International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 08 | Aug 2023 www.irjet.net p-ISSN: 2395-0072

Evaluation of Data Auditability, Traceability and Agility leveraging Data Vault Modeling in frequently changing Data ecosystem

Sayan Guha[1]

Sr. Data Architect, AI & Analytics Practice, Cognizant Technology Solutions, West Bengal, India

---

Abstract – *Maintaining the data audibility and traceability in the frequently changing agile business driven data ecosystem has been the pain area for every organization. Designing the storage architecture of a business which frequently needs to change its business strategy and rules, requires the agility and flexibility of Data Vault design & data modeling techniques which can accommodate frequent changes to the design without having to do much programming changes on the underlying processing logic. This paper discusses, the critical role of Data Vault modeling to address scenarios where frequent changes a business rule has minimal impact on the spend & An evaluation of the data auditability and traceability have also been studied with the help of business scenario comparing with other modeling techniques where Data Vault modeling it was not in practice.*

Key Words: Data Vault modelling, Auditability, Data traceability, third normal form data modelling, dimensional data modelling, hash key, hash key difference, data warehouse, Raw Data Vault, Business Data Vault, SQL,

1.INTRODUCTION

In the current time, as we have observed the continuous changes to business strategy for most organizations remains the pain area because implementation of such frequent changes to the business data sources, business rules and relationships between the business entities requires a considerable spend on the IT infrastructure, its design and storage architecture modelling techniques and to ensure IT design & execution doesn't become a bottleneck for seamless execution & adoption of Data Vault modelling in this context not only provides the required agility to the business but also empowers the organization of data governance through complete data auditability and traceability. In this paper, we will discuss a specific case study where agile business process of a subscription-based ecommerce business adopted data vault modelling technique to their benefit of agility We would try to evaluate how the changes in the business rules have minimal impact on the design architecture of the data model in the context using Data Vault modelling & how the data auditability and traceability .be maintained.

1.1 Motivation

My motivation here to write this paper is to help the readers understand about the flexibility, robustness and agility Data Vault modeling technique has to offer. As opposed by tradition third normal form or dimensional data modeling techniques where the data design practitioners can address the frequent changes & history data requirements respectively, Data Vault modeling provides the best of both schools of thoughts and hence if it is chosen it comes up the flexibility of adapting to the changes of business rules which is often required to have a competitive edge in today's business. All companies particularly often need to change the relationship between the business entities, get rid of them or add few more and the changes in relationships resulting many to many relationships established in the process between these entities. A Data Vault model duly designed not only addresses the same but also provides a benefit in complete data auditability and traceability in the process by leveraging a Data Vault modeling.

1.2 Aim of this paper

In this paper, I have aimed to discuss the specifics of business model of an Ecommerce Retailer based on a subscription-based business model as business scenario and sequential break-up of the business rules in terms of relationship between the customer & subscription. In this process, I will evaluate the Data Vault model suited for the purpose which will not undergo any change despite the changes in the business approach. In the end of the paper, I will share the evaluation metrics on how this approach have benefited on the data metrics e.g., auditability, traceability in scenarios when the data vault model has undergone changes.

2. LITERATURE REVIEW

A literature review was undertaken encompassing few academic and industry papers. This section reviews Data Vault as undertaken by organizations & addresses the key aspect of its agility and robustness.

The evolution of data modeling methodologies in organizations along with the business scenario is listed in the below table.

Table -1: Data Modeling techniques

Data modeling method	Construct	Scenario
3NF (third normal form)	Normalized & data stored in detailed level transactions in tables.	Operational systems.
Dimensional model	Denormalized & data stored in dimension & fact tables, can be aggregated	Business Intelligence, used for dashboards used by strategic business decision support.
Data Vault	3 major constructs are Hub (collection of business keys), Link (relationship between hubs) or potentially with other relationships (links) & SAT (Satellite stables with description data) – changes to the data reflects on SAT tables, other remain unchanged, in case of relationship change, reflects only on LINKS	Hybrid approach encompassing the best of breed between 3rd normal form (3NF) and star schema. Flexible to changes and less implementation cost

In [1], the author presented a case study where schema changes to the data source as an immediate need to the business scenario can be handled with complete traceability and auditability using Data Vault modeling technique. The ability to adapt to a changing business requirement, addition & removal of data sources without any change to the design and with ability to handle large scale database with history data were identified as the advantages to the Data Vault design in the author's research.

In [2], the authors have discussed the major challenges that are encountered by enterprises using traditional datawarehouse architectures encompassing complex updates, difficulties in source data reconciliation and integration, slow loading, brittle transformations (due to integration close to volatile sources and even more so among reporting viewpoints), and lack of an integrated system of record and how Data Vault conceptual data model is the potential answer to all the above pitfalls.

While understanding the work that has already being done in this field covering the areas of speed of loading, quick updates, addressing the scenarios of quick change of system of record, I have taken one step ahead to analyze the impact of the Data Vault design on a real-life business scenario with an Online Retailer who operates in a subscription-based business model in sections to follow.

Before considering Data Vault modeling, I have also studied the other streams of agile data modeling to check their suitability for purpose and business scenario. For example, Anchor modeling by which domain driven needs are prioritized over data driven needs as in Data Vault. While both forms belong to Ensemble modeling, Data Vault being data driven, audibility needs take precedence. In [3], the authors have narrowed down the specific guidelines for Anchor Modeling suited for the needs of domain driven needs. Inclusion of concepts covering business defined schemas, models suited to adapt to changes are similar between the 2 types of modeling techniques , however abilities of traceability & auditability are found to made for purpose in DV modeling .Every object in the Data Vault (Hub, Sat or Link entities refer Table-1) contains 2 additional attributes called **Record Source** and **Load_Datetime** which provides a detailed auditability and traceability back to record source along with history of changes. This is a recommendation from DV methodology [4] and I have followed the same in resolving the business scenario in my model & design.

In the context, where Data auditability definition, as I wanted to define can be phrased as the ability to provide a audit trail associated with a data transfer and important information, such as who sent the data, when they were sent, when they were received, what data structure (e.g., xls, csv, txt, xml) was used, how the data were sent (i.e., via what medium) and who received the data. I have referred to [5] as a definition particularly suited at the scenario. Henceforth while defining auditability in our evaluation section, I would try to break up the model into the above defined factors.

In connection to Data Traceability leveraging Data Vault, I will define and share a matrix which will contain the different hop points for the data flow and will have checkpointing enabled at every level. In real life data systems, this traceability was enabled by data lineage using standard Data governance tool and using Data Vault based architecture it ensured 100 percent traceability from the target database through to the source systems. In [6], the authors have defined and designed a similar data governance model based on data traceability and can get data feedback and revision through this model. The proposed method considers the different ownership of data and may form a closed-loop data service chain including effective data validation.

Therefore ,in the background of all of the above related work and having studied their research literature , I hereby promise that in subsequent section the reader will have detailed view of the business scenario of online retailer which we discussed initially in this section as our requirement for agile frequently changing data coming from 2 different sources (both sources providing data for online subscription) & the requirement will undergo a change which will result in the minimal change in DV model (solution will contain data vault layers in the data architecture). The data model design

diagrams also follow along with the framework which shows how data auditability and traceability is maintained in this ecosystem.

3. BACKGROUND AND BUSINESS SCENARIO

We have the business scenario of an online retailer that pertain to the customer to subscription relationships in the subscription-based model the company operates with. The below Business logic flow diagram(s) shows the business flow before and after a business rule changes.

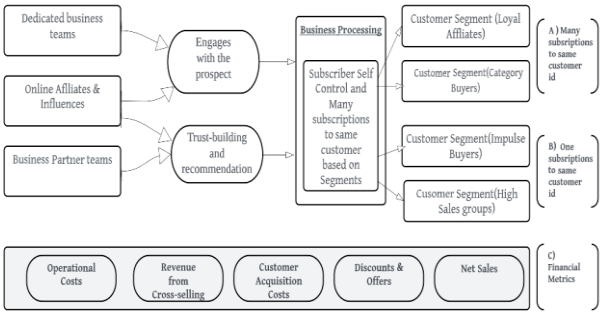

Fig. 1: Business logic flow diagram (before change)

Subsequently, the sales and marketing team decided to change one subscription model viz. One subscription to same customer applicable for Customer segments namely **Impulse buyers and High Sales groups** to many subscription & also same subscriptions can also be utilized by other customers belonging to the same family as base customer which essentially means, same customer id can be used by a family member of the customer to login to the Ecommerce website and order items and at the same time many subscriptions can be linked with same customer. As this process engages more with the portal , I have observed in the process of this project, based on the flexibility of using the same/different customer id(s) belonging to the base customer and his/her one /many family members the total time spent on the Ecommerce portal have increased, hence customer/customer groups can be eventually targeted by campaigns and promotions related to their buying habits as a family /group resulting in net sales increase in the scenario.

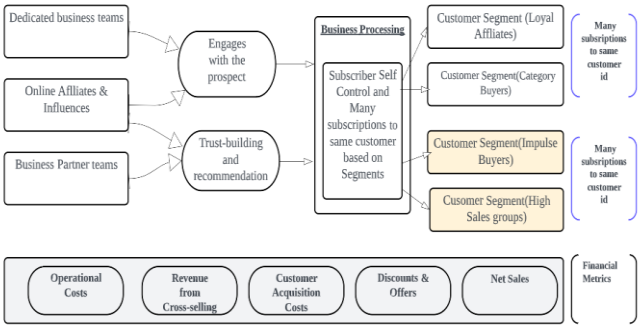

Fig. 2: Business logic flow diagram (after change)

In the context of this paper, however we are restricting our discussion on the agility of Data Vault modeling to address this frequently changing data and its impact on the data auditability and traceability. In Fig-2, the highlighted boxes, are the customer segments which have undergone the aforesaid changes in the business rule(s).

4. SOLUTION DESIGN AND DATA VAULT MODEL

The key design considerations for the Data Vault model are as follows:

- Resolve the conflict of relationships for frequent changing business rule(s) e.g., subscription model changed from one to one to one to many for 2 customer segments (Impulse buyers & High Sales Group)
- Change of any relationships because of frequent changing business rule(s) will have zero/minimal impact to the data model compared to conventional data modeling techniques.
- Change in any record source will have zero impact on the data model and data processing logic for new development.
- Notable improved price performance of storage, data traceability and auditability.

4.1 Architecture Solution

The standard Data Vault architecture solution [7] has been referred for the business use case in place, below is depicted which has been derived from standard Enterprise Data Warehouse which consists of the Data Vault model.

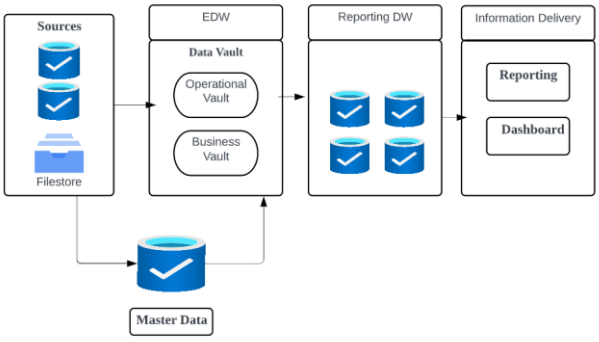

Fig. 3: Solution Architecture (DV approach)

The above architecture contains the below building blocks:

- **Data Sources:** The point of origin from where all data is brought to EDW (Enterprise Data Warehouse).
- **DV layer of EDW:** The DV model is composed of 2 data models. first is more source aligned called the **Operational Vault** & the second layer depicts the

system of record and oriented aligned towards the reporting side called the Business Vault.

- **Reporting DW**: Information Marts are constructed in this layer.
- **Information Delivery**: Users perform different data analysis tasks; this layer is an abstraction from IT for end users.

For evaluation of data traceability and auditability of the Data Vault model, the Raw Data Vault model (RDV) model is considered as below in Fig.4, the reason being the RDV, or Raw Data Vault contains Source system information & context without going much transformation changes.

For evaluation of agility of the data model subject to frequent changes in the business rule (in this context the subscription rule change for the online retailer), the Business Data Vault model (BDV) is considered in Fig.5, the reason being the BDV, or Business Data Vault is designed from RDV and added Master data and business rules in the context.

4.2 Raw Data Vault

In Fig.4, the Raw Data Vault model for the business scenario of the online retailer is shared. The business scenario of multiple subscription associated with the Customer Segments in the data model. The navigation will happen from Hub Customer to Hub Subscription using the Link Customer Subscription, designed in such a way to address many-to-many relationship, thus any change between any of the major entities will not have any change impact on underlying processing logic even with rapidly changing frequent business logic changes.

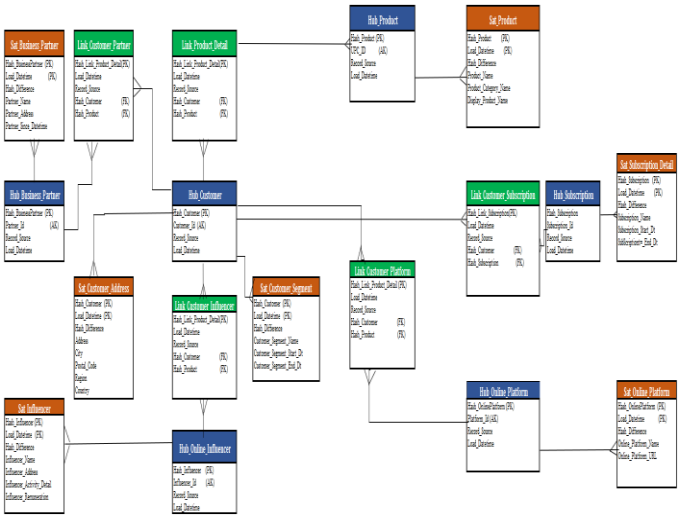

Fig. 4: Raw Data Vault model (RDV)

DV Entity	Color Code	Entity Name → Logical Name
Hub	Blue	**Hub Customer** -Customer **Hub Product**- Product **Hub Subscription**-Online Subscription **Hub Online Influencer**- Online Influencer associated with Customer. **Hub Online Platform**- Online Platform /Social Media handle **Hub Business Partner-** Partners associated with marketing and advertisements.
Satellite (Sat)	Brown	Sat Customer Address & Sat Customer Segment- Associated with the Customer Segments Sat Subscription Detail- Associated with Hub Subscription **Sat Product**- Associated with Hub Product. **Sat Online Platform**- Associated with Hub Online Platform **Sat Influencer** – Associated with Hub Influencer **Sat Business Partner-** Associated with Hub Business Partner.
Link	Green	**Link Customer Subscription** – Resolves the many-to-many rule change between Customer and Subscription Link Customer Platform. Link Customer Influencer and Link Customer Partner – kept as futureproof design for any rule change association with Customer. Link Product Detail – Captures Product line-item level

Table -2: Raw Data Vault Entities

4.3 Business Data Vault

In Fig.4, the snippet of the Business Data Vault model for the business scenario of the online retailer is shared. The business scenario of changes that are associated Customer Segments in the data model. The address of Customer is through a slowly changing attribute but as per the business scenario the customer is segment changes rapidly. Business Data Vault will ensure the agility to have minimum impact on the change and thereby attributing agility to the data vault model by separating the attributes out by rate of change, we help reduce the overall disk storage required for the Hub.

The problem statement requires us to understand how we get the data out of multiple Satellite tables when the Satellites are loaded independently and with **different rate of change.** The solution to the problem is Point in Time (PIT) tables in Business Data Vault.

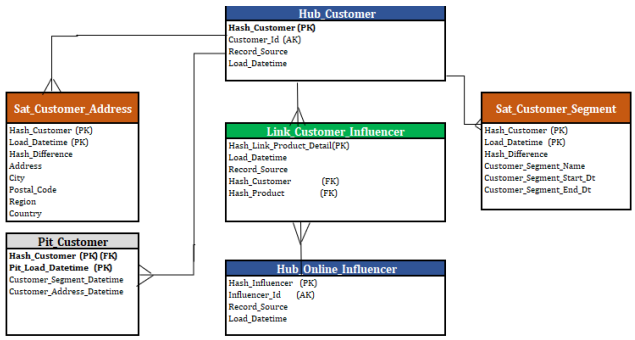

Fig. 5: Business Data Vault model (RDV) ~PIT Tables

The introduction of PIT tables (Point in Time) tables shown in the Fig. 5 on the portion of the Raw Data Vault Model where we have initially created more than one SAT tables viz Hub Customer and associated Sat_Customer_Segment and Sat_Customer_Address.

In [8], the author has researched the existing approaches of Bill Inmon and Ralph Kimball which lead to difficulties related to traceability of the of the entire data flow making it laborious modeling exercise. The author also states the complex ETL flow that are involved in preparation and refinement steps results in lack of agility in changing requirements-based data ecosystem like data warehouse

Hence the author is of an opinion of the using the power of design Data Vault 2.0 brings can substantially increase the speed & agility in such data flow and thereby resulting an agile data ecosystem more realistic.

For the example we have considered the retrieval SQL to be written using the PIT table **adds agility** to the overall Business Data Vault & the same being user interfacing layer, I have observed the complexity of SQL queries getting considerably lower & thereby supporting our initial approach of having increased agility using Business Data Vault. I have attached and highlighted the complexity of Joins in Table.3.

	Without Business Data Vault	**With Business Data Vault (PIT)**
Data Retrieval SQL	SELECT C.CUSTOMER_ID, CA. ADDRESS, CS. CUSTOMER_SEGMENT_NAME FROM HUB_CUSTOMER C JOIN SAT_CUSTOMER_ADDR CA ON C. HASH_CUSTOMER =CA.HASH_CUSTOMER WHERE CA. LOAD_DATETIME = (SELECT (MAX (CA2.LOAD_DATETIME) FROM SAT_CUSTOMER_ADDR CA2 WHERE C. HASH_CUSTOMER =CA2.HASH_CUSTOMER JOIN SAT_CUSTOMER_SEGMENT CS ON C. HASH_CUSTOMER =CS.HASH_CUSTOMER WHERE CS. LOAD_DATETIME = (SELECT (MAX (CS2.LOAD_DATETIME) FROM SAT_CUSTOMER_SEGMENT CS2 WHERE C.HASH_CUSTOMER=CS2. HASH_CUSTOMER	SELECT C.CUSTOMER_ID, CA. ADDRESS, CS. CUSTOMER_SEGMENT_NAME FROM HUB_CUSTOMER C JOIN SAT_CUSTOMER_ADDR CA ON C. HASH_CUSTOMER =CA.HASH_CUSTOMER JOIN SAT_CUSTOMER_SEGMENT CS ON C. HASH_CUSTOMER =CS.HASH_CUSTOMER JOIN PIT_CUSTOMER P ON P. HASH_CUSTOMER =C. HASH_CUSTOMER AND P.PIT_LOAD_DATETIME =’<PASS CURRENT DATE>’ AND P. CUSTOMER_SEGMENT_DATETIME= CS. CUSTOMER_SEGMENT_DATETIME AND P. CUSTOMER_ADDRESS_DATETIME= CA.P. CUSTOMER_ADDRESS_DATETIME
Complexity of Joins	High	Medium
Agility of design	Low	High

Table -3: Agility of SQL Joins in Business DV

Key takeaways from the data model designs and core analysis of the retrieval SQL I have framed above are

- **Raw Data Vault:** In the scenario I have specified, the simple usage of Raw Data Vault modeling standards addressed the entities by Hub-Link design considering many to many with a complete traceability & auditability of source data.
- **Business Data Vault:** In our scenario with the addition of PIT tables, point in time information is obtained adding agility benefits to the data

5. EVALUATION EXPERIMENT

5.1 Traceability in Data Vault as function of time

In [9], the authors in their article on requirements engineering have studied different frameworks on traceability & discussed the suitability of explicit traceability strategies for different companies and different projects. In their framework called TraciMo, they have defined the significance increased the correctness of identifying change sets for a given requirement, from the developer’s point of view. Hence drawing the ground definition from this existing literature, I find the motivation to define

Traceability (T)

= ***f*** (Connected checkpoints in Development) (Dα) X E*t* (Elapsed time in each checkpoint)

T = Dα X E*t*. **(i)**

5.2 Auditability in Data Vault as function of time

The guidance that is obtained from the latest from the international Data Vault & Ensemble modeling Enthusiast (DVEE) consortium, Data Vault Modeling should include the classic features of auditability at the granularity of every record mapped to every record source and record loading timestamp as the basic modeling standard in Raw Data Vault modeling. If we look back in Fig.4 and for any change in the Sat_Customer_Segment, the Record_Source in the transactional side is captured along with the Record_Datetime. This is an example in the current data ecosystem for most enterprise, where a customer can buy a product from multiple channels and the based on the same the segment of customer may change over time based on this Segment (business segment the customer belongs to), but Data Vault ensures that the whole auditability for the history trail of the entire transaction history is captured and auditable.

Auditability (**A**$_d$)

= ***f*** (Connected Record Attributes) (D_R) X T*s* (time measured at the lowest granularity)

A_d = D_R X T*s* **(ii)**

5.3 Agility in Data Vault as function of time

The ability to adapt to changes quickly and deliver value with point in time data to the end users with sustainable output produced continuously and repetitively makes Data Vault particularly suitable for Agile environments. In Fig.5 I have shared the PIT (point in time placeholder) which contains the point in time data in the user interfacing Business Data Vault. Therefore, adding and augmenting to the traceability and Auditability created in the Raw Data Vault layer with Agility.

In [10], the authors have made elaborate research, on definitions of agility many definitions of agility as there are agile practitioners and per the observations of the authors case, the understanding and definition of agility remarkably vary. In the conclusion section, the authors acknowledged that the concept of agility is complex and multidimensional (i.e. not simply about responsiveness to changes) It conceals many facets, the definitions of it vary considerably and as part of their research, various definitions of agility were gathered through a literature review.

To define my work in terms of agility I felt motivated and having studies the literature in [11], the researchers have devised an Agility Framework and illustrates the multi-faceted nature of Agility in terms of responsiveness, productivity, new innovations etc.

Hence, I have defined Agility in our context of Data Vault

Agility (**A**$_G$)

=*f* (Response time to connect all checkpoints in Development) (Dα) X Connected Record Attributes) (D_R)

A_G = d (Dα) (D_R) / dT_R **(iii)**

T_R**=** Response time in lowest granularity.

For the sake of our experiment and combining the understanding obtained for Traceability, Auditability and Agility from the above definitions the following approach is defined:

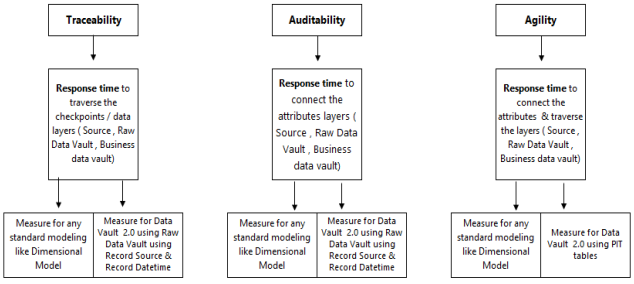

Fig. 6: Evaluation Approach

- **Validation with Dimensional Model:** Industry established Dimensional Model is taken as our standard to validate the response time for one use case scenario of Customer Segment change across the parameters of **Traceability, Auditability and Agility**. In the next section I have framed the response times for each of the above along with the design changes time being considered in the process.

- **Validate the Data Vault Model:** For the same changes, for a Customer Segment change as modeled in Fig 4 as part of Raw Data Vault and Fig 5 as par of Business Data Vault (PIT tables) we have captured the response times from the Data Vault Model across the parameters of **Traceability. Auditability and Agility** as defined in above section(s) where I have drawn the definitions.

5.4 Evaluation Results

- In our context of making 1 change in Customer Segment dimensional table Vs the same in Data Vault 2.0 for the change being discussed as our use case in Fig.4 -below results are obtained as the time taken

(time to design the change + time for getting response) as illustrated in Table 4 below

Serial #	Activity Description	Time in Dimensional Model	Time in Data Vault
1	Identify the business process	1800 seconds	1800 seconds
2	Identify the Source of the Change	900 seconds	10 seconds
3	Declare the grain of the customer Segment change	120 seconds	0
4	Identify the impacted tables with the change	900 seconds	500 Seconds
5	Add / redesign the dimensional model	300 seconds	0
6	Reload the data / refresh the data for the change	600 seconds	300 seconds
7	Time to traverse back to Source through all data layers	900 seconds	10 seconds
8	Provide a Point in time data (Dimensional Model requires querying at the time range granularity)	Not Possible	10 seconds

Table -4: Response Time Comparison Dimensional Model Vs Data Vault

- From the above evaluation results specifically across the parameters of Traceability, Auditability and Agility, as I observe, serial #2 and serial #7 for Traceability & Auditability and Serial #8 on Agility.

- As I observed, the way I have defined before in section 5.3 and considering various research study, the Dimensional Model is not capable to provide that Agility and hence Agility in terms of providing Point in Time data is not possible and hence it can be observed that Data Vault 2.0 is the one of the best suited modeling techniques to have agility of data in a changing data ecosystem. In the below graphs I have plotted Traceability & Auditability response times compared between Dimensional Model & Data Vault Model specific to our business scenario in Fig.7 and speed of response time plotted for Data Vault Model.

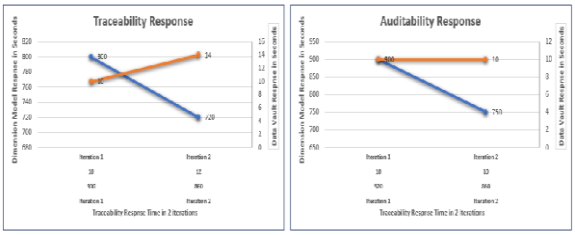

Fig. 7: Traceability & Auditability Response Time (Dimensional Model Vs Data Vault)

- Agility as defined and as per my experiment observation cannot be successfully satisfied by Dimensional Model design hence, I have not made any comparative visualizations between Dimensional Model and Data Vault. Agility as I derived from the existing literature has been defined in the context of this paper as responsiveness or response time to provide point in time data at any instance – the Data Vault model defines PIT tables in the Business Data Vault (reference to section 4.3 in this paper) defines agility within a timeframe of 10 seconds when queried to get the response. Such provisions are not part of Dimensional Model design.

The inability to provide agility and the slow response to traceability and auditability for the business scenario we have chosen of an ecommerce customer whose Customer Segment can change frequently owing to his changed buying habits established a business scenario of frequently changing data ecosystems , and I observe through the research already accomplished in this area of Data Vault Modeling and through my evaluation, Data Vault Modeling is found to be suitable and meeting the needs in this business scenario.

6. CONCLUSIONS

I expect that my work to augment to the existing research work in the field of exploration of features of frequent changing data ecosystem as achieved by Data Vault and traceability, auditability and agility in this respect is noteworthy.

With real business scenario obtained from a retail ecommerce organization and their business strategy to frequently changing the customer segment based on buying behaviors I have observed using Dimensional modeling school of thought not satisfying the needs of quick results expectations the organization expects to meet the downstream business & reporting demand. I acknowledge Dimensional modeling remains the de-facto standard for analytics and reporting with less frequent changing data, on the other hand Data Vault becomes a choice for scenarios like mine with a complete assessment and evaluation being done to align to the agile standards.

I have captured such data to predict from where we can seamlessly understand and acknowledge the usage of Data Vault requires minimal design changes and minimal response time owing to its unique design standards.

I, trust this work will motivate upcoming avenues of future research where data is stochastic & frequently changing, data driven approach to maximize needs of agility, traceability and auditability is necessary & an alternate thought process is expected to be applied from de-facto standards of dimensional modeling.

REFERENCES

[1] Zaineb Naamane and Vladan Jovanovic, "A Meta Data Vault Approach for Evolutionary Integration of Big Data Sets: Case Study Using the NCBI Database for Genetic Variation", pp-94-95

[2] Vladan Jovanovic and Ivan Bojicic, "Conceptual Data Vault Model", pp 131-132

[3] Lars Rönnbäck, Olle Regardt, Maria Bergholtz, Paul Johannesson,"Anchor modeling — Agile information modeling in evolving data environments", pp 8-10

[4] D. Linstedt and M. Olschimke, "Building a scalable data warehouse with Data Vault 2.0, 2016"

[5] Tommie W. Singleton, "Testing Controls Associated with Data Transfers", ISACA Journals.

[6] Guobao Zhang, data traceability method to improve data quality in a big data environment", 2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC), pp 2-5

[7] D.Linstedt, Supercharge Your Data Warehouse: Invaluable Data Modeling Rules to Implement your Data Vault. Create Space Independent Publishing Platform, USA, 2011

[8] Peter Gluchowski, "Data Vault as a Modeling Concept for the Data Warehouse" ,2022, pp-2-3

[9] Jan-Philipp Steghöfer, Paolo Bozzelli & Henry Muccini, TracIMo: a traceability introduction methodology and its evaluation in an Agile development team, August 2022, pp-59-60

[10] Necmettin Ozkan, Sahin Gok, Definition Synthesis of Agility in Software Development: Comprehensive Review of Theory to Practice, I.J. Modern Education and Computer Science, 2022, pp 26-44

[11] Petri Kettunen, Maarit Laanti, Combining agile software projects and large-scale organizational agility, Software Process Improvement and Practice, 2008, vol. 13, issue 3, pp. 183-193

BIOGRAPHY

Sayan Guha completed his Bachelors in Electronics & Communication Engineering in 2006. He has been serving Information Technology industry supporting Data & Analytics Space, Data Modelling & Architecture across business domains of Retail, Banking, and Insurance & Telecom. He is currently working with Cognizant Technology Solutions and his area of interests are in the fields of Artificial Intelligence, Cloud Solution architecture, Big Data Integration, Data Modeling, and Information Architecture.