Business Intelligence Tools for Big Data

Labhansh Atriwal[1], Parth Nagar[2], Sandeep Tayal[3] and Vasundhra Gupta[4]

[1,2,3,4]CSE Dept., Maharaja Agrasen Institute of Technology, Delhi, India
E-mail: [1]labhanshatriwal@gmail.com, [2]parth2603@gmail.com, [3]tayal.mait@mait.com, [4]vasundhra.gupta401@yahoo.com

Abstract—*This study examines various big-data techniques and technologies and giving a comprehensive comparison between various Business Intelligence tools currently in the market. Big data can mean big opportunities for organizations. Storing large amount of data is easy, but making sense out of it isn't. When we are talking about terabytes and petabytes of information, generated by social networking, sensors, financial transactions, mobile applications and so much more, this is no small task. On the other hand, Business Intelligence (BI), a concept that has been around for decades, allows easy interpretation of large volumes of data; identifying new insights and implementing effective strategies, thus, helping organizations in their long-term decision making and competitive market advantage.*

Keywords*: Big Data, Business Intelligence, Business Intelligence Tools, Software as a Service*

1. INTRODUCTION

Business Intelligence and Big Data have become increasingly important over the past two decades. Although there is a great advance in technology, but the rate at which business data is growing is much higher. According to the 2011 IDC Digital Universe Study, 130 Exabyte of data were created and stored in 2005. The amount grew to 1,227 Exabyte in 2010 and is projected to grow at 45.2% to 7,910 Exabyte in 2015. [1] All of this data is a mine of strategic knowledge which can be used for the betterment of the corporate world. So, there will always be a need for various efficient tools to analyze and monitor such vast datasets.

A. What is Business Intelligence?

Business Intelligence (BI) is a broad category of techniques used to gain strategic insights, by corporates, business analysts and other aspirants, for making future policies and decisions; for long term stability and to have a competitive edge in the market. [2] It includes techniques like reporting, visualization, OLAP, data mining, machine learning, analytics etc. So, as size of data increases over time, there is a need for competitive intelligence in the corporate world, for its better existence. [3]

B. What is Big Data?

Big data are datasets which are in size that is beyond the capacity of a single database to store, manage and analyze. [4] The definition of size is variable and subjective on the technology of that time. As technology advances over time,

the size of dataset to qualify for the big data also increases. Today, big data ranges from terabytes to petabytes, ranging from industry to industry.

C. What are Business Intelligence Tools?

Business intelligence tools (BI tools) are designed with the primary goal to retrieve, transform and monitor an organization's data to gain business intelligence. [5] But, getting the right information is not what makes a BI tool count. Delivering the same in the adequate amount of time is what makes it an ideal BI tool. It is basically a complete package of extracting, transforming and integrating data to produce insights using various techniques like mining, statistics and predictive analysis. [6]

BI tools can range from simple Excel-feed tools to Multidimensional data based tools. But in general, it can be categorized into generalized or big-data based tools that function on structured, semi-structured or unstructured data.

2. BIG-DATA TECHNIQUES

There are various techniques like statistics and sentiment analysis that can be used on big-data for analytics. These techniques [7], not all necessarily, form a part of a business intelligence tool. This section holds a list of such techniques applicable over a range of industries:

Classification

This technique is used to identify the set or category a particular data instance belongs to. Training datasets are used to determine the known sets or categories for classification.

Cluster analysis

It is a method of combining objects into clusters (groups), such that objects in the same group are similar. No training data set is required to ascertain groups.

Crowdsourcing

It is a technique of collaboration of evaluations from a large group of people to solve a problem related to big data, where computations do not work well.

Data fusion and data integration

This technique is used to integrate and analyze large data from different sources, by applying transformation methods, to produce useful outcome.

Data mining

It is a technique of discovering patterns from big data using concepts of artificial intelligence, machine learning and statistics.

Machine learning

It is an algorithm for predicting more accurate results in the form of patterns (Pattern recognition) or models (Predictive modeling); with the capability to learn from training datasets and previously produced results.

Natural language processing (NLP)

NLP provides an efficient way to analyze and derive meaning by processing human-computer interactions.

Optimization

It is a technique of selecting the best or optimal solution from a set of alternative solutions to a problem.

Regression

Regression is form of supervised learning used for establishing a relationship between dependents or outcome variables and the predictors or independent variables.

Sentiment analysis

It is a technique of measuring polarity (positive, negative or neutral) of subjective information contained in natural or human language.

Signal processing

It is a technique to analyze random signals (continuous or discrete) like sensor and radio signals etc., inherent in big data.

Spatial analysis

It is a technique to study data trends using its geographical, topological or geometric properties.

Statistics

This technique involves interpreting data, generally numeric, and its related computations to achieve more accurate analysis.

Simulation

This technique involves modeling real complex systems and to study actions and effects for predicting results for real systems.

Time series analysis

It is a technique of analyzing big data at successive intervals of time to forecast trends along the time axis.

Visualization

It is a tool for interpreting data into charts, diagrams and animations for better understanding and recognition of patterns.

Ad-hoc reporting

It is a reporting technique for non-technical business users, which can produce reports as per their requirement or occasion using simple queries, without much interference and help from technical people.

Dashboard

This technique is used as a graphical representation in various analysis tools pertaining to an organization's current and historic trends or performances.

ETL

Extraction-Transformation-Loading: It means extracting data from different sources, transforming it into a standard format and loading it into a data repository.

3. COMPARISON OF BI TOOLS FOR BIG-DATA

A. Free and Open Source Software (FOSS)

These softwares are freely licensed to be used by public. Their source code is also openly distributed and available for access by anyone. The free use attributes to the fast growth and constant improvisations.

Name of the tool	Big-data techniques included	Compatible Big-data technologies
Eclipse BIRT Project [8]	Statistical analysis, Visualization, Reporting, OLTP, OLAP	JDO Data stores, JFire Scripting Objects, POJOs, SQL, Web Services, XML, MongoDB, Cassandra, Microsoft Office, Hadoop
SpagoBI [9]	Includes Data mining, OLAP, Spatial analysis, Visualization, ad-hoc reporting and multidimensional analysis	SQL, Oracle, JBoss, Tomcat, Teradata, VectorWise, Netezza, Hive HBase, HDFS, Cassandra, OrientDB, MongoDB, Hortonworks, Cloudera, Impala, JasperReports, BIRT

KNIME [10]	Predictive modeling, Machine learning, Data mining, ETL, Visualization, sentimental analysis, Time series analysis, Cluster analysis, Reporting	R Project, RapidMiner, Hadoop via Hive, Web services, Microsoft office

B. Open Source Commercial Software

These softwares too have their source code openly available to users. However, commercial softwares involve monetary fees to ensure security and derivative distribution agreements.

Name of the tool	Big-data techniques included	Compatible Big-data technologies
Jaspersoft [11] (Also has a proprietary version)	Data integration, Visualization, Reporting, OLAP, Dashboard, Analytics	Hadoop, Cassandra, MongoDB, JBoss, SQL, NoSQL, Oracle, Tomcat Java/JavaEE, RTF/ODT, HTML/XML, Microsoft office
Pentaho [12] (Also has a proprietary version)	Data integration, Visualization, OLAP, Dashboard, Data mining, Ad-hoc reporting, ETL, Cluster analysis, Classification, Regression analysis, Analytics, Machine learning, Predictive modeling	Hadoop, Cassandra, MongoDB, Amazon Websevices, BIRT, JBoss SQL, NoSQL, Oracle, Google Analytics, R scripts, Splunk, Tomcat Java/JavaEE, MDX, ZIP XML, Microsoft office

C. Proprietary Free Software

These softwares are free to use but have restrictions placed on modification or updation, analysis and distribution. They are free but non-open source in some way or another.

Name of the tool	Big-data techniques included	Compatible Big-data technologies
Birst [13]	Visualization, Dashborads, Reporting	Excel, Tableau, R, Oracle, SQL, Google Analytics, SAP, Microsoft Services, Marketo, NetSuite, Hadoop
IcCube [14] (Also has a proprietary version)	OLAP, Dashboards, Visualization, Reporting, Modeling, Cluster analysis	RDBMS, Excel, CSV, MongoDB, MDX, JAVA, .NET, XMLA, Hadoop, Google BigTable

InetSoft [15] (Also has a proprietary version)	Visualization, Ad-hoc reporting, Dashboards, OLAP, Analytics, Spatial analysis	Java, XML, SAP, Excel, CSV, Oracle, MDDBs, Hadoop, Cloudera, Spark, SQL, PeopleSoft, Siebel CRM
Tableau [16] (Also has a proprietary version)	Visualization, Dashboards, Reporting, Natural Language Processing, Artificial intelligence, Data mining, Cluster analysis, OLAP, Statistical analysis, predictive analytics	RDBMS, MDDBs, Spreadsheets, Cloud sources, Google analytics, EDWs, NoSQL, Hadoop, Microsoft office
Splunk [17]	Dashboards, Reporting, Visualization, Learning, Analytics	Hadoop, NoSQL, Unix Piping, RDBMS/SQL/Oracle, Cloud data sources, Java/Python/C#/Ruby/PHP

D. Proprietary Software

These softwares are neither free nor open source. Their use, modification or updation, analysis and distribution require license agreements form vendors.

Name of the tool	Big-data techniques included	Compatible Big-data technologies
Domo [18]	Dashboards, Reporting, Visualization, Data integration, ETL, OLAP, Analytics, Predictive modeling, Machine learning	Hadoop, EDWs, RDBMS, SQL, NoSQL, Cloud sources, Microsoft Office
Dundas Data Visualization [19]	Dashboards, Reporting, Visualization, Data integration, OLAP, Statistical Analysis, Predictive modeling, Machine learning	Oracle, .NET, Sharepoint, SAP, SQL, CSV, Google Analytics, Salesforce, Web services, ODBC
IBM Cognos [20]	Dashboards, Reporting, Ad-hoc query, Visualization, OLAP, Analytics, Multidimensional analysis	RDBMS, SQL, MDX, Cloud sources, ODBC, JDBC, Apache Hive, Web services, Hadoop

Information Builders [21]	Dashboards, Reporting, Visualization, Data integration, Analytics, Predictive modeling, ETL	Hadoop, EDWs, RDBMS, SQL, Oracle, NoSQL, Web services, MongoDB, Cloudera, SAP, Teradata, IBM Netezza, Microsoft Office
Jedox [22]	OLAP, ETL, Data visualization, Dashboards, Reporting, Analytics	Salesforce, XML, JDBC, Oracle, SAP, C/Php/Java/.NET, SQL, R scripts, CSV, Excel
JReport [23]	OLAP, Data visualization, Dashboards, Reporting, Analytics	MongoDB, NoSQL, RDBMS/Oracle/SQL, Hadoop, Hive, Cloudera, Web services, Redhat, HTML/XML, CSV, Pdf/Excel
Klipfolio Dashboard [24]	Data visualization, Dashboards, Reporting, Analytics	Cloud Sources, Salesforce, SQL, Oracle, SAP
Lavastorm [25]	Data integration, Ad-hoc reporting, Statistical analysis, Dashboards, Data visualization, ETL, Pattern recognition	MongoDB, Hadoop/Hive, NoSQL, Web services, R/Python, JDBC, ODBC, XML
Logi Analytics [26]	Data visualization, Dashboards, Reporting, Analytics, ETL, Spatial analysis	MongoDB, SQL, Oracle, Salesforce, Excel, Amazon DB, HP Vertica
Looker [27]	ETL, Data visualization, Dashboards, Reporting, Analytics, Data mining	Amazon Redshift, Google BigQuery, HP Vertica, Netezza, Teradata, Hadoop/Cloudera, Impala, EDWs
MicroStrategy [28]	Data visualization, Dashboards, Reporting, Analytics, OLAP, Data mining, Predictive analysis, Data integration, Regression modeling, Simulation, Supervised learning, Clustering, Time series analysis	EDWs, Hadoop, Cloud sources, SAP, Salesforce, NoSQL/MongoDB, SQL, Google Big query, Spark, Hive, Web services, Oracle, Teradata, Cloudera, Excel/CSV

RapidMiner [29]	Data visualization, Dashboards, Reporting, Analytics, Machine learning, Data mining, Predictive analysis, Statistical modeling, ETL	Hadoop, Cloudera/Hive, R scripts, SQL, SPSS, Salesforce, Netezza, Teradata, Oracle, Excel/Access, Web services
Roambi [30]	Data visualization, Dashboards, Reporting, Analytics	Salesforce, Web services, Hive, Hadoop, SAP, Netezza, Excel
SiSense [31]	Data visualization, Dashboards, Reporting, Analytics, Crowdsourcing	Google Analytics, Salesforce, Hadoop, Teradata, Excel/CSV/Access
SAS [32]	Data visualization, Dashboards, Reporting, Statistical Analysis, Optimization, OLAP, Predictive modeling, Data mining, Ad-hoc reporting	Hadoop/Hive, SQL/Oracle, SAP, Oracle, Teradata/GreenPlum, Excel/XML
Spotfire (now Tib co) [33]	Data visualization, Dashboards, Reporting, Analytics	Hadoop, Hive/Hortonworks, Cloudera, Spark, JDBC, ODBC, Excel/Access
TARGIT Business Intelligence [34]	Data visualization, Dashboards, Reporting, Analytics, Data mining	Google bigquery, Cloudera, MongoDB, EDWs, SQL, Oracle, Hortonworks/Hive, Microsoft Analytics, CSV
Yellowfin Business Intelligence [35]	Data visualization, Dashboards, Reporting, Analytics, Ad-hoc analysis, Optimization, Predictive modeling	Hadoop, Hive, SQL, JDBC/ODBC, Oracle, RDBMS, MDDBs, SAP, Amazon RDS, Teradata, Excel
Zoho Reports [36]	Data visualization, Dashboards, Reporting, Analytics	RDBMS, NoSQL, Hadoop, Cassandra, Hortonworks/Hive, MongoDB, SQL, Oracle, Salesforce, Google Analytics, Cloud sources, Amazon RDS, Word/Excel/Access/CSV

4. CONCLUSION

Since business intelligence tools have been around in the industry for the better part of three decades in various forms and names. Hence the market today is full of such products, with proprietaries ruling the industry. But with the advent of digital and social media, big data is being generated at unprecedented rates. The need to analyze such datasets has struck many analysts. It would be interesting to observe what transformations the existing tools take to adapt the new challenges posed by big data. Thus, the focus is shifting towards developing more efficient big-data techniques and adopting free and open sources policies.

REFERENCES

[1] J. Gantz and D. Reinsel, "The 2011 Digital Universe Study : Extracting Value from Chaos," IDC, Sponsored by EMC, Massachusetts, US, 2011.

[2] I. J. D. Arnott and M.G., "Evaluating the intangible benefits of business intelligence: review & research agenda," IFIP TC8/WG8.3 International Conference, Toulouse, France, 2004.

[3] S. Negash, "Business intelligence," *Communications of the Association for Information Systems, vol. XIII, no. 1,* pp. 177-195, 2004.

[4] M. C. a. J. M. Brad Brown, "Are you ready for the era of 'big data'?," *McKinsey Quarterly, vol. IV, no. 1,* pp. 1-12, 2011.

[5] E. Dumbill, "Making Sense of Big Data," *Mary Ann Liebert,Inc Publishers , vol. I, no. 1,* pp. 1-2, 2013.

[6] H. Chen, R. H. L. Chiang and V. C. Storey, "Business Intelligence and Analytics : from Big data to Big impact," *MIS Quarterly vol. 36 no. 4,* pp. 1165-1188, 2012.

[7] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh and A. Hung Byers, "Big data: The next frontier for innovation, competition, and productivity," *McKinsey Global Institute,* pp. 27-36, 2011.

[8] "BIRT Project," 2015. [Online]. Available: http://www.eclipse.org/birt/. [Accessed October 2015].

[9] "SpagoBI," 2015. [Online]. Available: http://www.spagobi.org/homepage/product/big-data/. [Accessed October 2015].

[10] "KNIME," 2015. [Online]. Available: https://www.knime.org/knime. [Accessed October 2015].

[11] "Jaspersoft Community," 2015. [Online]. Available: http://community.jaspersoft.com/wiki/community-wiki. [Accessed October 2015].

[12] "Pentaho," 2015. [Online]. Available: http://www.pentaho.com/resources/ebooks. [Accessed October 2015].

[13] "Why Birst," 2015. [Online]. Available: https://www.birst.com/why-birst/. [Accessed October 2015].

[14] "IcCube," 2015. [Online]. Available: http://www.iccube.com/support/documentation/index.php. [Accessed October 2015].

[15] "InetSoft," 2015. [Online]. Available: https://www.inetsoft.com/. [Accessed October 2015].

[16] "Tableau Software," 2015. [Online]. Available: http://www.tableau.com/learn/whitepapers. [Accessed October 2015].

[17] "Splunk," 2015. [Online]. Available: http://docs.splunk.com/Documentation. [Accessed October 2015].

[18] "Learn Center," 2015. [Online]. Available: https://www.domo.com/learn/whitepaper-big-data-fueled-marketing-intelligence. [Accessed October 2015].

[19] "About Us," 2015. [Online]. Available: http://www.dundas.com/about-us/. [Accessed October 2015].

[20] "Cognos," 2015. [Online]. Available: http://www-01.ibm.com/software/analytics/cognos/. [Accessed October 2015].

[21] "Big Data Analytics," 2012. [Online]. Available: https://www.informationbuilders.com/pdf/factsheets/fs_part_em cgreenplumbigdata_2012.pdf. [Accessed October 2015].

[22] "Jedox," 2015. [Online]. Available: http://knowledgebase.jedox.com/knowledgebase/. [Accessed October 2015].

[23] "Big Data Visualization with JReport," 2015. [Online]. Available: http://www.jinfonet.com/resources/on-demand-webinars/603-big-data-visualization. [Accessed October 2015].

[24] "Klipfolio Dashboard," 2015. [Online]. Available: http://www.klipfolio.com/resources. [Accessed October 2015].

[25] "Lavastorm Analytics," 2015. [Online]. Available: http://www.lavastorm.com/products/analytics-engine/. [Accessed October 2015].

[26] "Logi Analytics," 2015. [Online]. Available: http://www.logianalytics.com/resources/bi-encyclopedia/. [Accessed October 2015].

[27] "Looker Blog," 2015. [Online]. Available: http://www.looker.com/blog/integrating-a-modern-big-data-and-analytics-platform-with-aws-services-looker-and-mortar. [Accessed October 2015].

[28] "MicroStrategy," 2015. [Online]. Available: https://www.microstrategy.com/us/learn/resource-library. [Accessed October 2015].

[29] "RapidMiner," 2015. [Online]. Available: https://rapidminer.com/learning/. [Accessed October 2015].

[30] "Roambi," 2015. [Online]. Available: http://roambi.com/analytics. [Accessed October 2015].

[31] "Sisense," 2015. [Online]. Available: http://www.sisense.com/features/. [Accessed October 2015].

[32] "Big Data Analytics," 2015. [Online]. Available: http://www.sas.com/en_us/insights/analytics/big-data-analytics.html. [Accessed October 2015].

[33] M. O'Connell, "Big Data Analytics: Scaling Up and Out in the Event-Enabled Enterprise," 2011. [Online]. Available: http://spotfire.tibco.com/assets/blt99d668dff27bf703/big-data-analytics.pdf. [Accessed October 2015].

[34] "TARGIT Business Intelligence," 2015. [Online]. Available: http://www.targit.com/en/resources/library?type=ebook&sort=d ate. [Accessed October 2015].

[35] "Yellowfin Business Intelligence," 2015. [Online]. Available: https://www.yellowfinbi.com/YFWebsite-Business-Intelligence-and-Analytics-Platform-24427. [Accessed October 2015].

[36] "Zoho Office Suits," 2015. [Online]. Available: https://www.zoho.com/google-apps/reporting-business-intelligence.html. [Accessed October 2015].