



Flow enhancement between Customer relationship management (CRM) and Data warehouse

Prayag Tiwari

Masters in Computer Science from MISIS, Moscow

ABSTRACT

“CRM is routine that integrates ideas of knowledge management, data mining, and data warehousing so as to support an organization’s determination-making method to maintain long term relationships with their customers. Better CRM analysis require a point by point data warehouse model that can bolster different CRM analysis and profound comprehension on CRM-related business questions. This exploration is a piece of study to analyze methodically CRM components that influence plan choices for CRM data warehouses to manufacture a scientific classification of CRM analysis and to decide the effect of those analysis on CRM data warehousing outline decisions. ETL is the center procedure of data integration and is normally connected with data warehousing. If source data taken from various sources is not cleanse, extracted, transformed and integrated in the proper way then our flow from CRM to Data warehouse will decrease. In this paper we presented advance approach by Query Cache which will increase ETL process then our flow enhancement will automatically increase.”

KEYWORDS : CRM (Customer Relationship Management), ETL (Extraction, Transformation and Load), Data Warehouse

1.0 INTRODUCTION

Clients are the key asset in accomplishing diverse objectives. So organization perceive the significance of enhancing their capacity to oversee client adequately. With the advancement of economy, ventures need to face more intricate and questionable business sector environment. As needs be, the past effective management concerning capitals or common asset is insufficient for undertakings to survive, create and achieve enough upper hand in rivalry, since client has turned into the conclusive principle asset, order to balance out existing customer and create potential customer, quick reaction and fast choice making are critical to achievement. This is the place data warehouse and CRM comes into the photo. Data warehouse is intended to bolster choice making. It contains both present and recorded data that are helpful for leaders and managers. Customer relationship management (CRM) is a management idea created by current need of modern organizations. Integration of Data warehouse and Customer relationship management (CRM) is a superior technique to achieve aimed stabilization in organizations.

2.0 CRM (Customer Relationship Management), Data warehouse and ETL:

Customer Relationship Management is an arrangement of techniques performed by an organization or association (business and non-business) to oversee and break down data about their clients, to stay in touch and correspond with their clients, to pull in and win new clients, to market items/administrations and behavior exchanges with their clients and to administration and backing their clients. It consist, guidelines, procedures, processes and strategies which provide organizations the ability to merge customer interactions and also keep track of all customer related information. The idea of Data Warehouse initially showed up in 1980s. It was not until 1993, William H. Inmon, “father of data warehousing”, firstly efficiently advances the thought of data distribution center and related hypothesis in his book Building the Data warehouse, and it is the establishment of data warehouse improvement. The attributes of a data warehouse were initially characterized by W.H. Inmon who expressed, “a data warehouse is subject-oriented, integrated, time-variant and non-volatile [data] collection in support of management decision making processes”. Contrasted and customary exchange parade framework, data warehouse center is a procedure which is situated in the endeavor where business data are incorporated, prepared and analyzed.

ETL (Extract, Transform and Load) is a process in data warehousing responsible for pulling data out of the source systems and placing it into a data warehouse. ETL involves the following tasks [3]:

- **Extract-** extracting the data from source systems (SAP, ERP, CRM and other operational systems), data from different source systems is converted into one consolidated data warehouse format which is ready for transformation processing.

- **Transform-** Transform is the process of converting the extracted data from its previous form into the form it needs to be in so that it can be placed into another database..
- **Load-** loading the data into a data warehouse or data repository other reporting applications

3.0 Decision Support System (DSS)

A decision support system (DSS) is a PC program application that breaks down business information and presents it with the goal that clients can settle on business decisions all the more effectively. It is an “informational application” (to recognize it from an “operational application” that gathers the information over the span of ordinary business operation).

4.0 Overall Flow Chart Diagram from Pre-Data Warehouse to Front-End Analytics

Data Warehouses are an imperative resource for organizations to look after effectiveness, profitability and upper hands. Organizations gather data through numerous sources - Online, Call Center, Sales Leads, Inventory Management. The data gathered have degrees of quality and business significance. As data is gathered, it is gone through a ‘transport line’, call the Data Life Cycle Management. Basically, Data Warehouses store a collection of an organization’s data. An organization’s data life cycle administration’s strategy will manage the data warehousing configuration and approach. This data and life cycle management is taken from [1]

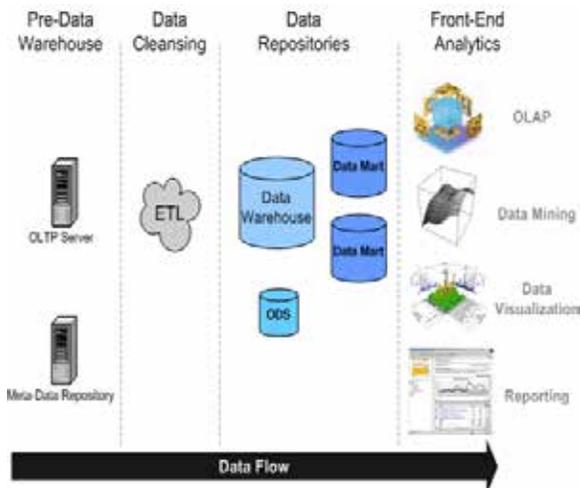


Figure 1 .Overview of Data Warehousing Infrastructure

The objective of Data Warehousing is to create front-end analytics that will bolster business administrators and operational managers.

4.1 Pre-Data Warehouse

The pre-Data Warehouse zone gives the data to data warehousing. Data Warehouse designers figure out which data contains business esteem for insertion. OLTP (Online transaction processing) databases are the place operational data are saved.

Metadata guarantees the sacredness and exactness of data going into the data life cycle process. Meta-data guarantees that data has the right organization and relevancy.

4.2 Data Cleansing

Before data enters the data warehouse, the extraction, transformation and cleaning (ETL) process ensures that the data passes the data quality threshold. ETLs are also responsible for running scheduled tasks that extract data from OLTPs.

4.3 Data Repositories

The Data Warehouse repository is the database that stores dynamic information of business quality for an organization. The Data Warehouse demonstrating configuration is upgraded for data analysis.

There are variations of Data Warehouses - Data Marts and ODS. Data Marts are not physically any not quite the same as Data Warehouses. Data Marts can be however of as littler Data Warehouses based on a departmental as opposed to on an company-wide level.

ODS are utilized to hold information that have a more profound history that OLTPs. Keep a lot of information in OLTPs can secure PC assets and moderate down handling - envision sitting tight at the ATM for 10 minutes between the prompts for inputs.

4.4 Front-End Analysis

Data Mining is the revelation of helpful pattern in data. Data mining are utilized for expectation analysis and order - e.g. what is the probability that a client will relocate to a contender.

OLAP, Online Analytical Processing, is utilized to dissect chronicled data and cut the business data required. OLAPs are frequently utilized by advertising administrators.

Data Visualization tools is utilized to show data from the data repository. Regularly data visualization is consolidated with Data Mining and OLAP tools. Data visualization can permit the client to control data to show importance and examples.

5.0 ETL Roles and Responsibility

The ETL planner must have the accompanying parts and obligations:

- Comprehend the ETL determinations and assemble the ETL applications like Mappings on consistent schedule.
- An ETL designer ought to not just see his or her single strung arrangement of projects. The designer must see the whole arrangement of projects, He/she should guarantee the specialized group comprehends the objective database design and its use so that the changes which change over the source information into the objective information structures are plainly recorded and caught on. [6]
- ETL planner must be to guarantee that the different programming tools expected to perform the distinctive sorts of information handling are appropriately chosen.
- Perform root cause analysis on all processes and resolve all production issues and validate all data and perform routine tests on databases and provide support to all ETL applications.

6.0 Problems arise during ETL processing

There are various issues to executing productive and dependable ETL process. [5]

- Technical challenges moving, integrating, and transforming data from disparate environments
- Short load windows, long load times
- Inconsistent, difficult to maintain business rules

- Source systems missing certain critical data
- Poor query performance

7.0 Query Cache

For the quick get to the database we utilize the query cache. Query cache will store all record of executed query. Query cache will keep record of recently executed inquiries. The major objective of the query cache is to diminish the reaction time of query. It will expand the mental aptitude of information product house so framework will retain the most recent work it has performed. This memory will be utilized a while later for answer the aftereffect of questions which has been before performed by the clients. The cache will keep up two state valid and invalid state. At the point when any query put together by the client, the cache memory is initially analyzed to check whether asked for query is as of now store in the cache. On the off chance that the query is saved, then check the state is valid or invalid. In the event that state is valid then data can be get to and if state is invalid then data can't be get to be that as it may, If client send a query of supplement, redesign, erase and drop then information will be change in database and condition of related query will be invalid. Presently invalid state data and query can't be access by client. This can spare vital time and enhance Data warehouse execution by not rethinking the inquiries which are now saved in the cache.

One of expert place a query to demonstrate to me the employee of an organization, who working under the manager_id is 110,102,202. The query will look like as takes after: -

```
SELECT emp_id, name, salary, manager_id
FROM employees
WHERE manager_id IN (110,102,202);
```

At the point when the query is submitted, query cache will be inspected to check whether this query is accessible or not and state is valid or invalid. In the event that it is not accessible, query will be assessed and result will be store in the query cache. The after effects of the query are appeared in the table 1.

TABLE 1 - OUTPUT OF THE ABOVE QUERY

| Emp_id | Name | Salary | Manager_id |
|--------|--------|--------|------------|
| 101 | Alex | 15000 | 202 |
| 105 | Vika | 18000 | 102 |
| 205 | Kostya | 21500 | 110 |
| 107 | Ivanov | 36400 | 202 |
| 104 | Krill | 23500 | 100 |

On the off chance that whatever other client presented the same query the outcome will be recovered from query cache since that query is as of now saved in the cache. We will call this Query1.

Let assume another client needs to the representative of an organization, whose salary more noteworthy than equivalent to 20000 AND manager_id is 100,101. The query will look like as takes after:

```
SELECT emp_id, name, salary, manager_id
FROM employees
WHERE manager_id IN (100,202) and salary >= 20000
```

At the point when the query is submitted, cache memory is analyzed. Same query is saved in the cache memory and state is valid then we can get the after effect of Query 2 as appeared in Table 2.

TABLE 2 - OUTPUT OF THE ABOVE QUERY

| Emp_id | Name | Salary | Manager_id |
|--------|--------|--------|------------|
| 205 | Kostya | 21500 | 110 |
| 107 | Ivanov | 36400 | 202 |
| 104 | Krill | 23500 | 100 |

Presently consequence of Query 2 will be produced from the Query 1 result set as opposed to experiencing from all the data saved in the data warehouse. This procedure will spare parcel of time and exertion required to experience every one of the records.

The SELECT query, the store first parcels the query to get to all the divided tables. At that point it checks whether each divided query is beforehand stored. If not, then it continues to check whether any already reserved query can give an incomplete answer.

At the point when the database gives back the outcome for the rest of, it consolidations the outcome with the incomplete result got from the reserve. At long last, it combines the outcomes from all the apportioned inquiries and returns the consolidated result back to the customer.

At the point when an update, or delete query is gotten, it is sent to the database. At the point when the reaction from the database is gotten, the reserve negates all store passages reliant on either the influenced tables or the influenced segments. These references are then deleted from the database mapping data structure.

7.1 Query cache state

Valid- on the off chance that query is put saved in query cache and not updated in database from any these query embed, update delete the state will be valid.

Invalid- In the event that query is not saved in query cache then state will be invalid. On the off chance that data is updated by client by any these query embed, update delete the state will be invalid.

Our issue is that we have a query and query result saved in the cache. In any case, if the warehouse is updated with the new information the cache query result will reflect to old information. We will make a system of state;

Query 1 is put together by the client and his outcome is saved in the query cache. At the point when next client present the same query on upgraded Data warehouse the query cache will check the state if state is invalid, it implies the information warehouse is redesigned with new information. Presently the query doesn't need to experience from the greater part of the records. It will get the last file of the query result put away in the query cache. At that point it will begin looking the records which meet the query criteria from ahead to that list. This can spare part of time and exertion required to seek the vast measure of information.

7.2 Experiment

In the analyses, we consider the customer emulator, Apache and Tomcat are co-situated on one machine. The query cache is running on a second machine and the MySQL database on a third machine. The same philosophy is utilized for measuring the execution of both cache forms (the C-JDBC cache with and without the semantic improvements).

In every analysis the principal portion of the run is utilized to warm-up the cache and is barred from the estimations. Every examination additionally begins with an indistinguishable database. Contrasts between rehashed keeps running of the same investigation were insignificant. To choose the heap for the analyses, we are driving the server without the cache with expanding the quantity of customers, until execution crests.

8.0 Conclusion

Customer relationship management is fundamental to contend adequately today's commercial center. The improvement of CRM advantages from the advancement of PC innovation, computerized. In this paper, we have acquainted a technique with enhance the execution and velocity of ETL in data warehouse by minimizing the reaction time essentially. The essential objective of this strategy is to store queries and their relating results. In the event that comparable inquiry is presented by whatever other client the outcome will be gotten utilizing cache memory. Inquiry Cache procedure is to store queries and their comparing results. On the off chance that comparable inquiry is put together by some other client the outcome will be acquired utilizing cache memory. The data from past queries is utilized to produce results. At last, the outcomes are productively converged for elite. This dynamic system extraordinarily enhances the execution and pace of ETL in Data product house. In this Paper, We increased the efficiency or flow enhancement from CRM to Data warehouse by using Query cache.

9.0 References

1. Data Warehousing Overview http://www.dwreview.com/DW_Overview.html
2. The Data Warehouse Toolkit third edition by Ralph Kimball and Margy Ross
3. ETL process <http://datawarehouse4u.info/ETL-process.html>
4. Jin-Yu WEI , Rui CHEN, He-feng YU " Study on the Design of CRM-Oriented Synthetic Decision Support System", 2008 IEEE Conference on Cybernetics and Intelligent Systems.
5. Data Warehousing Intermediate & Advanced Topics Common Problems, Uncommon Solutions by Eric Mellum
6. Fundamentals of database systems. 4th Edition. Persons international and Addison Wesley. Ramez Elmasri and Shamkant B. Navath.
7. Ideal Strategy to Improve Datawarehouse Performance by Fahad Sultan & Dr. Abdul Aziz. (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 02, 2010, 409-415
8. Shim J.; Scheuermann, P.; Vingralek, R.: Dynamic Caching of Query Results for Decision Support Systems, in: Proceedings of the 11th International Conference on Scientific and Statistical Database Management (SSDBM'99, Cleveland, Ohio, USA, July 28-30)