Cloud Computing Analysis for Environment Conditions and Sustainable Agriculture

Pratul Sharma, Brajesh Kumar Singh

Deptt. Of Computer Science & Engineering, R.B.S Engineering Technical Campus, Bichpuri, Agra

Abstract - In present time the data is transferred from source to destination in various forms and also the storage space is much needed. Cloud Computing Technologies describes a new generation of Technologies which are used to store a large amount of data and also extract some key values from the large data in an efficient manner. This Huge amount of data is hiding valuable information's. So Cloud computing is also used to extract the information from such the data ocean. So in this paper, we can utilize the environmental conditions for the data storage and also analyze some important facts about the data storage. Cloud computing provides enough capacity for storage and also processing power for handle large amount of data. In this paper, we provide the analysis over environment condition like winds, rainfall, Temperature etc. with the help of this analysis the farmers can utilize the environmental conditions in according to increase the productivity. Due to the use of particular cloud computing Analytical tool we get some understandable information from it which can be utilized by farmers for successful Agriculture.

Keywords - *Cloud Computing Analysis, Environment Conditions, Data Ocean, Agriculture*

I. INTRODUCTION

In cloud computing delivering information technology (IT) services are done. In which resources are retrieved from the Internet through web-based tools and applications, as opposed to a direct connection to a server. Instead of keeping files on a hard drive or local storage device, cloud-based storage makes it possible to save them to a remote database. As long as an electronic device has access to the web, it has access to the data and the software programs to retain it. In cloud computing Environment, Due to large data storage requirements the heavy server processing required to analyze large volumes of data economically. Cloud Computing refers to the ability to accumulate, structure, and interpret unstructured data. In cloud computing, the term Cloud generally refers to data sets whose size is larger the ability of commonly used software tools to capture, manage, and process within an endurable proceed time.

II. LITERATURE REVIEW

There are many institutions and researchers such as Monsanto, who is famous for seeds; they recently bought Climate Corporation, San Francisco Startup in $ 95 million. Climate Corporation focuses on crop information and insurance for farmers. Using Climate and Agriculture related information, Climate Corporation tells the farmers that it is best for water, their crops and chemicals and nutrients. In addition, the company collects water data from the marine and atmospheric administration and collects the temperature from the national climate services and clarifies both sets on the web site. Lamonica (2013). With the in association of Climate Corporation, Monsanto can start recommending some of their current products based on the data collected. Through the Web platform, farmers can monitor crops, reduce costs of farming, but more importantly make better and Far-sighted decisions to increase crop yield. As the local food and sustainable farming movements grow, Cloud Computing Analysis is helping farmers grow to produce more food at lower costs. By using the Cloud Storage Information a sensor can be developed due to this analysis of factors of production can be done. Infact other factors of production such as weather can't be changed, the sensors can match current weather patterns to previous weather and also select some better crops for the change of weather. Similar technology applications are in cattle management. The Animals are also monitored by sensors and notify farmers which is needs food, Cloud computing can be a powerful tool for scientists and researchers sharing large amounts of environmental data. Andrea (2013). In Durban (South Africa) , the European environment agency, geospatial software company Esri and Microsoft showed off the "eye on earth" network.At the United Nations climate conference (cop 17) The community uses Esri's cloud services and Microsoft Azure to create an online site and group of services for scientists, researchers, and policymakers to upload, share and analyze environmental and geospatial data. Google launched its own satellite and mapping service called Google earth engine, which combines an open API, a computing platform, and 25 years of satellite imagery available to researchers, scientists, organizations and government agencies. Google earth engine offers both tools and parallel processing computing power to groups to be able to use satellite imagery to analyze environmental conditions in order to make sustainability decisions. (At last year's u.n. climate meeting, cop 16). The government of Mexico created the first incredible, high-resolution map of Mexico's forests that incorporated 53,000 lands sat images to produce a 6 GB mapping product. The Mexican government and NGOs can use the map to make decisions about land use, sustainable agriculture, and species protection in combination with a growing population. Cloud computing research is helpful in both the business world and environmental purposes too. Scientists measure and monitor various attributes of lakes, rivers, oceans, seas, wells, and other water environments to support environmental research. Important research on water conservation and stability depends on understanding the underlying atmosphere and underwater environments and knowing how they change. Herwitz (2013). Changes in these natural environments can have an immense impact on the economic, physical, and cultural well-being of individuals and communities throughout the world. In order to improve their ability to forecast environmental impacts, researchers from universities and environmental organizations around the world have started to include analysis of speed data in their research. Scientific research includes the collection of large volumes of time-sensitive information about water resources and whether to help protect communities against risks and respond appropriately to disasters impacting these natural resources. Mathematical models are used to make predictions such as the severity of flooding in a particular location or the impact of an oil spill on sea life and the surrounding ecosystem. The data that can be used include everything from measuring temperatures, to measuring the current flow, to measuring the chemicals in water. In addition, it is helpful to be able to compare this newly obtain data with historical information about the same bodies of water. Many sophisticated research programs are in place to improve the understanding of how to protect natural water resources. Rivers and adjacent floodplains and lakes, for example, require security because they are important places for fish and wildlife. Many communities depend on rivers for drinking water, food, transportation, Power generation and tourism. Other than this, the rivers are monitored to provide knowledge about flooding and to give communities advance warnings about floods. By adding a real-time component to these research projects, scientists are expected to have a major impact on people's lives. In a research center in the United States, sensors are used to collect physical, chemical and biological data from the rivers. These sensors monitor spatial changes in temperature, pressure, salinity, debris, and water chemistry. Their goal is to create real-time monitoring networks for rivers and rivers. Researchers hope that in future, they will be able to predict the changes in rivers, the way weather predictions are made. Another research center located in Europe is using the sensible radio-equipped pulp to collect data about the ocean, including the waves and the measure of action. This streaming data is mixed with other environment and weather data to provide real-time information to fishermen and researchers on the ocean conditions. In both examples, sensors are used to collect large volumes of data as events are taking place. In both instances, the sensor is used to collect large amounts of data as the incidents occur. Although the infrastructure platforms are different, it is certain to include the middleware layer to integrate the data collected by the sensor with the data in the Data Mair. These research organizations are also using geographical information along with external sources like mapping databases and sensors coming from other places. Data is analyzed and processed because it exits from these different sources. An organization is building integrated networks of sensors, robotics and mobile monitoring.

This is using this information to create complex models like real-time, multiparty modeling systems. The model will be used to see dynamic interactions within local rivers and river ecosystems. Streaming technology opens up new areas of research and takes the concept of scientific data collection and analysis in a new direction. They are looking at the data that they have already collected in a new way and are able to collect new types of data sources. Although over time, changes of change in fixed intervals can learn a lot about temperature and water chemistry, you can forget about identifying changes or patterns. When you have the opportunity to analyze the streaming data, it is possible to pick up on the pattern you remember. Real-time data on river and weather is used to predict and manage river changes. Scientists are expected to predict the environmental impact and predict the weather. They are researching on the impact of global warming.

They are asking what can be learned from observing the activities of the fish that emerge from the fish, how can the pollutants be delivered to clean the future with pollution of the environment? If data scientists are capable of carrying data collected already, they can combine it with real-time data more efficiently. They have the ability to analyze more deeply and do better work to predict future outcomes.

Because this analysis has been completed, other groups need the same information so that the findings can be used in a new way to analyze the effect of various issues. This data can be stored in a data cloud environment so that access to researchers around the world, adding new data to the mix, and solving other environmental problems, Zik (2011).

Schadt et al (2010) Interest on computational solutions for collective data management and analysis, which provides cloud and asymmetrical computing as a solution to deal with mass and high-dimensional data sets. These technologies have been around for many years, raising questions: Why do not they use bioinformatics more often? The answer is that, apart from starting complexity, they are quickly broken when data is transmitted amidst the computing nodes. In his review, Schadt and colleagues have said that computational analysis in biology is high-dimensional, and estimates that the data of the extras, even the exabytes will be stored and analyzed soon. We agree with this prediction scenario and explain it through a simple calculation, how current computational technologies really are suitable for such a large amount of data.

bottleneck in this process is the input/output (IO) hardware that links data storage to the calculation node. All nodes are idle for long periods, waiting for data to arrive from storage; shipping the data on a hard disk to the data storage would not resolve this bottleneck. We estimate that 1,000 cloud nodes each processing 1 petabyte (1 petabyte to 1 exabyte of total data) would take 2 years, and cost $6,000,000. In the calculations, 1,000 computational nodes each Currently, it takes minimally 9 hours for each of 1,000 cloud nodes to process 500 GB, at a cost of US$3,000 (500 GB to 500 TB of total data). The processing 500 GB would take 9 hours (at a rate of 15 MB/s) using large nodes at US$0.34/h. The total cost for a single analysis run would be 1,000 × 9 × 0.34 = $3,060. In reality, throughput will be lower because of competition for access to data storage caused by parallel processing. There are significant throughput instability and abnormal delay variations, even when the network is lightly utilized. In the illustrated example, 1,000 cloud nodes each processing a petabyte would take 750 days (at 15 MB/s) and cost 1,000 × 750 × 24 × 0.34 = $6,120,000. A less expensive option would be to use heterogeneous computing, in which graphics processing units (GPUs) are used to boost speed. A similar calculation shows, however, that GPUs are idle 98% of the time when processing 500 GB of data. GPU performance rapidly degrades when large volumes of data are communicated, even with state-of-the-art disk arrays. Furthermore, GPUs are vector processors that are suitable for a subset of computational problems only. Which is the best way forward? Computer systems that provide fast access to petabytes of data will be essential. Because high-dimensional large data sets exacerbate IO issues, the future lies in developing highly parallelized IO using the shortest possible path between storage and central processing units (CPUs). Examples of this trend are Oracle Exadata2and IBM Netezza, which offer parallelized exabyte analysis by providing CPUs on the storage itself. Another trend for improving speed is the integration of photonics and electronics.To fully exploit the parallelization of computation, bioinformaticians will also have to adopt new programming languages, tools, and practices, because writing correct software for concurrent processing that is efficient and scalable is difficult. The popular R programming language, for example, has only limited support for writing parallelized software. However, other languages can make parallel programming easier by, for example, abstracting threads and shared memory. So, not only do cloud and heterogeneous computing suffer from severe hardware bottlenecks, they also introduce (unwanted) software complexity. It is our opinion that large multi-CPU computers are the preferred choice for handling big data. Future machines will integrate CPUs, vector processors and random access memory (RAM) with parallel high-speed interconnections to optimize raw processor performance. Our calculations show that for petabyte- to exabyte-sized high-dimensional data, bioinformatics will require unprecedented fast storage and IO to perform calculations within an acceptable time frame. Golpayegani et al(2009) paper on Cloud Computing for Satellite Data Processing on High-End Compute Clusters where they focused on Hadoop and MapReduce framework of which they suggested improvement without having to change its coding.They agreed that these tools are well suited to the analysis of big data in the cloud. Baraglia et al(2010) discussed the huge deluge of data in the cloud and the fact that it can be handled by the latter's architecture albeit with shortcomings which he gave algorithmic solutions.

III. ANALYTIC ALGORITHMS

Parallel computing is a well-adopted technology seen in processor cores and software thread-based parallelism. However, massively parallel processing— leveraging thousands of networked commodity servers constrained only by bandwidth—is now the emerging context for the Data Cloud. If distributed file systems, such as GFS and HDFS, and column-oriented databases are employed to store massive volumes of data, there is then a need to analyze and process this data in an intelligent fashion. In the past, writing parallel code required highly trained developers, complex job coordination, and locking services to ensure nodes did not overwrite each other. Often, each parallel system would develop unique solutions for each of these problems. These and other complexities inhibited the broad adoption of massively parallel processing, meaning that building and supporting the required hardware and software was reserved for dedicated systems. MapReduce, a framework pioneered by Google, has overcome many of these previous barriers and allows for data-intensive computing while abstracting the details of the Data Cloud away from the developer. This ability allows analysts and developers to quickly create many different parallelized analytic algorithms that leverage the capabilities of the Data Cloud. Consequently, the same MapReduce job crafted to run on a single node can as easily run on a group of 1,000 nodes, bringing extensive analytic processing capabilities to users in the enterprise. Working in tandem with the distributed file system and the multi-dimensional database, the MapReduce framework leverages a master node to divide large jobs into smaller tasks for worker nodes to process. The framework, capable of running on thousands of machines, attempts to maintain a high level of affinity between data and processing, which means the framework intelligently moves the processing close to the data to minimize bandwidth needs. Moving the compute job to the data is easier than moving large amounts of data to a central bank of processors. Moreover, the framework manages extrapolative errors, noticing when a worker in the cloud is taking a long time on one of these tasks or has failed altogether and automatically tasks another node with completing the same task. All these details and abstractions are built into the framework. Developers are able to focus on the analytic value of the jobs they create and no longer worry about the specialized complexities of massively parallel computing. An intelligence analyst is able to write 10 to 20 lines of computer code, and the MapReduce framework will convert it into a massively parallel search—working against petabytes of data across thousands of machines—without requiring the analyst to know or understand any of these technical details. Tasks such as sorting, data mining, image manipulation, social network analysis, inverted index construction, and machine learning are prime jobs for MapReduce. In another scenario, assume that terabytes of aerial imagery have been collected for intelligence purposes. Even with an algorithm available to detect tanks, planes, or missile silos, the task of finding these weapons could take days if run in a conventional manner. Processing 100 terabytes of imagery on a standard computer takes 11 days. Processing the same amount of data on 1,000 standard computers takes 15 minutes. By incorporating MapReduce, each image or part of an image becomes its own task and can be examined in a parallel manner, Farber et al (2011).

Tsuchiya 2012 examined two processing technologies of big data in the cloud and referred to works done in Fujistu laboratories. Consumption of big data can be made as easy as online media according to Foster (2009) after he studied geosciences data on the cloud.Note that two technical entities have come together. First, there's big data for massive amounts of detailed information. Second, there's advanced analytics, which is actually a collection of different tool types, including those based on predictive analytics, data mining, statistics, artificial intelligence, natural language processing, and so on. Put them together and you get big data analytics, the hottest new practice in BI today. Of course, businesspeople can learn a lot about the business and their customers from BI programs and data warehouses. But big data analytics explores granular details of business operations and customer interactions that seldom find their way into a data warehouse or standard report. Some organizations are already managing big data in their enterprise data warehouses (EDWs), while others have designed their DWs for the well-understood, auditable, and squeaky clean data that the average business report demands. The former tend to manage big data in the EDW and execute most analytic processing there, whereas the latter tend to distribute their efforts onto secondary analytic platforms. There are also hybrid approaches.

IV. THE RESEARCH PROBLEM

Because of its massive datasets, today environmental sustainability meets big data. The problem is not data but interpretation. To start, environmental sustainability is a complex idea. Generally speaking, most people are not well-versed in the language concerning the idea. Overly complex and hard to understand, executives, are unable to make intuitive decisions based on the information because they cannot interpret the data. The data lacks context and therefore holds little meaning to top-level executives. Executives are lost in a sea of information, laced with confusing terminology. This lack of clarity makes business decisions even more difficult. To an outsider, resource consumption isn't measured in traditional terms, but rather in a set of foreign "currencies" without a currency converter:

- Water in kgal
- Electricity in kWh
- Heating in mBTu
- CO2 in tons

The problem that appears to be halting the harmonious relationship between Big Data and resource consumption information is not the data. The root of the problem rests in the interpretation of the information and datasets. The numbers are already difficult to digest and unintuitive, so it doesn't help the case that there isn't a commonly shared, universal language in the measurement of environmental sustainability. The Solution to this problem is through the development of a mathematically rigorous, yet simple and intuitive way to interpret the different data streams, companies may be able to make better business decisions. Companies currently use business intelligence and analytics to better understand and predict future trends. Through the use of Big Data, similar techniques can be applied to better understand businesses processes and environmental sustainability efforts. For instance, businesses might transform each domain-specific resource into the energy used to create that resource. This can then expressed in terms of what we call Energy Points, using the equivalent of the embodied energy of a gallon of gasoline as a unit and factoring in local parameters such as water scarcity. So instead of wondering what is the relative importance of 1,000 kgals and 1,000 kWh, businesses can simply treat each domain like Weight Watcher's treats calories – based on efficiency points. This is one way we collapse Big Data to a common metric to address resource consumption decisions. There are other benefits that come along with using Cloud computing. Companies are able to share and access real-time analytics and data sets, allowing progressive companies and organizations to release data to a broader ecosystem. A relatively simple task for IT, companies can exponentially increase efficiencies and provide the new material for building tangential business models and interactions. Cloud computing is playing an increasingly critical role in decision making, given simple and rigorous interpretation, which is the power to help transform the sustainability from an exercise in feel-good terminology to a quantifiable approach that has a true impact on our environment. The road to get there is paved with mathematics – developing intuitive, yet mathematically rigorous ways to interpret the data says Dr. Zik.

V. METHODOLOGY

In this study, a private cloud is built using Ubuntu and Eucalyptus open source software on two quad processor machines with 8GB Ram. Apache™ Flume used is a distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of streaming data into the Hadoop Distributed File System (HDFS). It has a simple and flexible architecture based on streaming data flows, and is robust and fault tolerant with tunable reliability mechanisms for failover and recovery. Flume lets Hadoop users make the most of valuable log data. Specifically, Flume allows users to:

 Stream data from multiple sources into Hadoop for analysis
l Collect high-volume Weblogs in real time
 Insulate themselves from transient spikes when the rate of incoming data exceeds the rate at which data can be written to the destination
 Guarantee data delivery

l Scale horizontally to handle the additional data volume

Flume’s high-level architecture is focused on delivering a streamlined codebase that is easy-to-use and easy-to-extend. The project team has designed Flume with the following components:

 Event – a singular unit of data that is transported by Flume (typically a single log entry)
 Source – the entity through which data enters into Flume. Sources either actively poll for data or passively wait for data to be delivered to them. A variety of sources allow data to be collected, such as log4j logs and syslogs.
 Sink – the entity that delivers the data to the destination. A variety of sinks allow data to be streamed to a range of destinations. One example is the HDFS sink that writes events to HDFS.
 Channel – the conduit between the Source and the Sink. Sources ingest events into the channel and the sinks drain the channel.
 Agent – any physical Java virtual machine running Flume. It is a collection of sources, sinks, and channels.
 Client – produces and transmits the Event to the Source operating within the Agent

A flow in Flume starts from the Client. The Client transmits the event to a Source operating within the Agent. The Source receiving this event then delivers it to one or more Channels. These Channels are drained by one or more Sinks operating within the same Agent. Channels allow decoupling of ingestion rate from drain rate using the familiar producer-consumer model of data exchange. When spikes in client-side activity cause data to be generated faster than what the provisioned capacity on the destination can handle, the channel size increases. This allows sources to continue normal operation for the duration of the spike. Flume agents can be chained together by connecting the sink of one agent to the source of another agent. This enables the creation of complex data flow topologies. Given the rigorous demands that big data places on networks, storage, and servers, it's not surprising that some customers would outsource the hassle and expense to the cloud. Although cloud providers say they welcome this new business opportunity, supporting in the cloud is forcing them to confront various, albeit manageable, architectural hurdles. The elasticity of the cloud makes it ideal for big data analytics -- the practice of rapidly crunching large volumes of unstructured data to identify patterns and improve business strategies -- according to several cloud providers. At the same time, the cloud's distributed nature can be problematic for big data analysis. "If you're running Hadoop clusters and things like this, they put a really heavy load on storage, and in most clouds, the performance of the storage isn't good enough," said Robert Jenkins, co-founder and chief technology officer of CloudSigma, a Zurich-based Infrastructure as a Service (IaaS) provider. "The big problem with clouds is making the storage perform to a level that enables this kind of computing, and this would be the biggest reason why some people wouldn't use the cloud for big data processing." But Jenkins and other cloud providers emphasized that these challenges aren't insurmountable, and many providers already have plans to tweak their cloud architectures to improve the capacity, performance, and agility of all their cloud services -- moves that they expect will also provide better support for big data in the cloud. Devising an architecture that supports big data analysis in the cloud is no more daunting than meeting the challenges of satiating the rapidly growing appetite for cloud services in general, according to Henry Fastert, chief technologist and managing partner at SHI International, a large reseller, managed service provider (MSP) and cloud provider based in Somerset, N.J.

VI. CHALLENGES OF CLOUD STORAGE

The cloud storage challenges in Cloud computing analytics fall into two categories: capacity and performance. Scaling capacity, from a platform perspective, is something all cloud providers need to watch closely. "Data retention continues to double and triple year-over-year because customers are keeping more of it. Certainly, that impacts us because we need to provide capacity," Corvaia said. Storage performance in a highly virtualized, distributed cloud can be tricky on its own, and the demands of Cloud computing analysis only magnify the issue, several cloud providers said.SHI International's cloud strategy is built on the company's vCore model, it's branding for a proprietary "finite collection of servers, storage, and switching elements," replicated across SHI's cloud, Fastert said. The distributed storage architecture enables SHI to "really optimize the performance of our infrastructure because it's set up in that granular fashion," he said. "Storage is something that's also impacted by specific types of virtualization loads, and so the way in which you spread tasks across your storage will always impact your performance," he said. "The [vCore] model allows us to spread loads based on the characteristics of those loads, and so we constantly look at the characteristics of customers' loads across our vCore infrastructure … and then we do load balancing across them from a storage-performance point of Hadoop is an open-source legend built by software heroes. Yet, legends can sometimes be surrounded by myths—these myths can lead IT executives down a path with rose-colored glasses. Data and data usage is growing at an alarming rate. Just look at all the numbers from analysts—IDC predicts a 53.4% growth rate for storage this year, AT&T claims 20,000% growth of their wireless data traffic over the past 5 years, and if you take at your own communications channels, its guaranteed that the internet content, emails, app notifications, social messages, and automated reports you get every day has dramatically increased. This is why companies ranging from McKinsey to Facebook to Walmart are doing something about big data.

Just like we saw in the dot-com boom of the 90s and the web 2.0 boom of the 2000s, the big data trend will also lead companies to make some really bad assumptions and decisions. Hadoop is certainly one major area of investment for companies to use to solve big data needs. Companies like Facebook that have famously dealt well with large data volumes have publicly touted their successes with Hadoop, so it's natural that companies approaching big data first look at the successes of others. A really smart MIT computer science grad once told me, "when all you have is a hammer, everything looks like a nail." This functional fixedness is the cognitive bias to avoid the hype surrounding Hadoop. Hadoop is a multi-dimensional solution that can be deployed and used in a different way.

VII. DISCUSSION

Dr. Zik(2009) suggested that the way to better analysis lies in mathematics and his studies found out that analytics would be improved by application of some mathematical modeling. In the studies of Lamonica (2013) which focused on crop agriculture information, improvement of inclusion of weather conditions that would greatly enhance its performance and be generally beneficial to all forms of agriculture. Similar studies would improve by integrating any big data analytics in the market today like Hadoop.This would bring business sense from the confused form of Cloud computing. Applications of sensors to track animals Andrea (2013) would be enhanced if such applications would be mounted on clouds so as to have a wide coverage accessibility. Computational techniques break down when applied to large unstructured data, Schadt et al (2010).This is due to traditional IS limited architectures a problem eliminated by cloud computing and analytics.Elasticity and scalability is the limiting factor of big data on traditional IS. This is sorted by Meng-Ju Hsieh et al (2011) by the use of MapReduce and SQL data processing methodologies. This also gives a solution to Abadi (2009) where cloud compatible databases are required.

VIII. CONCLUSION

The Cloud computing concept is surely the new elephant in the block. Organizations just have to embrace it to have a competitive advantage over their peers especially if they can integrate their big data into a cloud data analytics to produce a sensible relative output which can be used by different sectors positively. The huge deluge of data provides a great opportunity for business intelligence to those who will keep up with technology. There are still lots of research opportunity to study the cloud computing analytics and database frameworks. Environmental Cloud computing would doubtlessly enrich agricultural endeavors and quality of living.

IX. REFERENCES

[1]. Cloud Computing for Satellite Data Processing on High-End Compute clusters. The Golpayegani University of Maryland, Baltimore County golpa1@umbc.edu Prof. M. Halem University of Maryland, Baltimore County halem@umbc.edu

[2]. Large-scale Data Analysis on the Cloud Ranieri Baraglia1, Claudio Lucchese1, and Gianmarco De Francisci Morales1;2 1 ISTI-CNR, Pisa, Italy 2 IMT - Institute for Advanced Studies, Lucca, Italy

[3]. Massive Data Analytics and the Cloud, A Revolution in Intelligence Analysis by Michael Farber et al

[4]. Wagging the long tail of earth science: Why we need an earth science data web, and how to build it, Ian Foster, Daniel S. Katz, Tanu Malik Peter Fox Computation Institute Tetherless World Constellation University of Chicago Rensselaer Polytechnic Institute

[5]. SQLMR: A Scalable Database Management System for Cloud Computing Meng-Ju Hsieh; Inst. of Inf. Sci., Acad. Sinica, Taipei, Taiwan; Chao-Rui Chang; Li-Yung Ho; Jan-Jan Wu Published in: Parallel Processing (ICPP), 2011 International Conference on Date of Conference: 13-16 Sept. 2011 Page(s): 315 – 324 ISSN: 0190-3918 E-ISBN: 978-0-7695-4510-3 Print ISBN:978-1-4577-1336-1 INSPEC Accession Number:1231

[6]. Big data, but are we ready?: Correspondence by Schadt et al. | Review by Schadt et al. Oswaldo Trelles1,5, Pjotr Prins2,5, Marc Snir3 & Ritsert C. Jansen4, Author affiliations, Oswaldo Trelles is at the Computer Architecture Department, University of Malaga, Campus de Teatinos, E-29071, Spain. Pjotr Prins and Ritsert C. Jansen are at the Groningen Bioinformatics Centre, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands.

[7]. Wagging the long tail of earth science: Why we need an earth science data web, and how to build it, Ian Foster, Daniel S. Katz, Tanu Malik Peter Fox Computation Institute Tetherless

World Constellation University of Chicago Rensselaer Polytechnic Institute
[8]. Big data analysis in the cloud: Storage, network and server challenges by Jessica Scarpati
[9]. Computational solutions to large-scale data Management analysis by Schadt et al(2010) White Paper Big data: What is and Why You Should Care Sponsored by: AMD Richard L. Villars Carl W. Olofson Matthew Eastwood
[10]. https://www.ibm.com/blog/cloud-computing

Pratul Sharma
B.Tech (C.S.E), M.Tech (C.S.E)
R.B.S Engineering Technical Campus, Bichpuri, Agra
Research Area:- Cloud Computing, Big Data Analysis