Next Level Data Warehouse with OLAP and OLTP

Kunal Kumar[1]
[1]M. Tech
[1]Computer science Department
[1]S.R.M university NCR campus, M Mohan Faculty of Department of Computer Science & Engg. India

Abstract---In the data warehouse we are using two important process that is OLTP(Online Transaction Processing) and OLAP(Online Analytical Processing) ,with the help of OLTP in the data warehouse we take data from different type of sources and store in row view in database and after some interval like three or six month that data update in data warehouse because in data warehouse regular updation is difficult because data warehouse contain large amount of data .OLAP is perform analysis data that is present is data warehouse. OLAP help to show our business performance means it show that our business is goes profit or loss or any kind of analysis .But the problem is OLA P take data from data warehouse and in the data warehouse contain historical data because not regular updation.so the problem is analysis report is based on old or historical data and we know business means change every day and result is not in real time. The problem is OLTP and OLAP we work on different type of database .So in this paper we try to solve this problem with the help combining OLTP and OLAP and then get real time result for better business analysis. So because of changing business trend we try to resolve the problem of historical analysis process proposed new techniques that help to analysis historical as well as current data.

Index Terms—In-memory analysis, row store database, Column store database, Database management.

I. INTRODUCTION

Objective of project is merging of OLTP (Online transaction process) & OLAP (Online analytical process) in same database .OLTP is work in DBMS & OLAP is in data warehouse. If we use same database for OLTP & OLAP then the business intelligence get in real time.

Existing Problem, The concept of extracting the data from an OLTP system to OLAP system in various intervals seems to provide data with delay of month or even more.

Proposed Methodology, software technology name In-Memory computing. It is help to merge both OLTP & OLAP in same database.

In data warehouse data are come from different source and also in different form that means every source station may be different format for storing a data and after that because of source data is not in one format ,we are using ETL(Extraction Transformation and Loading) with the help of this we arrange a source data .The OLTP is store data in row oriented database because inserting a new store in the database in row view take less time as compare to column view ,so OLTP work in row oriented database and when we analysis a record column oriented database using because it is more efficient then row oriented database in the term of analysis and data warehouse store data in column store database so it is very difficult to update data warehouse therefor data is transfer from source to data warehouse in a particular interval like one month or more .OLAP perform analysis operation in data warehouse and the data that is store in data warehouse that is one month old even more so analysis report is not in real time , in today's market there is a need to analysis the business as it happens. Even an hour delay of the data for the OLAP system will have impact on the overall decision making .Even a hour delay that effect the business because business trend are frequently change, the nature of customer or buyer is not static it is change very fast habits are change.so if there is possible that is OLAP perform analysis operation in both type of data means historical as well as current data .so result become in real time and business intelligence are also done. So the problem is how we are performing analysis in both data. The current data is available in OLTP and all historical data are available on data warehouse but in OLTP data are store in row view and in data warehouse data is store in column view so the questions is how we take data form different place and different storing format

If it is possible to achieve this type of data warehouse system, it will be "one new combined data warehouse " and this will enable business and users to react to business events more quickly through real –time analysis and all old data and then new business decisions in real time.

Online Transaction Processing (OLTP) systems have been using the RDBMS concept for taking data and perform write operation. Online Analytical Processing (OLAP) system for analytical reporting is optimized to aggregate many records which involve more read operations. Hence to enhance the performance in data warehouse combinations are important. But major problem is both OLTP and OLAP work on different database.

If there is an option to combined OLTP and OLAP systems then the business intelligence system can get the real time data for the analytical query reporting. So we are using one technique that help to combine both and this type of data warehouse become one of new step in data warehouse because data warehouse is in real time analytics.
Objective of project is merging of OLTP (online transaction process) & OLAP (online analytical process) in same database.

The business trends are frequently changing, customer buying habits are constantly changing there is a need for almost real time data which could help to better analysis and to assist in making better decisions which helps organization to be best running business. The concept of extracting the data from an OLTP system to OLAP system in various intervals seems to provide data with delay of day or even less.

Use techniques that help to perform operation in both databases for business intelligence system can get the real time data for the analytical query reporting.

II. PROBLEM DESCRIPTION

RDBMS is best way to manage the business and data will be store in Data warehouse so data store in form of row oriented database and after a fix interval it transfer to data warehouse and for analysis purpose data we take from data warehouse, so the problem is data is in data warehouse that is not update day by day so the analysis report is not in real time because data that present in data warehouse is not in real time or current data. This is very big problem, OLAP report are wrong because of data is transfer in some interval.

III. PROPOSED METHODOLOGY

The data is coming and store in row oriented database because in row oriented database insertion is easy and if any new row want to insert in database then only one row to update and analysis or any aggregation function is easy to apply in column oriented database because analysis (max, min, avg etc) is perform only one column so apply both database in data warehouse. In data warehouse, problem is regular update is not possible. So we use one memory in between OLTP database and data warehouse and the size of memory depend on what is the time interval for perform updation in data warehouse so particular that time of period how much memory is used, that is the size of memory. We define a name of memory that is called temp/change memory. In temp/change data is store in column-oriented, regular updation is perform from OLTP to temp/change memory. So when OLAP perform analysis operation in data warehouse it first take data from temp/change memory that contain all current data and after it take data from data warehouse that contain old or historical data .so after the analysis the result is in real time because we analysis both type of data means old data as well as historical data.

IV. PRESENT DATA WAREHOUSE MODEL

A data warehouse is simple a collection of pieces of information in a summerized or detail way that manage and direct the business for the most profitable outcome.

In definition "The data warehouse is more than just data, it is also the process involved in getting the data in the form of table to analysis."

In data warehouse architecture, first we understand how data is store in data warehouse.

Data warehouse architecture

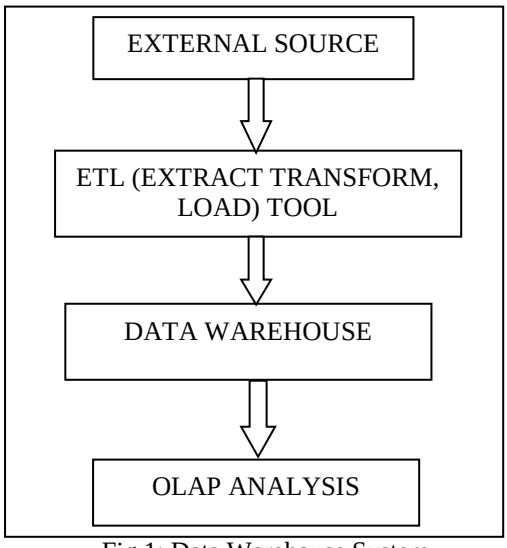

Fig.1: Data Warehouse System

In the first part there is source system e.g. people soft, sap, Siebel ,oracle application .extraction of data from different source like ERP(enterprise resource planning) a,d CRM(customer relationship management) and other type of operational system .

In second part that is ETL tool for extraction loading and transforming .OLTP is working in this layer There are different type of source like ERP or CRM all are store data in different format, so ETL tool is first extract the data after that perform transform that convert all different type of format in one format and in last load the data.

In third part that is data warehouse means repository layer, in this layer data is store in data warehouse .in this part the responsibility of ETL tool to load data in data warehouse. There are different tool like oracle, Sql server, Teradata, Sybase IQ etc

In fourth part that is presentation layer ,in this layer OLAP ,data mining ,data mart are working .In this layer OLAP perform analysis process ,OLAP servers deliver warehouse applications such as performance reporting, sales forecasting, product line and customer profitability, sales analysis, marketing analysis, what-if analysis and product line and customer profitability, sales analysis, marketing analysis etc.

With OLAP servers' robust calculation engines, historical data is made more useful by transforming it into derived and projected data.

data is coming from different source like different type of system or ERP (enterprise resource planning) or many type of external source after that ETL (extraction transform load) are apply and convert data in reader and in a one format .The work of data warehouse is store data in efficient manner and use to analysis that done with the help of OLAP.

V. CHANGE IN CURRENT MODEL

Use same database for the OLTP and OLAP systems then the business intelligence system can get the real time data for the analytical query reporting because Online Transaction Processing (OLTP) systems have been using the traditional relational Database management system for many years. Online Analytical Processing (OLAP) system for analytical reporting is optimized to aggregate many records which involve more read operations. Hence to enhance the performance in OLAP systems various modeling options like star schema, extended star schema was implemented. But major problem is both OLTP and OLAP work on different database.

A. *A step towards next generation data warehouse design*

- Use same database for the OLTP and OLAP systems then the business intelligence system can get the real time data for the analytical query reporting .Perform migration of an OLTP system to my SQL and comparative performance evaluation.
- IN-MEMORY Technology- In memory technology move data and information source from remote database into local memory so the result of analysis and transaction are immediately
- ROW Vs COLUMN Data storage- Relational database typically use row based data storage. However, column storage database is more suitable for some business application.

In row store data are stored in the disk tuple by tuple.

Where in column store data are stored in the disk column by column

Most of the queries do not process all the attributes of a particular relation.

For example the query

Select c.name and c.address
From CUSTOMES as c
Where c.region=Mumbai;

Only process three attributes of the relation CUSTOMER. But the customer relation can have more than three attributes.

Column-stores are more I/O efficient for read-only queries as they read; only those attribute which are accessed by a query.

Row Store - Easy to add/modify a record
- Might read in unnecessary Data

Column store - Only need to read relevant data
-Tuple write require multiple accesses

So column stores are suitable for read-mostly, read intensive, large data repositories.

Why Column Stores?

- Can be significantly faster than row stores for some applications
- Fetch only required columns for a query
- Better cache effects
- Better compression (similar attribute values within a column)
- But can be slower for other applications
- OLTP with many row inserts
- Long war between the column store and row store camps

This paper tries to give a balanced picture of advantages and disadvantages, after adding/ subtracting a number of optimizations for each approach

B. *COMPERESSION*

1) Trades I/O for CPU

- Increased column-store opportunities:
- Higher data value locality in column stores
- Techniques such as run length encoding far more useful

2) Schemes

Null Suppression
Dictionary encoding
Run Length encoding
Bit-Vector encoding
Heavyweight schemes

Table-Employee

	EMP ID	Name	Area
Row1	A101	Kumar	UP
Row 2	A102	Kunal	Delhi
Row 3	A103	Himan	Delhi
Row 4	A104	Aman	Rajesthan

Row Oriented in Memory

A101	Kumar	UP	A102	Kunal	Delhi

Column Oriented in Memory

A101	A102	A103	A104	Kumar	Delhi

VI. THE PROPOSED ARCHITECTURE

In the architecture data come from different source and store in a one memory according to the row oriented database the size of memory depend on the how much time interval for transfer the data in memory to data warehouse .After coming data(source data) store in row oriented database and then we have taken one another memory that work according to column oriented database and data transfer to row oriented database names 'write' to the column oriented database names 'delta' so overall now data is store according to column oriented database that means column view .Now in 'delta' memory have current data and in the data warehouse have historial data so if analysis the business then we take data from both place means from data warehouse that contain historial data and 'delta' memory that contain present or current data so analysis perform both of data historial and present so result is in real time and business become intelligence .

So we remove the problem of analysis result that not in real time with the help of using column/row oriented database

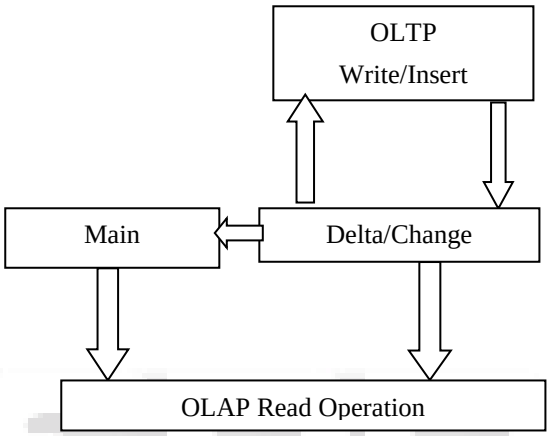

Fig.2: In-memory columnar Database

A. *In the model*

"write/insert" work in row oriented database.
The "delta/change" work in column oriented database.
The "main memory" work in column oriented database.
OLAP work is analysis the data.

B. *In data warehouse is buffer/memory, in memory we store data in two types.*

<1>Row oriented database
<2>Column oriented database

All update, insert is done in update/insert buffer that is work in row oriented database.

All aggregation (max, min, ave, count, sum) is done in "delta/change" that is work in column oriented database.

1) STEP 1: OLTP want data then fetching the record from different source.
2) STEP 2: OLTP use and store data in buffer as row oriented. (Write/store)
3) STEP 3: write/store buffer transform data to delta/change as column oriented.
4) STEP 4: main memory (data warehouse) is also store data in form of column oriented.
5) STEP 5: OLAP is perform analysis in both main memory and "delta/change" buffer.
6) STEP 6: finally analysis result is in real world because main memory have historial data and delta buffer have current data.
7) STEP 7: finally business is in real time.

In the result we get the answer of the query ,result is in real time because we analysis the both of the area that means the regular update data that is store in delta and the

historial data that is store in the main memory (data warehouse). OLAP check both areas so our business is in real time. Use same database for the OLTP and OLAP systems then the business intelligence system can get the real time data for the analytical query reporting basically OLTP are capturing or inputing the data and OLAP is utilizing the data. OLTP operation perform row view and OLAP operation perform column view .Finally result is in real time

VII. CONCLUSION

Now business need to work on real time and the traditional system is not in real time so result is not depend on current data so we have a analysis report according to only historial data .In this system result is in real time because we analysis the both of the area that means the regular update data that is store in delta and the historial data that is store in the main memory (data warehouse). OLAP check both areas so our business is in real time. Use same database for the OLTP and OLAP systems then the business intelligence system.

VIII. ACKNOWLEDGMENT

Our sincere thanks to M. Mohan, Dept. of Computer Science SRM University NCR Campus. Our thanks to Hasso Plattner Institute for IT Systems Engineering articles that helped us a lot to explore more in this area.

REFERENCES

[1] Hasso plattner, hasso plattner institute for it systems engineering, a common database approach for oltp and olap using an in-memory column database.
[2] W. H. Inmon. Building the Data Warehouse, 3rd Edition. John Wiley & Sons, Inc., New York, NY,USA, 2002
[3] Jens Krueger, In-Memory Data Management for Enterprise Applications, Hasso Plattner Institute for Software Engineering
[4] Whitepaper, SAP HANA® Databas e for Next-Generation Business Applications and Real-Time Analytics
[5] Jens Krueger, In-Memory Data Management for Enterprise Applications. Hasso Plattner Institute for Software Engineering
[6] OLTP and OLAP data integration, Samuel s.comm Regis University school for professtional studies.
[7] M. Stonebraker et al. C-Store: A Column-oriented DBMS. In Proc. VLDB, 2005.

.

