International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

IJIEMR Transactions, online available on 23rd Nov 2019. Link :http://www.ijiemr.org/downloads.php?vol=Volume-08&issue=ISSUE-11

Title **ANALYZE BIG DATA AND COLLABORATION RESEARCH**

Volume 08, Issue 11, Pages: 121–126.

Paper Authors

SHAIK ANJIMOON, MD.JOHN SAIDA

Eswar College of Engineering, Narasaraopet

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

ANALYZE BIG DATA AND COLLABORATION RESEARCH

SHAIK ANJIMOON[1], MD.JOHN SAIDA[2]
[1] PG Student, Eswar College of Engineering, Narasaraopet
[2] Asst.Professor, Eswar College of Engineering, Narasaraopet

Abstract
Does it make sense to make connections between big data and collaboration research? I was asked such a question during one of my presentations in China months ago when I mentioned that role-based collaboration (RBC)and the Environments-Classes, Agents, Roles, Groups, and Objects (E-CARGO) model might contribute to innovative research topics, including big data. I hope to clarify this opinion and answer that question.

1. INTRODUCTION
Today the term 'big data' draws a lot of attention, but behind the hype—especially for research organizations—there's a simple story. For decades, traditional research organizations have been collecting and working with data on a daily basis. As computing technology has evolved, so has the ability to gather, aggregate, analyze, and store increasing volumes of data. But even many of the most forward-thinking research technologists underestimated how fast these volumes of data would grow. In large part, the data deluge—or big data phenomenon—in research has been fueled by the proliferation of unstructured, non-traditional data generated through collaboration tools and social networking sources as well as the global sharing among researchers of observational data, simulation models, and experimental data. In addition, libraries have continued to digitize huge volumes of archived bodies of research that once were available only to a handful of researchers—or not at all.In addition, the cost of storage, compute power, and capacity has decreased, making it more affordable to aggregate, share, and analyze data in ways that may not have been feasible for many research organizations just a few years ago. And the proliferation of smart phones, GPS and other mobile devices have supported the immediacy of capturing observation and location research data—including large multi-media files.

2. LITERATURE SURVEY
Cloud computing has promoted the success of big data applications such as medical data analyses. With the abundant resources provisioned by cloud platforms, the QoS (quality of service) of services that process big data could be boosted significantly. However, due to unstable network or fake advertisement, the QoS published by service providers is not always trusted. Therefore, it becomes a necessity to evaluate the service quality in a trustable way, based on the services' historical QoS records. However, the evaluation efficiency would be low and cannot meet users' quick response requirement, if all the records of a service

are recruited for quality evaluation. Moreover, it may lead to 'Lagging Effect' or low evaluation accuracy, if all the records are treated equally, as the invocation contexts of different records are not exactly the same. In view of these challenges, a novel approach named Partial-HR (Partial Index Terms—big data, cloud, context-aware service evaluation, historical QoS record, weight Historical Records-based service evaluation approach) is put forward in this paper. In Partial-HR, each historical QoS record is weighted based on its service invocation context. Afterwards, only partial important records are employed for quality evaluation. Finally, a group of experiments are deployed to validate the feasibility of our proposal, in terms of evaluation accuracy and efficiency.

The existing work either only considers partial context elements, or lacks quantitative weight model for historical QoS records. Therefore, it becomes a challenging task to develop a quantitative weight model that considers all the context elements, for evaluating the quality of big data services accurately and efficiently. In view of this challenge, a novel service evaluation approach Partial-HR is proposed in this paper. Partial-HR not only considers all the important context elements of service invocation (i.e., invocation time, input size and user location), but also satisfies the Volatility Effect and Marginal Utility. Through Partial-HR, we can select partial important historical QoS records for service evaluation, so that the evaluation accuracy and efficiency could be improved. Through a set of experiments, we validate the feasibility of our proposal.

In cloud environment, the advertised QoS information of big data services is not always trusted. Therefore, it becomes a necessity to evaluate the service quality based on historical QoS records. Today, many researchers have studied this problem and given their proposals. In the problem of QoS credibility is firstly put forward, and the historical QoS records are suggested to be considered for evaluating the real quality of service. In the literature the service's QoS credibility is calculated, by comparing the historical QoS data with the SLA (Service Level Agreement) promised by service providers. Afterwards, it became popular to utilize the historical QoS records of services for various trustable service-oriented applications, such as service recommendation, service evaluation, service selection and service composition. However, in the above literatures, the weight problem of different historical QoS records is discussed.

Due to the unstable network or fake advertisement, the QoS information of services that process big data in cloud, is not always trustable as advertised by service providers. Therefore, it becomes a necessity to evaluate the service quality in a trustable way, based on the historical QoS records. However, it may lead to low efficiency if all the records are considered in service quality evaluation. Moreover, evaluation accuracy would be low if all the historical QoS records are treated equally, as their service invocation contexts are not exactly the same. In view of these challenges, a novel evaluation approach named Partial-HR is proposed in this paper, which not only considers the service invocation context, but

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

also satisfies 'Volatility Effect' and 'Marginal Utility' simultaneously. Through a set of experiments, we validate the feasibility of Partial-HR in terms of evaluation accuracy and efficiency. In the future, we will introduce more context elements into our weight model for historical QoS records, so as to further improve the evaluation accuracy of big data ser-vices in cloud.

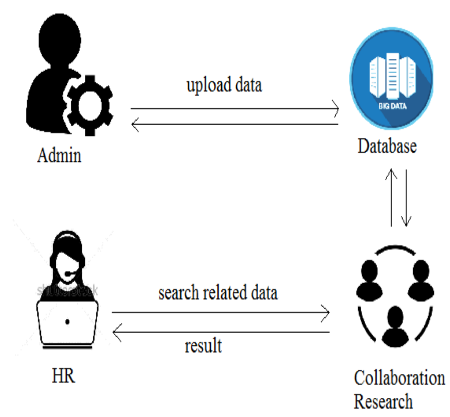
3. EXISTING SYSTEM

Collaboration is required when there is a task that cannotbe accomplished by an individual. Even highly intelligentindividuals may not handle collaboration well enough,and learning how to collaborate successfully with otherscan be a life-long task. It is also necessary to systematicallyand thoroughly investigate the challenges present in collaboration.Therefore, theories, technologies, practices,and systems should be proposed, proved, tested, verified, and validated to support collaboration. This requirementimposes a grand challenge to the IEEE Systems, Man, andCybernetics Society, and it is our responsibility to face thischallenge by putting significant effort into investigatingthis matter deeply and broadly.

4. PROPOSED SYSTEM

Collaboration, or most collaborative activities, is complex because it is difficult to express using formal or mathematical expressions. It is almost impossible for us to handle all of the related elements, because each, such as intelligence, emotions, or people, is complicated enough for a group of researchers to investigate during their whole lives. Therefore, dividing and conquering is the fundamental way to perform collaboration research, and we need to provide a pertinent, abstract level and concentrate on the aspects that can be handled by individual researchers or a well-formed team. Big data brings new challenges and opportunities for collaboration research.

5. ARCHITECTURE



6. IMPLEMENTATION

Collaboration

Collaboration, or most collaborative activities, is complexbecause it is difficult to express using formal or mathematicalexpressions. It is almost impossible for us tohandle all of the related elements, because each, such asintelligence, emotions, or people, is complicated enough fora group of researchers to investigate during their wholelives. Data collaboration means visualizing data from all relevant different data sources and delivering them to the right people, in the right format, and in time to use them in making effective decisions [2]. In this article, data collaboration is used to describe the process that transforms scattered large data sets into meaningful data sets, i.e., big data. Collaboration research manages a complex

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

world that includes physical objects, resources, people, information, intelligence, capital, emotions, and organizations.

Real Big Data

Big data brings new challenges and opportunities for collaboration research. Big data is defined as data sets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze. Data sets need to be developedinto big data through collaboration. Consequently,these matters reveal a connection between big data andcollaboration research. Theimmense quantity of big data changes the property ofsheer raw data. Therefore, big data is data mapping of thereal-world activities, i.e., collaboration among people andtheir environments. If datasets may produce the class forobjects, then they are big data.Large volumes of duplicated dataare not to be considered big data.

Collaboration Model

From the aforementioned discussion, big data presentsboth challenges and resources. If we consider big datachallenges, we need innovative methodologies or technologiesto process data, i.e., data collaboration. If we take bigdata as resources, we need to mine from data those elementsthat are valuable for collaboration. In both aspects,RBC and E-CARGO are promising contributions. Based onthis requirement, the E-CARGO model is an effectivemodel to support data collaboration. With the E-CARGOmodel, a system S can be defined as a nine-tupleR : : = < C,O,A,M,R,E,G,S0,H >, where C is a set ofclasses, O is a set of objects, A is a set of agents, M is aset of messages, R is a set of roles, E is a set of environments,G is a set of groups, S0 is the initial state of thesystem, and H is a set of users. In such a system, A andH and E and G are tightly coupled sets.

7. ALGORITHM

1.MapReduce

The MapReduce algorithm contains two important tasks, namely Map and Reduce.

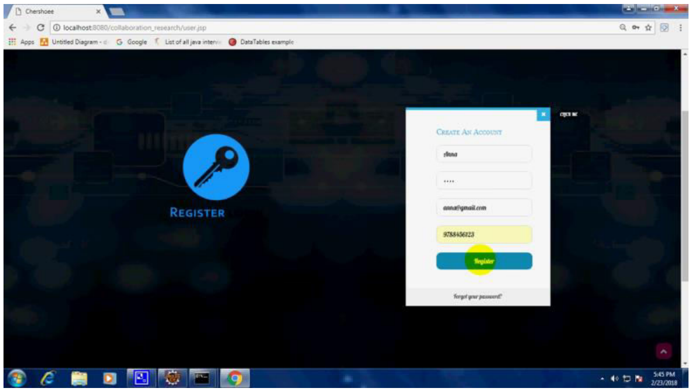
- The map task is done by means of Mapper Class
- The reduce task is done by means of Reducer Class.

Mapper class takes the input, tokenizes it, maps and sorts it. The output of Mapper class is used as input by Reducer class, which in turn searches matching pairs and reduces them.

2.Clustering

Hierarchical clustering involves creating clustersthat have a predetermined ordering from top to bottom. For example, all files and folders on the hard disk are organized in a hierarchy. There are two types of hierarchical clustering

8. RESULTS



9. CONCLUSION

Agent evaluation can be versatile and broad. There aremany aspects for us to evaluate, and the evaluation of differentaspects highly affects the optimal role assignment in thesense of group performance. Data mining from big data producesindicators to evaluate agents on roles, and datacollaboration may provide more big data for evaluation. Big data is thesource for us to extract such cooperation/conflict factors.Role execution and transfer can also acquire benefits frombig data, because big data provides patterns, rules, and conditionsfor role transfer and role execution.In summary, big data brings in the fundamental componentsof collaborative systems. At the same time, datacollaboration creates real big data. It is valuable toapply big data to collaboration research and to applycollaborative technologies to data analysis to make bigdata. The conclusion is that collaboration needs data,and data require collaboration to be big data.

REFERENCES

[1] R. Fuller. (2014, Jan. 29). Leveraging big data to redefine collaboration in the enterprise. Digitalist Mag. [Online]. Available: http://www.digitalistmag.com/ technologies/big-data/leveraging-big-data-redefine-collaboration-enterprise-01243242

[2] Honeywell. (2011, July). Data collaboration turning data into decisions. Advanced Solutions. [Online]. Available: https://www.honeywellprocess.com/library/ marketing/ whitepapers/HoneywellOperationalInsight_ DataCollabora tion_WP636.pdf

[3] E. Kitzelmann, "Inductive programming: A survey of program synthesis techniques," in Proc. Approaches and Applications of Inductive Programming Third Int. Workshop, 2009, pp. 50–73.

[4] J. Manyika, M. Chui, B. Brown, J. Bughin, R. D. Charles Roxburgh, and A. H. Byers. (2011, May). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. [Online]. Available: http://www.mckinsey.com/insights/ business_technology/big_data_the_next_frontier_for_innovation

[5] D. Partridge, "The case for inductive programming," Comput., vol. 30, no. 1, pp. 36–41, Jan. 1997.

[6] M. Swan, "The quantified self: Fundamental disruption in big data science and biological discovery," Big Data, vol. 1, no. 2, pp. 85–99, June 2013.

[7] J. M. Tien, "An SMC perspective on big data," IEEE Syst., Man, Cybern. Mag., vol. 1, no. 2, pp. 27–29, Apr. 2015.

[8] H. Zhu and M. C. Zhou, "Role-based collaboration and its kernel mechanisms," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 36, no. 4, pp. 578–589, July 2006.

AUTHORS PROFILE

SHAIK ANJIMOON is a student pursuing M.Tech(CSE) in Eswar college Of Engineering, Narasaraopet, Guntur

MD.Jhon Saida M.Tech in Computer Science & Engineering. He is currently working as an Asst Professor in Eswar

College of Engineering, Narasaraopet, Guntur, India. He is having about 10 years of teaching experience in different Engineering Colleges.