Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM
Volume 7 Issue 27 (September 2022) PP. 168-186
DOI: 10.35631/JISTM.727014

JOURNAL OF INFORMATION SYSTEM AND TECHNOLOGY MANAGEMENT
(JISTM)
www.jistm.com

AN IMPACT ANALYSIS OF EXTRACT TRANSFORM LOAD PROCESS FOR MAINTAINING THE SYSTEM OF DATA WAREHOUSE

Azman Ta'a[1*], Norzila Ishak[2], Ezanee Mohamed Elias[3], Norlila Mahidin[4]

[1] School of Computing, College of Arts and Science, UUM, 06010 Sintok, Kedah
Email: azman@uum.edu.my
[2] Seberang Perai Polytechnique, Jalan Permatang Pauh, 13500 Permatang Pauh, Pulau Pinang
Email: norzilaishak1851@gmail.com
[3] School of Technology Management & Logistics, UUM, 06010 Sintok, Kedah
Email: ezanee@uum.edu.my
[4] Universiti Utara Malaysia
Email: 3dnorlila@gmail.com
[*] Corresponding Author

Article Info:

Article history:
Received date: 19.04.2022
Revised date: 12.06.2022
Accepted date: 13.07.2022
Published date: 05.09.2022

To cite this document:

Taa, A., Ishak, N., Mohamed Elias, E., & Mahidin, N. (2022). An Impact Analysis of Extract Transform Load Process for Maintaining the System of Data Warehouse. *Journal of Information System and Technology Management, 7* (27), 168-186.

DOI: 10.35631/JISTM.727014

Abstract:

The Extract Transform Load (ETL) process involves extracting data from database sources, transforming them into a suitable form for research analysis, and then loading it into a data warehouse (DW) to support effective decision-support implementation. To maintain the target of the data warehouse, several issues are discussed in the DW life cycle, ETL processes, and impact analysis for maintaining the DW. This research focuses on the issue high frequency of data changes makes ETL processes difficult to propagate the data changes and to maintain the changes history in the DW life cycle. Therefore, the focus issues of this research are identifying factors for frequent data changes that occur in data sources and the DW structure that need to be modified by performing impact analysis. The general factors of data changes in DW were identified by questionnaire from 41 respondents, and the factor of impact analysis was evaluated using the statistic test method called Kruskal Wallis H Test to make a comparison between the impact analysis factor and the category of users. The aims of this research are to perform the impact analysis of the DW process in order to maintain the DW system and to achieve DW with the right requirements definition. In addition, this will help users working with the DW to understand the elements of impact analysis in DW, especially on how to ensure the DW process runs efficiently and successfully. Therefore, the database administrators, data analysts, and DW developers can utilize these research findings as a guideline to deal with the data changes in the DW process.

Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM
Volume 7 Issue 27 (September 2022) PP. 168-186
DOI: 10.35631/JISTM.727014

Keywords:

Impact Analysis, DW Life Cycle, ETL Process, Data Changes

Introduction

A data warehouse (DW) is a huge collection of business data. It contains business current as well as historical data that is used to benefit and effectively support an organization's decision-making. Nowadays, numerous business organizations (BO) have turned to data warehousing to ensure the accuracy and quality of their strategic business decisions. This is because, by utilizing the DW and other possible drivers in business decisions, the BO can distinguish whether the findings of the analysis can be falling into the profit or loss category (Gonzales et al., 2015).

The DW's life cycle involves many components and processes. The main component of a DW system contains data sources, the Extract Transform Load (ETL) process, DW storage, and the front-end applications (Ta'a, Abdullah, & Norwawi, 2011). Importantly, the data is gathered from a variety of sources. It can be sourced from heterogeneous data and then transported for reporting and analytics purposes after going through the extraction, transform, and loading (ETL) process (Gupta & Sahayadhas, 2020). There are a lot of steps in the DW development architecture. These steps include looking at operational systems, figuring out what needs to be done, designing, implementing, and testing, then deploying and maintaining the system after it's done (Asrani & Jain, 2016). Depending on how frequently the BO receives new data, data changes may occur. Any changes in data sources or the target DW are an onerous responsibility to the ETL process (Ralph Kimball & Joe Caserta, 2004). Therefore, a systematic process to examine and evaluate the changes in these data storage objects is crucial to maintaining the success of DW implementation. This systematic process referred to as impact analysis needs to be performed together by the project manager, database administrator (DBA), and DW modeler to ensure the appropriate actions can be taken.

Moreover, most current research in this area focuses deeply on the technique required for data changes issue and the lack of knowledge and awareness to identify general factors the impact analysis of data changes in the DW environment. This is because there are fewer studies that look specifically at users' perceptions of the impact of changes in data sources for DW implementation. As a result, the implementation of the DW is incomplete. The main objective of this research is to identify the general factor impact analysis of the DW process for maintaining the DW. And there are three sub-objectives; to discover the general factor for impact analysis in the DW, to perform an impact analysis factor for maintaining the DW processes, and to evaluate the factor of the impact analysis process)

Literature Review

This section presents a necessary background of information on the research's relevant literature. It is going to focus on the topic related to the theoretical issues of DW life cycle, data changes, ETL process, and impact analysis that covers the research problems.

Data Warehouse Life Cycle

A DW is a repository that integrates information from heterogeneous and autonomous sources in order to effectively implement the decision-support like reporting, visualization, business intelligence, etc (Thi, 2011). Figure 2.1 shows the DW processes.

Journal of Information System and Technology Management
EISSN: 0128-1666
Volume 7 Issue 27 (September 2022) PP. 168-186
DOI: 10.35631/JISTM.727014

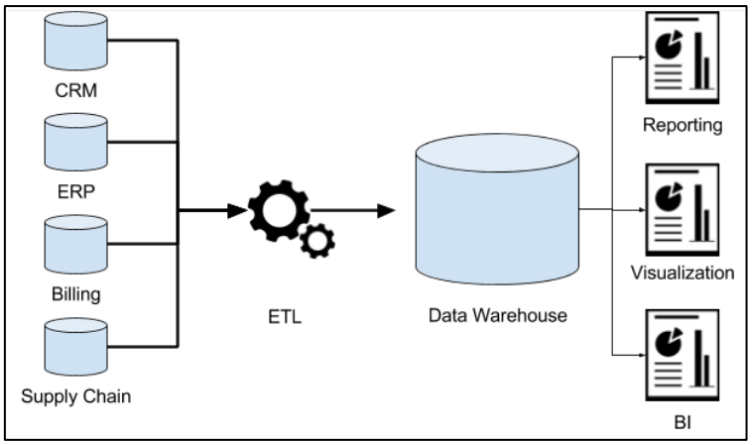

Figure 1: DW Processes (Pereira De Oliveira, 2019)

The challenges in DW environments include integrating, rearranging, and building up big quantities of data from many systems, therefore it is necessary to create a new collective information base for business intelligence (Vyas & Vaishnav, 2017). Thus, it will be able to provide data availability for BO performance, freshness, and prediction capabilities, as well as to provide an integrated information repository for strategic and tactical decisions (Bouaziz et al., 2017). So many approaches are proposed to manage the DW issues. All these issues can be defined as a system that represents the characteristics and the actual situation of the organization.

The challenges to developing a DW are:

- *Data Quality:* Data originates from a variety of diverse sources, spanning all facets of an organisation. When a data warehouse attempts to combine conflicting data from many sources, errors are discovered. Inconsistent data, logic conflicts, and missing or duplicated data all contribute to data quality difficulties. Due to poor data quality, optimal decisions are required, which results in ineffective reporting and analysis (Pandey, 2014);
- *Comprehend Analytics:* The excellent analytics tools and reports made available through integrated data will give credit to leaders who produce correct results that have a positive impact on the future success of their enterprises. To accomplish this, the user must be aware of the exact nature of the analysis that will be performed (Chen & Zhao, n.d.; Mukherjee & Kar, 2017; Tirumala et al., n.d.);
- *Performance:* To meet the comprehensive performance conditions, a DW must be prudently designed. Although the final product can be customized to adapt to the organization's performance requirements and to provide a stable foundation, the initial overall design must be carefully thought out (Bogale & Ababa Ethiopia, 2016; Lee et al., 2004; Tiwari et al., 2017);
- *DW Design and Cost:* People usually don't want to waste time defining the necessary parameters for a proper DW design, but they have a strong sense of what they want from DW. Nonetheless, they are unaware of all the implications of this perspective, making the time frame difficult to quantify, and there is significant confusion between technicians constructing DW and business users. The end result is a DW that does not produce the expected outcomes for the user. As a result, there has been a significant increase in development fees, as well as a significant increase in the cost of development (Maule, 2009);

Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM
Volume 7 Issue 27 (September 2022) PP. 168-186
DOI: 10.35631/JISTM.727014

- *Human Factor:* People are not interested to exchange their routines especially if the new process is not intuitive. This hesitation can be alleviated by having an inclusive user training program, but it will require additional planning and resources (Berhane et al., 2020);
- *Stakeholders and Designer's Gap:* DW actions are being carried out in an uncertain manner due to a lack of communication amongst the numerous parties involved in the process (Grillo, 2018), and *Data Warehouse Model and Schema:* The regular of DW schema changes have increased although tolerance for downtimes has almost disappeared. DW schema evolution eternal as an error-prone and time-consuming undertaking, because the Database Administrator (DBA) not have the suitable methods and tools to manage and automate this effort such as not forecasting and evaluating the effects of the suggested schema changes, rewriting queries, and requests to operate on new schema, and not migrate the DW.

Extract Transform Load (ETL)

ETL is a critical process in the DW process to make the BO strategies successful. The data in the DW system will keep in a variety of data storage systems, locations, and formats (Leslie Hendrie Spits Wamars, 2016). Usually, the data needs to be merged, cleaned, adjusted and summarized. ETL process contains three main processes, it is the procedure of extracting data from different data sources, transforming it according to the DW location demands, and loading it effectively into the DW destination (Tirumala et al., 2015). In the transformation process, data also need the cleaning operation to make data standardized and compatible with the destination database (Zekri et al., 2018a). To learn more about the ETL data process's functionality, BO can refer to Isabella's research (Isabella Johanna Swart, 2016). It clearly explains the functionality that is required for the data of ETL to start from the vestige to the DW environment. The following subsections explain the general ETL process. There are many issues of ETL that can be discussed but there are a few general issues that will make errors in the ETL process. When mistakes arise during the ETL process, the user should investigate whether data is missing from the data sources or if the information has been wrongly translated by the data transformation tool.

There are a variety of elements that can influence the issues that can generally cause errors in the ETL execution (Ardianto Wibowo, 2015; Jaiteg Singh & Kawaljeet Singh, n.d.; Romero & Abello, 2014) such as:

- Missing Requirements Information: Business requirements should be clear in stating how the system should behave and what actions it should and should not take under different conditions. Furthermore, the criteria should include every area of the ETL system, including its design and implementation. As a result, it is necessary to conceptually characterize the ETL operations at the outset to ensure that the user needs have been effectively understood;
- Data Is Coming from Multiple Sources: DW uses multiple sources of information to process the data, including information from existing database tables, information from external files (.txt,.xls,.xml), and so on. As a result, it is critical to keep an eye on and manage the information that is being acquired, as well as to ensure that all this data is processed in the same manner and, eventually, is connected to each one based on the join qualities;
- Incomplete Data: In the system, one of the difficulties is creating failure errors in the absence of certain properties from the data sources. Consider the following scenario: we

Journal of Information System and Technology Management
EISSN: 0128-1666
Volume 7 Issue 27 (September 2022) PP. 168-186
DOI: 10.35631/JISTM.727014

are utilizing a custom property to connect two data sources together. If this property is absent from one of the two sources, the execution of the ETL flow will be slowed;

- Incompatible and Duplicate Data: ETL process analysis can be affected by the entries that are generated by a system more than once, which can affect the accuracy and consistency of the results. However, it can be expected that the information supplied by the learning platforms does not contain any duplicate values unless otherwise stated, and
- High Volume or Capacity of Data: When dealing with a large volume or capacity of data, the performance of the system that creates the ETL operations might have a detrimental impact on the consistency of the data. Even though the volume of data from the data sources is not deemed to be big in this context, this performance issue has the potential to negatively impact the overall system. One possible option is to test and, if necessary, optimize the queries that were used during the fact table's generation.

Impact Analysis in Data Warehouse

Impact analysis examines the metadata that is connected to an object like a table or column. It determines the probable repercussions of a change or calculates what needs to be changed to effect a change (Tomingas, 2018). By doing a change to its structure or content, it will determine what is affected (Rajendrani Mukherjee et al., 2017). To properly load the DW, the processes of breaking and changing data-staging objects are crucial. The achievement of a project becomes harmful if ad-hoc changes to objects' data staging are allowed. Before any changes are made, impact analysis must be completed before a table is prepared in the staging area. Most ETL tool vendors provide an impact analysis function, but this function is often overlooked during ETL product evidence concepts. DW developers assume that it is not so important.

Analysis of the impact and function of ETL is a burdensome responsibility as there are continuous modifications to the resource system and DW goals need to be maintained. Only the ETL process can be used to know exactly what the changes are. Communication between the DW modeling team members like the ETL project manager, source system DBA, and others is important to make sure any system in DW is dependent on changes, and if so, suitable impact analysis needs to be performed. In the DW environment, the impact analysis permits to list of all of the attributes that would be affected by a suggested change (Ralph Kimball & Joe Caserta, 2004). Therefore, the impact of a change that occurs in any of the DW components should be able to be analyzed and be able to list all of the attributes in all other components. The questions about data change activity should be able to answer by an impact analysis solution (Ralph Kimball & Margy Ross, 2013). Some ETL tools are designed to respond to all of those questions and may use some technique. The ETL tool is very helpful in solving the questions because, without ETL tools, the DW team is essential to sustainably capturing all table and column changes from the source systems and mapping them into the DW to make sure data stay up to date.

Methodology

The research has passed in different phases and needs to follow all the phases. These contain three phases starting with finding the component of impact analysis, followed by performing the impact analysis and evaluating the results. In Phase 1, the initial phases of the study, renowned journals, publications, conference proceedings, and books were reviewed to clarify and understand the issues in impact analysis. In addition to these sources, the findings of numerous empirical studies were researched and analyzed. Based on the literature review, problem statements are defined, and study objectives are established in accordance with the

Journal of Information System and Technology Management
EISSN: 0128-1666
Volume 7 Issue 27 (September 2022) PP. 168-186
DOI: 10.35631/JISTM.727014
narrowed scope. Three research question is constructed based on the objectives of the study to be achieved, and they are listed below:

- RQ1: What is the general factor for impact analysis in a data warehouse?
- RQ2: How to perform an impact analysis factor for maintaining the DW processes?
- RQ3: How to evaluate the impact analysis process?

The method used in this phase is a literature survey where some fields such as DW life cycle process, ETL process, and the impact analysis of data changes issues. In phase 2, a questionnaire was developed to be the best instrument for collecting the data in this study. The questionnaire, which is used in this study, is based on previous studies and the researchers' assessments and adapted from Adeoye et al. (2011) and Asrani and Jain (2016) to suit the objectives and requirements of the study. The questionnaire's main goal is to determine the relevance of hypotheses to perform an impact analysis factor for maintaining the DW. It is vital to get sufficient information and responses from potential respondents. The category of the respondents is top management, middle management, developer, and operational from any BO involved in any economic sector. According to the central limit theorem (CLT), it states the minimum sample size is equal to or greater than 30. Therefore, for this survey, the minimum number of respondents from BO will be 30 as it is considered sufficient for the CLT to stand.

The questionnaire was designed based on a web-based concept using the online google forms. The questionnaire is broken into two parts. The first section inquiries about the respondent's background, such as the respondent's organizational category (position), experience working in DW implementation, and the DW tools that the respondent employs at their organization. The second section of the questionnaire contains the twelve questions. Each question offers a single hypothesis and requires the user to respond using a Likert scale. The Likert scale evaluation is represented by 1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, and 5 = Strongly Agree. Respondents fill out the questionnaire based on their personal experience with a data warehouse deployment in their organization. The list of questionnaires can refer to the explanations in Table 1.

In phase 3, the study's hypotheses were statistically examined in order to answer the study's questions, achieve its goals, and determine whether or not there is a positive link between independent and dependent variables. This study is non-parametric since the data collected is of the nominal and ordinal types, and the statistical tests to be performed are the Kruskal-Wallis tests. The test examines whether two or more groups' medians differ. The Kruskal-Wallis's test will determine whether or not there is a statistically significant difference between groups. As a result, the data were statistically analyzed in the final stages of the study using the Python language and a few libraries in Python such as Pandas, Pingouin, Seaborn, Statsmodels, Scipy, and Scikit_posthocs to support the analysis in order to get appropriate descriptive and inferential statistics, such as means, standard deviation and p values.

Result and Analysis

This section discusses the findings based on the three research questions formulated in this study. They are RQ1: What are components for impact analysis in DW? RQ2: How to perform an impact analysis factor for maintaining the DW processes? and RQ3: How to evaluate the factor of impact analysis process?

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 27 (September 2022) PP. 168-186
DOI: 10.35631/JISTM.727014

The General Factor of The Impact Analysis

Based on Table 1, it shows the list of general factors studied by some reaches about DW life cycle, ETL process, and impact analysis for data change issues. From the factor of the impact analysis, few hypotheses were generated to support the idea. A hypothesis's goal is to find an answer to a question RO1. It is designed to support the idea that all general factors significantly affect the impact analysis of the DW process. Based on the factor in Table 2, the study's hypotheses were formulated, and they are listed below:

- H1: H_0 = DW user knowledge factor about data changes has a significant effect on the DW implementation.
  H_a = DW user knowledge factor about data changes not significant effect on the DW implementation.
- H2: H_0 = DW data source changes factor has a significant effect on the DW process.
  H_a = DW data source changes factor does not have a significant effect on the DW process.
- H3: H_0 = A good DW factor has a significant influence on making better business decisions.
  H_a = A good DW factor does not have a significant influence on making a better business decision.

Table 1: Finding Factor of Impact Analysis.

No	Factor	Explanation	Hypothesis Classification
1	Data source	As many data sources as feasible should be supported by the DW, which should also have the capability of integrating more data sources with the reporting, modeling, and analytical approach soon.	H2
2	Organization Functionality	Close synchronization between functional area units within a company can assist managers in making judgments about how the units work together as a whole and how they affect one another in the process of doing business.	H3
3	Data Access	Users should be able to obtain data changes more simply and quickly, and they should be able to extract specific reports that will help them make more informed decisions faster.	H2
4	Classification Data Changes	Users should be able to obtain data changes more readily and quickly, and they should be able to extract single reports that can assist them in making decisions in a shorter amount of time. Through the structured queries and reporting created by DW, organizations should be able to discover which aspects of data changes in the business are generating more money.	H1
5	Awareness	Users should be aware of the data changes so that the data process can be adjusted to support the required data analysis.	H1

Journal of Information System and Technology Management
EISSN: 0128-1666
Volume 7 Issue 27 (September 2022) PP. 168-186
DOI: 10.35631/JISTM.727014

6	Report Generation	A single report that can depict the changing state of the business and its components in real-time must be available to senior management.	H1
7	Adjustment Support	Data warehouses are capable of adapting data changes to new business processes and responding rapidly to future and new requirements.	H2
8	Real Data	Objects and events specified in the data warehouse are accurately represented by the data values stored in the data warehouse.	H3
9	Query Simultaneous	The DW can deal with the complexity and the number of simultaneous query data changes without affecting the overall performance of the system.	H2
10	Data Correctness	In DW, the level of data accuracy should be as high as possible.	H3
11	Frequency	The frequency of data changes often occurs in the data sources.	H1
12	Consistent and Accuracy	Consistent and accurate data in DW will produce better reporting or dashboarding.	H3

The Results

The results of the study are presented in this section, which was developed with the support of the Python and Pandas data frames. This study examines respondents' feedback based on descriptive statistics, Cronbach's alpha, and inferential analysis, all of which are explored in depth. It gives the demographic background of 41 respondents for descriptive analysis, and then the study analyzes the Cronbach's alpha, which can demonstrate the reliability of all the variables when compared to the demographic background. Finally, to perform an inferential analysis, the Kruskal-Wallis test will be used to quantify the analysis based on the category of respondents' opinions for each factor. After that, a post-hoc test like the Dunn test will be used to identify which category of respondents is significant. The last analysis is to indicate how strong the effect size of respondents' opinions using Epsilon square ($\boldsymbol{\varepsilon^2}$).

The Respondent's Demographic

There are 41 respondents from various organizations in Penang, who have experience in DW development. There are four types of category respondent positions in the BO. All the respondents agree to be part of the survey after reading and agreeing with the consent form provided prior to the survey. Figure 2 and 3 shows the demographic respondents' details of the survey. The percentage of responders in each category who completed the questionnaire is shown in Figure 2. According to the pie chart, the largest category of responses is operational, with 41.5 percent, middle management is second with 34.1 percent, developers are third with 22 percent, and top management is the least involved with 2.4 percent. Respondents come from a wide range of industries and used the DW in their business decisions. The data warehouse installation experience of respondents is depicted in Figure 3. The pie chart indicates that no respondent has more than 20 years of job experience, with 48.8 percent having less than five years, 31.7 percent having six to ten years, and 19.5 percent having eleven to twenty years.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 27 (September 2022) PP. 168-186
DOI: 10.35631/JISTM.727014

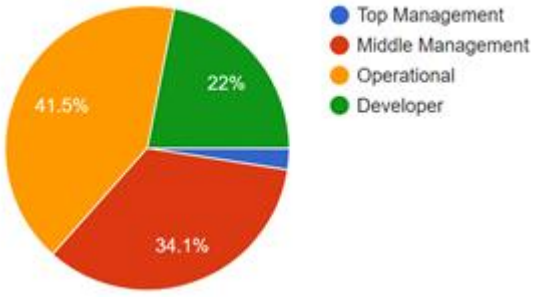

Figure 2: Category of Respondents

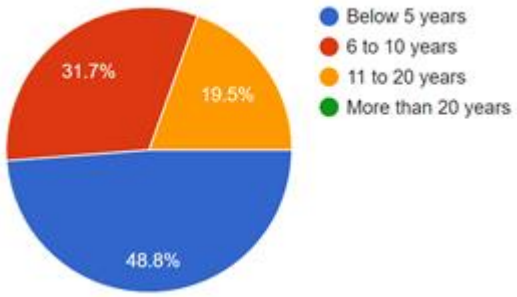

Figure 3: Experience Working In Data Warehouse Implementation

The Reliability Tests

The Cronbach's alpha (CA) test is used to measure the reliability of a multi-question Likert scale survey (Bonett & Wright, 2015). A high alpha score may indicate that the variables under examination are inextricably related (Taber, 2018). First, the examination of CA for all questions used to check the reliability of all questions is 0.8987851134706583, array (0.846, 0.939) is a good level for reliability value. The second is the CA test based on the hypothesis analysis item. Table 2 shows the CA result.

Table 2: CA Results

Item	Item No	CA Values	Level of Reliability
H1	4	0.6931022992335888	Acceptable
H2	4	0.7138233934350439	Good
H3	4	0.8216045732308299	Good

The CA value of the dependent variable of H1 is 0.693, as shown in Table 3. In addition, there are four other dependent variables items: Classification Data Changes, Awareness, Report Generation, and Frequency. The CA value of H2, with 0.713 and 4 items there are four other dependent variables items: Data source, Data Access, Adjustment Support, and Query Simultaneous. The CA value of H3, with 0.821 and 4 items, is the highest, indicating that it is the most dependable variable item: Organization Functionality, Real Data, Data Correctness, and Consistent and Accuracy.

Table 3 shows the percent of data for each level of the question. It may be considered that the whole category of respondents strongly agreed on all of the hypotheses for the factor of impact analysis. To ensure the accuracy of the result hypothesis analysis, a statistical test will be performed on this data set. This analysis will determine the significance of the hypothesis's independent and dependent variables. The independent of the analysis refers to the category of the respondent, while the dependent refers to the factor of impact analysis. The values of the dependent variable are not normally distributed. Consequently, a non-parametric (Kruskal-Wallis Test) test is suitable to be used for analyzing the data (Lalanne & Mesbah, 2016; Xia, 2020).

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 27 (September 2022) PP. 168-186
DOI: 10.35631/JISTM.727014

Table 3: Result of The Questionnaire Based on The Percentage

Questions	Percentage of Data				
	1 = Strongly Disagree	2 = Disagree	3 = Neutral	4 = Agree	5 = Strongly Agree
Q1	0.00	0.00	0.00	31.71	68.29
Q2	0.00	0.00	9.76	48.78	41.46
Q3	0.00	0.00	0.00	41.46	58.54
Q4	0.00	2.44	0.00	39.02	58.54
Q5	0.00	2.44	7.32	34.15	56.10
Q6	0.00	0.00	0.00	48.78	51.22
Q7	0.00	0.00	2.44	36.59	60.98
Q8	0.00	0.00	9.76	34.15	56.10
Q9	0.00	0.00	9.76	29.27	60.98
Q10	0.00	0.00	12.20	29.27	58.54
Q11	0.00	2.44	7.32	39.02	51.22
Q12	0.00	0.00	0.00	26.83	73.17

The Kruskal Wallis Test

Table 4 shows the result of all the questions using the Kruskal Wallis Test. The following sample code in Figure 4 and Figure 5 is used to process data and generate the output seen below. Replace the DV value to measure other questions.

```
kq1 = kruskal(data=df, dv="Q1", between = "Category")
print("Q1= ",kq1)
```

Figure 4: Code Test Using Kruskal Wallis

```
Q1=              Source  ddof1         H     p-unc
Kruskal  Category       3  2.607761  0.456131
```

Figure 5: Output of test using Kruskal Wallis

Table 4: Result of Kruskal Wallis testing

Question	Type of Values		
	DF	H	P
Q1	3	2.607761	0.456131
Q2	3	4.296365	0.231189
Q3	3	1.596044	0.660287
Q4	3	3.145709	0.369701
Q5	3	1.130162	0.769798
Q6	3	1.11827	0.772666
Q7	3	1.50966	0.680043
Q8	3	7.374723	0.060866
Q9	3	4.855696	0.182674
Q10	3	0.849131	0.837683
Q11	3	1.087744	0.780033
Q12	3	4.455762	0.21627

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 27 (September 2022) PP. 168-186
DOI: 10.35631/JISTM.727014

Results of Test for Q1

Figures 6, 7, and 8 are the results that show the demographics of Q1 that respondents answered from the survey based on the category of respondents that can be concluded about the data source as a factor of impact analysis regarding the survey component can be accepted.

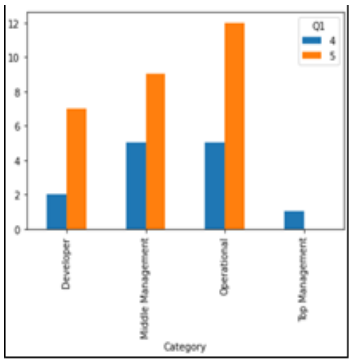

Figure 6: Data Source factor

	Developer	Middle Management	Operational	Top Management
Developer	1.000000	1.0	1.000000	0.703807
Middle Management	1.000000	1.0	1.000000	1.000000
Operational	1.000000	1.0	1.000000	0.872187
Top Management	0.703807	1.0	0.872187	1.000000

Figure 7: Data Source factor in Dunn Test

```
esq1 = HQ1*(n+1)/(n**2-1)
print("esq1 = ", esq1)
esq1 =  0.06519402632847988
```

Figure 8: Data Source factor in Epsilon square Test

The Kruskal-Wallis's test showed that all categories of respondent opinions about the data source factor significantly had a relatively moderate effect on the DW process, $\chi 2(3, N = 41)$ H = 2.607761, p > .05, ε^2 = .0651. Using Dunn's test with Bonferroni correction as a post-hoc test, it was discovered no different significance for all categories of respondent opinion about data source factor, p > .05.

Results of Test for Q2

Figures 9, 10, and 11 are the results that show the functionality factors of Q2 that respondents answered from the survey based on the category of respondents.

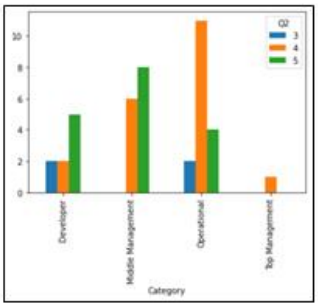

Figure 9: Organization Functionality factor

	Developer	Middle Management	Operational	Top Management
Developer	1.0	1.000000	1.000000	1.0
Middle Management	1.0	1.000000	0.301134	1.0
Operational	1.0	0.301134	1.000000	1.0
Top Management	1.0	1.000000	1.000000	1.0

Figure 10: Organization Functionality factor in Dunn Test

```
esq2 = HQ2*(n+1)/(n**2-1)
print("esq2 = ", esq2)
esq2 =  0.10740912981809358
```

Figure 11: Organization Functionality factor in Epsilon square Test

The Kruskal-Wallis's test showed that all categories of respondent opinions about organization functionality factor significant relatively very strong effect influence in making better business decisions, $\chi 2(3, N = 41)$ H =4.296365, p > .05, ε^2 = .1074. Using Dunn's test with Bonferroni correction as a post-hoc test, it was discovered no different significance for all categories of respondent opinion about organization functionality factor, p > .05.

Results of Test for Q3

Figures 12, 13, and 14 are the results that show the data access factor of Q3 that respondents answered from the survey based on the category of respondents.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 27 (September 2022) PP. 168-186
DOI: 10.35631/JISTM.727014

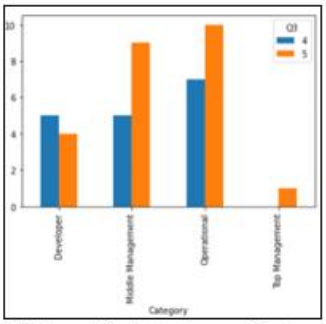

Figure 12: Data Access factor

	Developer	Middle Management	Operational	Top Management
Developer	1.0	1.0	1.0	1.0
Middle Management	1.0	1.0	1.0	1.0
Operational	1.0	1.0	1.0	1.0
Top Management	1.0	1.0	1.0	1.0

Figure 13: Data Access factor in Dunn Test

```
esq3 = HQ3*(n+1)/(n**2-1)
print("esq3 = ", esq3)
esq3 =  0.039901091155417005
```

Figure 14: Data Access factor in Epsilon square Test

The Kruskal-Wallis's test showed that all categories of respondent opinions about the data access factor significantly had a relatively moderate effect on the DW process, $\chi 2(3, N = 41)$ H = 1.596044, p > .05, ε^2 = .0399. Using Dunn's test with Bonferroni correction as a post-hoc test, it was discovered no different significance for all categories of respondent opinion about data access factor, p > .05.

Result of Test for Q4

Figures 15, 16, and 17 are the results that show the classification data change factor of Q3 that respondents answered from the survey based on the category of respondents.

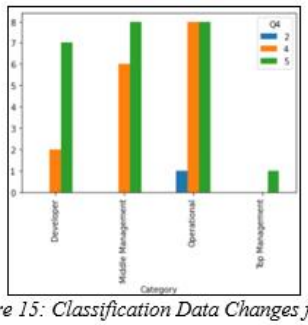

Figure 15: Classification Data Changes factor

	Developer	Middle Management	Operational	Top Management
Developer	1.000000	1.0	0.707546	1.0
Middle Management	1.000000	1.0	1.000000	1.0
Operational	0.707546	1.0	1.000000	1.0
Top Management	1.000000	1.0	1.000000	1.0

Figure 16: Classification Data Changes factor in Dunn Test

```
esq4 = HQ4*(n+1)/(n**2-1)
print("esq4 = ", esq4)
esq4 =  0.07864271983303105
```

Figure 17: Classification Data Changes factor in Epsilon square Test

The Kruskal-Wallis's test showed that all categories of respondent opinions about classification data changes factor significant relatively moderate effect on the DW implementation, $\chi 2(3, N = 41)$ H = 3.145709, p > .05, ε^2 = .0786. Using Dunn's test with Bonferroni correction as a post-hoc test, it was discovered no different significance for all categories of respondent opinion about classification data changes factor, p > .05.

Result of Test for Q5

Figures 18, 19, and 20 are the results that show the awareness factor of Q5 that respondents answered from the survey based on the category of respondents. The Kruskal-Wallis's test showed that all categories of respondent opinions about awareness factors significantly had a relatively weak effect on the DW implementation, $\chi 2(3, N = 41)$ H = 1.130162, p > .05, ε^2 = .0282. Using Dunn's test with Bonferroni correction as a post-hoc test, it was discovered no different significance for all categories of respondent opinion about awareness factor, p > .05.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 27 (September 2022) PP. 168-186
DOI: 10.35631/JISTM.727014

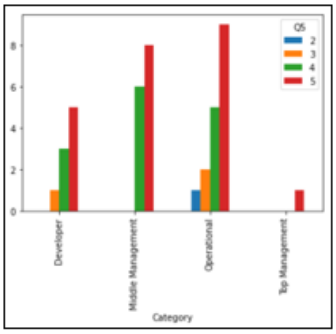

Figure 18: Awareness factor

	Developer	Middle Management	Operational	Top Management
Developer	1.0	1.0	1.0	1.0
Middle Management	1.0	1.0	1.0	1.0
Operational	1.0	1.0	1.0	1.0
Top Management	1.0	1.0	1.0	1.0

Figure 19: Awareness factor in Dunn Test

```
esq5 = HQ5*(n+1)/(n**2-1)
print("esq5 = ", esq5)
esq5 =  0.02825404920220848
```

Figure 20: Awareness factor in Epsilon square Test

Result of Test for Q6

Figures 21, 22, and 24 are the results that show the report generation factor of Q6 that respondents answered from the survey based on the category of respondents.

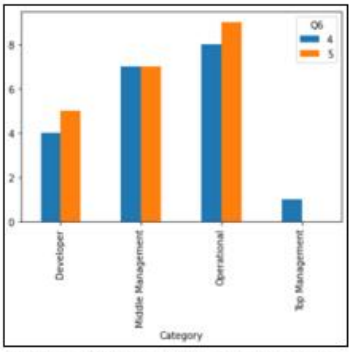

Figure 21: Report Generation factor

	Developer	Middle Management	Operational	Top Management
Developer	1.0	1.0	1.0	1.0
Middle Management	1.0	1.0	1.0	1.0
Operational	1.0	1.0	1.0	1.0
Top Management	1.0	1.0	1.0	1.0

Figure 22: Report Generation factor in Dunn Test

```
esq6 = HQ6*(n+1)/(n**2-1)
print("esq6 = ", esq6)
esq6 =  0.027956738250856004
```

Figure 23: Report Generation factor in Epsilon square Test

The Kruskal-Wallis's test showed that all categories of respondent opinions about the report generation factor significantly had a relatively weak effect on the DW implementation, $\chi 2(3, N = 41)$ H = 1.11827, p > .05, ε^2 = .0279. Using Dunn's test with Bonferroni correction as a post-hoc test, it was discovered no different significance for all categories of respondent opinion about report generation factor, p > .05.

Result of Test for Q7

Figures 24, 25, and 26 are the results that show the adjustment support factor of Q7 that respondents answered from the survey based on the category of respondents.

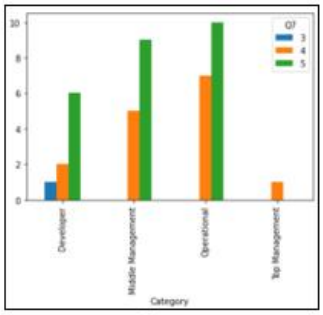

Figure 24: Adjustment Support

	Developer	Middle Management	Operational	Top Management
Developer	1.0	1.0	1.0	1.0
Middle Management	1.0	1.0	1.0	1.0
Operational	1.0	1.0	1.0	1.0
Top Management	1.0	1.0	1.0	1.0

Figure 25: Adjustment Support in Dunn Test

```
esq7 = HQ7*(n+1)/(n**2-1)
print("esq7 = ", esq7)
esq7 =  0.03774150685915341
```

Figure 26: Adjustment Support in Epsilon square Test

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 27 (September 2022) PP. 168-186
DOI: 10.35631/JISTM.727014

The Kruskal-Wallis test showed that all categories of respondent opinions about the adjustment support factor significantly had a relatively weak effect on the DW process, $\chi 2(3, N = 41)$ H = 1.50966, p > .05, ε^2 = .0377. Using Dunn's test with Bonferroni correction as a post-hoc test, it was discovered no different significance for all categories of respondent opinion about adjustment support factor, p > .05.

Result of Test for Q8

Figures 27, 28, and 29 are the results that show the report generation factor of Q8 that respondents answered from the survey based on the category of respondents.

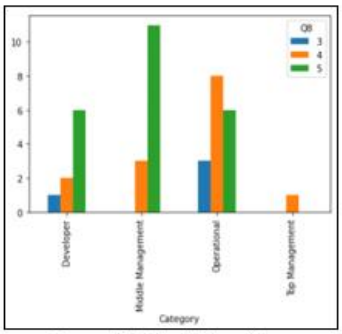

Figure 27: Real Data factor

	Developer	Middle Management	Operational	Top Management
Developer	1.00000	1.000000	0.861340	1.0
Middle Management	1.00000	1.000000	0.072915	1.0
Operational	0.86134	0.072915	1.000000	1.0
Top Management	1.00000	1.000000	1.000000	1.0

Figure 28: Real Data factor in Dunn Test

```
esq8 = HQ8*(n+1)/(n**2-1)
print("esq8 = ", esq8)
esq8 =  0.1843680748909505
```

Figure 29: Real Data factor in Epsilon square Test

The Kruskal-Wallis's test showed that all categories of respondent opinions about real data factors significantly had a relatively very strong effect influence on making better business decisions, $\chi 2(3, N = 41)$ H = 7.374723, p > .05, ε^2 = .1843. Using Dunn's test with Bonferroni correction as a post-hoc test, it was discovered no different significance for all categories of respondent opinion about real data factor, p > .05.

Result of Test for Q9

Figures 30, 31, and 32 are the results that show the query simultaneous factor of Q9 that respondents answered from the survey based on the category of respondents.

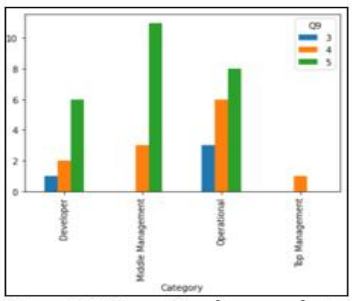

Figure 30: Query Simultaneous factor

	Developer	Middle Management	Operational	Top Management
Developer	1.0	1.000000	1.000000	1.0
Middle Management	1.0	1.000000	0.316431	1.0
Operational	1.0	0.316431	1.000000	1.0
Top Management	1.0	1.000000	1.000000	1.0

Figure 31: Query Simultaneous factor in Dunn Test

```
esq9 = HQ9*(n+1)/(n**2-1)
print("esq9 = ", esq9)
esq9 =  0.12139240986857613
```

Figure 32: Query Simultaneous factor in Epsilon square Test

The Kruskal-Wallis's test showed that all categories of respondent opinions about query simultaneous factor significantly had a relatively very strong effect on the DW process, $\chi 2(3, N = 41)$ H = 4.855696, p > .05, ε^2 = .1213. Using Dunn's test with Bonferroni correction as a post-hoc test, it was discovered no different significance for all categories of respondent opinion about query simultaneous factor, p > .05.

Journal of Information System and Technology Management
EISSN: 0128-1666
Volume 7 Issue 27 (September 2022) PP. 168-186
DOI: 10.35631/JISTM.727014

Result of Test for Q10

Figures 33, 34, and 35 are the results that show the report generation factor of Q3 that respondents answered from the survey based on the category of respondents.

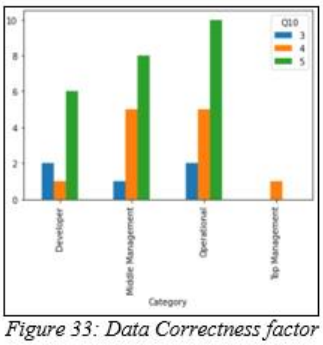

Figure 33: Data Correctness factor

	Developer	Middle Management	Operational	Top Management
Developer	1.0	1.0	1.0	1.0
Middle Management	1.0	1.0	1.0	1.0
Operational	1.0	1.0	1.0	1.0
Top Management	1.0	1.0	1.0	1.0

Figure 34: Data Correctness factor in Dunn Test

```
esq10 = HQ10*(n+1)/(n**2-1)
print("esq10 = ", esq10)
esq10 =  0.021228283214721543
```

Figure 34: Data Correctness factor in Epsilon square Test

The Kruskal-Wallis's test showed that all categories of respondent opinion about data correctness factor significantly had a relatively weak effect influence on making better business decisions, $\chi 2(3, N = 41)$ H = 0.849131, p > .05, ε^2 = .0212. Using Dunn's test with Bonferroni correction as a post-hoc test, it was discovered no different significance for all categories of respondent opinion about data correctness factor, p > .05.

Result of Test for Q11

Figures 35, 36, and 37 are the results that show the frequency factor of Q11 that respondents answered from the survey based on the category of respondents.

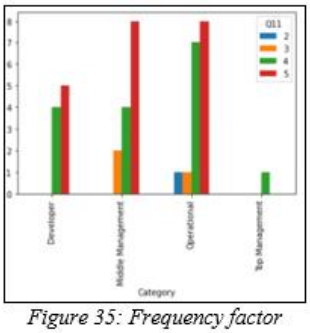

Figure 35: Frequency factor

	Developer	Middle Management	Operational	Top Management
Developer	1.0	1.0	1.0	1.0
Middle Management	1.0	1.0	1.0	1.0
Operational	1.0	1.0	1.0	1.0
Top Management	1.0	1.0	1.0	1.0

Figure 36: Frequency factor in Dunn Test

```
esq11 = HQ11*(n+1)/(n**2-1)
print("esq11 = ", esq11)
esq11 =  0.027193605762673694
```

Figure 37: Frequency factor in Epsilon square Test

The Kruskal-Wallis test showed that all categories of respondent opinion about frequency factor significantly had a relatively weak effect on the DW implementation, $\chi 2(3, N = 41)$ H = 1.087744, p > .05, ε^2 = .0271. Using Dunn's test with Bonferroni correction as a post-hoc test, it was discovered no different significance for all categories of respondent opinion about frequency factor, p > .05.

Result of Test for Q12

Figures 38, 39, and 40 are the results that show the report generation factor of Q12 that respondents answered from the survey based on the category of respondents. The Kruskal-Wallis's test showed that all categories of respondent opinions about consistent and accuracy factor significantly had a relatively very strong effect influence on making better business decisions, $\chi 2(3, N = 41)$ H = 4.455762, p > .05, ε^2 = .1113. Using Dunn's test with Bonferroni correction as a post-hoc test, it was discovered no different significance for all categories of respondent opinion about consistent and accuracy factor, p > .05.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 7 Issue 27 (September 2022) PP. 168-186
DOI: 10.35631/JISTM.727014

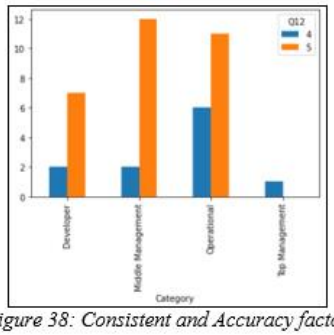

Figure 38: Consistent and Accuracy factor

	Developer	Middle Management	Operational	Top Management
Developer	1.000000	1.000000	1.000000	0.599932
Middle Management	1.000000	1.000000	1.000000	0.389335
Operational	1.000000	1.000000	1.000000	0.965787
Top Management	0.599932	0.389335	0.965787	1.000000

Figure 39: Consistent and Accuracy factor in Dunn Test

```
esq12 = HQ12*(n+1)/(n**2-1)
print("esq12 = ", esq12)
esq12 =  0.11139405257052266
```

Figure 40: Consistent and Accuracy factor in Epsilon square Test

The Findings

Based on the results in Table 3, the degree of freedom (DF) is 3 for all the questions. By using the chi-square (X^2) table with an alpha value of 0.05, then the value of X^2 is 7.815. Therefore, based on the method used by Kruskal-Wallis's testing, if the H (Chi-square) value < X^2, then the null hypothesis is retained. Retaining the null hypothesis is reinforced again by analyzing the p-value for all the questions with the Dunn's test and Epsilon Square test. This proves that Kruskal-Wallis's testing showed no significant difference between the categories because of a p-value greater than 0.05 for all questions. Thus, the null hypothesis for all hypotheses is not rejected. That means, the good knowledge of DW, the data source changes, and the good DW factors have a significant impact on the DW implementation.

Assessment and Discussion

At the beginning of the study, three main hypotheses have been identified. Based on the results in Table 4, the H1 is measured based on four factors in terms of DW user knowledge about data changes and has a significant effect on the DW implementation. First, users' knowledge is significant for an organization to identify which parts of data changes in the business to bring more revenue through the structured queries and reporting generated by a data warehouse. Second, users' knowledge is significant with awareness about the data changes so that the data process can be adjusted to support the required data analysis. Third, user knowledge is significant for an organization's ability to generate a report that can describe the changing state of business data at all times. Lastly, the user knowledge is significant with the frequency of data changes occurring frequently in data sources.

The H2 is a measure based on four factors based on the DW data source changes that have a significant effect on the DW process. First, the DW data source changes are significant, with the DW supporting as many data sources as possible and having the ability to integrate more data sources with the reporting, modeling, and analysis strategy in the near future. Second, the DW data source changes are significant, with users being able to access data changes more readily and quickly, as well as extract unique reports that can help them make faster decisions. Third, DW data source changes are significant, with DW being able to adapt data changes to new business processes and support future and new needs easily. Lastly, DW data source changes are significant, with DW able to handle complexity and the number of simultaneous queries of data changes without impacting system performance.

Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM
Volume 7 Issue 27 (September 2022) PP. 168-186
DOI: 10.35631/JISTM.727014

The last hypothesis is H3, it measures based on four factors based on the good DW has a significant influence in making better BO's decisions. First, a good DW is significant with close synchronization among functional area units within the business, so that it can help managers make decisions on the overall working together of the units and how they affect each other. Second, having a good DW is significant with Data values in the data warehouse correctly representing the real-world objects and events being described. Third, a good DW is significant, with the level of data correctness in a DW being as high as possible. Lastly, a good DW is significant with consistent and accurate data in the data warehouse will produce better reporting or dashboarding. Lastly, a good DW is significantly consistent, and accurate data in a DW will produce better reporting or dashboarding.

Conclusion

This research can assist new users who want to use the DW as a tool to support the management in making a decision. Although the criteria examined are generic in general factor, if they are not taken carefully, the users might contribute to a failure by making inappropriate decisions. This is due to the fact that data changes are an ongoing process in the use of the DW system. Thus, the general issues or factors that have been discussed should not be taken lightly and seen as one of the most important parts of running the DW system. Moreover, the emergence of big data will drive the data changes more challenging and require an advanced method to manage the impact of data changes.

References

Adeoye, T., Raufu, O., & Omodara, O. (2011). Design of Data Warehouse and Business Intelligence System A case study of a Retail Industry. www.bth.se/com

Ardianto Wibowo. (2015). Problems and Available Solutions On The Stage of Extract, Transform, and Loading In Near Real-Time Data Warehousing (A Literature Study).

Asrani, D., & Jain, R. (2016). Designing a Framework to Standardize Data Warehouse Development Process for Effective Data Warehousing Practices. International Journal of Database Management Systems, 8(4), 15–32. https://doi.org/10.5121/ijdms.2016.8402

Astriani, W., & Trisminingsih, R. (2016). Extraction, Transformation, and Loading (ETL) Module for Hotspot Spatial Data Warehouse Using Geokettle. Procedia Environmental Sciences, 33, 626–634. https://doi.org/10.1016/j.proenv.2016.03.117

Ta'a, A., Abdullah, M. S., & Norwawi, N. M. (2011). Goal-Ontology ETL Processes Specification. Journal of Information and Communication Technology (JICT), 10, 15–43.

Bala, M., Boussaid, O., Alimazighi, Z., Bala, M., Boussaid, O., & Alimazighi, Z. (2015). Big-ETL: extracting-transforming-loading approach for Big Data. https://www.researchgate.net/publication/319302361

Berhane, A., Nabeel, M., & Grose, C. (2020). The impact of business intelligence on decision-making in public organisations. IEEE International Conference on Industrial Engineering and Engineering Management, 2020-December, 435–439. https://doi.org/10.1109/IEEM45057.2020.9309763

Bogale, T., & Ababa Ethiopia, A. (2016). ASSESSMENT OF WAREHOUSE PERFORMANCE: A CASE OF ETHIOPIAN TRADING ENTERPRISE.

Bonett, D. G., & Wright, T. A. (2015). Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. Journal of Organizational Behavior, 36(1), 3–15. https://doi.org/10.1002/job.1960

Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM
Volume 7 Issue 27 (September 2022) PP. 168-186
DOI: 10.35631/JISTM.727014
Bouaziz, S., Nabli, A., & Gargouri, F. (2017). From traditional data warehouse to real time data warehouse. Advances in Intelligent Systems and Computing, 557, 467–477. https://doi.org/10.1007/978-3-319-53480-0_46

El-Sappagh, S. H. A., Hendawi, A. M. A., & el Bastawissy, A. H. (2011). A proposed model for data warehouse ETL processes. Journal of King Saud University - Computer and Information Sciences, 23(2), 91–104. https://doi.org/10.1016/j.jksuci.2011.05.005

Grillo, A. (2018). Developing a Data Quality Scorecard that Measures Data Quality in a Data Warehouse.

Homayouni, H., Ghosh, S., & Ray, I. (2018). An approach for testing the extract-transform-load process in data warehouse systems. ACM International Conference Proceeding Series, 236–245. https://doi.org/10.1145/3216122.3216149

Isabella Johanna Swart. (2016). Inmon versus Kimball: The agile development of a data warehouse.

Jaiteg Singh, & Kawaljeet Singh. (n.d.). Statistically Analyzing the Impact of Automated ETL Testing on the Data Quality of a Data Warehouse. 2009.

Jamaluddin, M. S., Firdaus, N., & Azmi, M. (2016). EXTRACTION TRANSFORMATION LOAD (ETL) SOLUTION FOR DATA INTEGRATION: A CASE STUDY OF RUBBER IMPORT AND EXPORT INFORMATION (Vol. 78). www.jurnalteknologi.utm.my

J. Anitha. (2014). ETL Work Flow for Extract Transform Loading. In International Journal of Computer Science and Mobile Computing (Vol. 3, Issue 6). www.ijcsmc.com

Kabiri, A., & Chiadmi, D. (2013). SURVEY ON ETL PROCESSES. Journal of Theoretical and Applied Information Technology, 20(2). www.jatit.org

Lalanne, C., & Mesbah, M. (2016). Measures of Association, Comparisons of Means and Proportions for Two Samples or More. In Biostatistics and Computer-based Analysis of Health Data using Stata (pp. 25–57). Elsevier. https://doi.org/10.1016/b978-1-78548-142-0.50002-x

Lee, S. M., Hong, S., & Katerattanakul, P. (2004). IMPACT OF DATA WAREHOUSING ON ORGANIZATIONAL PERFORMANCE OF RETAILING FIRMS. In International Journal of Information Technology & Decision Making (Vol. 3, Issue 1). www.worldscientific.com

Leslie Hendrie Spits Wamars, H. (2016). Datawarehouser: A Data Warehouse artist who have ability to understand data warehouse schema pictures. www.wellcomecollection.org

Maule, A. (2009). Impact Analysis of Database Schema Changes.

Mukherjee, R., & Kar, P. (2017). A Comparative Review Of Data Warehousing ETL Tools With New Trends And Industry Insight. https://doi.org/10.1109/IACC.2017.183

Ong, T., Pradhananga, R., Holve, E., Iii, ;, & Kahn, M. G. (n.d.). A Framework for Classification of Electronic Health Data Extraction-Transformation-Loading Challenges in Data Network Participation.

Pandey, R. K. (2014). Data Quality in Data warehouse: problems and solution (Vol. 1). www.iosrjournals.orgwww.iosrjournals.org18|

Pereira De Oliveira, P. A. (2019). Data Warehouse Automation Trick or Treat?

Rajendrani Mukherjee, Pragma Kar, & Institute of Electrical and Electronics Engineers. (2017). A Comparative Review Of Data Warehousing ETLTools With New Trends And Industry Insight.

Ralph Kimball, & Joe Caserta. (2004). The Data Warehouse ETL Toolkit.

Ralph Kimball, & Margy Ross. (2013). The Data Warehouse Toolkit.

Romero, O., & Abello, A. (2014). ETL Testing Analyzer.

Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM
Volume 7 Issue 27 (September 2022) PP. 168-186
DOI: 10.35631/JISTM.727014
Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. Research in Science Education, 48(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Thi, D. (2011). Impact Analysis for On-Demand Data Warehousing Evolution.

Tirumala, S. S., Nandigam, D., & Katragadda, R. (2015). ETL tools for Data Warehousing: An empirical study of Open Source Talend Studio versus Microsoft SSIS Speaker Identification using Deep Learning Technologies View project ETL tools for Data Warehousing: An empirical study of Open Source Talend Studio versus Microsoft SSIS. https://www.researchgate.net/publication/271207443

Tiwari, P., Mishra, A. C., Kumar, S., Kumar, V., & Terfa, B. (2017). Improved performance of data warehouse. Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2017, 94–99. https://doi.org/10.1109/ICICCT.2017.7975167

Tomingas, K. (2018). Semantic Data Lineage and Impact Analysis of Data Warehouse Workflows.

Vassiliadis, P., & Simitsis, A. (n.d.). EXTRACTION, TRANSFORMATION, AND LOADING.

Vyas, S., & Vaishnav, P. (2017). A comparative study of various ETL process and their testing techniques in data warehouse. Journal of Statistics and Management Systems, 20(4), 753–763. https://doi.org/10.1080/09720510.2017.1395194

Xia, Y. (2020). Correlation and association analyses in microbiome study integrating multiomics in health and disease. In Progress in Molecular Biology and Translational Science (Vol. 171, pp. 309–491). Elsevier B.V. https://doi.org/10.1016/bs.pmbts.2020.04.003

Zekri, A., Massa Â Bi, M., Layouni, O., & Akaichi, J. (2018a). Trajectory ETL modeling. Smart Innovation, Systems and Technologies, 76, 380–389. https://doi.org/10.1007/978-3-319-59480-4_38

Zekri, A., Massa Â Bi, M., Layouni, O., & Akaichi, J. (2018b). Trajectory ETL modeling. Smart Innovation, Systems and Technologies, 76, 380–389. https://doi.org/10.1007/978-3-319-59480-4_38