IV. The implementation of generating the dominant dimension from the hidden dimensional hierarchy

A . Steps

(1) With the conditions described in A of part Ⅲ, ensure the values correctness of tuples property which wants to generate the dimension hierarchy.

(2) Based on the condition of B of part III and the definition of A of part II, determine the dimension class feature of property which wants to generate the dimension property hierarchy, it means dividing tuples in dimension tables on certain condition.

(3) Based on the condition of C of part III and the definition of B of part II, to determine the every equivalence class of value range of property which wants to generate the dimension property hierarchy and whether the result of the aggregation of the metric on every equivalence class is a partial order after some mapping process.

(4) With the definitions of C of part II and D of part II, to determine the feature of hidden dimension hierarchy of the generated dimension hierarchy from step (3).

(5) According to the functionality and performance of the system OLAP, to determine whether to form the hidden dimension hierarchy generated from step (4) to apparent dimension hierarchy. If the forecast can be identified that OLAP probability related to the apparent dimension hierarchy is very low, can choose to give up generation, and terminate the design of the dimension hierarchy. Otherwise, turn to steps (6) and (7).

(6) To define the Meta data in the system and describe new dimension hierarchy in the dimension table.

(7) To form the aggregation tables by the dimension hierarchy generating in step (5) for further use in OLAP

B . An Application

The CDCDH method is used in the establishment of data warehouse and decision analysis system design process in university. In the data warehouse, the teacher dimension table TeacherD holds Teacher ID, Department ID, Sex, Title ID, Education ID, Age ID properties, and design TeacherD with three sub-dimension elements Sub-TitleD, Sub-EducationD and Sub-AgeD. The table Sub-tileD generally divides into five categories Responsibility Professor, Professor, Associate Professor, Lecturer and Teaching Assistant respectively. The table Sub-EducationD can be divided into four categories Post Doctorate, Doctor, Master and Bachelor respectively. The table Sub-AgeD divides into Senior Teacher (50 years old), Middle-Aged Teacher (40 to 50 years old) and young teachers (40 years old). Obviously, all these are dimension hierarchies generating from mapping operations of the metric of value range. So can determine the hidden dimension hierarchies and generate apparent dimension hierarchies.

(1) Any tuple in the teacher dimension table has clear property value of title, educational background, age, etc.

(2) According to the same title, same age, same degree, tuples can be classified into each hierarchy in title sub-dimension table Sub-TitleD, degree sub-dimension table Sub-EducationD and age sub-dimension table Sub-AgeD.

(3) Define the mapping:

f: $x \mapsto$ Count(x), where $x \in$ {{Responsibility Professor}, {Professor}, {Associate Professor}, {Lecturer}, {Teaching Assistant}}

g: $x \mapsto$ Count(x), where $x \in$ {{Post Doctorate}, {Doctor}, {Master}, {Bachelor}}

h: $x \mapsto$ Count(x), where $x \in$ {{Senior Teacher}, {Middle-Aged Teacher}, {Young Teacher}}

Obviously.

A partial order is formed with the number of assistant teachers above> the number of lectures above >the number of vice professors above >the number of professors above >the number of responsibility professors above, which holds the hidden dimension hierarchy feature.

A partial order is formed with the number of bachelor degree above> the number of master degree > the number of doctorate above, which holds the hidden dimension hierarchy feature.

A partial order is formed with the number of Young teachers>the number of middle-aged teachers>senior teachers, which holds the hidden dimension hierarchy feature.

So the above hidden dimension hierarchy can be converted to apparent dimension hierarchy on demand for OLAP.

V . Conclusions

CDCDH method is introduced and a comprehensive discussion is given on dimension analysis and designs, which can avoid the performance decline caused by defects in design process and reduce the development cost. Of course CDCDH method has its shortcomings. In example B of part IV, the dimension table structure includes not only the dimension class attribute fields but also sub dimension hierarchy tables generating from dimension class, the larger storage space is needed. With the development of Mapreduce, a lot of storage space is shared on the network, OLAP efficient response will become an important research direction.

Reference

[1] Wang Shan, Data Warehousing and On-Line Analytical Processing. Beijing: Publishing House of Science,1999,pp.50-128.

[2] Leng Fangling, Bao Yubin, Yu Ge, Gao Wei, "Closed Data Cube Based on MapReduce",Journal of Computer Research and Development,vol.48, no.12, March 2011,pp .232-238.

[3] SHI Jingang, BAO Yubin, LENG Fangling, YU Ge, "Efficient Parallel Dwarf Data Cube Using MapReduce", Journal of Frontiers of Computer Science & Technology, vol.5,no.5,May 2011,pp. 398-409

[4] Xi Jian-qing, You Jin-guo, Tang De-you, Xiao Wei-ji, "A Parallel Closed-Cubing Algorithm Based on MapReduce", Journal of South China University of Technology(Natural Science Edition), vol.37,no.1, December 2009,pp.91-94.

[5] Leng Fangling, Bao Yubin, Gao Wei, Yu Ge, "MapReduce-based data aggregation algorithms Sciencepaper Online",vol.6,no.7,June 2011, pp.469-481.

.