THE COMPUTERTECH
(An International Peer Review Journal)
VOLUME 4; ISSUE 2 (JULY-DEC); (2018)
WEBSITE: THE COMPUTERTECH

# The Role of Data Quality Scorecards in Measuring Business Success

**Bharath Kishore Gudepu[1], Divya Sai Jaladi[2]**
[1]Senior in Formatica Developer, Transamerica, 10100 N Central Expy Ste 595, Dallas, TX 75231
[2]Senior Lead Application Developer, SCDMV, 10311 Wilson Boulevard, Blythewood, SC 29016, UNITED STATES

**Abstract**

The primary objective of this thesis is to create a data quality scorecard (DQS) that connects the data quality requirements of the Data Warehouse at Brewage Ltd Group with certain data quality dimensions. In the initial iteration, utilizing Design Science Research (DSR) as a paradigm, the artefact was constructed based on the findings of a comprehensive and thorough literature research conducted. A data quality scorecard (DQS) was devised. The initial iteration of the System Usability Scale (SUS) indicates that the data warehouse stakeholders deemed the Data Quality System (DQS) beneficial. The second iteration was performed to further assess the DQS by executing a trial in the FMCG sector, followed by a semi-structured interview. The theme analysis of the semi-structured interviews revealed that the stakeholders perceived the DQS as transparent, an extra reporting tool, integrative, user-friendly, consistent, and confidence-enhancing about the data. The timeliness data component was deemed superfluous, requiring an amendment to the DQS. The third iteration was executed using the same procedures as the second iteration, although with the revised DQS in the oil and gas sector. The findings from the third iteration indicate that DQS is an effective and user-friendly tool for everyday application. The study advances theoretical understanding by presenting an innovative methodology for DQS design. This was accomplished by ensuring the DQS architecture corresponds with the data quality concerns of the DW stakeholders and the data quality dimensions. This research establishes a robust foundation for the future by creating a DQS model that may serve as a basis for subsequent development.

**Keywords:** Data Quality Scorecards, Data Quality, Business Success, Data Governance, Data Management, Data Accuracy, Data Profiling, Metadata, Compliance, Analytics, Business Intelligence, Data Integrity, Enterprise Data, Performance Metrics, Decision-Making

## Introduction

### Brewing Ltd

Brewing Ltd is a brewing enterprise situated in Europe. Brewing Ltd is a prominent global company primarily engaged in the manufacturing, distribution, and sale of beer and soft beverages. Brewing Ltd is a subsidiary of the Brewing Company Group, which encompasses several other businesses. The Company employs more than 125,000 individuals and operates sixty-seven facilities across forty nations. The R&D (research and development) efforts are essential to the Brewing Company

group for enhancing the international brand's position, particularly Brewing Ltd, alongside the advancement of distinct regional brands. The principal aim of Brewing Ltd is to assemble a world-class team including highly qualified, motivated, ambitious, and open-minded individuals, dedicated to continuous development. The Company places significant focus on establishing a transparent organizational structure that delineates the roles of all employees [1-4].

Brewing Ltd is a GloCal organization that must strike an appropriate balance between global collaboration and the promotion of robust local projects and brands. The Company consistently generates substantial local value in each market by fostering proximity to consumers and customers, while simultaneously leveraging the advantages of group affiliation, disseminating best practices, capitalizing on scale, and centralizing and standardizing processes and functions across borders. The value development approach of Brewing Ltd fosters growth, enhances efficiency, and refines their methods. Achieving the appropriate GloCal equilibrium within the matrix structure is essential for Brewing Ltd to realize its ambitions and attain success. This study identifies firm facts, investors, media, careers, markets, contact information, and corporate social responsibility (CSR) as elements that remain relatively stable. This research aims to provide a case analysis of the products of Brewery Ltd [5-8].

**The Data Quality Issue at Brewery Ltd**

The value of data typically corresponds with a company's growth. The primary issue is the necessity for regular updates to enhance product development input for end consumers. Generally, the absence of even a single data point might precipitate significant issues for the firm. Given the complexity of the data warehouse, it is mostly impractical for a single function to maintain the entire database; a more effective approach is to partition the burden into different components for management. The Data Manager, Data Producer, and Data Custodian are crucial to executing this task securely. The delineation of these jobs is crucial, although it is not presently upheld by the system at Brewery Ltd.

**• Data Producer:**

The data producer at Brewery Ltd gathers raw data from many source systems and subsequently transmits it to the Data Custodian. The data producer is the one who engages with several operational units to methodically gather comprehensive data. The data producer initially gathers the region type in a distinct file from the others. The product type and its associated qualities cannot be concurrently connected, as the latter may vary on a daily basis. Consequently, it requires frequent updates.

**• Data Custodian:**

The data custodian at Brewery Ltd is responsible for gathering information from the data producer and subsequently transferring it into the database. The data custodian manages the company's database server. An SAP BI system is utilized for data warehousing. The organization emphasized the necessity for the data custodian to have an effective technical solution to manage the data within the data warehouse. SAP is the preeminent market leader and the most extensively utilized ERP system globally. The following is a detailed benefit chart of the SAP BI system:

**• Database Administrator:**

The Data Manager at Brewery Ltd is responsible for overseeing all data within the data warehouse. The primary role of the Data Manager is to verify the security of the Brewery Ltd data warehouse database. It is essential to adhere to particular rules during the testing procedure.

The data manager should focus on the following key points:

Will the additional data sources need any audit limits and the implementation of new security measures?

• Is the data on newly added users with limited access currently available?

Cryptography: Data encryption is a crucial procedure necessary for its transmission into the data warehouse. It is vital to store the info utilizing a distinctive encryption procedure. Access to the administrative part is limited to end users; only the Company's administrator is permitted to enter. The finance manager is restricted to accessing just the financial database. Each division of the organization possesses distinct administrative positions for various administrative functions. A comprehensive roles and authorization matrix is established to guarantee that only authorized data is accessed [9-12].

**• Information Recipient:**

A critical concern for Brewery Ltd is end-user security and authorization. The EEM (End User Experience Monitoring) solution, offered by SAP, is utilized by the Company to simulate the behavior of users authorized to access central servers at many locations and to execute business activities. Similar to the administrator, data consumers may monitor system availability and connection performance from the end user's perspective in real time.

**Chosen Data Dimension: -**

• Completeness: - The raw data from the operational house has been fully acquired and subsequently encrypted through an encryption process prior to its transmission to the data warehouse.

• Validity: It is important to maintain the database current and ensure it has a valid dataset when a new product is introduced to the market.

The organization emphasized that the precision of the data in the data warehouse is paramount. It was noted that every data in the data warehouse underwent many manual quality checks to verify the authenticity of the raw data sourced straight from the plant.

• Timeliness: Timeliness is employed to assess the age of the data within the data warehouse.

Consequently, the chronology of the data is invariably contingent upon external variables. The company's product alters the season. New products, flavors, and beer strengths are regularly produced or modified to align with the current season. In addition to these factors, the prices of the items exhibit seasonal fluctuations.

• Consistency: In Brewery Ltd., it was noted that brand mixing and flavor consistency are vital to the organization. This location was omitted from our investigation due to the lack of authorization for analyzing this specific data.

• Integrity: The Data Custodian is responsible for managing the database. Integrity is mostly employed to ascertain the validity of data. Consequently, it is imperative for data managers to verify the validity of the transmitted data. In addition to this, they must also assess the efficiency.

The evaluation of the implemented framework demonstrated that both criteria were met. The researcher specifically noted instances of merchandise with a total number of requests.

The alterations did not correspond with the figures computed for the legacy system. Consequently, the established quality objective is: "Attain 100% data consistency for the data warehouse views."

**Assessment of DQS**

Subsequent to the DQS evaluation, participants were questioned to obtain direct insights into the DQS's efficacy. The interview was documented with a dictation device and subsequently transcribed for additional research. The participants' details and the analytical subtleties are elaborated in the following chapters [13-17].

**Participants**

The semi-structured interview was conducted with participants from each stakeholder group. Twenty people from various data warehouse stakeholder groups participated in the interview. The major objective of the semi-structured interview was to obtain verbal feedback about the utilization of the data measurement scorecard. The participants also participated in the case study.

Table 12 below presents an overview of the participants' demographics.

| Measure | Item | Frequency | Percentage (%) |
|---|---|---|---|
| Stakeholder Role | Data Custodians | 5 | 25 |
| | Data Producers | 3 | 15 |
| | Data Managers | 2 | 10 |
| | Data Consumers | 10 | 50 |
| Age | 21-30 | 4 | 20 |
| | 30-45 | 13 | 65 |
| | 45 and Over | 3 | 15 |
| Years in Industry | 0-3 years | 2 | 10 |
| | 3-7 years | 3 | 15 |
| | 7 years + | 15 | 75 |

Each participant was informed about the evaluation's nature and provided their agreement for the interview. The researcher communicated to the stakeholders that the objective of the interview was to obtain their comments on the DQS to facilitate its future improvement.

**Data Acquisition Methodology**

The data gathering was conducted using a qualitative methodology. Subsequent to the completion of the run-through, participants from the stakeholder groups were interviewed to obtain their

perspectives on the efficacy of the DQS. The semi-structured interview is further upon in section below.

The main objective of the offered results is to address the following inquiries;

1. Does the implementation of the DQS significantly influence the perceptions of data warehouse stakeholders on the quality of data within their data warehouse?

2. Does the utilization of the DQS provide precise information on the quality of data within the data warehouse?

The interview questions were examined utilizing a qualitative data analysis method known as theme analysis. The complexities of the analytical technique and findings are detailed.

**Semi-Structured Interviews**

The semi-structured interviews enabled participants to offer insights into the utilization of the scorecard. Five individuals from each stakeholder group participated in the interview. Arthur and Nazroo (2003) emphasize the significance of meticulous preparation for interviews, and specifically the formulation of a "topic guide". Their main emphasis is on classifying themes to address rather than formulating particular inquiries for the interview. Having crucial questions written verbatim might be beneficial, not to enforce rigorous adherence to the wording, but to provide a reference for phrasing, particularly useful in the event of an interviewer's memory loss during the interview. Arthur and Nazroo recommend structuring the topic guide inside a framework that includes the following elements:

**• Introduction; • Initial inquiries; • Fundamental comprehensive inquiries; and • Conclusion.**

This planning aligns with the phases of an interview process delineated by Legard et al. (2003), who offer two perspectives on in-depth interviewing. One begins with the assumption that knowledge is 'given,' and that the researcher's role is to uncover it; although the phrase was not explicitly used, this aligns with a positivist perspective. The alternative method is constructivist: information is generated and mutual understanding is achieved via dialogue between the interviewer and interviewee. Legard et al. (p.143) underscore the significance of cultivating a rapport, asserting that the interviewer functions as a "research instrument." They also contend that researchers must possess "a degree of humility, the capacity to receive the participant's insights without the necessity to compete by showcasing their own." Subsequently, some participants sought clarification on the meanings of certain terms or phrases.

**Evaluation Outcomes**

A thematic analysis method was utilized to examine the qualitative data items. A combination of inductive and deductive methodologies was employed. The deductive theme analysis is an approach guided by the researcher's analytical or theoretical goals, whereas inductive thematic analysis is not.

*The analytical method is predominantly data-driven and entirely reliant on the participants' replies. The decision to employ a mixed methodology is driven by the need to mitigate study bias by facilitating the identification of possible new variables.

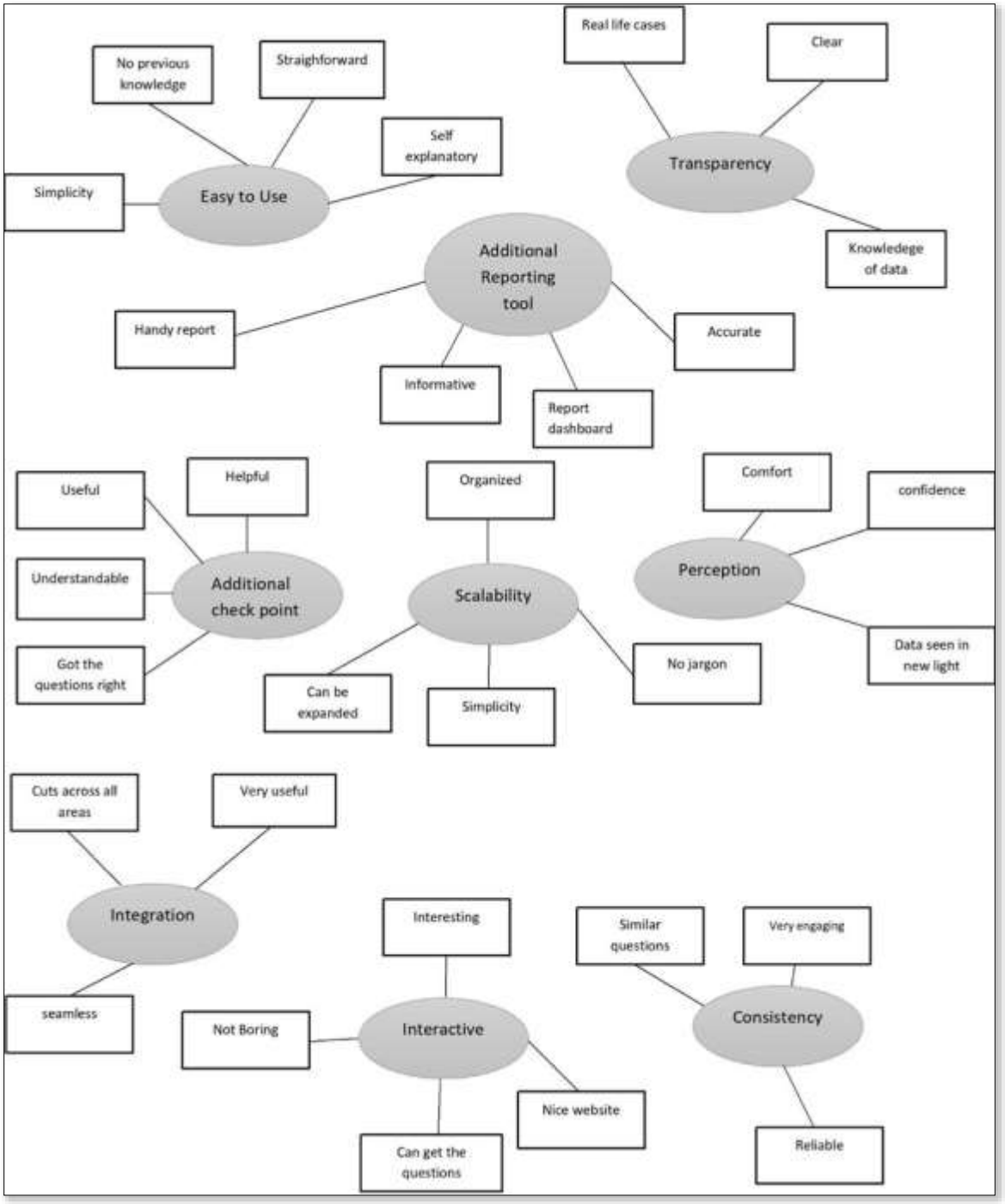


Figure 1: Thematic map of initial 9 central themes

The preliminary ideas illustrated in figure 1 were subsequently examined, with closely related and overlapping topics consolidated into singular themes to create the final thematic map presented in figure 2 below.

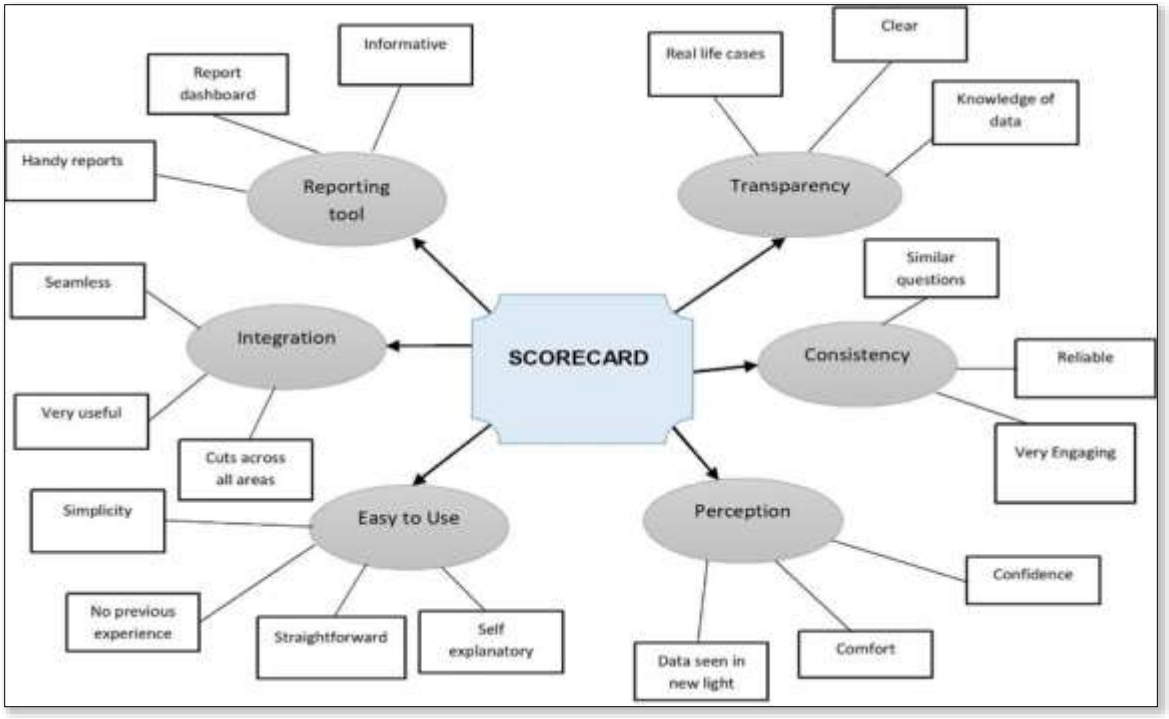


Figure 1: Shows the scorecard

The architecture of the Scorecard, which assesses data quality at many phases of data warehouse building and utilization. To facilitate the assessment of the scorecard, a case study was executed, allowing various stakeholders to utilize the scorecard inside their data warehouse development environment. Following the case study, semi-structured interviews were utilized to get the perspectives and perceptions of the participants. Twenty individuals engaged in the case study and semi-structured interview sessions. Each participant dedicated around one hour to the study. The completed scorecards from the case study demonstrated substantial added value, as participants regarded the data quality scorecard as a beneficial instrument.

Furthermore, the participants provided verbal comments on their perspectives on the scorecard following the case study. Their comments corroborated the findings of the finalized scorecards from the case study. The majority of participants recognized that the scorecard enhanced their confidence about the data in the data warehouse. The findings from the semi-structured interview indicate that the scorecard is both user-friendly and enhances their sense of data quality. The thematic analysis indicates that the scorecard reveals six principal themes: (a) supplementary reporting tool; (b) integration; (c) user-friendliness; (d) transparency; (e) consistency; and (f) perception [17-19].

**Conclusion**

The stakeholder conducted a review of the scorecard. Subsequent to the DQS run-through, participants were questioned to obtain direct insights into the DQS's performance. The interview was recorded with a dictation device and subsequently transcribed for additional study. The DQS was shown to enhance several of the evaluated data aspects. Several stakeholders observed that the timeliness factor is superfluous, as a timestamp is a typical feature in their data warehouse. The

findings indicate that the data quality scorecard fulfills the necessary criteria for a data quality measuring instrument. Based on assertions by the majority of participants endorsing this viewpoint. The subsequent chapter assesses the modified DQS within the Oil and Gas sector.

**References:**

[1] Power, M. (2004). The risk management of everything. The Journal of Risk Finance, 5(3), 58-65.
[2] Stulz, R. M. (2008). Rethinking risk management. In Corporate risk management (pp. 87-120). Columbia University Press.
[3] Stoneburner, G., Goguen, A., & Feringa, A. (2002). Risk management guide for information technology systems. Nist special publication, 800(30), 800-30.
[4] Pasham, S.D. (2017) AI-Driven Cloud Cost Optimization for Small and Medium Enterprises (SMEs). The Computertech. 1-24.
[5] Chen, D., & Zhao, H. (2012). Data security and privacy protection issues in cloud computing. International Conference on Computer Science and Electronics Engineering, 647-651.
[6] Garg, P., Verma, D., & Kaushal, V. (2018). A study on data migration techniques for cloud computing. International Journal of Advanced Research in Computer Science, 9(1), 45-52.
[7] Sai, K.M.V., M. Ramineni, M.V. Chowdary, and L. Deepthi. Data Hiding Scheme in Quad Channel Images using Square Block Algorithm. in 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI). 2018. IEEE.
[8] Pasham, S.D. (2018) Dynamic Resource Provisioning in Cloud Environments Using Predictive Analytics. The Computertech. 1-28.
[9] Ahmed, T., & Smith, M. (2018). Cloud data migration: Challenges, solutions, and future directions. Journal of Cloud Computing, 7, 12-29.
[10] Tallon, P. (2013). Corporate data migration strategies: Managing risks and maximizing benefits. MIS Quarterly, 37(4), 1125-1147.
[11] Grolinger, K., Higashino, W. A., Tiwari, A., & Capretz, M. A. M. (2013). Data management in cloud environments: NoSQL and NewSQL data stores. Journal of Cloud Computing: Advances, Systems and Applications, 2(1), 1-24.
[12] Inmon, W. H. (2005). Building the data warehouse (4th ed.). Wiley.
[13] Khine, P. P., & Wang, Z. (2018). Data lake: A new ideology in big data era. Proceedings of the 2018 IEEE 6th International Conference on Future Internet of Things and Cloud Workshops, 37-42.
[14] Kimball, R., & Ross, M. (2013). The data warehouse toolkit: The definitive guide to dimensional modeling (3rd ed.). Wiley.
[15] Dageville, B.,andDias, K. (2006). Oracle's Self-Tuning Architecture and Solutions. IEEE Data Eng. Bull., 29(3), 24-31
[16] Malhotra, I., Gopinath, S., Janga, K. C., Greenberg, S., Sharma, S. K., & Tarkovsky, R. (2014). Unpredictable nature of tolvaptan in treatment of hypervolemic hyponatremia: case review on role of vaptans. Case reports in endocrinology, 2014(1), 807054.
[17] Shakibaie-M, B. (2013). Comparison of the effectiveness of two different bone substitute materials for socket preservation after tooth extraction: a controlled clinical study. International Journal of Periodontics & Restorative Dentistry, 33(2).
[18] Gopinath, S., Janga, K. C., Greenberg, S., & Sharma, S. K. (2013). Tolvaptan in the treatment of acute hyponatremia associated with acute kidney injury. Case reports in nephrology, 2013(1), 801575.
[19] Shilpa, Lalitha, Prakash, A., & Rao, S. (2009). BFHI in a tertiary care hospital: Does being Baby friendly affect lactation success?. The Indian Journal of Pediatrics, 76, 655-657.