Research on Multidimensional Data Storage Model Based on Relational Database

Yan Zhang, Dong Xie, Guangjun Guo, Tong Jiang

Information School, Hunan University of Humanities, Science and Technology, Loudi, 417000, China

Keywords: Multidimensional Data, Storage Model, Relational Database

Abstract. This article describes a relational database for the ultimate data storage management model, through the establishment of a multi-dimensional data storage management support module between the data management and relational database, users achieve multidimensional data objects and storage management method which expounded the key idea of the algorithm, an application example, and evaluated.

Introduction

Traditional relational database RDB is a two-dimensional relational database, it can directly support the description and data relationships are two-dimensional table, each attribute must be atomic level, that the table can not have a table. However, in practice application areas, some of the complex multi-dimensional relationship is ever-present, in order to describe these complex data relationships necessary and multidimensional, programmers often need to spend a lot of energy to do some sophisticated design, the multi-dimensional relationship to defuse a two-dimensional relationship and define the corresponding database table. hand, this approach can not effectively help the programmer to define and describe the multi-dimensional relationship, on the other hand when the multi-dimensional uncertainty relation, the programmer can only be guaranteed by reserving alternate models and other methods of property availability and flexibility of object-oriented programming ideas so that people in the design of a complex program to an instance of the class as a whole to consider and understanding, but in fact really effective to an object as a whole when storage management currently relational database could not be reached.

Construct a multidimensional relationship maintenance and data manipulation environment so that users can easily manage this data with complex relationships is necessary. For the ultimate data storage management model herein to a relational database, by user relational database of established between a multidimensional database management module, the user data on a multidimensional view of data manipulation translate into a relational database manipulation, and the results returned by a relational database in the form of multi-dimensional data to show users; realization of complex multi-dimension user transparent management of relational data manipulation and query.

Multidimensional Data Storage Model and Its Description

Multidimensional OLAP data storage model is realized on the basis that it is different from the traditional two-dimensional data model (relational data model). Analysis of multidimensional data storage model based on data analysis needs to implement different data in multiple angles.

Multidimensional data storage model has the following characteristics: high-speed integrated speed embedded powerful analysis and calculation functions, and automatically shut-level meter model; a similar "hypercube" configuration, dimension and combination of attributes, metrics produce the intersection where the variable views. User custom "sheet" available online Random flexible query to the data needed further analysis. "Rotate the data cube" and "slice dice" is the main technical multidimensional data analysis; time series: usually only store in a data storage unit in a few years of data, simplifying application development and data storage and processing time; vector operations: time array operation fast, easy synthesis, one step to complete the operation; data tracking: drill through the data and change the angle.

Multi-Dimensional Data Storage Model

In the multi-dimensional data storage model, the data structure emphatically described "metrics" and "Victoria." Metrics are used to track the numerical data, which can be stored in the important "fact table". Dimension is defined for each transaction of natural service parameters may be stored in dimension "satellite list" and the fact table association, called dimension tables. Dimensional object model is viewed in accordance with the business analyst perspective of the main organizational information. All dimension table stores all the information related to a particular dimension. The main record for each dimension hierarchy and description attributes.

Often organized dimension to level. Such as the time dimension, daily accumulation into weeks, weeks and accumulated quarterly. Ratings for the model dimension are very important because it provides a detailed and functional structure comes down. The so-called refinement is to get a data set of more detailed view of the process, is attributed to a reverse process.

Multidimensional analysis provides users with a variety of multi-dimensional data views, these views reflect the information hidden in the data behind the business and the nature of specific operations multidimensional analysis include: multi-level data based on one or more specific statistical dimensions (total, average value, variance, etc.), top-down in the data hierarchy that is, from coarse to fine and bottom-up that is, from fine to coarse data mining, various projection, rotation, cross table and prediction (causal forecasting and time series forecasting). In order to support these OLAP operations can be a star and snowflake-shaped pattern in a relational database to organize data.

Star Schema. The organization of multidimensional data in a star pattern, the middle is a fact table, connect to multiple radial dimension description in which fact table to store the actual end-user business data and key value of interest-dimensional description of a plurality of these health It composed of a combination of the fact table primary key; dimension description: store the dimension key value and other non-health property description Kin values (such as product dimension table features, price, etc.) the main alive fact table dimension described in the table. It is alien code.

Star pattern reflects the final user of the multi-dimensional analysis of the way. A simple logical star schema consists of a fact table and several dimension tables and complex star schema can contain hundreds of fact and dimension tables. Fact table contains the basic business data, which may include thousands confidant record; dimension tables include a business property SOL query conditions, generally small general fact sheet is for a particular topic, are private; while dimension table for multiple shared theme, dimension tables are common, especially in the time dimension table, it is a standard dimension, for all subjects sharing. users can from time, store and product category for product sales, etc. are analyzed in the fact table contains the key value, product sales as well as sales of each dimension in national currency, US dollars, and may be in the form of currency dimension tables described in the non-health property so the fact table include only sales.

A star pattern in favor of multi-dimensional analysis of the structure, namely hypercube, standardized SQL statement connecting operation, reducing the level and complexity of the connection, improve query performance. Multidimensional database does not support multi-table joins, multi-table joins is the outstanding characteristic of the star structure.

Snowflake Pattern. for complex multidimensional data dimensional model, a star schema and not ideal. Because a lot of dimension tables is detrimental to performance, partly offset by the performance data is gathered. In some cases, the property store dimensional element will consume a lot of space, a lot of time when dimension elements, more prominent. Compared with the star schema, snowflake pattern main drawback is normalized relational data structure of snowflakes is relatively complicated. If you use a typical database query tool to generate queries, which will be difficult to navigate multidimensional database based snowflake pattern. Further, since the data model becomes more complicated, and the overall maintenance of the loader more difficult.

Snow mode, each level of each dimension table dimension level, that each element has a keyword. The lowest level of keywords associated dimension tables and fact tables, and dimension attributes associated with the lowest dimension table contains a description of the elements of

information. The remaining keywords dimension tables associated with the corresponding attribute table. Snow mode can improve performance and save disk storage, so at the expense of increased complexity is worth it. For table has many rows, many properties are stored in a low-level dimension table level or limited disk capacity, the snow mode is effective. The best solution is the one or two-dimensional into snow mode, the other dimension is still a star schema, called local snowflake pattern.

Aggregate Data. Aggregation is the low-level data are summarized in advance and place the middle of the table. These aggregate tables allow applications to predict the user's query, to avoid duplicate computing resources concentrated. Measure variables in the form of hard details are stored in the fact table to a later stage of data analysis used.

Details chromatography fetches data from and do not provide the best performance, even when using the most advanced software and hardware. The fact sheet will become increasingly large, resulting in significant performance degradation. Because statistical query is very resource-intensive, it is usually to access the data in advance generalizations are useful. These intermediate results or the use of aggregated data can effectively reduce the consumption of resources at the same time to get the final results.

Therefore needed is intelligence gathering. Suppose you have a query to get annual sales, while sales by each business stored in the aggregate table on a monthly basis in the fact table, the query should be gathered twelve table records are added, rather than the fact that thousands of records in the table add up. If the query index can do that, we no longer need an annual aggregate sales data. With this intelligence gathering tool, aggregation structure automatically adapts to the foreground program use data, and the data obtained from this intelligence gathering tool foreground program do not have to change a line of code.

Metadata. In ROLAP, the data is based on a relational format (two-dimensional table) memory, indeed dimension given by the user, in order to hide the storage format, you need to create a semantic data layer, this layer is to map the user dimension to relational tables, the semantic layer from Yuan data to describe metadata (metadata) is another type of data stored in the data warehouse, similar to relational database data dictionary, is data about data storage metadata purpose is to create, maintain, manage, use database.

Model and Evaluation

First, the application of this model can be achieved on the object storage management users can use to read and write data objects directly in the database, without having to start reading object database of all properties, and then again one by assignment; in the opinion reach the user the effect object database.

Second, users can transparently and easily manage complex data relationships, achieve records management function. Since the reference relationship between tables, so data hierarchy between tables more distinctive. In a real application environment, can greatly improve the user's data management efficiency, help users to examine data from different data levels of internal relations, which greatly facilitates data mining layer progressive style. our existing common SQL Server as the relational database management system, in object-oriented programming language Visual Basic environment, the initial realization of this model, and the model is applied to the definition of a generic product production management model, in general production management model, users tend to be based on their production of different products and product update changes dynamically define their own specifications of each product composition; users of products, raw materials and semi-finished products specifications are different, we have created for each category of products, semi-finished products, raw materials, a multidimensional property sheet, the product contains a corresponding class or categories of complex properties of semi-finished raw materials such as multi-dimensional table: a color TV manufacturers to define the TV table (make, model (primary key), size, picture tubes, display), kinescope table (whether Color, size , specifications, brand), the display table (whether flat, size, manufacturer), the picture tube TV table reference table and display table, on the basis of a user can be quite easy to define new models of color TV, but just

give the TV model to complete the model used in color television picture tubes and display data read and write. on the other hand if the user later added such as "manufacturers" and other relevant attributes to kinescope kinescope table original program information when viewing the TV, but also to find new dynamic "kinescope. Manufacturer" attribute allows the user to put the picture tube as a whole, and not worry about missing the kinescope new information; the use of this model we provide to users flexible and efficient production management model can complete the definition of products, raw materials and semi-finished work.

Acknowledgements

Conclusions

This article describes a relational database-based storage management, object-oriented programming language environment for running programs to read and write data, to achieve the objects and complex multidimensional data storage management method, briefly introduces the realization of its algorithm ideas and its application and its effect was evaluated, this has proven to be an effective way can a flexible and dynamic management of complex data.

References

[1] Rajesh Narang, K.D.Sharma: Computer Journal of Science and Technology, Vol. 2 (2009) No 53, p.25-26

[2] Fishman D H: ACM Transaction on Office Information System, Vol. 12 (2010) No45, p.83-85

[3] Meng Li IEEE Computer Technology, Vol. 2 (2013) No 59, p.102-105

[4] Shuang Wu: Computer Information, Vol. 2 (2011) No 43, p.29-43

[5] Koen De Bosschere: The Journal of System and Software, Vol. 12 (2014) No 22, p.55-60