DESIGNING A FRAMEWORK TO STANDARDIZE DATA WAREHOUSE DEVELOPMENT PROCESS FOR EFFECTIVE DATA WAREHOUSING PRACTICES

Deepak Asrani[1] and Renu Jain[2]

[1]Department of Computer Science Engineering, Teerthanker Mahaveer University Moradabad, U. P., India
[2]Department of Computer Science Engineering, University Institute of Engineering & Technology, Kanpur, U.P., India

ABSTRACT

Data warehousing solutions work as information base for large organizations to support their decision making tasks. With the proven need of such solutions in current times, it is crucial to effectively design, implement and utilize these solutions. Data warehouse (DW) implementation has been a challenge for the organizations and the success rate of its implementation has been very low. To address these problems, we have proposed a framework for developing effective data warehousing solutions. The framework is primarily based on procedural aspect of data warehouse development and aims to standardize its process. We first identified its components and then worked on them in depth to come up with the framework for effective implementation of data warehousing projects. To verify effectiveness of the designed framework, we worked on National Rural Health Mission (NRHM) project of Indian government and designed data warehousing solution using the proposed framework.

KEYWORDS

Data warehousing, Framework Design, Dimensional modelling, Decision making, Materialized View

1. INTRODUCTION

Data warehousing solutions have always been an information asset for organizations. They facilitate taking correct and timely decisions. With the exponential growth of data volumes and ever increasing competition among business houses, it has become a challenge to maintain a quality information base that could effectively aid decision making of the organization. To cater with such issues, effective implementation of data warehouse and choosing suitable Business Intelligence (BI) tools is must. Moreover, large numbers of data warehousing projects fail. It is clear from the below cited references. [7] claim that, a significant percentage of data warehouses fail to meet out their business objectives. The authors argue that requirement analysis is typically overlooked in real world Data Warehousing projects. Maintenance is cited as one of the leading causes of data warehouse failures. Warehouses fail because they do not meet the needs of the business or are too difficult to change with the evolving needs of businesses [25]. Recent statistics indicate that 50% to 80% of CRM initiatives fail due to inappropriate or incomplete CRM processes, poor selection and design of supporting technologies (e.g. data warehouses)[3]. If your data warehouse is not driven by a significant and legitimate business need is not worth the investment. The "build it and they will come" approach is technology looking for a solution, and

DOI: 10.5121/ijdms.2016.8402

is responsible for the 70 % of failed data warehouse efforts. [5] 40% to 50% of data warehouse initiatives end in costly failure. The search for root causes conversed on not understanding the user's business problems [11]. Despite the booming data warehousing market, a large number of costly data warehouse initiatives are ending in failure [24]. Connor [4] estimated failure rates at 40%.

Such findings of many research groups and surveys motivated us to look into reasons behind these failures and design a framework for developing data warehousing solutions that could meet out organization's decision making needs. Chances of successful implementation are higher when data warehousing project is planned, committed to, and managed as a business investment, rather than a technology initiative [23].

In our study, we found following major reasons responsible for high failure rates of data warehousing project implementations.

1. Very big size and complexity of projects
2. Lack of established standards for data warehouse development process
3. Lack of user interest towards implementation of data warehousing solution
4. Need of different database management techniques with which most of the developers community is not much familiar. e.g. type of indexing required, nature and frequency of data updating, schema de-normalization, materialized view updates, and data population through ETL
5. Ineffective identification of key business processes and business user's reporting needs
6. Non compulsion on usage of such solutions in the organizations

Although, many research articles have expressed concern about high failure rates of data warehousing projects but many others have presented work on most advance concepts like real time issues in data warehousing, web warehousing, and need of flexible data visualization tools & techniques. Availability of solutions for these issues reflects that, the technology of data warehousing has substantially matured. It is capable of delivering solutions that would never fail but would facilitate business users with much enhanced reporting facilities. Thus, we only need to utilize the available techniques effectively to develop useful data warehousing solutions. Our paper is aligned with this objective.

There is no standard method or model that allows to model all aspects of a data warehouse [12]. [18] Argue that most existing modelling approaches do not provide designers with an integrated and standard method for designing the whole DW. Data warehousing is not fully different from standard software development process. It also requires thorough investigation of client's needs, design of its solution followed by implementation. We can ensure quality of design & implementation, provided basic steps of standard software development life cycle are followed effectively. We have followed such approach for designing the framework for data warehousing. We have followed standards of software development process to standardize the process of data warehousing with some extra customization required for data warehousing projects.

Data warehouse development life cycle consists of following macro phases:

1. Analysis of operational system
2. Requirement analysis
3. Data warehouse designing

4. Implementation and Testing
5. Deployment
6. Post implementation maintenance

1.1 Extraction Transformation and Loading (ETL):

ETL is an important activity of data warehousing which is responsible for populating data warehouse with the required data from the operational system during implementation and testing for the first time after performing data cleansing process and then it is again executed as per decided schedule based on data updating frequency requirements of the organization. Data cleansing deals with detecting and removing errors and inconsistencies from data in order to improve its quality, and is typically required before loading the transformed data into the data warehouse [8]. The Extract-Transform-Load (ETL) processes efficiently update the data warehouse with batches of new records. Since ETL builds de-normalized tables and transforms existing tables for integration, it is considered a part of data warehouse modelling. ETL is time-consuming and difficult in data warehousing, but easier and faster in MapReduce/DFS [20]. ETL process design is generally given very less time and there is no standard model for ETL process because of the fact that operational systems are different, user expectations are different, and accordingly data warehouse modelling techniques to be followed are also different. We need to work in the direction of ETL process standardization. We need to propose a model for ETL standardization based on adaptation techniques where ETL process can be dynamic and can change the ETL parameters based on usage pattern and changing user requirements on a continuous basis.

Organization of this paper is as follows: Section 2 covers literature review broadly categorized in four different aspects of data warehousing. Section 3 discusses proposed methodology for data warehousing and gives steps of the proposed framework for effective data warehousing. Section 4 is of data warehouse design for NRHM project. Section 5 concludes the paper.

2. LITERATURE REVIEW:

Data warehousing has witnessed huge research efforts in multiple areas, be it the design of data warehouses, or its implementation, or the maintenance. Literature mainly points to two major issues. These are high failure rates of data warehousing projects and secondly the lack of standardization of data warehousing practices. Few researchers have developed frameworks but have included only limited aspects of data warehousing while some other researchers have worked on data warehousing practices. None of them carries complete practical approach of developing data warehousing solution. Area wise references of some research undertakings are as follows.

2.1 Data Warehousing Practices:

“The process of developing data warehouse starts with identifying and gathering requirements, designing the dimensional model, followed by testing and maintenance. The first phase is analysis of operational systems whose aim is to collect the information concerning the pre existing operational system. Conceptual modelling is the necessary foundation for building a database [12].” Juan Trujilio proposed the use of UML for design of data warehouse. He defined four different profiles for modelling different aspects of data warehousing namely UML profile,

Data mapping profile, ETL profile, and Database deployment profile [12]. Stefano Rizzi gave a semi automated methodology to build a data warehouse from the pre existing conceptual or logical schemas [12]. [26] has put up report on efforts of various researchers on querying data warehouses or OLAP databases, data warehouse modelling, data warehouse design, and query processing and view maintenance. "[21] proposed the *R-cube*, a type of OLAP cube based on (i) specifying the relevance of each fact in a query, and (ii) defining the related documents that provide information on the selected facts. In this way, users can query a traditional Data Warehouse (DW) (with the corresponding MD terms) and obtain further information stored in related documents." [13] identified and classified main dimensional patterns that normally occur in specifying dimensions. Data warehousing methodologies share a common set of tasks including business requirements analysis, data design, architecture design, implementation, and deployment [10, 15]. Data cleansing deals with detecting and removing errors and inconsistencies from data in order to improve its quality, and is typically required before loading the transformed data into the data warehouse [8]. Effective CRM analyses require a detailed data warehouse model that can support various CRM analyses and deep understanding on CRM-related business questions [3]. "Most data warehouses are designed with the ER model, complemented by object-oriented software engineering methods like UML. The Extract-Transform-Load (ETL) processes efficiently update the data warehouse with batches of new records. Since ETL builds de-normalized tables and transforms existing tables for integration, it is considered a part of data warehouse modelling. ETL is time-consuming and difficult in data warehousing, but easier and faster in MapReduce/DFS [20]." "The primary data collection methods used for the study were semi-structured interviews and document analysis. Chances of successful implementation are higher when the data warehousing project is planned, committed to, and managed as a business investment, rather than a technology initiative [23]". "The Inmon's approach requires the creation of a data warehouse ER model as a first step. The result of this process can then be used in subsequent steps as a basis for modelling dimensional and non-dimensional extracts. In the Kimball approach, dimensionally modelled structures are generated without creating an underlying ER model for them [14]." By selecting the most cost effective set of materialized summary views, the total of the maintenance, storage and query costs of the system is optimized, thereby resulting in an efficient data warehousing system [2]. Executives must have the right data to make strategic, tactical, and operational decisions [11]. [1] mentioned that decision makers require concise, dependable, information about current operations, trends, and changes.

2.2 Lack of Standardization:

Though several conceptual models have been proposed, none of them has been accepted as a standard so far [12]. There is no standard method or model that allows us to model all aspects of a data warehouse. Interest on physical design of a data warehouse has been very poor [12]. [18] argue that most existing modelling approaches do not provide designers with an integrated and standard method for designing the whole DW. There are no agreed upon standardized rules for how to design a data warehouse to support CRM and how to effectively use CRM technologies [3].

2.3 Data Warehouse Framework:

[12] proposed a two level framework containing requirement level and design level for requirement gathering and constructing UML designs respectively for data warehouse conceptual design. [22] focuses on the problem of representing OLAP databases and their query language.

To this aim, the author first defines a framework based on functional symbols annotated by typing information. Then, once the basic multidimensional database has been defined, query constructs are specified as higher-order polymorphic functions, and queries are expressed as complex functional expressions. [17] present a framework in which end users (i) specify their information preferences by ordering the different parts of an OLAP query (e.g. dimensions, classification hierarchy levels, and so on), and (ii) define their visualization constraints. [9] Identified 11 critical success factors for data warehousing projects with their relative importance. Among most important were quality of source data, clarity of business needs and objectives, and ways to measure benefits drawn from the developed data warehousing solution. By selecting the most cost effective set of materialized summary views, the total of the maintenance, storage and query costs of the system is optimized, thereby resulting in an efficient data warehousing system. [2]

2.4 Data Warehouse Failure:

[7] claim that, a significant percentage of data warehouses fail to meet their business objectives. The authors argue that requirement analysis is typically overlooked in real world DW projects. Maintenance is cited as one of the leading causes of data warehouse failures. Warehouses fail because they do not meet the needs of the business or are too difficult to change with the evolving needs of businesses [25]. Recent statistics indicate that 50% to 80% of CRM initiatives fail due to inappropriate or incomplete CRM processes, poor selection and design of supporting technologies (e.g. data warehouses)[3]. If your data warehouse is not driven by a significant and legitimate business need is not worth the investment. The "build it and they will come" approach is technology looking for a solution, and is responsible for the 70 % of failed data warehouse efforts [5]. 40% to 50% of data warehouse initiatives end in costly failure. The search for root causes conversed on not understanding the user's business problems [11]. Despite the booming data warehousing market, a large number of costly data warehouse initiatives are ending in failure [24]. Connor [4] estimated failure rates at 40%. "Every organization that initiates a data warehousing project encounters its own unique set of issues around a common set of factors. They include the business climate in which the organization exists, project sponsorship, organizational issues, the information intensity of the organization, the technological sophistication of the organization, the age and quality of the operational systems, the quality of the data, and the existing decision-support environment [19]".

3. PROPOSED METHODOLOGY FOR DATA WAREHOUSING:

In this paper we have standardized data warehouse development cycle by giving a framework that covers the complete process of data warehousing as a whole. We have followed an approach where standard steps of software development life cycle (SDLC) are applied with some adjustments for data warehousing projects. Data warehousing methodologies share a common set of tasks with SDLC including business requirements analysis, data design, architecture design, implementation, and deployment [10, 15]. Our methodology is based on fundamental approach of software development process. We have also integrated four step data warehousing development process introduced by Ralph Kimball [16]. The key to data warehousing success is to effectively identify data analysis needs of the organization. We propose to start by analyzing the reporting expectations of the business users then design solution that could meet out the reporting needs of the organization. During requirement analysis, we also need to determine infrastructure requirements, user training requirements, and expected project duration. Our focus has been on identifying techniques to be followed during each step of data warehousing to get an effective

solution. We have applied multiple methods of system investigation like Use of Questionnaire, Personal Interviews, document study, operational system input and output investigation, and user query expectations for designing data warehousing solution. Thorough investigation of operational system is done because it is the source of data for data warehousing solution.
Following are broad steps considered for design and implementation of data warehousing solutions.

1. Investigating Design of Operational System
2. Design & Implement dimensional model for Data Warehouse
3. Design & Implement ETL process
4. Select Data Warehouse accessing tools & techniques

In the above approach, interaction between consecutive steps is the key of the whole process. e.g. during dimensional modelling and design, operational system design is input and during ETL process dimensional model works as input and so on. Though there are many issues that come across while developing data warehousing solutions, but the most important is schema design as it has to hold data required for analysis. To get effective design of data warehousing solutions, we need to primarily focus on schema design for which we need to identify sources of input for schema design and deal with them effectively. In general following are sources of input:

1. Organization's business documents
2. Operational system of the organization
3. Business users and analysts

Above sources are highly useful and supplement each other in the overall schema design for data warehousing solution. Business documents help in understanding the vision, goals and objectives of the organization, details of existing infrastructure and future expansion plans. With the help of vision, goals and objectives of the organization, we can identify type of information required to analyze and monitor organization's performance. Details of infrastructure availability help in identifying size and geographical setup of the organization and volume of transactions. We have designed dimensional model and data warehouse schema for NRHM project by applying the proposed methodology.

Table 1 summarizes steps of the proposed approach for data warehouse development. This approach forms basis for the framework and sets guidelines for effective data warehousing. We have identified activities to be performed during each step and their expected outcome based on experimental verification on a live business case of NRHM.

Table 1: Steps for Data Warehouse Development Process as proposed in the framework

Step No.	**Step Details**	**Activities to be performed**	**Purpose/Objective/Outcome**
1.	Analyze the organization	Know goals and objectives of the organization, investigate key business processes, and identify organizational structure.	Helps in identifying reporting needs. Knowledge of business processes gives idea about facts and dimensions. Knowledge of organizational structure gives idea about different business users.

2.	Analyze Operational System of the organization	Collect details of operational system in use. Find out the functionality supported and the reports generated. Also identify reports the system is not able to generate	Finds quality of operational system to determine its dependability on data warehouse design. Identified problems of the operational system could be resolved to get quality data inputs in future for the data warehouse. Determines further reporting needs and accordingly helps in designing dimensional model for data warehousing solution.
3.	Assess current level of user satisfaction	Interact with different user groups to know their work profile, reporting requirements, points of satisfaction and dissatisfaction with the operational system	Helps in understanding different user types, the issues they are not able to deal with the existing operational system and their expectations from the data warehousing solution.
4.	Analyze operational system data structure	Investigate data input forms of the application, physical document formats, database structure of the application with an objective to know data items that are maintained	Helps in identifying data items that are useful for data warehousing but are missing in the operational system.
5.	Analyze query expectations of business managers	Interact with business managers to know type of queries they need to run and kind of information they want	Helps in deciding design of data warehousing schema, grain of measurements to be maintained.
6.	Identify key business processes to be considered for data warehouse design	Interact with different user groups to know what business processes are important in their organization that need to be analyzed	To determine fact tables and numeric measurements the organization is interested in
7.	Identify dimensions of analysis for each business process	Interact with different user groups to know the parameters that are important for analyzing key business processes	To determine dimension tables and their attributes
8.	Determine grain of data to be considered	Assess reporting requirements in terms of lowest level of data aggregation	Determines level of reporting required in terms of atomicity and grain for fact tables

9.	Identify update-frequency required in the data warehouse	Assess nature of reporting requirements in terms of time intervals	Helps in identifying real time data warehousing requirements and frequency of ETL execution
10.	Design dimensional model for the data warehouse	Design star, snowflake and fact constellation schema	To design dimensional model (logical schema) for the data warehouse
11.	Design ETL process	Identify all sources of data, Design data cleansing rules and its code	To design and implement ETL code for data warehouse population
12.	Implement data warehouse	Create tables and run ETL code to populate data in the warehouse	Data warehouse physical implementation is done to make it ready for business analysis
13.	Test data warehouse	Run data warehouse accessing tools to generate various reports	Helps in identification and rectification of any design or implementation issues in the data warehouse schema, ETL code or data accessing tools to ensure effective data warehousing
14.	Deploy the solution at the client site	Setup the infrastructure at the client site, install the required software, create and populate data warehouse schema	To ensure uses of deployed data warehousing solution by the business managers of the organization

Table 2: Steps for post development maintenance process as proposed in the framework

Step No.	**Step Details**	**Activities to be performed**	**Purpose/Objective/Outcome**
1.	ETL Execution	Identify frequency of ETL execution and run its code accordingly	Data updation in the data warehouse as per pre determined schedule to make data available for analysis
2.	Maintenance of multiple materialized views	Identify query execution patterns and accordingly update materialized views available in the data warehouse	To ensure availability of appropriate materialized views in the data warehouse for better data analysis
3.	Index updating	Monitor query execution performance and indexes in use and accordingly update the indexes	To facilitate fast query execution and provide high performance to business users

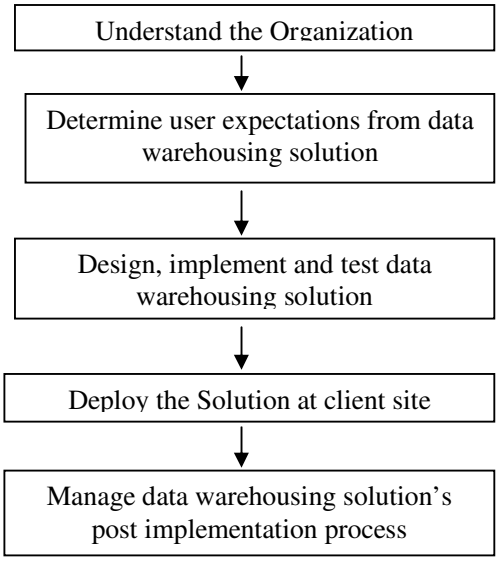

Figure 1: Basic Milestones for Data Warehouse Implementation

Figure 1 shows Basic Milestones for Data Warehouse Implementation. Our methodology is based on these milestones. We have followed an approach of data warehouse design and implementation in which the first step is to analyze user requirements and expectations from the data warehousing solution. We then identified the type of reports user wants for what kind of analysis.

Following are basic components of the Framework for developing data warehousing solutions:

1. Data warehouse models
2. Data warehouse design
3. Query processing and view maintenance

Our proposed framework is based on procedural aspect of data warehousing. Its components are identified on the basis of key steps to be taken. We worked on finding the effective methods to be followed during each of the following steps:

1. Analysis of the organizational needs
2. Design of solution
3. Implementation
4. Post implementation maintenance of the solution

3.1 Data Sampling and Validation of Questionnaire:

For data warehouse implementation of NRHM project, universe or data population is house hold survey data of whole country, patient registration and services data at various government and private hospitals that are providing different health facilities as per goals and objectives of Indian

government under NRHM project. Data population is very big. This project was originally focussed on rural population of our country but has now expanded its horizon to cover both rural and urban population of India. NRHM run seventeen health programmes out of which mother and child health is most focussed programme. We have considered mother ante natal care (ANC), Delivery, and post natal care (PNC) as sample. Data studied and collected is concerned with mother health during and post pregnancy. NRHM project is running nationwide in all states through different primary health centers (PHCs), community health centers (CHCs), health sub centers (SCs), sub district hospitals (SDHs), and district hospitals (DHs). Taking all health activities nationwide for research work would be very lengthy. Thus we have considered mother health care for our work which is representative sample. Based on guidelines of the proposed framework, investigation was carried out at different facility centres in multiple states like Uttar Pradesh, Madhya Pradesh, and Rajasthan to have quality representation of requirements across the whole population nationwide. Questionnaire was prepared to carry out investigation. Validation of questionnaire was done by putting the thoroughly prepared questionnaire based on study of NRHM project using multiple investigation techniques before faculty at state training institute of NRHM project to ensure quality of the prepared questionnaire. Questionnaire was distributed to the NRHM officers with whom personal interaction was done. We interacted with Chief Medical Officers (CMO), Principal Medical Officers (PMO), District Programme Managers (DPM), District Monitoring & Evaluation Officers, Block accounts managers (BAM), Assistant Research Officers (ARO), Health Education Officers (HEO), and Monitoring & Evaluation Officers MEO) at different facility centers or health department offices in states of Uttar Pradesh, Madhya Pradesh and Rajasthan to have representative sample of responses.

4. DATA WAREHOUSE IMPLEMENTATION FOR NRHM:

Data warehouse implementation of NRHM was proceeded as per our proposed framework as given in table no.-1. We have shown the dimensional model designed based on the steps and techniques of the proposed framework.

Questionnaire preparation is very important for requirement gathering. We categorized the questions in the following types:

1. Organizational questions
2. Questions on general awareness about functioning
3. Questions on organization's operational system
4. Questions about data collection and reporting methods
5. Questions on decision making requirements

For understanding requirements methods such as investigation of data input forms in print formats and in operational system software were referred. Various organizational documents like Health Management Information System (HMIS) user's manual, HMIS managers manual, HMIS reporting formats, HMIS Five years report, Mother and Child Tracking System (MCTS) data entry forms and reporting formats were also referred to identify data items that were being maintained and those which were required for data warehousing. HMIS is a reporting application where as MCTS is an operational system to track health and health facilities provided to pregnant women and newly born child. HMIS system is used for aggregated reporting from facility centres, district head quarters, state head quarters to the central government. MCTS application is another operational system used for tracking of mother and child health.

It was seen during the design that all dimensions are not required for all fact tables. For each dimension we have analyzed if it is required for a fact table. For e. g health provider medico may not be required for ANC registration therefore its id should not be referred in ANC checkup table. In delivery it was required hence was referred.

We explored each report type and input form in the operational system followed by listing out expected reports, assessing level of user satisfaction with the existing system, identification of missing reports, and shortcomings in the existing reports. It helped us in identifying dimensions and facts. We used attribute relevance techniques to eliminate the attributes which were not required for analysis. We took standard date dimension table. We analyzed the fields required in the user's queries for analysis based on which we retained only those attributes that were useful for analysis of NRHM data. Figure 2 shows steps of requirement analysis for data warehouse design.

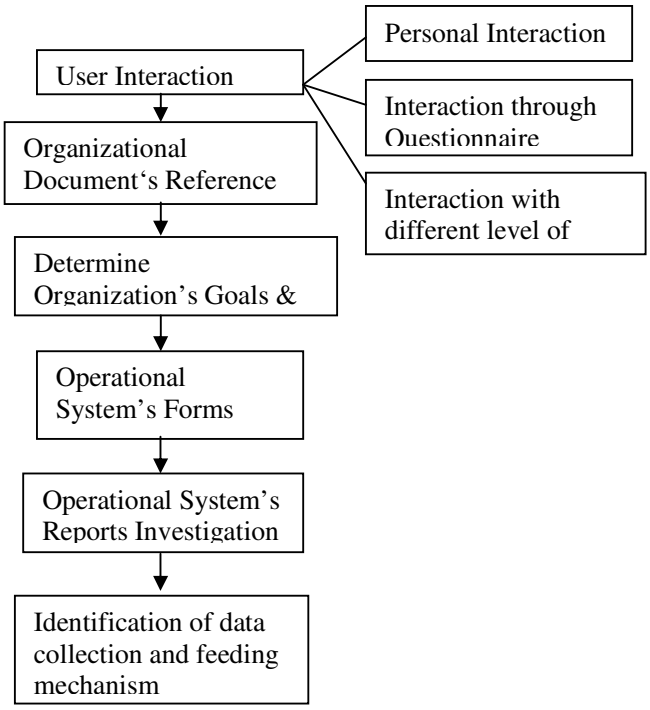

Figure 2: Steps of Requirement analysis for data warehouse design

4.1 Schema Design for NRHM:

For designing fact table we need to select the appropriate grain. We considered two models taking data at atomic level of grain and with daily grain respectively. Both the models have their own pros and cons. e.g. with atomic grain daily counts cannot be stored, whereas with daily grain the cases of individual women could not be extracted.

For designing dimension table, we initially took pregnant women as a dimension table and found that daily count maintenance was not possible in the fact table. We needed a solution to solve this problem. We determined that the womendemographic was the correct dimension table instead of women dimension.

We prepared a questionnaire for NRHM project. Sample questions of each category are given in table 3 to table 7.

Table 3: Sample Organizational questions

Sr. No.	Organizational Questions
1.	What are goals and objectives of your organization?
2.	What type of organizational activities you undertake? Please list out.
3.	What is organizational structure at your centre and also at national level? What is role at each level?
4.	Please list data items you record activity wise.
5.	Please specify type of decisions you need to take.
6.	What is basis of your decision making? Is it data based or experience based?
7.	Have you ever been provided any IT related training? Please specify details of trainings provided.

Table 4: Sample Questions on general awareness about organization's functioning

Sr. No.	Questions on General Awareness and Functioning
1.	Are you able to work effectively on your work profile? If no, then what are the reasons? Please specify.
2.	What is average frequency of patient visit at your facility center per day? Please specify category wise.
3.	What type of problems you face while making decisions?
4.	What facts you want to measure about your business activities and why?
5.	What are parameters on which you want to analyze data?
6.	Is historical data available to you for decision making? How old data is available to you?
7.	How do you monitor progress of working? How do you identify area where improvements have taken place? How do you identify factors responsible for the improvement?
8.	How do you identify areas that need further improvements? How do you identify factors required for such improvements?

Table 5: Sample Questions on organization's operational system

Sr. No.	Questions on Organization's Operational System
1.	Which operational system(s) is(are) in use in your organization? Please specify its role.
2.	What do you do with these operational systems?
3.	What are limitations of these operational systems?
4.	What kind of data is input in operational systems? Atomic, Daily count, Weekly count, Monthly count, or quarterly count?
5.	Is your organization able to collect and feed complete data in the computerized system? If no, what are reasons?
6.	What type of reports you are able to generate from the operational system?
7.	What kinds of reports are not available in the existing system?
8.	What are the problems you face while using these operational systems?

Table 6: Sample Questions about data collection and reporting methods

Sr. No.	Questions on Data Collection and Reporting Methods
1.	When do you get data from fields? What are data formats for data collection in the field?
2.	What are data formats used for data collection at facility centers?
3.	What type of reports you are able to generate from the current system? What kinds of reports are not available in the existing system?
4.	How data is managed? Manual system, computerized system, or mix of both?
5.	How do you ensure quality of data collection and management?
6.	Is your organization able to collect and feed complete data in the computerized system? If no please specify reasons.
7.	Do you collect data of those cases also that are not taking services of facility centers?

Table 7: Sample Questions on decision making requirements

Sr. No.	Questions on Decision Making Requirements
1.	What are data analysis needs of top, middle, and junior management?
2.	Specify details of decisions, you are required to take and why?
3.	How often you access business reports? Daily, weekly, monthly, quarterly, or annually?
4.	What type of reports you want from the computerized system?
5.	How do you identify areas where improvements have taken place?
6.	How do you identify factors behind these improvements?
7.	How do you identify areas where improvement is required?

Responses for some important questions are given in table 8.

Table 8: Responses of some Important Questions

Sr. No.	Responses During Interaction with Managerial Staff
1.	Data collection is manual to computerized which is gradually converting to fully computerized.
2.	Validation of data is on the basis of experience and some thumb rules that are approximate.
3.	We first started using HMIS and later for mother and child care additional application named MCTS was introduced. HMIS is reporting system for all 17 programmes.
4.	HMIS reporting is monthly, quarterly, and yearly whereas MCTS is online application.
5.	Data format for HMIS is XL files which are uploaded as per schedules.
6.	Very few facility centres upload data in HMIS system. Thus available data is incomplete.
7.	Field workers do not have idea about meaning of data he/she simply collects it. Supervisor corrects the collected data and compiles it.
8.	Supervisors are not computer savvy. Training is provided to supervisors to improve data quality.
9.	BP, Haemoglobin, ANC checkups, and weight of pregnant women is recorded and monitored. Low, medium, and highly anaemic cases are identified. 25 % ANM are not

	able to take correct readings, which affects the quality of data inputs.
10.	Data entry manager feeds data into the system. BAM and BPM monitor it.
11.	Block sends data to district and district send data to state head quarters.
12.	HMIS is developed with focus on mother and child care, immunization, and family planning. Other programmes are ignored. In every state additional s/w are developed locally.
13.	Some states use FMIS for financial management and SDMIS for store drug management. Janini express is developed for maternal transportation management.
14.	Man power shortage is a big problem. Behavioural attitude is a major problem in staff. Base level data quality is not good. Transaction level data entry at facility centre is not in practice. Operational system do not record many useful Input indicators required for decision making.
15.	Five major indicators are 3 ANC checkups, Institutional delivery, full immunization, sterilization, and IUDs. Additional indicators are Home delivery by TBA non TBA, weight, breast feeding in first hour, health follow-ups etc.
16.	Measured facts are ANC checkups count, TT1, TT2, Booster, IFA 100, HV level 7-11, <7, complicated cases, referred cases, treated cases, home deliveries by SBA, by non SBA, follow-ups within 24 hrs, BPL cases, 48 hrs follow-ups, and 11-14 days follow-ups count.
17.	In Rajasthan PCTS(Pregnant and child tracking system) is in use in addition to HMIS. In Rajasthan stock management, OPD, and IPD modules are computerized. 90 % data is correct. Early identification of pregnant women is poor. ANM is not able to cover specified area.
18.	Gap Analysis needs to be performed i.e. what the current system is able to give or how it is doing and what is the future expectations in terms of performance. Difference between the two is important to know for all segments of health services.
19.	Most of the NRHM staff is contractual. Multiple S/Ws are in use instead of a single integrated solution.
20.	Comparison of performance between two similar levels like districts is done.
21.	Ranking needs to be generated for various services among similar units of operations.
22.	Some facilities do not report data. This data becomes missing results in incomplete data.
23.	Weekly expected quantities and the actual weekly figures are compared. Quality of data is checked on the basis of expectations. There are some data integrity issues like double entry of child through ANM and also during institutional delivery.
24.	Monitoring performance of ANC, Deliveries, High risk pregnancies, maternal deaths, JSY etc. is required.
25.	Monitoring of live birth, still birth, child immunization, MTP, and FTP is required.
26.	Improvements are required in decision making capabilities.
27.	Increase of man power , IT personals and IT facilities is required.
28.	Sufficient health indicators are available for analysis.
29.	Weekly reports are not been generated and needs to be provisioned for. Data availability is not 100 %.
30.	Comparison of old data and new data against targets is done to monitor progress.
31.	We need grass root data in anaemic or high risk pregnancy cases.
32.	Improvement is required in reporting.
33.	We work as per priorities of government authorities.
34.	There are two types of profiles medicos and non medicos. Medico discharges duties in terms of health facilities whereas non medicos are involved in monitoring activities.

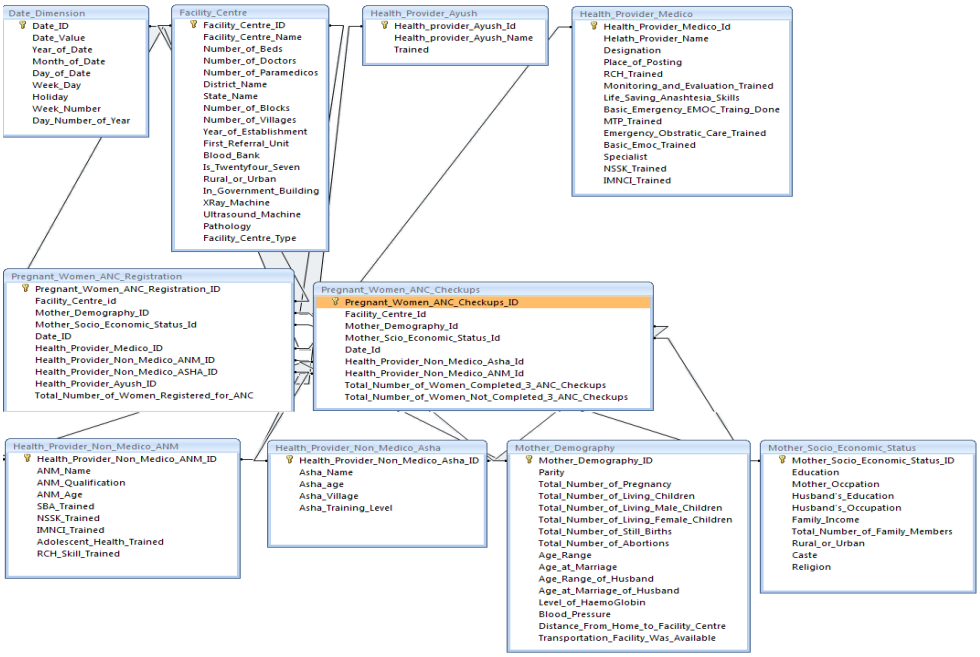

Figure 3: Dimensional Model for Mother Health Care during and Post Delivery

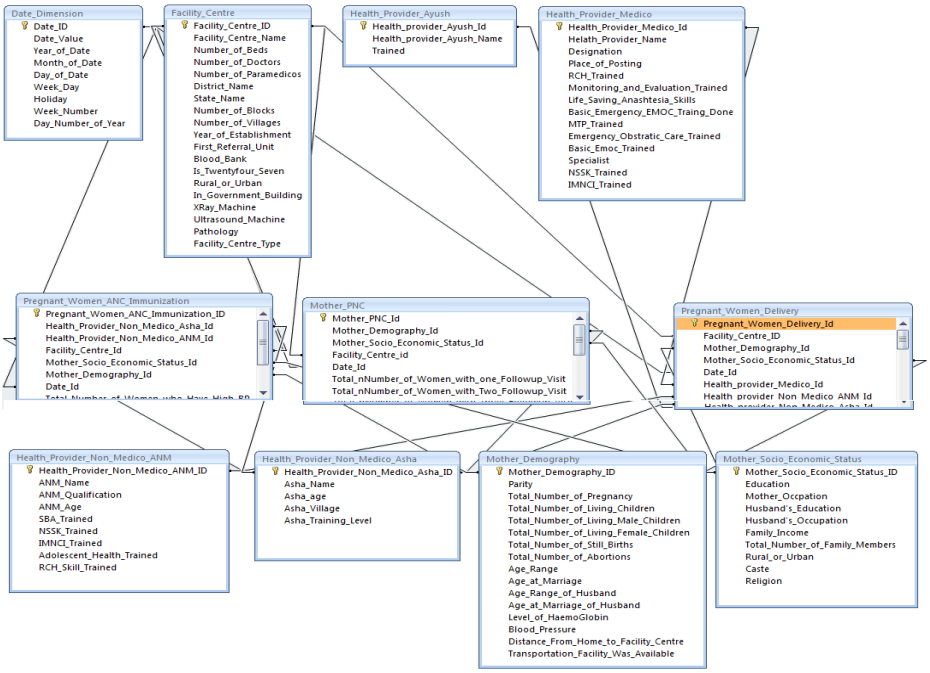

Figure 4: Dimensional Model for Mother Health Care during and Post Delivery (continued)

Figure 3 and Figure 4 show the dimensional model for mother health care during and post pregnancy.

Table 9: List of Fact Tables

Sr. No.	Name of Fact Table
1.	Pregnant_Women_ANC_Registration
2.	Pregnant_Women_ANC_Checkup
3.	Pregnant_Women_ANC_Immunization
4.	Pregnant_Women_ANC_Delivery
5.	Mother_PNC

Table 10: List of Dimension Tables

Sr. No.	Name of Dimension Table
1.	Date_Dimension
2.	Facility_Centre
3.	Health_Provider_Ayush
4.	Health_Provider_Medico
5.	Health_Provider_Non_Medico_ANM
6.	Health_Provider_Non_Medico_Asha
7.	Mother_Demography
8.	Mother_Socio_Economic_Status

List of fact tables is given in table 9. List of dimension tables is given in table 10. Grain of fact measurements was daily counts. The dimensional model for NRHM included measurements for pregnant women ANC registration, antenatal care checkups, ANC immunization, delivery, and Post natal care in five different fact tables. On the basis of inputs collected from different level of managerial staff about parameters of data analysis eight dimension tables were designed. For deciding grain of measurements initially we considered two different grains namely atomic grain and daily grain. Upon investigation, it was revealed that kind of queries the managerial staff was interested in could be executed with the help of data on daily grain and the estimated size of fact tables with atomic grain would have been significantly higher without giving any extra benefits. Thus we selected daily grain of measurements for all fact tables.

5. CONCLUSION & FUTURE SCOPE:

All enterprises need an information system which could aid in their decision making tasks. Data warehousing solutions are useful in meeting out such needs. Data warehouse research community has concern over high failure rates of data warehousing projects resulting in loss of heavy investments of time and money. In this paper we proposed a framework to overcome such issues. We took an approach of standard software development life cycle to standardize the process of data warehousing. We have designed a framework for effective data warehousing practices. Our framework is based on procedural aspects of data warehouse development and has included details of activities to be performed during each step and the objective/output at each of the stage. We designed dimensional model for NRHM project of Indian government to verify the quality of the proposed framework. The work in this paper mainly focused on defining steps for data warehousing and activities to be performed during each of these steps so as to get effectiveness, investigation of methods for system investigation such as Questionnaire, Personal Interviews,

Document Reference, Operational system Investigation, and Designing data warehouse schema for NRHM project. In future, based on responses of questionnaire from different industry domains and scales of business volumes specific guidelines for different industry and volume segments can be devised. Our future plans include population of the designed schema of the current paper with the data collected from different facility centers and other higher operational entities to test its performance for different data analysis needs of the project. Based on response of the developed solution, we shall try to further refine the proposed Framework. We have plans to work on a model of data warehousing in which integration of developing operational system and data warehousing solution is to be considered to give the organization a complete system of business operations and analysis. This approach would ensure effective implementation of both types of systems. It would also ensure better quality in the developed solution.

REFERENCES

[1] Lee, S.hyun. & Kim Mi Na, (2008) "This is my paper", ABC Transactions on ECE, Vol. 10, No. 5, pp120-122.[

[2] Chan. G. K. Y., Li. Q., and Fenf. L., "Design and Selection of Materialized Views in a Data Warehousing Environment: A Case Study", DOLAP' 99, Kanas City MO USA, ACM 1999 1-58113-220-4/99/11

[3] Colleen Cunningham and Il-Yeol Song, "A Taxonomy of Customer Relationship Management Analyses for Data Warehousing", Twenty-Sixth International Conference on Conceptual Modelling ER 2007 Tutorials

[4] Connor, D. (2003). Report: Data Warehouse Failures Commonplace. Network World, 20(3), 24.

[5] Ewen. E. F., Medskar. C. E., and Dusterhoft. L. E., "Data Warehousing in an Integrated Health System;Building the Business Case" DOLAP '98 Washington DC USA, ACM 1999 l-581 13-120-8/98/l 1

[6] Finnegan, P., Murphy, C., and O'Riordan, J. (1999). "Challenging the Hierarchical Perspective on Information Systems: Implications from External Information Analysis," Journal of Information Technology, Forthcoming.

[7] P. Giorgini et al. "Goal-Oriented Requirement Analysis for Data Warehouse Design", in "Eighth ACM International Workshop on Data Warehousing and OLAP", ISBN: 1-59593-162-7, ACM Press, NY, 2005...

[8] Hao Fan, Alexandra Poulovassilis "Using AutoMed Metadata in Data Warehousing Environments" DOLAP'03, November 7, 2003, New Orleans, Louisiana, USA.

[9] Hwang, M.I. & Xu, H., (2005, fall). "A Survey of Data Warehouse Success Issues", Business Intelligence Journal, 10(4), 7-13.

[10] Inmon W. H. "Building the Data Warehouse, 3rd Edition, Willey, 2002.

[11] James F. Kimpel "Critical Success Factors for Data warehousing: A Classical Answer to A Modern Question" Issues in Information Systems, Volume 14, Issue 1, pp. 376-384, 2013.

[12] Jindal. R. & Taneja. S., "Comparative Study of Data warehouse Approaches: A Survey", International Journal of Database Management Systems (IJDMS) Vol. 4, No. 1, February 2012

[13] M.E. Jones and I-Y. Song. "Dimensional Modelling:Identifying, Classifying, and Applying Patterns ", in "Eighth ACM International Workshop on Data Warehousing and OLAP", ISBN: 1-59593-162-7, ACM Press, NY, 2005..

[14] Kim. H., Lee. T., Lee. S., and Chun. J., "Automated Data Warehousing for Rule-based CRM Systems", (ADC 2003), Fourteenth Australian Database Conference, Adelaide Australia.

[15] Kimball R., Reeves. L., Ross. M., and Thronthwaire. W. "The Data Warehouse Life Cycle Kit" Willey. New York, 1998.

[16] Kimball R., & Ross. M., "The Data Warehousing Tool Kit" Willey, New York, 2005."

[17] L. Bellatrach et al. "A Personalization Framework for OLAP Queries", in "Eighth ACM International Workshop on Data Warehousing and OLAP", ISBN: 1-59593-162-7, ACM Press, NY, 2005...

[18] J-N. Mazón et al. "Applying MDA to the Development of Data Warehouses", in "Eighth ACM International Workshop on Data Warehousing and OLAP", ISBN: 1-59593-162-7, ACM Press, NY, 2005..
[19] MET/S Methodology Documentation Book Release 2.3, Cork: Data Warehouse Network.
[20] Ordonez. C., Song. Y., & Alvarado. C. G., "Relational versus Non-Relational Database Systems for Data Warehousing", DOLAP'10, October 30, 2010, Toronto, Ontario, Canada.ACM 978-1-4503-0383-5/10/10.
[21] J. M. Pérez et al. "A Relevance-Extended Multidimensional Model for a Data Warehouse Contextualized with Documents", in "Eighth ACM International Workshop on Data Warehousing and OLAP", ISBN: 1-59593-162-7, ACM Press, NY, 2005..
[22] K. Q. Pu "Modelling, Querying and Reasoning about OLAP Databases: a Functional Approach", in "Eighth ACM International Workshop on Data Warehousing and OLAP", ISBN: 1-59593-162-7, ACM Press, NY, 2005...
[23] Sammon, D. & Finnegon, P., "The Ten Commandments of Data warehousing", "The DATA BASE for Advances in Information Systems - Fall 2000 (Vol. 31, No. 4)"
[24] Sen, A., Ramamurthy, K. (Ram), & Sinha, A.P. (2012) "A Model of Data Warehouse Process Maturity", IEEE Transactions on Software Engineering", 38(2), 336-353. doi:http://dx.doi.org/10.1109/TSE.2011.2
[25] Sen. A, and Sinha. A.P. "A Comparison of Data warehousing Methodologies", " Communications of the ACM, March 2005, Vol. 48, No. 3"
[26] Trujillo. J. & Song. Y., "A Report on the Eighth ACM International Workshop on Data Warehousing and OLAP (DOLAP' 05), SIGMOD Records, Vol. 36, No. 4, December 2006.
[27] Business Intelligence: A managerial approach. (2011) (2nd ed.) Boston: Prentice Hall.
[28] http://nrhm.gov.in
[29] http://nrhm-mcts.nic.in

AUTHORS

Deepak Asrani is a Ph. D scholar of Teerthanker Mahaveer University, Moradabad, U.P., India. He is a B.E. (Electronics and Telecommunication), and M.S (Software Systems) from BITS, Pilani, Rajasthan, India. He has Software industry and academic experience of 24 years in various organizations and academic institutions.

Dr. Renu Jain is Director of University Institute of Engineering & Technology, Kanpur, U.P, India. She has worked as Head in the Department of Computer Science Engineering in the same institutions. She is an eminent personality in academics and holds Ph. D in Computer Science from BITS, Pilani, Rajasthan, India. She has rich academic experience of more than 30 years and has published around 50 research papers in various reputed national and international journals.