

“The Need for World Class Documentation, Data and Record Centres in Universities”

¹Venkata Subrahmanyam C. V._{M.B.A., M.Phil., (Ph.D.)}, ²Dr. K. Ravichandran_{M.B.A., Ph.D.}

¹Research Scholar, Department of Entrepreneurship Studies, Madurai Kamaraj University,
MADURAI – 625 021 Tamilnadu.

²Professor & HEAD, Department of Entrepreneurship Studies, Madurai Kamaraj University,
MADURAI – 625 021 Tamilnadu.

Abstract: The primary purpose of this paper is - To discuss the current state of Data Management in Universities & To Suggest the ways in which Universities can manage their Data in a more effective & Efficient way rather than the conventional methods.

This paper deals with the Concepts of Record Centre, Data Centre, Data Warehouse, Data warehouse in University setup & Steps in the Implementation of Data warehouse in universities.

This paper also deals with the Challenges faced by Universities especially, Indian Universities in handling Data & Records & Suggests ways to deal with the problems of not only present but in the future too. This paper further deals with the Methods of Digitization of Documents and Records in the university setup.

Key Words: Record Centre; Data Centre; Data Warehouse; Data Warehouse in Universities; 21st Century Challenges for Universities; Methods of Digitization

I. Introduction:

The problem of storage of information is, as old as, the evolution of Communication, in the history of mankind, itself. Early man produced information using stone as the writing surface, used pictures & symbols to communicate and stored them in naturally formed caves.

With the invention of Paper and then Printing Press, huge volumes of information started flowing, and to store them araised, the concept of **Record Centres** which, came into existence in the late 1800s and became very popular in the early 1900s all throughout the World.

II. Record Centre:

A place where all Information / Records / Data is Classified, Categorized, Stored, Retrieved & Preserved, very scientifically & Safely under the Conspicuous observation of a Curator is called a **Record Centre**.

Record Centres classify Records, based on the need of the access of the records, and store them accordingly, in a scientific manner, in a best safe guarded place, under some conspicuous observation. In Record Centres, the Data / Records will be classified into 3 basic categories namely – Active, Inactive & Passive Data.

Data that is accessed very frequently in a regular interval of time is called **Active Data**. Daily use cash books, receipt books, letter pads etc. are all examples for Active Data.

Data that is used less frequently, but, is essential to run / transact with the system, occasionally is called **Inactive Data**. Annual financial statements, Company Constitutional documents, Employee HR records etc. are all good examples for Inactive Data.

Data that is never used, but is stored purely for the purpose of Preservation and for the purpose of giving access to the Future generations is called **Passive Data**. National Archives, University Preserved Libraries, Preserved Scripts Etc. are all good examples for Passive Data.

The biggest disadvantage with the Record Centres is their accession times. Often, it is very difficult to find the required piece of information that is buried in the huge stakes or volumes in the Record Centres. With the Destruction of stupendous amounts of records, during World War - I & World War - II, Nations started searching for ways to safeguard the information beyond destruction.

In, the Late 1960s & the early 1970s, with the Development / Evolution of 1st & 2nd generation Computers & their Peripherals, to store and Retrieve Information (Called Data) with fewer efforts and with greater speeds, given birth to the Concept of Data Centres.

Data Centre:

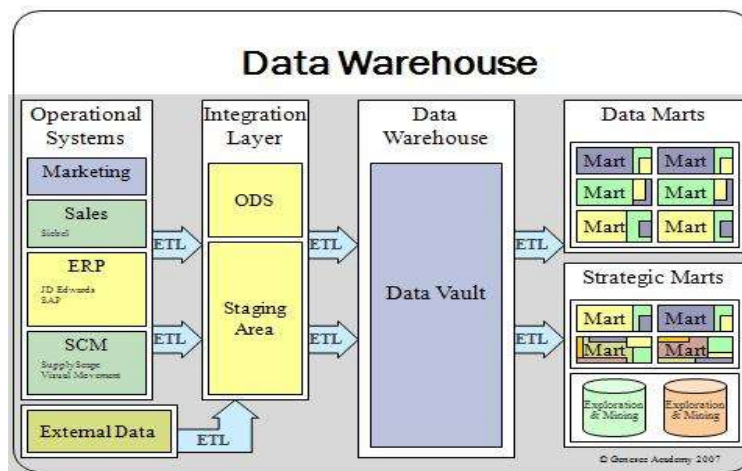
A place where the Information, in the form of Data, is Classified, Categorized, Recorded, Stored, Retrieved and Preserved with the assistance of Electrical / Magnetical / Electro-Magnetical / Electronic Devices under the conspicuous observation of a Curator called Administrator.

With the aggressive evolution of successive (3rd, 4th & 5th generation of Computer processors) generations of computers & their peripherals, and with the start of the information age, by 1990s, the shape and size of Data Centres have changed dramatically, aggressively & stupendously.

With the evolution of 5th & 6th Generation of Computer processors, which made Real-Time Computing possible; With the evolution of Internet & With the emergence of Global Data, the Data storage volumes have started growing significantly with time, from Giga bytes to Tera Bytes to Peta Bytes and more, and accessing these bulkier and heavier Databases became a herculean task which, gave birth to the Integrated Data Ware Concepts called Data Warehousing (popularly denoted as *DWH*), by late 1990s and became very popular by early 2000s all throughout the world.

III. Data Warehouse:

A *Data Warehouse (DW or DWH)* is a database used for reporting and data analysis. The following is the simple representation of DWH in pictorial form.



The data stored in the warehouse are uploaded from the operational systems (such as marketing, sales etc., shown in the figure to the Left). The data may pass through an operational data store for additional operations before they are used in the DW for reporting.

Data Mart - is the access layer of the data warehouse environment that is used to get data out to the users. The data mart is a subset of the data warehouse that is usually oriented to a specific business line or team. Usually, Data Marts will be given to read the Data in various reporting format with a restriction on editing the Data.

Strategic Data Mart(s) - is the access layer of the Data Warehouse environment with Limited / Full Edit access that is used to get data out to the user as well as to Edit / correct Data based on User inputs with restricted / exclusive access. In some deployments, each department or business unit is considered the owner of its data mart including all the hardware, software and data. These are called *Strategic Data Marts*.

Data Vault is the centralized Hardware / Peripheral environment (Physical Setup) that is used to Store / Retrieve data from the Data Warehouse system.

It is interesting to note that even today's ultra-modern & World Class Data Warehouses use the same classification of Data that is used in the Record Centres, to categorize & store the Data, in a Particular place & in a Particular Format within the Data Warehouse.

Advantages / Benefits of a Data Warehouse:

A data warehouse maintains a copy of information from the source transaction systems. This architectural complexity provides the opportunity to:

- Maintain data history, even if the source transaction systems do not.
- Integrate data from multiple source systems, enabling a central view across the enterprise. This benefit is always valuable, but particularly so when the organization has grown by merger.
- Improve data quality, by providing consistent codes and descriptions, flagging or even fixing bad data.
- Present the organization's information consistently.
- Provide a single common data model for all data of interest regardless of the data's source.
- Restructure the data so that it makes sense to the business users.

- Restructure the data so that it delivers excellent query performance, even for complex analytic queries, without impacting the operational systems.
- Add value to operational business applications, notably customer relationship management (CRM) systems.
- Presents Data in almost Real-Time which will be very much useful in taking Quick Managerial Decisions.

Need for aData Warehouse in Universities:

Universities are like a slice of the Society and almost resemble the society in a small (By Volume viz.) way. Every university offers Hundreds of Courses to Thousands of Students, using Hundreds of Resource Personnel & Thousands of Books / Materials etc. Managing all these interested groupssingly is, a difficult task, by itself. But, managing all these interested groups simultaneously, and at one place, in all possible permutations, and Combinations is, a real Herculean Task.

To make this job easier, many types of software like University Management System (UMS), Integrated Course Management System (ICMS) and many likes are available today. But, to run these softwares, an effective Data Warehouse is required without which even these high-end softwares become dumb and inoperable. So, it is the Need-Of-The-Hour for all the universities to maintain a proper & effective Data Warehouse.

Clients / Parties to University Data System:

In a data system environment, based on the nature of transactions, the different parties associated with the system are basically classified into 3 groups namely – Primary Users, Secondary Users & Tertiary Users.

Primary Users– generate data into the system, as well as, access the data from within the system.

Secondary Users– don’t generate any data into the System but, access the data from the system regularly.

Tertiary Users– neither generate any data into the system nor access the data regularly from the system but demand information according to their taste & need, as and when required.



In a university data system – Students, Employees & Affiliated colleges / institutions are Primary Users; Banks & Suppliers are Secondary Users; and Authorities, Government, Press & Other interested parties are Tertiary Users of Data/ information.

Data Warehouse in a University Setup:

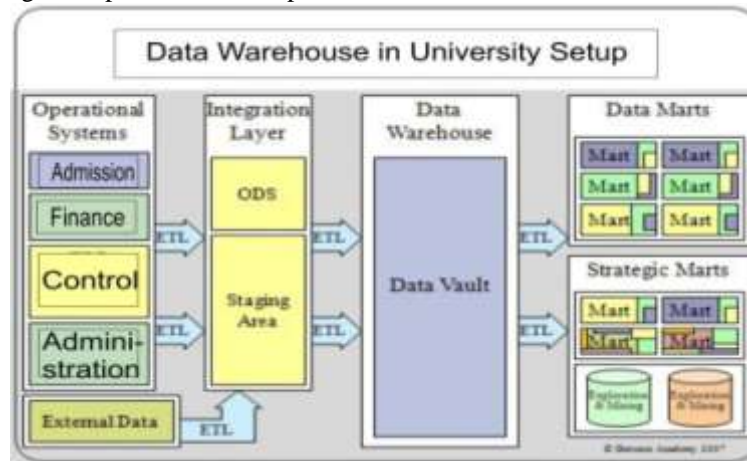
In a university Data Warehouse setup, Students and Employees are the prime sources of data. They generate data using software sub-systems, also called operational systems, like Academic Management System (AMS), Financial Management System (FMS), Learning Management System (LMS), Learning Content Management System (LCMS), Computer Aided Assessment system (CAA), Electronic Performance Support System (EPSS) and other such operational systems.

Unified Platform softwares like Integrated Course Management System (ICMS) or University Management System (UMS) act as an integration layer and pass / conduct all transactions through them to the Data Warehouse.

Student Access terminals or other interested parties terminals act as the Data Marts and takes output Data, in the form of reports, from the Data Warehouse.

Special controlled operations, like Computer Aided Assessment or Performance Valuation etc., will be given special permissions to Edit or to Append Data entries inside the Data Warehouse, act as Special Data Marts.

The Following is the pictorial block representation of Data Warehouse in university setup.



Steps in the Implementation of Data Warehouse in University Setup:

The following are the steps in the implementation of Data Warehouse in the university Setup / Environment.

1. **Business Case Analysis:** This task determines the business case for investing in a Data Warehouse. It identifies / determines the projected returns from the information content of the Data Warehouse over a 3 – 5 year period (ideally).
2. **Requirement Analysis:** This task identifies the information requirements for the short and medium term; specifically, the logical model for the Data Warehouse, and the kinds of Query that need to be supported after the 1st build phase. This task also identifies the source systems that will be providing the information for the Data Warehouse, and the business rules that need to be applied to the data.
3. **Technical Blue Print:** This task will define the technical architecture for the whole data warehouse, not just the 1st deliverable. It determines over all solutions, including –
 - ✓ Technical Architecture of solution (Database Architecture, System Architecture & Infrastructure);
 - ✓ Database sizings and query performance expectations;
 - ✓ Identification of the mechanisms for data transfer and load;
 - ✓ Recommendations for the Sizes and configuration of hardware;
 - ✓ Access Control, Backup and Recovery Guidelines.
4. **Extract and Load the Data:** This task will extract source data from the source system, and load it into the temporary area of the Data Warehouse. Other tasks will then clean up and convert the Data into appropriate structures.
5. **Design and Build Load Manager Process:** This task will deliver an automated system process that will initiate and carry out the extracts on all the required source systems. This system process will then load the data into the temporary staging area within the data warehouse.
6. **Design and Build Database Structure:** This task will deliver all the programs and database scripts that populate the data warehouse with the source data in the transient area. All user views, indexes, synonyms and aggregations are also created. This task can be complex, because the transformations to the data warehouse database structures may be involved.
7. **Validate and Clean Data:** This task will clean the Data in the transient storage area of the Data Warehouse, and make sure that it is fully consistent.
8. **Design and Build Operational Infrastructure:** This task delivers the programs and scripts that provide the backup / recovery, access control, and archiving facilities.
9. **Design and Build Warehouse Manager:** This task will deliver the system process that performs all the required activities of the warehouse manager. All the database and infrastructure operations are packaged into a system process that can be fully automated. These facilities must be capable of recovering in the event of failure, and of restarting without or with minimal human intervention.
10. **Design of Query Manager:** This task will design the Query Manager required to support the full data warehouse. Nothing is physically built within this phase; that is, it is typically deferred to a later phase, because the need to have a query manager can be bypassed by creating queries that access the required tables directly. As the complexity of the data warehouse and/or the number of the users increases, query management becomes more significant in order to optimize the use of system resources.
11. **End-User Reporting:** This task will produce the canned queries and reports, and configure the user access tool to access the data in the Data Warehouse.

12. **TESTING:** This task will specify and perform all the system, integration, volume and acceptance tests of the Data Warehouse.
13. **Transition to Production:** This task transitions the Data warehouse into a production environment. The initial set of data is preloaded, and the Data warehouse is performance tuned, if necessary.
14. **Project and Technical Management:** This is an on-going task, which performs project and technical management of the data warehousing project as a whole. It begins at the very start of the build phase, and continues until the data warehouse is delivered (that is, the end of the *Building the Vision* phase).

Challenges Ahead for Universities:

The biggest challenge, that is being faced, not only by universities but other organizations too, lies in dealing with the past data. With some kind of resources, many organizations are somehow successfully able to roll through the process of Automation and hence present data will be available within the system from the point of automation or so. But, it requires past data too, in order making the system completely online and accessible. However good a Data warehouse may be, but without a proper and complete database, it can only give dumb results.

In a recent Survey, conducted by McKinsey, it is estimated that, of the available world data (or archives), around 7% only got digitized so far and the rest of the data still remains in Analog form like books or Papers etc.

In spite of projects like Gutenberg and Many likes, working aggressively on digitizing the world archives, from the past few years, the volume of Digitization still remains at a pathetic level of 7 – 8% only.

This is the biggest challenge; almost any university is facing, in the current situations.

Methods of Digitization of Data:

The process of conversion of Hard copies into system readable files is called Digitization and there are so many methods available worldwide for converting the physical data into the intangible data form or to the Digital Data Form.

In a university setup, largest volume of Data lies in the form of books. So it is required to convert those volumes of books into system readable files using some method. The most popular and the cost effective (cheapest) method is scanning the documents.



The picture above shows the high resolution book scanner which is used for professional scanning of Books & other one side stitched documents. Using this equipment, the physical books & other Documents can be scanned and stored in searchable PDF (Portable Document Format) which consumes less volume of space to store and yet becomes searchable.

Suggestions:

- Universities' processes need to be automated and fully integrated using single platform integration software like University Management System (UMS) or Integrated Course Management System (ICMS) in order to bring speed, accuracy and precision into the organizations.
- Universities must equip themselves with a World Class Data Warehouse in order to complete the Data Integration with the System.
- Universities are advised to concentrate on the Digitization of Existing as well as past physical data into computer readable Digital form.

IV. Conclusion:

By implementation of Data Warehouse & Single Platform integrated Software, universities can speed up their transactions and can save nearly 60 – 70% of Time and human resources. Further, by DWH implementation, universities can save up to 35% of their total costs in administration.

References:

- [1] William H. Inmon (2005) “Building the Data Warehouse” 4th Edition Wiley Dreamtech Publishers.
- [2] Sam Anahory, Dennis Murray (2003) “Data Warehousing in the Real World” 6th Edition Pearson Education.
- [3] Patil, Preeti S.; SrikanthaRao; Suryakant B. Patil (2011). "Optimization of Data Warehousing System: Simplification in Reporting and Analysis". IJCA Proceedings on International Conference and workshop on Emerging Trends in Technology (ICWET) (Foundation of Computer Science) 9 (6): 33–37.
- [4] "The Story So Far". 2002-04-15. Retrieved 2008-09-21.
- [5] Kimball 2002, pg. 16
- [6] Inmon, Bill (1992). Building the Data Warehouse. Wiley. ISBN 0-471-56960-7.
- [7] Kimball, Ralph (1996). The Data Warehouse Toolkit. Wiley. ISBN 0-471-15337-0.
- [8] Hillard, Robert (2010). Information-Driven Business. Wiley. ISBN 978-0-470-62577-4.
- [9] http://mike2.openmethodology.org/wiki/Small_Worlds_Data_Transformation_Measure
- [10] Kimball 2002, pg. 310
- [11] "The Bottom-Up Misnomer". 2003-09-17. Retrieved 2012-02-14.
- [12] Ericsson 2004, pp. 28–29 "Data Warehouse".
- [13] Abdullah, Ahsan (2009). "Analysis of mealybug incidence on the cotton crop using ADSS-OLAP (Online Analytical Processing) tool, Volume 69, Issue 1". Computers and Electronics in Agriculture 69: 59–72
- [14] http://en.wikipedia.org/wiki/Data_warehouse
- [15] http://en.wikipedia.org/wiki/File:Data_warehouse_overview.JPG
- [16] http://en.wikipedia.org/wiki/Data_mart
- [17] http://en.wikipedia.org/wiki/Master_data_management
- [18] http://en.wikipedia.org/wiki/Data_center
- [19] <http://en.wikipedia.org/wiki/Digitizing>
- [20] http://upload.wikimedia.org/wikipedia/commons/6/65/Internet_Archive_book_scanner_1.jpg
- [21] <http://en.wikipedia.org/wiki/Communication>
- [22] <http://www.d7shred.com/records-center.php>
- [23] PaulrajPonniah. Data Warehousing Fundamentals for IT Professionals. Wiley, 2010, pp. 29–32. ISBN 0470462078
- [24] McQuail, D (2000) McQuail's Mass Communication Theory (4th edition), Sage, London, pp. 16-34
- [25] Flew, Terry. 2008. New Media An Introduction. South Melbourne. 3rd Edition. South Melbourne: Oxford University Press.