



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

Survey on Process in Scalable Big Data Management Using Data Driven Model Frame Work

Dr. G.V. Sam Kumar

Faculty of Computer Studies, Ministry of Education, Republic of Maldives

ABSTRACT: Data in rapid evolution scalable huge volume with multiple heterogeneous sources in all domains. The data grows expanding continue to grow as firms seek to integrate more source. The Big Data is been characterized increasing volume in the order of zetta bytes and their data generation velocity which shall climb market size upto 53.7 billion in 2017. The data objects need a scheme of efficiency modeling with security over the massive with vast distribution. This paper present the data driven frame work model processing on heterogeneous property of data characterized in scaling data encourages data demand routing data on validating security concerns with the distributed data. The higluc data driven model characterized the features of data revolution with processing data scaling and routing data immediately on validating privacy concerns proposed significantly outperform the state of processing.

KEYWORDS: Heterogeneous property, Data demand, Routing data, Data scaling

I. INTRODUCTION

Today is the era of information technology millennium, data are very huge quantity appear with wide variety of formats. The data management incrementing scalable with privacy preserving on efficient, data rapidly expanding in fields of domains as engineering, biological, physical, science and technology, even in the simple application used by every individual. In recent decade every individual person supposed to roughly create new data about 1.8 megabytes every second of every day. In 2017, even mobile traffic expected to grow to 10 Zetta Byte per month. Ability to process this kind of data demand analytics on all data. The data comes in many dimension and with variety of formats. Every individual, various domains and many application require the solid scientific foundation in design. New efficient, scalable and implementable algorithms are to meet this challenging task. The data structures also need to be understandable and scalable for these techniques. The new objects need modeling and simulation which are complex and massive in nature with vast distributions. Strong computational paradigms orienting towards distributed and parallel computing, data visualization and simulation are the need for the day. Data context in Big data is an imminent part due to the increasing number of users and devices. Efficient content-aware routing of data is required since, data driven involves demand-driven aggregation from sources, analysis, mining with the security and privacy consideration. The higluc data driven model algorithm that characterized the features of data revolution as with processing scaling data encourages routing data immediately on validating privacy concerns in the distributed data.

II. LITRATURE SURVEY

The data demand on size of the data largeness grow to span more and different data sources with variety of types. The scalable data becomes much complex and very hard to handle in process. Jeffrey Dean and Sanjay Ghemawat et al proposed model for processing data from the large data sets. The Map Reduce programming model can process the from high scalable and easy to use. Kuangyu Zheng, Xiaodong Wang et al proposed the optimization that the concept of workload correlation analysis with minimize the power consumption of server. Since the traffic consolidation from data center networks. Due to processing demanded data from massive scalable with optimal. Thus, the solution of frame work for finding demanded knowledge data from massive scalable nature needed of processing with maximum optimize model over the variety of types in real time ensures the quick response on dynamic occurrence of

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

many complex relation. The process has to be manage with regular data in real time in big data related application with quick response and nonstop services.

III. KEY FEATURES OF SCALABLE BIG DATA

The feature of scalable big data with huge volume, heterogeneous mixture of data, distributed data with autonomous nature, dynamic changes in complex relation. Over the time and usage data applications the key is to take the complex as non-linear, many-to-many data relationships along with their evolving changes discover pattern from the collected big data. In present and future scenario is to carry the data demand analytics on all types of data comes from multiple sources. The data retrieval and placement is biggest issue which calculate the latency time for data transition and migrate the needed data. So then optimization over the distributed environment is more important for the routing for the huge scaling data. Ever since the information impact in all the fields of usage the scaling factor grows many times fold. Many phases of different formats and types has to be correlated in analytical way to methodical energize the factor to routing from the source.

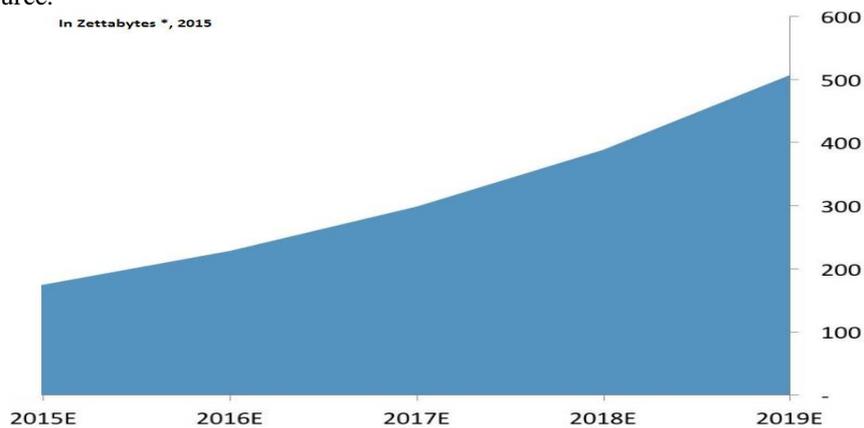


Figure 1 Data created by connected people and things worldwide, ie., 1 zettabytes = 1 trillion gigabytes

The data demands kept increasing in real time. The growth in data significantly increases global data traffic volume triggered new insight about data flow. In future things connected to the internet exceeded the number of people on earth. The rapid increase in data traffic continue to feature for process to increase operational efficiency becomes more critical for massive scaling data. Large data require fast and high performance processing for demanded data.



Figure 2 Global data traffic growth exa bytes per month growth increases 10 fold 2014 – 2019

IV. SCALABLE BIG DATA CHALLENGES IN PROCESSING AND ROUTING

Operational approach for the knowledge secure challenges with data availability, data access, data utilization and data stability. The necessary design model for the maintaining principle operation on data inflation, traffic and growth.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

A. HYBRID NEED OF USER SPECIFIC

The accurate analysis for the key generated respond to hybrid needs of use case specific. The factors of volumes, speed of data and aggregation level and frequency proceed for the data strategy. The procedure require computational intensive perspective on the diverse data sources. The data key enable the fusion mechanism and information exchange to ensure that appropriate data in domain knowledge. The key data knowledge produced coordinate with the data and computing process in sight of the capitalize relationships. The anonymous approaches are to use demand on the key propose on the hybrid need. The key generated based on categorical variable evaluate the demand criteria on data value of their availability and usability. The key generated on algorithms which maintained with structure for handling demand specific.

B. SEARCH MODEL IN HETEROGENEOUS AND DIVERSE DIMENSIONALITY

The routing criteria on heterogeneous mixture produce many groups with semantics and application knowledge. The appropriate group are determined with methods determined by the set of data with the relationships applicable. The group can be large data set are divided into many subsets which have dynamic nature from multiple information source. In the methods constraint are posed for reduction of inappropriate group and error produced in that group. The groups can be applicable for demand analytics on the data sharing, domain and application knowledge. The group anonymization approaches are to use generalization, suppression, permutation and perturbation to generate an altered version of uncertain data.

C. ADEQUATE DATA DEMAND ANALYTIC PROCESS

The data demand system of filter mechanism ensure the analytic mechanism on the anonymous distributed sites. Filtering and federated condition are mined on correlation ensure the models or patterns from multiple sources to meet the objective of data demand in the uncertain. The demand criteria insight the dimension for the size maintained in data set. The data group set process complex intrinsic semantics associations in data with high- performance clusters, technical capability in capturing and processing the data on regular stream within the time and reasonable cost limits. The method of optimizing with dynamic version the routing of labeled analytics and protecting suppression which deleting sensitive values, data swapping where switching values of sensitive data to prevent from matching and mining. Efficiency of source data discovery generated from values synthetically generated from distributions.

V. CHALLENGING ISSUES IN MINING WITH SCALABLE BIG DATA

A. RECOMMENDATION MODEL FRAMEWORK OF PROCESSING IN SCALABLE

The proposed high level framework mechanism from a multi-source perspective provides three layer where the first layer on key management on knowledge discovery, second layer represents the search model and third layer about data demand analytical process. The first layer sources the knowledge of deterministic aspect of knowledge, dynamic, privacy and application. The prediction key which capable of high accuracies and occurrence of the promising candidate process. The second layer search opinion on partitioning the working set database been used to minimize data transfers with compute node across different data domains. The analytic layer with group demand on generated set, API, dynamic data with the prediction constraint, cross source from different phase and with optimizing by using the principle of semantic generation of system supporting the structure which changes from the time to time on the demand needs on data and the usage of the data in process. The analytic third layer capable to adaptive mechanism on various sources like web logs, social network, media data and sensor data. The filtering process of analytic on item based and user based with the matrices of dynamic data. The data level on local data sources and global data distribution demand on the privacy, optimization with the structure basic. The search analytic access by value with the access control interface.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

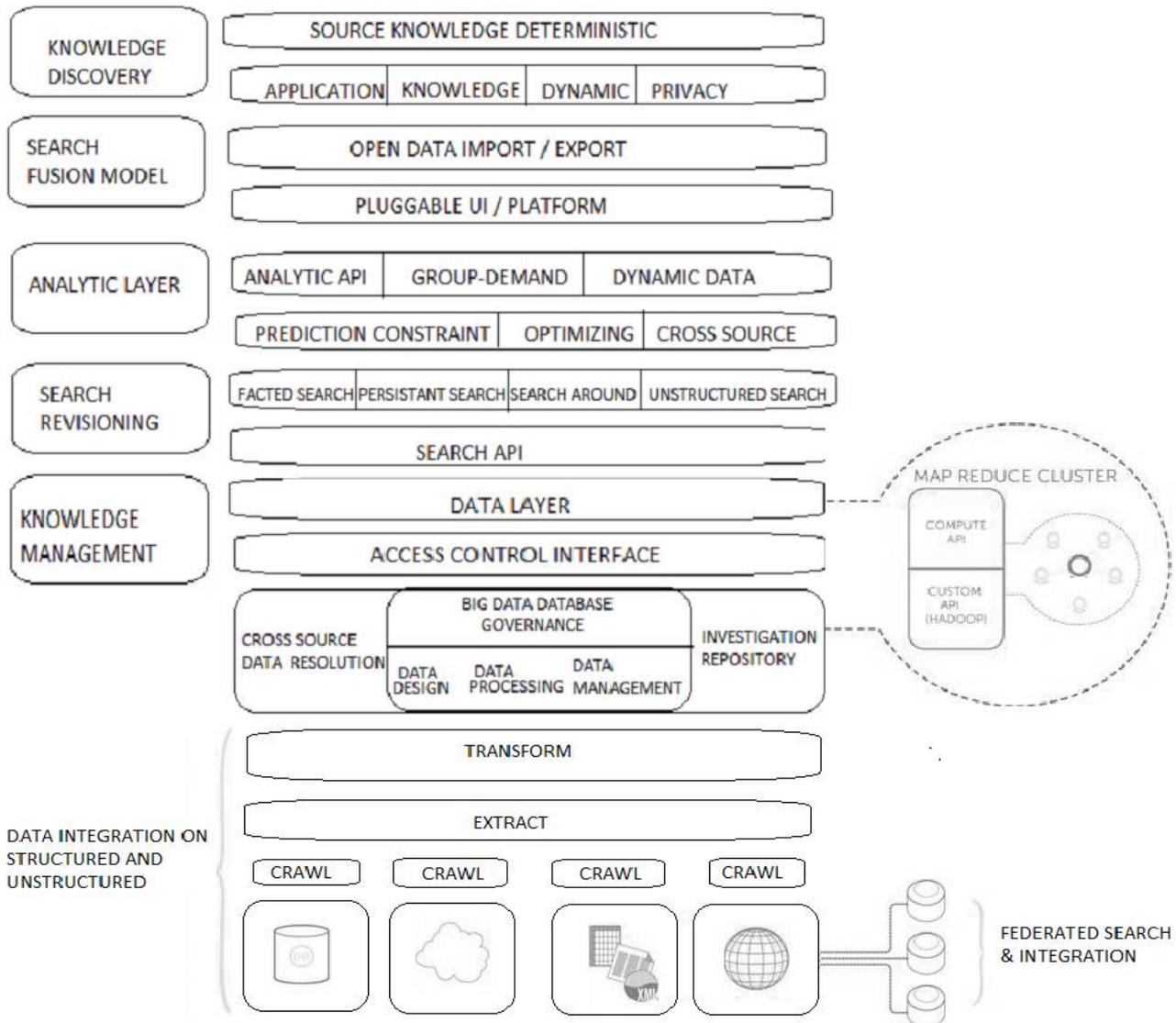


Figure 3 HighLuc frame work for process in scalable big data

The analytic property measures organise the data values for transform and extract from repository enable to ensure the correct domain data of having high dimensionality, uncertain and vagueness from the multiple data sources. The data from the different phase of have many formats with structured and unstructured between the distributed view. The process effective distributed set in vertical scaling for high demand rates of analytic and horizontal scaling over multiple servers.

VI. CONCLUSION

Inspired by data revolution in real world by recent, emerging trends in access from scalable big data and discussion on relevant open problems and directions important aspect to consider in current and future efforts . The system imposes



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

systematic design on essential factors such as autonomous with distribution on multiple sources. The data driven model manage to access on the demand data on growth of huge data volumes , heterogeneous and diverse data source with various formats on the maximum optimizing which essential for dynamic occurrence of complex relation The system designed to handle different structures, unstructured can be linked through the patterns and volume relationship of patterns to support the trend and future.

REFERENCES

- [1] C. Wang, S.S.M. Chow, Q. Wang, K. Ren, and W. Lou, "Privacy- Preserving Public Auditing for Secure Cloud Storage" IEEE Trans Computers, Vol. 62, no. 2, pp. 362-375, Feb. 2013.
- [2] J. Veiga, R. R. Expósito, G. L. Taboada, J. Touriño, "Analysis and evaluation of MapReduce solutions on an HPC cluster", Computers & Electrical Engineering, Vol. 50, pp. 200-216, 2016.
- [3] P. Jakovits, S. N. Srirama, "Evaluating MapReduce frameworks for iterative scientific computing applications", Proc. of the International Conference on High Performance Computing & Simulation (HPCS'14), pp. 226-233, 2014.
- [4] J. Samosir, M. Indrawan-Santiago, P. D. Haghighi, "An evaluation of data stream processing systems for data driven applications", Proc. of the International Conference on Computational Science (ICCS'16), pp. 439-449, 2016.
- [5] X. Wu and S. Zhang, "Synthesizing High-Frequency Rules from Different Data Sources," IEEE Trans. Knowledge and Data Eng., Vol. 15, no. 2, pp. 353-367, Mar./Apr. 2003.
- [6] J. Veiga, R. R. Expósito, G. L. Taboada, J. Touriño, "MREv: An automatic MapReduce Evaluation tool for Big Data workloads", Proc. of the International Conference on Computational Science (ICCS'15), pp. 80-89, 2015.
- [7] Das S., Sismanis Y., Beyer K.S., Gemulla R., Haas P.J., McPherson J., Ricardo, " Integrating R and Hadoop", In: Proceedings of the 2010 ACM SIGMOD International Conference on Management of data (SIGMOD '10), 2010, pp. 987-998.
- [9] H. Bal et al., "A medium-scale distributed system for computer science research: Infrastructure for the long term", Computer, vol. 49, no. 5, pp. 54-63, 2016.[10] M. Bertoni, S. Ceri, A. Kaitoua, P. Pinoli, "Evaluating cloud frameworks on genomic applications", Proc. of the 2015 IEEE International Conference on Big Data (IEEE BigData 2015), pp. 193-202, 2015.
- [11] K. Su, H. Huang, X. Wu, and S. Zhang, "A Logical Framework for Identifying Quality Knowledge from Different Data Sources," Decision Support Systems, Vol. 42, no. 3, pp. 1673-1683, 2006.
- [12] S. Huang, J. Huang, J. Dai, T. Xie, B. Huang, "The Hi-Bench benchmark suite: Characterization of the MapReduce-based data analysis", Proc. of the 26th IEEE International Conference on Data Engineering Workshops (ICDEW'10), pp. 41-51, 2010.
- [13] G. Cormode and D. Srivastava, "Anonymized Data: Generation, Models, Usage", Proc. ACM SIGMOD Int'l Conf. Management Data, pp.1015-1018, 2009.
- [14] M. Zaharia et al., "Resilient Distributed Datasets: A fault-tolerant abstraction for in-memory cluster computing", Proc. of the 9th USENIX Conference on Networked Systems Design and Implementation (NSDI'12), pp. 15-18, 2012.
- [15] J. Dean, S. Ghemawat, "MapReduce: Simplified data processing on large clusters", Communications of the ACM, Vol. 51, no. 1, pp. 107-113, 2008.
- [16] Y. Lindell and B. Pinkas, "Privacy Preserving Data Mining," J. Cryptology, Vol. 15, no. 3, pp. 177-206, 2002.

BIOGRAPHY

Dr. G.V. Sam Kumar received Ph.D in the Year 2013, He received master degree M.Tech [IT] and MCA degree in Computer Applications. Always passionate in teaching to the young minds in the field of Information Technology. He is working as Faculty of Computer Studies in Ministry of Education, Republic of Maldives. His research interest are in the area of Big Data. He has published technical paper in International conferences. Four papers in the International journal and International Conferences.