

An Iterative and Incremental Data Quality Improvement Procedure for Reducing the Risk of Big Data Project

Sen-Tarng Lai*

Department of Information Technology and Management, Shih Chien University, No.70, Dazhi St., Zhongshan Dist., Taipei City 104, Taiwan (R.O.C.)

* Corresponding author. Tel.: +886-2-25381111 ext. 8954; email: stlai@mail.usc.edu.tw

Manuscript submitted September 18, 2017; accepted October 17, 2017.

doi: 10.17706/jsw.12.12.945-956

Abstract: Big data applications can enhance the market competitive advantages of enterprises and organizations and can improve people's quality of life. However, by the impact of many factors, failure rate of big data project is higher than the IT project. In order to reduce the risk of failure, big data projects must overcome a serial of challenges. Ambiguous requirements, poor data quality, and lacking changeability and extensity will directly affect the results of big data analytics. And even cause the wrong decision, inaccurate prediction and improper planning to make the big data projects with potential high risk. For this, this paper migrates iterative and incremental development (IID) features to the data preprocessing, and draws up the iterative and incremental data quality improvement (IIDQI) procedure. IIDQI procedure applies data preprocessing task frame to repeatedly detect and identify the defects of data quality, and incrementally strengthen big data quality and control the factors of failure risk. Iterative inspection activities can effectively enhance data quality, intercommunication efficiency, and precision requirement and objective to reduce the risk of big data project failure.

Key words: Big data, data preprocessing, failure risk, IID, quality improvement.

1. Introduction

Big data applications enhance enterprises and organizations market competitive advantage and improve the quality of people's daily lives. However, the fail rate of big data projects have over 50%. In a recent report, Gartner predicted that "through 2017, 60% of Big Data projects will fail to go beyond piloting and experimentation and will be abandoned." [1] Many papers have surveyed and discussed the failure factors of big data projects [2]-[4]. Summarized the failure factors of big data projects can be divided into five items: (1) Ambiguous requirements; (2) Lacking well communication channel; (3) Data quality cannot get effective and timely improvement; (4) Lacking experienced professional persons; (5) Project plan lacks extensibility and changeability. For getting the advantages of big data applications, big data projects must overcome the challenges of project failure risks. One of major failure factor of big data project is the critical data quality that cannot effectively and timely improve. It is because that most of the network data are unconfirmed contents, unverified information and incompatible data format [4]-[8]. Poor data quality of big data is bound to affect the results of big data analysis and increase the failure risk. How to improve data quality of big data have become the subject of enterprises and organizations must focus on [4]-[8].

Agile software development neglects the formal analysis and design phases, uses non-document oriented development, and pay more attention to refactoring and communication [9]-[12]. The features of agile

software development effectively reduce the software development projects risks and increase project success rate [13]. It is because that agile software development uses IID (Iterative and Incremental Development) [14], [15] to reduce the requirement complexity, refactoring to increase the requirement modification flexibility, and non-document oriented can reduce the cost of requirements change [13], [16]. Incremental workable versions let the client can detailed test, validate requirement items of new version and precisely understand the development progress. Iterative extension and incremental validation increase the integration effect and intercommunication ability, and strengthen the unambiguous requirements and objective. In development change impacts, IID greatly decreases schedule delay, cost over budget and quality unsatisfied user requirement situations. IID makes software development change risk can be effectively reduced. The features of IID should help control the failure risk of big data project.

In big data applications, poor quality data causes incorrect analysis results and may form the wrong decision or inaccurate prediction. Data quality will directly affect the results of the analysis of big data. In order to enhance the big data applications effectiveness and practicality, application project must confirm the data quality in advance. Combines the IID, this paper designs the Iterative and Incremental Data Quality Improvement (IIDQI) procedure to overcome the development risks of big data projects. IIDQI procedure combines the key features of IID into data preprocessing tasks. Corresponding to the preprocessing tasks, incremental data preprocessing mechanism can stepwise identify data quality defects and repeated inspection activities can increase opportunities of intercommunication among DE (Data Engineers), users and stakeholders for improving the precise requirements and objective. Continuous communication and iterative data quality improvement can increase big data analysis efficiency and quality and overcome the development risks of big data projects. In Section 2, discusses the advantages and development risks of big data project. In Section 3, describes data preprocessing steps and importance characteristics of data quality. In Section 4, introduces the advantages of IID and designs the IIDQI procedure to overcome the development risks of big data projects. In Section 5, the improvement effects of IIDQI procedure compare to others two data preprocessing methods. In Section 6, emphasizes the contributions of IIDQI procedure and makes a conclusion of this topic.

2. Advantages and Development Risks of Big Data Projects

The advantages of big data applications make most of enterprises and organizations actively introduce into the related technology of big data.

2.1 Big Data Trends and Advantages

In the network and IoT (Internet of Things) age, making all kinds of data and information generated speedily and continuously, just like data explosion age. From the business intelligence view, big data becomes the important and useful assets of enterprises and organizations to help them enhance market competitiveness advantages. Effectively collection, appropriately management and suitably storage all kinds data of the EC transactions and social network activities. Using high efficiency statistic and analysis tools for speedy generating the visualization information to help enterprises and organizations make correct decision and accurate prediction [17]. Big data widely applications make that most of enterprises and organizations actively import the related technology of big data. Customer relationship viewpoint, big data applications technology can forecast the trend of consumer group. Assisting management level make the best planning and decisions to increase the advantage of market competition and increase the incomes and profits [18]-[20]. Daily life viewpoint, big data applications to traffic control can reduce time of traffic jam, accurate meteorological prediction may be reducing the calamities losses, suitable crop planting can improve the quality and yield, in health management can control the disease etc., can efficiently improve in people's quality of life. In network and IoT age, big data applications have become the important topics that

should further explore [18], [20]. However, many development challenges need to overcome in big data applications. Analysis tools, data quality, extensibility and changeability are the critical failure factors to affect the big data applications.

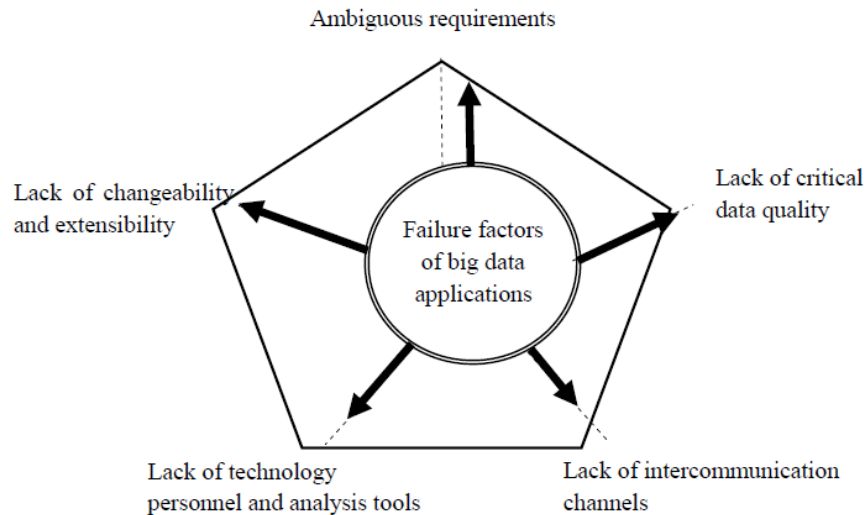


Fig. 1. Failure factors of big data applications.

2.2. Development Risk of Big Data Projects

Big data applications enhance enterprises and organizations market competitive advantage and improve the quality of people's daily lives. However, big data application projects always have high failure risk. Over 50% big data projects belong to failure project. Big data projects failure rate have exceeded IT projects. In a recent report, Gartner predicted, "through 2017, 60% of Big Data projects will fail to go beyond piloting and experimentation and will be abandoned." [1] Many papers have surveyed and discussed the failure factors of big data projects [2]-[4]. Summarized the failure factors of big data projects as follows:

- 1) Ambiguous requirements: big data application and IT software development have the similar situations. User requirements often existed ambiguous or incomplete description. The requirement specifications of project cannot reach a consensus and always cause the project failure.
- 2) Plan cost overrun and schedule delay: big data project plan should accurately estimate the development cost, time and resources. However, additional requirements and environment change always cause project plan cost overrun and schedule delay.
- 3) Lack of extensibility and changeability: big data applications planning process, data maintainability quality is generally very little consideration as the application requirements. Therefore, often lead to big data projects lack of application of extensibility and changeability.
- 4) Decision making challenges of managers: lack of communication and interaction among users, stakeholders and developers, causing managers prefer to rely on their own intuition as the decision making, and do not want to believe that big data analysis results.
- 5) Poor data quality: data quality may affect the results and efficiency of data analysis. Big data with poor quality will lead to difficult to predict the results.
- 6) Lack of technical resources: Lack of experienced personnel, and professional technology or practical analysis tools, often leading to large data projects cannot successfully achieve the desired goal.

Failure factors of big data project can be inducted into 5 items (shown as Fig. 1): (1) Ambiguous requirements and objective; (2) Lack of intercommunication channels; (3) Lack of critical data quality (ex. changeability and extensibility); (4) Lack of technology personnel and analysis tools; (5) Lack of experienced project manager. Ambiguous requirements and objective, lack of intercommunication channels, and Lack of critical data quality are more serious than the others two shortcomings. IT projects have many failure factors that include lack user involvement, lack executive management support, lack changeability and extensibility, and ambiguous requirements [13], [21]. Failure factors of big data application project have many same items as IT failure factors. The failure factors contrast table between big data application and IT project shown as Table 1. In software development project, agile software development applied IID process to reduce the software development risk. Iterative extension and incremental validation increase the integration effectiveness and intercommunication ability, and effectively improve the unambiguous requirements and objective. In change management, IID has flexibility and changeability to decrease schedule delay, over budget and quality unsatisfied user requirement situations. IID makes software development change risk can be effectively reduced. The features of IID can also control and reduce the failure risk of big data project.

Table 1. Comparison of Failure Factors of Big Data Application and IT Projects

Failure factors	Big data project	IT project
Lack of user involvement	X	V
Lack of executive management support	X	V
Ambiguous requirements	V	V
Lack of intercommunication channels	V	V
Requirements & specifications frequent change	V	V
Lack of critical technology	V	V
Lack of experienced professionals	V	X

3. Data Preprocessing and Data Quality

The objective of data preprocessing in big data application is to enhance the data quality and increase the efficiency of data management, analysis and modeling.

3.1. Major Tasks of Data Preprocessing

Big data applications collect many kind network transactions and activity records that have large amount data, multiple styles and formats. Big data applications able to complete the data analysis on schedule and show the data visualization results to assist the decision making and subsequent processing jobs. Applications must have high efficient analysis methods and tools to meet the requirements of enterprises and organizations and reach the objectives of big data applications. For getting the correct decisions, accurate predictions, perfect planning and arrangements, before data analysis, big data application project must previously improve the data quality. Data preprocessing is an important procedure to increase the data quality of big data. Five tasks of data preprocessing of big data application [7], [8] discusses as follows:

- 1) Data cleansing: For assuring data quality, at first data cleansing need identify the data defects and problems. Then, according the types of data defects and problems, fill in missing values, smooth noisy data, recheck or remove abnormal data, and adjust the incomplete or inconsistent data.
- 2) Data transformation: Big data collected from widely areas and different environments. It needs take more time and resource to handle and analyze the different formats data. Therefore, the collected data should normalize into a unified format and aggregate into the suitable clusters.
- 3) Data integration: For increasing data analysis efficiency, data integration of multiple databases, data

storages or files is an important task [22]. High quality data integration can help reduce and avoid redundancies and inconsistencies in the stored data set. Data integration can help improve the accuracy and speed of the subsequent data analysis process.

- 4) Data reduction: The collected data may exist the duplicate or similar contents. For reducing data volume and data analysis time, the duplicate or similar data should identify for timely merging or removing. It is necessary to use the duplicate or similar data recognition tools for avoiding the incorrect data deleting.
- 5) Data discretization: This task divides the range of a continuous attribute into intervals for reducing data size by discretization. Big data applications decide the necessity of data discretization task.

3.2 Data Quality Impact for Big Data Applications

In order to get the practical benefits of big data applications, data management and analysis must have high efficiency. Therefore, data preprocessing is a necessary job before data analysis of big data. Data preprocessing can inspect data defects and exclude abnormal data, cannot only enhance the efficiency of data analysis, but also can strengthen the data quality. In order to enhance the efficiency of data analysis and the practicality of big data applications, four critical data quality characteristics of big data applications discussed as follows:

- 1) Data collection from same one data source should have the correctness/veracity, completeness and consistency quality. Basic quality of big data is a necessary quality characteristic for confirming the efficient and practical data.
- 2) Usability: Large amount data collection must invest much resource for data classification, storage, management. Therefore, data usability should be confirmed to remove most of inapplicable data and avoid wasted resources investment. Data usability should care three confirmable items: data sources with high credibility, times of data generated and the expression of content.t.
- 3) Maintainability: Data collected from many different websites or devices need complete data classification, storage, management, then proceed suitable data merge, integration and transformation etc. maintenance operations. Data maintainability can enhance the efficiency of maintenance operations.
- 4) Manageability: Data of network or IoT not only has not uniform format, but also without same data interpretation and parsing methods. The meaning or intelligent information of data cannot be parsed or extracted, the data will belong to useless assets. Therefore, data need have manageability to increase analysis capability and applications effects. Data manageability should consider data format transferability, data assessment ability and comparability, data semantic analyzability quality factors.

Table 2 shown the cross-reference table of data quality and data preprocessing tasks.

Table 2. Cross-Reference Table of Data Quality and Data Preprocessing Tasks

Data Quality Characteristics	Quality factors	Preprocessing Tasks
Basic quality	<ul style="list-style-type: none"> • correctness • completeness • consistency 	<ul style="list-style-type: none"> • Data cleansing
Usability	<ul style="list-style-type: none"> • data source credit • data creation date • data expressed contents 	<ul style="list-style-type: none"> • Data cleansing
Maintainability	<ul style="list-style-type: none"> • data extensibility • data integration ability • data non-duplicate 	<ul style="list-style-type: none"> • Data reduction • Data integration
Manageability	<ul style="list-style-type: none"> • data format transferability • data access and 	<ul style="list-style-type: none"> • Data transformation • Data discretization

	comparability • data semantic analyzability	
--	--	--

4. IIDQI Procedure

Data collection of big data application is the continuous job. Therefore, collected data quality improvement also need to continuously process for ensuring the quality of big data application development.

4.1 Data Preprocessing Tasks Frame

IID is an important development process in the agile software approach [9]. IID effectively improves the efficiency of intercommunication among the user, developers and stakeholders. IID process is a practical and effective way that periodically released the workable version to reduce the risk of software development. Continuous workable versions increase users and stakeholders' confidence in software development projects. In big data project, many development risks and challenges are the same as the software development project. Using continuous communication and progressive development can control the failure risk of big data project. Therefore, it is a practical approach to migrate the characteristics of the IID to the data preprocessing operations. The data incremental work items of each step must have the extensibility and integration capabilities. Before data analysis and modeling, the specific improvement of the data quality can effectively enhance the efficiency of follow-up big data application project. For this, this section to develop a Data Preprocessing Task Frame (shown as Fig 2). Each preprocessing task items must sequentially complete pre-conditions, sub-work items, problems and defects solution, and post-conditions etc. four work tasks. The pre-condition and post-condition are set up to increase the effectiveness of data preprocessing:

- 1) First, set the pre-condition:
 - Ensure the previous task items have completed.
 - Identify the detailed tasks and objective of the data preprocessing through an intercommunication manner.
 - Data engineer should accept the recommendations of users and stakeholders to formulate the main handling rules of task items.
- 2) Completion of five sub-items according to the pre-conditions and the task rules:
 - Ensure that the preprocessed data has the quality characteristics of correctness, completeness and consistency.
 - Ensure that the preprocessed data has usability.
 - Ensure that the preprocessed data has manageability.
 - Ensure that the preprocessed data has maintainability.
 - Handle the exception situations that lack the task rules.
- 3) According to task of data preprocessing, the task items that need to be discussed or unreached the objective should be clearly listed.
- 4) Finally, set the post-condition of the task items according to the pre-conditions:
 - Ensure the completed task items meet the task objectives and meet the task requirements.
 - Complete the task items according to handling rules.

Each time a task of data preprocessing is completed, an inspection activity is taken to determine the quality and results of the task items. The results of the inspection activity divide into three levels (shown as Fig. 2):

- Passed review: the procedure can directly enter the next task item.

- Conditional passed: According to inspection report, the incomplete or lacking items must be remedied, and the revised task items must be re-inspected.
- Reject and redo: many serious errors or incomplete task items, the inspectors determine the defects impossible improvements and the task should be redone.

This paper combines IID features with data preprocessing task frame to improve the data quality of big data application. For the uninterrupted and incremental data, Iterative data detection and inspection can efficiently identify the data defects. Iterative data defects improvement operations make the big data with correctness, completeness and consistency quality and usability, manageability, and maintainability characteristics. The iterative and incremental data quality improvement operation is called an IIDQI (Iterative and Incremental Data Quality Improvement) procedure.

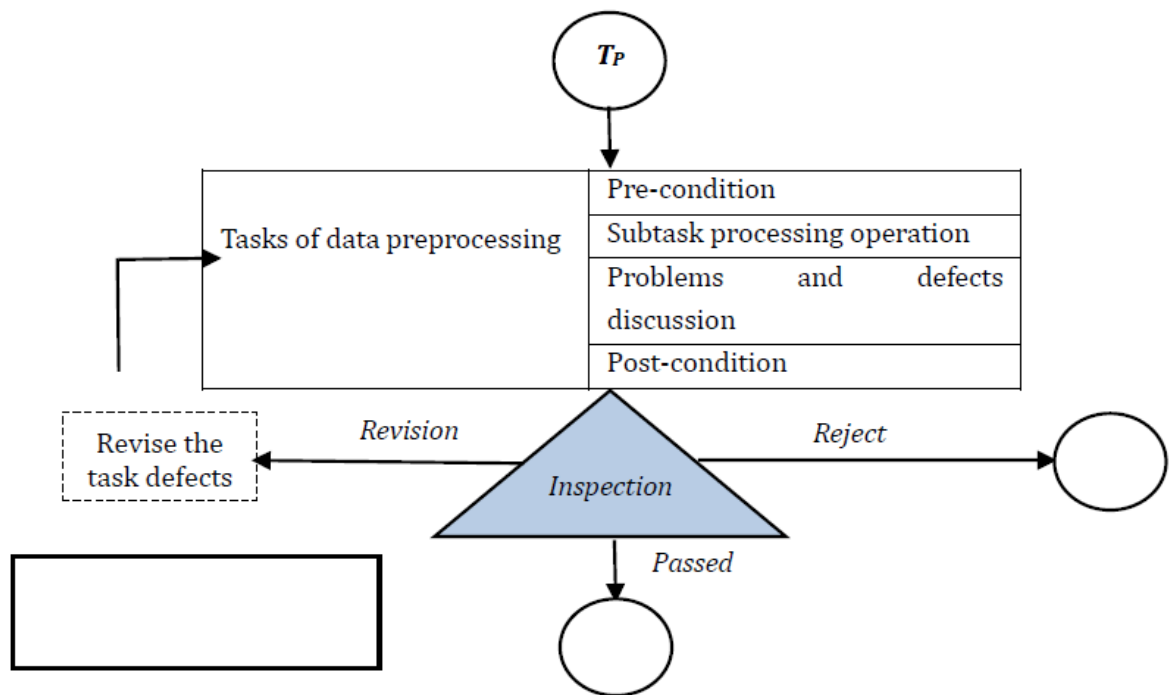


Fig. 2. Inspection activity of data preprocessing task frame.

4.2 Data Quality Improvement Flow

Task items inspection activities same as the acceptance test of new release version of the software. User can deeply understand the work items to complete the task, including the contents of the pre-conditions and post-conditional completion of the content, and the encountered problems. User can fully understand the effectiveness and processing progress of data quality improvement in data preprocessing. Inspection activities not only can achieve the effect of communication and interaction, but also increase the confidence of user and stakeholders. Based on the task frame, IIDQI procedure is divided into five task items that include data cleansing, data transposition, data integration, data reduction and data quality assurance (shown as Fig. 3). Each task based on the task described in Section 4.1 to define the contents of task frame. Before task finish, each task item must be inspected to determine whether the task item is completed or not. The post-condition of the following five task items are the core of the inspection:

- 1) Post-condition for data cleanup: Make up incomplete field, modify data content, confirm check fields, or delete exception data, and adjust incomplete or inconsistent data.
- 2) Post-condition for data transposition: Based on the data attributes, the collected data must be normalized to a uniform format and be aggregated into the appropriate cluster.

- 3) Post-condition for data integration: Data with the same purpose, the same attribute, or the same target must be integrated for effectively improving the quality and efficiency of the data management and analysis.
- 4) Post-condition for data reduction: The duplicate or similar data needs to be identified and be merged or be deleted for reducing the unnecessary data. It is necessary to use duplicate or similar data recognition tools to determine the data that should be deleted.
- 5) Post-condition of data quality confirmation: In addition to requiring the correct, complete and consistent basic quality, it also requires the key qualities such as availability, manageability and maintainability.

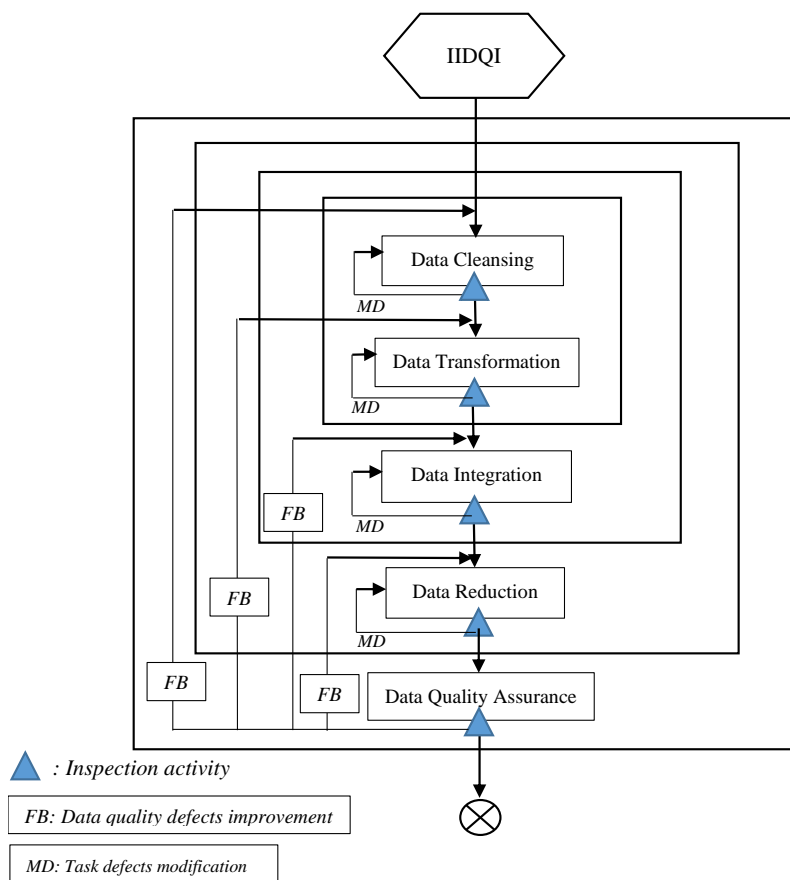


Fig. 3. Operation flowchart of IIDQI procedure.

In order to reduce the big data projects failure risk, the IIDQI procedure must have the following five features:

- 1) Pre-condition of the task item should integrate the suggestions of data engineer, user, stakeholders and other personnel.
- 2) Before the completion of the task, task items must be inspected to confirm to the complete task items, progress and quality.
- 3) According to task items, each task of data preprocessing should enhance the specific quality characteristics of big data.
- 4) After the completion of all preprocessing tasks, big data should have usability, manageability, and maintainability etc. quality characteristics.
- 5) Combining the relevant personnel to discuss and help handle the unresolved problems and quality

defect.

In data preprocessing, IIDQI procedure identifies data quality defects and measures four critical quality characteristics that include basic quality, usability, maintainability and manageability. IIDQI procedure apply the DQM (Data Quality Measurement) model [23] and blends four critical quality characteristics into big data application. Four critical quality characteristics enhance the data quality, intercommunication channels, changeability, and extensibility of big data project for efficiently reducing the big data project failure risk.

5. Advantages and Efficiency Evaluation of IIDQI Procedure

In this section, IIDQI procedure compares to others two data preprocessing methods to show the efficiency of IIDQI procedure in big data application project.

5.1 Improvement of Big Data Failure factors

The big data application projects exist the high failure rate. For overcoming the challenges of big data development risk, IIDQI procedure migrates the characteristics of iterative and incremental development to the data preprocessing operation mechanism. In this section, five important factors that affect the big data development risk are identified and discussed as follows:

- Lack of critical data quality: IIDQI procedure applies iterative data preprocessing mechanism that can effectively reduce the defects of critical data quality of big data application.
- Lack of intercommunication ability: IID process provides iterative data check and data inspection that can improve the intercommunication ability among DA (Data Analysts), stakeholder and DE (Data Engineers).
- Lack of changeability: Before enter the analysis phase of big data application development, the data preprocessing tasks can provide the uniform format and remove the bad data for improving the flexibility and changeability of big data application.
- Lack of Technology personnel & analysis tools: This factor belong to the complex technology issue. Therefore, any data preprocessing methods cannot provide suitable improvement effects.
- Lack of suitable PM: This factor belong to the management experience and technique that need to cultivate long time. Therefore, any data preprocessing methods cannot get useful improvement effect.

Based on the five important factors, the IIDQI procedure is compared with others two big data application data preprocessing methods (shown as Table 3).

Table 3. Data Preprocessing Methods Improvement Effect for Big Data Failure Factors

Data preprocessing methods	IIDQI procedure	Complete data preprocessing tasks	Partial data preprocessing tasks
Failure factors of big data project			
Lack of critical data quality	<i>High</i>	<i>High</i>	<i>Middle</i>
Lack of intercommunication ability	<i>High</i>	<i>Middle</i>	<i>Low</i>
Lack of changeability	<i>High</i>	<i>Middle</i>	<i>Low</i>
Lack of Technology personnel & analysis tools	<i>Low</i>	<i>none</i>	<i>none</i>
Lack of suitable PM	<i>none</i>	<i>none</i>	<i>none</i>

5.2 Data Quality Improvement for Big Data Application

In big data application development, the purpose of data preprocessing is to increase the data quality, and assist data management, data analysis, and data modeling tasks. However, general data preprocessing method just consider the follow-up jobs (ex. data storage, data analysis, and data modeling). The critical

quality characteristics measurement do not consider as a necessary job. Therefore, the quality characteristics of big data application existed incompletely inspect and measure. Some data quality defects still remain on the application development. In Section 4.2, data quality impact for big data applications have discussed the critical data quality that data preprocessing should consider the quality factors and characteristics. IIDOI procedure identifies data quality defects and measures four critical quality characteristics that include basic quality, usability, maintainability and manageability. In order to efficiency reduce the project failure risks, IIDOI procedure applies the DQM (Data Quality Measurement) model [23] and blends four critical quality characteristics into big data application. In general data preprocessing methods, the task focuses on the data defects identification and revision. The critical quality characteristics measurement often is omitted or doesn't consider. Collecting and finishing three data preprocessing methods improvement effect for data quality characteristics, the comparison results shown as Table 4.

Table 4. Data Preprocessing Methods Improvement Effect for Data Quality Characteristics

Data preprocessing methods	IIDQI procedure	Complete data preprocessing	Partial data preprocessing
Data quality characteristics			
Basic quality	High	Middle	Middle
Usability	High	Middle	None
Maintainability	High	Middle	None
Manageability	High	Middle	None

6. Conclusion

Big data applications can enhance the enterprises and organizations competitive advantage, can effectively improve the quality of people's lives. However, big data application projects existed over 50% failure rate. Poor data quality, ambiguously application requirements, lack communication, interactive and extensibility are major project risk factors of big data. How to reduce the risks of big data applications has become an issue that enterprises and organizations must take care. For this, this paper introduces IID features into the data preprocessing operations and designs the iterative and incremental data quality improvement (IIDQI) procedure. Applied IIDQI procedure, iterative inspection activities identify the defects of data quality and control the failure risks, and incrementally strengthens big data quality. Before big data analytics, the specific improvement of big data quality, can effectively improve the results of big data analytics and reduce the failure risk of big data project. IIDQI procedure migrates the characteristics of IID to the data preprocessing operations and generates the following contributions:

- Incremental data processing can efficiently improve data change and extension ability of big data application.
- Iterative and continuous inspection manner can increase the opportunity of communication and interaction among users, DE and stakeholders.
- IDQI procedure can enhance the critical data quality of big data application that include basic quality, usability, maintainability and manageability.

Acknowledgment

The research was supported by Ministry of Science and Technology research project funds (Project No.: MOST 105-2221-E-158-002)

References

- [1] Cartner. (2015). The gartner business intelligence and analytics summit. Retrieved July 23, 2017 from

<http://www.gartner.com/newsroom/id/3130017>

- [2] Almquist, E., Senior, J., & Springer, T. (2015). Three promises and perils of big data, bain and company, Inc.
- [3] Meng, X. F., & Ci, X. (2013). Big data management: Concepts, techniques and challenges. *Journal of Computer Research and Development*, 50(1), 146–169.
- [4] Lackey, D. A. (2016). *The Big, Big Data Workbook*. blazon.online .
- [5] Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data Science Journal*.
- [6] Saha, B., & Srivastava, D. (2014) Data quality: The other face of Big Data. *Proceedings of the 2014 IEEE 30th International Conference on Data Engineering*.
- [7] Taleb, I., Dssouli, R., & Serhani, M. A. (2015). Big data pre-processing: A quality framework. *Proceedings of the 2015 IEEE International Congress*.
- [8] Deshpande, Bala: 5 situations which drive data pre-processing before data mining. (2013). Retrieved from <http://www.simafore.com/blog/bid/116618/5-situations-which-drive-data-pre-processing-before-data-mining>
- [9] Larman, C. (2004). *Agile and Iterative Development: A Manager's Guide*. Addison-Wesley Professional.
- [10] Szalvay, V. (2004). *An Introduction to Agile Software Development*. CollabNet, Inc.
- [11] Cockburn, A. (2002). *Agile Software Development*. Boston: Addison-Wesley.
- [12] Beck, K., Beedle, M., Van, B. A., Cockburn, A., Cunningham, W., Fowler, M., ... & Kern, J. (2001). *Manifesto for Agile Software Development*.
- [13] Clancy, T. (2014). Chaos report. *The Standish Group Report*.
- [14] Larman, C., & Victor, R. B. (2003) Iterative and incremental developments. A brief history. *Computer*, 36(6), 47-56.
- [15] Martin, R. C. (1999). Iterative and incremental development (IID). *Engineering notebook column*.
- [16] Duvall, P. M., Matyas, S., & Glover, A. (2007). Continuous integration: improving software quality and reducing risk. *Pearson Education*.
- [17] Zikopoulos, P., & Eaton, C. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media.
- [18] Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences*, 275, 314-347.
- [19] Wagner, D. (2014, October). The importance of big data analytics in business, World of tech, Retrieved August 20, 2017 from <http://http://www.techradar.com/news/world-of-tech/the-importance-of-big-data-analytics-in-business-1267606/2>
- [20] Elgendy, N., & Elragal, A. (2014). Big data analytics: A literature review paper. *Lecture Notes in Computer Science*, 214-227.
- [21] Al-Ahmad, W., Al-Fagih, K., Khanfar, K., Alsamara, K., Abuleil, S., & Abu-Salem, H. (2009). A taxonomy of an IT project failure: root causes. *International Management Review*, 5(1), 93.
- [22] Dong, X. L., & Srivastava, D. (2013). Big data integration. *Proceedings of the 2013 IEEE 29th International Conference on Data Engineering (ICDE)*, 1245–1248.
- [23] Lai, S. T. (2016). Data quality measurement model for improving big data analysis efficiency and quality. *Proceedings of the International Symposium of Quality Management (ISQM)*, Taiwan.



Sen-Tarng Lai was born in Taiwan in 1959. He received his BS from Soochow University, Taiwan in 1982, the master from National Chiao Tung University, Taiwan in 1984 and the PhD from National Taiwan University of Science and Technology, Taiwan in 1997. His research interests include software security, software project management, and software quality. He is currently an assistant professor in the Department of Information Technology and Management at Shin Chien University, Taipei, Taiwan.