International Journal of Engineering & Technology, 8 (1.6) (2019) 44-46

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Data Warehouse Performance Efficiency using Snowflake and Bitmap Index

M Benjelloun[1], M El Merouani[2], E Aoulad Abdelouarit[3*]

[123] *Laboratory modeling and information theory Abdelmalek Essaadi University, Tétouan, Morocco*
**Corresponding author E-mail: elamin@hotmail.es*

Abstract

Designing a Data warehouse is the main mission done by DBA; Testing the data warehouse performance efficiency is also done while going to production. Actually, hardware has developed in term of memory volume, speed and storage, but for the specialists, the objective is still to minimize the execution time and the storage space. The paper shows that using bitmap index and partitioned fact tables in big data warehouses volumes based on a snowflake schema is advantageous based on query execution time.

Keywords: *Data warehouse, Business Intelligence, Index.*

1. Introduction

As is considered the first preoccupation for designers. The Data warehouse efficiency depends on the SQL running delay.
To do this, the data warehouse conception is the supreme task for the designers, and how to optimize access to it data.
Toward minimizing the time execution, the type of conception should be optimized (number of joins). So, we attempt in improving query execution query time using snowflake and bitmap index in comparison with the star schema.

2. Using analytic process

The relational scheme of the data warehouse is opted, which is composed of metrics and labelling data. The metrics are the business calculations and a link to each label. The role of this is the reuse of production DBMS which minimize the implementation charges [7][9].
We will compare between two conception types: Star and Snowflake.

3. Snowflake conception

It reveals the hierarchies linked with each dimension. And it is divided into multiple classifications. This are optimizing the labels measurement [7][8].
All representing the labels measurements are linked to the metrics

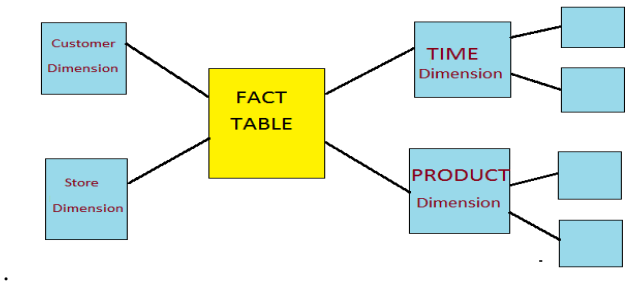

Fig. 1: Sample of Snowflake conception

4. Thesis

Fast access to data is satisfied by using the bitmap index [1][2][3][4][6], but when the fact table has large data, the star schema is not efficient [5]: small count of relations and hence faster SQL running, for that, tables should be partitioned, this will minimize the table size: data will be distributed in many tables, then small foreign keys number.

To get it, we should split the metrics table in multiple portions in function of label measurements, For the timing dimension, it should be divided chronologically (in months, years, decade…).
Our Goal is to show the use of index type bitmap and dividing tables is more efficient than other configuration with a star schema.
We opt for Microsoft SQL Server as RDBMS to achieve this experimentations (Two homogenous environments).

5. Analysis and Results

A big Data warehouse volume sample Schemas has been created to satisfy our study:
-Star conception: one metrics table, three labels tables.
-Snowflake conception: 1 metrics table, 3 labels tables and 2 sub lables under dim1.
-10 million insertion in both metrics tables.
Some queries have been used for both data warehouse scheme in order to get an intense analyse.

5.1. lab environment specs:

It has been realized in the same IT specifications: (Table 1)

Table 1: IT specs of experimentation lab

RAM	HDD	OS	RDBMS
16 Gb	8 TB	MS Windows 2012 Srvr	Microsoft SQL 2014

Table 2: Star join query used

```
SELECT dim1.a2, dim3.a2, dim1.a1, SUM(f.msr1)
from Fct1 f inner join dim1 on dim1.dimkey1 = f.dmkey1
GROUP BY dim1.a1
```

Table 3: Snowflake join query used

```
SELECT  Fct2.Fct2_attr1,  Fct2.Fct2_attr2,Fct2.Fct2_attr3,  dim11.dim11_attr1,  Fct2.Fct2_attr3, Fct2.Fct2.a3
FROM  dim11 LEFT JOIN  d1
ON  dim11.dim11_att1 = dim1.dim11_att1  LEFT JOIN Fct2 ON dim1.dim11_att1 = Fct2.Fct2_att1
GROUP BY Fct2.Fct2_att1;
```

5.2. Used Queries

Otherwise, the both SQL statements return same results.
We should apply pressure on these 2 conception types before and after introducing the bitmap index and dividing table to star schemas, this will be called in our paper as optimum jobs.
Our comparison will be based on SQL running delay and RAM exploitation.

5.3. Results Pre-optimum jobs

Before applying the optimum jobs, both SQL Statement have been tested; and results are showed in the next tables: (Table 4) (Table 5)

Table 4: Results before applying the optimum jobs

SQL statement	Occupied RAM	Running Time
Star + Bitmap	48%	40 min
Snowflake + Bitmap	69%	59 min

After the results analysis:
-Snowflake conception is advantageous when labelling table is large in size, because of space occupation reduction.
-In a large data case, Query performance is slower when using star schema.
-The star conception uses more RAM than the snowflake one because of data redundancy.

5.4. Results Post-optimum jobs

Following results, after dividing table in a star conception based on time dimension, and used bitmap index:

Table 5: Results before applying the optimum jobs

SQL statement	Occupied RAM	Running Time
Star + Partitionned Tables + Bitmap	45%	33 min
Snowflake + Partitionned Tables + Bitmap	40%	21 min

After applying modification, we found that in the big data warehouses case, the snowflake schema is recommended to use.

5.5. Final Analyse

The RAM occupation, and the SQL running time become ultimate after dividing tables and using bitmap index for:
-Smaller data on dividing per labeling
-multiples dimensions' level
-Bitmap index effectiveness because of big data and smaller different values.
Finally, dividing tables and the use of bitmap index are considered some tune operations that allows a best exploitation of snowflake conception on a big data warehouses volume.

6. Conclusion

To conserve it lead like quick access to data by many level of dimensions, we have to opt of the utilisation of the snowflake conception in our data warehouse.
Actually, most of data warehouses have the target to analyse the business activity.
So if data warehouses are designed by organization activity, they are more efficient. But our paper target is to show that the snowflake schema might be deployed even in big data cases, to create "data marts" time oriented we should use the bitmap index and partitioning metrics.

References

[1] M. Benjelloun, M. El Merouani and E. A. Abdelouarit. Using Snowflake Schema and Bitmap Index for Big Data Warehouse Volume. International Journal of Computer Applications 180(8):30-32, December 2017
[2] M. Benjelloun, M. El Merouani and E. A. Abdelouarit. Impact of using Snowflake schema and Bitmap index on Data warehouse querying. International Journal of Computer Applications 180(15):33-35, January 2018
[3] E. Abdelouarit, M. El Merouani, A. Medouri, The bitmap index advantages on the data warehouses American Academic & Scholarly Research Journal Vol. 6, No. 4, July 2014
[4] E. Abdelouarit, M. El Merouani, A. Medouri, Data Warehouse Tuning: The Supremacy of Bitmap Index International Journal of Computer Applications (0975 – 8887) Volume 79 – No7, October 2013.
[5] E. Abdelouarit, M. El Merouani, A. Medouri, The impact of indexes on data warehouse performance IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013.
[6] E. Abdelouarit, M. El Merouani, A. Medouri, OPTIMISATION DES PERFORMANCES DES ENTREPÔTS Rev. Ivoir. Sci. Technol., 20 (2012) 35 - 67 ISSN 1813-3290, http://www.revist.ci
[7] S. Chaudhuri, U. Dayal,An Overview of Data Warehousing and OLAP Technology., ACM SIGMOD RECORD. 1997
[8] E. E-O'Neil and P. P-O'Neil, Bitmap index design choices and their performance impli-cations, Database Engineering and Applications Symposium. IDEAS 2007. 11th International, pp. 72-84.
[9] R. Kimball, L. Reeves, M. Ross, The Data Warehouse Toolkit. John Wiley Sons, NEW YORK, 2nd edition, 2002.
[10] W. Inmon, Building the Data Warehouse., John Wiley Sons, fourth edition, 2005.
[11] C. DELLAQUILA and E. LEFONS and F. TANGORRA, Design and Implementation of a National Data Warehouse. Proceedings of the 5th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases, Madrid, Spain, February 15-17, 2006 pp. 342-347.
[12] R. Kimball and K. Strehlo. Why decision support fails and how to fix it. SIGMOD Record, 24(3) :92–97, September 1995.
[13] E. Sidi, E. Abdelouarit, M. El Merouani, Star Schema Advantages on Data Warehouse: Using Bitmap Index and Partitioned Fact Tables, International Journal of Computer Applications (973-93-80890-95-3) Volume 134 – No13, January 2016.
[14] E. Sidi, E. Abdelouarit, M. El Merouani, The Impact of Partitioned Fact Tables and Bitmap Index on Data Warehouse Performance, International Journal of Computer Applications (973-93-80891-16-1) Volume 135 – No13, February 2016.