



International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 6, June 2017

Improving and Tackling Big Data Quality utilizing Process Pattern Model

Rohit Negi

Department of Computer Science and Engineering, Galgotias University, Yamuna Expressway Greater
Noida, Uttar Pradesh, India

Email Id: rohit.negi@Galgotiasuniversity.edu.in

ABSTRACT: Data only here and there make a value without anyone else. It should be connected and joined from different sources, which can regularly accompany variable information quality. The assignment of enhancing data quality is the common test. In this paper, it utilize a contextual analysis of an enormous telecom organization to build up a conventional process pattern model to improve the data quality. The model of process pattern is characterized as a demonstrated arrangement of activities, planned for enhancing the data quality provided a specific set, a specific aim and a certain context of beginning conditions. Four distinct patterns are determined to manage the varieties in the data quality of the datasets. Rather than having to discover the best approach to enhance the big data quality for every circumstance, the process model gives data clients with conventional patterns that can be employed as the reference model to enhance the big data quality.

KEYWORDS: Big Data, Data Processing, Data Quality, Process Pattern and Process Pattern Model

I. INTRODUCTION

The present associations gather an exceptional measure of data because of datafication that includes digitalization of business exercises and objects as a feature of the associations' processing chains. Datafication covers a scope of regular routine errands, for example, contract digitalization and sensor reading. Also, the inescapability of later advances, for example, mobile computing, parallel computation, internet of things and social media have empowered associations to store up information from its foundations and its clients. Notwithstanding, with data originate from different sources, data quality regularly changes, and it makes hard for associations to control, specifically when the data are not consistently cleaned or adulterated. Some present a perfect set of data though others might be undermined because of missing traits, particular mistakes, etc. Data quality is commonly estimated by its level of qualification for use by consumers or data users, catching a wide point of view of the degree to which intrinsic and utilization estimation of the big data could be acknowledged and further tackled. Numerous investigations propose that associations can pick up profits by the information on the off chance that it prevail with regards to opening a value from data[1]. These advantages include: more prominent effectiveness and benefits, and upper hands. Be that as it may, the inquiry remains, how associations can understand potential value from the data. Integral to this worth creation is data client's point of view of how to guarantee top notch datasets can be effectively handled. In spite of the fact that the noteworthiness of the data processing is by all accounts instinctive, numerous associations have neglected to execute it.

In this paper, it look to build up the process pattern model so that an association can use to manage the data of the variable quality. Model will give an orderly way to deal with distinguish, survey quality, combine and curate data. Which processing ought to be followed will rely upon the data quality, operational and context goals. A portion of variable data quality are conventional to a larger part of information driven activities and are definitely not one of a kind to a particular sort of association. Data quality gives the underlying arrangement of conditions for choosing the procedure steps that are important to set up data for utilization. Such use patterns could be seen as a training, that can be reutilized or from that others can learn[2]. It characterize the process pattern model like a common grouping of steps that outcomes in achieving the particular operational objective. Given the unique situation and certain beginning



International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 6, June 2017

conditions, models can be followed to make esteem from data. Process pattern model ought to be autonomous of the executed innovation and should empower associations to make value from data. The goal of this exploration is to build up the process pattern model to tackle variable data quality. The process pattern model would be characterized in its investigation as a demonstrated arrangement of exercises to conquer a repetitive issue in a specific setting against set of targets, and under a particular arrangement of beginning conditions[3]. It employed an approach of case study to deal with look at the regular practice in an information driven organization.

II. LITERATURE BACKGROUND

To determine the process pattern model to tackle variable data quality, it survey various ideas from the surviving writing on data processing, process patterns and data quality. These ideas are fundamental to accomplishing information driven tasks what's more, objectives. Specifically, data quality is part factor, which influences the activities of data quality in this exploration. An exhaustive portrayal of the data quality is needed not exclusively to comprehend the idea better yet in addition to give a benchmark reference model for future and present research[4]. Furthermore, with variable data quality, despite the fact that the necessary procedures may vary from each other, it may share certain recognizable steps in the rising patterns. Consequently, it additionally incorporate the writing on the process patterns and the models in its review.

➤ *Data Quality:*

1. *Concepts of Data Quality:*

DQ i.e. Data quality has been generally recognized to be a noticeable test in big data writing. As depicted by researcher, low DQ can adverse effect operational, tactical and strategic levels of organizations. It incorporate significant expense, poor basic leadership, and expanded troubles in detailing strategy. High data quality is the instrumental to esteem creation as the high-quality data ought to be characteristically acceptable, relevantly fitting for the errand, obviously spoke to, and available to data consumer.

2. *Assessment of Data Quality:*

Assessment of data quality is regularly performed by building up a benchmark and occasionally checking status across databases, organizations and stakeholders. The status is spoken to by quantifiable parameter from the certain dimension of data quality, for example, liberated from blunder rating as the metric of accuracy[5]. Survey of data quality evokes assessments of numerous dimensions of data quality from various stakeholders in the association. The subsequent evaluation reflects subjectivity, for example the view of stakeholders.

➤ *Data Processing:*

Because of the fluctuation of the data quality in the datasets, there is not any uniform manner to process them. Accordingly, which procedure ought to be followed relies upon data quality. Ordinarily, data are prepared consecutively in the data lifecycles that envelop all aspects of data generation for knowledge creation. There are numerous models of the data lifecycles in writing. Some unmistakable ones involve: DDI i.e. "Data Documentation Initiative" "Joined Lifecycle Model", "ANDS informing Sharing Verbs" and "Data One Information Lifecycle". Although these models utilize various phrasings, all models of information lifecycles share regular exercises, which mirror an information client's point of view[6].

Its examination looks to analyse variation forms to improve data quality from the point of view of information clients. It condense information processing lifecycle from writing in Fig. 1. From an information client's point of view, the initial step of the data processing is for find pertinent information from information suppliers. It utilizes accessible interfaces to find the information or by making concurrences with information providers. Finally, mechanistic techniques are employed for deterministic estimation. These techniques advantage clients from the investigation in different media. Numerous reports can be produced to impart the discoveries.

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 6, June 2017



Fig.1: Data Processing Lifecycle

➤ *Process Patterns & Models:*

The previously mentioned the data lifecycle gives bases to data processing that may fluctuate dependent on data quality of data. Variable quality of external and internal data will bring about the utilization of isolated arrangements of protocols. The variation utilization of the protocols comparable to the data quality forms premise of its process pattern model. Model is involved pattern and process. In the writing, there is no agreement about what ought to be involved for the process pattern. A setting of a pattern depicts a structure circumstance that offers ascend to a plan issue[7]. The issue depicts a solid circumstance, which may rise in the relevant application. Process patterns of defeating the challenges of data quality help associations in making an incentive from the information.

III. RESEARCH METHODOLOGY

The aim of exploration is to determine process patterns for how an association can form value by tackling variable big data quality. It utilized an approach of case study to deal with inductively infer its process pattern model. Subjective case study look into is generally utilized in research of information systems and is appropriate to comprehend the connections between organizational contexts and informational technology based innovations. Such methodology permits us to inspect the ordinary act of tackling the data quality in the real life contexts what's more, investigate contemporary issues in-situ[8]. As indicated by researcher, case study incorporates an assortment of information assortment instruments to guarantee develop validity. Following criteria were employed in the determination of case study association:

- It was an information driven association both as far as its tasks and business procedure.
- It utilized what's more, consolidated data from various sources to achieve its objectives.
- Gave priceless bits of knowledge into tackling variable data quality as a conventional issue.

IV. PROCESS PATTERNS MODEL TO TACKLE THE BIG DATA QUALITY

In view of the ideas from the past section, it propose process pattern model that is delineated in Fig. 2. Forming value from the big data needs a fine-grained portrayal of the fundamental nonexclusive activities of data processing with clear pathways for achieving certain objectives and targets. The build stage and design of data processing begins by

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 6, June 2017

considering the tasks and contexts that speak to the goals of information reuse. It incorporate variation forms that utilization same data, but with various targets and outcomes. Depend on evaluated data quality, data processing can vary with one dataset then onto the next. For instance, a dataset containing incorrect perceptions should be cleaned earlier to being misused and joined with different datasets.



Fig.2: Process Pattern Model for Tackling Big Data Quality

V. CASE STUDY

The aim of case study was for derive the process pattern model. Hence, case study included numerous techniques for data assortment, involving ethnography, document and interviews. It investigated the essential procedures of CDMA marketing office in handling with different data sources of the variable data quality. A portion of the programs of marketing didn't have adequate support and were difficult to quantify and assess its outcomes. Initially, removing value from data appeared to be clear utilizing functions gave in information system. Be that as it may, it was seen as perplexing regarding tackling the variable data quality. Because the telecom consolidated numerous datasets, it's changed significantly[9]. Internal partner's data and big data ordinarily accompanied high intrinsic data quality since these information were self-guided.

There were numerous manners that the data processing can be sent to tackle different issues of data quality. For instance, exactness issue could be settled by purging the information before use. Some purging schedules were shown in case study for underline a particular nonexclusive answer for a specific data quality issue. From these, the process patterns which were intermittent could be related to every data processing to a specific issues of data quality. The connection between data quality and process pattern issue given a fundamental building block of architecture for its process pattern models.

VI. PROCESS PATTERNS

In recognizing the process pattern model to tackle the variable data quality, its case study association demonstrated that in spite of the fact that it built up information system with different functionalities, data were handled in a grouping, following data lifecycle from significant level of deliberation. From case study and literature, it determined the accompanying crucial data lifecycles[10]. The ostensible data lifecycle combined with functionalities employed to process data is appeared in Fig. 3. Organizations frequently need to manage datasets with the low quality. This happens for the most part with the datasets from outside sources. The hole because of low natural quality causes interior shortfall, for example a condition where inward clients see the interior properties of low quality data, for example,

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 6, June 2017

inaccurate, untrusted, biased and so forth[11]. To defeat this issue, associations can improve the genuineness of datasets by leading exercises such as surveying its precision and representativeness, rating validity of information suppliers, pre-processing data, etc.

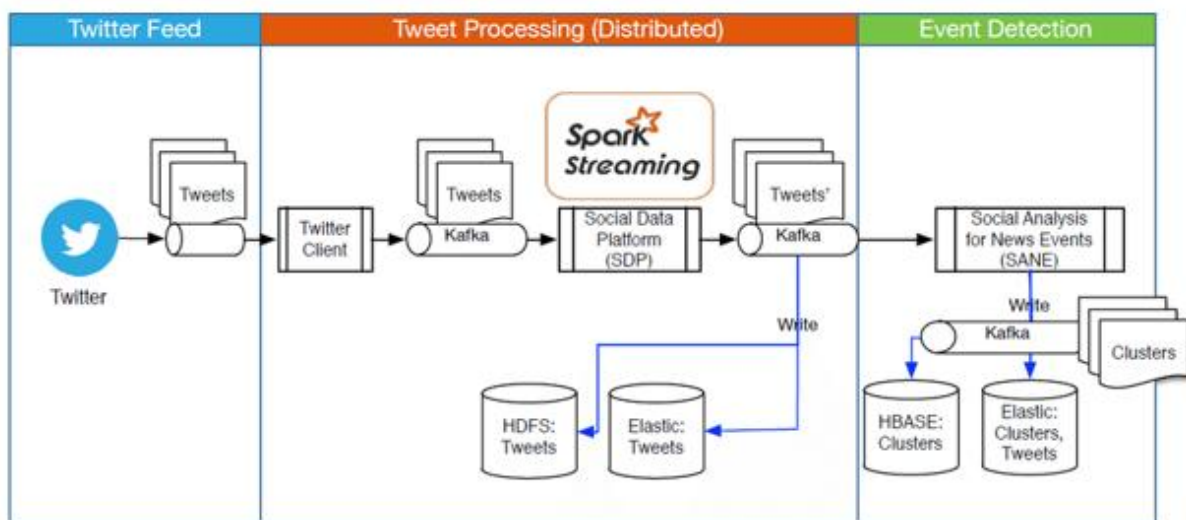


Fig.3: Process Architecture for Processing Big Data

VII. CONCLUSION

Associations are struggling with employing big data. Every time, it need to discover methods for tackling data with the variable quality. Rather than beginning without any preparation, it build up a summed up system of the process patterns models for assisting creating value from the data in the data driven associations. Its model envelops 11 variations of the process patterns. Every pattern is identified with variable quality of data. It characterize the process pattern as the proven arrangement of activities that should conquer a repetitive issue in a certain setting, specific target, and explicit starting condition.

The patterns are included activities of data processing employing data from different sources to achieve certain objectives and targets from the one context to other. Such patterns are defined by variable quality of data being employed. Besides patterns, the model to beat the issue of data quality was additionally proposed. Any associations can profit by patterns and employ model to settle the issue of data quality. Along these lines, associations can utilize patterns as the 'best practices', save resources and time and abstain from discarding ventures by reutilizing patterns. The strategy and patterns may advance the archive of the big data strategy and process pattern in the literature. Moreover, it can be employed as a beginning stage to further refine patterns and generalise with other circumstances.

REFERENCES

- [1] A. Wahyudi, G. Kuk, and M. Janssen, "A Process Pattern Model for Tackling and Improving Big Data Quality," *Inf. Syst. Front.*, 2018.
- [2] Q. Liu, G. Feng, N. Wang, and G. K. Tayi, "A multi-objective model for discovering high-quality knowledge based on data quality and prior knowledge," *Inf. Syst. Front.*, 2018.
- [3] L. V. Le, D. Sinh, L. P. Tung, and B. S. P. Lin, "A practical model for traffic forecasting based on big data, machine-learning, and network KPIs," in *CCNC 2018 - 2018 15th IEEE Annual Consumer Communications and Networking Conference*, 2018.
- [4] B. M. Ghodki and T. K. Goswami, "Modeling of granular heat transfer in cryogenic grinding system: Black pepper seeds," *Chem. Eng. Res. Des.*, 2019.
- [5] 6(1). <http://doi.org/10.1186/s13613-015-0104-6> Deye, N., Vincent, F., Michel, P., Ehrmann, S., Da Silva, D., Piagnerelli, M., ...Laterre, P.-F. (2016). Changes in cardiac arrest patients's temperature management after the 2013 trial: Results from an international survey. *Annals of Intensive et al.*, "The EMEA ADHES survey in schizophrenia: Nurses' views on the treatment options for improving partial or non-adherence in patients with schizophrenia," *Schizophr. Res.*, 2012.



ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 6, June 2017

- [6] W. J. O'Brien, S. Ponticelli, and American Society of Civil Engineers. Computing and Information Technology Division, *Computing in civil engineering 2015 : proceedings of the 2015 International Workshop in Civil Engineering, June 21-23, 2015, Austin, Texas*. 2015.
- [7] F. Deye Vincent Michel, P., Ehrmann, S., Da Silva, D., Piagnerelli, M., ... Laterre, P.-F. (2016). Changes in cardiac arrest patients's temperature management after the 2013 trial: Results from an international survey. *Annals of Intensive Care*, 6(N et al., "Mobile health (mHealth) approaches and lessons for increased performance and retention of community health workers in low- and middle-income countries: a review.," *Journal of medical Internet research*. 2013.
- [8] 37–41. <http://doi.org/10.1037/a0022390> Tuma, J. M., & Pratt, J. M. (1982). Clinical child psychology practice and training: A survey. *Journal of Clinical Child & Adolescent Psychology*, 137(August 2012) et al., "Excellence Framework 2014: The results," *Ref2014*, 2014.
- [9] 6(1). <http://doi.org/10.1186/s13613-015-0104-6> Deye, N., Vincent, F., Michel, P., Ehrmann, S., Da Silva, D., Piagnerelli, M., ... Laterre, P.-F. (2016). Changes in cardiac arrest patients's temperature management after the 2013 trial: Results from an international survey. *Annals of Intensive Care et al.*, "Building capacity through implementation of a clinical pharmacy training model for HIV medicine," *Pharmacotherapy*, 2015.
- [10] 37–41. <http://doi.org/10.1037/a0022390> Tuma J. M. & Pratt J. M. (1982). Clinical child psychology practice and training: A survey. *Journal of Clinical Child & Adolescent Psychology* 137(August 2012) et al., *Detecting diseases in medical prescriptions using data mining tools and combining techniques*. 2016.
- [11] 6(1). <http://doi.org/10.1186/s13613-015-0104-6> Deye, N., Vincent, F., Michel, P., Ehrmann, S., Da Silva, D., Piagnerelli, M., ... Laterre, P.-F. (2016). Changes in cardiac arrest patients's temperature management after the 2013 trial: Results from an international survey. *Annals of Intensive Care et al.*, "'Shadows and Light' project: Improving the access to health services and fostering increased and accelerated action on sexual and reproductive health and HIV linkages for key populations," *Trop. Med. Int. Heal.*, 2015.