Designing the Data Warehouse for Disaster Management: A Case Study of Korean Government

Young-Jai Lee
Dongguk University
Seoul, Korea
yjlee@dongguk.edu

Keywords: Disaster, Data warehouse, Data model

Abstract

Disaster management organization of korean government has faced with a big problem about an immediate response in natural disasters such as flood, snow storm, and hurricane etc. The problem is not to provide timely information from many operational databases to decision makers as disasters were happened. The operational databases were mainly built according to preparation, prevention, response, and recovery stage and have stored the results whenever disasters happened. Thus, it is difficult to retrieve lots of analyzed information that decision makers need to decide a policy such as resource allocation etc.

In order to solve the problem, the organization makes a decision to build a data warehouse so that can produce a variety of information. This paper illustrates a framework for building disaster data warehouse and a data modeling which is the most important of the process of building it. The framework consists of three parts: data collection, data storage, and data analysis. And the data model examples a star schema in terms of a disaster including people, facilities, damaged assets, and weather etc.

Basic Framework

Data warehousing is a process, not a product, for assembling and managing data from various sources for the purpose of gaining a single, detailed view of part or all of a business (Gardner, 1998). Figure 1 shows a basic framework of a data warehouse.

A data warehouse contains summary, historical, and detail data to support tactical and strategic decision making. Data is extracted from operational data sources, transformed, cleansed, reconciled, aggregated, and summarized in preparation for warehouse processing. Data might be denormalized to achieve acceptable performance for queries. This data is extracted from among several sources: internal and external to the organization.

Data marts are often connected to a centralized warehouse in a three-tier configuration in which clients are connected to specific data marts that draw their data from a data warehouse. Metadata is popularly defined as data about data. In data warehousing, metadata refers to anything that defines a data warehouse object, such as a table, a column, a query, a report, a business rule, or a transformation algorithm (Inmon, 1994).

To support OLAP, many users perform multidimensional analysis of warehouse data mart data. The market has been a proliferation of multidimensional database tools representing two approaches to the storage and management of their supporting data bases: multidimensional OLAP (MOLAP) and relational OLAP (ROLAP) (Bontempo, 1998). MOLAP uses specialized data structures designed for performance; ROLAP enables greater scalability and full-function DBMS support, since it relies on conventional RDBMSs for storage and management of warehouse data, usually organized as a star schema.

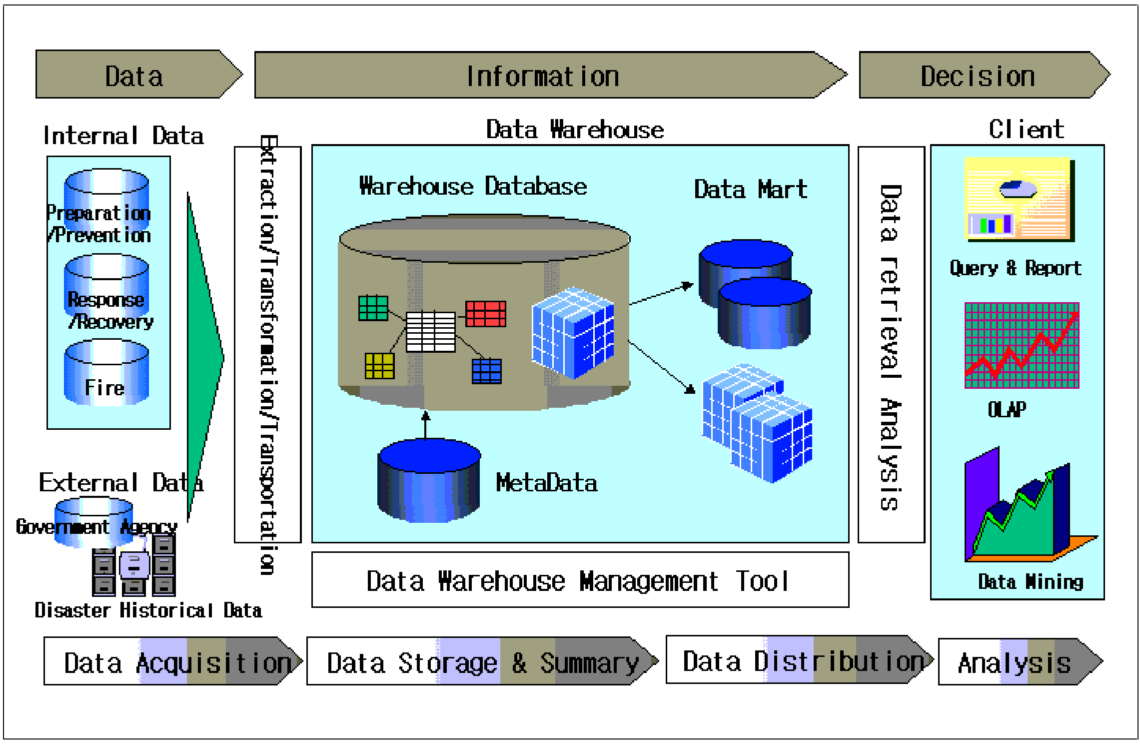

Figure 1. Data Warehouse Framework

Data Warehouse Architecture of National Disaster Management

Figure 2 shows NDM warehouse architecture. The architecture specified an RDBMS such as Sybase ASIQ for the storage and management of warehouse data. Data is extracted from NDM information system and Government agency system using a data extraction tool such as Power Stage. Summarizing data is the processing of raw input data or detail data for more compact storage in a form useful for analysis in the particular application recording the data.

Data is analyzed by OLAP tool (e.g. business objects) and statistical analysis tool (e.g. SPSS) to produce information to support decision-making. The OLAP tool includes web, report, manager, designer, and user module.

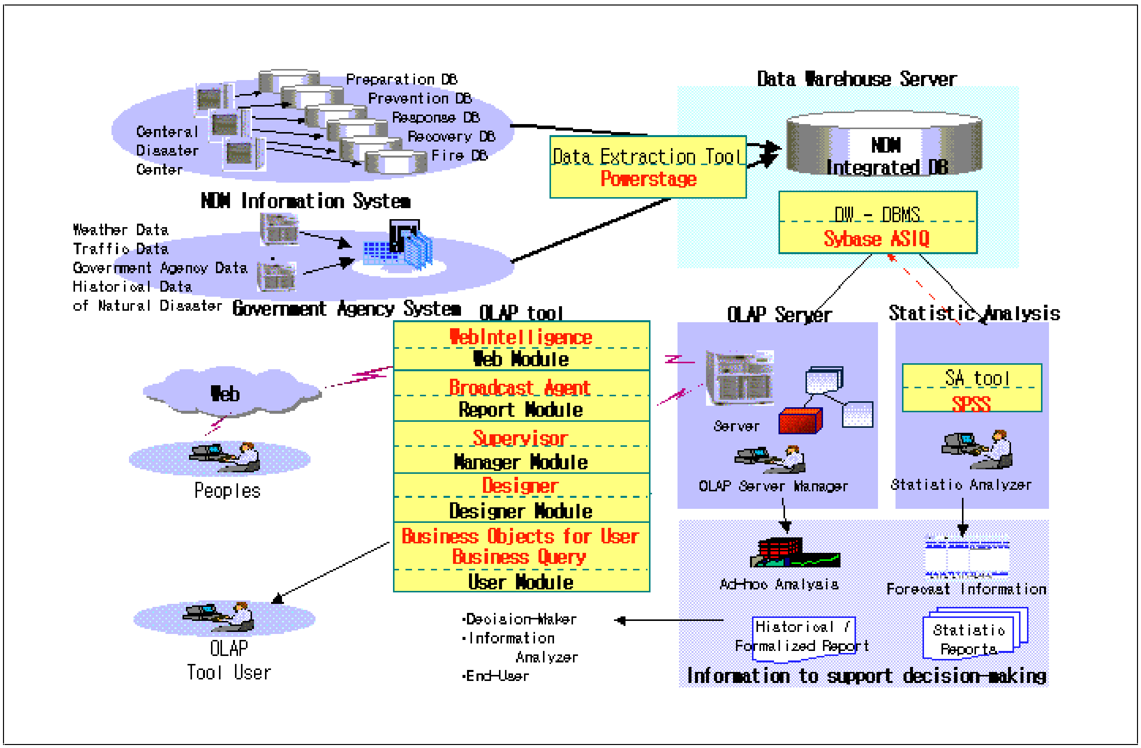

Figure 2. National Disaster Management Warehouse Architecture

Analytical Information from NDM Data Warehouse

Data enters the data warehouse from among several sources. A data mart typically contains a narrower scope of data, characterized by a single subject, a single business function, or even a single application (Gardner, 1998). This topology as shown in Figure 3 exploits locality of reference to provide optional performance to the warehouse's data mart clients. Data mart categorized by disaster types or a single subject such as weather, resource, facility, and danger spot etc. Using OLAP, statistical package, and data mining tool, many different kinds of information such as resources, statistics, trend, and forecast etc., are produced from the data mart environment.

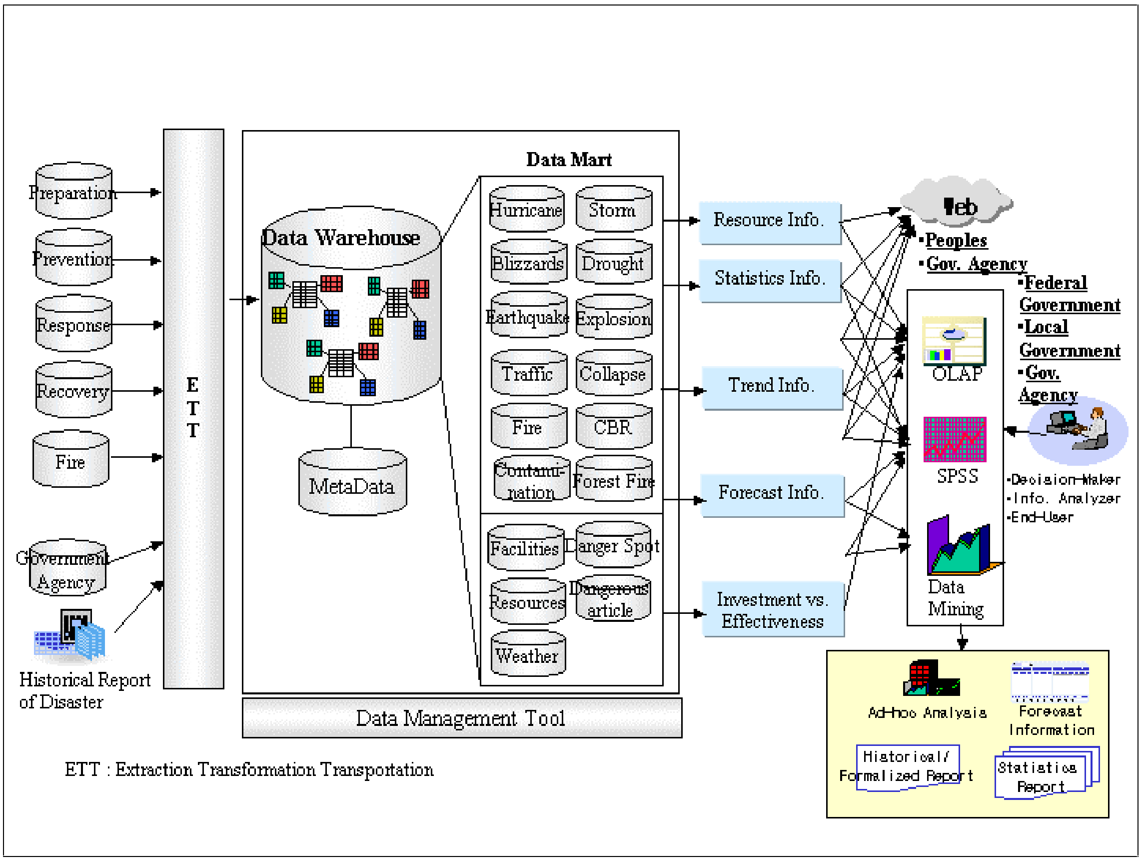

Figure 3. Disaster Related Information Types

The data warehouse environment enables the users to perform efficient multidimensional and accurate data analysis by providing an integrated environment that can be easily accessed through subjects on demand. As sown in Figure 4, resource information and damage information in terms of a type of disaster such as flood are provided to decision makers from subject/data-driven orientation of the data warehouse.

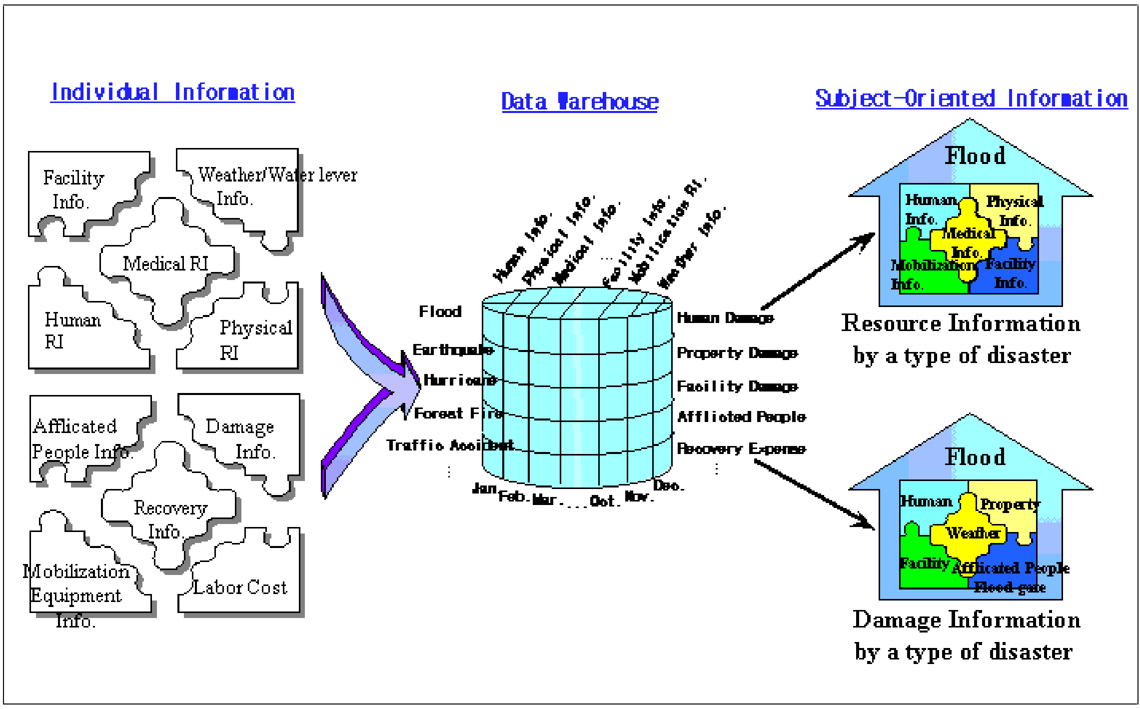

Figure 4. Integrated Environment and Access to Subject-oriented Information

Multidimensional Analysis

Perhaps the simplest user perspective of analysis is that of creating summaries of data. The decision-maker has looked at raw data in the data warehouse and has created a summary for management. Management finds summaries useful because it capsules data for them. Summarization presents data to management quickly and concisely.

However the conciseness of summarization is received, management typically wants to be able to “peel the onion back” and look at successively lower levels of detail. This peeling back of the onion is called “drill-down analysis” (Inmon, 1996) and is shown in Figure 5.

Subject-oriented, detailed flood data allows corporate users to drill down in the heart of their business operations, not only to find answers to specific questions but to then show how and why they got each answer.

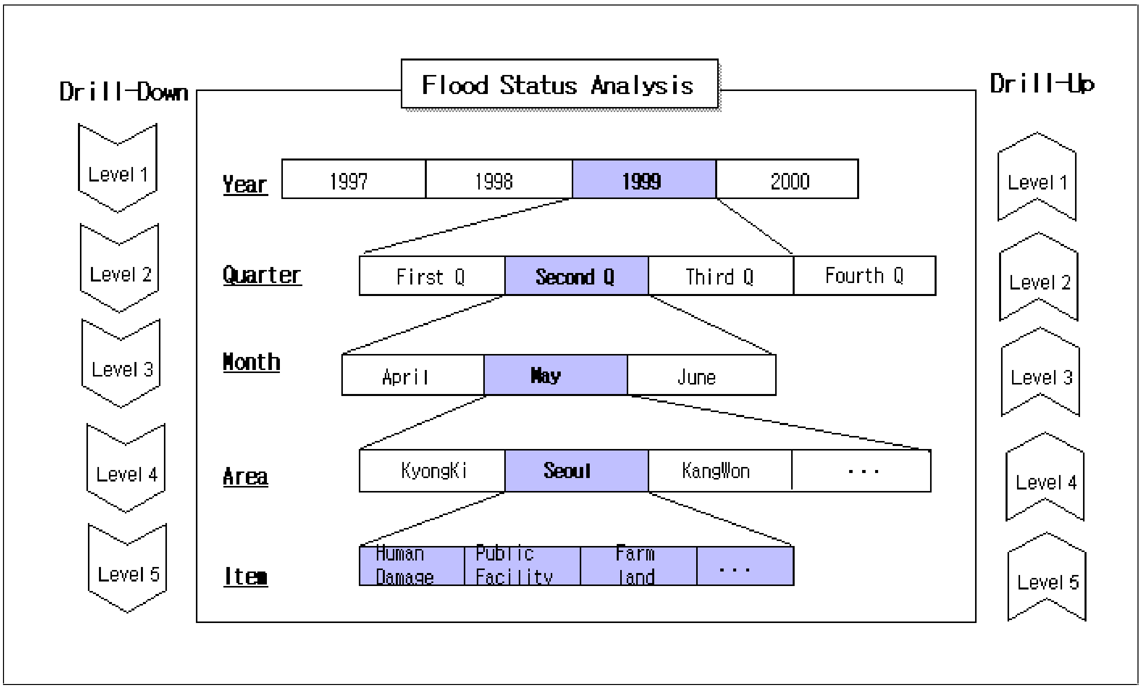

Figure 5. Multi-Dimensional Analysis

Data Modeling

At the heart of the relational OLAP technologies is dimensional modeling. This technique organizes information into two types of data structures: measures, or numerical data, which are stored in "fact" tables; and dimensions, which are stored in satellite tables and are joined to the fact tables (Matthias, 2000).

A star schema consists of one central fact table and several dimension tables. The measures of interest for OLAP are stored in the fact table (e.g. disaster information). For each dimension of the multidimensional model there exists a dimensional table (e.g. time code, disaster cause). This table stores the information about the dimensions.

Figure 6 is an example of a star schema. The table DISASTER INFORMATION at the center of the star is the fact table with the foreign keys date, town code, DM number, damage factor code, water system section code and disaster cause code to the corresponding dimension table, which carry the information on the hierarchical structure of the respective dimensions. Each dimension table consists of several attributes describing one dimension level. For example, the time code dimension is organized in a date, month, quarter, half year, half year name, season, season name, and year. The hierarchical structure of a dimension can then be used for drill-down and roll-up operations. Every hierarchy is represented by one or more attributes in the dimension table. Dimension tables have a denormalized structure.

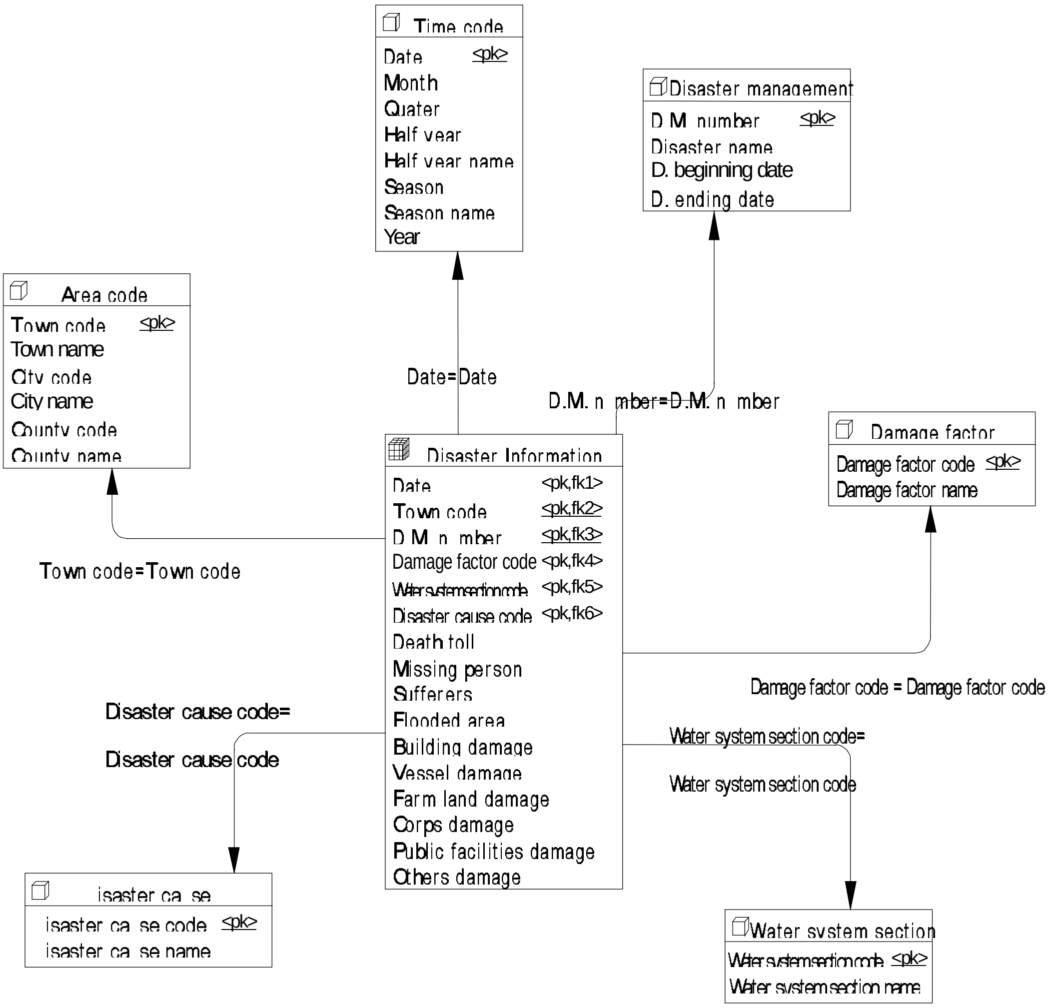

Figure 6. Disaster Information Star Model

Other examples are shown in Figure 7 and Figure 8. Figure 7 illustrates data model related with disaster recovery information. And Figure 8 depicts weather information star data model.

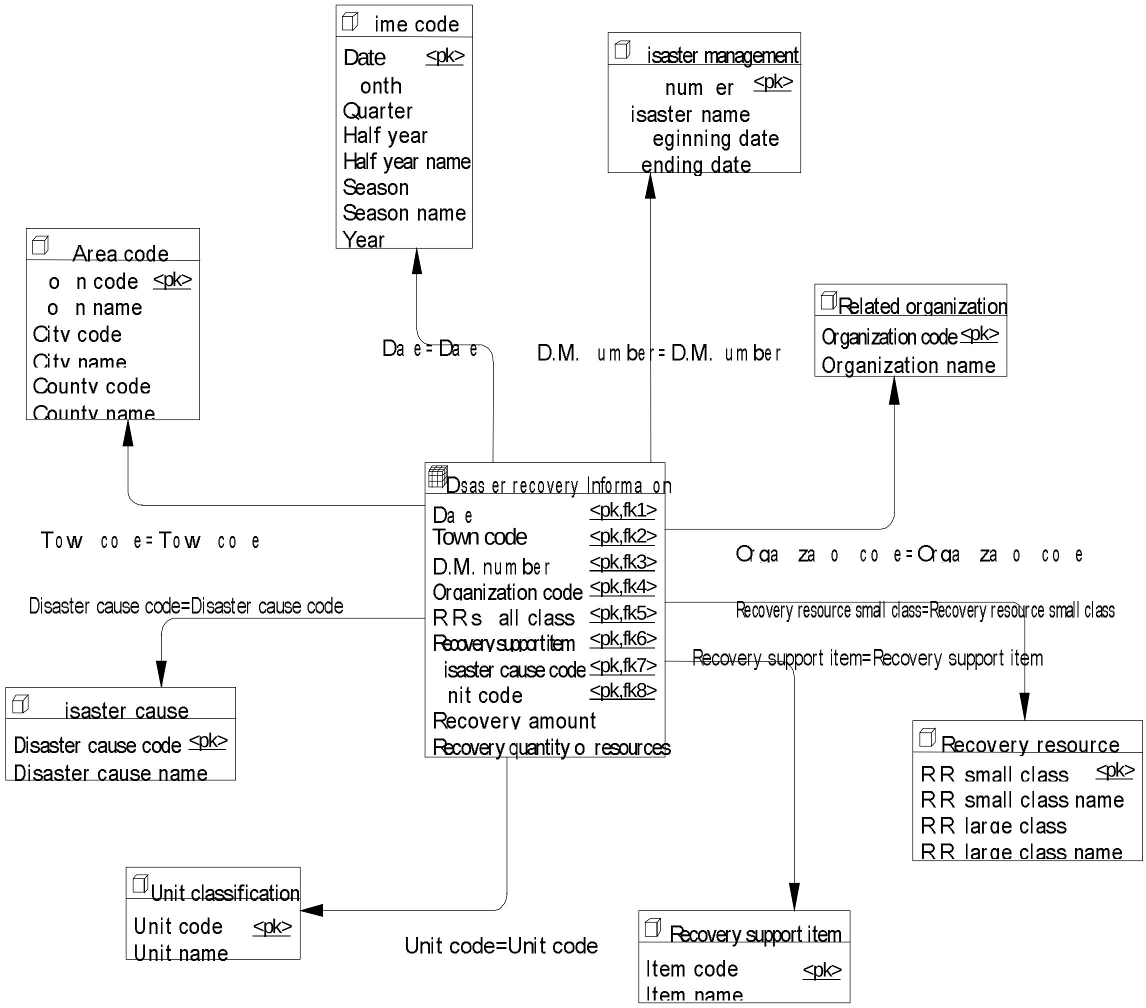

Figure 7. Disaster Recovery Information Star Model

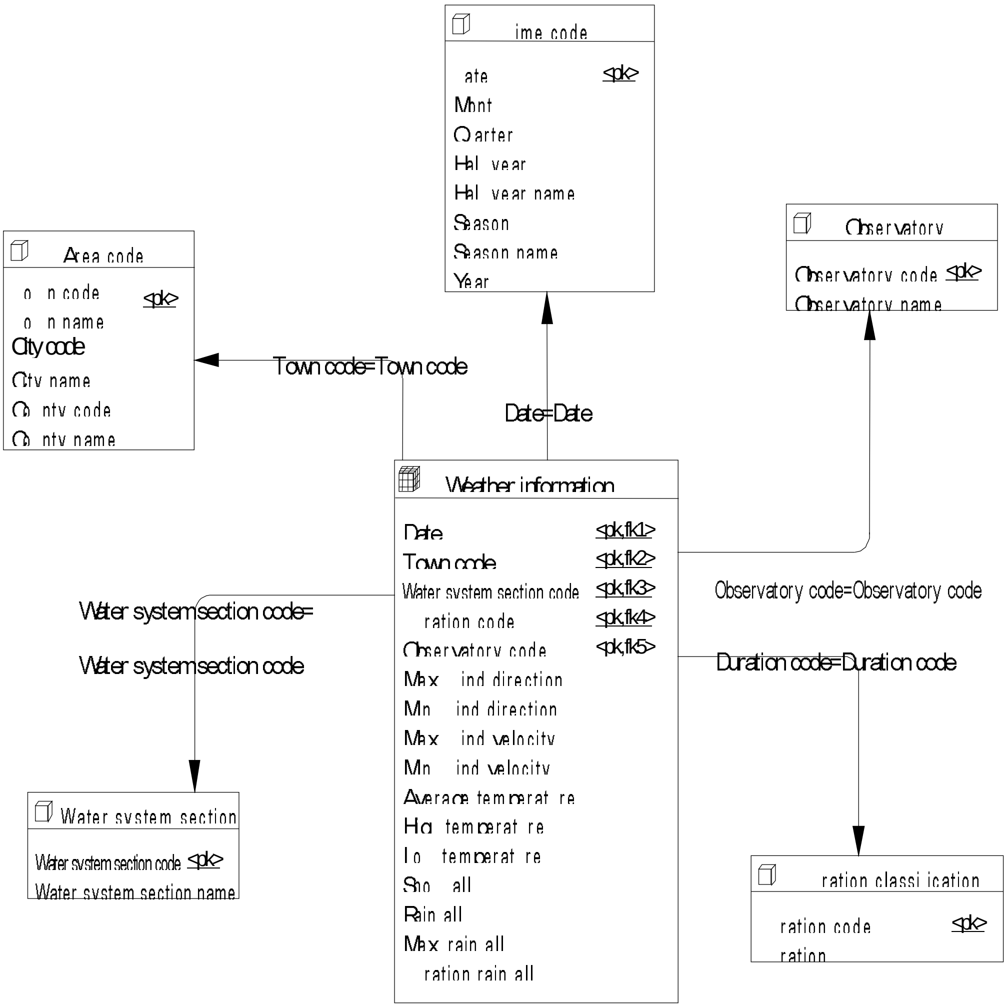

Figure 8. Weather Information Star Model

Expected Effects and Conclusion

The following effects by building national disaster management warehouse are expected.

1) systematic disaster response/recovery by forecasting potential disaster possibility and following damage through prompt decision-making information support.
2) Efficient preparedness and damage reduction through the management of disaster factors and by building a database about the results of past case data analysis in a pro-active manner.
3) Public service improvement
4) Efficient operation of workers related with disaster such as re-allocation and re-usability.
5) Efficient business process such as prompt decision-making and simple report generation.
6) Data integration with consistent data sharing for a different type of multi-dimensional analysis.
7) Various analytical techniques such as multi-dimensional, trend, and forecasting analysis.

Building a warehouse is a complex process requiring careful planning and alignment between the information technology department and business users. Especially, it is essential to recognize the importance of design features enabling tight integration of data warehouse components to simplify deployment and administration.

Reference

Charles Bontempo and George Zagelow, “The IBM Data Warehouse Architecture,” Communication of the ACM, vol 41 no 9, Sept. 1998.

Matthias Jarke, etc., Fundamentals of Data Warehouses, Springer, 2000.

National Disaster HQ, System Analysis and Design Report, Minister of Administration, Korean Government, 2000.

Stephen R. Gardner, “Building the Data Warehouse,” Communication of the ACM, vol 41 no 9, Sept. 1998.

W.H. Inmon and R.D. Hackathorn, Using the Data Warehouse, John Wiley & Sons, Inc., 1994.

W.H. Inmon, Building the Data Warehouse, 2nd Ed., John Wiley & Sons, Inc., 1996.