TO GET BEST PERFORMANCE FOR ANALYTICS DATABASE SYSTEMS BY MERGING "IN MEMORY DATABASE"AND "COLUMN ORIENTED DATABASE" TECHNOLOGIES

Dr. Raghav Mehra
Associate Professor & Dean,
Bhagwant Institute of Technology,
Bhagwantpuram,Muzaffarnagar U.P., India.

Ram Babu
Ph.D. Research Scholar,
Bhagwant University, Ajmer, India.

Received Feb. 27, 2016 Accepted March. 10, 2016

ABSTRACT In order to get best performance for an analytic system or data warehouse systems, two technologies, column oriented database management systems and main memory database management system can be combined to get advantages of these two. Both technologies give its best performance to opponent database system, for example Main memory database management systems are faster due the fact that these reside in main physical memory as compared to disk resident database systems. This is because RAM is faster than hard drive/disk. The main memory databases systems are100 timesfaster than that of disk resident databases. The column oriented databases are faster than the row based database because in column oriented databases, data is stored in columns and indexed as compared to row based database systems. Similarly column based database systems are 100 times faster than that of row based database systems By combining these two technologies, we can achieve 200 times better performance for analytics systems they needs to be high performing in analysis and computation.

Key words: Analytics, Column Based Database, Columns, Data Warehouse, Database, database management, Hard Disk, Main memory database, Memory, real time database, storage memory and products,

1. DEFINITION

1.1 Analytical database Systems

Analytic database or Analytical database is a database system which stores historical read only data on business dimensions such for example revenue for each quarter each year or profit of organization quarter on quarter or year on year etc. Database can be queried by business users which have access to generate and analyze report based on their own custom requirements or they can generate pre-defined reports based on the SQL statement they have to generate reports.

These databases are a component of data warehouse systems or data mart, specially designed to support decision support system such as business intelligence and analytical application.

These are different from online transactional database and have better read and analytical performance while executing a report.

There are five analytical database solutions available

i. Column Based Database Systems
ii. Main Memory Database Systems
iii. Data warehouse Applications
iv. Massively Parallel Processing Database Systems
v. Online Analytical Processing Database Systems

1.2 Main Memory Database

Main memory database system (MMDB) also called real-time database (RTDB) or in memory database (IMDB) is a database management system that reside in main memory rather disk. Since physical memory is considerably faster than a hard disk or a solid state disk, which are faster than hard disk, complex decision support queries can be satisfied much more rapidly and high-end computers can be configured with terabytes of memory. Due to faster speed of RAM, In-Memory Database are way faster in comparison of disk storage database systems it is contrasted with database management systems which applies to a disk storage mechanism. In-memory databases are faster than disk-optimized databases due to the internal optimization algorithms are simpler and execute lesser CPU instructions.

Accessing in memory data reduces the Disk seek whilereading the data which provides quicker and better predictable read performance in comparison that of disk.

1.2.1 ARCHITECTURE

Architecture of main memory database system is shown in Figure 1

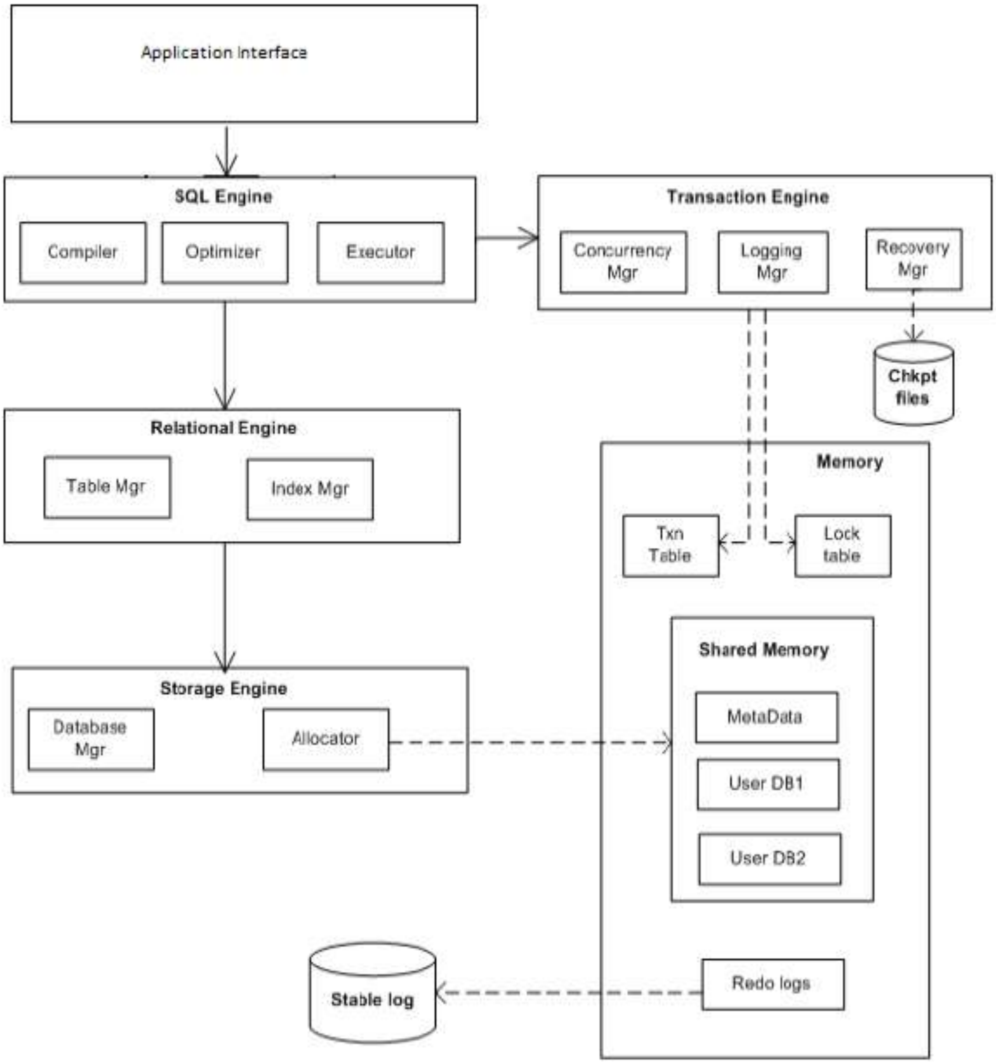

Figure 1

1.2.2 Advantages of In Memory Database Management Systems over Disk StorageDatabase Systems

1. Another variation is to have huge amounts of nonvolatile memory on hosts/servers. E.g. Flash memory chips as addressable memory rather than structured as disk arrays. Databases in this form of memory combines very quick access speed with persistence over reboots and power losses.
2. In-memory databases are often used for applications which demandshigh response time is, such as mobile advertising networks and telecommunications network equipment.
3. IMDBs have gained a huge traction, predominantly in the data analytics, beginning mid-2000s mainly due to cheaper RAMs.
4. Apart from providing extremely fast data read / response times, in-memory analytics eliminate the need of data indexing and storing of pre-aggregated data in OLAP cubes or aggregate tables. This capacity minimizes IT costs and enables faster implementation of BI and BA applications.
5. The developments in recent past have made in-memory analytics database increasingly possible: 64-bit computing, multicore hosts and cheaper RAM prices
6. Main memory databases store data on volatile memory devices. These devices lose all stored information due to power off or is reset. In this case, MMDBs can be said to lack support for the durability of the ACID (atomicity, consistency, isolation, durability) properties. The Volatile memory-based MMDBs can, and often do, support the atomicity, consistency and isolation of ACID properties.

1.3 Column Oriented Database

A Database Management Systems which stores its contents/data in columns instead of rows. This is all about data store technology, here database column separately so that the attributes belonging to the same column would be stored contiguously, compressed and packed densely in the disk. The advantage of this technology is in reading the records faster as compared to classical row-stores where in rows are stored one after another in the disk.

These databases are more beneficial for analytics systems to get analysis done faster as data is stored in columnar form. Indexes are much faster in column oriented databases hence the faster data retrieval for data analysis. This is an alternate database technology over row oriented databases.

There are two technologies to map database tables onto a one dimensional interface: store the table row-by-row or store the table column-by-column. The row-by-row technology keeps all information about an entity together.

Take an example of author table

Author Table

Author ID	Author Name	Age	City
1	Ram Singh	30	Pune
2	Rudra Singh	4	Delhi
3	Riddhima Singh	3	Faridabad
4	Nirmal Singh	30	Faridabad

Table 1

Row by row storage technology

Author table will stored as below using row-by-row methodology or technology storage
1, Ram Singh, 30, Pune; 2, Rudra Singh, 4, Pune; 3, Riddhima Singh, 3, Faridabad; 4,Nirmal Singh, 30, Faridabad;

Column –by-column storage technology

Author table will stored as below using row-by-row methodology or technology storage 1, 2, 3, 4; Ram Singh, Rudra Singh, Riddhima Singh, NirmalLodhi; 30,4,3,30, Pune, Delhi,Faridabad, Faridabad;
In the author example above, it will store all information about the first Author, and then all information about the second Author, etc.

The column-by-column approach keeps all attribute information together: the entire author id will be stored consecutively, then all of the author names, et cetera. Both approaches are reasonable designs and typically a choice is made based on performance expectations. If the expected work load tends to access data on the granularity of an entity, then the row-by-row storage is preferable since needed information will be stored together.

Column based database management systems work by advancing the simple nature of key or value based ones. Despite their complicated-to-understand, these databases functions very simply by creating collections of key / value pairs that match a record.

Unlike the traditional database defines schemas of relational databases, column-based solutions do not require a pre-structured tables to work with the data sets. Each record comes with one or more columns containing the information and each column of each record can be different.

Basically, column-based NoSQL databases are two dimensional arrays whereby each key has one or more key / value pairs attached to it and these management systems allow very large and un-structured data to be kept and used (e.g. a record with tons of information).

These databases are commonly used when simple key / value pairs are not enough, and storing very large numbers of records with very large numbers of information is a must. DBMS implementing column-based, schema-less models can scale extremely well.

Column-oriented databases are very cool, and they have a role to play in data that is much larger than available RAM, or where naive aggregates or stream processing is paramount. They’re often found for specific purposes in the financial industry, and there are some absolutely fantastic products.

1.3.1 Advantages of Column Oriented Database Management Systems over Row OrientedDatabase Systems

1. High performance on aggregation queries (like COUNT, SUM, AVG, MIN,MAX)
2. Highly efficient data compression and/or data partitioning
3. True scalability and fast data loading for Big Data
4. Accessible by many 3rd party BI analytic tools
5. Fairly simple systems administration

Data warehouses Applications, Massively Parallel Processing Database Systems, online Analytical Processing Database Systems are out of scope of this article.

2. PRODUCTS

Couple of products available in market is give below

2.1 SAP HANA

HANA DB takes advantage of the low cost of RAM, data processingabilities of multicore CPUs and fast data retrievalfrom solid-state drives relative totraditional hard drives to provide better performance of analytical and transactionalsystems. It enables a multiengine query processing environment that allows it tosupport relational databases (with both row-oriented and column-oriented physical representations within ahybrid engine) as well as text and graph processing for semi and unstructured databasesystems in the same system. HANA follows 100% ACID properties.While HANA has been called variously an acronym for HAsso's New Architecture (this is areference to SAP founder HassoPlattner) and High Performance Analytic Appliances,HANA is a name not an acronym.

2.2 EXASolution

EXASolution is a parallelized relational DBMS runs on a cluster of standardhardware hosts. Following the SPMD model, identical code isexecuted simultaneously on every node. The data is stored in a columnar way and proprietary In-Memory compression methods are used. Company claims that tuning efforts are notrequired as the database includes self-optimization (likeautomatic indices, statistics, and distributing of data).EXASolution is designed to run In Memory, although data is stored persistently on disc following the ACID rules. EXASolution supports the SQL Standard 2003and can be integrated through standard interfaces like JDBC, ODBC or ADO.NET.Additionally, a SDK is provided for native integration.License model is based on the RAM allocated for the database software (per GBRAM) and independent to the physical infrastructure. Customers get the maximumperformance if compressed active data fits into that licensed RAM, but it can also behuge.

EXASOL implemented a Cluster Operating System (EXACluster OS). It isbased on standard Linux and have supports parallelism functionality. It can becompared to Virtualization, but instead of virtualizing the hardware of a server, itvirtualizes a cluster of nodes to a single node. The Cluster management algorithms areprovided such asfailover mechanisms or auto cluster installation.

3. CONCLUSION

There has been always a requirement of a databases which provides faster query execution and analysis for business purpose. As the size of databases are increasing dayby day, the query execution and performance of these databases is getting slower and sloweror the maintenance cost is going higher and higher.

To overcome these issues, we can take advantages of two technologies called column based database and Main Memory Database management systems. This way we can achieve adatabase management system having 200 times higher performance in comparison of DRDB/ RODBs.This has been achieved by Oracle Corporation by implementing these two technologies in oracle 12 con exadata platform is just an example to show the improved performance.

REFERENCES

1. O.S. Tezer on “A Comparison Of NoSQL Database Management Systems And Models” February 21, 2014
2. Dennis Forbes on “The Real Advantage of Column-Oriented Databases” April 4, 2014
3. Joseph Foley on “Comparison of Data Warehousing DBMS Platforms” , November 10, 2013
4. John Appleby , on “What In-Memory Database for SAP BW? A Comparison of the Major Vendors”, November 13, 2013
5. Hector Garcia-Molina, Kenneth Salem on “Main Memory Database Systems: An Overview” ed vol. 4, NO 6. December 1992 IEEE

6. Stephen Swoyer, article “What "In-Memory" Truly Means” in TDWI on May 21, 2013, pp 1-3.
7. Stephen Tu, Wenting Zheng, Eddie Kohler† , Barbara Liskov, and Samuel Madden MIT CSAIL and †Harvard University in “Speedy Transactions in Multicore In-Memory Databases”
8. Mohit Kumar Gupta , Vishal Verma , Megha Singh Verma on “In-Memory Database Systems - A Paradigm Shift” (IJETT) – Volume 6 Number 6- Dec 2013
9. Hector Garcia-Molina, and Kenneth Salem, “Main Memory Database Systems: An Overview”, IEEE Transactions on Knowledge and Data Engineering, Vol. 4, No. 6, pp. 509-516, Dec. 1992.
10. Main Memory vs. RAM-Disk Databases, McObject LLC. (2003) [Online]. Available: http://www.mcobject.com/130/EmbeddedDatabaseWhitePapers.htm.
11. Y. Huang, Y. Zhang, X. Ji, Z. Wang, and S. Wang, “A Data Distribution Strategy for Scalable Main-Memory Database”, Advances in Web and Network Technologies, and Information Management, Lecture Notes in Computer Science, Vol. 5731, pp. 13-24, 2009.
12. Gorine, “Building durability into data management for real-time systems” boards and Solutions pp 20-21
13. Daniel J. Abadi Yale University, New Haven, CT , Peter A. Boncz CWI, Amsterdam, The Netherlands, Stavros Harizopoulos HP Labs, Palo Alto, CA on “Column-oriented database systems” Volume 2 Issue 2, August 2009 , pp 1664-1665
14. S. Ghemawat, H. Gobioff, and S.-T. Leung.: The Google file system. SIGOPS Oper. Syst. Rev., vol. 37, pp. 29-43(2003)
15. Chin-Chao Huang, WenchingLiou National Chengchi University, Taiwan on “A STUDY ON THE TRANSLATION MECHANISM FROM RELATIONAL-BASED DATABASE TO COLUMN-BASED DATABASE”
16. VandanaBhagat ; Arpita Gopal on “Comparative Study of Row and Column Oriented Database”
17. Gheorghe MATEI, Column-Oriented Databases, an Alternative for Analytical Environment, Database Systems Journal vol. I, no. 2/2010
18. Gheorghe MATEI on “Column-Oriented Databases, an Alternative for Analytical Environment”, Database Systems Journal vol. I, no. 2/2010
19. Daniel J. Abadi on “Query Execution in Column-Oriented Database Systems”, February 2008 , pp 17-38.

Seeing yourself as you want to be is the key to personal growth.
~ Anonymous