Approach and development of data warehousing for sorghum multi locational database

P. Mukesh¹, Sameen S. Fathima², Vasumathi, D³ and Kalaisekar⁴

¹,⁴Indian Institute Millets Research (IIMR), Rajendranagar, Hyderabad
²College of Engineering, Osmania University, Hyderabad
³College of Engineering, JNTU, Hyderabad

Abstract - **Sorghum bicolor (L.) Moench** is one of the major crops for food grains under impoverished conditions in the semi arid tropics and is the fifth most important staple food crop for more than 500 million people in about 30 countries (Kelley et al. 1993) for food, feed, fodder and fuel. Sorghum is a C4 plant to be eco-friendly in role of climate change. To improve sorghum yield and quality there has been huge amount of multiplication data is generated every year in the sorghum research system which formulates thousands of data sets to formulate large data base which lead into a data warehousing for using data base for make decisions. To develop data warehouse for repository of integrated information, available for queries and analysis. This makes it much easier and more efficient to run queries over data that originally came from different sources of agricultural disciplines.

**key words:** Data warehousing, Sorghum bicolor (L.) Moench, sorghum data base, OLAP.

I. INTRODUCTION

Data mining and data warehousing are increasingly becoming popular among professional users. A data ware house is a decision-support environment that leverages data stored in different sources, organizing it and delivering it to decision makers across the enterprise, regardless of their platform or technical skill level. Data warehouse systems are valuable tools in today’s competitive and fast evolving world. A data ware house is a new approach to enterprise wide computing at strategic or architectural level and provide a central repository for large amounts of diverse and valuable information. Data ware house is different from data base that is supposed to be a place where data get stored so that applications can access and share it easily and data ware house is a different kind. A large amount of sorghum data source data ware house is a abnormal data base.

Data ware house, is a single, complete and consistent store of data obtained from a variety of different sources made available to end users in a what they can understand and use in a business context. A process of transforming data into information and making it available to users in a timely enough manner to make a difference.

A data ware house is of course a data base, but it contains summarized information. To access information by sending a query to ware house retrieving data and information and data base quires are very complex. Data ware house is a read only data as a place where we store many different things for the sake of convenience. The data ware house is a subject oriented, integrated, time varying, non-volatile collection of data in support of the management’s decision making process (WH Inmon 1993). To understand the data warehouse data model consider the way the statistical tables are traditionally represented.

2. Methodology

Of sorghum crop thousands of data sets are available for the production and pest etc and form traditional
data base for transaction to select various discipline wise traits.

The data warehouse of sorghum have the distinguishing characteristics/ traits which that they are mainly intended for decision support applications. Applications that data warehouse supports are: **OLAP** (Online Analytical Processing) is a term used to describe the analysis of complex data from the data warehouse. **DSS** (Decision Support Systems) also known as EIS (Executive Information Systems) supports organization’s leading decision makers for making complex and important decisions. **Data Mining** is used for knowledge discovery, the process of searching data for unanticipated new knowledge.

Data mining in that dimension modeling is special technique for structuring (numeric measures) data around the research concepts which describes entities and relationships.

The available core areas are Plant breeding, Entomology, Agronomy, Pathology and Extension but for the ware house the plant breeding only has taken the multilocations coordinated data for the development of data ware house..

**Plant Breeding: Data structures and organized form**

**Trial:**

**Kharif trials data :**
- Advanced Hybrid trial (AHT)
- Advanced Varietal Trial (AVT)
- Initial Hybrid Trial (IHT)
- Initial Varietal Trial (IVT)
- Advanced Varietal Trial - Late Kharif (AVT-LK)
- Advanced Hybrid Trial - Late Kharif (AHT-LK)

**Rabi trials Data :**
- Advanced Varietal & Hybrid Trial - Deep Soil
  - Initial Hybrid Trial-Deep Soil
  - Initial Varietal Trial-Deep Soil

Initial Varietal & Hybrid Trial - Shallow Soil
- Parental line trial

**Location:** say 11, 12, 13 ……..Ln


**Genotype/ Entry:**
- Genotype name 1
- Genotype name 1
  ……
- Genotype name n

Dimension hierarchy helps us in view of the data in multidimensional data in several views of data representations. Conceptually multi dimensional data can be viewed as lattice cuboids.

There are four major processes that contribute to a data warehouse:

- Extract and load the data.
- Cleaning and transforming the data.
- Backup and archive the data.
- Managing queries and directing them to the appropriate data sources.

This is probably the first attempt of data warehousing of agricultural resources in the world. This will provide systematic and periodic information to research scientists, planners, decision makers and...
developmental agencies in the form of On-line Analytical Processing (OLAP) decision support system. The above project was in collaboration with 13 other ICAR institutions, namely NBSSLUP Nagpur (for soil resources), CRIDA Hyderabad (for agro-meteorology), PDCSR Modipuram (for crops and cropping systems), NBAGR Karnal (for livestock resources), NBFGR Lucknow (for fish resources), NBPGR New Delhi (for plant genetic resources), NCAP New Delhi (for socio-economic resources), CIAE Bhopal (for agricultural implements and machinery), CPCRI Kasargod (for plantation crops), IISR Calicut (for spices crops), ICAR Research Complex for Eastern Region Patna (for water resources), NRCAF Jhansi (for agro forestry) and IIHR Bangalore (for horticultural crops).

In all 59 databases on agricultural technologies generated by council, research projects in operation and the related agricultural statistics from published sources at least from the year 1990 onwards at the district level have been integrated into this information system.

The system is currently under development phase with subject-wise data marts being created, multi-dimensional data cubes being developed for publishing on Internet/Intranet and the validation checks being implemented. The data warehouse system has been developed and the data mining process on the developed data warehouse would be taken up to provide decision support to the research managers.

The above system has been developed keeping in view the three groups of users i.e. (1) research managers and planners (2) research scientists and (3) general users. The information of this data warehouse will be available to the user in the form of decision support system in which the all the flexibility of the presentation of the information, its online analysis including graphic is inbuilt in to the system. The system also provides the facility of spatial analysis of the data through web using functionalities of Geographic Information System (GIS). Apart from this, the subject wise information system has been developed for the general users. The user of this system has the access of subject wise dynamic reports through web. The facilities of data mining and generation of ad-hoc querying will also be extended to limited users. Therefore, the dissemination of information from this data warehouse for different categories users will be through web browser with proper authentication of the users. The web site of the project is already launched (www.inaris.gen.in) and the multidimensional cubes, dynamic reports, GIS maps and some of the information systems are already available to the users.

A data warehouse database contains the above data of multi location that is organized and stored specifically in a specific format for direct user queries and reports. It differs from an OLTP database in the sense that it is designed primarily for reads not writes. Usually handles queries of an ad-hoc nature supports more complex and demanding transactions contains logically organized data in multiple dimensions.

Fig2: Data in flow and data out flow including general functions/tasks

Dimensional modeling is a term used to refer a set of data modeling techniques that have gained

www.ijaegt.com
popularity and acceptance for data warehouse implementation. Dimensional modeling is one of the key techniques in data warehousing. Two types of tables are used in dimensional modeling: Fact tables and dimensional tables. A schema is a fact table plus its related dimensional table.

Applications that data warehouse supports are OLAP (Online Analytical Processing) are a term used to describe the analysis of complex data from the data warehouse. Data Warehouse for the multi location data where cleaning the data by correcting misspellings, resolving domain conflicts, dealing with missing data elements, and parsing into standard formats. Purging selected fields from the legacy data that are not useful for data warehouse. Combining data sources of multi location by matching exactly on key values or by performing fuzzy matches on non-key attributes. Creating surrogate keys for each dimension record in order to avoid dependency on legacy defined keys, where the surrogate key generation process enforces referential integrity between the dimension tables and fact tables. Building the aggregates for boosting the performance of common queries.

Data warehouse database contains the data that is organized and stored specifically for direct user queries and reports. It differs from an OLTP database in the sense that it is designed primarily for reads not writes. An OLAP application is a system designed for few but complex (read only) request. An OLTP application is a system designed for many but simple concurrent (and updating) requests. In design concern dimensional modeling is a term used to refer a set of data modeling techniques that have gained popularity and acceptance for data warehouse implementation. Dimensional modeling is one of the key techniques in data warehousing. Two types of tables are used in dimensional modeling: Fact tables and dimensional tables.

The system is currently under development phase with subject-wise data marts being created, multi-dimensional data cubes being developed for publishing on Internet/Intranet and the validation checks being implemented. The data warehouse system has been developed and the some of data mining process on the developed data warehouse data base would be taken up to provide decision support to the research managers.

Conclusion:

The sorghum crop data ware house would be applied to all other core areas other than plant breeding using many other advanced data ware house operational techniques for better results for strategic research plans for supporting in crop improvement. This project will be taken in clubbing with data mining technique application processes.

REFERENCES:

2. Kavita Pabreja; Rattan K. Datta, A data warehousing and data mining approach for analysis and forecast of cloudburst events using OLAP-based data hypercube, Birla Institute of Technology and Science, Vidya Vihar Campus, Pilani, Rajasthan, 333031, India. Gwalior Academy of Mathematical Sciences, A-9, USO Road, Shaheed Jit Singh Marg, New Delhi, 110067, India 2012 Vol.4, No.1


4. DSR, Annual Reports of AICSIP group meeting (AGM,12), Rajendra nagar, Hyderabad

