COMPUTERS AND ELECTRONICS IN AGRICULTURE 64 (2008) 111–119

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/compag

Original papers

Design and development of data mart for animal resources

Anil Rai, Vipin Dubey, K.K. Chaturvedi, P.K. Malhotra*

Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi 110012, India

ARTICLE INFO

Article history:
Received 23 August 2006
Received in revised form
27 March 2008
Accepted 19 April 2008

Keywords:
Animal resources
Data warehouse
Dimensional modeling
OLAP
Data mart
Decision support system

ABSTRACT

Planners, researchers, development agencies and farmers require information on animal resources for further studies and evolving realistic strategies for improvement and rearing of livestock and poultry. Data is also required for keeping watch on prices and movement of animal products, animal feed and establishment of services such as veterinary hospitals, artificial insemination (AI) centers, meat and dairy industries, etc. Further, there is a need to study animal resources in relation to other aspects of agriculture, such as soils, vegetation, agro-meteorology, socio-economic, land use, water resources for overall development of agricultural production system. Indian Agricultural Statistics Research Institute (IASRI), New Delhi has designed and implemented a Central Data Warehouse (CDW) under a National Agricultural Technology Project (NATP) Mission Mode sub-project entitled "Integrated National Agricultural Resources Information System (INARIS)". In this CDW, 13 different data marts related to various subjects in agriculture were designed, implemented and integrated. In this article, attempt was made to discuss concepts and problems of dimensional modeling of the animal data mart in relation to available source data on livestock resources in the country. Alternative solutions to specific problems were also discussed along with the solution implemented for modeling of this data mart. The complete process of building On-line Analytical Processing (OLAP) system for research managers including inbuilt technique of data quality and consistency checks to be implemented is being described with special reference to animal resource management. This article will provide guidelines for design and development of similar complex data marts in agricultural sector, particularly in the field of livestock management.

1. Introduction

Planners, researchers, development agencies and farmers require information on animal resources for further studies and evolving realistic strategies for improvement and rearing of livestock and poultry. Data is also required for keeping watch on prices and movement of animal products, animal feed and establishment of services such as veterinary hospitals, artificial insemination (AI) centers, meat and dairy industries, etc. Further, there is a need to study animal resources in relation to other aspects of agriculture, such as soils, vegetation, agro-meteorology, socio-economic, land use, water resources for overall development of agricultural production system.

Animal wealth of India is represented by a broad spectrum of native breeds of cattle (30), buffalo (10), goat (20), sheep (42), equines (6) and camels (8). In addition, several other animals such as poultry, duck, geese, quails, yak, mithun, pigs are also

* *Corresponding author*. Tel.: +91 11 25847122 25x4290; fax: +91 11 25841564.
E-mail addresses: anilrai@iasri.res.in, anilrai64@gmail.com (A. Rai).
0168-1699/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.compag.2008.04.009

important components of animal wealth of India. The livestock resources of India contribute 29% in the gross domestic product of the country. Statistical information on livestock resources in the country is being collected by different organizations, through integrated livestock survey, livestock census and ad hoc studies. Livestock census is being conducted after every 5 years in India. This provides age and sex wise data at district level for different categories of animals. State Animal Husbandry Departments and National Sample Survey Organization are conducting integrated livestock survey to get information on livestock products, prices and utilization of animal products all over the country. Surveys are also being conducted by different research organizations on estimation and assessments related to animal diseases, animal genetic resources, migration of livestock, etc. Still huge data gap persists in this sector. Information on livestock population and products such as poultry meat production, production of milk by products, numbers of animals from different breeds of a species, etc. are not available. Data gaps in terms of geographical coverage for collection of various information related livestock numbers and products still exist in the country. The problems of identification of data gaps, monitoring of data quality, integrated analysis of livestock sector and information dissemination to stakeholders exists due to non-availability of digital centralized repository of information for livestock sector.

A data warehouse is a read only analytical database which is used as a foundation of a decision support system. In other words, a data warehouse is a repository of integrated information of operational systems, available for queries and analysis to provide decision support. In general, data warehouses are subject oriented, integrated, time variant and non-volatile (Inmon, 1995, 2002). Data mart is a logical subset of an organizational data warehouse. The dimensional data marts are organized by a specific domain or by subject area. Basically, data warehouses are developed to enhance profitability, transparency and visibility of an organization. The design and implementation of a data warehouse present many challenges, i.e. data quality, accessibility and appropriate design (Agosta, 2001). It is reported that 40% of the data warehouse development efforts fails to meet design objectives (Kelly, 1997; Whiting, 2003). Presently, popularity of this technology is in banking, retail marketing, telecommunication, manufacturing, transportation etc. (Inmon, 2002; Whiting, 2003; Hoffer et al., 2005). However, adoption of this technology in government sector is slow due to data ownership, data privacy, data security and mainly due to lack in attitude (Bieber, 1998; Harper, 2004). United States Department of Agriculture (USDA) has developed one of the earliest data warehouse in government sector. This includes information from primary as well as secondary sources through various agricultural surveys, census and agribusiness (Yost, 2000). The information on pests, pesticides, and meteorological parameters from government of Pakistan has been integrated in the form of data warehouse (Abdullah et al., 2004).

Design and development of animal resource data mart is very important not only for decision making but also for understanding and solving problems related to livestock sector for sustainable development. Relationships of animal resources with other sectors of the country can also be studied through integration of information related to various sectors of agriculture with livestock sector. A mission mode sub-project entitled "Integrated National Agricultural Resources Information System (INARIS)" was taken up under World Bank funded National Agricultural Technology Project (NATP). The mission set for this project was to design and develop a state-of-art, flexible Central Data Warehouse (CDW) of agricultural resources of the country at Indian Agricultural Statistics Research Institute (IASRI), New Delhi (http://agdw.iasri.res.in/). This project has been implemented with active collaboration and support from 13 other organizations from Indian Council of Agricultural Research (ICAR), New Delhi (Rai et al., 2007).

In this article process of design and development of a data mart of animal resources in India has been described. The concepts and problems of dimensional modeling of the animal data mart in relation to available data on livestock resources in the country has been discussed. Alternative solutions to specific problems were also discussed along with the solution implemented for modeling of this data mart. The process of fact builds adopted for building various fact tables and brief description about each fact has been mentioned in this article. In order to maintain the data quality of the data mart, number of consistency checks on the source data was applied while moving it to the staging area. The process of extraction and data transformation at the various stages of data mapping are also described. The process of automation for updating data mart and On-line Analytical Processing (OLAP) functionalities are presented in this article. Results of end user evaluation surveys of INARIS project have been presented to demonstrate the utility of this data warehouse.

2. Dimensional modeling of animal data mart

The data warehouse database is maintained separately from operational databases. The organizational data warehouses are projected to hundreds of gigabytes or terabytes in size. Therefore, basic design and implementation issue of a data warehouse is query performance. In data warehouse, query, mostly ad hoc in nature, can access millions of records and can perform a lot of scans, joins and aggregates (Gupta and Mumick, 1995; O'Neil and Graefe, 1995; O'Neil and Quass, 1997). In this system, query throughput and response time are more important than transaction throughput.

In order to improve the query performance, the data in a data warehouse is typically modeled for multidimensional perspective. In case of data warehouse technology, the key performance indicators are known as fact. Facts are additive, non-additive and semi-additive. These are the numeric data items used to satisfy all calculation options that are of interest to the end user. The table which provides context to the fact is called dimensional table. The details of information can be visualized further to the lower levels, which are also known as grain level. Granularity of information is the lowest level of available fact information. Data storage in the data warehouse database is in the form of a multidimensional model, known as a cube.

In a data mart design, grain level of fact tables is to be decided first (Kimball, 1996, 1998; Kimball and Ross, 2002). The

decision regarding granularity depends on the level of detail at which fact need to be addressed (Bonifati et al., 2001).

Animal resources in India are highly diverse. Therefore, information on different aspects of animal resources is collected through a large number of data collection agencies. The problem is further compounded by the fact that there are no common standards that are applied in data collection procedures. Designing a data mart to integrate the collected information poses a formidable challenge to any data modeler. The available information of animal resources has been extracted from following operational source systems:

(i) Livestock breed database.
(ii) Livestock population database.
(iii) Livestock infrastructure and production database.
(iv) Livestock products and utilization database.
(v) Livestock import/export database.

Livestock breed database at national level has information related to breeds of different species of livestock animals and their characterizations. Livestock population database at district level has information from livestock census, which is being held after every 5 years. The information of livestock are categorized with respect to Sex, Age, Working Categories, etc. Information related to livestock infrastructure is being collected occasionally at district level. Information related to different livestock products and its utilization is being collected every year through integrated livestock survey at district level. Data related to import and exports of various livestock products are available on yearly basis at country level. These databases were developed and populated in Oracle 9i at National Bureau of Animal Genetic Resources (NBAGR), Karnal (India).

In order to use the information at micro/macro planning and decision-making level (Chen et al., 2003) data is to be integrated and aggregated properly. Further, it is also important to capture information at its lowest level so that there is no loss of information due to its aggregation from lower level to higher level data flow hierarchy. Therefore, several fact tables were designed to capture information at its lowest level. The information from these databases were extracted from the source system and after suitable cleaning, transformation and aggregations, moved into staging area of the central data warehouse. Consistency checks were also applied while extracting information from the source system.

2.1. *Dimensions and hierarchies in the data mart*

Three important dimensions were identified for development of animal resource data mart, i.e. (i) Animal (ii) Location and (iii) Time. Animal is a subject dimension. The hierarchical structure of this dimension has been given in Fig. 1.

Fig. 1 shows alternate multiple path hierarchy of animal dimension. It has three alternate paths. These paths were identified after careful analysis of information available in source systems. This is a non-covering hierarchy (Pedersen et al., 2001; Malinowski and Zimanyi, 2004) which is a generalized hierarchy with the additional restriction that at the schema level the alternative paths are obtained by skipping one or more intermediate levels. Animal hierarchy has three different paths. First path starts from Animal and ends on Working Category via Species, Gender and Age. The figures in brackets indicate cardinalities of respective levels. Here, Working Category cardinality depends on animal species. Therefore, w_i indicates the number of categories of ith species. Second path of this hierarchy is from Animal to Maturity via Species and Gender. Third and last path is from Animal to Working Category via Species, Breed, Gender and Age.

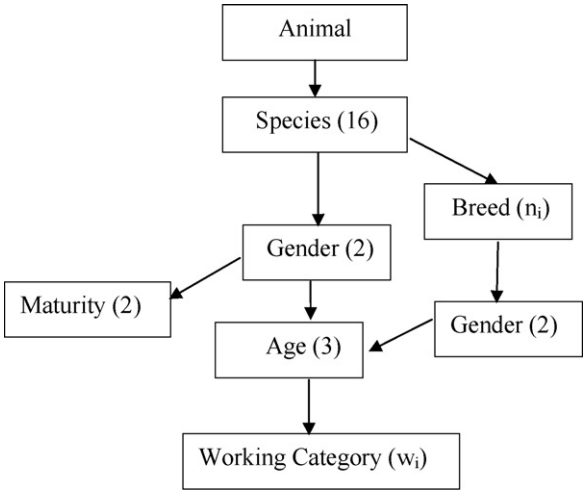


Fig. 1 – Hierarchical structure of animal dimension.

Several approaches were proposed in the literature to deal with these kinds of hierarchies, i.e. (i) by creating separate tables for each level of hierarchies based on different paths (Jagadish et al., 1999), (ii) using null values for absent levels of a attribute (Cabibbo and Torlone, 2000; Lehner et al., 1998), (iii) separate star schema for each path (Bauer et al., 2000; Hahn et al., 2000; Kimball and Ross, 2002) and (iv) creating one table for common levels and separate for specific levels (Bauer et al., 2000). In this data mart, dimension table were designed for every possible path in the above hierarchy.

Fig. 2 shows the Location hierarchical structure of the livestock information available in the source system. Similar to the Animal hierarchy, it also has three different paths.

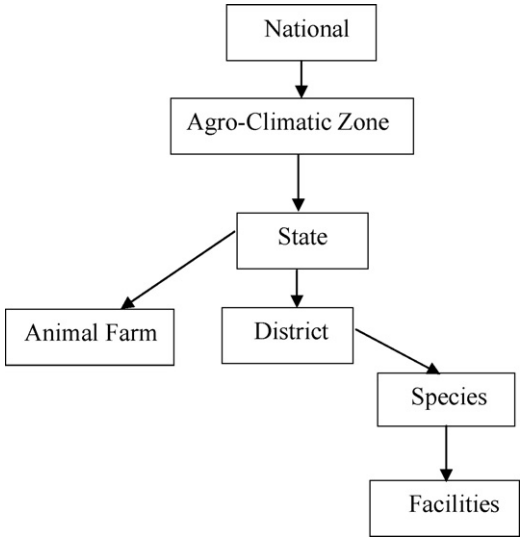


Fig. 2 – Hierarchical structure of Location dimension.

First path starts from National and ends at the District level via Agro-climatic zone and State. Second path starts from National and ends at Animal Farm instead of District. Third and last path starts from National and ends at Facilities via Agro-climatic zone, State, District and Species. In this case also separate dimension tables were designed for every possible path of the above hierarchy.

In this data mart three different definitions were followed for year in the TIME dimension. Therefore, TIME dimension shows all three definitions, i.e. financial year, agricultural year and calendar year associated with different information. Months in calendar year are from January to December whereas in agriculture year of India, it is from July to June. Similarly, months in financial year of India are from April to March. Therefore, particular agricultural and financial years are part of two calendar years. As a consequence of this, first month of calendar year i.e., January is seventh month of agricultural year and 10th month of financial year. In other words, first month of financial year i.e., April is fourth month of calendar year and 10th month of agricultural year. Also, first month of agricultural year i.e., July is seventh month of calendar year and fourth month of financial year. It can be seen that there are three different year systems followed for collection of data for agricultural sector of the country. Therefore, integration of data at monthly grain level is a major issue. Also, in case the data is available at grain levels of quarterly, half yearly, can also be integrated across years as quarters are offset by 3 months. The major problem of integration is for weekly and annual grain level information with respect to different year domains. This information can be integrated through some apportionment techniques at appropriate grain level.

3. Fact design

Designing of dimension models are followed by identification and designing of fact tables of the data mart. Initially, fact tables from single source data were identified. Then fact tables, which were to be generated from multiple data sources, were identified. Keeping in view, the user requirements and data availability at the source data, grain level of each fact table was decided. Declaration of grain level also provided information about the individual record level of each fact table. A good and clear declaration of grain level for each fact table, made it easy to choose appropriate dimensions, which can be associated with a particular fact table. Logical fact diagram of each selected fact table were prepared. The fact diagram shows not only the specifics of a given fact table but also shows the context of the fact table in overall data mart. The fact table diagram provides information about the name of the fact table, its grain level and dimensions connected to this fact. This serves as introduction to the overall model. Fig. 3 shows fact diagram of livestock production fact table. In this case, measure is "quantity of livestock products" such as milk, meat, wool, egg, etc. These quantities are measured either in weight or in number depending on the kind of product. For example, quantity of milk, meat, wools produced in a particular district of a state in a particular year are measured in units of weight while eggs produced with reference to same dimensions are measured in numbers.

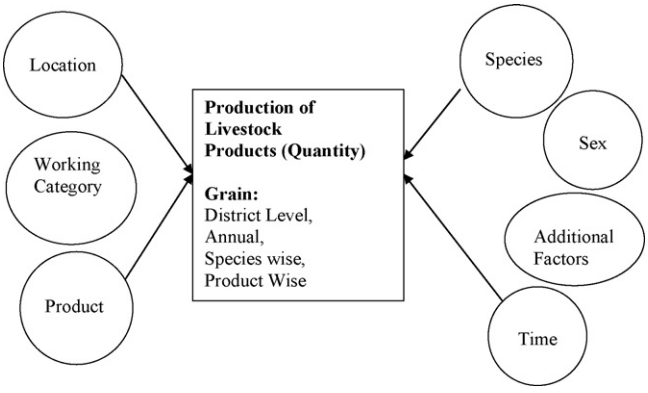


Fig. 3 – Fact diagram of livestock production fact.

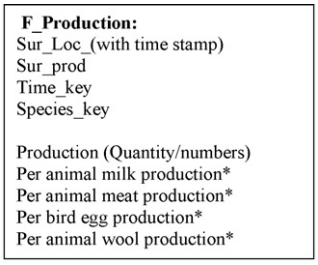


Fig. 4 – Fact table diagram showing dimension keys and facts.

It can be seen from this fact diagram that the availability of information about livestock products are at district grain level in "LOCATION" dimension, at the annual grain level in "TIME" dimension, at species grain level in "SPECIES" dimension and product grain level in the "PRODUCT" dimension. Some of the dimensions such as "SEX", "WORKING CATEGORY", etc. which may be relevant, but information is not available at these levels are also shown in the fact table. These can be integrated in the future as per availability of information.

The details of each fact were described in the fact table. The fact table provides a complete list of all the facts. This list includes actual facts in the physical table, derived facts and other facts that are possible to calculate from the first two. The fact table of the corresponding fact diagram is shown in Fig. 4. This diagram shows base facts and derived facts. Here derived facts are shown with asterisks (*) marks. Similar fact diagram and fact table diagram were designed for each of the fact tables of the data mart. The fact tables which were identified from the source databases for livestock data marts are presented in Table 1.

4. Data staging

Data staging area lies between data warehouse and data source systems. In this data mart, data staging has been extensively used to apply all consistency checks on each variables moving from source system to the data mart. All transformations and recoding functions for various indicators were applied in this data staging area to make data consistent and

Table 1 – Description about fact tables identified for animal data marts

Fact table name	Description
F_BREED	This fact table consists of detail information related to breed characterization of different animal species. Since most of the information is in textual form so there is no dimension associated with this fact. The information content of this table are breed name, species name, color, number of horns, horn's shape and size, visible characteristics, height of male and female, weight of male and female, fiber type, synonym, location of availability, main use, origin, adaptation to special environment, tract, mobility, etc.
F_CENSUS_BREED	This fact table contains district level data of livestock population census for few states where breed wise population of different species were recorded for few years in limited states only. Therefore, two hierarchy, i.e., district and breed are attached with this fact. It provides district wise, year wise information about male and female breeds count from livestock census
F_CENSUS_SPECIES	This fact table is similar to F_CENSUS_BREED table. In this, count (livestock population) fact related to various livestock census were captured at district level. Year, District, Species are dimensions in this fact table. Species is a hierarchy consists of Sex, Age (group) and Category (utilization) as levels
F_FARMS	This fact is related to animal farms. It provides information about the number of registered poultry farms in a state along with their location. Some information is available for different breeds also. It is associated with State lookup table and Breed hierarchy. It provides state wise, place wise and date wise information about farms, animals, breeds, category, sex, age and count (animal population)
F_INFRASTRUCTURE	This fact table contains information about availability of various livestock facilities such as artificial insemination centers, veterinary dispensaries, veterinary hospitals, veterinary institute, animal farms etc. at district level in a particular year. This also contains information related area coverage by these veterinary centers
F_POP_SPECIES	It contains information about species, age group, sex and category wise population of different animal species at district level for various years. It is associated with year, species lookups and district dimensions
F_POULTRY_PERFORMANCE	It provides poultry bird information about breed wise performance of different species
F_PRODUCTION	It provides district wise information of different livestock products such as milk, egg, meat and wool production, number of slaughter houses for different species of animal. This information is available for each year
F_PROD_BYPROD	This fact provides information of different animal products and by-products such as dung, egg, and milk at state level. This also provides information of utilization of these products and by products such as manure, cake for dung, consumed spoiled, hatched for eggs, sold as fluid, converted, etc. for milk. This has information for quantity and %value of the product and by product

uniform before migrating to data warehouse database. Further, different processes of data aggregations were also run in this data staging area.

Fig. 5 shows the process of mapping of data from original source system to target data mart in the data warehouse database. In this mapping process, initially data were extracted from the sources data bases (which may be in any digital format such a RDBMS data base, spreadsheet, text file, etc.) and moved in the staging area for suitable transformation, aggregation, cross-validation, maintenance of consistency and uniformity. In the staging area itself some of the derived parameters were generated using various single source variables. In the second step, data coming from different sources were integrated with each other using conformed and non-conformed dimensions and lookups. Again, important derived parameters were generated using multiple source data. Finally, suitable aggregation mechanisms were applied to generate different aggregation levels before populating the target data source.

The staging area server of IASRI, New Delhi receives on-line periodic updates from database server established at National Bureau of Animal Genetic Resources, Karnal (India). Data of the staging area is being processed through ETL server and uploaded to CDW. Further, information is being transformed in form of dynamic reports, OLAP cubes and Web GIS deployed on web portal for end users (Fig. 6)

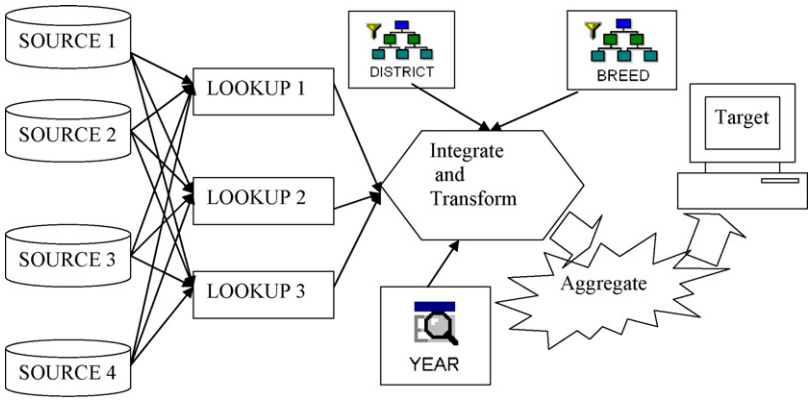


Fig. 5 – Mapping of a data from source system to the target data mart.

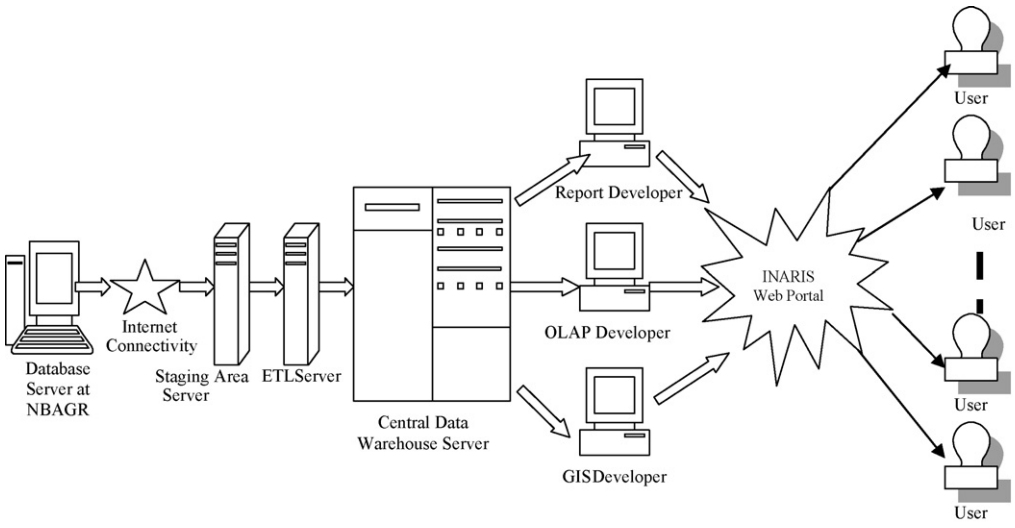


Fig. 6 – Infrastructural diagram of animal data mart.

5. Automation of data mart

Data mart in any data warehouse needs periodic updating. The updating policy of a data mart depends on process of data generation. In case of highly dynamic transaction processing system of a business process such as ATM transactions in a banking system there is need for daily updating the corresponding data mart. However, in case of livestock data mart data generation mechanism is quite slow. For example, production data of livestock products is being collected annually through integrated sample surveys of livestock products, livestock population census is being conducted biennially (once in a 5 year). Since, data generation time frame is different, so quarterly updating of this data mart is sufficient. In order to minimize human intervention in this updating process, it is desirable to automate whole process. This requires automation of data extraction, data transformation, data loading, exception/error handling, and logging/notification process of the warehouse to function properly. Data warehouse administrator can be benefited from a well-constructed job control process that comprises the following:

- Build status notification (send e-mail, if fails).
- Coordinates facts and shared dimension builds.
- Data staging and cleansing prior to create a data mart.
- Pre-processing and post-processing SQL.
- Different arrival rates of source data, etc.

The complete process of building animal data mart has been automated. This automation process was used to coordinate groups of builds, processing instructions, conditions and SQL into an operational process. The job control processes needed to keep data warehouse or data mart functioning more smoothly were implemented. A node was created for each type of operation and all operations were performed in a predefined sequence to maintain consistency of the system. The sequential or parallel nodes were created, arranged and executed accordingly. In fact the ultimate warehouse operation would run the regular load processes in the lights-out manner, i.e., completely unattended. Complete automation is difficult to obtain but it is possible to get closer to this situation (Kimball, 1998). Fig. 7 shows mapping of this automation process of animal data mart.

In order to automate the whole process of updating animal data mart, process has been scheduled in scheduler which can initiate the updating process according to fixed schedule. The process of updating starts from updating dimensional builds. These dimensions are connected with a node at the center, which checks the consistency of the dimensional table. In case, any inconsistency is detected by condition node in the incoming data from the source, certain procedures needs to be executed by procedure node to remove it. If it is not possible to remove this inconsistency then alert node is to be executed to send the alert signal and e-mail node will send e-mail notification to the system administrator. The alert signals and e-mail notification is to be send in number of situations when normal execution is not performed in any of the dimensions. After execution of all dimensions, facts are executed. Again, same notification system, as in case of dimensional builds, has been implemented for updating fact builds also (Fig. 7).

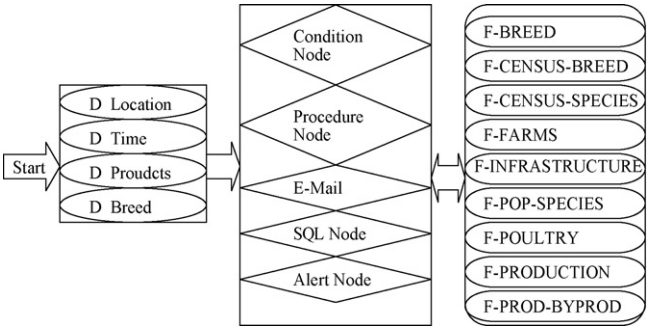
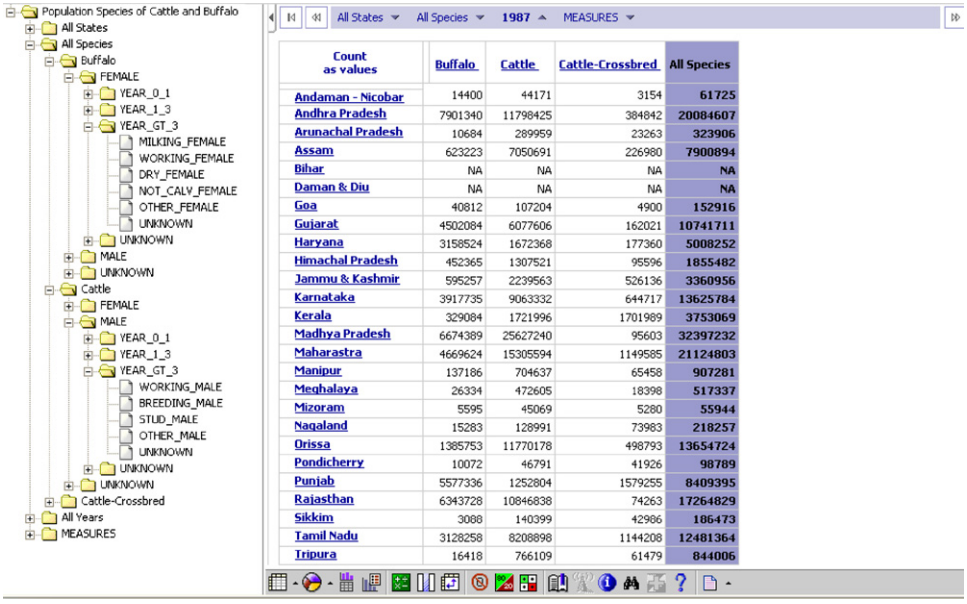


Fig. 7 – Mapping of Automation Process of animal data mart.



Count as values	Buffalo	Cattle	Cattle-Crossbred	All Species
Andaman - Nicobar	14400	44171	3154	61725
Andhra Pradesh	7901340	11798425	384842	20084607
Arunachal Pradesh	10684	289959	23263	323906
Assam	623223	7050691	226980	7900894
Bihar	NA	NA	NA	NA
Daman & Diu	NA	NA	NA	NA
Goa	40812	107204	4900	152916
Gujarat	4502084	6077606	162021	10741711
Haryana	3158524	1672368	177360	5008252
Himachal Pradesh	452365	1307521	95596	1855482
Jammu & Kashmir	595257	2239563	526136	3360956
Karnataka	3917735	9063332	644717	13625784
Kerala	329084	1721996	1701989	3753069
Madhya Pradesh	6674389	25627240	95603	32397232
Maharastra	4669624	15305594	1149585	21124803
Manipur	137186	704637	65458	907281
Meghalaya	26334	472605	18398	517337
Mizoram	5595	45069	5280	55944
Nagaland	15283	128991	73983	218257
Orissa	1385753	11770178	498793	13654724
Pondicherry	10072	46791	41926	98789
Punjab	5577336	1252804	1579255	8409395
Rajasthan	6343728	10846838	74263	17264829
Sikkim	3088	140399	42986	186473
Tamil Nadu	3128258	8208898	1144208	12481364
Tripura	16418	766109	61479	844006

Fig. 8 – Visual representation of multidimensional (OLAP) cube.

6. On-line Analytical Processing

A multidimensional OLAP cube provides on-line decision support system. Multidimensional cube is a structure that organizes data on the basis of dimensions. Data access perspective from user's point of view is to be modeled from dimensional tables. Hierarchies of dimensions represent logical flow and aggregations of the data from bottom to top and dis-aggregation from top to bottom. Building models for developing multidimensional cube mainly depends on users requirements. These cubes are posted on web in a ready to browse form for OLAP (Kimball, 1996; Craig et al., 1999; Scalzo, 2003; Kambayashi et al., 2004) analysis by the decision makers and can be accessed from anywhere with the help of web browser. Fig. 8 shows livestock population census multidimensional cube accessed through Internet browser for OLAP. In this cube hierarchies are All States, All Species and All Years. All States has state names as a top level and district as bottom level of data flow hierarchy. All Species has top level as species name, second level as sex, third level as age group and bottom level as working categories of animals. All Years has only one level, i.e. years. Similarly, Measures has only count (livestock population) information. All States, All Species and All Years were modeled through dimensional builds, whereas Measures are associated with fact builds.

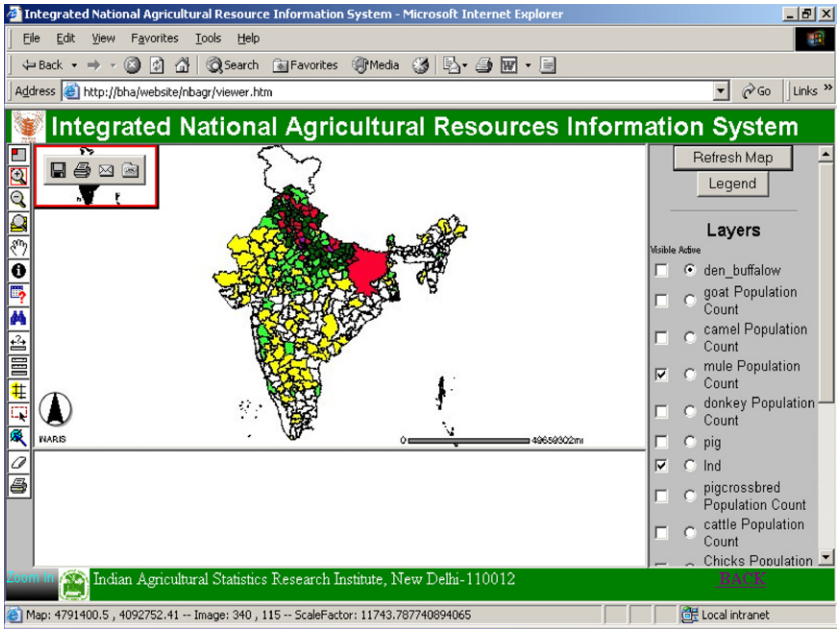


Fig. 9 – User front-end of web GIS from animal data mart.

This on-line system has drag and drop option for creation of nested tables, drill up and drill down functionalities based on hierarchies of various dimensions. System has simple calculation options on tabular data, hide and show option to hide certain undesirable rows or columns to be displayed on the screen, graphical representation options such as line, bar, charts, three-dimensional graphs, etc. through single click of a button, aggregations, segregations, slicing and dicing options are available with out any additional requirements for data exploration tools.

The zero suppression option suppresses all rows/columns with no information. It also has 80/20 options, which display major rows/columns contributing to 80% of grand total of respective row/column. This may be used to identify major contributor members of a particular level in dimensional hierarchy to total. It has option of exceptional highlights in which well-defined exceptions in the tabular information can be highlighted with different text and cell colors. The history of a particular session of analysis is recorded and can be seen whenever required by single click of a button. This will help the user to look into list of operations performed by him/her in a particular session. The find and search options are available for finding a particular information in tabular data of a cube. It also has help option to assist user in working with this decision support system. The most important option in this is to export visible data/graphs to number of commonly used file formats.

Apart from this option, it has functionalities of disseminating information through dynamic reports. In dynamic reports, a query is fired through web browser to the data warehouse database and pre-formatted reports are generated on-line in PDF/HTML format

This decision support system also has functionality of on-line spatial analysis of information in the data mart. In this case user has access to web based Geographic Information System (GIS). It has all simple and routine functionalities including layer analysis, spatial querying functionalities, etc. (Fig. 9).

In order to access this system, a user need not require to install any GIS software. The authenticated user would access this system using simple Internet browsers. This system helps in providing access to GIS technology to general user and decision makers with no extra cost and efforts.

7. User evaluation

Animal data mart was developed under National Agricultural Technology Project as a Mission Mode sub-project "Integrated National Agricultural Resources Information System (INARIS)". This project was implemented keeping in view the requirements of three broad groups of users, i.e. (i) research managers (ii) research scientists and (iii) general users. Depending on information requirement of these different groups of users, information is delivered in different formats using various reporting tools. For example, the OLAP and GIS system is only accessible to the research managers through proper authentication, as some of the contents are sensitive in nature. The information delivery to the research scientists are in the form of dynamic reports and information systems. However, 13 subject wise information systems have been developed to deliver basic information to the general users. Therefore, to evaluate the satisfaction level of various users, on-line summing technique proposed by Chen et al. (2000) was used. This on-line survey consisted 16 items which were measured in a seven point Likert type scale. Total 59 subjects were identified for responses out of which 42.4% responded. Majority of users (68%) have been using this data warehouse since more then 2 years. Also, most of the users (87.50%) use this data warehouse frequently. According to users, information accuracy level is high (Cronbach's alpha (α) = 0.944) and it provided high satisfaction level to the users (α = 0.906). Since its deployment, this data warehouse receives average 500 queries per months including hits from general users (Nilakanta et al., 2008).

8. Conclusion

Development and implementation OLAP system in the field of agriculture in India is a challenging task. Number of factors influenced the development process. The most important factor is diversification and complexity of this sector in the country. Therefore, number of organizations and departments were involved in collection and compilation of agricultural statistics. As a consequence of this, there are problems of enforcing uniformity and standardization of various concepts and definitions in this data collection process. In order to develop a OLAP system, integration of these information through designing data marts has number of challenges for data modeler, such as integration of information collected following different definitions of year, dimensions with unbalanced and un-covering hierarchies, differences in the grain levels, etc. Second important aspect is data quality and data consistency. In order to ensure the quality and consistency of data stored in data warehouse, it is important to apply quality and consistency checks while moving data from source system to staging area. Generally, OLAP systems are developed for research managers, therefore these systems should be designed with highly flexible and in a simple manner, so that most of the user requirements are fulfilled. OLAP based on this data mart has all these features including GIS analytical functionalities which makes it very powerful tool for decision making. This article provides guidelines for design and development of similar complex data marts in agricultural sector, particularly in the field of livestock management.

Acknowledgements

This article is based on data mart from National Agricultural Technology Project (NATP) Mission Mode sub-project "Integrated National Agricultural Resources Information System (INARIS)". The financial assistance received from PIU NATP for this project is duly acknowledged. We are also grateful to Dr. S.D. Sharma, Director, Indian Agricultural Statistics Research Institute, New Delhi for providing constant guidance in development of this data mart. We are thankful to Dr. P.K. Vij, National Bureau of Animal Genetic Resources (NBAGR), Karnal (India) and his team for their association in providing source databases for this data mart.

REFERENCES

Abdullah, A., Brobst, S., Umer, M., Khan, M.F., 2004. The case for an agri data warehouse: enabling analytical exploration of integrated agricultural data. In: Proceedings of the IASTED International Conference on Databases and Applications (DBA 2004), Innsbruck, Austria.

Agosta, L., December 13, 2001. Top 10 Data Warehousing Challenges. Forrester Research.

Bauer, A., Hummer, W., Lehner, W., 2000. An alternative relational OLAP modeling approach. In: Proceedings of the Second International Conference on Data Warehouse and Knowledge Discovery, pp. 189–198.

Bieber, M., 1998. Data Warehousing in Government, DM Rev.

Bonifati, A., Cattaneo, F., Ceri, S., Fuggetta, A., Paraposchi, S., 2001. Designing data marts for data warehouse. ACM Trans. Software Eng. Meth. 10 (4), 452–483.

Cabibbo, L., Torlone, R., 2000. The design and development of logical system for OLAP. In: Proceedings of the Second International Conference on Data Warehouse and Knowledge Discovery, pp. 1–10.

Chen, R., Chen, C., Cheng, C., 2003. A web-based ERP data mining system for decision-making. Int. J. Comp. Appl. Technol. 17 (3), 156–158.

Chen, L.D., Soliman, K.S., Mao, E., Frolick, M.N., 2000. Measuring users satisfaction with data warehouses: an exploratory study. Info. Manage. 37, 103–110.

Craig, R.S., Vivona, J.A., Bercovitch, D., 1999. Microsoft Technologies. John Wiley & Sons, New York.

Gupta, A., Mumick, I.S., 1995. Maintenance of materialized views: problems, techniques, and applications. Data Eng. Bull. 18 (2).

Hahn, K., Sapia, C., Blaschka, M., 2000. Automatically generating OLAP schemata from conceptual graphic models. In: Proceedings of Ninth International Conference on Information and Knowledge Management and Third ACM International Workshop on Data Warehousing and OLAP, pp. 9–16.

Harper, F.M., 2004. Data warehousing and the organization of governmental databases. In: Digital Government: Principles and Best Practices. IGI Publishing, pp. 236–247.

Hoffer, J.A., Prescott, M.B., McFadden, F.R., 2005. Modern Data Base Management. Pearson Education, Inc., Upper Saddle River, New Jersey.

Inmon, W., 1995. What is a Data Warehouse? PRISM Tech. Top. 1 (1).

Inmon, W., 2002. Building the Data Warehouse. John Wiley & Sons, Inc.

Jagadish, H., Lakshmanan, L., Srivastava, D., 1999. What can hierarchies do for data warehouse? In: Proceedings of the 25th Very Large Databases Conference, pp. 530–541.

Kambayashi, Y., Kumar, V., Mohania, M., Samtania, S., 2004. Recent advances and research problems in data warehouse. Lecture Notes in Computer Science. 1552, pp. 81–92.

Kelly, S., 1997. Data Warehousing in Action. John Wiley & Sons, Inc.

Kimball, R., 1996. The Data Warehouse Toolkit: Practical Technique for Building Data Warehouses. John Wiley & Sons, New York.

Kimball, R., 1998. The Data Warehouse Lifecycle Tool Kit. John Wiley & Sons, New York.

Kimball, R., Ross, M., 2002. The Data Warehousing Toolkit. John Wiley & Sons, New York.

Lehner, W., Albrecht, J., Wedekind, H., 1998. Normal forms of multidimensional databases. In: Proceedings of the 10th International Conference on Scientific and Statistical Database Management, pp. 63–72.

Malinowski, E., Zimanyi, E., 2004. Hierarchies in a multidimensional model: from conceptual modeling to logical representation. Data Knowledge Eng. 59 (2), 348–377.

Nilakanta, S., Scheibe, K., Rai, A., 2008. Dimensional issues in agricultural data warehouse designs. Comput. Electron. Agric. 60 (2), 263–278.

O'Neil, P., Graefe, G., 1995. Multi-table joins through bitmapped join indices. In: Proceedings of the SIGMOD Conference.

O'Neil, P., Quass, D., 1997. Improved Query Performance with Variant Indices. In: Proceedings of the SIGMOD Conference.

Pedersen, T., Jensen, C., Dyreson, C., 2001. A foundation for capturing and querying complex multidimensional data. Inform. Syst. 26 (5), 383–423.

Rai, A., Dubey, V., Chaturvedi, K.K., Malhotra, P.K., 2007. Issues of design and development of agricultural data warehouse in India. CSI Commun. 31 (1), 43–51.

Scalzo, B., 2003. Oracle DBA Guide to Data Warehousing and Star Schema. Prentice Hall, New York.

Whiting, R., 2003. The Data Warehouse Advantage, Information Week 949, pp 63–66.

Yost, M., 2000. Data warehousing and decision support at the national agricultural statistics service. Soc. Sci. Comput. Rev. 18 (4), 434–441.