

CLOUD COMPUTING IN ANALYTICS

Bidyutlata Sahoo

Assistant Professor (MCA)
Aurora's P.G. College,
Paniagutta, Hyderabad

Abstract: Big Data and Cloud Computing as two important technologies are at the center of concern in the IT industry. Every single day an enormous amount of data is producing from different sources. This data is so huge in size that traditional processing tools are unable to deal with them. Besides being huge in size, this data travels so fast and has a lot of diversity. Big Data is a concept that deals with storing, processing and analyzing large amounts of data. Cloud computing on the other hand is about offering the infrastructure, platform and software to enable such processes in a cost-effective and efficient manner. Many sectors, including among others businesses (small or large), healthcare, education, etc. are trying to leverage the power of Big Data. In healthcare, for example, Big Data is being used to reduce costs of treatment, predict outbreaks of pandemics, prevent diseases etc. This paper, presents an overview of Big Data Analytics as a crucial process in many fields and sectors and how cloud computing helps for this purpose.

Keywords: Big data, DASaaS, DAPaaS, DAaaS, Cloud Computing

INTRODUCTION

The expansion of data warehouses, webpages, audio and video streams, tweets and blogs is generating a massive amount of complex and extensive digital data. Coherent means are now available for creating, storing, and sharing this information, which also leads to data growth [1]. However, extracting useful knowledge from huge digital datasets requires smart and scalable analytics services, programming tools, and applications.

Big data analytics use compute-intensive data mining algorithms that require efficient high-performance processors to produce timely results. Cloud computing infrastructures can serve as an effective platform for addressing both the computational and data storage needs of big data analytics applications. Much big data already resides in the cloud, and this trend will increase in the future. For example, IT research and advisory firm Gartner estimates that, by 2016, more than half of large companies' data will be stored in the cloud. This trend requires that clouds become the infrastructure for implementing extensive and scalable data analytics platforms. Coping with and gaining value from cloud-based big data requires novel software tools and innovative analytics techniques.

RESEARCH OBJECTIVE

The main objective of the study was how Cloud Computing helps in analytics.

LITERATURE REVIEW

The amount of data each organization deals with today has been rapidly growing. However, analysing large datasets commonly referred to as "big data" has been a huge challenge due to lack of suitable tools and adequate computing resources. What is the current state of big data solutions and service providers? How effective are some of the solutions that have been put into real world practices? What is the current state of cloud computing technologies? [Zhiming xue-Cloud computing & big data computing]. How advanced data mining techniques and associated tools help extract information from large, complex datasets that is useful in making informed decisions in many business and scientific applications including tax payment collection, market sales, social studies, biosciences, and highenergy physics? [D. Talia-Cloud for scalable big data analytics 2013].

CLOUD BASED BIG DATA ANALYTICS

Big data refers to large, heterogeneous, more complex data sets and often unstructured digital content especially from new data sources that is difficult to process using traditional data management tools and techniques. The term encompasses the complexity and variety of data and data types, real-time data collection and processing needs, and the value that can be obtained by smart analytics. Advanced data mining techniques and associated tools can help extract information from large, complex datasets that is useful in making informed decisions in many business and scientific applications including tax payment collection, market sales, social studies, biosciences, and high-energy physics. Combining big data analytics and knowledge discovery techniques with scalable computing systems will produce new insights in a shorter time. Although few cloud-based analytics platforms are available today, current research work anticipates that they will become common within a few years. Some current solutions are based on open-source systems such as Apache Hadoop and SciDB, while others are proprietary solutions provided by EMC, BigML, Splunk Storm, Kognitio, and Insights One. As more such platforms emerge, researchers will port increasingly powerful data mining programming tools and strategies to the cloud to exploit complex and flexible software models such as the distributed workflow paradigm.

SERVICE MODELS IN DATA ANALYTICS

Data Analytics as a Service (DAaaS) is an extensible analytical platform provided using a cloud-based delivery model, where

various tools for data analytics are available and can be configured by the user to efficiently process and analyze huge quantities of heterogeneous data.

Developers and researchers can adopt the software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS) model to implement big data analytics solutions in the cloud. The SaaS model offers complete big data analytics applications to end users, who can exploit the cloud's scalability in both data storage and processing power to execute analysis on large or complex datasets. The PaaS model provides data analytics programming suites and environments in which data mining developers can design scalable analytics services and applications. Researchers can exploit the IaaS model to compose a set of virtualized hardware and software resources for running data analysis frameworks or applications.

Data analytics software as a service (DASaaS)—It is a single and complete data mining application or task (including data sources) offered as a service. It provides a well-defined data mining algorithm or ready-to-use knowledge discovery tool as an Internet service to end users, who can access it directly through a Web browser. Here the clients may be End users, analytics managers or data analysts.

data analytics platform as a service (DAPaaS)— It is a data analysis suite or framework for programming or developing high-level applications, hiding the cloud infrastructure and data storage. It provides a supporting platform that developers can use to build their own data analytics applications or extend existing ones without concern about the underlying infrastructure or distributed computing issues. Here the users may be data mining application developers or data scientists.

data analytics infrastructure as a service (DAIaaS)—It is a set of virtualized resources provided to a programmer or data mining researcher for developing, configuring, and running data analysis frameworks or applications. It provides a set of virtualized resources that developers can use as a computing infrastructure to run data mining applications or to implement data analytics systems from scratch. Here the users may be data mining programmers, data management developers or data mining researchers.

WORKFLOWS BIG DATA ANALYTICS

Developers can use workflows, which consist of complex graphs of many concurrent tasks, to address the complexity of scientific and business applications. This approach supports data analytics design by providing a paradigm that encompasses all the steps of data analytics, from data access and filtering to data mining and sharing produced knowledge. Workflow-based data mining frameworks that run on cloud platforms and use a service-oriented approach offer a flexible programming model, distributed task interoperability, and execution scalability that reduces data analytics completion time. Application developers can design data analysis tasks, scientific computation methods, and complex simulation techniques as workflows that integrate single Web services and execute them concurrently on virtual machines in the cloud.

Figure 1 shows a data analysis workflow application designed using the Data Mining Cloud Framework's graphical programming interface recently developed in our laboratory. Data sources and tools such as data mining algorithms, filters, and data splitters are connected through direct edges that define specific dependency relationships among them. When creating an edge between two nodes, the system automatically attaches a label to it that represents the relationship between them. To ease workflow composition and allow users to monitor its execution, each resource icon has an associated tag—the checkmarks in Figure 1—representing the status of a corresponding resource. The experimental results of a set of studies using the framework to analyze genomics, network intrusion, and bioinformatics data demonstrated its effectiveness, as well as the linear scalability achieved through concurrent execution of the workflow tasks on a pool of virtual servers. Current research focuses on the workflow composition interface, with the aim of extending supported design patterns such as conditional branches and iterations and evaluating its functionality and performance during the design and execution of complex data mining workflows on large datasets in the cloud.

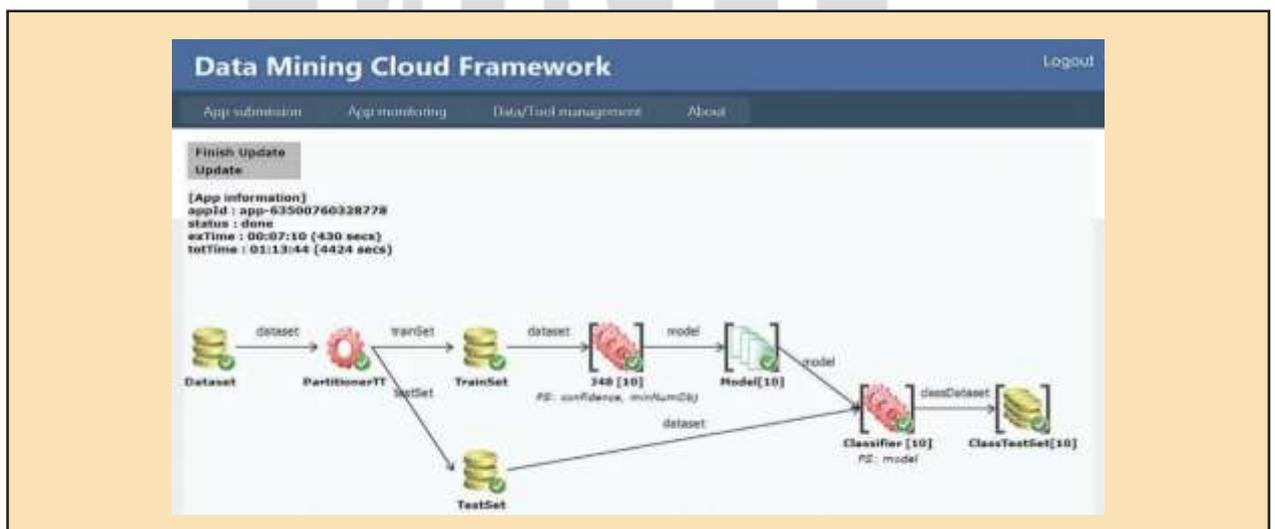


Figure 1. Data analysis workflow application designed using the Data Mining Cloud Framework's graphical programming interface

RESEARCH RECOMMENDATIONS

Cloud-based data analytics requires high-level, easy-to-use design tools for programming large applications dealing with huge, distributed data sources. This necessitates further research and development in several key areas. Programming abstracts for big data analytics. Big data analytics programming tools require novel complex abstract structures. The MapReduce model is often used on clusters and clouds, but more research is needed to develop scalable higher-level models and tools. Data and tool interoperability and openness. Interoperability is a main issue in large-scale applications that use resources such as data and computing nodes. Standard formats and models are needed to support interoperability and ease co- operation among teams using different data formats and tools. Integration of big data analytics frameworks. The service- oriented paradigm allows running large-scale distributed workflows on heterogeneous platforms along with software components developed using different programming languages or tools. The Web and cloud services paradigms can help manage worldwide integration of multiple data analytics frameworks. Data provenance and annotation mechanisms. Provenance is captured as a set of dependencies between elements that researchers can use to interpret data and provide reproducible analysis. Research is needed to develop innovative techniques for visualizing and mining provenance data. These solutions, together with others addressing data privacy and security concerns, will promote cloud-based data analytics in large companies, and eventually will benefit users such as independent research teams, start-ups, and small enterprises that aren't deeply skilled in cloud programming and management.

CONCLUSION

Advancing the cloud from a computation and data management infrastructure to a pervasive and scalable data analytics platform requires new models, tools, and technologies that support the implementation of dynamic data analysis algorithms. Using a synergic approach that integrates the use of clouds and data analysis techniques to investigate key issues will help researchers achieve this goal.

REFERENCES

- [1]. Mostapha Zbakh-“Cloud Computing and Big Data: Technologies, Applications and Security” July 2018.
- [2]. Zhiming xue-“Cloud computing & big data computing”
- [3]. R. Gupta, H.Gupta, M. Mohania-“Cloud computing and big data analytics”- 2012
- [4]. D. Talia-“Cloud for scalable big data analytics”- 2013

