

## Emergence of Big Data in Business Intelligence

Ashna Rejee<sup>1</sup> Triveni Deshmukh<sup>2</sup>

<sup>1,2</sup>Btech CSE, Shri Vaishnav Institute of Information Technology, SVVV, Indore.

### Abstract

In today's digital era, the world is revolving around technologically coveted ideas and techniques which can be transformed into business logics in corporate world. The entire corporate world is performing to maximum extents and striving to stay forward from their competent, they do 'different things, differently'!. (1) The most different thing the companies follow is the Business Intelligence Technology to leave an impact before their competent. Like for eg:- To improve Intel's business intelligence (BI), Intel IT is putting in place the systems and skills for analyzing big data. A major portion of this data consists of the large, unstructured data sets that constitute up to 90 percent of enterprise data. The ability to mine and analyze big data, in any form, from many sources, gives us deeper and richer insights into business patterns and trends, helping drive operational efficiencies and competitive advantage in manufacturing, security, marketing, and IT. (2)

### Introduction

Worldwide, the amount of raw data is growing exponentially, due in part to the explosion of connected devices, Internet services, social media, cameras, sensors, and user-generated content. Moreover, up to 90 percent of corporate data, including documents, web pages, and email, is unstructured. The sheer volume and complexity of data is overwhelming typical database software, and this situation is calling for a new approach. Business intelligence (BI) is an ideology of rendering out the right information or decision to the people requiring them at the right time, to improve business situations and

performance. It is canopy of concepts, which includes various softwares, data structures, big data concepts, algorithms etc. It is like the other data interpretation concepts, which requires data as it's main object and undergoes usual processes such as data collection, storage, processing and extraction of information too. (1)

Developing the organizational skill to mine and process big data to perform predictive and prescriptive analytics will be a key driver of performance in the future, enabling companies to: (2)

- Make better decisions
- Increase business velocity
- Accelerate the pace of innovation
- Discover and tap new markets

Advances in parallel computing now make it possible to handle big data, to the point where it is now becoming standard practice to capture and store information well before its value is completely understood, and tackle many business problems that have been previously too large to handle. (2)

The results of BI have a direct effect on company's business operations, working strategies and tactics as the companies and organisations change their operating methods accordingly and adopt different strategies based upon the decisions of BI. So, it is mandatory for the decisions to be reliable and accurate as they are what going to decide the future of the company and helps the management to act before their competent. (1)

## History

Business intelligence existed before technology.

Today it's understood as a set of analyses that derive value and insight from data.

The concept has been in existence all times through-out, right from a native shop-keeper remembering his frequent customers and the product they ask-for to a larger company producing a range of products that is in demand in the market! (1)

Richard Miller Devens' 1865 work, *Cyclopaedia of Commercial and Business Anecdotes* contains the first known usage of the term "business intelligence."

It was with the 1958 publication of a landmark article on the subject, written by IBM computer scientist Hans Peter Luhn, that the potential of BI was recognized.

IBM's invention of the hard disk in 1956 revolutionized data storage.

A 1988 international conference aimed to streamline data processes. The Multiway Data Analysis consortium, held in Rome, was a landmark in simplifying BI analysis.

BI tools in the present day are often designed with a very specific industry in mind, be it healthcare, law enforcement, or even professional sports. Known as "software verticalization," this growth of industry-specific tools has contributed significantly to increased adoption of business intelligence. (3)

Tangential to the cloud is the rise of mobile-empowered platforms, which allows users to work with BI on-the-go on smartphones, tablets, and other devices.

## Proposed Study

### *Role of Big Data in Business Intelligence*

The cardinal element of business intelligence is data. Big Data points on the volume of both structured and unstructured data collected

from the sources. The size of data relies upon the sources of data considered, the company's establishment in the market, it's short and long-term goals to be achieved, knowing it's customers' need, it's business model etc..

It is simple to explain the big data in just three words, variety, velocity and volume of data. Big data involves in these main activities with data i.e: collection, storage, integration.

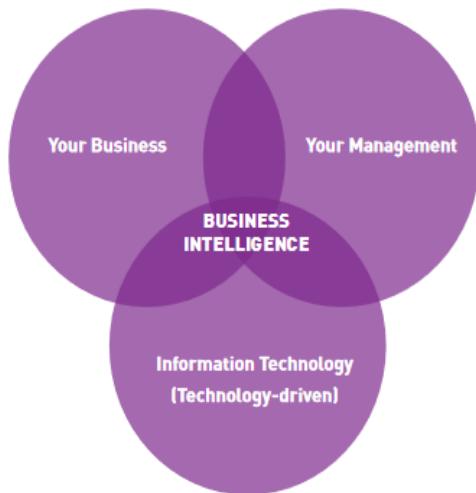
**Collection:** The techniques of collecting the data by the companies have transformed versatility. The feedback method has now become an old technique. Now a day the customer requirements are learnt just by observing them and not even taking the method to their knowledge. Those few tricks are using high sensitive cameras with accurate motion sensing, by tracking the online purchasing method of user, by noticing the online transaction data etc. (1)

**Storage:** As the amount of data collected have become vultuous, the warehousing facility should also be bulged up. As the size of big data is said to range between few dozen terabytes to exabytes, the storage area may be required with double of their sizes predicted. (1)

Few methods of storing big data are *Hadoop and Mapreduce* (they usually do analysis with SAS, Splunk and SAP Hana), *Edge Computing* (the simultaneously generating data has to be stored in an expanding storage space), *Multi-Cloud* (the public cloud platforms offering online computing opportunities), *Storage Intelligence* (this the software's themselves harnessing the required storage space).

**Integration:** This is the joining the closely relevant processed data together, based on their relativity. This stage usually comes across various challenges to be faced. Few are, the information extracted might haven't managed properly, finding the right place to place the data in big data, synchronisation of

data sources, talent lack in handling the data while incorporating them in the technologies and other un-expected miscellaneous challenges. (4)



### Techniques/Algorithms used for Decision making

i) **Hypothesis Testing:** It is a simple idea to analyse to check whether the hypothesis is true or false based upon the data given. With the results of this hypothetical testing, we could either accept or reject the hypothesis. This method is useful in analysing whether an event happened is either by chance or an upcoming trend. It is of two types; T-Test, a statistical test based upon the comparison of two different means or variances, how far they are different from each other. And the other one is Chi-Square Test, to test variability of two distributions of variables. The test formula is,

$$X^2 = \frac{\sum(O-E)^2}{E}$$

Where,  $X^2$  Chi Square obtained

O=observed score

E=expected score

ii)**Random Forest:** Decision trees in collection is called a Random Forest. It is used to classify the new objects based on it's attributes, thus with that each tree is classified. The forest chooses the classification having

the most number of votes. For a better understanding, let us consider a dataset with N sample cases, every case is chosen at random and this set is the training set for growing the tree. If the number of cases is M, all cases are tested without pruning and the tree is grown to the largest extent possible. (4) (1)

iii)**Gradient Boosting & AdaBoost:** When predictions with higher accuracy must be made even though the data load are massive it is suitable to go with these algorithms. It is an ensemble learning algorithm that combines the predictive power of several base evaluators to improve robustness. To achieve high rate of accuracy it is being advised to use along with Python and R. These could also be partnered with other multiple weak or average predictors to build a better predictor. (1)

iv) **Market Basket Analysis(Apriori algorithm):** Apriori algorithm, to draw data insights on which products are likely to be brought together and the things which are most responsive to promotion. For example, a shopkeeper may apply this idea to predict that the people who buy tea and milk are likely to buy sugar and lemon to make tea. This ideology is applied by giants like Amazon, Walmart etc.

v) **Logistic Regression:** It is the method of statistically analyzing a dataset in which one or more independent attributes that determine the outcome. It predicts the outcome that could only have two values, i.e: dichotomy. For example, in predicting win or lose type of events.

$$Z = B_0 + B_1 \times X_1 + B_2 \times X_2$$

### Tools

Algorithms in the form of a software is itself a biggest business as they are themselves the core functionality of a coded product. Few of

those software buzzing in the machines and looting away huge sum are below: (5)

i) **Tableau**: It is considered as the most advanced business software as it connects to various sources of data with its productivity suite. It is said to be the most technological and high speed as it offers 10 to 100 times faster analytics than its competitors. (1)

ii) **Power BI**: The Microsoft's product is the other leadingly adopted analytics software. It offers cloud-based analytics facility too through which the users could build re-usable data models with the existing data. It also gives self-servicing access to the other major third-party cloud sources such as GitHub, Zendesk, Marketo, Salesforce etc.

iii) **Cognos**: The oldest ever business intelligence company in the field since 1969 by IBM. It initiated its services as a consulting company for Canadian federal government. Further it re-newed itself by developing its own web technologies as its competitors were adopting new technologies then. Now it is producing software compatible to both mobile and desktop devices.

iv) **Qlik Sense**: The speciality of this tool is, its capability to integrate automatically the visualizations and respond changes in context from anywhere in the software. It provides user a responsive user experience and intelligently adapts all the functionalities required from any sort of devices with touch.

v) **Si Sense**: This is the first company in the field to provide services with in-memory technology called in-chip analytics. This was

done to maximise disk space, memory and CPU performance. It also launched Prism 10X, which helps in processing the data with 10 more efficiency.

### **Future of Business Intelligence- Embedded Intelligence:**

As everything has its own future, let us look what BI has got to it. It is said to be embedded intelligence. In this softwares are said to have the capabilities of reporting and analytical functionalities. This feature may be present outside the software but must be accessible from the inside without forcing the user to shift between the systems. The addition of these features in a software with business intelligence platforms helps users to choose where in business methods, analytics should be embedded. (1)

### **Architecture of Embedded BI**

This would help the users to make out aptful decisions to increase their revenue and incorporate value to software applications. The elevation in business can't alone be achieved by making out decisions or just by predicting the outcome alone. It is necessary to take actions such as: using the analysis for making out better decisions, efficiently accessing the analysing data and business effectiveness through integration of analysed data with CRM, ERP, sales, marketing and other such functional bodies of the company. (1)



Fig 2: Architecture for Embedded BI

## Bibliography

1. *Mining Big Data in the Enterprise for Better Business Intelligence*. July 2012.
2. *Big Data in BI*. **Balakrishnan, S.**
3. History of Business Intelligence. <https://www.betterbuys.com/bi/history-of-business-intelligence/>. [Online]
4. *A Review on the Role of Big Data in Business*. **Alam, Jafar Raza**. 4, Vol. 3.
5. *“Turn Big Data into Big Value, A Practical Strategy*. **Intel**.