International Journal of Engineering and Techniques - Volume 3 Issue 3, May-June 2017

RESEARCH ARTICLE OPEN ACCESS

Automated Predictive Big Data Analytics Using Ontology Based Semantics

Karthikeyan R [1], Dr.T.Geetha [2] , Vasanthakumar K[3] , Vetrivel A[4]

[1,2] Asst.Prof, Dept of MCA, Gnanamani college of Technology, Namakkal, INDIA.

[3,4] P.G.Scholar, Dept of MCA, Gnanamani college of Technology, Namakkal, INDIA.

Abstract:

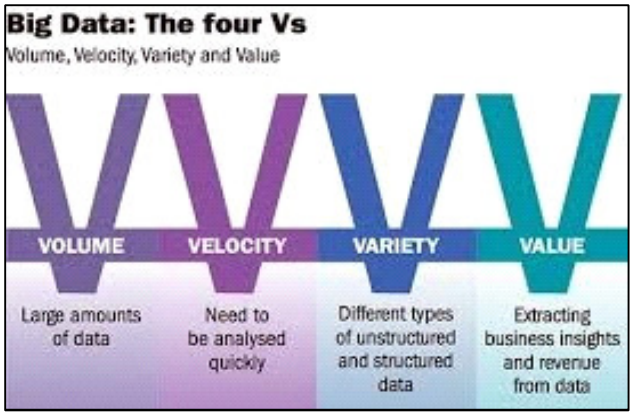
Study the growth history of galaxies by following their 'merger trees' in large-scale astrophysical simulations. The service uses the addition of lambda expressions and the Stream API in Java 8, Java has gained a powerful and expressive query language that operates over in-memory collections of Java objects, making the transformation and analysis of data more convenient. There is much enthusiasm currently about the possibilities created by new and more extensive sources of data to better understand and manage cities. Here, I explore how big data can be useful in urban planning by formalizing the planning process as a general In this paper we discuss Pig SPARQL, a competitive yet easy to use SPARQL query processing system on Map Reduce that allows adhoc SPARQL query processing on large RDF graphs out of the box. Instead of a direct mapping, Pig SPARQL uses the query.

INTRODUCTION

Big data and analytics are vast topics in both the popular and business process. Today, many organizations are collecting, storing, and analyzing massive amounts of data. This data is commonly referred to as “big data” because of its volume, the velocity with which it arrives, and the variety of forms it takes. Big data is creating a new generation of decision support data management Collecting and storing big data creates little value.What is new is the coming together of advances in computer technology and software, new sources of data (e.g., social media), and business opportunity.This confluence has created the current interest and opportunities in big data analytics. It is even spawning a new area of practice and study called “data science” that encompasses the techniques, tools, technologies, and processes for making sense out of big data.

I.WHAT IS BIG-DATA?

One perspective is that big data is more and different kinds of data than is easily handled by traditional relational database management systems (RDBMSs). Some people consider 10 terabytes to be big data, but any numerical definition is likely to change over time as organizations collect, store, and analyze more data.



Describes the four Vs of big-data. Like as Volume ,Velocity, Veriety,And Value.

A. Big-Data Analytics

By itself, stored data does not generate business value, and this is true of traditional databases, data

warehouses, and the new technologies such as Hadoop for storing big data. Once the data is appropriately stored, however, it can be analyzed, which can create tremendous value. A variety of analysis technologies, approaches, and products have emerged that are especially applicable to big data, such as in-memory analytics, in-databaseanalytics, and appliances (all discussed later).

B. Big-Data Sources

Big data has many sources. For example, every mouse click on a web sitecan be captured in Web log files and analyzed in order to better understand shoppers buying behaviors and to influence their shopping by dynamically recommending products. Social mediasources such as Facebook and Twitter generate tremendous amounts of comments and tweets. This data can be captured and analyzed to understand, for example, what people think about new product introductions.Image, voice, and audiodata can be analyzed for applications such as facial recognition systems in security systems.

II.EXAMPLES OF BIG-DATA ANALYTICS

A. Introducing a New Coffee Product at Starbucks

Starbucks was introducing a new coffee product but was concerned that customers would find its taste too strong.The morning that the coffee was rolled out, Starbucks monitored blogs, Twitter, and niche coffee forum discussion groups to assess customers reactions

B.Drilling for Oil at Chevron

Each drilling miss in the Gulf of Mexico costs Chevron upwards of $100 million. To improve its chances of finding oil,Chevronanalyzes 50 terabytes of seismic data.The geologists at Chevron took this time to seize the opportunity offered by advances in computing powerand storage capacity to refine their already advanced computer models.

C.Monitoring Trucks at U.s.Xpress

U.S. Xpress is a transportation company. Its cabs continuously stream more than 900 pieces of data related to the condition of the trucks and their locations [Watson and Leonard, 2011]. This data is stored in the cloud and analysed in various ways, with information delivered to various users, from drivers to senior executives, on iPads and other tablet computers.

II BENEFITS OF BIG-DATA ANALYTICS

Research shows the benefits of using data and analytics in decision making. One study of 179 large publicly tradedfirms found that companies that have adopted data-driven decision making have output and productivity that is 5%to 6% higher than that of other firms. The relationship extends to other performance measures such as asset utilization, return on equity, and market value. Imagine a world with an expanding population but a reduced strain on services and infrastructure; dramatically improved health care outcomes with greater efficiency and less investment; intensified threats to public safety and national borders, but greater levels of security; more frequent and intense weather events, but greater accuracy in prediction and management. Imagine a world with more cars, but less congestion; more insurance claims but less fraud; fewer natural resources, but more abundant and less. expensive energy. The impact of big data has the potential to be asprofound as the development of theInternet itself.

III.RELATED WORKS

cloud-based conference[1], Data warehousing [2],

A. A Clear Business Need

It is common knowledge that projects should be business rather than technology driven. They should address a business need such as solving a problem or seizing an opportunity

- ✓ Automobile insurance
- ✓ Telecommunications
- ✓ Manufacturing, distribution, and retail
- ✓ Transportation and logistics
- ✓ Utilities
- ✓ Gaming
- ✓ Law enforcement

In many organizations, the initial business case for big data analytics focuses on customer-centric objectives and uses existing and newly accessible

internal sources of data. Big data analytics can be especially helpful for companies that seek to understand customers better, develop meaningful relationships with customers, and improve operations that enhance the customer experience.

B. Strong, Committed Sponsorship

Without solid sponsorship, it is difficult to succeed with any IT project, and this includes big data analytics projects. If the project is departmental, sponsorship can reside at the departmental level. However, projects that are more strategic and enterprise wide should have senior management support.

C. A Fact-Based Decision-Making Culture

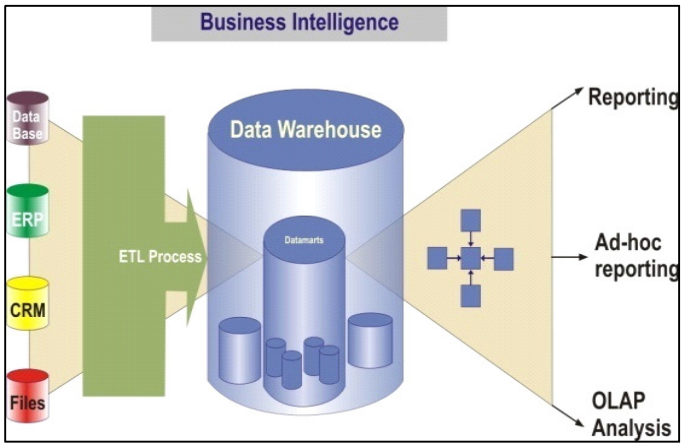
To benefit from big data analytics, decisions must be based on "the facts" (generated by analytics) and there should be constant experimentation to see what works best. Changing the organizational culture associated for how decisions are made can be more challenging than solving technical issues. As the CEO explained: "Their idea of marketing was giving balloons and suckers along the teller line and running focus groups, but marking has become very analytical."Senior management can do other things to change the culture

D.A Strong Data Infrastructure

Data is critically important to BI and analytics. When a strong data infrastructure is in place, applications can often be developed in days. Without a strong data infrastructure, applications may never be completed. IT understands the importance of the data infrastructure, but the business units sometimes assume it is a given and don't fully appreciate what is required to create and maintain it.

IV DATA WAREHOUSES

For many organizations, data warehouses provide the single version (or source) of the truth for decision support data. The data is extracted from source systems (e.g., operational systems, ERPs), transformed (e.g., consistent formats), integrated (e.g., around a common key, such as a customer ID), and loaded into the data warehouse. The data can be thought of as "squeaky clean" because of the care taken to ensure its accuracy.



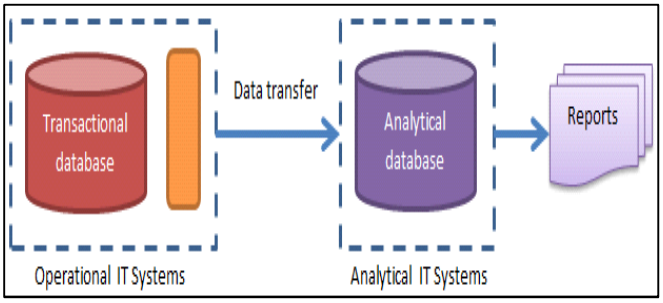
Users and applications access the data from the warehouse to support decision making. Data warehouses are primarily designed for the storage and analysis of structured data—that is, data easily stored in the rows and columns of relational databases. The data is used for queries, reporting, online Analytical processing (OLAP), dashboards/scorecards, data visualization, and regulatory and compliance requirements

A. In-Memory Analytics

In-memory technology comes in two forms: either on the platform or with the BI tool. When implemented on theplatform, the server stores the data in-memory, and data is accessed by the BI tools. In-memory BI tools also move data between disk (e.g., in the data warehouse) and the local desktop memory so that the most frequently used data (so called "hot data") is available in memory.

B. In-Database Analytics

A change is taking place as to where analytics is performed. In the past, data was moved to a server (think of asandbox) and the analysis was performed there.



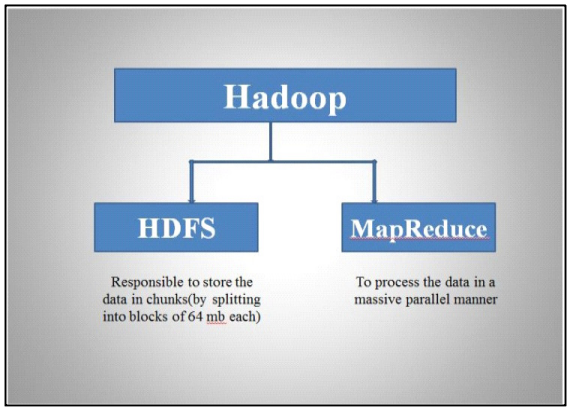
In describes for Data has been transferred from transactional analytics to database analytics to generate a report.

V CLOUD-BASED SERVICES

The cloud is now in the mainstream of computing. The potential benefits of the cloud include access to specialized resources, quick deployment, easily expanded capacity, the ability to discontinue a cloud service when it is no longer needed, cost savings, and good backup and recovery. private clouds are implemented within a company's firewall. Concerns about data security is a primary reason that private clouds are sometimes preferred over public clouds. We will discuss public clouds—although the same approaches and technologies are used with private clouds. Cloud services are available as software-as-a-service (SaaS), platform-as-a-service (PaaS), or infrastructure-as-aservice(IaaS), depending on what software is provided.

VI .HADOOP/MAP REDUCE

Apache Hadoop is a software framework for processing large amounts of data across potentially massively parallel clusters of servers.



.

To illustrate, Yahoo has over 42,000 servers in its Hadoop installation. The key component of Hadoop is the Hadoop Distributed File System (HDFS), which manages the data spread across the various servers. It is because of HDFS that so many servers can be managed in parallel. HDFS is file based and does not need a data model to store and process data. It can store data of any structure, but is not a RDBMS. HDFS can manage the storage and access of any type of data (e.g., Web logs, XML files) as long as the data can be put in a file and copied into HDFS.

CONCLUSION

- Should be effective and strong .
- Summarize the main program.
- Suggest future avenues of research.
- Chance to give opinion.

With the advent of Hadoop the new release of Hadoop known as Yet Another Resource thinking has been solidified. As is explained in this chapte, Hadoop YARN separates the resource scheduling part from the MR paradigm. It should be noted that in the first-generation Hadoop, the scheduling was tied with implying that the only processing that was possible on data was the MR type or its orchestrations.

REFERENCES

1. Y. Feng, B. Li and B. Li, "Airlift": Video conferencing as a cloud service using inter-datacenter networks, in Proceedings of the IEEE International Conference on Network Protocols(ICNP'12), (2012).
2. R.Karthikeyan," Improved Apriori Algorithm for Mining Rules" in the International Journal of Advanced Research in biology Engineering science and Technology Volume 11, Issue 4, April 2016, Page No:71-77.
3. R.Karthikeyan,Dr.T.Geetha "Honeypots for Network Security", International journal for Research & Development in Technology.Volume 7.Issue 2 ,Jan 2017,Page No.:62-66 ISSN:2349-3585
4. R.Karthikeyan,"A Survey on Position Based Routing in Mobile Adhoc Networks" in the international journal of P2P Network Trends and Technology, Volume 3 Issue 7 2013, ISSN:2249-2615
5. R.Karthikeyan,"A Survey on Sensor Networks" in the International Journal for Research & Development in Technology Volume 7, Issue 1, Jan 2017, Page No:71-77
6. R.Karthikeyan,Dr.T.Geetha "Web Based Honeypots Network",in the International journal for Research & Development in Technology.Volume 7.Issue 2 ,Jan 2017,Page No.:67-73 ISSN:2349-3585.

7. R.Karthikeyan,Dr.T.Geetha,"A Simple Transmit Diversity Technique for Wireless Communication",in the International journal for Engineering and Techniques. Volume 3. Issue 1, Feb 2017, Page No.:56-61 ISSN:2395-1303.
8. C.Ganesh,B.Sathyabhama,Dr.T.Geetha " Fast Frequent Pattern Mining using Vertical Data Format for Knowledge Discovery "International Journal of Engineering Research in Management & Technology. Vol.5,Issue-5,Pages:141-149.
9. R.Karthikeyan,Dr.T.Geetha "Strategy of Trible – E on Solving Trojan Defense in Cyber Crime Cases", International journal for Research & Development in Technology.Volume 7.Issue 1 ,Jan 2017,Page No.:167-171
10. R.Karthikeyan,"A Survey on Position Based Routing in Mobile Adhoc Networks" in the international journal of P2P Network Trends and Technology, Volume 3 Issue 7 2013, ISSN:2249-2615.
11. K.Ramya and K.Pavithradevi "Effective Wireless Communication",International journal of Advanced Research, Vol 4(12), pp.1599-1562 dec 2016.
12. R.Karthikeyan,Dr.T.Geetha "FLIP-OFDM for Optical Wireless Communications" in the international journal of Engineering and Techniques, Volume 3 Issue 1, Jan - Feb 2017, ISSN:2395-1303,PP No.:115-120.
13. R.Karthikeyan,Dr.T.Geetha"Application Optimization in Mobile Cloud Computing" in the international journal of Engineering and Techniques, Volume 3 Issue 1, Jan - Feb 2017, ISSN:2395-1303,PP No.:121-125.
14. "Eckerson", W. (2004) "Gauge Your Data Warehousing Maturity", DM Review, (14)11, pp. 34.
15. R.Karthikeyan,Dr.T.Geetha"Estimating Driving Behavior by a smart phone" in the international journal of Engineering and Techniques, Volume 3 Issue 2, March 2017, ISSN:2395-1303,PP No.:84-91.
16. R.Karthikeyan,Dr.T.Geetha"Advanced Honey Pot Architecture for Network Threats Quantification" in the international journal of Engineering and Techniques, Volume 3 Issue 2, March 2017, ISSN:2395-1303, PP No.:92-96.