DSS based on Data Warehouse

Decision support system is a complex system engineering. 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.

An Intelligent Decision Support System: Data Warehouse, Online Analytical Processing, Data Mining and Decision Supporting based on models: DSSBDW.

- A DSSBDW system enables knowledge workers to make strategic and tactical business decisions based on factual data stored in a business intelligence data warehouse. DSS must allow users to intuitively, quickly, and flexibly manipulate data using familiar terms, in order to provide analytical insight.
- DSSBDW are implementations of an informational database used to store sharable data sourced from an operational database-of-record. They are typically a subject database that allows users to tap into a company's vast store of operational data to track business trends, facilitate forecasting and data mining, respond to data patterns, aid planning efforts, and measure performance against strategic goals—all with the goal of aiding and supporting decision making.

Yang Bao LuJing Zhang - Decision Support System Based on Data Warehouse - <u>World Academy of Science, Engineering and Technology 71 2010</u>
 MicroStrategy / Support - Decision Support Data Warehouse - <u>www.microstrategy.com/decision-support-data-warehouse/</u>
MicroStrategy

Decision Support and Data Warehouse (DW)

Decision Support System (DSS) generally is consisted of three bases model, and Intelligent Decision Support System (IDSS) is consisted of four bases structure, 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.

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

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.

DSS Architecture based on DW

DSSBDW 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.



THE KEY TECHNOLOGY OF DW IN DSS

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.

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.

management	Warehouse Process						Warehouse Operate				
analysis	Transform	OLAP		Data Mining		Information Vis		isualization Co		Commercial Terms	
resource	UML	Relational Resource Re				Reco	ording	g Multidimension		XML	
foundation	Business Information D		Data	Туре	Expression		Keys Index		Type Mapping		Software Release
	UML 1.3 (Foundation,Behavior Element,Model Management)										

The composition of CWM Metamodel (*Package Structure of CWM Metamodel*).

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.

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



Main Steps of Data Mining

Designing and Implementing 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 next figure shown:



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, develops DSS application, strengthen management, take the active participation with end users, decompounds the complex needs, prepare data, provide OLAP and data mining tools, etc.

THE DSS DESIGN WITH DW TECHNOLOGY

DSS design adopts the architecture shown as next figure. Data 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.



Architecture of DSS:

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 next figure:



Realizing Scheme of Decision Support System Based on Data warehouse

ETL Design

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

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.

Data Pipeline is a data migration tool, it is used to migrate from a database to another database, these database can be the same DBMS, as well as different DBMS. When creating data pipeline object must specify some pipeline parameters.

OLAP Design

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 establishing 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.

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, constraining 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.

... OLAP Design

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 connecting 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.

Data Mining Design

 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 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. 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 its 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 characterized and the classification border is very clear. So the property and its related values pack should be included in the conjunction normal form.

... Data Mining Design

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 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.

End of ... 13

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 Application 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

 $\dots C_{13} / 7.01.2020$ 15/15