World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:4, No:11, 2010

Decision Support System Based on Data Warehouse

Yang Bao LuJing Zhang

***Abstract*—**Typical Intelligent Decision Support System is 4-based, its design composes of Data Warehouse, Online Analytical Processing, Data Mining and Decision Supporting based on models, which is called Decision Support System Based on Data Warehouse (DSSBDW). This way takes ETL,OLAP and DM as its implementing means, and integrates traditional model-driving DSS and data-driving DSS into a whole. For this kind of problem, this paper analyzes the DSSBDW architecture and DW model, and discusses the following key issues: ETL designing and Realization; metadata managing technology using XML; SQL implementing, optimizing performance, data mapping in OLAP; lastly, it illustrates the designing principle and method of DW in DSSBDW.

***Keywords*—**Decision Support System, Data Warehouse, Data Mining.

I. Decision Support System and Data Warehouse

A. Decision Support System (DSS)

Traditional DSS generally is consisted of three bases model, and IDSS (Intelligent Decision Support System) is consisted of four bases structure[1], based on three bases increasing the knowledge base system and its reasoning system, the IDSS is considered of an increase of expert system in the general DSS. Knowledge base is intelligent component in IDSS, used to simulate some smart activity in the human decision making process, that can make DSS supporting to decision makers been greatly enhanced.

B. Data Warehouse (DW)

B.1 Composition of the DW system

DW system contains three levels of architecture, as shown in figure 1. The three levels respectively: data sources, data storage and management , OLAP (On-Line Analytical Processing) and data mining tools.

B.2 Data organization structure of DW

The data organization style of DW contains three kinds:virtual database relations based on the storage and multidimensional databases. Data Warehouse, the data is divided into four categories: Early details, the details, light degree integrated, highly integrated level. Data Warehouse, there is an important data-metadata (metadata). The data storage environment, there are two main metadata: The first is the operational environment to the data warehouse and conversion of $data, including all source data of members, properties and the data warehouse in the transformation and the two million data for the OLAP and Data Warehouse building map.Figures

C. DSS Architecture based on DW

DSSBDW[2] (Decision Support System Based on Data Warehouse) compared with traditional DSS, it provided three kinds of decision tools:MDSS(Model DSS), DM, OLAP. The figure 2 can be seen, DSSBDW including DW, Model Analyzing, metadata, OLAP, user interface,etc.

II. The Key Technology of DW in DSS

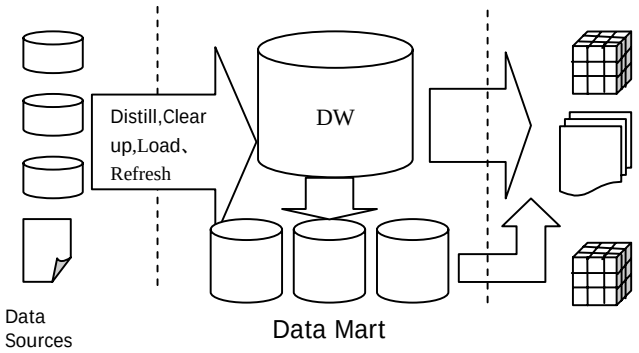


Fig. 1 Architecture of Data Warehouse

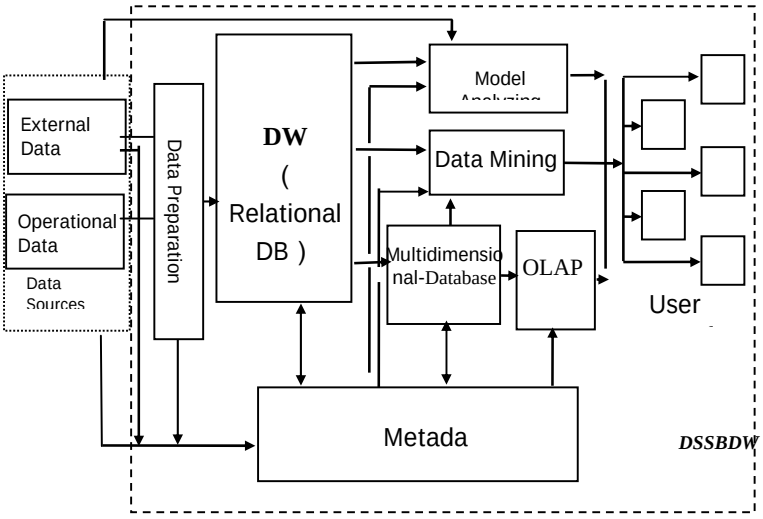


Fig. 2 Architecture of Decision Support System Based on Data Warehouse

A. Data Preparation

Data preparation, also known as data pretreating, is the basic premise on building DW. It majorly includes extraction, transformation, loading, commonly known as ETL (Extract, Transform, Load). Data preparation is the data entry for the whole DW, this link will complete the task of data acquisition, cleaning, certification, amalgamation and integration, loading, filing, data reproducing, reducing paradigm degree and summary.

B. Constructing metadata model

Open Science Index, Computer and Information Engineering Vol:4, No:11, 2010 publications.waset.org/10424/pdf

Open Science Index, Computer and Information Engineering Vol:4, No:11, 2010 publications.waset.org/10424/pdf

Metadata takes on extremely important role in the design, operation of DW, it describes the various objects of DW in all its aspects, which is the core of DW. For it,there is two major metadata standards to be used in the DW areas,: OIM (Open Information Model) Standards of MDC (Meta Data Coalition) and CWM (Common Warehouse Model, CWM) standards of OMG.Such as the figure3 is the composition of CWM Metamodel:

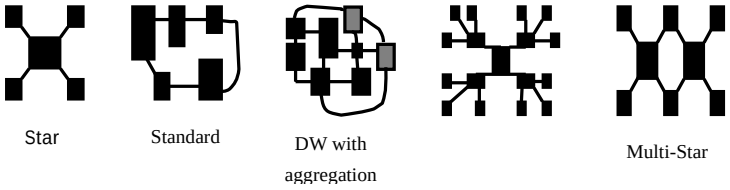


Fig. 5 Data Model of Data Warehouse

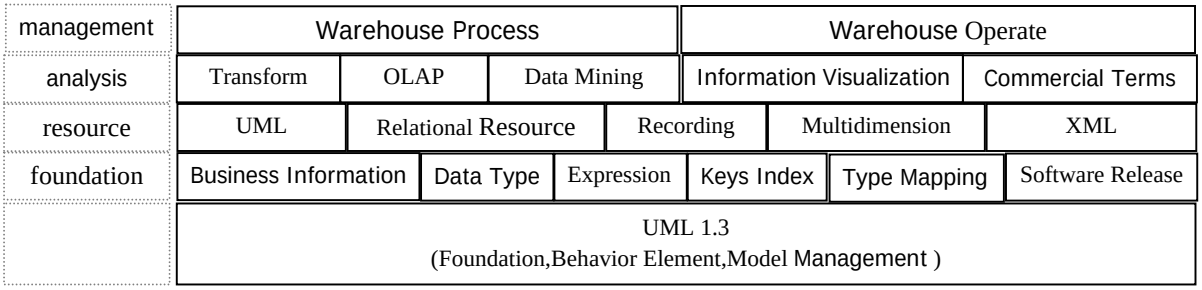


Fig. 3 Package Structure of CWM Metamodel

C. OLAP Operation

The OLAP operating mainly aims to online data access and analysis of specific problems, the object is to meet the need of specific query and reporting of decision support or multidimensional environment, focusing on the decision support of decision makers and senior managers, and so it is a powerful tool to analyze and make decision.It contains Slice,Dice,Rotate and Drill.

D. Data Mining Operation

Typical Data mining methods in accordance with the different apply mode contain:Related analysis, Time order mode,Cluster,Classification,Deviation testing and Forecasting models[3]. **It** is an orderly and complete process. fundamental processes of data mining with ETL, OLAP of DW have many coincide ways, shown in figure 4 below.

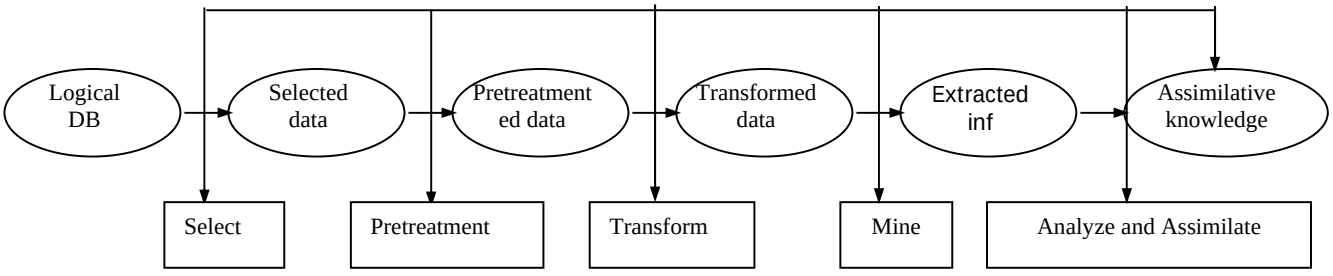


Fig. 4 Main Steps of Data Mining

E. Designing and Implementing of DW

E.1 Three levels of model of DW

Three levels of model of DW contain:Conception Model, Logical Model and Physical Model.In these,Logic model is the important link in the implementation of DW, There are five kinds of relative Modeling Technology commonly like figure 5 shown:

E.2 The main implementing methods and steps of DW

To highlight the development process with its uncertain demand , DW design method is described as CLDS methods,and SDLC on the contrary,the keystone is through adopting the External Data Access to complete DW modeling, data acquisition integrating, DW building , DSS application programming,system testing, requirement understanding, and then going trace back to the DW modeling. The implementing steps of DW contain:collect and analyzing business needs,establish the data models and the physical design of Dw , select DW technique and platform,define data sources,develope DSS application, strengthen management, take the active participation with end users,decompound the complex needs , prepare data, provide OLAP and data mining tools,etc.

III. The DSS design with DW technology

A. Architecture design of DSS

Open Science Index, Computer and Information Engineering Vol:4, No:11, 2010 publications.waset.org/10424/pdf

DSS design adopts the architecture shown as figure 6.Data

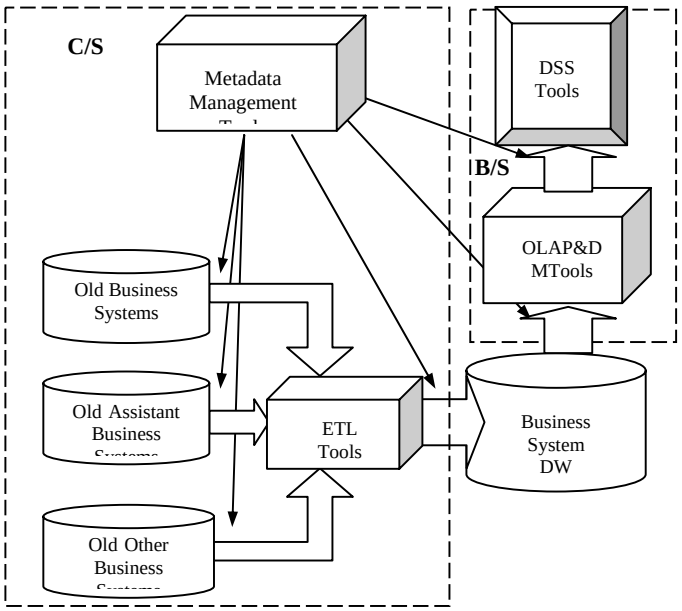


Fig. 6 Achitecture of Decision Support System

source is from the original operational system database and stored into the Data Warehouse Center storage through ETL tools. OLAP and DM tools provide multidimensional analyzing and mining means in DSS, and show to the final decision makers by performance tools.

B. The typical realizing scheme of System Architecture

DSS can be used "Three layers plus two layers" structure model, namely that ETL/system management module use Client/Server mode, OLAP module use Browser/Server mode. Its main design scheme in figure 7.

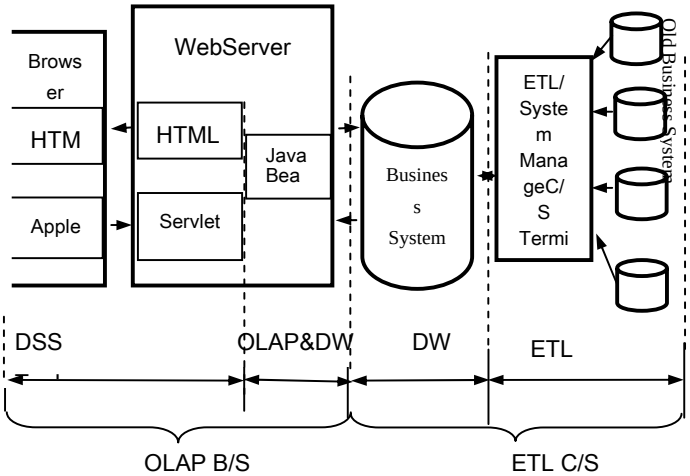


Fig. 7 Realizing Scheme of Decision Support System Based on Data arehouse

C. ETL Design

C.1 Logic flow of ETL

Operating steps in ETL are shown as Table I

TABLE I
CLASSIFICATION OF ETL OPERATION FLOW

Functional Categories	Processing Steps	Frequency
Process Management	1．set data information in data source	once
Process Management	2．set target data information	once
Process Management	3．data mapping on source data and target data	once
Process Management	4．data replication mode definition	once
Process Management	5．schedule ETL task	times
Data Acquisition	6．achieve source data	times
Data Transformation	7．transform between source data and target data based on mapping regulation	times
Data Loading	8．load data to target data	times
Process Management	9．revise execution definition	according to the needs

C.2 ETL elementary COMPONENTS

The ETL integrated components include: Data Access Component, Data Conversion Component, Data Loading Component, Process Management Component, Systems Management Component.

C.3 Data Conversion mapping relation

In DSS, the source field should be related to the property field of target data object, and be defined the length,data type of property field in the source field record, and so on. Targets objects and properties are the table and field in DW or Data In data objects and properties in transaction component. Rule information describe the corresponding data conversion regulation and transaction. Completed regulation description and the corresponding definition information shown as table II.

TABLE II
DEFINITIONS AND DESCRIPTIONS OF DATA MAPPING REGULATIONS

Regulation Description	Value Field
Source Field	
Location	computer/internet information
DataEntity name	database or other datasets
Connection Information	user/password
DataObject name	table or other data files
Field name	data fields or file fields
Length	integer
Data type	Integer/Char/Varchar/Float/Double,etc
DataObject type	Oracle8i/SQLServer/Excel/Acess/Txt,etc
Target Field	
Location	computer/internet information
DataEntity name	database or other datasets
Connection Information	user/password
DataObject name	table or other data files
Property name	object property or tavle fields
DataObject type	Oracle8i/SQLServer/Excel/Acess/Txt,etc
Default value	identical to this field type
Conversion function	MAX/MIN/COUNT/AVG/

C.4 ETL Realization

Data Pipeline is a data migration tool,it is used to migrate from a database to another database, these database can be the

Open Science Index, Computer and Information Engineering Vol:4, No:11, 2010 publications.waset.org/10424/pdf

same DBMS, as well as different DBMS. When creating data pipeline object must specify some pipeline parameters, for example:

Pipelineobject.Start(sourcetrans,destina-tiontrans,errordata windows{,arg1,arg2,…,argn})

The first three parameters are necessary. P1: sourcetrans is the name of transaction processing object in source database;P2: destinationtrans is the name of transaction processing object in target database; P3: errordatawindows is the DataWindow control name in the pipeline handling process, others may be default parameters. The following is some program code of migrating the source data pipeline into the target database:

```
// Defining data pipeline object----pipeline EtlPipe
// Defining source and target databse transaction----transaction SourceTrans;transaction TargetTrans
// Building connection to source database----SourceTrans = create transaction;SourceTrans.dbms = "odbc";
SourceTrans.dbparm = "connectstring='DSN='" + SourceDsnString + "'"
// Building connection to target database----TargetTrans = create transaction,TargetTrans.dbms="odbc";
TargetTrans.dbparm = "connectstring='DSN='" + TargetDsnString + "'"
//Building pipeling----EtlPipe = create pipeline;connect using SourceTrans;connect using TargetTrans;EtlPipe.start(SourceTrans,TargetTrans.dw_err);disconnect using SourceTrans;disconnect using TargetTrans;destroy EtlPipe;destroy SourceTrans;destroy TargetTrans
```

D.Building the Metadat

D.1 Building Dimensionality Metadata

Such as dimensionality defination. Document type definition (DTD) of application XML can have dimension information (Dimension, DimLevel, and Member) be defined as follows:

```
<!ELEMENT Dimensions(Dimension*)> <
Dimension(DimID,DimName,DimFlds)>
<!ELEMENT DimID(#PCDATA)) >        <
DimFlds(KeyFldCode,DesFld+)>
<!ELEMENT DimName(#PCDATA) >       <
KeyFldCode(#PCDATA)>
<!ELEMENT DimFldName(#PCDATA)>     <
DesFld(DimFldName,DataType,DataVal)>;
<!ELEMENT DataType(#PCDATA)>       <
DataVal(#PCDATA)>
```

D.2 Managing Subject metadata

Useing XML DTD to describe and define subject, measurement and the relationship between dimensions,it's as the following:

1. **Subject**：<!ELEMENT Subjects(Subject+)> <
SubjectID(#PCDATA)><!ELEMENTSubjectName(#PCDATA)><Describe(#PCDATA)><!ELEMENTCreator(#PCDATA)><Date(#PCDATA)><!ELEMENTSubject(SubjectID,SubjectName,Describe,FactTbl,Creator,Date,Dimension+)>

2.**Fact**<!ELEMENTFactTbl(FactTblName,DimPrmkevFldName+,Measure+)><!ELEMENT FactTblName(#PCDATA)>

3. **Measure**：<!ELEMENT MsrFldName(#PCDATA)>
<
MsrName(#PCDATA)><!ELEMENTAggregateType(#PCDATA)><Measure(MsrFldName,MsrName,AggregateTvpe)>

4.**Dimension**<!ELEMENTDimID(#PCDATA)><Dimension(DimID,DimPrmkevFldName,DimDescibe)><DimPrmkevFldName(#PCDATA)><DimDescibe(#PCDATA)>

E. OLAP Design

E.1The typical multidimensional star-type model design

Generally,the OLAP design of DSS will transform indicators entities in the system theme domain into fact tables in the relational database,and transform dimension entities into in the relational database. By testablishing the relationships between fact tables and dimension tables in the form of primary key and foreign key, multidimensional star-type framework can be formed with all the fact tables as the center.

E.2 Multidimensional analysis

DSS needs to use multidimensional analysis ways to provide multidimensional analysis services for user level.

1.Multidimensional analysis session management:The multidimensional analysis Management is to import users' multidimensional analysis demands, specifically in the querying dialog,which includes themes,constrainting conditions, aggregate levels be observed in all various of dimensions, and what dimension checked. According to the multidimensional results collection, it can have various multidimensional operations (Slice.Dice.Rotate.Drill, etc.), so get some angle of data in a comprehensive degree.

2.Analyzing and querying by use of SQL:Through the star-type model dimensionality and multidimensional theme management are constructed,so the connection between dimension tables and center fact tables exists, and can use SQL statements to realize the querying and analyzing to the data,it includes three ways:constructing Material View, constructing the connectiong between tables,and constructing temporary tables.

3.Optimizing query performance:Considering the features of star-type model, the analyzing records number of general dimension table is limited, but fact table is large,which maybe several order of magnitudes more than that,so it can use multi-connection, aggregation operating algorithm to raise query performance, such as the way of Material View expansion.

F. Data Mining Design

1. Data Classification Standard[4]:Class should be internal high cohesion and low cohesion between classes. On classifying, it need be high comparability between isomorphic class objects, namely as high cohesion, and there is a great difference between different kind of class objects, namely as low cohesion. It need adopt different data conceptual level on different classification tasks in different datasets, set different class valve, select different classification properties, there will be completely different classifying results.

2. Data Conception Arrangement: It contains two ways:implication type representation way and tree structure representation way.

3. Main Induction Relation: In order to get clear classification mode, the first is to induce the original data into

Open Science Index, Computer and Information Engineering Vol:4, No:11, 2010 publications.waset.org/10424/pdf

the higher conceptual level. This can be done on the way of using induction facing to property in the data related to assignment. It contains: increasing, classification threshold, main induction relation, coverage number of main induction relation.

4. Classifying Mode:Based on classification properties in main induction relations to set off tuple, and then export one conjunction normal form for each class. If the values of some certain property(except classification property) in main induction relation cover all the concepts on its respective concept level,this property should not be included in the the certain conjunction normal form; Otherwise, the property should be included in the conjunction normal form,and all the values of some certain property are included in the property set.

5. Induction Facing to Property: In the classification rules facing to property, knowledge engineers or field knowledge experts generally put out the data conceptual level. Some data conceptual level can be automatically extracted. Generally, the type of data in the database can be divided into character, date, boolean and numerical.

6. Arithmetic of Main Induction Relation and Classifying Mode:

Creating arithmetic of main induction relation and classifying mode is the key to control data mining technologies in itsd using whether the effect is good or bad.

The first is to achieve induction facing to the properties. That is, a property without a higher level concept will be deleted, otherwise to induce the properties on the way of upgrading conceptual level,namely as replacing the property value with a higher level concepts.

The next is to create classification mode, and show by conjunction normal form. Conjunction normal form, namely as if the data values present to some properties are covering all the data on its respective concept level, these properties should be deleted from conjunction normal form, otherwise, these properties and the pack of related values should be included in the category of conjunction normal form. On the other hand, if the data values present to some properties are not covering all the data on its respective concept level, which shows that there are only some specific values in this class. By containing these specific values in conjunction normal form,this classification is precisely charactered and the classification border is very clear. So the property and its related values pack should be included in the conjunction normal form.

IV. Conclusion

Decision support system is a complex system engineering. This article gives thorough research on building DSS with data warehouse technology, points out the role and significance of DSS in modern management , as well as the deficiencies of traditional DSS approachs.At the same time, research DW composition, DW structure and DSS Architecture based on DW, puts forward the theoretical basis and key technologies of building DSS, including data preparation, Metadata Model Building, OLAP, data mining, and give a set of DSS design solution scheme in accordance with typical DW technology.

References

[1] Fan Jin-Sheng,Zhang Xiao-Ning. Imports and exports decision support system based on data warehouse [J]. COMPUTER ENGINEERING AND DESIGN ,2008,(12).

[2] Zhang Zhi-Jun. Enterprise Management Decision Support System Based on Data Warehouse[J] Computer Appication and Software,2005,(6):41-43.

[3] SHAMS,KAMRUDDIN,FARISHTA,et al.Data Warehousing:Toward Knowledge Management[J].Topics in Health Information Management,2001,21(3):24-32.

[4] FORGIONNE,GANGOPADHYAY,ADYA.Cancer Surveillance Using Data Warehousing,Data Mining,and Decision Support Systems[J].Topics in Health Information Management,2000,21(1):21-34



Yang Bao was born in ShenYang City,LiaoNing Province,China in 1978.He received the bachelor in 2001 and the master in 2004,both in computer application major from HARBIN INSTITUTE OF TECHNOLOGY,HARBIN,China. Since 2004,he is working in Beijing Institute of System Engineering,BeiJing,China as the research associate,mainly engaging in software engineering research. He is the member of China Computer Federation(CCF).His papers such as Research on the Data Quality of the Large-scale Software System published on Computer Engineering & Design,Third-Party Software Testing and Evaluation published on JOURNAL OF COMPUTER RESEARCH AND DEVELOPMENT,and etc.His research interests include software developing,computer software architecture research, software testing,database technology.



LuJing Zhang was born in 1975 in BeiJing, China. She received the bachelor in 1996 and the master in 2003,both in computer science and science major from Nanjing University, China. From 1996 to 2003, She worked in Beijing Electronics Standardization Institute. Since 2003,She has been woking in Beijing Institute of System Engineering as associate researcher. She is the member of China Computer Federation(CCF).Her papers such as The application of fuzzy token bucket terminology in data management system published on Computer Engineering and Applications,software configuration management tools and its using in testing labs published on Computer Engineering And Design, and etc. Her research interests include Software Engineering, Network Security, Data Management and software testing.