ISSN:2320-0790

COMPUSOFT
An International Journal of Advanced Computer Technology

A methodical and adaptive framework for Data Warehouse of Salary Management System

Manzoor Ahmad

Scientist ‘D’, Department of Computer Sciences, University of Kashmir, Srinagar, J&K, India,190001
Email: manzoor@kashmiruniversity.ac.in

Abstract: Years of experience as an employee of University of Kashmir has always desired us to have a typical solution where most of the activities related to salary are fully automated without checking across the files whenever there is a need e.g. individual month’s salary report , web based information submission, filing of returns , increment information etc. After thorough analysis , taking employee satisfaction , sensitivity and security of data , a long term solution was to develop a centralized University salary management system and its data warehouse . In this paper the design and implementation of an adaptive data warehouse is presented which supports large volume of data and saves the cost effectively. It also enable decision makers pose queries and question to the system. However decision support systems only support a set of queries and operations that are to be performed.

Keywords: Operational systems , STAR Schema , Data Marts , Data Granularity , Data Fusion , OLAP (Online Analytical Processing)

I. INTRODUCTION

Analytical processing of the data has become an important activity for its optimum utilization in the decision making of an organization. A 'data warehouse' is a repository of selected organization's data from its online transaction databases as well as other disparate data sources which are designed to facilitate reporting and analysis[1] for the decision makers of an organization..This classic definition of the data warehouse focuses on data storage. However, the extraction, transformation and loading data, and to manage dictionary data and analysis are essential components of a data warehousing system [2]. A data warehouse is a subject-oriented, integrated, time-varying, non-volatile (mostly read only)collection of data that is used primarily in organizational decision making. A decision support database that is maintained separately from the organization’s operational databases. Data Warehouse is designed around “subjects” rather than processes. These operations depend more on the way the data is stored. There are two main approaches to for designing the tables where the data can be stored i) Dimensional Approach and ii) Normalized approach. In the dimensional approach, transaction data are partitioned into "facts", which are generally numeric transaction data and "dimensions", which are the reference information that gives context to the facts[3]. The advantages to this approach are that the results are given in less time since there are minimum joins and it is easier for the user to understand. In the normalized approach, the data in the data warehouse are stored in the. tables which are split based on the subject areas that reflect general data categories. The disadvantage with this approach is the amount of time spent for retrieving the results and in turn performing the lossless joins and at the same time preserving the dependencies. These approaches are not exact opposites of each other. Dimensional approaches can involve normalizing data to a degree[4].

The Salary management system is optimized to deliver a regular function i.e monthly preparation of payroll of all the employees of University of Kashmir. Every month large volume of salary data is created and stored separately in files and not aiding in any decision making or disclosure. Discrepancies if any are unresolved and the process goes on every month. In this paper we have implemented a Star Schema Model of Data Warehouse of the salary management system in a university which will enable us to build a Decision Support Database for future analysis.

II. .EXISTING SYSTEM

In University of Kashmir, more than 2500 employee work on regular basis either in teaching or non teaching jobs and more than 500 employees work either on temporary or adhoc basis. Every year some of the employee either reach their superannuation or voluntarily retire and also new employees join the organization. For these employees the University of Kashmir has a payroll system developed in 1980 by the computer center , University of Kashmir based on File Based information system developed in Fortran. In this system the data related to employees details and his salary information is input from files and the program then computes the salary for the month which is then stored again in a individual months salary files. Lot of Manual intervention and paper based record management has to be done.

If the employee has a increment it has to be manually updated in his salary information file. This system has various limitations in terms of Income Tax Returns ,Insurance Policy details and Bank Reconciliation . Whenever there is a change in the Dearness Allowance a special program has to be written for its calculation which is again stored in a separate file for many months if the effect is from some prior date. And thus no update is made to the monthly salary files leading to inconsistency. Moreover no employee information system is available to the employees of all ranks to check their salary or to report any sort of discrepancy. No Administrative decision making is possible in case the funds available are insufficient or the salary has to be released or not released depending upon late receipt of the activity duty by the individual departments. Supplementary salary has to done for the contractual or Adhoc employees manually and also involves in lot of paper record maintenance. Summary data about the Payments as well as Deductions , reductions or contributions is not available. Tax information is also not stored anywhere

III. REQUIREMENTS

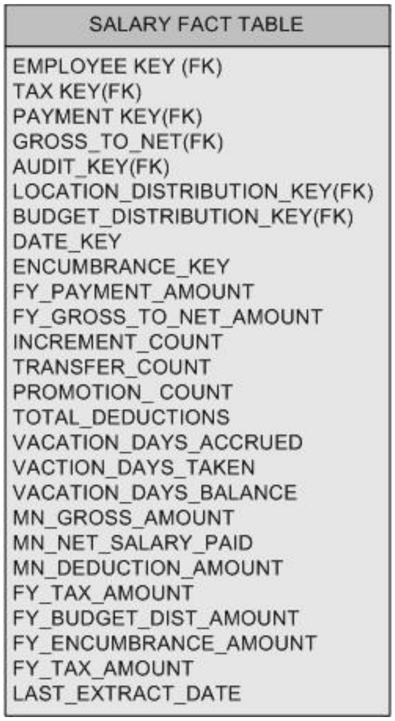
- The data must be intuitive and obvious to the user, be it employee, clerk or administrator.
- To change the file based information system to relational database management system to take advantage of RDBMS features like consistency , Concurrency Control ,Security and Recovery system in case of failures.
- To develop an intranet application where more than one user can use the application.
- To address the university authorities imperatives (more inline regulatory compliance , added accountability and efficiencies with paperless Processing and reports.
- Addressing the primacy of employee needs (Expanded online information extraction capabilities).
- Supporting administrative needs of performing inter disciplinary research needs e.g. Web based salary , Web Based Research Administration and Effort certification & reward system
- Slicing and Dicing of data must be possible in what everway the user (with appropriate right) wants.
- Access tools must be simple and easy to use with acceptable performance.
- If two performance measures have the same name, then they must mean the same thing. If two records relate to same employee they should not contradict each other. Consistent information means high quality information. Consistency also means that common definitions for the contents of the data warehouse are available.
- Data warehouse for the Salary Management System must be adaptive and resilient to change, changes to data warehouse must be graceful, meaning that they do not invalidate the existing data or applications.
- Data Warehouse for Salary management System must be secure that protects our information assets, Data Warehouse contains sensitive information pertaining to salary which if goes in the hands of wrong people can be potentially harmful.Data warehouse must effectively control access to the Salary confidential information.
- The data warehouse must serve as the foundation for improved decision making.
- The university community must accept the data warehouse if it is to be deemed successful.

IV. DATA WAREHOUSE PROCESS

The design of the data warehouse is a challenging task and comprises of 4 main steps which include i) identifying the Business Process ii) declaring the Granularity of the data iii) identifying the Dimensions iv) identifying the Facts. We are using a star flake schema for the dimension modeling of the data warehouse for the salary management system wherein we represent the data in a standard framework. The framework is easily understandable by end users and contains same information as ER model. This model packages data in symmetric format , is resilient to change and facilitates data retrieval and analysis. Within a data warehouse, information is split into two distinct classes, the basic factual information event, and the reference information that is used to analyze the factual event. The factual event contains the physical information that describes a factual event that occurred within the business (Salary Management System) in this case declaration of Salary is the best example. The star schema design consists of two types of tables the dimension table and the fact table. The dimension table stores the dimensions which contains textual descriptors of business. The fact table stores the facts or measure. The facts s are "numeric" & "additive. Fact and dimension tables form a Star Schema. "BIG" fact table in center surrounded by "SMALL" dimension tables. The Fact table contains numerical measurements of the business which are taken at the intersection of all dimensions. Intersection is the composite key. The fact table presents many-to-many relationships between dimensions.

Fact data within the data warehouse will form the bulk of the database volume. Fact data is the major database component of a typical data warehouse. Fact data represents a physical transaction that has occurred at a point in time and as such is unlikely to change on an ongoing basis during the life of the data warehouse.

Now the question arises which facts will appear in the fact table, the facts collected are shown below



SALARY FACT TABLE
EMPLOYEE KEY (FK)
TAX KEY(FK)
PAYMENT KEY(FK)
GROSS_TO_NET(FK)
AUDIT_KEY(FK)
LOCATION_DISTRIBUTION_KEY(FK)
BUDGET_DISTRIBUTION_KEY(FK)
DATE_KEY
ENCUMBRANCE_KEY
FY_PAYMENT_AMOUNT
FY_GROSS_TO_NET_AMOUNT
INCREMENT_COUNT
TRANSFER_COUNT
PROMOTION_ COUNT
TOTAL_DEDUCTIONS
VACATION_DAYS_ACCRUED
VACTION_DAYS_TAKEN
VACATION_DAYS_BALANCE
MN_GROSS_AMOUNT
MN_NET_SALARY_PAID
MN_DEDUCTION_AMOUNT
FY_TAX_AMOUNT
FY_BUDGET_DIST_AMOUNT
FY_ENCUMBRANCE_AMOUNT
FY_TAX_AMOUNT
LAST_EXTRACT_DATE

Examples of facts, Increment_Count, Promotion_count, MN_net _Salary. The grain chosen here is the net salary , it's the most atomic information captured, Atomic data is the most detailed information collected. Why we have chosen Salary as the grain is very obvious and true to the definition of the data warehouse.

Employee is appointed only once to the university, he/she can be given salary for each month at a time, and for each month he/she may be paid many allowances and also can offer deductions no of time. In data warehouse language the allowances and deductions are dimensions data that is used to analyze the actual data that is factual information.

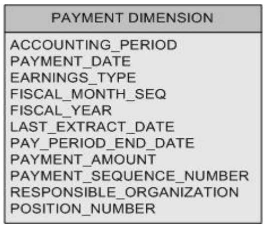
The facts collected by the Salary System include Monthly gross salary , Fiscal year gross salary , Fiscal year gross to net salary , Monthly total reductions or deductions or contributions , total yearly increments taken from the date of joining , total no of transfers in departments , total job activities handled at a time (e.g A employee may be a professor in a department but he may also be serving as the chief proctor and he may also be chief investigator for some project.)

Total no. of promotions , vacations taken in this case earned leave , vacations remaining , total vacations accrued .This table is populated with the declaration of every employees summary salary information .

Dimension Tables contains attributes for dimensions . Normally there are 50 to 100 attributes common and best attributes are textual and descriptive. Dimensional data is the information that is used to analyze the elemental transaction, example being Date, Employee etc. Structuring of the data in this way makes it possible to optimize query performance. Dimension data differs from fact data in a number of ways, which affect the way we treat it. First of all, dimension data will change over the time. This may be due to changes in the University, even Subscription code for some fund can change.

Because its invertible that there will be requirement to change dimension data, dimensions have been structured with a view to allowing rapid change. The dimensions of our concern are identified and one of them is shown below:

1. Employee Dimension : The Employee dimension table contains the most current biographic/demographic data for each employee appointed in the University, whether employee is a faculty or non teaching staff coming in the Plan or Non Plan category or the employee is an Adhoc or contractual employee.
2. Payment Dimension : The payment dimension table Contains data on payments to employees, and reallocations. There is one record per payment or reallocation (per person, check date, account number and earnings type).



PAYMENT DIMENSION

ACCOUNTING_PERIOD
PAYMENT_DATE
EARNINGS_TYPE
FISCAL_MONTH_SEQ
FISCAL_YEAR
LAST_EXTRACT_DATE
PAY_PERIOD_END_DATE
PAYMENT_AMOUNT
PAYMENT_SEQUENCE_NUMBER
RESPONSIBLE_ORGANIZATION
POSITION_NUMBER

3. The Gross to Net Dimension : The Gross to net dimension table Contains reductions, deductions, and contributions effected in this pay cycle. There is one record per person, payroll cycle, gross to net code, special transaction indicator, and source code.

4. The Location Distribution Dimension : The Location Distribution Dimension stores the place of posting details of all the employees appointed in the university.

5. Date Dimension : The DATE dimension is a unique and powerful dimension in every data mart and enterprise data warehouse Date dimension is the one dimension nearly guaranteed to be in every data mart because virtually every data mart is a time series.

6. Tax Dimension: The tax dimension contains data about tax and benefits information and amounts, by employee and pay cycle. This table is commonly used to determine taxable entities, benefits amounts, and clearing house (ACH) values by employee and pay cycle.

7. Job Table : The job table Contains data about employees and their jobs. There is one record per employee/job. Multiple versions of job records are included and are organized by job sequence number and extract date. This table is commonly used to determine what jobs an employee holds (or has held), the salary for each job, when the job appointment begins and ends, when the employee first began working in the job, the pay cycle, the percentage of full-time hours spent in the job, and the organization responsible for the job appointment.

8. Job_Class Table : The Job_Class table contains basic data about each job class, including title and several category codes. There is one row per job class. This table is commonly used for displaying the employee's job title, rather than the Job_Class code that appears in a related table. ("Don't show the codes in the report. Show the English instead.") Another common use is classifying employees by category (for example, Faculty_Class)..

9. Retirement_benefits table : The Retirement_benefits table contains benefits information relating to both retirement and health/welfare benefits This table is commonly used to obtain benefits-related information for individual employees.

10. The budget distribution Table : The budget distribution table contains fiscal year budget based salary distributions. The data is drawn from the Budget Office. This table is commonly used to locate the budgeted base salary distributions of employees

11. Encumbrance Dimension: The Encumbrance dimension contains data on the amounts currently or previously encumbered for the salary for each employee for each account. Encumbered funds have been set aside but have been spent for a specific planned approved expenditure (in this case, salaries).

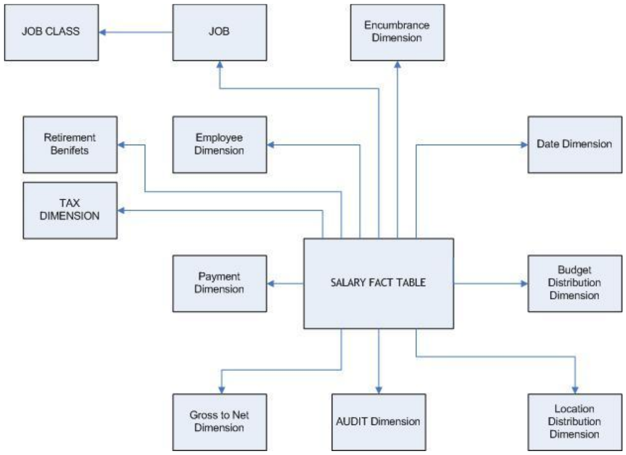
12. AUDIT Dimension: The AUDIT dimension describes our confidence in the quality of data for the Salary fact table . For example

    - Whether a specific "Not applicable "fact column is unknown , impossible , corrupted , or not available yet.
    - Whether a specific fact was altered after the initial load and if so why.
    - Whether the row contains facts more than 2,3,or 4 standard deviations from the mean or equivalently , outside various bounds of confidence derived from some other statistical analysis.

V. DATA WAREHOUSE DESIGN

One of the major technical challenges within the design of a data warehouse is to structure a solution that will be effective for a reasonable period of time.

This implies that data should not have to be restructured when the business changes or the query profiles change. Star schemas exploit the fact that the content of factual transaction is unlikely to change, regardless of how it is analyzed. Because the bulk information in the data warehouse is represented within the facts, it can be very effective to treat fact data as primarily read-only data with rear exceptions, and reference data as data that will change over period of time. If and when reference information needs to change, the underlying fact data should not have to change as well. Star schemas are physical database structure that stores the factual data in the center surrounded by the reference data. The method of normalizing the dimension tables in a star schema is called snow flaking When all the dimension tables are normalized the resultant structure resembles a snowflake with the fact table in the middle. In the snowflake schema the attributes with the low cardinality in each original dimension table are removed to form separate table.



VI. CONCLUSION

The data warehouse for salary management system for the University of Kashmir has been loaded with data which has been collected over the years in the form of text files and data from SQL Server. The data warehouse supports the quality management practice that consists of fact table which contains the summary data and the related information in the form of dimension tables. The summaries in the fact table were generated and proved to be correct after verification.. This dat warehouse can be extended with additional security features.

VII. REFERENCES

[1] Inmon, W. H., "Building the Data Warehouse", Second Edition, John Wiley & Sons, Inc 1996

[2] Larry, Greenfield, LGI Systems Inc., "The Data Warehousing Information Center," 1997 pp http://pwp.starnetinc.com/larryg/index.html.

[3] Kimball, Ralph, "The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses", John Wiley & Sons, Inc, 1996

[4] A. Gupta, V. Harinarayan, and D. Quass. Aggregate query processing in datawarehousing environments. In Proc. 21th Int. Conf. on Very Large Data Bases, Zurich, Switzerland, 1995.

[5] ACM/ IEEE CS Joint Task Force for Computing urriculum 2005. "Computing Curriculum 2005". The Over view report" 30 Sep, 2005

[6] C. Fahmer, and G. Vossen. A survey of database transformations based on the Entity Relationship model. Data & Knowledge Engineering, vol. 15, n. 3, pp. 213-250. 1995.

[7] CAI Yong, HE Guangsheng, "Designing Model of Data Warehouse with OO Method [J]", Computer Engineering and Applications, 2003.6.[5]

[8] Fon Silvers, "Building and Maintaining a Data Warehouse," AN AUERBACH BOOK", CRC Press is an imprint of the Taylor & Francis Group, an informa business

[9] Jeff Lawyer, Shamsul Chowdhury, " Best Practices inData Warehousing to Support Business Initiatives and Needs", Proceedings of the 37th Hawaii International Conference on System Sciences 2004

[10] Jorge Bernardino, Pedro Furtado, Henrique Madeira," ACost Effective Approach for Very Large Data Warehouses", Proceedings of the International Database Engineering and Applications Symposium, 2002

[11] Krishna. "Principles of Curriculum Design and Revision: A Case Study in Implementing Computing Curricula CC2001". ITiCSE '05, June 27–29, 2005

[12] LIN Yu,etc, "The Principles and Applications of Data Warehouse [M]", Posts & Telecommunications Press, 2003.1

[13] R. Barquin, and S. Edelstein. "Planning and Designing the Data Warehouse",. Prentice Hall, 1996.

[14] REN Jinluan, GU Peiliang, ZENG Zhenxiang, "Research on the Methods of Designing Data Structure of Data Warehouse ", Computer Engineering and Applications, 2001.22.

[15] SvetlozarNestorov, NenadJukic, "AdHoc AssociationRule Mining within the Data Warehouse", Proceedings of the 36th Hawaii International Conference on System Sciences, 2002

[16] Syed Najamul-Hassan, MaqboolUddinShaikh, Uzair Iqbal Janjua," Data Warehousing an Academic Discipline "Curriculum Development Approach, Methodologies and Issues", 2006

[17] Wu Shuning, Cui Deguang, Cheng Peng ,"The Four-stage Standardized Modeling Method in Data Warehouse System Development" Proceedings of the IEEE International Conference on Mechatronics & Automation Niagara Falls, Canada • July 2005

[18] YUAN Hong, HE Houcun, "Online Analysis and Data Warehouse Modeling Technologies [J]", Computer Application Research, 1999.12.