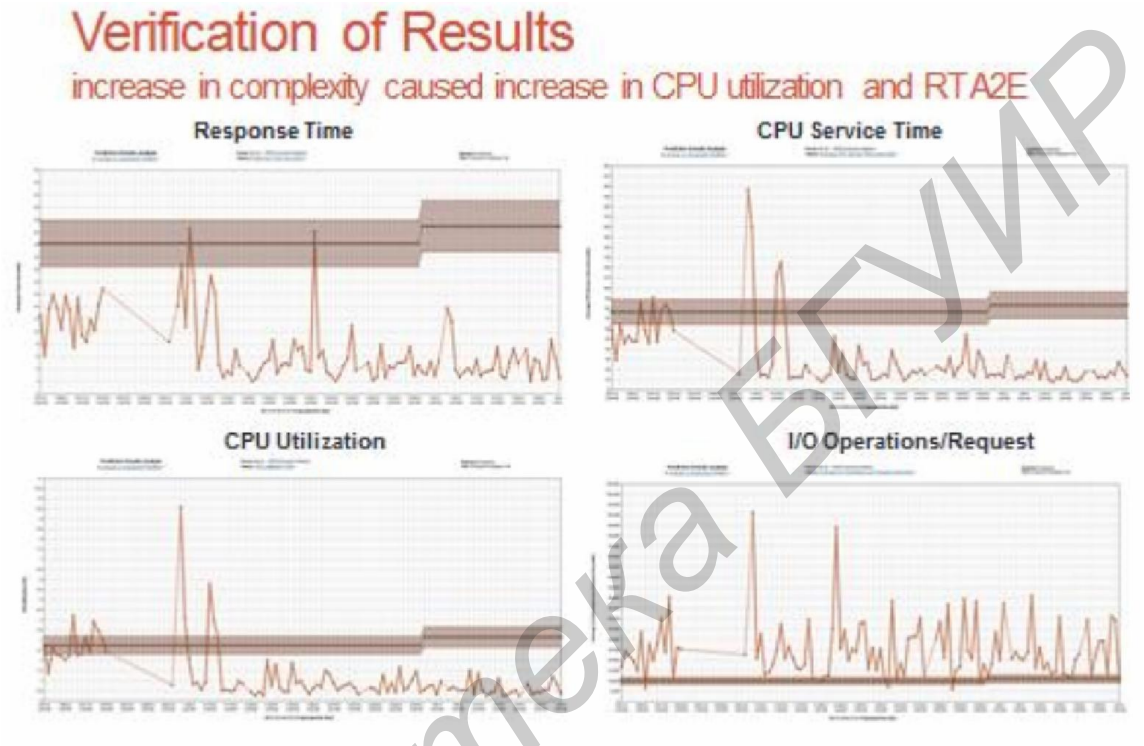

Fig. 8.

Conclusion

Big Data Predictive Analytics is in big demand, but rate of implementation of Big Data applications is slower than expected.

We reviewed the role of Big Data Predictive Analytics for business and IT, challenges of implementations and opportunities in this area.

One of the problems is a shortage of experienced people. Companies looking for Big Data scientists and demand many unique skills:

- Ability to operate analytics technology

- Build predictive models that help business decision-makers chart smarter courses for their organizations

- Be able to understand the business model and culture of the company

- Strategic and creative thinking, the ability to collaborate with colleagues across the business, and strong communication skills to present findings to senior decision-makers in a compelling way

During the workshop we will further discuss architecture, algorithms and challenges of building Big Data Predictive Analytics applications and value of the applying performance assurance methodology during application life cycle.

About the Presenter

Boris Zibitsker is a founder and CEO of BEZNext. He is focusing on applying predictive and prescriptive analytics for performance optimization and performance assurance of Big Data applications.

Prior to that he was Founder, CTO and Chairman of the Board at BEZ Systems (1983 - 2010), acquired by Compuware in 2010, and co-founder of Computer Systems Institute (1989). He managed development of the modeling and optimization software supporting capacity management of mix workloads in multitier, distributed, virtualized, parallel processing systems, including Big Data Clusters, Teradata, Oracle, DB2 and Application servers.

He was working on applying Machine Learning algorithms for performance and availability problems detection, root cause analysis and risk prediction for internet and cloud computing applications.

Boris managed consulting projects for over 200 of Fortune 500 companies and taught seminars in North and South America, Europe, Asia, and Africa.

Since 1983 till 1990 he was an Adjunct Associate Professor at DePaul University in Chicago where he developed and taught graduate courses on Modeling of Computer Systems, Queueing Theory with Computer Applications, Computer Communication Systems Design and Analysis. Boris taught seminars at Northwestern University, University of Chicago and Relational Institute.

He is the author of many papers on modeling, performance prediction, workload characterization, performance management, workload management and capacity planning.

Boris has MS in CS and his PhD research was done at BSUIR and NIIEVM.