

BIG DATA: CHALLENGES AND OPPORTUNITIES FOR STORAGE AND MANAGEMENT OF DATA TO BE AVAILABLE AND ACTIONABLE FOR BUSINESS SURVIVAL AND GROWTH

Mr. RameshJadhav
Sinhgad Institute of Management,
Pune, Maharashtra, India

Mr. SunilKhilari
Sinhgad Institute of Management,
Pune, Maharashtra, India

Abstract—Technological advancement and usage of Information Technology has created huge challenges and opportunities for data storage and management to make data as a resource which is available and actionable for the business survival and growth of all types of business. This research gives insight on upcoming big data tools and technologies for database management system as challenges and opportunities. Practicing the big data tools and technologies will give the chance to the business for survival from current competitive age and also for future growth. The study focuses on the upcoming big data tools and technology usage in database management system used in business.

I. INTRODUCTION

Information technology is whole about it's easy to hold onto the concept that database is all about systems, networks, and application software. Big Data is a comfortable and suitable for all concerns of any type of business. For decades accumulated data was considered as a simply bi-product of information processing tasks [1]. However, there is growing usage that stored information is not just digital "raw material", but a corporate asset containing big amounts of hidden value.

Currently storage was considered primary peripheral that was considered only when new systems and applications were being deployed. Also today storage has an identity of its own that is independent from the other systems and software in the software engineering.

Big Data forces focus from IT assets deployment and operation to the management of high-value data assets. It emphasizes that data should be secure, recoverable, easy manipulated, assessable and performing means its "ACID" property should be always in consistence [2] assets sit at the center of concentric rings, ensuring security, recoverability, accessibility, performance, data manipulation, and other aspects of data retention are addressed as abstract requirements with unique requirements. Now information must be captured, identified, valued, classified, assigned to resources, protected, managed

according to policy, and ultimately purged from the system after its value to the organization has been expended.

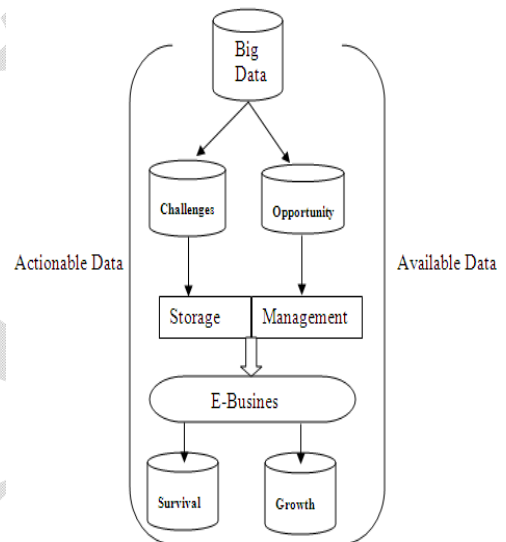


Fig 1: Big Data and business Integration

This requires a fundamental change in corporate culture. As we move into an era of "big data" the entire organization must be aware of information's value as an asset, and the shift from technology-centric approaches for IT management. Just like gold in the above analogy, users must recognize that all data is not "created equal" and delivers different levels of value to an organization for specific periods of time [3].

For example, financial records typically have a high level of inherent value, and retain a level of value for some defined period of time. (The Sarbanes-Oxley act requires publicly-traded companies to maintain related audit documents for no less than seven years after the completion of an audit. Companies in violation of this can face fines of up to \$10 million and prison sentences of 20 years for Executives.)

Big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the strictures of your database architectures.

To gain value from this data, you must choose an alternative way to process it.

The hot IT buzzword of 2012, big data has become viable as cost-effective approaches have emerged to tame the volume, velocity and variability of massive data. Within this data lie valuable patterns and information, previously hidden because of the amount of work required to extract them [4].

To leading corporations, such as Wal-Mart or Google, this power has been in reach for some time, but at fantastic cost. Today's commodity hardware, cloud architectures and open source software bring big data processing into the reach of the less well-resourced. Big data processing is eminently feasible for even the small garage startups, who can cheaply rent server time in the cloud.

The value of big data to an organization falls into two categories: analytical use, and enabling new products. Big data analytics can reveal insights hidden previously by data too costly to process, such as peer influence among customers, revealed by analyzing shoppers' transactions [5], social and geographical data. Being able to process every item of data in reasonable time removes the troublesome need for sampling and promotes an investigative approach to data, in contrast to the somewhat static nature of running predetermined reports.

The past decade's successful web startups are prime examples of big data used as an enabler of new products and services. For example, by combining a large number of signals from a user's actions and those of their friends, Face book has been able to craft a highly personalized user experience and create a new kind of advertising business. It's no coincidence that the lion's share of ideas and tools underpinning big data has emerged from Google, Yahoo, Amazon and Face book.

1.1. Big Data Applications

Just look across any industry and you're going to find some reoccurring themes. One of those themes is more data, because we believe we can make accurate decisions when you have access to more data. More data that's persisted for longer periods of time leads to better models [6].

So that's definitely a recurring Big Data theme: "We want to keep more and more data to get better and better insight, and we able to have analysis on the data that—when it's NOT only structured" There's unstructured and semi-structured to fold into our mostly structured analytics of today and ALL industries are facing this challenge today.

Following are the some major big data applications

- ✓ Bioinformatics: Sequencing Analysis
- ✓ Web: Social Network Analysis
- ✓ Media: Content Optimization
- ✓ Telco: Network Analytics
- ✓ Retail Promotion Analysis
- ✓ Financial: Fraud Analytics
- ✓ Business :Trade Reconciliation
- ✓ Telecommunications – More reliable networks
- ✓ Government– E-Governess services

Smarter Healthcare	Search Quality	Traffic Control	Manufacturing
Finance	Homeland Security	Retail: Churn, NBO	Trading Analytics
Telecom	Multi-channel	Log Analysis	Fraud and Risk

Fig 2 : Various application areas – generate big data

1.2. Big data Characteristics

Behavioral characteristics upon which quality of Big data is majored and controlled with help of three key components

Variety: The data is in both structured and unstructured forms; ranges across the spectrum of E-mail messages, document files, tweets, text messages, audio and video; and is produced from a wide number of sources, such as social media feeds, document Management feeds and, particularly in government, sensors.

Velocity: The data is coming at ever increasing speeds — in the case of some agencies, such As components of the Defense Department and the intelligence community, at Millisecond rates from the various sensors they deploy.

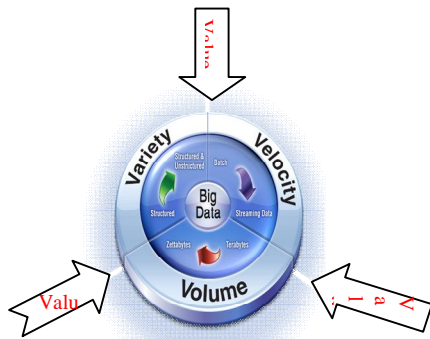


Fig 3 : Quality of big data measuring and controlling parameters

Volume: The data has to be collected, stored and distributed at levels that would quickly overwhelm traditional management techniques. A database of 10 terabytes, for Example, is an order or two less than would be considered normal for a big data Project [7].

Value: The data can be used to address a specific problem or can address a particular mission objective that the agency has defined.

1.3 Definition of Big Data

As the rate of data growth continues to grow and more and more sources of data become relevant for businesses, organizations are forced to look beyond traditional databases. Companies need to store data inexpensively, access it quickly and be able to apply analytics to it to understand its meaning for their business. With the uninterrupted growth of data volumes ever since the primitive ages of computing.

Definition; Big Data can be defined by size/volume, rate of growth, and variety of structure of that data. The definition of Big Data and technologies are inextricably linked, where Value of the latter may force Big Data to peter out into a ‘damp squib.’(Dev Patel, CEO, Bitlyota). Traditional relational database systems were not designed to cope with Big Data. It is the combination of the three dimensions that has given rise to new Big Data technologies broadly in the areas of NoSQL, NewSQL, Storage and Analytics [8].

Datasets are growing larger and larger every minutes. The goals of big data tools and techniques that can be used for managing and analyzing large datasets. These tools and techniques manage databases, NoSQL databases, and distributed file systems, such as Hadoop, Hana. And explorer, analyses big data set.

1.4 Big Data Challenges

Big data presents more number of challenges relating to its complexity.

- ✓ How we can understand and use big data when it comes in an .unstructured format,
- ✓ How we can capture the most important data as it happens and deliver that to the right people in real-time.
- ✓ How we can store the data,
- ✓ How we can analyze and understand it given its size and our computational capacity.
- ✓ Privacy
- ✓ Security to access
- ✓ Deployment.

Table no.1 Big data challenges and there complexity

Sr. No.	Challenges	Technological value	Business value	Social Value
1	Unstructured formats	1	2	3
2	Deployment	2	1	3
3	Security	2	1	3
4	Store data	1	2	3
5	Privacy	3	2	1
6	On budget	3	1	2

**1-High, 2-Medium, 3-Low

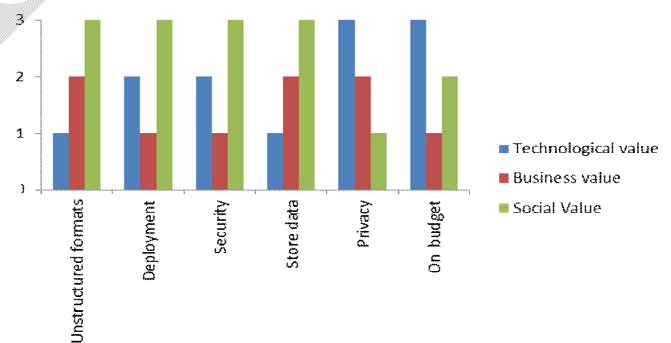


Fig 4 : Big Data Challenges and Domain Values Impact

II. BIG DATA OPPORTUNITIES

Greater than the challenges are the opportunities that big data presents. McKinsey calls big data “the next frontier for innovation, competition and productivity.” We can answer questions with big data that were beyond reach in the past. We can extract insight and knowledge, identify trends and use the data to improve productivity, gain competitive advantage and

create substantial value for the world economy. The challenges with big data are limited compared to the potential opportunities, which are limited only by our creativity and ability to make connections among the trillions of bytes of data we have access to [9]. Big data provides an opportunity to find insight in new and emerging types of data. Some are listed below

Table no.2 big data opportunities and their formulation

Opportunity	Area	Formulation	%
Innovation	Health care	enhancing life expectancy, quality of life,	30%
Productivity	Manufacturing	Agile to react to fluctuations in market demand or supply.	25%
competitions	Global & local	Survival and beat the competitor	20%
Knowledge	Commercial	Business re-engineering	15%
Insight	Research	commercialization	10%

Above table indicates that there will be 30% of opportunities for the innovation in the health care domain.

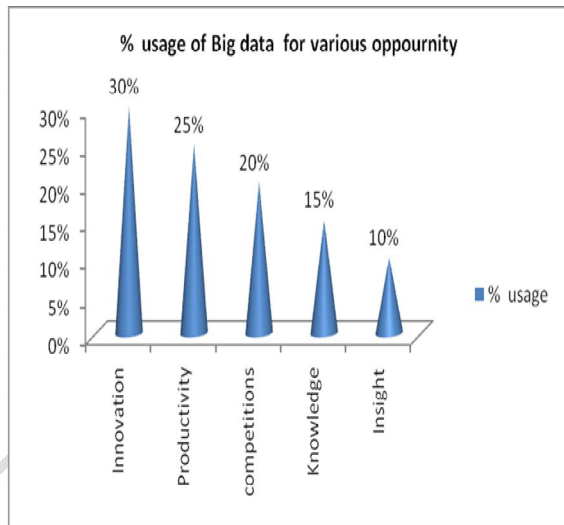


Fig 5 : Area wise Big Data usage

III. BIG DATA TOOLS & TECHNOLOGIES

There are various emerging tools and technologies available for reengineering spotting out business trends by high level analysis on big data .Current trend in the database technology is towards new way of storing data, without eliminating the ACID properties that business applications need, and without sacrificing the SQL language. These tools, there usage and benefit for business on the basis of demand are listed below.

Table no.3 Big Data tools and their benefit

Tool Name	Usage	Benefits	Priority for business	Year of introduction
Hadoop	storing large volumes of unstructured data	framework that supports data-intensive distributed applications	high	2009
Hana	Multi core Data processing ability	multi-engine query processing	high	2011
Cass Abdra (CQL)	self-tuning memtables,	support for zero-downtime upgrade	high	2011
MonogoDB	document-oriented database system	stores structured data as documents with dynamic schemas	Medium	2007
NoSQL	Not only structured query languages	Highly optimized for retrieval and appending operations.	High	2010

Using big data tools to make data actionable for business Column stores, Columnar databases group data by column. Within a column, generally speaking, all the data is of the same type. A columnar store, therefore, stores data of a single type all together, which can give advantages such as the possibility for significant compression. Good compression can lead to reduced disk space requirements, memory requirements, and access times.

Using columnar or other compression approaches, even larger data sets can be loaded entirely into main memory. With high-speed access to memory-resident data, more users can be supported on a single machine. Also, with an in-memory database, both transactional and decision-support queries can be supported on a single machine, meaning that there can be zero latency between data appearing in the system, and that data being actionable and available to the business for decision-support applications.

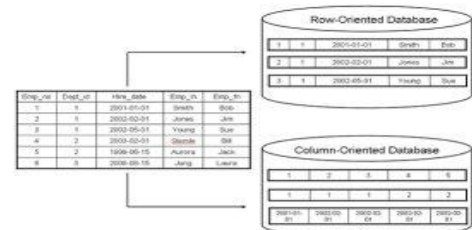


Fig 6 : Grouping of Data

Using big data tools to make data available for business Data availability can be measured in terms of how often the data is available and how much data can flow at a time. Big data enables small businesses to have exciting and interesting conversations with customers

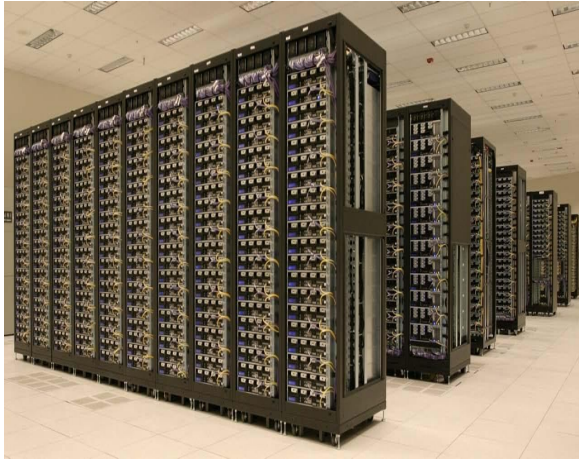


Fig 7 : Data availability for business

- **Big data helps you know your customer:** Consumers leave traces of their preferences in their online interactions. Every tweet, Facebook post, or search term defines and documents consumer tastes or sentiment. Big data gathers this information and helps you know your customer.

It simplifies, categorizes, and segments customers. Using big data, you can identify customer traits and preferences. For example, you can identify triggers to boost your conversion rates for a particular demographic using big data about their preferences.

- **Big data helps you implement strategy:** Majority of small scale businesses, business strategy is a product of intuition and guesswork. However, you can make your approach scientific by using information or metrics derived from big data to implement business strategy.
- **Big data helps you scale business internally and externally:** Big data helps you make sense of the important internal drivers of your business. By making data available to your internal employees, you empower them to make informed decisions.

IV. BENEFITS OF BIG DATA TOOLS AND TECHNOLOGIES FOR BUSINESS

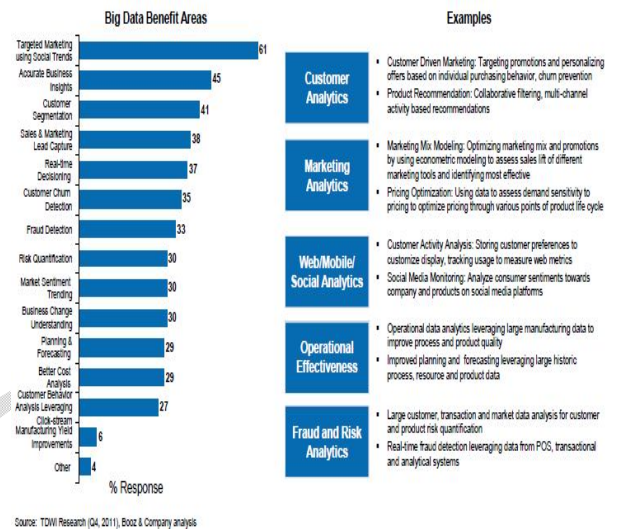


Fig 8 : Area wise Big Data Benefit

To obtain the maximum business impact, this process also requires a precise combination of people, process and analytic tools. Some of the potential business benefits from implementing an effective big data insights methodology include:

- Timely insights from the vast amounts of data. This includes those already stored in company databases, from external third-party sources, the Internet, social media and remote sensors.
- Real-time monitoring and forecasting of events that impact either business performance or operation
- Ability to find, acquire, extract, manipulate, analyze, connect and visualize data with the tools of choice (SAP HANA, SAP Sybase®, SAP Intelligence Analysis for Public Sector application by Palantir, Kapow®, Hadoop).
- Convergence of the BDI solution for variety with the speed of SAP HANA for velocity, and the capability of Hadoop for volumes to manage vast amounts of data, in or out of the Cloud, with validation and verification.
- Identifying significant information that can improve decision quality
- Mitigating risk by optimizing the complex decisions of unplanned events more rapidly

V. CONCLUSION

The data is the “rich” but information is “poor” this scenario is get shifted to rich –rich situation that is rich data and rich information on time and on budget to the business for enhancing competitiveness of business. So carry out various processing on an available big data as store large volumes of unstructured data, Multi core Data processing, self-tuning memtables, make document-oriented database system and Not only structured query languages techniques are used to make data actionable and available within the business. This will create new opportunities in the areas Innovation, Productivity, competitions, Knowledge, research Insight. Big Data tools Hadoop, Hana, Cassabdra (CQL), MonogoDB, and NoSQL are focused on the technology and on the business outcome.

References

- [1] <http://www.adexchanger.com/online-advertising/big-data/>
- [2] Sai Wu, Dawei Jiang, et al. Efficient B-tree Based Indexing for Cloud Data Processing. VLDB 2010
- [3] David Chiu, A. Shetty, et al. Evaluating and Optimizing Indexing Schemes for a Cloud-based Elastic Key-Value Store. In 2011 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing.
- [4] J. Wang, S. Wu, et al. Indexing Multi-dimensional Data in a Cloud System. SIGMOD 2010.
- [5] D. Kossmann, T. Kraska, et al. An Evaluation of Alternative Architectures for Transaction Processing in the Cloud. SIGMOD 2010.
- [6] T. Kraska, M. Hentschel, et al. Consistency Rationing in the Cloud: Pay only when it matters. VLDB 2009.
- [7] H. T. Vo, C. Chen, et al. Towards Elastic Transactional Cloud Storage with Range Query Support. VLDB 2010.
- [8] H. Killapi, E. Sitaridi, et al. Schedule Optimization for Data Processing Flows on the Cloud. SIGMOD 2011.
- [9] M. K. Aguilera, W. Gloab, et al. A Practical Scalable Distributed B-Tree. VLDB 2008 .