ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug061

A Proposal: From Failure to Success for Data Warehousing in Investment Banking Projects

Tejas Sriprasad[1]

[1]ESGCI: Doctoral Research Studies
UPGRAD
East Windsor, NJ

Publication Date: 2026/08/12

Abstract: The background of this paper will explore data governance and types of implementations of Data warehousing technologies in the Investment Banking Industry. Many implementations use the data warehousing technologies connected to the cloud in investment banking; our research will aim to set the frontiers of Data Governance and expand the horizons for Data Science Research. In different implementations and varying organization structures data warehousing, data science projects are in huge demand. There are many issues with data warehousing implementations, including high failure rates of data warehousing projects, trying to fix them in a way never done before, will be the secondary aim of this research paper. Accidental data redundancy, data quality issues, causing high failure rates, this research paper will look at possible solutions to alleviate this issue. There are different implementations of the warehouse which are highly successful, they include data lakes, data meshes, and data ponds. Data governance includes full data management and the way we manage both big data and analytics.

Keywords: *Data Warehousing, Dimension Modeling, Data Governance, Data Warehouse Architecture Planning, Requirement Gathering,, Azure, Aws, Cloud Data, Lakehouse, Data Science, Data Analytics, Cloud Computing, Data Science Success, Investment Banking.*

How to Cite: Tejas Sriprasad (2026) A Proposal: From Failure to Success for Data Warehousing in Investment Banking Projects. *International Journal of Innovative Science and Research Technology*, 11(8), 171-175. https://doi.org/10.38124/ijisrt/26aug061

I. INTRODUCTION

The idea of this research paper will be to solve the issues of failure in data warehousing projects in the banking sector set the frontier of human knowledge in data warehousing technologies. There are many reasons data warehouses fail, the idea for data warehouses originally is to store and further organize operational data. To understand data warehouses, a concept that started in the 1980's (Farhan et al., 2024) that there are core decision support systems that evolved from major corporations. We must first see what a data warehouse is, (Ponniah, 2011) and establish the fundamentals of what a data warehouse is, the benefits of the data warehouse are wide ranging and storage, organization and retrieval of data, control modern day life. Modern day data warehouses contain thousands and thousands of rows of banking data and need are cloud hosted, take less space and are easy to move around, (Asrani et al., 2017) states that data warehouses act as a decision support system and are the lifeblood of the organization.

In Investment banking, data warehouses contain highly critical information. Everything is data, both transactional and dimensional data. They are so important that as they contain important information such as balances, transaction history. The data must be loaded and accurate. The literature around data warehouse in modern landscape establish what a data warehouse is, how data is loaded into the data warehouse, how it is categorized.

II. ISSUES

The data lacks completeness, causing failure of the Datawarehouse, completeness of data specifications indicate whether the data warehouse is a sound implementation and follow-ups are done to make sure data accurate, if data is not refreshed or historical data are not updated then the data warehouse is said to lack completeness, therefore failing, our research will make sure that if were to match data requirements for completeness then will it succeed.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug061

➢ *Consistency:*

Consistency of Data means that data across the enterprise should be in sync with each other and should not provide conflicting information. For examples a credit card is cancelled, and inactive, but the card billing status shows 'due'. Such kind of inconsistent data across data sets should not prevail." Having consistency across the data is critical and could cause logical breaks in getting accurate and information. So not having consistent information across table would and banking data would be inconsistent. Therefore, causing failures in upstream systems as well as downstream systems, would be caused, no sense would be made of data. Inconsistent and repeated data that take extra space would be created, read and used.

➢ *Validity:*

Data Validity means the correctness and reasonableness of data. For example, a bank account number should fall within a specific range, numeric data should be all digits, dates should have a valid month, day and year and the spelling of names should be proper." Validity of information in banking data can be critical, solving for small inaccuracies can create huge pockets of misinformation, data of deposit and loans is kept in the Datawarehouse across transactions and accounts and if small number of errors do occur, it can cause major failure, so validity is really a focus on making sure Datawarehouse projects succeed.

➢ *Integrity:*

Data integrity verifies that data has remained unaltered in transit from creation to reception. Appropriate relationship linkages among records are very important else it might introduce unnecessary duplication throughout the systems." Data integrity is also crucial, meaning the latest data is present and it matches the user requirements. Accidental redundancy is caused by adding the same data again and finding

➢ *Conformity:*

This dimension verifies whether data is expected to adhere to certain standards and how well it's represented in an expected format. For example, date may be represented either as 'dd/mm/yyyy' format or as 'mm/dd/yyyy' format. So, conformance to specific data format is essential." Rules must be set around reference data, so we can know we are successful in uniformity, and the warehouse matches a certain conformity and has the right dimensional and factual modeling to match user requirements.

➢ *Accuracy:*

Checks for the accurate representation of the real-world values. For example, the bank balance in the customer's account is the real value that the customer deserves from the Bank. Any inaccuracy in the existing data can have a worse impact on the operational and analytical applications."

Data accuracy tells the whole story of whether a Datawarehouse has failed and eventually be scrapped if it not accurate and the design and dimensional modeling don't match user requirements of accuracy and reliability.

So, measuring for this common reason the data warehouse and data governance implementation fail, we can hypothesize the solution for the failures. There's also another dimension that is important to understand as background of the research:

Data warehouses feed different type of data including ETL processes that transform data from columns, tables, and databases turn into data analytics and reports that are consumed for information. There is a tremendous need for cloud services which include hosting the Datawarehouse in the cloud, so failure rates for projects could be more.

- Data analytics rely on accuracy of data and successful implementations of the Datawarehouse.
- Trends currently demand data warehousing to be accessible, especially in the cloud. There are also implementations of overall data governance that include metadata, static data, reference data and lookup data which are connected to overall ell implementation and data science.
- Exploring the full end to end data governance to successfully implement the Datawarehouse in the Cloud would be in Certain successful implementations of the Data warehouse are huge, they have a layout of Data Marts, backup data.
- Data warehouses in the banking industry contain a lot of data that is private, so protecting privacy and is of utmost importance so make sure the Datawarehouse implementation succeed.

The objectives of the research will be to set the frontiers of data research to another level. After analyzing current trends in data warehousing data governance as whole in banking, the objective will be to hypothesize the solutions of data warehouse failure the research questions will try to acquiesce the main reasons that data governance and Datawarehouse implementation fail and suggest possible solutions;

- The objectives will be tracing the original foundations of the data warehouse and further the frontiers of data warehousing research even further, essentially exploring Data Governance, implementations of the Datawarehouse in banking, the reasons that some data warehousing projects fail and alleviating them. The research objectives will aim to seek reasons of failures in data warehousing, data governance in banking and garner and elicit the right solutions to be a beacon for data research.
- The research objective will be to hypothesize set state of the art dimensional modeling techniques which reduce failure in data governance and data wa1rehousing. We will aim to test the hypothesis and come up with and confirm the solution works.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug061

III. RESEARCH QUESTIONS

The research questions will delve into what the possible causes of failure of Data warehousing and data governance in banking in current time are and how could we possibly resolve them. The research questions will be as follows:

- What are current failures of Data governance and data warehouse in banking and how can we resolve them?

- Many implementations exist of the Datawarehouse, due to issues, some data warehousing projects fail in banking. Some due to quality issues, some due to lack of meeting user expectations. Our research question will explore into what the main reason data warehousing in the cloud and on prem fail and how to resolve these issues. Issues range from erroneous data, to lack in inconsistency, validity. There is implementation of data lakes, data meshes, data ponds, S3 and Redshift, which is an Aws based Datawarehouse.

- What could be the possible solutions where we can fix failures in data warehousing in the banking industry?

✓ Tremendous amount of Datawarehouse projects have been failures, our question will be to decipher why and what causes the failure of Datawarehouse projects. Is it data quality, lack of clear requirements. There is also factual and dimensional modeling in Datawarehouse where data is highly organized by dimensional and facts, so therefore data marts can access specific records quickly. Our research questions will focus on which theory fits the solution for Datawarehouse failure in banking. Having a good dimensional modeling techniques could potentially be the solution avoiding the failures in data warehousing.

- How does one successfully implement data governance and data warehouse in Banking to properly to include cloud and data analytics?

Exploring how a data warehouse implementation is successfully done in the cloud especially with banking data and presenting solution that will alleviate issues will be one the main questions addressed in this research. With the advent of cloud, there are many ways to access Datawarehouse, store customer data and run intelligent reports. With banking data there's also a tremendous need for data privacy which is important for us to protect client's data. Also, in the range of importance is how we make sure make sure the projects succeed. So, the question will to be to see if we can implement state of art solutions

- What is Data Governance, how can you tell a successful implementation of Data Governance is done in Banking?

Data Governance is end-to-end management of data in an organization, it includes management of static data, referential data, lookup data, etl and overall banking data, the research question will focus on collecting data about data governance and failures of Data warehouses in banking and how we can solve them. Can Data Governance overall as a field be improved? Can the banking as a field be improved as well?

So, the research questions and objective center around successfully implementing the data warehouse and manage end to end data governance without failing. These questions will comprise of a thorough understanding why data warehouse fail, and we can learn to navigate them to success.

IV. RESEARCH DESIGN BEING CONSIDERED

The research design will essentially be a mixed mode type of research that will include primary convenient data collection and sampling as well as secondary data.

The qualitative data research will include semi-structured interviews of data professionals, data collected from existing databases like EBSCO, and Sci-hub. There will also be observation techniques employed of data professionals, who work with extracting transforming and loading banking data.

The quantitative data research will include multiple linear regression, hypothesis testing of multiple variables including size of data, number of successful implementations,

We propose a be collection of interviews that will be conducted with Data warehousing experts. We will also conduct structured and semi structured interviews, check the data for data quality, code the data and then perform analysis, including hypothesis testing. We are considering the use of convenience sampling.

Observational Studies will include collecting data from secondary sources: The research design will include existing observational studies of Data Governance. We will collect a range of secondary data and collect data from anthropological studies including Smithsonian, Harvard Dataverse which include the origins of the Datawarehouse.

We also are considering using linear regression and multiple linear regression analysis on successful implementations of the data warehouse versus the size of the Datawarehouse. Observing several successful implementations versus the size the subject matter, cleaning the data and then quantitatively analyzing the data.

The research design will include exploring secondary data on failures of Data Warehouses and how they can be successfully solved We will collect literature on data warehousing technologies, including expert opinion, conducted studies, including anthropological studies on origins of data, cleaning, coding the data the data and

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug061

V. DATA ABOUT DATA

(Bilal et al., 2016) states the "banking industry is categorized as one of the highest information demanding industries in the business world. With the advancement in information technology sector, the role of business intelligence (BI) increases with great number in the process of banking operations [54]. The increased business speed and growing competition has shown the need of banking intelligence dramatically. Bank intelligence is the ability to gather, manage, and analyze a large amount of data on bank customers, products, operations, services, suppliers, partners and all the transactions. As data increases, it becomes difficult to collect, handle and transform it into useful knowledge and DW solves this problem." In the banking industry tons of operational data is stored regards deposits, loans, transactions, investment banking where the banks acts as a custodian on deals are stored in the Data warehouse. (Bilal et al., 2016) is accurate in the fact, due to the nature of that data, there is a tremendous amount of Extract transform and load (ETL) that needs to happen between the transactions the loans and deposits.

Clients' data is stored and important to be archived and kept for a long time so storage and data retention for a long period of time become important, the need for storage in Azure and Aws also become important as that data is important and needs to be backed up, creating more and more need for the data warehouse. (Bilal et al., 2016) mentions the importance of the data warehouse in banking

(Bilal et al., 2016) describes the importance of the data warehouse in the banking, many customers have bank account for years and all that needs a master data warehouse, which is basically an accounts master data, holding all the account information. This means that the data warehouse needs be maintained in a modern pipeline that collects data from different sources such as loan data, deposits data, holdings, investments and combines them and store in a data warehouse. From that data warehouse more important reports, Additionally, more metadata is created, which is data about the data, managers use metadata to analyze trends, number of tables created read used and deleted and in highly important in the future for data analytics. The nature of banking is that modern banks like Citibank, Toronto Dominion Bank, use data warehouses to store financially protected data. As (Bilal et al., 2016) mentions a variety of customers utilize the data warehouse from the buy-side to the sell-side Investment banks store tons of data which have still untapped potential that is to made use of. The research design will test the hypothesis that successful implementations of the data warehouse include, good data quality, dimensional modeling, clear requirements and clear access to conduct ETL to and from the Datawarehouse to Data marts.

- Timeframe: The research design will include the timeframe of the 2 and half years to complete the research as subscribed by ESGCI.
- Budget: The budget will be self-funded and convenience method of sampling and interviewing will be employed.

VI. LIMITATIONS

Limitations of the research include, not be able to cover the full spectrum of the data warehouse data marts, data governance and cloud. Also not being able to judge failures of the data warehouse or be able to acquiesce how big the scope of solution might be might limit the research. Also included are lack of resources to interview survey and observe 100-150 data professionals so that might limit the perspective of the research. There are also qualitative and quantitative limitations that might skew the results of the research like lack of secondary data in data warehousing. On a grander scale, we might not have the accurate data causing errors in epistemology.

VII. CONCLUSION

In conclusion, the research will aim to set the frontier of intersections of data governance, data warehouse knowledge, using and solve issues of Datawarehouse failure in banking. Data feeds AI models is the core crux of These include garnering solutions in banking data, successful and state of the art data warehousing and data governance structure and procedures and routines. Being able to design research that will trace the anthropological roots of how data was founded will be the strengths of the research, and eventually we will be able to set a new level of accomplishment in human research of data.

ACKNOWLEDGMENT

A special acknowledgment to my loved ones and my supervisor Dr. Raveendran and family.

REFERENCES

[1]. Adams, Karen., & Brace, I. (2006). An introduction to market & social research: planning & using research tools & techniques. Kogan Page.

[2]. Bilal, M., Sheikh, U., Raza, B., & Javaid, Q. (2016). Application of Data Warehouse in Real Life: State-of-the-art Survey from User Preferences' Perspective.
[3]. International Journal of Advanced Computer Science and Applications, 7(4). https://doi.org/10.14569/ijacsa.2016.070455
[4]. Chandra, P., & Gupta, M. K. (2018). Comprehensive survey on data warehousing research. International Journal of Information Technology (Singapore), 10(2). https://doi.org/10.1007/s41870-017-0067-y
[5]. Dinesh, L., & Devi, K. G. (2024). An efficient hybrid optimization of ETL process in data warehouse of cloud architecture. Journal of Cloud Computing, 13(1). https://doi.org/10.1186/s13677-023-00571-y
[6]. Hwang, M. I., & Xu, H. (2005). A survey of data warehousing success issues. Business Intelligence Journal, 10(4).
[7]. (Citigroup)
[8]. Research Methods for Business Students. (n.d.). www.pearson.com/uk.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug061

[9]. Saddad, E., El-Bastawissy, A., Mokhtar, H. M. O., & Hazman, M. (2020). Lake data warehouse architecture for big data solutions. International Journal of Advanced Computer Science and Applications, 11(8). https://doi.org/10.14569/IJACSA.2020.0110854

[10]. Wani, A. A., & Raina, B. L. (2019). Issues and handy solutions addressed at every stage in real time data warehousing, I.E. ETL (Extraction, transformation & loading). International Journal of Engineering and Advanced Technology, 8(5 Special Issue 3). https://doi.org/10.35940/ijeat.E1100.0785S319

[11]. R. J. Santos, J. Bernardino and M. Vieira, "A survey on data security in data warehousing: Issues, challenges and opportunities," 2011 IEEE EUROCON - International Conference on Computer as a Tool, Lisbon, Portugal, 2011, pp. 1-4, doi: 10.1109/EUROCON.2011.5929314.

[12]. Ponniah, P. (2011). Data warehousing fundamentals for IT Professionals, Second edition. In Data Warehousing Fundamentals for IT Professionals, Second Edition. https://doi.org/10.1002/9780470604137

[13]. Satyanarayana Reddy, G., Srinivasu, R., Poorna, M., Rao, C., & Rikkula, S. R. (2010). Data Warehousing, Data Mining, Olap and Oltp Technologies Are Essential Elements to Support Decision-Making Process in Industries. International Journal on Computer Science and Engineering, 02(09)

[14]. Wixom, B. H., & Watson, H. J. (2001). An empirical investigation of the factors affecting data warehousing success. MIS Quarterly: Management Information Systems, 25(1). https://doi.org/10.2307/3250957

[15]. Hendayun M, Yulianto E, Rusdi JF, Setiawan A, Ilman B (2021) Extracttransform load process in banking reporting system. MethodsX 8:101260

[16]. Farhan, M. S., Youssef, A., & Abdelhamid, L. (2024). A Model for Enhancing Unstructured Big Data Warehouse Execution Time. Big Data and Cognitive Computing, 8(2). https://doi.org/10.3390/bdcc8020017

[17]. Georgiev, A., & Valkanov, V. (2024). CUSTOM DATA QUALITY MECHANISM IN DATA WAREHOUSE FACILITATED BY DATA INTEGRITY CHECKS Mathematics and Education in Mathematics, 53. https://doi.org/10.55630/mem.2024.53.067-075