ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

Datasets Preparation in SQL Using Horizontal Aggregation

Sonali Karle[1], Swati avhad[2], Ashlesha Shelar[3], Prof. Suvarna Pawar[4], Kajal Dighe[5]

[1] *B.E.(I.T), Pune University, India*

[2] *B.E.(I.T), Pune University, India*

[3] *B.E.(I.T), Pune University, India*

[4] *H.O.D.(I.T),Pune University, India*

[5] *B.E.(I.T), Pune University, India*

Abstract— ***Data mining is largely used in preparing data sets for data mining analysis. But it is so much time consuming process. It requires large amount of manual effort. Data mining is largely used domain for getting the patterns from historical database or stored database. Large amount of effort is required to prepare datasets that may be input for data mining algorithm. As we already have some aggregation function MAX,MIN,SUM,COUNT,AVG which are not efficient for making datasets in data mining analysis. This aggregate function have drawback as they return single value single value per aggregated group in that table.In data mining analysis when we requires data in horizontal layout that tine we require hard effort.So we are developing simple but powerful tool to get SQL code to return combined columns in horizontal layout form,which returns group of numbers instead of one number per row. This new group of tool or function is said to be horizontal aggregation.From third queries we will get output output data which is suitable for various data mining operations. It means this paper gives horizontal aggregation using some constructs that include SQL queries. Here we are using three functions which is Grouping column, Horizontal column, Aggregate column. User have to give this as input.So that user get the output which is suitable for data mining analysis.***

Keywords— ***PaaS, Private Cloud, Middleware, load balancing, resumption of work, E-learning.***

I. INTRODUCTION

Data mining is the tool which is used to extract the useful information in the form of datasets.In a relational database data present in the normalized format. So huge amount of effort required to prepare short summarized data sets as a input for data mining algorithm. Most of the algorithm requires data sets in horizontal layout format, which is not present in available database.That is the problem in models likeclustering,classification,regression and various other algorithms.Different research areas uses various concept to explain data sets.This paper represents a new group of aggregate function that user may used. To create data sets in horizontal format.This helps automation in SQL code writing and extention.In existing SQL capability.In data mining algorithm input is required in the form of table.Extra effort is required for relational database to predict the data in classified form.For obtaining the details of particular application for further analysis data is required in denormalized format.Using the standarg SQL queries users able to perform various aggregation functions on tables and can achieve the output in vertical and horizontal format[6].

This paper describes three horizontal aggregation operators these are SPJ,PIVOT and CASE.SPJ aggregation is using the standard SQL constructs, which are selection, projection and joins. it is the set of SQL operations. PIVOT operator is built in operator in some relational operator and it is used to transform the rows into columns. CASE method can be performed by combining group by and case statement[9].Using this we provide the condition. so we give some extension to normal functionalities to CASE,PIVOT and

ISSN: 2321-9653

SPJ to obtain result in horizontal layout. We have predicted the present method of horizontal aggregation is complex and not efficient to prepare data set and this is challenging problem. Thus we introduced different strategy for their efficient analysis. It is useful to prepare data sets in horizontal layout format [10].

Motivation

As horizontal aggregation can produce output data sets that are useful in various real time applications but it takes more time to give an output in required format.This task requires large and complex SQL code which is very difficult to remember and it requires huge manual effort.There are two important methods used in SQL code:these are JOINS and AGGREGATIONS.Aggreagtion is mostly used to obtain the data sets in summarized form.So we directly go to introduce aggregation[2].Aggregation is defined as collection or gathering of things together, considered as a whole. Oracle provides a number of predefined aggregate functions such as MAX,MIN, SUM,AVG,COUNT for performing a operations on database and among this SUM function is mostly used. Aggregate functions MAX is used to return maximum value.MIN function return minimum value.AVG is used to return average of values.COUNT is used to count the number of rows.Their are certain limitations in preparing the data sets using aggregation function for data mining analysis. Normally the data sets stored in relational database.Comes from real time or online transaction processing systems.Where database tables are present in highly normalized form.But various data mining,machine learning and statistical algorithms requires data in summarized format.When user requires data in horizontal tabular format a large amount of effort is require using current available functions in SQL.User don't get the output for data mining algorithm.Such endeavor is due to large amount of SQL code and its complexity.There are some other issues to obtain aggregate functions in horizontal layout.Some OLAP tools are used to transpose the result.This sometimes said to be PIVOT.PIVOT is more beneficial if it can provide the facilities of aggregating and transposing the rows into column combined together.It is very difficult to get the data sets when there are large number of rows present in database.With consideration of all this limitations,we introduce a new method of aggregate functions that aggregate numeric values of given expression and transpose rows into column so to give horizontal format output.Horizontal aggregation some sort of extension in existing SQL aggregation. Traditional aggregation returns the single value per row but horizontal aggregation returns the set of values [8].

II.AGGREGATION

Database is nothing but the collection of large amount of data.To extract the relevant information or data from various types of sources Structured Query Language is used.Mainly the SQL is used in aggregation of large amount of data.Aggregation is used to combine or aggregate rows over a number of columns.Various aggregation functions are used to gain information in summarized form.Simply it is collection of several things group together consider as whole.In general Database management an aggregation function is a function where the values of multiple rows are grouped together as input on certain criteria to form a single value of more significant meaning or measurement such as a set,a bag,or list[9].

A. Vertical Aggregation

Normal SQL aggregation is same as vertical aggregation. In vertical aggregation result is predict in the form of vertical layout.Result of vertical aggregation contains more number of rows.

B. Horizontal Aggregation

In Horizontal Aggregation result is produce in horizontal layout. To represent output in horizontal format small syntax extension to aggregate function is required. In contrast, we call standard SQL aggregation vertical aggregation since they produce tables with vertical layout[6].The problem of horizontal aggregation number of column may exceed than the

allowed number of column of DBMS.That means reaching the maximum number of maximum column name length when column are automatically named.To elaborate on this, the

III.LITERATURE SURVEY

A.SPJ Method

Left outer join queries are used to join all the projected tables SPJ method can produce tables in horizontal layout an optimized SPJ method can produce more efficient result.The performance of SPJ approach is very low when there is large number of rows .This can perform aggregation with the help of basic SQL quries.This is easier to support by any database. The SPJ method is interesting from a theoretical point of view

ISSN: 2321-9653

because it is based on SPJ method is based on relational operators. Only the main purpose is to create a table with vertical aggregation for each result column and then combine all those tables to produce horizontal table.We aggregate table using select,project,join and aggregation queries. We SPJ method for standard relational algebra operator.We can use left outer join, right outer join and inner outer join[9].Relational operators only.The idea is to create one table with a vertical aggregation for each result column and join all those tables to produce horizontal aggregation [6].

B.CASE Method

For this method we use the"case" programming construct that are present in SQL.case gives us some of the values on the basis of condition from a set of values based on boolean expressions and return values from the selected set of values. CASE statement put the result to NULL when there is no matching row is found. This also produce resultant table in horizontal layout. We produce two basic sub-categories to compute FH[6]. In a similar way to SPJ,the first one directly aggregates from F and the second one computes the vertical aggregation in a temporary table FV and then horizontal aggregations are predicted from vertical aggregation table i.e FV. CASE method can be performed by GROUP BY and condition statement.It is more efficient and wide applicability.CASE statement. We represent the direct aggregation method:Horizontal aggregation queries To overcome those problems in existing system,we are going for our proposed horizontal aggregations which provide several unique features and benefits.this gives us pattern to generate SQL code from this method.this gives us SQL code without writing,minimize them and to test them whether it is correct or not[9].

C.PIVOT Method

PIVOT operator which is a built-in operator in some of the DBMS. This method can transform rows into columns which is known's as transposition which indirectly helps to produce the output in horizontal form[9].The PIVOT method mainly require to determine how many columns are needed to store the transposed relation and it used with the GROUP BY clause. We cannot used single PIVOT operator for that we have to use CASE and SPJ method. PIVOT operator is used with standard select statement by using small syntax extention.PIVOT operator is perform well even though the dataset is very large. The major advantage of PIVOT operator is that it can solved the upper limit limitation of DBMS.

IV. EXISTING SYSTEM

Existing system consist of SPJ, CASE,PIVOT operators. Using SPJ, CASE, PIVOT we can get the result in horizontal layout format but only SPJ or CASE user cannot use it needs PIVOT operator for transposition. And code for PIVOT is so long and hard so it not efficient for data mining algorithms and it is time consuming task. In existing system to creating a data set for analysis is generally requires more time in a data mining project, it needs many complex SQL queries,joining tables and aggregating columns so it becomes a very time consuming task. Existing SQL aggregations have some certain limitations to prepare data sets in data mining because they return only one column per aggregated group. In Existing SQL aggregations a significant manual effort is required to build data sets, where a horizontal layout is required.

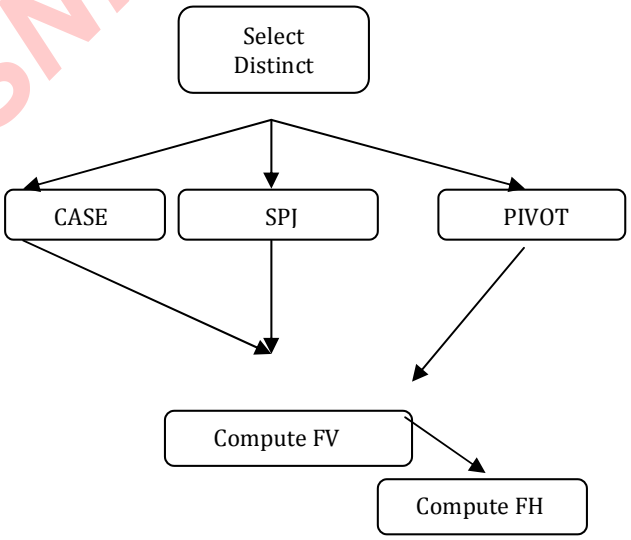

Fig1:Existing System

Suppose we have relations R1....Rk.Then using CASE,SPJ and PIVOT we can compute vertical tabular form. Using CASE and SPJ we cannot easily get table in horizontal format. It needs PIVOT operator for transposition [15].

ISSN: 2321-9653

1. Disadvantage

1. Existing SQL aggregations have limitations to prepare data sets.

2. To return one column per aggregated group

3. Manual effort is required to build data sets.

4. Disadvantage is that vertical aggregation increase the number of rows and columns. Thus increases the complexity.

V. PROPOSED SYSTEM

To overcome those problems in existing system,we are going for our proposed horizontal aggregations which provide several unique features and benefits. It represents a pattern to create SQL code from this method. We get SQL code without writing the code, minimizing them and to test whether it is correct

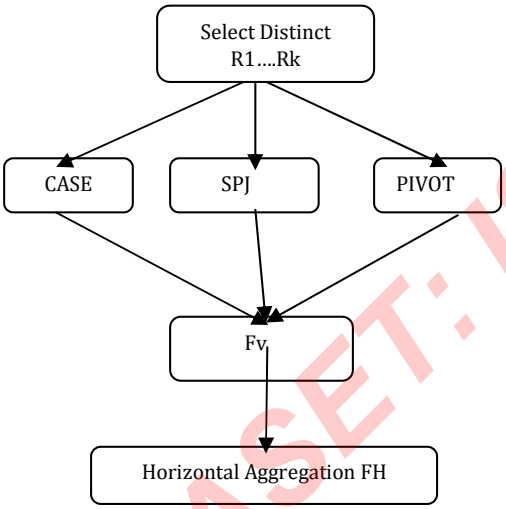

Fig.2:Proposed System

V. CONCLUSION

We proposed a new method of extended aggregate functions i.e extension to standard aggregation function called horizontal aggregations which provide efficient way for preparing data sets which can be given as a input for various data mining algorithms. Output table with horizontal layout is more efficient for creating data sets as mostly required for data mining analysis. Commonly this approach of horizontal aggregation is giving output as set of numbers instead of a single record for each group. We analyzed three query evaluation strategies. The first one SPJ is based on standard relational operators. The second approach of CASE is based on the SQL CASE construct. The third approach PIVOT is nothing but, it is a built-in operator I.e present in some of a commercial DBMS that is not usually available. The SPJ method consists of selection.projection and join queries.CASE construct used by combining GROUP-BY and CASE statements. We proved that these all the three methods giving the same result. Our proposed horizontal aggregations can be used as a database method to automatically generate efficient SQL queries with three sets of parameters: grouping columns, Horizontal columns and aggregated column. The database obtained from horizontal aggregation is analyzed with the help of aggregating column, grouping column, horizontal column and generate the output. This paper present the horizontal aggregation through some method like SPJ,CASE and PIVOT method.

REFERENCES

[1]. PradeepKumar, Dr.R.V.Krishnaiah, IEEE, ”Horizontal Aggregations in SQL to Prepare Data Sets for Data Mining Analysis” vol.2, ISSN: 2278-0661, ISBN: 2278-8727 Volume 6,PP 36-41,Nov - Dec. 2012.

[2]. Mohd Abdul Samad,Md. Riazur Rahman,Syed Zahed, Mohd Abdul Fattah, International Journal of Computer Applications in Engineering Sciences, ”Creation of Datasets for Data Mining Analysis by Using Horizontal Aggregation in SQL” VOL III,ISSN: 2231-4946, pp.46-51, March.2013

[3]. Karana Hanirex.D,Durka.C,International Journal of Advanced Research in Computer Science and Software Engineering, ”An Efficient Approach for Building Dataset in Data Mining” Volume 3,ISSN: 2277,pp.156-160,128X Issue 3, March 2013.

[4]. Carlos Ordonez, Zhibo Chen, University of Houston”Horizontal Aggregations in SQL to Prepare Data Sets for Data Mining Analysis”, pp.1-14.

[5]. Carlos Ordonez and Zhibo Chen, IEEE,”Horizontal Aggregations in SQL to Prepare Datasets for Data Mining Analysis” VOL. 24, NO. 4, pp.678-691, APRIL 2012.

[6]. Mr.Prasanna M.Rathod Prof. Mrs. Karuna G. Bagde, IJARCET”Workload Optimization by Horizontal Aggregation in SQL for Data Mining Analysis” Volume 1, pp.144-147, Issue 8,October 2012.

[7]. Mrs Krishna Veni,Mr Ranjith Kumar K,Int.J.Computer Technology Applications,, ”PREPARE DATASETS FOR DATA MINING ANALYSIS BY USING

HORTIZONTAL AGGREGATION IN SQL" Vol 3(6),ISSN:2229-6093,pp.1945-1949, Nov-Dec 2012.

[8]. B.Susrutha1, J.Vamsi Nath2, T.Bharath Manohar3, I.Shalini4.International Journal of Modern Engineering Research (IJMER),"Horizontal Aggregation in SQL for Data Mining Analysis to Prepare Data Sets" Vol 3, ISSN: 2249-6645,pp.1861-1871,Jul - Aug. 2013.