International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Using Semantic Query Analysis Integrating the Big Data: An Approach to Convert Big Data into Smart Data

Surya Kant Bansal[1], Naresh Kumar[2]

[1, 2]Amity Institute of Information Technology, Amity University, Noida (UP), India

Abstract: ***Late looks into on big data are demonstrating the difficulties identified with the semantic web: how to connection data on the web and joining the accessible data on web so that is made at high velocity in expansive volume in diverse variability or mixture that guarantee the quality and accuracy of that data so it can comprehend for people and machines. Individuals are investing a considerable measure of time to concentrate, convert and incorporating the data for their scientific reason however their master devices needs data into a particular organization and reshaped data so they can recovered better come about. It will empower the new advancement of the semantic web as joining, preparing, and dissecting all that data gets better and less expensive [3]. The destination to this paper to get the quick query reaction time by reshaping the data without access all data and select the parcel of data which to be look at or to change it into the brilliant data getting sagacity of data. However big data is characterized in term of 3v Velocity, volume and mixed bag yet shrewd data is indicated in term of volume just so it can bode well and might be broke down or settle on better choice for enhancing the result.***

Keywords: Big Data, Query analysis, Semantic web, RDF, JSON, WSDL, Data Wrangler

1. Introduction

All the database innovations lately being created for enormous data due to its drifting innovation, for example, Hadoop by Apache establishment, Mongodb or numerous other open source devices or the strategy or the engineering created by big data designers or new businesses, are indicating how the improvement of the semantic web as connecting, examining and transforming, all that data gets better and less expensive. It additionally indicates how big data reflects the semantic web or can help make the semantic web [1], [22]. Anyway issue is with big data where data is not in a few megabytes or gigabytes all things considered in terabytes or may be zeta bytes where it is not conceivable to get to and reviving the particular data and convert them into as indicated by necessity .The extensive volume of data is, no doubt produced step by step for each one movement that may be on Web and out in the open or private division and by their records [5]. There is need of smart data not to get to all data and select just a part of data to Process the inquiry so query could be quick and more adaptable.Figure-1 data transformation process show how to transform data into data reshaping model.

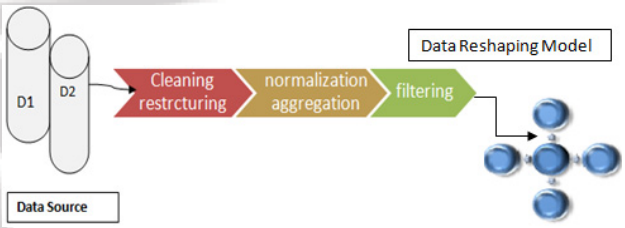

Figure 1: Data Transformation Process

2. Approach

We are proposing the new data transformation issue by permitting the client to characterize a right data approach on a little example of data and afterward execute that data plan into true data. To accomplish this we are creating an data transformation approach that might be handled on vast volume of true data which could be any sort of data structure like organized data or semi structure data and investigating the right Data reshaping plans so get to time could be spared. We are tending to the new approach that will create a semantic model of the given data source. This model maps the data source as XML, JSON, WSDL or organized database in the wake of displaying it is changed over into diverse sort of organization, for example, RDF and converts the data into smart data.

In this procedure firstly we perform convert the data utilizing a little partition of data and Afterward reshape the entire data (Big data) with smart data.Figure-2 defines using reshaping process converting big data into smart data.

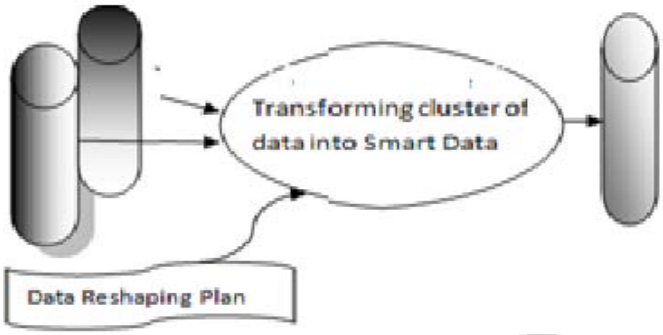

Figure 2: Extracting the smart data using a partition of data with data reshaping plan

3. How it will be Implemented?

Data change is a persistent procedure begins from concentrating the data, cleaning, standardization, sifting, rebuilding and conglomeration that change over data into semantic model. There will be an interface where client can straightforwardly cooperate with model or can adjust to their necessity firstly it spares the model data into general script then it will be executed on bunch of data.

Client gathers data from a few data source or commercial enterprises and will download the data to the interface called SNK. SNK is an intuitive stage where client can connect or change the semantic model .SNK is some piece of data reshaping model with the goal that it will naturally undercover .Source data into script data in type of their classes and records ,after that client can undoubtedly break down data of distinctive class or join together them in proficient and steady way.

In the process that SNK does not have any current semantic model it will create new query arrange and add it to with existing arrange and return once again to client. It maps data into class and lists in the metaphysics. For instance in the process that somebody gather data of DNA and genes of human it will maps them on the premise of sort of DNA and classes of genes.

After that it will make table outline for various query plan data may be put away in diverse table with distinctive area however their record class will be there for mapping on metaphysics at whatever point client execute any inquiry firstly it will seek in set of inquiry in the process that it victories to discover it, will return come about rapidly overall.

- Select test data
- Execute inquiry data
- Mapping with existing data or create new inquiry plan

There is an alternative of alteration to any semantic arrange so more solid example data will be mapped. In the further study test data will be made on the premise of client query with the goal that Manual procedure will be less and improvement will be progress. Smart data intends to speak to data in term of semantic. It makes data enlightening that could be reasonable by machine [27], [28]

4. Related work

We are approaching a great data coordination display however there is need of a ton of work to do get more adaptability and improvement. This model speak to the brilliant data however to expand versatility there are now numerous arrangements are executed or proposed one of then Google refine which take care of this sort issue yet into little partition of data and not on joining of data.[29,30]. Google refine works with grouped data, it perform cleaning, converting (one state to an alternate), amplifying it with web administration and interfacing it to other database [10]. Data Wrangler is additionally identified with it this is more systematic instrument instead of organizing device [11].

Both tools perform the accompanying capacities:

- Exploring The Data
- Cleaning And Changing Data
- Reconcile And Match Data

But both of them having inconvenience of numerous database combination

- Available For Just Structure And Non Structure Data[22]
- For Various Leveled Data Like Xml, Json, Rdf Does Not Work[4]
- More Time Devour Process
- Need Of Focal Vault
- Lack Of Joining Of Data
- Data Wrangler Is Self-Loader Process[24]

Some other related model likewise composed yet every one of them need more time intensive procedure to change the data and there is additionally need of a unified archive [8].

In today basically engineer or developers invest a ton of time to set up the data into data they are utilizing existing system or some data warehouse instrument yet these are not worth fit to work with big data semantics we have characterized a combination arrange that proposed data reshaping robotize plan.

Yet in this paper a superior methodology is presenting:

- Using all kind of data structure including various leveled data structure likewise
- No need of focal vault data could be anyplace just record mapping will be regular
- Less procedure time because of indexing
- Able to process on big data likewise

5. Conclusion

We accept that this methodology will diminish the inquiry reaction time and build a solid and legitimate reshaping arrangement .it will be give result at scale level so client can undoubtedly examination and envision the data and yield of one data treated as info of an alternate .For big data there is no as being what is indicated device that give result of all kind of database however Google refine is intuitive with

data ware house item which is rigid. Nobody give thusly simple methodology to build the data reshaping arrange its end to end answer for any issue. However still it is a approach there is a ton of future work must be carried out to build the reshaping arrange there is need of more enhance way such that less data ought to be get to when we plan the reshaping arrangement, it is not an improve approach to consider entire data .There is need of handy usage of this methodology in future.

References

[1] Big data: The next frontier for innovation, competition, and productivity. James Manyika,Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. McKinsey Global Institute. May 2011, Kandel, S.; Paepcke, A.; Hellerstein, J.; and Heer, J.(2011).
[2] Knoblock, C. A. Szekely, P. Ambite, J. L.; Goel, A. Gupta, S. Lerman, K.; Muslea, M. Taheriyan, M. and Mallick, P. 2012. Semi-automatically mapping structured sources into the semantic web. In the Extended Semantic Web Conference.
[3] Wrangler: interactive visual specification of data transformation scripts. 2011 annual conference on Human factors in computing systems, (CHI '11).
[4] Thusoo, A. Sarma, J. S. Jain, N.Shao, Z.; Chakka, P.Zhang, N.Antony, S. Liu, H. and Murthy, R. 2010. Hive -a petabyte scale data warehouse using hadoop technology. In the 26th International Conference on Data Engineering (ICDE), 996–1005.
[5] Tuchinda, R.; Knoblock, C. A.; and Szekely, P. 2011. Building mashups by demonstration. ACM Transactions on the Web (TWEB) 5(3).
[6] Wu, B.; Szekely, P.; and Knoblock, C. A. 2012. Learning data transformation rules through examples: Preliminary results. In 9th International Workshop on Data Integration on the Web
[7] Taheriyan, M.; Knoblock, C. A.; Szekely, P.; and Ambite J. L. 2013. A graph-based approach to learn semantic descriptions of data sources. In the 12th International Semantic Web Conference (ISWC 2013).'
[8] http://openrefine.org/
[9] http://vis.stanford.edu/wrangler/
[10] http://ebiquity.umbc.edu/blogger/2014/01/14/2014 ontology-summit-big-data-and-semantic-web-meet-applied-ontology/
[11] http://databasemanagement.wikia.com/wiki/Semantic_ Data_Model
[12] http://stko.geog.ucsb.edu/s4bd2013/
[13] http://www.ischool.berkeley.edu/newsandevents/events/ 20140226yingding
[14] http://www.ibmbigdatahub.com/blog/converting-big-data-and-analytics-insights-results-ibv-study-findings
[15] http://wiki.knoesis.org/index.php/Smart_Data
[16] http://www.fastcocreate.com/1682757/turning-big-data-into-smart-data-5-lessons-for-marketers-from-the-obama-campaign
[17] http://www.slideshare.net/apsheth/big-data-to-smart-data-keynote
[18] http://amitsheth.blogspot.in/2013/06/transforming-big-data-into-smart-data.html
[19] Navin Kabra, David J. DeWitt: An Object-Oriented Implementation for Extensible
[20] Database Query Optimization [J]. VLDB J. 8 (1):
[21] Yuan LIN, Hongfei LIN, Li HE, A Cluster-based Resource Correlative Query Expansion in Dis-tributed Data Retrieval, Journal of Computational Data Systems
[22] Michael J. Franklin, Donald Kossmann, Sukriti Ramesh, Reynold Xin. CrowdDB: Answering queries with crowdsourcing. Proceedings of the SIGMOD
[23] G. Ladwig and A. Harth. CumulusRDF: Linked data
[24] Management on nested key-value stores. In SSWS, 2011
[25] E. Gabrilovich and S. Markovitch, Computing semantic relatedness using Wikipedia-based explicit semantic
[26] Analysis, in *Proc. International Joint Conference on Artificial Intelligence*, 2001
[27] Akrivi Vlachou, Christos Doulkeridis, Neoklis Polyzotis. Skyline query processing over joins. Pro-ceedings of the SIGMOD 2011.
[28] Xavier Martinez-Palau, David Dominguez-Sal, Josep-Lluis Larriba-Pey. Two-way ReplacementSelection. (VLDB 2010).
[29] Li YE, Zhiguang QIN, Uncertain Range Queries for Revised Bead Model, Journal of Computational Data System
[30] Using Data for Systemic Financial Risk Management. Mark Flood, H V Jagadish, Albert Kyle, Frank Olken, and Louiqa Raschid. Proc. Fifth Biennial Conf. Innovative Data Systems Research, Jan. 2011.
[31] Pattern-Based Strategy: Getting Value from Big Data. Gartner Group press release. July 2011
[32] Understanding individual human mobility patterns. Marta C. González, César A. Hidalgo, and Albert-László Barabási.
[33] Computational Social Science. David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer,Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall
[34] Big data: The next frontier for innovation, competition, and productivity. James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. McKinsey Global Institute. May 2011.
[35] Materials Genome Initiative for Global Competitiveness. National Science and Technology Council. June 2011.