



International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Best Practice For Big Data Analytics

Aakash Singh

(MCA-I)

ABSTRACT :

Today's world is highly impacted by technological developments. One of the crucial aspects is the power of making decisions. And, these decisions are made on the basis of the objective data which is generated by the usage and application of technological advancements involving Information Technology. With the proliferation of data from various devices, data management and data analytics are the key components of decision-making processes. Thus, the purpose of this article is to examine the best practices for big data analytics today.

KEYWORDS: Big Data; Datasets; Data Quality; Data Governance; Real-Time Analytics

I. INTRODUCTION

Today data is ruling the business operations. Everywhere we are bombarded by data in varied forms. On account of the importance of data, we have the concept of Big Data. In simple parlance big data is the concept wherein we associate the contexts of data generation and its management in terms of 5V's. these 5V's are enumerated below

Volume refers to the volume of data created and collected from various sources each day. The volume can be in terms of number of people accessing the data sources. For example, the booking of railway tickets during the Tatkal period of the number of people watching the world cup football final match. In both these instances, the number of people watching or booking the tickets is invariably huge as compared to normal operations

Value is the end result of processing the data. In other words, value is the result which assists the people in taking decisions. For example, the trends and patterns of the runs scored by an individual player in a cricket tournament.

Variety is the classification of the data into different categories. It can be structured, unstructured, and semi- structured formats.

Velocity is the speed and pace at which the data is being processed. It involves creation, saving and retrieval of data.

Veracity is the quality of the data in terms of truthfulness and dependability of information and data.

By applying these methods, we can examine the data by combining the above elements and gain valuable insights in terms of handling large data efficiently and effectively. These help them to gain valuable experience and insights that empower businesses to innovate, cut expenses, and improve their business processes.

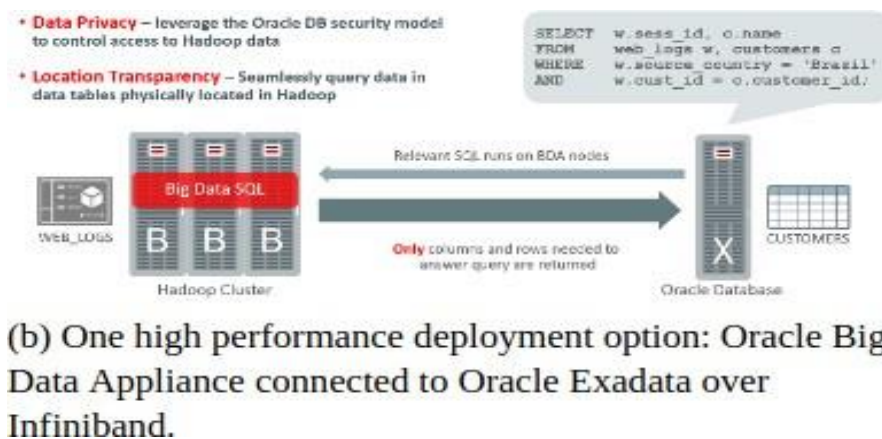
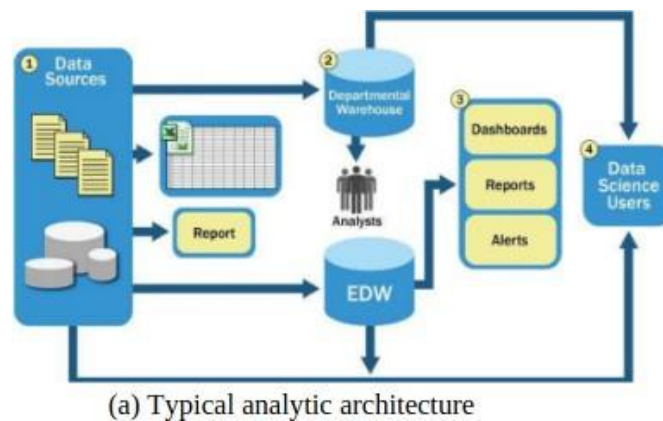
Guidelines for conducting extensive data analysis on a large scale within organizations includes

Data governance which is used to establish guidelines and processes for managing the data in terms of quality and security.

1. Cloud-based infrastructure which includes the aspects of scalability of solutions for an effective data management.
2. Real-time analytics for utilizing the tools for rapid analysis and decision-making processes
3. User-centric metrics for prioritizing key performance indicators (KPI) that directly impact user experience.
4. Evaluation of user interaction. This includes the aspects of evaluating the concepts of how the user is utilizing the data to perform various tasks
5. Visualization tools: these are used to depict the trend and patterns through dashboards about the performance of the business processes

II LITERATURE REVIEW

By taking an example of a author who has written a book. the entire analytic architecture is depicted in the figure (a) given below. [1] provides information about the author of the book. Highlights about big data cloud services and oracle big data. Data design.



III. RESEARCH METHODOLOGY

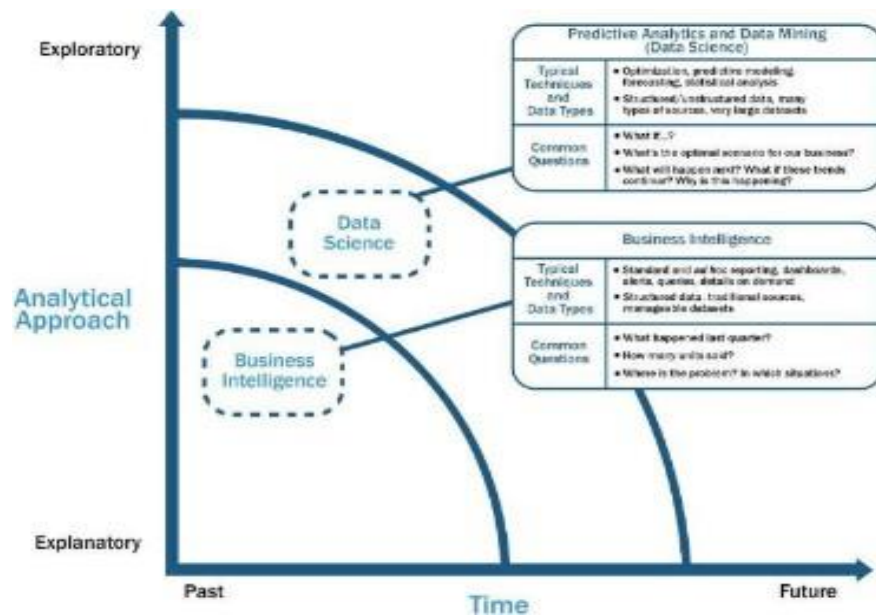
In this paper, we examine the various aspects and the techniques for analyzing large datasets. We examine various methods such as techniques for utilizing big data analytics to achieve optimal results so as to make sound choices.

IV Big Data Analytics

The term big data is applied to a concept where data in varied forms is stored and processed so as to enable the stakeholders to make decisions.[4]. The data as such grows exponentially and is constantly increasing in size. An example of big data is the data wherein the live streaming of a football match is being shown in television, you tube and other forms of social media. The data size in this form of streaming is huge and is measured in terms of tera-bytes (TB) and petabytes (PB). Organizations today are processing and analyzing this data to generate information to make decisions. [5].

BI Versus Data Science

To evaluate the difference between BI and Data Science, the following is the enumeration. BI has a tendency to provide reports, dashboards, and queries on commercial enterprise questions for the modern- day length or within the past. BI systems make it clean to reply questions associated with area-to-date sales, progress towards quarterly targets, and understand how tons of a given product become bought in a prior region or year. those questions tend to be closed-ended and explain modern or beyond behavior, usually through aggregating historical information and grouping it in a few way. BI provides hindsight and some perception and typically answers questions associated with “while” and “wherein” events occurred. records technology has a tendency to be extra exploratory in nature and may use situation optimization to cope with more open-ended questions. This technique provides perception into cutting-edge pastime and foresight into future activities, whilst usually that specialize in questions associated with “how” and “why” occasions arise. in which BI problems generally tend to require exceptionally based information organized in rows and columns for accurate reporting, statistics technological know-how projects tend to use many sorts of information assets, which include huge or unconventional datasets. relying on an organisation’s goals, it may select to embark on a BI task if it’s miles doing reporting, growing dashboards, or acting easy visualizations, or it could pick statistics science initiatives if it desires to do a greater state-of-the-art analysis with disaggregated or various datasets



How to Enable Big Data Best Practices

In order to enable big data best practices, the following are the recommended practices.

- Gather the information pertaining to the available dataset. The quality can in terms of the types of data such as doc file, mp file, pdf file audio video file and the like
- Also gather the type of the technical information pertaining to the dataset such as meta data, the constraints and the type of the linkages of the data in the dataset
- Once this is obtained, brain storm amongst the team members with respect to the likely concepts and the contexts in which the user may demand.
- Develop the dashboard based on the requirements.
- Study the best practices which are adopted and implemented by leading organizations
- Finally document the whole process

VI. CONCLUSION

Developing a big facts strategy isn't a walk inside the park; it calls for essential inputs from statistics scientists, IT teams, analytics leaders, and records managers. To ensure which you aren't flushing cash down the drain with the aid of nose-diving into technological improvements, you ought to take a look at large fact as an asset with business enterprise-centered analytics.

VII. REFERENCES

- [1.]Elgendy, N.: Big Data Analytics in Support of the Decision Making Process. MSc Thesis, German University in Cairo, (2013)
- [2.]EMC: Data Science and Big Data Analytics. In: EMC Education Services, (2012)
- [3.]Oracle (2015) An Enterprise Architect's Guide to Big Data
- [4.]Kubick, W.R.: Big Data, Information and Meaning. In: Clinical Trial Insights,(2012)
- [5.]Russom, P.: Big Data Analytics. In: TDWI Best Practices Report,(2011)