An Implementation of Big Data Processing to Separate the Payload Based on Classification Tree Model

G. Renukadevi, K. Selvakumar, S. Tamilarasan, S. Venkatakrishnan

Abstract: ***The process of distinguishing different types of data in the SQL server is the challenging task for further processing of big data. The big data is available in the Webpages, social media networks and cloud based web servers. In this implementation, the data can be retrieved from the cloud based web services. The data is temporarily posted in the REST API, and the data stored permanently in the SQL Server. The stored data is processed using the Classification Tree Model. Based on this method, the separation of types of payload is possible. With the help of this implementation, the types of the documents are automatically categorized using the trained data. Previously the training set has to be prepared for distinguishing different payloads and documents.***

Keywords : Big data Processing, Classification Tree Model, Separation of Payload, REST API, SQL Server, Machine Learning.

I. INTRODUCTION

The big data posting, storing, processing and preparing report based on the big data is the task in the web pages. There are many methods are available to process the big data in the web pages as well as in some SQL servers. It is a major application of IoT for preparing reports for future references and predictions. One of the methods in the big data processing is the Classification Tree Model, which helps to separate the different types of data and documents in the SQL servers and also it performs maintenance of user profiles in the social media and some web sites. The classification tree model also helps for grouping the similar kinds of payload for easy processing without using defragmentation. In this implementation, the payload and documents stored in the SQL Server is taken as a big data, and this Classification Tree Model separates the document type and categorizing the different payloads.

II. PROPOSED METHODOLOGY

The first step of acquiring data is to establish the access between REST (Representational State Transfer) and SQL Server. Then, the data posting in the REST web services can be automatically stored in the particular SQL Server. The REST API (Application Program Interface) can be easily accessed with the help of HTTP Gateway protocol. The HTTP gateway helps to access the REST API by GET the data, PUT the data, POST the data and DELETE the data. All the data and documents which are available in the REST API will be automatically stored in the SQL Server using some syntax and commands. The remote server can also accessed by the user for processing this big data. The Classification Tree Model uses machine learning for separating the different types of payloads and documents using the reference and the trained datasets.

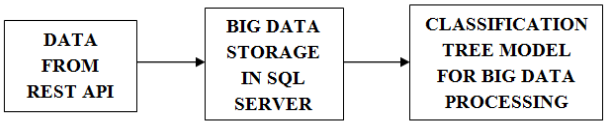

Fig. 1. Flow Diagram of the Proposed Method

The REST API is created for maintaining the profiles of different users enrolled in the online courses as an example.

Table 1. Registered candidates for Online Courses

User ID	Personal Profile	Academic Profile	Career Profile with Resume
	Course 1		
Student	1110	1103	+107
Engineer	1223	1214	+109
Professional	1197	1120	+277
Researcher	1134	1121	+113
Industrialist	1202	1210	+308
	Course 2		
Student	124	120	+104
Engineer	143	153	+210
Professional	13	111	+308
Researcher	19	14	+405
Industrialist	153	152	+501
Total	**6498**	**6408**	**2190**

The above table shows the created REST API for POST the data of users who enrolled in the online courses. The Classification Tree Model is used to process the above data using some decision tree and trained data.

Revised Manuscript Received on January 06, 2020.

G. Renukadevi, Research Scholar, Department of Computer Science & Engineering, Annamalai University, TN, India.
Dr.K.Selvakumar, Department of Information Technology, Annamalai University, TN, India.
S.Tamilarasan, Department of Computer Science, Bharathiar University, TN, India.
Dr. S. Venkatakrishnan, Department of Computer Science, Annamalai University, TN, India

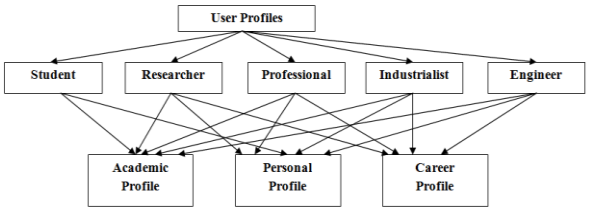

Fig. 2. Classification Tree Model for above example in REST API

III. RESULTS AND DISCUSSIONS

After retrieving the data from the SQL server, the processing has to be started. The method of accessing the remote server for processing is shown below,

```
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation.  All rights reserved.

C:\Users\Participants>mongo
MongoDB shell version: 2.6.5
connecting to: test
> _
```

Fig. 3. Connecting to SQL Server from REST API using test command

```
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation.  All rights reserved.

C:\Users\Participants>mongo
MongoDB shell version: 2.6.5
connecting to: test
> use bcc
switched to db bcc
> _
```

Fig. 4. Switching to SQL Server database

```
> var mapFunction1= function(){
... if(this.material == 'Fish')
... {
... emit(this.material,this)
... }
... }
>
> var reduceFunction1= function
... return values.length;
... }
>
> db.trips.mapReduce(
```

Fig. 5. Categorizing the similar payload using 'Fish' Command

```
        "result" : "map_reduce
        "timeMillis" : 27,
        "counts" : {
                "input" : 60,
                "emit" : 15,
                "reduce" : 1,
                "output" : 1
        },
        "ok" : 1
}
>
```

Fig. 6. Consolidated Output of the Classification Tree Model

IV. CONCLUSIONS

The payload and documents stored in the directory of the SQL Server is clearly distinguished with the help of Classification Tree Model used in the proposed method. The processing of big data made simple by this implementation. The server can store huge data, but the process of separating those payloads and documents are easily classified using this model.

REFERENCES

1. De Mauro, A., Greco, M. and Grimaldi, M. (2016), "A formal definition of Big Data based on its essential features", Library Review, Vol. 65 No. 3, pp. 122-135.
2. Lee J et.al, Intelligent prognostics tools and e-maintenance. Comput Ind 2006;57(6):476–89.
3. Murdoch TB, Detsky AS. The Inevitable Application of Big Data to Health Care. JAMA. 2013;309(13):1351–1352. doi:https://doi.org/10.1001/jama.2013.393
4. Lee J et.al, Prognostics and health management design for rotary machinery systems – reviews, methodology and applications. Mech Systems and Signal Proc; 2013. Article in Press.
5. Jay Lee, Edzel Lapira, Behrad Bagheri, Hung-an Kao,Recent advances and trends in predictive manufacturing systems in big data environment, Manufacturing Letters, Volume 1, Issue 1, 2013, Pages 38-41, ISSN 2213-8463, https://doi.org/10.1016/j.mfglet.2013.09.005.
6. Chui M et.al, The internet of things. Mckinsey quarterly No. 2; 2010.
7. National Institute of Standards and Technology. Workshop report on foundations for innovation in cyber-physical systems. January 2013.
8. Lee J et.al, TPM gets smart. Manufacturing. Engineering. June 2011, 156(6):70–5.
9. Lapira E. (2012) Fault detection in a network of similar machines using clustering approach. (Doctoral Dissertation). 2012.
10. Safran C et al; Expert Panel. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. J Am Med Inform Assoc. 2007;14(1):1-9.
11. Lee J, et.al, Predictive manufacturing system trends of next generation production systems. In: the Proceedings of the 11th IFAC workshop on intelligent manufacturing systems, vol. 11(1); 2013. p. 150–156.
12. MT Connect Overview. MT Connect Institute. Retrieved on September 4, 2013.
13. Lee J et.al, Predictive factories: the next transformation. Manufacturing leadership journal. Frost Sullivan; 2013.
14. Murff HJ et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. JAMA. 2011; 306(8):848-855.

AUTHORS PROFILE

G. Renukadevi, Research scholar, Department of Computer Science and Engineering , Annamalai University. She has graduated in Bachelor of Information Technology from Vivekanandha college of Engineering, Tiruchengode in May 2008 and received her Master's degree in Computer science and Engineering from Jayam college of Engineering, Dharmapuri in June 2012. Her area of Interest include Big data , Business Intelligence and Data Analytics. She has extended her interest in implementing Big Data Processing on classification tree model where the server can store huge data and the process of separating payloads and documents are easily classified.

Dr. K. Selvakumar, is currently working as a Professor in Department of Information Technology, Annamalai University . He has more than 20 years of teaching experience and a year of industrial experience. His area of Interest includes Cryptography and wireless networks. He has also published papers in many international and national journals and also attended many international conferences. His interests further extend on Intelligent systems, Routing protocols for mobile networks, Utilization of Wireless Multi-hop Networks for improving Transmissions Rate and Power, Maximal Entropy Design in Space Time blick Codes for future wireless technology.

S. Tamilarasan, Research scholar, Department of Computer Science, Bharathiar University. He has graduated in Bachelor of Sanskrit and Computer Science from Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya University, Kanchipuram in May 2003 and received his Master's degree in Computer science from Vinayaka Missions University, Salem in May 2013. His area of Interest include Cloud Analytics, IoT and Big data . He has extended his interest in a new approach using IoT sensor to oversight the smart home environment. His research area also include Big data Exploration, Visualization and Analytics.

Dr. S. Venkatakrishnan, Annamalai University. His area of specialization is Image processing. He has more than 15 years of Teaching and Research experience. His area of interest also extended to the applications of image processing in the field of science and technology including computer vision, remote sensing, feature extraction, forecasting and optical sorting. He has done a study on Image acquisition, storage, transmission, Image Enhancement and Restoration, Image Understanding and Image Recognition.