

Distr.
GENERAL

WP.13
31 March 2010

ENGLISH ONLY

**UNITED NATIONS ECONOMIC COMMISSION
FOR EUROPE (UNECE)
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE EUROPEAN
UNION (EUROSTAT)**

**ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE**

Meeting on the Management of Statistical Information Systems (MSIS 2010)
(Daejeon, Republic of Korea, 26-29 April 2010)

Topic (iii): Issues for Asian statistical organizations and ways to increase international cooperation

Setting up a National Warehouse of Official Statistics in India

Prepared by P.C. Mohanan, Central Statistical Office, India

I. Introduction

1. India has a highly decentralized statistical system both at the centre and in the provinces. Thus official statistics are generated and disseminated at various levels of administration. Large volumes of official statistical data consisting of both administrative statistics and statistics from surveys and censuses are available with the Government agencies. The Government of India in 1999 announced a National Data Dissemination Policy mandating the Ministry of Statistics 'to set up a data warehouse collecting data from various source agencies and integrate the data in to logical subject areas, store the data in a manner that is accessible and understandable and deliver the data through report writing and query tools'. The paper briefly explains the experience of the Ministry of Statistics in setting up the National Data Warehouse of Official Statistics.

Background

2. A statistical system consists of organisations performing statistical tasks to produce statistical data/products. The types of data produced by a statistical system consist of derived statistics like national accounts statistics, indicators like price index, data bases from census and surveys. These are made available to the users in printed reports, press announcements, website and in other stand alone electronic media. The Indian Statistical System is a highly decentralized one following the administrative structure mandated in the Indian constitution. The statistics on various subjects are produced by different line Ministries and other Government agencies both at the Centre and in the States. The statistical system at the Centre has the National Statistical Office playing a nodal and coordinating role among the various central agencies. In the provinces, the system is a replica of the system at the centre with a nodal agency called the Directorate of Economics and Statistics playing the role of NSO.

3. In this decentralized system, the data collection, processing, production, dissemination and archiving are performed at various points. In such a system, it becomes difficult to maintain standards in respect of metadata, data layout, data access policy besides the difficulties in identifying contact points.

II. Dissemination of official data

4. The major types of official data produced by the statistical system at the national level are:
 - (a) Socio-economic indicators;
 - (b) Macro aggregates like national accounts statistics;
 - (c) Periodical thematic reports of the Central Statistical Office;
 - (d) Periodic census and survey data;
 - (e) Administrative data produced from statutory returns.

In the past the system did not allow for dissemination of the basic data available with the government agencies. Official statistics were disseminated through publications and press announcements etc. With the availability of internet, most of the statistical agencies started making available the data in their websites hosted by the concerned agencies. There was very little attention paid to providing metadata and access to unit level basic data.

5. It was in this context that the Government of India announced the National Data Dissemination Policy. After the policy was announced the Ministry of Statistics, which was made the nodal agency, started disseminating the unit level data from its surveys like the National Sample Survey, Annual Survey of Industries and the Economic Census etc. This was well appreciated by the users and there was spurt in the research using these data.

6. As part of the data dissemination policy all the reports published by the Ministry of Statistics are made available to users free of cost. These reports are either regular publications or survey reports. The users are required to register their names through the web and then download the reports they select using the user name and password provided to them at the time of registration. On an average about 1500 users registers every month.

7. While the reports are disseminated freely, the dissemination of unit level data of national sample surveys, annual survey of industries and the economic census is done on request after payment of the requisite charges. The unit level data are provided on CDs with full documentation. These are voluminous data in text format. However using these large databases requires large computing resources besides a deep knowledge of the data generating mechanism followed in the surveys. Changes in the classification like the Industrial classifications, geographical domains covered etc also posed immense difficulties to the data users. It was therefore found essential to make use of the advances in IT and develop data repositories that will address the problems and make user friendly data warehouses providing full access to metadata and micro data. A pilot project for creation of a National Data warehouse was therefore taken up by the Ministry of Statistics.

III. Objective of the National Data Warehouse for Official Statistics

- To develop an integrated repository of current and historical data;
 - This repository over a period of time will also encompass data generated by various Central Ministries, State Governments and UTs;
 - To build a state-of-the-art data warehouse with online analytical processing (OLAP) capabilities, which includes providing a multi-dimensional and subject oriented view of the database;
 - To enable web-based access to the Data Warehouse.
8. For software support, COGNOS 8 Data Warehouse tools with the following Components were procured:

Decision Stream (ETL)
Framework Manager

Power Play Transformer
ReportNet with SDK

- Cognos 8 BI is a complete enterprise business intelligence software solution with integrated reporting, and analysis capabilities that let you effectively and efficiently query, analyze, and report on data.
- Cognos Data Manager is a conventional ETL (extract, transform and load) tool that has extended capabilities for supporting data warehousing environments and data marts that make the management of those data stores much simpler than would otherwise be the case.
- Framework provides the metadata model development environment for Cognos8. Using Framework Manager one can build the model of business presentation of data derived from one or more data source as per the requirement of business need.
- Transformer models help you structure data from various sources into multidimensional Power Cubes. One can use Transformer by assessing their data and deciding how will organize it to support the types of business analysis that are important. Transformer accepts data from Impromptu files (iqd), flat files, and most spreadsheet and database applications, with or without column headings, and compresses the data into a compact Power Cube.

9. The source data normally available are in ASCII (text) form along with supporting documents including schedule, record layout etc. These are initially loaded in the ORACLE IQG database tables. By using IBM Cognos8 Data Manager tool, ORACLE tables are accessed and used for necessary ETL (Extraction Transformation & loading) process to establish the Star Schema/Data Mart. Cognos Power Play transformer tool is used to develop multi-dimensional power cubes having scope of quick retrieval of data. These data Marts/power cubes being the component of National Data Warehouse are transformed into various data models and published as packages in the Data Warehouse Server using Cognos Framework Manager tool. Cognos Report net tool is finally used by the Report authors to design and develop various web enables reports.

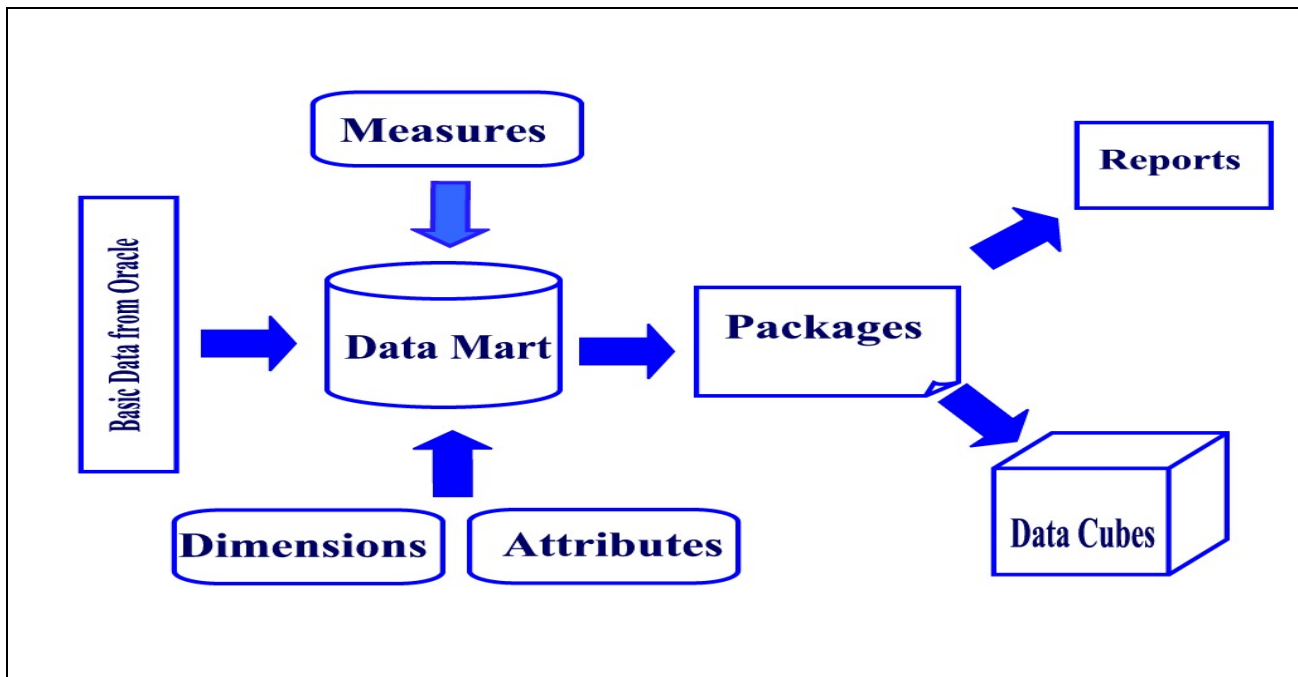


Figure 1: Data Warehouse Model

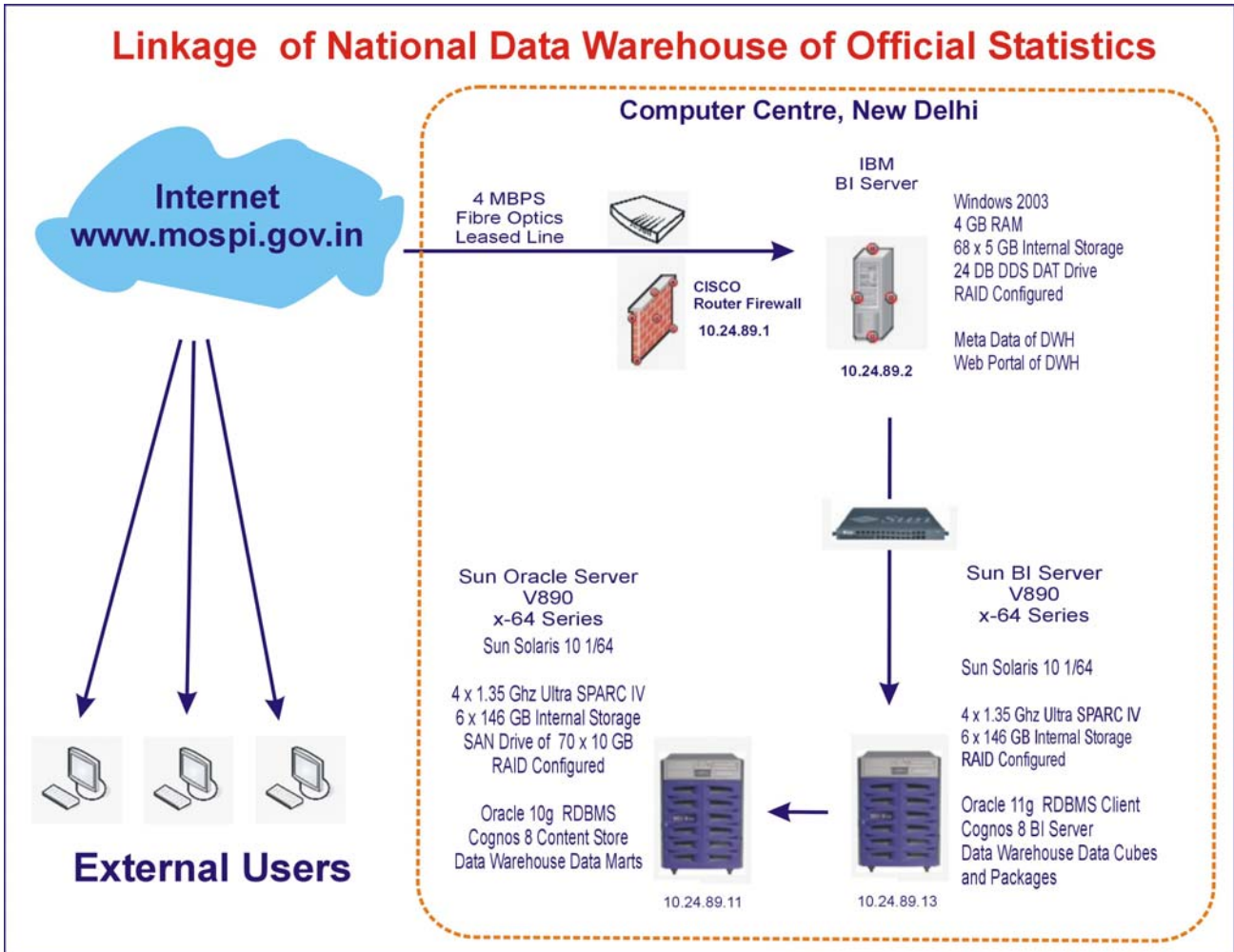


Figure 2: Accessing the data warehouse

A. Main features

10. The main features of the data warehouse are:

- (a) User friendly interface
- (b) Dynamic /pull down menus
- (c) Search based Report
- (d) Secured web access
- (e) Bulletin board
- (f) Complete Metadata
- (g) Parametric and Dynamic report in exportable format (Excel, text and PDF formats)
- (h) Power Cubes

B. Access to the Data Warehouse

11. Currently access is provided at two levels. At level 1, the users have free and complete access to all metadata, published reports and pre-formatted tables from different surveys and censuses. At level 2, restricted usage is allowed to generate dynamic tables, use of power cubes etc from the data available in the warehouse.