ABSTRACT

To validate the application effect of distributed MOLAP with large data analysis. Methods: The influence of large data scale effect is inevitable during the process of data storage and analysis. In order to solve these problems, a multi-distributed file system is adopted to deal with, and a distributed MOLAP technique for large data analysis is analyzed by establishing a data model. Conclusion: Distributed MOLAP techniques for large data analysis are more efficient.

KEYWORDS: Method of the Characteristics, Water Hammer, Pipe Network Discretization.
1. INTRODUCTION

With the development of modern information technology, big data times has quietly come, big data makes fields such as internet technology, electronic computer technology, science technology obtain wide development space, but under the background of universal development in every field, which may cause information data bursting and affect processing ability of information data in every field. Therefore, MOLAP technology has been universally applied; it can realize quick storage and processing on data information on the basis of data structure with multi-dimension. MOLAP uses one expert multi-dimension data base to store data required by MOLAP analysis, data is stored by way of multi-dimension and it supports multi-dimension view of data. The position of data in multi-dimension space is determined by dimension nature, data value is the value of measurement nature. In MOLAP structure, data scattered in every OLTP data base within enterprises is submitted to multi-dimension data base through extracting, cleaning, converting etc. When these data is stored in multi-dimension data base, it makes a series of pre-operation (calculation and integration) according to its belonging dimension and stores results into multi-dimension data base according to certain hierarchical structure. Users submit analysis demand to OLAP server through interface of application software in customer terminal, and OLAP server searches MDDB data base and feeds back results to customers. Multi-dimension data base(MDD) stores data into one array of n-dimension, it has plenty of sparse data, in the part of event occurrence, data is integrated together and has big dense, which is regarded as dense data. Therefore, it needs to solve problems of sparse data and data integration. MDD separates sparse data from the dense data; it makes compression and storage on dense data so as to reduce occupancy of storage space. The summary data stored in MDD derives from the original detail data, so it needs one kin of mechanism to extract data from data source. The main advantages of MOLAP structure is that it can quickly respond to analysis request of decision analysis personnel and quickly feeds back analysis result to users, this is profits from its unique multi-dimension data base structure and data(generally pre-processing degree is above 85% with high pre-processing degree.

In the actual application of MOLAP, it is usually regarded as one part of data warehouse application, its unique and stable cubic structure is completely separated from the relational database, while online analysis and processing based on MOLAP, data is stored in the physical cube, because cube includes data subset calculated in advance, so data analysis and processing efficiency based on MOLAP technology is higher, and it is easily to be realized. In MOLAP, it usually regards multi-dimension image as one super-cube in concept,
and organizes it as one multi-dimension database in physics, the nature value of dimension is reflected into index bound of multi-dimension data, and stores the data of multi-dimension array unit, which is the core of organizing data by using multi-dimension array to store data, while multi-dimension data will form into cube structure in the multi-dimension data, drilling, rotating, section and switch are the most basic technologies.

2. DISTRIBUTED MOLAP TECHNOLOGY OF BIG DATA ANALYSIS

The involved OLAP of distributed MOLAP technology of big data analysis is divided into 5 operations: scroller, drill-down, cube, section and rotation. Of which, search of data range includes cube and section, data range and integration can be divided into scroller and drill-down, different images of creaming cube includes rotation. While one LLAF in distributed MOLF technology can be divided into 4 abstract tuple, of which it respectively includes Target representing data cube waiting for analysis, Flang C of data range waiting for analysis data in the cube, aggregation function Aggregation represents data analysis and Result represents data cube result. In the distributed MOLAF of big data analysis, in the process of involving with OLAF in technology will use 4 kinds of most basic estimation method, of which includes data search, block option, change of dimension grade and aggregation of information data. While block option is the key of these 4 basic estimation method, it refers to data information can meet all the search conditions and process to be quickly determined, so it can use this algorithm to reduce the input number required by OLAF operation to the minimum. Which means quick option algorithm enhances the operation performance of OLAP by way of reducing search space. In the distributed MOLAF of big data analysis, when technology is making OLAF operation, it can data range Range represents waiting for analysis data in the cube set as \{α1, α2, ..., αn\}, \{β1, β2, ..., βn\}, the block coordinate of setting can meet and represent data range Range waiting for analysis data is \{c1, c2, ..., cn\} block scale is \{λ1, λ2, ..., λn\}, so the formula is as follows: \((α_i/λ_i) ≤ c_1*(β_i/λ_i)\). The block can meet the above formula has no necessary to input all the data information block, while it just needs to make input in the process of technology making OLSF operation in the distributed MOLAF of big data analysis, which reduces search space of OLAF to great extent, it can complete it through simple coding. The detailed option algorithm is indicated by the following figure:

![Figure 2 Block option algorithm](image_url)

Make OLAP on distributed MOLAP of big data analysis, the data range waiting for analysis in
the cube is represented in the operation process, which is set as \((\{x_1, y_1, z_1\}; \{x_2, y_2, z_2\})\), the coordinate in the corresponding diagram is F and D, form this it can induce that data information range needs search is between ABCD-EFGH. Make the whole cube as reference and it can find select most unrelated information to OLAP. Take the OKAP involved in the distributed MOLAP technology of big data analysis for example, its algorithm can be divided into 4 parts, which is input formatter, mapper, reduce and output formatter, its detailed scroller operation procedure of OLAP is indicated by the following diagram.

![Figure 3 Scroller operation procedure of OLAP](image)

In MapReduce, before program procedure model is working, the scroller operation involved in OLAP in the distributed MOLAP of big data analysis, its 4 parts will be submitted, while system terminal will make effective verification on data information of these 4 parts, this is to prevent program invalid due to result in search task. After this operation is completed, it obtains the required input list by block algorithm. MOLAP technology is to use Map Reduce as basis, it will make processing on electronic computer by many ways in the actual operation process, for example, it makes mapping on dimension and measurement of information data, this is implemented by multi-dimension model, such as operation on dimension level of information data, this operation undergoes traversal algorithm and operation of dimension coding. In a word, MOLAP technology plays important effect in the whole information technology processing; it is also the key point of present analysis.

3. DATA MODEL

MOLAP technology adopts multi-dimension data model, its content has dimension and fact. The key point in the operation process is to seek mapping relations between dimension and fact. In the actual operation process, traditional ROLAP technology adopts relational database and star model, which realizes respective storage of dimension information and fact information, it realizes mapping of both through foreign key, this makes operation efficiency in the actual operation low. MOLAP needs one kind of integrated way to maintain mapping of dimension and fact, because one dimension can include many layers, every layer can include several grades; dimension and fact are relation of one to more, so dimension model has complication. In order to avoid additional storage and maintenance cost, DOLAP firstly makes simplification on dimension so as to adapt to distributed environment, meanwhile it reduces complication of OLAP algorithm.

(1) Dimension: in the multi-dimension model, after classification on all the data, dimension will place all the even data in the data structure
without overlapping; meanwhile, it provides item selection method and organization method among every data. In this research, we firstly makes simplification on connotation of dimension and data model with multi dimensions, the model after simplification should obey restrictions of the following conditions, suppose a is dimension then a has dimension layer, but the number of dimension layer is 1. a is the assemble composed of n dimension grades, it is marked as { }. suppose (i∈[1, n]) is any dimension grade, then there is one dimension nature having one value. Now it regards a as the structure composed of dimension nature value of every grade, the node of the same grade includes the same sub-node number.

Figure 4 Example of data cube

Figure 4 is the cube composed of 3 dimensions (x, y, z), the smaller grid represents cell, the bigger grid represents block. In the actual operation, block may include some empty cells, which means this cell has no measurement. In the actual application, in order to reduce physical space of cube, if cell has any no measurement, this file dose not store record of this cell.

(2)Measurement: suppose measurement is one independent variable, and is regarded s MOLAP technology analysis object according to dimension value, of which the measurement with the finest particle refers to the dimension value with low grade in the dimension.

(3)Cell: in the whole logic view, cell is composed of different measurement, these measurements use the same dimension value as reference object, so cell can be regarded as the combination of measurement.

(4)Data cube: according to the above-mentioned definitions, data cube is the multi-dimension structure of MOLAP, it is composed of many cells.

(5)Definition of block: block is the logic diagram of data cube, every data cube can take different values according to different dimensions.

4. LAYERED DIMENSION INTEGRATION AND DATA STORAGE OF MOLAP

A.Layered dimension integration of AMOLAP

It adopts multi-dimension array to realize MOLAP based on multi-dimension database. Constructing MOLAP, if it wants to establish dimensions with layer grade, multi-dimension array and cubic structure of corresponding cluster operation. In the design of category, it establishes base class of Dimension, it derives subclass of dimension. For example, it further distinguishes Standard Dimension of dimension with different structures, and Non Standard Dimension of dimension with special structure. Under the standard structure dimension class, it is distinguished as the usual Star Dimension, Snow Dimension, because dimension includes
layer relations, so it should establish one base class Dimension Status which can represent dimension information under the state of different layer and structure, when making processing on dimension with designated layer, it uses derived class of this base class, such as Standard Dimension. Sion Status, Non Standard Dimension Status and base class Level of defining grade member. As for the detailed structure pattern, it can derive Star Hierachy, Star Level etc. In the aspect of database storage, it considers storage of layer data. This paper uses Hashtable to store hierarchical relationship. The members are respectively member and parentIndex.

```java
/**
 * The corresponding relations of father node in level element
 */
Private H asht able parents= new Hashtable( ) ;
......
/**
Set number of father node of level element
@param member level element
@param parentIndex number of father node.
*/
This, it can set father node number of level element through parents, put0, when it needs to use information of level relations of dimension, it can obtain father node sequence of designated level element through parents, put0. Here it discusses realization of dimension integration with level. As for aggregation operation, it can establish corresponding base class of aggregation calculation; different aggregation calculation operation is derived from it. Dimension aggregation is divided into 2 kinds of conditions, aggregation and level aggregation of dimension. As for the aggregation of the whole dimension, it can be regarded as calculation of old cube, stores the obtained result into new cube. As for algorithm of aggregation, it can consider referring to array of 1 dimension to simulate ergodic algorithm of multi-dimension array. Under the condition of unknowing multi-dimension array dimension, if it uses the usual way to indicate multi-dimension array. Generally speaking, we think that if it is 3 dimensions, it makes 3 circulations, every dimension one circulation, but it may be 4 dimensions and 18 dimensions, the multi-dimension array of one dimension change can not make traversal through designated corresponding dimension. Since this, we can also use 1 circulation to solve problem. The principle is from the lowest level, starts from 0 and adds 1 every time; it adopts way of carry to indicate changes of dimension. 0. 1. 10. 11..., the detailed code section is as follows:

```
Lei Cheng, Yajun Zhang. 33(2017) 47-56

//
Update idxNewArray, so as to make traversal
idxNewArray[newDimensionCount-1]++;
for (int newDimensionCount = 1; i > 0; i--) {
    if (idxNewArray[i] >= newDimensionLength[i]) {
        idxNewArray[i] = 0;
        idxNewArray[i-1]++;
    } else {
        break;
    }
}

If it considers the cluster problem of level, calculation of dimension aggregation becomes further complicated. We not only need the present level information, but also need the information of the last level. Traversal level aggregation needs the following steps: || traversal from element of the bottom level to element of the highest level.

while (column < table.getColumnCount0) {
    / / level sequence (the highest level is 0)
    Object cell = table.getCell(row, column);
    if (levels[levelIndex].indexOf(cell) == 1) {
        / / element is not existed grade object of this level, so it is added to grade object of this level.
        levels[levelIndex].addMember(cell);
        if (levelIndex > 0) {
            / / is not the highest level element, it needs to construct relations of father node
            Object parent = table.getCell(row, column + 1); // father node element
            int parentIndex = levels[levelIndex - 1].indexOf(parent); // the sequence of father node element in this level grade
            if (parentIndex != 1) { // father node element is existed in the corresponding level grade
                levels[levelIndex].setParent(cell, parentIndex);
                break;
            } else { // father node element is not existed in the corresponding level grade
                levels[levelIndex - 1].addMember(parent);
            }
        } else {
            / / first data Cache, command Cache, star MMU etc.
            if (isPowerSwitch) {
                / / board interrupt
                BoardInterruptPendingRegister.BIPR = 0;
                BoardControlRegister.BCR = 0;
                FLASHB0WP and FLASHB1WP respectively controls Flashbank0 and bank. if here is 0, the corresponding bank can be used.
            }
        }
    }
}
DOLAP is obtained by calculation without memory, so storage cost of data cube is very small, DOLAP makes simplification on dimension, it can guarantee the code of the same level is the continuous decimal, meanwhile, every brother node has the same sub-node, so dimension information only needs to store scale of every dimension, which greatly reduces storage cost of dimension. Suppose dimension d is composed of m dimension levels, which is marked as \{li|i∈[1, m]\}, then physical storage of d can be indicated as assemble composed of dimension level and ordered pair of dimension level scale. \{\langle li, |li⟩ |i∈[1, m]\}, of which, li indicates the name of this dimension level. Realization system of DOLAP can use XML file to store information and stored in the main node of cluster.

In logic, structure of cube and its cell or cube and its block can be analoged as data structure of multi-dimension array and its element. In physics, block is the memory cell of cube, makes linearization on cell in the block, block can be regarded as the independent file to make memory. In order to find address, block and cell need support linearization and reverse-linearization calculation. and this calculation is the same with linearization and reverse-linearization of this calculation. Suppose there is one array of n-dimension, its scale is marked as \langle A1, A2, ..., An⟩, coordinate of element X in multi-dimension is marked as (X1, X2, ..., Xn) in multi-dimension space. the coordinate after linearization is marked as index(X), its linearization method is indicated by formula (1), reverse-linearization method is indicated by formula (2):

\[
\text{index}(X)=((X_n \times A_n-1+X_n-1)\times A_n-2+...+X_3)\times A_2+X_2)\times A_1+X_1 \quad (1)
\]

\[
\text{temp1= index, X}_1=\text{temp1} \bmod A_1, \text{temp2=}[\text{temp1}/A_1] \\
\text{X}_2=\text{temp2} \bmod A_2, \text{temp3=}[\text{temp2}/A_2] \\
\text{X}_n=\text{temp}_n \bmod A_n \quad (2)
\]

As for cell, the dimension of constructing cube is marked as d1, d2, ..., dn. Suppose x is one cell in the cube and the corresponding dimension value of x in dimension di is v′, code(v′)=xi, then coordinate of x is (x1, x2, ..., xn), formula (1) repalces (X1, X2, ..., Xn) as \langle x1, x2, ..., xn⟩, \langle A1, A2, ..., An⟩ is replaced as \langle |d1|, |d2|, ..., |dn|⟩, formula (2) is also the same. linearization and reverse-linearization algorithm of cell, cell is stored as one piece of record in MapFile of Hadoop HDFS, this record includes the coordinate and the included measurement of this cell after linearization.

As for block, suppose y is one block and it is the division of cube, |y|= \langle λ1, λ2, ..., λn⟩, coordinate of y is (y1, y2, ..., yn), coordinate of any one cell in y is (x1, x2, ..., xn), then \(y_i=[x_i/\lambda_i]\), formula (1) in which, replace (X1, X2, ..., Xn) as (y1, y2, ..., yn), \langle A1, A2, ..., An⟩ is replaced as \langle λ1, λ2, ..., λn \rangle, formula (2) is the same. it can linearization and reverse-linearization algorithm of cell. In storage reality, one block is stored as one MapFile file in Hadoop HDFS, the coordinate after linearization is regarded as the file name of block file, so that it can seek address of block file. Every record in the block file stores one cell in this block and block file is stored in the distributed file system Hadoop HDFS. In the file, it stores all the cell data by way of Key-Values, of which, Values is corresponds to reality data. Key is the search value after linearization of cell coordinate according to formula (1). This search value is one decimal positive integer, in the big data environment, cube scale of data is very large, take Java language for example, and search value will exceed the maximum range of \(2^{64}\) of long. In addition, we adopt character string to indicate the number data. In addition, if we adopt Key-Values to store, storage expenditure of Key is larger than the corresponding storage expenditure of Values. which wastes plenty of storage, therefore, as for one block file, it only record the minimum cell search value while storage of Key corresponds to the deviation amount of this value.
5. ALGORITHM OF DIMENSION CODING

Algorithm of dimension coding includes 2 ways, which are binary coding algorithm and decimal coding algorithm, binary coding algorithm can realize summary on dimension information of every level through connection of every coding and realize dimension review by position movement, but binary coding algorithm has certain omission, decimal coding is relatively detailed, which can realize coding calculation of grade dimension value, but it can not directly realize mapping between coding and dimension value. In order to avoid the omission in the actual work, MOLAP technology adopts decimal coding algorithm, suppose t is one dimension grade in dimension a, then calculation is as follows:

Input: Dimension a : A target dimension
Function: Dimension Coding
1. FOR f=1 TO |t(a)|
2. FOR g=0 TO |t(())|-1
3. Dimension value of ∈ ( )
4. Code= g
5. END FOR
6. END FOR

In the actual application, most data values are of value form of dimension, such as height and price etc. The dimension with data form can make different division according to difference of its co-domain. Steps with different division can confirm different dimension grade.

With the development of modern information technology, we have already come into bid data times, data has already deepened into every field of our life, the approaching of big data times makes data processing much more convenient, but human society will gradually developing and makes data quantity drastically increase, the traditional data processing way, data settlement, storage way can no longer meet demand, so it urgently needs one new data processing way under this condition, it is based on program editing of May Reduce.

6. REFERENCE

[8] Hu Yi. Hua Jing. Research on ROLAP and