Big Data Solutions with Cloud Computing: Recent Trends and Approaches

Sadhana Pandey
Assistant Professor
Department of Computer Science and Engineering
SAGE university Indore
(M.P) INDIA

Abhay Kothari, PhD
Professor
Department of computer Science and Engineering
SAGE university Indore
(M.P) India

Jyotsana Goyal
ASSISTANT PROFESSOR
Department of Computer Science and Engineering
SAGE university Indore
(M.P) India

ABSTRACT
Big Data and Cloud Computing has become trends and technologies of the day. The daily explosion of data means that it's better to have big data included in the applications. Whereas cloud computing is allowing users to use platforms according to their time, convenience and affordability. Cloud computing seems to be a perfect vehicle for hosting big data workloads. However, working on big data in the cloud brings its many challenges of reconciling two contradictory design principles. Cloud computing is based on the concepts of consolidation and resource pooling, but big data systems (such as Hadoop) are built on the shared nothing principle, where each and every node is independent and self-sufficient. The integration of big data with cloud, businesses and educational institution can have a better direction to the near future. Various analytics and technology involved in coupling of big data with cloud computing, the challenges involved in this process, trends applications of the domain and security issues involved have been discussed in this paper.

Keywords
Cloud Computing, Big Data, Efficiency,Virtualization, Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS).

1. INTRODUCTION
One of the main challenges in the context of Cloud computing is the management of very large volumes of data. This is totally independent of the resource type which is shared in the Cloud – data bases are either directly visible or accessible to clients as part of the Infrastructure, or are hidden behind service interfaces. Big Data needs a large amount of storage space. While the price of storage continued to decline, the resources needed to leverage big data can still pose financial issues for small to medium sized businesses. Cloud computing is an extremely successful paradigm of service oriented computing, and has revolutionized the way computing infrastructure is used Now a days. Three most famous cloud models include: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). The concept however can also be extended to Database as a Service or Storage as a Service. Platform as a Service (PaaS) provides a runtime environment. Programmers can easily create, test, run, and deploy web applications. Anyone can purchase these applications from a cloud service provider on a pay-as-per use basis and access them over the Internet . Infrastructure as a service (IaaS) provides the customer with high-level APIs used to dereference various low-level details of underlying network infrastructure like backup, data partitioning, scaling, security, physical computing resources, etc. software as a service is a cloud-based software delivery model that allows SaaS applications to run on SaaS providers' servers instead of installing and maintaining software on-premises. The SaaS provider is responsible for managing and access to the application, including security, availability, and performance. Elasticity, pay-peruse, low upfront investment, low time to market, and transfer of risks are some of the major enabling features that make cloud computing a universal paradigm for deploying various applications which were not previously feasible in a traditional enterprise infrastructure.

In this paper we have described both cloud computing and big data systems, focusing on the various issues and challenges. We discuss security issues when hiring a big data vendor: data privacy, data governance, and data heterogeneity; cloud data uploading methods; and how cloud computing speed and scalability poses a problem regarding exaflop computing. Our contributions to the current state-of-the-art is done by providing an overview over the issues to improve or have yet to be addressed in both technologies[2].

Data storage using cloud computing is a viable option for small to medium sized businesses considering the use of Big Data analytic. Cloud computing is on-demand network access to computing resources which are provided by an outside entity and require little management effort by the business or enterprises[3]. A number of architectures and deployment models exist for cloud computing, and these architectures or models are able to be used with other technologies and design approaches[4].

2. BIG DATA AND CLOUD COMPUTING
Big Data in cloud[5] refers to large size of the dataset perhaps in few dozens of terabytes and peta bytes and working with them in a local computer based Database Management System becomes much difficult. Furthermore, big data is viewed as a mean to store and process a large scientific datasets. So, Big data refers to the data which is huge in size and also increasing rapidly with respect to time[6]. Big data consist of structured data, unstructured data as well as semi-structured data. Big data is not possible to be stored and processed in traditional data management tools that is why it needs specialized big data management tools.
distinguishing

2.1 Characteristics of Big Data:

Variety of Big data – Structured, unstructured, and semi structured data
Velocity of Big data – Speed of data generation
Volume of Big data – Huge volumes of data that is being generated
Value of Big data – Extracting useful information and making it valuable
Variability of Big data – Inconsistency which can be shown by the data at times.

Volume means the size of datasets that a big data system deals with. Processing and storing big volumes of data is rather difficult. Velocity refers to the different rates at which data streams may get in or out the system and provides an abstraction layer so that big data systems can store data independently of the incoming or outgoing rate as and when required[7]. Value refers to the true value of data (i.e., the potential value of the data regarding the information they contain). Huge amounts of data become meaningless unless they provide value. Veracity implies the trustworthiness of the data, addressing data confidentiality, integrity, and availability. Organizations or enterprises are needed to ensure that data as well as the analysis performed on the data are correct.
Cloud computing has many useful features and advantages [8]. A standard definition of cloud computing, in which resources (e.g., CPU and storage devices) are provided as general utilities that can be leased and released by practitioners via Internet in an on demand basis. Cloud computing is the on-demand availability of computing resources as services over the internet. It removes the requirement for enterprises or organizations to procure, configure, or manage resources themselves, and they only pay for what they use. Infrastructure providers & service providers are the two service providers in a cloud computing environment[9] .The prominence of cloud computing has made a large impact on the Information Technology (IT) industry over the past decade, where majority of search engines such as Google, Amazon and Microsoft try to provide more powerful, reliable and cost effective cloud platforms, and business enterprises look to reshaping their business models to gain benefit from this new paradigm[10].

Types of cloud:

- Public cloud
- Private cloud
- Hybrid cloud
- Community cloud
- Multi-cloud

Public clouds are managed by third parties which offer cloud services via the internet to the public, these services are available as pay-as-you-go billing models. They offer solutions for minimizing IT infrastructure costs and become a good option for handling peak loads on the local infrastructure[11]. Public clouds are the go-to option for small enterprises, which can start their businesses without large investments by completely relying on public infrastructure for their IT requirements and services. Private clouds are distributed systems that work on private infrastructure and provide the users with dynamic provisioning of computing resources. A hybrid cloud is a heterogeneous distributed system formed by combining the various facilities of the public cloud and private cloud. Hence, they are also called **heterogeneous clouds[12].** Community clouds are distributed systems created by integrating the services of different clouds to address the specific needs of an industry, a community, or a business sector. But sharing responsibilities among the organizations is difficult. Multi-cloud is the use of multiple cloud computing services from different providers, which allows organizations to use the services for their specific requirements and avoid vendor lock-in [13].

This allows organizations to take advantage of the different features and capabilities offered by different cloud providers as and when required.

3. STORAGE MANAGEMENT AND CLOUD

Storing and processing big volumes of data requires scalability, fault tolerance and availability. Cloud computing delivers all these by the implementation of hardware virtualization[14]. Thus, big data and cloud computing are two compatible concepts as cloud enables big data to be available, scalable and fault tolerant at the same time. Business regards big data as a valuable business opportunity. As such, several new companies such as Cloudera, Hortonworks, Teradata and many others, have started delivering Big Data as a Service (BDaaS) or DataBase as a Service (DBaaS)[15]. Companies such as Google, IBM, Amazon and Microsoft also provide ways for customer to consume big data whenever required.
Next, we have discussed two examples, Nokia and RedBus, which describe the successful use of big data within cloud environments [16].

3.1 Nokia

Nokia was one of the first companies to understand the advantage of big data in cloud environments (Cloudera, 2012). Many years ago, the company used individual DBMSs to accommodate each application requirement [18]. However, realizing the advantages of integrating data into one application, the company decided to migrate to Hadoop-based systems, integrating data within the same domain, leveraging the use of analytics algorithms to get proper insights over its clients and various aspects.
As Hadoop uses commodity hardware, the cost per terabyte of storage was cheaper than a traditional RDBMS.
Since Cloudera Distributed Hadoop (CDH) bundles the most popular open source projects in the Apache Hadoop stack into a single, integrated package, with stable and reliable releases, it embodies a great opportunity for implementing Hadoop infrastructures and transferring IT and technical concerns onto the vendors' specialized teams. Nokia considered Big Data as a Service (BDaaS) as an added advantage and put a trust on Cloudera to deploy a Hadoop environment that copes with its requirements in a short period of time. Hadoop, and in particular CDH, strongly helped Nokia to fullfil their needs.

3.2 RedBus

RedBus is the largest company in India active in online bus ticket and hotel booking. This company wanted to implement a powerful data analysis tool to gain insights over its bus booking service. Its datasets could easily stretch up to 2 terabytes in size. The application would have to be able to analyze the booking and inventory data of thousands of bus operators serving many routes. On the other hand, the company needed to avoid setting up and maintaining a complex in-house infrastructure, extract-transform-load (ETL), data processing, data archival, and deletion. Before moving one or more stages of data lifecycle to the cloud, the following factors needs to be considered:
1. *Availability Guarantees*: Each cloud computing provider must ensure a certain amount of availability guarantees.

Transactional data processing requires quick real-time answers whereas for data warehouses long running queries are used to generate reports. Hence, one may not want to put its transactional data over cloud but may be ready to put the analytics infrastructure over the cloud.
2. *Reliability of Cloud Services*: Before offloading data management to cloud, enterprises want to ensure that the cloud provides required level of reliability for the data services. By creating multiple copies of application components the cloud can deliver the service with the required reliability of service.
3. *Security*: Data that is bound by strict privacy regulations, such as medical information covered by the Health Insurance Portability and Accountability Act (HIPAA), will require that users log in to be routed to their secure database server.
4. *Maintainability*: Database administration is a highly skilled activity which involves deciding how data should be organized, which indices and views should be maintained, etc.
One needs to carefully evaluate whether all these maintenance operations can be performed over the cloud data. Cloud has given enterprises the opportunity to fundamentally shift the way data is created, processed and shared. This approach has been shown to be superior in sustaining the performance and growth requirements of analytical applications and, combined with cloud computing, offers significant advantages.

4. SOLUTIONS AND METHODOLOGIES FOR DATA STORAGE

Although big data solves many current problems regarding high volumes of data, it is a constantly changing area that is already in development and still poses various issues. In this section we present some of the issues not yet addressed by big data and cloud computing. As the amount of data grows at a rapid rate, keeping all data is physically cost-ineffective[19]. Therefore, corporations must be able to create policies to define the life cycle and the expiration date of data (data governance). Moreover, they must define who accesses and with what purpose clients' data is accessed. As data moves to the cloud, security and privacy becomes very crucial which is the subject of broad research.
Within this section we provide an overview over these problems.

4.1 Security

a) Who is the owner of the data and who can access it?
In cloud service models the clients pay for a service and share their data onto the cloud. Now the question is, to which one of the two stakeholders does data really belong? Moreover, are provider permitted to use the client's data? What level of access can be implemented to it and with what purposes? Can the cloud provider benefit from that data?
b) Where is the data?
Sensitive data that is considered legal in one country may not be legal in another country, therefore, for the sake of the client, there must be an agreement upon the location of data, as its data may be considered illegal in some countries and lead to prosecution.

4.2 Privacy.

The harvesting of data and the use of analytical tools to mine information raises several privacy concerns. Ensuring data security and protecting privacy has become extremely difficult as information is spread and replicated at very fast pace. Analytics often mine users' sensitive or private information such as their medical records, energy consumption, online activity, supermarket records etc.

4.3 Heterogeneity

Big data refers to big volumes of data with different velocities (i.e., data comes at different rates depending on its source output rate and network latency) and great variety.

4.4 Data Governance

The belief that storage is cheap, and its cost is likely to decline further, is true regarding hardware prices. However, a big data DBMS does also needs to take into considerations other expenses such as infrastructure maintenance, energy, and software licenses.

4.5 Disaster Recovery

Data is a very valuable business and losing data will definitely result in losing its value. In case of emergency or hazardous accidents such as earthquakes, floods and fires, data losses must be minimal. To avail data, in case of any incident, data should be quickly available with minimal downtime and loss.
Although a large part of the data produced may be un-structured, relational databases have been the choice most organizations have to store data about their customers, sales, and products, among other things. As data managed by traditional DBMS ages, it is moved to data warehouses for analysis and for sporadic retrieval. Models such as MapReduce are generally not the most appropriate to analyze such relational data. Attempts have been made to provide hybrid solutions that incorporate MapReduce to perform some of the queries and data processing required by DBMS's . The researchers provide a parallel database design for analytics that supports SQL and MapReduce scripting on top of a DBMS to integrate multiple data sources. A few providers of analytics and data mining solutions, by exploring models such as MapReduce, are migrating majority of the processing tasks very closer to where the data is stored, and thereby trying to minimize surpluses of data preparation, storage, and processing. Data processing and analytics capabilities are moving towards Enterprise Data Warehouses (EDWs), or are being deployed in data hubs to facilitate reuse across various data sets . Another distinctive trend in Cloud computing is the increasing use of NoSQL databases as the preferred method for storing and retrieving information. NoSQL adopts a non-relational model for data storage and management. Many researchers argues that non-relational models have been avail- able for more than 50 years in forms such as object-oriented, hierarchical, and graph databases, but recently this paradigm started to attract more attention with models such as key-store, column- oriented, and document-based stores. The causes for such raise in interest, are better performance, capacity of handling unstructured data, and suitability for distributed environments .

5. MANAGEMENT OF RESOURCES AND DATA

5.1 Hadoop

Hadoop, is a free Java-based programming framework supports the processing of large sets of data in a distributed computing environment. It is a part of the Apache project sponsored by the Apache Software Foundation. Hadoop cluster uses a Master/Slave structure. Using Hadoop, large data sets can be processed across a cluster of servers and applications can be run on systems with thousands of nodes involving thousands of terabytes. Distributed file system in Hadoop helps in rapid data transfer rates and allows the system to continue its normal

operation even in the case of some node failures. This approach lowers the risk of an entire system failure, even in the case of a significant number of node failures. Hadoop enables a computing solution that is scalable, cost effective, and flexible and fault tolerant.

5.2 Map Reduce

Hadoop Map Reduce is a framework used to write applications that process large amounts of data in parallel on clusters of commodity hardware resources in a reliable, fault-tolerant manner. A Map Reduce job first divides the data into individual chunks which are processed by Map jobs in parallel. The outputs of the maps sorted by the framework are then input to the reduce tasks

5.3 Hadoop Distributed File System (HDFS)

Hadoop comes with a distributed file system called HDFS. In HDFS data is distributed over several machines and replicated to ensure their durability to failure and high availability to parallel application.

It is cost effective as it uses commodity hardware. It includes the concept of blocks, data nodes and node name.

HDFS Concepts

Blocks: A Block is the minimum amount of data that it can read or write.HDFS blocks are 128 MB by default and this is configurable.Files n HDFS are broken into block-sized chunks,which are stored as independent units.Unlike a file system, if the file is in HDFS is smaller than block size, then it does not occupy full block?s size, i.e. 5 MB of file stored in HDFS of block size 128 MB takes 5MB of space only.The HDFS block size is large just to minimize the cost of seek.

Name Node: HDFS works in master-worker pattern where the name node acts as master. Name Node is controller and manager of HDFS as it knows the status and the metadata of all the files in HDFS; the metadata information being file permission, names and location of each block.The metadata are small, so it is stored in the memory of name node, allowing faster access to data. Moreover the HDFS cluster is accessed by multiple clients concurrently,so all this information is handled bya single machine. The file system operations like opening, closing, renaming etc. are executed by it.

Data Node: They store and retrieve blocks when they are asked for; by client or name node. They report back to name node periodically, with list of blocks that they are storing. The data node being a commodity hardware also work as block creator, deletion and replication as stated by the name node.

6. RESEARCH CHALLENGES

Despite all of these advantages of the integration between cloud computing and big data, there are many challenges and risks that are required to be considered while deploying big data on a cloud environment. The fundamental issue that should be considered is the security of the big data cloud environment. There are some security vulnerabilities that arise because of integrating both and creating a new platform. One of the most known Big Data cloud security vulnerability is platform heterogeneity. One more challenge is the nature of data and its location, as in Big Data, the data can be in various locations. The cloud environment may include these locations or not. The type of processing that should be applied to the data, the parallelism of the processing, and where the processing should take place either the data is moved to a processing environment or the processing is performed on the location of the data. All of above challenges that needs to be taken into consideration while integrating the Big Data to a cloud system environment.

In addition to that, another challenge is the optimization of the Big Data cloud topology as it specifies the configuration, the size of the clouds, clusters and nodes that should be included to reach the optimal Big Data cloud model.

7. EXISTING SOLUTIONS

All large software vendors either have already started offerings in cloud space, or are in the process.

In addition there are many startups that have interesting products in cloud space. Here we have a list of major vendors of cloud computing.

Few of the cloud providers are google, citrix, netmagic, redhat, rackspace etc. Amazon (aws) is the most popular cloud provider amongst all. Microsoft is also providing cloud services.

Infrastructure as a Service cloud computing companies:

Amazon's offerings include S3 (Data storage/file system), SimpleDB (non-relational database) and EC2 (computing servers).

Rackspace's offerings include Cloud Drive (Data storage/file system), Cloud Sites (web site hosting on cloud) and Cloud Servers(computing servers).

IBM's offerings include Smart Business Storage Cloud and Computing on Demand (CoD).

AT&T's provides Synaptic Storage and Synaptic Compute as a service.

Platform as a Service cloud computing companies

Googles App Engine is a development platform that is built upon Python and Java.

com's provides a development platform that is based upon Apex.

Microsoft Azure provides a development platform based upon .Net.

Software as a Service companies

In SaaS, Google provides space that includes Google Docs, Gmail, Google Calendar and Picasa.

IBM provides Lotus Live iNotes, a web-based email service for messaging and calendaring capabilities to business users on demand basis.

Zoho provides online products same as Microsoft office suite.

As per [19] The group of private clouds should not only ensure the interoperability of various layers of infrastructures and collaborations. It should also comply with the range of security, legal and assurance requirements that deal with the data sharing and processing, satisfying particular requirements of public administrations.

As per this study main enablers that will facilitate the secure and transparent federation of private clouds in the public sector. These include models and architectures for access control in distributed environments, data and security policy languages and cryptographic methods [23] and services that support secure and efficient storage and processing of data in distributed and semi-trusted systems.

Then access control approaches were reviewed and architectures that enable data sharing and interoperability in connected clouds.

Bojan et. Al.[19] also reported that the collaboration in multi-layer federated environments requires the storage, sharing and processing of data that exhibits various levels of sensitivity, raising the issues of conformance with security and legislative requirements or obligations.

In the form of solutions to above mentioned challenges they presented the approaches that enable data sharing and

processing in outsourced environments that support different levels of trust and access granularity. The authors have considered two approaches of solutions: first- that enable the data sharing using advanced encryption schemas. Second- using novel architectures that deliver cryptographic services on various platforms and protocols.

8. BIG DATA APPLICATIONS

The technology known as Big Data is one of the most impactful innovations of the digital age. Patterns and correlations hidden in massive collections of data, revealed by powerful analytics, are informing planning and decision making across nearly every industry. In fact, within just the last decade, Big Data usage has grown to the point where it touches nearly every aspect of our lifestyles, shopping habits, and routine consumer choices.

8.1 Transportation

Big Data powers the GPS smartphone applications most of us depend on to get from place to place in the least amount of time. GPS data sources include satellite images and government agencies.

Airplanes generate enormous volumes of data, on the order of 1,000 gigabytes for transatlantic flights. Aviation analytics systems ingest all of this to analyze fuel efficiency, passenger and cargo weights, and weather conditions, with a view toward optimizing safety and energy consumption.

Big Data simplifies and streamlines transportation through:

- **Congestion management and traffic control**
  Thanks to Big Data analytics, Google Maps can now tell you the least traffic-prone route to any destination.
- **Route planning**
  Different itineraries can be compared in terms of user needs, fuel consumption, and other factors to plan for maximize efficiency.
- **Traffic safety**
  Real-time processing and predictive analytics are used to pinpoint accident-prone areas.

8.2 Advertising and Marketing

Ads have always been targeted towards specific consumer segments. In the past, marketers have employed TV and radio preferences, survey responses, and focus groups to try to ascertain people's likely responses to campaigns. At best, these methods amounted to educated guesswork.

Today, advertisers buy or gather huge quantities of data to identify what consumers actually click on, search for, and "like." Marketing campaigns are also monitored for effectiveness using click-through rates, views, and other precise metrics.

For example, Amazon accumulates massive data stories on the purchases, delivery methods, and payment preferences of its millions of customers. The company then sells ad placements that can be highly targeted to very specific segments and subgroups.

8.3 Banking and Financial Services

The financial industry puts Big Data and analytics to highly productive use, for:

- **Fraud detection**
  Banks monitor credit cardholders' purchasing patterns and other activity to flag atypical movements and anomalies that may signal fraudulent transactions.
- **Risk management**
  Big Data analytics enable banks to monitor and report on operational processes, KPIs, and employee activities.
- **Customer relationship optimization**
  Financial institutions analyze data from website usage and transactions to better understand how to convert prospects to customers and incentivize greater use of various financial products.
- **Personalized marketing**
  Banks use Big Data to construct rich profiles of individual customer lifestyles, preferences, and goals, which are then utilized for micro-targeted marketing initiatives.

8.4 Government

Government agencies collect voluminous quantities of data, but many, especially at the local level, don't employ modern data mining and analytics techniques to extract real value from it.The Centers for Disease Control tracks the spread of infectious illnesses using data from social media, and the FDA deploys Big Data techniques across testing labs to investigate patterns of foodborne illness. The U.S. Department of Agriculture supports agribusiness and ranching by developing Big Data-driven technologies

8.5 Media and Entertainment

The entertainment industry harnesses Big Data to glean insights from customer reviews, predict audience interests and preferences, optimize programming schedules, and target marketing campaigns.

Two conspicuous examples are Amazon Prime, which uses Big Data analytics to recommend programming for individual users, and Spotify, which does the same to offer personalized music suggestions

8.6 Healthcare

Big Data is slowly but surely making a major impact on the huge healthcare industry. Wearable devices and sensors collect patient data which is then fed in real-time to individuals' electronic health records. Providers and practice organizations are now using Big Data for a number of purposes, including these:

Prediction of epidemic outbreaks

Early symptom detection to avoid preventable diseases

Electronic health records

- Real-time alerting
- Enhancing patient engagement
- Prediction and prevention of serious medical conditions
- Strategic planning
- Research acceleration
- Telemedicine

Enhanced analysis of medical images

9. CONCLUSION

The application of Big Data in Cloud Computing certainly has a huge potential in the coming years. While using Software as Service, typically, big data plays significant role in giving

insight, in various cloud computing applications. Big Data when applied in cloud computing, has many applications in different areas. The major advantage with the cloud computing and big data integration is the data storage and processing power availability, the cloud has access to a large pool of resources and various forms of infrastructures that can accommodate such integration in the best suitable way possible; with minimum effort the environment can be set up and managed to allow an excellent work space for all the big data needs. The paper also introduced the characteristics, trends and challenges of big data. In addition to that, it investigates the benefits and the risks that may rise out of the integration between big data and cloud computing. We have also tried to provide an overview of big data in cloud environments, highlighting its advantages and showing that both technologies work very well together but also presenting the challenges faced by the both technologies.

10. REFERENCES

[1] Agrawal, Divyakant & Das, Sudipto & Abbadi, Amr. (2011). Big Data and Cloud Computing: Current State and Future Opportunities. ACM International Conference Proceeding Series. 530-533. 10.1145/1951365.1951432.

[2] Yadav S., Sohal A. (2017) "Review Paper on Big Data Analytics in Cloud Computing" in International Journal of Computer Trends and Technology (IJCTT) V49(3):156-160, July 2017. ISSN:2231-2803.

[3] Hariharan, U. & Kotteswaran, Rajkumar & Pathak, Nilotpal. (2020). The Convergence of IoT with Big Data and Cloud Computing. 10.1201/9781003054115-1.

[4] D. Borthakur, "The hadoop distributed file system: Architecture and design," Hadoop Project Website, vol.Ltd.

[5] "Big data: science in the petabyte era," Nature 455

[6] (7209):1, 2008.

[7] S. Ghemawat, H. Gobioff, and S. Leung, "The Google file system," in ACM SIGOPS Operating Systems Review, vol. 37, no. 5. ACM, 2003, pp. 29–43.

[8] Douglas and Laney, "The importance of 'big data': A definition,"2008.

[9] J. Dean and S. Ghemawat, "Map reduce: simplified data processing on large clusters," Communications of the ACM,vol. 51, no. 1, pp. 107–113, 2008.

[10] B. P. Rao, P. Saluia, N. Sharma, A. Mittal, S. V. Sharma, Cloud computing.

[11] D. Kossmann, T. Kraska, and S. Loesing, "An evaluation of alternative architectures for transaction processing in the cloud," in Proceedings of the 2010 international conference a. Rabkin and R. Katz, "Chukwa: A system for reliable largescale log collection," in USENIX Conference on Large Installation System Administration, 2010, pp. 1–15.

[12] ISO/IEC JTC 1. Information technology Big data, Preliminary Report 2014. ISO/IEC 2015.

[13] Marchiori, Massimo. (2017). Learning the way to the cloud: Big Data Park. Concurrency and Computation: Practice and Experience. 31. 10.1002/cpe.4234.

[14] Brevini, Benedetta. (2015). Book Review: To the Cloud: Big Data in a Turbulent World. Media, Culture & Society. 37. 1111-1113. 10.1177/0163443715596318a.

[15] R. Cumbley, P.Church, Is Big Data creepy? Comput.LawSecur.Rev. 29 (2013)601–609.

[16] Y. Cao, C. Chen, F. Guo, D. Jiang, Y. Lin, B. Ooi, H. Vo, S. Wu, and Q. Xu, "Es2: A cloud data storage system for supporting both oltp and olap," in Data Engineering (ICDE),2011 IEEE 27th International Conference on. IEEE, 2011,pp. 291–302.

[17] Reimair, F. & Feichtner, J., 2015. Attribute-based Encryption goes X.509. s.l., s.n.

[18] Reimair, F., Teufl, P., Kollmann, C. & Thaller, C., 2015. MoCrySIL - Carry Your Cryptographic Keys in Your Pocket. s.l., s.

[19] Bojan Suzica *, Andreas Reitera , Florian Reimaira , Daniele Venturib , Baldur Kuboc, Procedia Computer Science 68 (2015) 116 – 126" Secure Data Sharing and Processing in Heterogeneous Clouds," ScienceDirect. www.sciencedirect.com.