Solving Scheduling Problem in Data Warehouse Using OLAP and OLAP

Authors

Himanshi Kataria¹, Amit Garg²

¹M.Tech (CS) Computer Science Department Indus Institute of Engineering and Technology
²Asst. Professor Computer Science Department Indus Institute of Engineering and Technology

ABSTRACT:

Data in the warehouse and data marts is stored and managed by one or more warehouse servers, which present multidimensional views of data to a variety of front end tools: query tools, report writers, analysis tools, and data mining tools. Finally, there is a repository for storing and managing metadata, and tools for monitoring and administering the warehousing system. An operational database undergoes frequent changes on a daily basis on account of transactions that taken area. Think a business management wants to analyze previous feedback on any data such as a product, a supplier, or any consumer data, then executive will had no data available to analyze because previous data has been updated due to transactions. A data warehouse helps business executives to organize, analyze, & use their data for decision making. A data warehouse serves as a sole part of a plan-execute-assess "closed-loop" feedback system for enterprise management. Data warehouses are widely used in following fields: Financial services. Banking services. Consumer goods. Retail sectors. Controlled manufacturing.

1 INTRODUCTION

Meaning of Data Warehouse¹ was firstly coined by Bill Inmon in 1990. This data helps analysts to take informed very important decisions in group. An operational database undergoes frequent changes on a daily basis on account of transactions that taken area. Think a business management wants to analyze previous feedback on any data such as a product, a supplier, or any consumer data, then executive will had no data available to analyze because previous data has been updated due to transactions.

Data Warehouse Applications²

A data warehouse helps business executives to organize, analyze, & use their data for decision making. A data warehouse serves as a sole part of a plan-execute-assess "closed-loop" feedback system for enterprise management. Data warehouses are widely used in following fields:

1. Financial services
2. Banking services
3. Consumer goods
4. Retail sectors
5. Controlled manufacturing

It includes tools for extracting data from multiple operational databases & external sources; for cleaning, transforming & integrating this data; for loading data into data warehouse; & for periodically refreshing warehouse to reflect updates at sources & to purge data from warehouse, perhaps onto slower archival storage. In addition to main warehouse, there might be several departmental data marts. Data in warehouse & data marts are stored & managed by one or more warehouse servers, which present multidimensional
views of data to a variety of front end tools: query tools, report writers, analysis tools, & data mining tools. Finally, there are a repository for storing & managing metadata, & tools for monitoring & administering warehousing system. Warehouse might be distributed for load balancing, scalability, & higher availability.

Fig 1 Data Warehouse Architecture

Fig 2 Data Warehouse Architecture within Staging Area & Data Marts

2 LITERATURE REVIEW

Another research in 2013 by Mr. Dishek Mankad, Mr. Preyash Dholakia titled “The Study on Data Warehouse Design & Usage” was published. In this authors explain that data warehousing was a booming industry within many interesting research problem. data warehouse was concentrated on only few aspects. Here they are discussing about data warehouse design & usage. Let’s look at various approaches to data warehouse design & usage process & steps involved. Data warehouse could be built using a top-down approach, bottom – down approach or a combination of both. In this research paper they are discussing about data warehouse design process.

Research on topic “DATA WAREHOUSING & OLAP TECHNOLOGY” in 2012 was published by Manya Sethi in which she was of opinion that DATA WAREHOUSING & Online Analytical Processing (OLAP) are essential elements of decision support, which has been increasingly become a focus of database industry.

Data Warehouse provides an effective way for analysis & statistic to mass data & helps to do decision making. Many commercial products & services are now available & all of principal database management
system vendors now had offerings in these areas. paper introduces data warehouse & online analysis process within an accent on their new requirements. I describe back end tools for extracting, cleaning & loading data into data warehouse, tools for metadata management & for managing warehouse.

Another research titled “Realistic Analysis of Data Warehousing & Data Mining Application in Education Domain” was published by Manjunath T. N., Ravindra S. Hegadi, Umesh I. M., & Ravikumar G. K. In 2012.

Data-driven decision support systems, such as data warehouses could serve requirement of extraction of information from more than one subject area. Data warehouses standardize data across organization so as to had a single view of information. Data warehouses could provide information required by decision makers. Developing a data warehouse for educational institute was less focused area since educational institutes are non-profit & service oriented organizations.

### 3 TOOLS & TECHNOLOGY

Online Analytical Processing Server (OLAP) are based on multidimensional data model. It allows managers, & analysts to get an insight of information through fast, consistent, & interactive access to information. This chapter cover types of OLAP, operations on OLAP, difference between OLAP, & statistical databases & OLTP.

**Pivot**

The pivot operation are also known as rotation. It rotates data axes in view in order to provide an alternative presentation of data. Consider following diagram that shows pivot operation.
4 PROPOSED WORK
Data warehousing projects are one of its kinds. All data warehousing projects do not pose same challenges & not all of them are complex but they are always different. Knowing these challenges upfront are your best bet to avoid them. Data warehousing are different. For most part of it, these projects are heavily dependent on backend infrastructure in order to support front-end user reporting. But these are not only reasons why doing data warehousing are difficult. In below list we show top 5 reasons which actually make things complex on practical ground.

5 RESULT & DISCUSSION
Here we had chosen a huge database of MLM Company. Records of Approximate 8000 people had been maintained here along within their daily payout & regular buying.
Handling Challenges in data ware house management using query optimization
There are several Factors that would effect query processing
1. Data has been extracted from local or Remote server

Fig 5 Here in above diagram we had extracted records from remtoter server & it took 10 seconds in case of 5298 records.
Fig 6 Here in above diagram we had extracted records from local server & it took 0 seconds in case of 5298 records.

2. Number of columns extracted

Fig 7 If 3 columns had been retrieved from remote server than it will take about 2 seconds.

Fig 8 If 51 columns had been retrieved from remote server than it will take about 12 seconds.

to solve scheduling problem

We had two options to schedule query

Using Application program interface in programming language like java, C#, VB.net (using if condition are true than run this query).

Using trigger mechanism to schedule a particular query in particular circumstances.
Suppose a person wants to update counter in one table when entry are made to another table. Then he needs to perform the following steps:

Step 1: Create a table where records are to be inserted.

```sql
create table aa(name varchar(9), salary int)
```

Step 2: Insert records into the table.

```sql
insert into aa values('sunil', 40000)
```

Step 3: Create a table where counter would be stored.

```sql
create table cc(count int)
```

Step 4: Insert initial counter.

```sql
insert into cc values(1)
```

```sql
select * from cc
```

Step 5: Create a trigger to update counter to increment by 1 where insertion is made.

```sql
create trigger t1
on aa
after insert
as
begin
update cc set count=count+1
end
```

When the user makes some entries in `aa`, the counter in table `cc` would increase automatically.

Simulation of Normal Query & Optimized Query in Matlab

Here is the code of `gettime.m`.

Here `x` is the time taken to connect to the remote server, this time will vary as per connection speed & bandwidth of the database.

Here `x1` is the number of records extracted during 1 second by a normal query.

Here `x2` is the number of records extracted during 1 second by an optimized query.

```matlab
function a=gettime(x,x1,y1)
a=(x+x1)/y1;
end
```
Following code will call above function in case of different number of records & will write result in a text file for further plotting

```matlab
function x=wreading(x1,y1,y2)
    fid=fopen('x.txt','w');
    for r=1000 : 100:10000
        fprintf(fid,'%d %f %f
',r,gettime(r,x1,y1),gettime(r,x1,y2));
    end
    fclose(fid)

Result would be added in x.txt file
```

![Figure 5.8](image.png)

**Fig 5.8** Result would be added in x.txt file

The above result would be plotted in matlab as follow

```matlab
fid1=fopen('x.txt');
c=textscan(fid1,'%d %f %f');
a=c{1};
br=c{2};
cc=c{3};
plot(a,br,'r-');
hold on
plot(a,cc,'b*');
legend('Normal','Optimized');
fclose(fid1)
```
6 CONCLUSION
Testing in data warehousing are a real challenge. A typical twenty percent time allocation on testing are just not enough. One of reasons why testing are tricky are due to reason that a top level object in data warehouse typically has high amount of dependency. Because of such high dependencies, regression testing requires lot of planning. Making data available for re-testing for a certain component might not be possible as fresh data loading often changes surrogate keys of dimension tables thereby breaking referential integrity of data. Relational OLAP servers is placed to relational back-end server & client front-end tools. To store & manage warehouse data, relational OLAP uses relational or extended-relational DBMS. ROLAP servers are highly scalable. ROLAP tools analyze big range of data across multiple dimensions. ROLAP tools store & analyze highly volatile & changeable data.

REFERENCES
3. Manjunath T. N. wrote on Realistic Analysis of Data Warehousing and Data Mining Application in Education Domain
4. Weiss, Sholom M.; and Indurkhya, Nitin (1998); Predictive Data Mining, Morgan Kaufmann


15. Ye, Nong (2003); The Handbook of Data Mining, Mahwah, NJ: Lawrence Erlbaum

16. Cabena, Peter; Hadjinian, Pablo; Stadler, Rolf; Verhees, Jaap; Zanasi, Alessandro (1997); Discovering Data Mining: From Concept to Implementation, Prentice Hall, ISBN 0-13-743980-6