Check 26. Measurement Metric Proposed For Big.pdf

by Arta Sundjaja

Submission date: 17-May-2019 02:55PM (UTC+0700)
Submission ID: 1102513921
File name: 26._Measurement_Metric_Proposed_For_Big.pdf (215.03K)
Word count: 3909
Character count: 21335

Measurement Metric Proposed For Big Data Analytics System

Ridha Sefina Samosir
Doctor of Computer Science[1]
Information System Study Program[2]
Bina Nusantara University[1]
Institute Technology and Business Kalbis[2]
Jakarta, Indonesia
ridha.samosir@binus.ac.id[1]
ridha.samosir@kalbis.a.id[2]

Harco Leslie Hendric[1], Ford Lumban Gaol[2], Edi Abdurachman[3]
Doctor of Computer Science
Bina Nusantara University
Jl. Kebon Jeruk No.27,
Jakarta, Indonesia
shendric@binus.edu[1],
fgaol@binus.edu[2],
edia@binus.edu[3]

Benfano Soewito
Binus Graduated Program
Bina Nusantara University
Jl. Kebon Jeruk No.27, Jakarta
Indonesia
bsoewito@binus.edu

ABSTRACT

Big data is defined as a very large data set (volume), velocity and variety. Big data analytics systems must be supports for parallel processing and large storage. The problem of this research is how to identify measurement metric based on big data analytics system characteristic. One device that support big data platform is Hadoop. Measurement is a process for assigning values or symbols to the attributes of an entity. The purpose of measurement is to distinguish between entities one to another. Indicator for software measurement represented with a metric. The aim of this research is to proposes some measurement metric for big data analytics system. This research using UML exactly a class diagram in system modelling to identify the measurement metric. Both of dynamic and static metric is proposed as solution to measure big data analytics system. Result for this researh are some measurement ndicator both of dynamic and static metric based on class diagram for big data analytics.

CCS Concepts

• Software and its engineering→ Software creation and management • Design Software Software Design Engineering.

Keywords

Metric; Measurement; Big Data Analytics; Software

1. RESEARCH BACKGROUND

Big Data terminology is now widely used as a research issue. The rapid development of data acquisition technology is currently one contributor to the emergence of data in the form of big data. There are so many current data acquisition devices like sensors, internet, multimedia and more. Various data sources are producing data in addition to very large also very fast changes as well as many variations of its data format [1]. Older platforms or approaches are perceived to be incompatible with the size, velocity and variety of data. So the research for develop a flatform and algorithm for big data processing is growing nowadays. Processing of a set of data that meets the criteria of volume, variety and velocity requires devices that can work very quickly. One application for super computers that can be used for big data analaytics is Hadoop. Hadoop is an open source software that licensed by Apache. Hadoop are divided into two parts, namely the data storage and data processing. The data storage part is called Hadoop Distributed File System (HDFS). For data processing, Hadoop using map reduce algorithm. HDFS system in hadoop devide large files into smaller parts and then distributed to cluster system. This HDFS file system is designed to be fault tolerance and supports very large data set processing. Fault tolerance means that on a computing system if any node is damaged or dead then another node will takes over the process without disrupting the service. Support for large data set processing is accordance with the Hadoop method. Files processed by HDFS, devided into smaller pieces and distributed to multiple nodes in the system cluster. Data is stored in the cluster. One cluster consists of many nodes or servers. Every node in the cluster must be installed with hadoop

Measurement is a process of assigning a value or symbol to an attribute of an entity in order to distinguish between entities. Measurement also aims to describe the entity in accordance with clear rules. For Example, mileage and speed of our vehicle is usually used when we measure our journey. The amount of distance and speed becomes a measure that can be used to explain the journey. Measurement results can be used for analysis or conclusions related to the entity [2]. In generally, software measurement based on the needs of users of the system, graphical user interface design of the system and System completion based on specified time. Software measurement can be done for every stage in software development such as analysis, design, implementation and maintenance [3].

According to explanation about big data above then it will be suite if we use object oriented approach for develop the big data analytics system. Its because, big data means volume, variety and velocity of data, . If we talk about large volume and variety of

CSAI 2017, December 5–7, 2017, Jakarta, Indonesia

ACM ISBN 978-1-4503-5392-2/17/12…$15.00
DOI: https://doi.org/10.1145/3168390.3168425

data then we need encapsulation to cover all kind data. We need inheritance too because there is same characteristic among data both attribute and method. Polymorphisme needed while we choose the operation for data.

One stages in software development is system modelling. Unified modelling language (UML) can used for system modelling. There is some schema in UML there are use case, activity, class diagram, sequence diagram and etc. For whole of diagram in UML, there are three diagram are most widely used: use case, sequence and class diagram. Class diagram is used to visualize all class and relation among the class involved to the system. From related work of this research showed that design model give some advantages while it used for metric measurement identify there are: relatively in expensive and likely much more practical in nature, event be adapted to work on program component, create valid result, early decision making and all possible runtime information will be reported [4]. Booch [5] introduced four major step in OO design process:

1. Identification of Classes (and object)
2. Identify semantec of classes (and object)
3. Identify relationship between classes (an object)
4. Implementation of classes (and object)

Chidamber and Kemerer summarizes the three static object oriented metric above into six part, there are :Weighted method per class (WMC), Response For Class (RFC), Lack of cohesion (LCOM), Coupling between object (CBO), Depth of inheritance tree (DIT), Number of children (NoC)

Previous research showed that static approach have some weaknesses like less accurate, observe mainly program structure, lacks in handling of object oriented programming features, and relatively less amount of data to handle. Static metric are aimed at design phase in OO development not specific at implementation phase [6]. Static metric have been useful in predicting quality attribute like fault proneness and maintainability [7].

To handle all weaknesses above, some research proposes dynamic metrics for object oriented software like number object created during exsecution, number of coupling at run time and number of cohesion at run time [8]. An object is representation of class and occupies memory spaced at run time. A class instantiated of objects at run time depending upon of specific execution scenario simulated by an input data and events. During execution of program, a class can be instantiated using more than one object. Dinamic specific on implementation phase. Here is dynamic metric proposed from previous research:

1. Number of object created at run time
2. Number cohesion created at runtime [9]

Based on object oriented and big data analytics explanation then we proposes metric measurement combination both of static and dynamic for big data analytics system. This combination metric is for complete one each other. Static is used to measure complexity trough design stage and dynamic metric would be used while the program executed.

Content structure for this research start from background of research in section 1, literature review in section 2, research result in section 3and conclution in section 4.

2. LITERATUR REVIEW

In this section will described some concept related to this research they are big data, measurement, object oriented, UML and class diagram.

2.1 Big Data

Big data is defined as a very large set of data and meets at least three characteristics of data: high volume, velocity and variety. This three characterictic called as 3V concept. The volume characteristics are defined as amount of very large data. Its usually in the Terabytes size or more. The second characteristic is variety, big data is a combination of large data consist of various formats both structured and unstructured. Velocity is a representation of data with high frequency of data changes. In other words, data growth is very fast in very large numbers and a short period of time. For example are social media data, customer, bank and promotion data. Form three characteristic of big data will affect to the storage (storage) and speed. The following are challenges in big data technology: data capture capabilities, data storing capabilities, related capabilities to ensure data curation and the ability to search, share, transfer, analyze and visualize big data. Other concept about big data called as 5V which consist of volume, velocity, variety, veracity, and value. Veracity means data quality. Big data analytics must ensure that data used is correct or valid. And Value concept in big data analytics means that obtained data from analytics process gives many value for other process.

Big data analytics is some process to manage data with large volume, data with many format and data with high frequency growth. Trough this process will be obtained some pattern of data, correlation between data and other valuable information. Technology for big data analytics is needed today in line with increasing data growth. Technology sophistication is one of triggers of data growth. There are so many product offering products for big data analysis like IBM, SAS, PENTAHO, AMAZON, Goole and others.

2.2 Software Measurement

Measurement is a process of assigning a value or symbol to an attribute of an entity in order to distinguish between entities. Measurement also aims to describe the entity in accordance with clear rules. For Example, mileage and speed of our vehicle is usually used when we measure our journey. The amount of distance and speed becomes a measure that can be used to explain the journey. The unit of measurement is called the scale. Measurement results can be used for analysis or conclusions related to the entity [10]. Software measurement has an important role in software development. In generally, software measurement based on the needs of users of the system, graphical user interface design of the system and System completion based on specified time. Software measurement can be done for every stage in software development such as analysis, design, implementation and maintenance [11]. Measurements are made based on predetermined metric measurements. Metrics are the empirical value of an entity that aims to describe the specific character of an entity. Software metrics can help design development, prediction, final assessment of product quality, data utilization, cost estimation, system error prevention and reduced testing time. In the end, can be used to develop better software for the next time [12].

2.3 Unified Modelling Language

Unified modelling language (UML) is a diagram which used for system modelling in object oriented approach. Object oriented (OO) is one of approach in software development. OO have some characteristic there are encapsulation, polymorphism, and inheritance. Encapsulation means a process to cover many item in one unit. Encapsulation cover attribute and method in one class. OO is one of the most successfull techniques to design and implementation applications and computer programs. The purpose of this process is for information hiding from external environment. There is some diagram in UML there are use case, activity, class diagram, sequence diagram and etc. For whole of diagram in UML, there are three diagram are most widely used: use case, sequence and class diagram. Class diagram represents structural view, use case represents functional view and sequence diagram represents behavioral view. Use case diagram, visulize all function and actor do in system. Trough use case diagram, we can identify all system function or how many modul will be created. Class diagram represents all class and relationship among the class. With class diagram, we know all attribute and method involved in the system. Attribute means as instance of variable or class while method means system behaviour.

2.4 Class Diagram

Class diagram is used to visualize all class and relation among the class involved to the system. A class define set of object with same characteristic. Its mean an object with same attribute and method. A class is a blue print of objects. A class encapsulates two elements there are attribute and method. Attribute is instance of variable while method means class behaviour. Class diagram also visualize the relationship among the class called multiplicity. Atrribute means value or characteristics from the object of class like name for class of students while method explaine all function will be operated to the class. For example that registration can be method for student class.

Design of classes declared to be central to the OO paradigm. Class design is highest priority in OO development. Class diagram can be used to identify the metric measurement for Object Oriented Software. Number of class/method/attribute and cohesion/coupling will affect to software size or complexity .

Many software metric has been proposed for object oriented software both static and dynamic aspect. Static aspect like size, complexity, coupling, cohession, etc. Static aspect consist of:

1. Number of class and method and atrribute.
   From number of class, we can detect size and complexity of the software.
2. Structural measurement (coupling and cohesion) .
3. Inheritance depth.

2.5 Static Object Oriented Metric

Chidamber and Kemerer summarizes the three static object oriented metric above into six part, there are :

1. Weighted method per class (WMC)
   WMC represent number of method in a class. Number of method affect to understandability, reusability and maintenance. The more method in the class, it is more difficult to understand the class and more difficult to maintainance. In addition more method within the class, the class are less reusable.
2. Response For Class (RFC)
   RFC is number of methods that used to response all message from a class. Include all methods in class hierarchy.
3. Lack of cohesion (LCOM)
   LCOM used to count cohesion level in a system. Low cohesion affect to high complexity. It will be potential for error being development. LCOM represents design and reusability complexity.
4. Coupling between object (CBO)
   CBO define number of relation among the class but not the inheritance class. More number of coupling can decrease the usability of system. More independent the class then it will be more easy to use by other application [10].
5. Depth of inheritance tree (DIT)
   Depth of class hierarchy. Its means length of node until to root the tree. Deeper the class tree, more better the complexity for design. DIT affect the reusability.
6. Number of children (NoC)

Number of class that have direct inheritances in class hierarchy. NoC as an indicator for reusability and testing error .

3. RESEARCH METHODS

To identify metric measurement for big data analytics system, this research used a proposed class diagram for big data analytics systems . This following is research step done :

1. Problem identification
2. Literatur Review
3. Data Collection
4. Analysis
5. Identify Measurement Metric

For the first time we identify the problem how to measure the software for big data analytics. We do literatur review through some article in journal or conference. Literature review is done to find any method proposed from previous research to measure any object oriented system. Data collected involve some diagram plan that will be used, some process or modul for big data analytics in general and any data utilities needed. After that, researcher do experiment with any diagram in UML according to process and data in big data analytics generally. Output for this step is class diagram determined as data modelling . And the final steps is through class diagram, we identify the measurement metric corresponding to object in big data analytics. This research does not convert the class diagram proposed into script or any language code. Researcher identify the measurement metric just based on the class diagram proposed.

4. RESULTS

This section will be described about proposed class diagram for big data analytics system and identify metric measurement according to the system.

4.1 Proposed Class Diagram

In the previous explanation say that this research proposes class diagram to identify metric measurement for big data analytics. According to the characteristic of big data and big data analytics process then here some constraint which needed to the system:

a. Big data means large volume of data, many format or type of data and high frequency data growth.
b. The system will process big data with parallel processing and distributed storage system.

From both of that constraint above then here we proposes the class diagram for big data analytics system. Its showed by Figure 1.

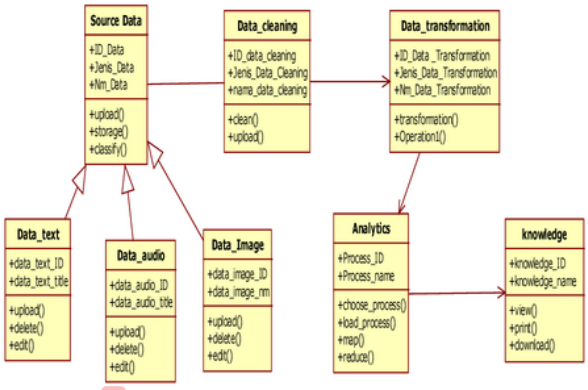

Figure1. Class Diagram For Big Data Analytics System.

Figure 1 show there is eight class for big data analytics system proposed. Data_text class, data_audio class, data_image class is inheritance from data_source class. Data_cleaning class is a class which consist of three attribute and two operation. Data_cleaning represents cleaning data process from noise like empty raw data and etc. Just the same with data_transformation, here three attribute and two operation too. Data_transformation represent integration process among data. This is the way to convert all data to be syncron with system format used. Finally all analytic process will do after all pre process done. There many process on analytic process like classification, clustering and etc.

From class diagram above, we can identify Weighted method per class (WMC), Response for a class (RFC), Lack of cohesion (LCOM), Coupling between object (CBO), Depth of inheritance tree (DIT), Number of children (NoC). Each class represent each process in big data analysis system. So it will be easy to use all the class by other application according to the requirement. Its means reusability the systems. Other words, few coupling and high cohesion will affect the reusability.

4.2 Software Metric Proposed for Big Data Analytics System

According to explanation above then table 1 show metric measurement proposed using both of static and dynamic measurement.

Table1. Proposed Static and Dynamic Metric For Big Data Analytics System

No.	Metric Type	Detail Metric
1	Static	Weighted method per class (WMC) Lack of cohesion (LCOM) Coupling between object (CBO) Response for a Class (RFC) Depth of inheritance tree (DIT) Number of children (NoC)
2	Dynamic	Number of object created at run time [dufor 2003] Number cohesion at run time

Number of cohesion is to instrument the source code and save occurences log of interaction among object members at run-time.

Object Oriented static size measure have been useful in predicting quality attributes like fault proneness and maintability [6]. Static size measureshave also been shown to have confounding effect on the validity of other OO Design measures when predicting external software quality attributes. On static metric, cohesion metric represent measurement object interaction from whole class.

Dynamic mesure are usually obtained from execution trace the code or from executable design model [13]. Dynamic metric that proposed is the number of object in a system during actual run-time program code. It take into account what likely happen when program executed. On dynamic metric, cohesion metric measurement can be spesific to single object belonging to a class at run time [14].

5. CONCLUSION

Big data analytics involve high volume data, many format of data and velocity of data. All of class and method must be fit with data OO characteristic polymorphisme, inheritance and encapsulation are frequently used to improve internal reusability and maintanability.

Proposed metric measurement which consist of static and dynamic metric can be considere to used while develop big data analytics system. Static metric is obtained from design phase and dynamic metric is obtained after execution or at run time.

Run time program profile different with profile obtained from design. Some information will be created and occur while program executed like method called and instances variable. It will be an advantages while identify the measure for big data analytics systems .

Static metric can used to predict the potential object and interaction that will occure at run time while dynamic metric measure actual object and interaction happen at run-time [9].

For the future, it is suggested to convert the class diagram into any language programming so that it can be calculated the measurement.

6. ACKNOWLEDGMENTS

Thanks to our lecturer who helped a lot in writing this article. Thanks to all of my friends for your opinion and sources of literature that used in this paper. Also thank you to Institute Technology and Business Kalbis for many support I have receipt.

7. REFERENCES

[1] Bayati.Shahabedin, Parsons.David, Susnjak.Teo, Heidary. Marzieh. (2015) "Big Data analytics on Large Scale Socio Technical Software Engineering Archives", In *Proceedings of the 3rd Workon Conference on International Conference on Information and Communication Technology*: 65-69..

[2] Fenton.Norman, Bieman.James, Software Metric: A rigorous and Practical Approach, CRC Press, 2005.

[3] Abran.Alain, Sellami.Asma, Suryn.Witold. (2003) "Metrology, Measurement and Metrics In Software Engineering", In *Proceedings of the 9th International Software Metrics Symposium*, IEEE Computer Society.

[4] V.Gupta, Jitender Kumar Chhabra. (2008) " Measurement of Dynamic Metric Using Dynamic Analysis of Program", In

Proceedings Workon Conference on Applied Computing Conference: 81-88.

[5] Booch, et al. (2007). Object-Oriented Analysis and Design With Application. 3rd Edition. New York: Addison Wesley. 27-91.

[6] V.Gupta, Jitender Kumar Chhabra. (2011) " Package Level Cohesion Measurement in Object-Oriented Software", J Braz Comput Soc. DOI= 10.1007/s13173-011-00524

[7] Curtis B, Carleton (1994) Two Software Measurement Conundrums. In: *Proceedings of 8th international software metric symposium*. IEEE Computer Society Press. Pp.96-105.

[8] Shyam R, Chidamber and Chris F.Kemerer. (1994) "A Metric Suite for Object Oriented Design", In *Proceedings Workon Conference on IEEE Transaction of Software Engineering*. Vol.20. No.6: 476-493

[9] V.Gupta, Jitender Kumar Chhabra. (2011) " Dynamic Cohesion Measures for Object-Oriented Software", *Journal of Systems Architecture*. pp.452-462 DOI= 10.1016/j.sysarc.2010.05.008

[10] Xenos,Michalis (2006) "Software Metric and Measurement", in Encyclopedia of e-commerce, e-government and mobile commerce, Mehdi Khosrow-Pour, Idea Group Publishing.

[11] F.Garcia, M.Serrano, J.Cruz Lemus, F.Ruiz, M.Pianttini. (2007) "Managing Software Process Measurement:A Metamodel-based Approach", *Information Sciences: An International Journal* **177**: 2570-2586

[12] Rawashdeh. Adnan, Matalkah. Bassem. (2006) "A New Software Quality Model for Evaluating COSTS Components". *Journal of Computer Science* **2 (4)**: 373-381. DOI= http://doi.acm.org/10.1145/964696.964697.

[13] A.Gosain, G. Sharma (2017). " A Dynamic Size Measure For Object Oriented Software "Int J Syst Assur Eng Manag.

DOI = 10.1007/s13198-017-0588-6

[14] V.Gupta, Jitender Kumar Chhabra. (2009) " Package Coupling Measurement in Object-Oriented Software", *Journal of Computer Science and Technology*. 24(2). pp.273-283.

Check 26. Measurement Metric Proposed For Big.pdf

ORIGINALITY REPORT

17%	6%	16%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Ridha Sefina Samosir, Harco Leslie Hendric, Ford Lumban Gaol, Edi Abdurachman, Benfano Soewito. "Measurement Metric Proposed For Big Data Analytics System", Proceedings of the 2017 International Conference on Computer Science and Artificial Intelligence - CSAI 2017, 2017 Publication	7%
2	Anjana Gosain, Ganga Sharma. "A dynamic size measure for object oriented software", International Journal of System Assurance Engineering and Management, 2017 Publication	3%
3	Submitted to Universiti Putra Malaysia Student Paper	3%
4	"Trends and Advances in Information Systems and Technologies", Springer Nature, 2018 Publication	1%
5	Submitted to University of Maryland, University College Student Paper	1%

6	Submitted to Nottingham Trent University Student Paper	1%
7	www.cs.put.poznan.pl Internet Source	1%
8	www.scribd.com Internet Source	1%
9	Varun Gupta, Jitender Kumar Chhabra. "Dynamic cohesion measures for object-oriented software", Journal of Systems Architecture, 2011 Publication	1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		