Mining the Data Using Aggregator From the Dynamic Web Page

Mrs. S.Abirami, M.E
M.E.Computer Science and Engineering
A.S.L Pauls college of Engineering and Technology, Coimbatore.
Guidance by G.Shenbhagamoorthy, Assistant Professor
A.S.L Pauls college of Engineering and Technology, Coimbatore.

ABSTRACT- Typically a user wishes to acquire the value of some aggregation gathering over distributed data items, for example, to know value of collection for a client; or the AVG of temperatures sensed by a set of sensors. In these queries a client specifies a coherency condition as part of the query. We present a inexpensive, scalable technique to answer constant aggregation queries using a set of connections of aggregators of active data items. In such a network of records aggregators, each data aggregator serves a set of data items at exact coherencies. Just as various garbage of a active web-page are served by one or more nodes of a content sharing network, our system involves decayed a client query into sub-queries and executing sub-queries on sensibly chosen data aggregators with their individual sub-query incoherency limits. We provide a procedure for getting the best possible set of sub-queries with their incoherency bounds which satisfies customer query's coherency requirement with slightest number of energize messages sent from aggregators to the client. For estimating the number of refresh messages, we build a query cost model which can be used to guesstimate the number of statement required to satisfy the client specified incoherency hurdle. feat results using real-world traces show that our cost based query progress leads to queries being executed using fewer than one third the number of messages required by offered schemes.

1. INTRODUCTION

Application such as auctions, personal portfolio valuations for financial decisions, sensors-based monitoring, Route planning based on traffic information, etc., make extensive use of dynamic data. For such applications, data from one or more independent data sources may be aggregated to determine if some action is warranted. Given the increasing number of such applications that make use of highly dynamic data, there is significant interest in systems that can efficiently deliver the relevant updates automatically.

A. Proposed System:

Uninterrupted queries are used to observe changes to time varying data and to provide effect helpful for online result making. We present a economical, scalable technique to respond unbroken aggregation queries using a content sharing network of energetic recorded things.

Advantage:

1. It saves the time and the user spending low cost.
2. A continuous query cost model which can be used to estimate the number of messages required to satisfy the client specified incoherency bound.
3. We present to implementations of Continuous Aggregation in optimized query.

B.Existing System:

Many data intensive applications delivered over the Web suffer from performance and scalability issues. Content distribution networks (CDNs) solved the problem for static content using caches at the edge nodes of the networks. CDNs continue to evolve to serve more and more dynamic applications. The static fragments are served from the local caches whereas dynamic fragments are created either by using the cached data or by fetching the data items from the origin data sources.

One important question for satisfying client a request through a network of nodes is how to select the best node(s) to satisfy the request.

For static pages content requested, proximity to the client and load on the nodes are the parameters generally used to select the appropriate node.

Disadvantage:

1. For data item which needs to be refreshed at an incoherency.
2. The exact data value at the corresponding data source need not be reported as long as the query result satisfies user specified accuracy requirements.

2. THE GREEDY ALGORITHM FOR RESERVATION ARRANGEMENT SELECTION

```
Result ← Ø
While A ≠ Ø
Choose a sub-query a ϵ A with criteria ψ
Result ←Result υ a
A ← A-{a}
For each data element e ϵ a
For each b ϵ A
b ← b-{e}
If b = Ø
A ← A-{b}
Else
Estimate sumdiff for modified b
Return Result
```

S.Abirami M.E (CSE), E-mail:abirami.24@gmail

3. LITERATURE SURVEY

A.The Scope of Data Mining

Data mining derives its name from the similarities between searching for valuable business information in a large database — for example, finding linked products in gigabytes of store scanner data — and mining a mountain for a vein of valuable ore. Both processes require either sifting through an immense amount of material, or intelligently probing it to find exactly where the value resides. Given databases of sufficient size and quality, data mining technology can generate new business opportunities by providing these capabilities:

B.Automated prediction of trends and behaviors.

Data mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data — quickly. A typical example of a predictive problem is targeted marketing. Data mining uses data on past promotional mailings to identify the targets most likely to maximize return on investment in future mailings. Other predictive problems include forecasting bankruptcy and other forms of default, and identifying segments of a population likely to respond similarly to given events.

C.Automated discovery of previously unknown patterns.

Data mining tools sweep through databases and identify previously hidden patterns in one step. An example of pattern discovery is the analysis of retail sales data to identify seemingly unrelated products that are often purchased together. Other pattern discovery problems include detecting fraudulent credit card transactions and identifying anomalous data that could represent data entry keying errors.
The most commonly used techniques in data mining are:

1. Artificial neural networks**:** Non-linear predictive models that learn through training and resemble biological neural networks in structure.

2. Decision trees**:**
Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID).

3. Genetic algorithms**:**
Optimization techniques that use process such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution.

4. Nearest neighbor method**:**
A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where k ³ 1). Sometimes called the k-nearest neighbor technique.

5. Rule induction**:**
The extraction of useful if-then rules from data based on statistical significance.

The ideal starting point is a data warehouse containing a combination of internal data tracking all customer contact coupled with external market data about competitor activity. Background information on potential customers also provides an excellent basis for prospecting. This warehouse can be implemented in a variety of relational database systems: Sybase, Oracle, Redbrick, and so on, and should be optimized for flexible and fast data access.

4. PROJECT MODULES

A. Protection Module**:**

This module is used to help the user to provide the security of access. Because once the user to logout or leave our account automatically user password is changed and server to send the password in our mail ID. Whenever the user to logout the account automatically the security key is changed based on the arbitrary function.

B. Aggregation Module:

This module is used to help the user to view the BSE and NSE value in bar flow chat based on the date. This chat to display the aggregated value based on the companies sharing values constantly. The company's value is changed without human intervention chat value is changed.

C. Modernize Module:

This module is used to help the user to view the BSE and NSE value to update in every minute. So the user waiting time is reduced and sees the efficient value in every minute the server to set the time when our form is updated.

D. Consumer / Server Module:

This module is used to help the consumer and server communication to the database. This module is used to with dynamism create the table based on the server entering value. These values are conveying to the chat x and y position and display the client. These values are changed in dynamically based on the server entering values.

5.SYSTEM ARCHITECTURE DESIGN

Data refresh from data sources to clients can be done using push or pull based mechanisms. In a push based mechanism data sources send update messages to clients on their on where as in pull based mechanism data sources send messages to the client only when the client makes a request.
The push based mechanism for data transfer between data sources and client .In this data aggregator maintains its configured incoherency bounds for various data items. The given client request query can be divided by sub queries and sent it to corresponding data aggregators to finding the data items.

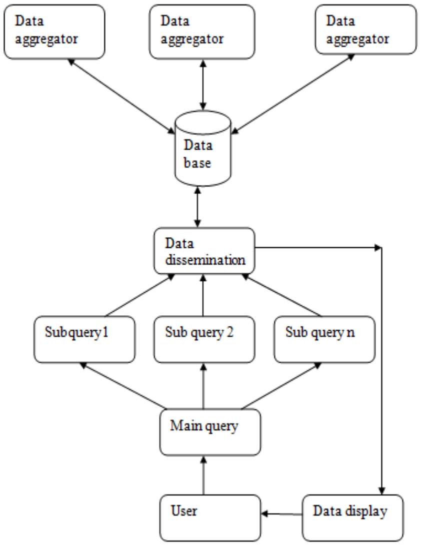


Figure 5:System architecture design

All data items from the different aggregators to be combined and Produce aggregated results.

6. IMPLEMENTATION

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective.

The implementation stage involves careful planning, investigation of the existing system and it's constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

A. Input Design

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy.
Objectives:

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

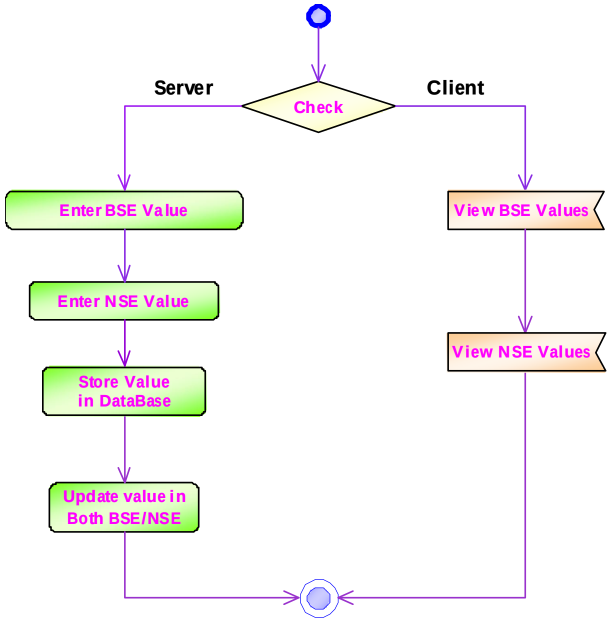
B. Output Design

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.
1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.
2. Select methods for presenting information.
3. Create document, report, or other formats that contain information produced by the system.

7. SYSTEM DESIGN

Our NSE and BSE applications are based on the Data mining concepts to increase performance in time and cost .First server has to register using Registration form and automatically generating random security key for server password to login next time in the server side. Same as client also has to register the registration form and give his own password to login separately. Both information to be stored in the System database and to refer the details for login authentication. Server has store the BSE and NSE share values in the databases. Whenever the client enters to their login he can get the recent updated aggregation value of share values to be displayed. Client need not to update every time changes in the share values automatically gets update by server. These values can be stored in dynamic table created in databases to display in flow chart for updated values to client.

A.Activity Diagram:



8. SYSTEM STUDY

Feasibility Study

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- Economical Feasibility
- Technical Feasibility
- Social Feasibility

9. SOFTWARE ENVIRONMENT

FEATURES OF. NET

The .Net Framework:

The .NET Framework has two main parts:
1. The Common Language Runtime (CLR).
2. A hierarchical set of class libraries.
The CLR is described as the "execution engine" of .NET. It provides the environment within which programs run. The most important features are

Conversion from a low-level assembler-style language, called Intermediate Language (IL), into code native to the platform being executed on.

- Memory management, notably including garbage collection.
- Checking and enforcing security restrictions on the running code.
- Loading and executing programs, with version control and other such features.
- The following features of the .NET framework are also worth description:

FEATURES OF SQL-SERVER

The OLAP Services feature available in SQL Server version 7.0 is now called SQL Server 2000 Analysis Services. The term OLAP Services has been replaced with the term Analysis Services. Analysis Services also includes a new data mining component. The Repository component available in SQL Server version 7.0 is now called Microsoft SQL Server 2000 Meta Data Services. References to the component now use the term Meta Data Services. The term repository is used only in reference to the repository engine within Meta Data Services
SQL-SERVER database consist of six type of objects,
They are,
1. TABLE
2. QUERY
3. FORM
4. REPORT
5. MACRO

Table:

A database is a collection of data about a specific topic.

Views of Table:

We can work with a table in two types,
1. Design View
2. Datasheet View

Design View

To build or modify the structure of a table we work in the table design view. We can specify what kind of data will be hold.

Datasheet View

To add, edit or analyses the data itself we work in tables datasheet view mode.

Query:

A query is a question that has to be asked the data. Access gathers data that answers the question from one or more table. The data that make up the answer is either dynaset (if you edit it) or a snapshot (it cannot be edited).Each time we run query, we get latest information in the dynaset. Access either displays the dynaset or snapshot for us to view or perform an action on it, such as deleting or updating.

10. CONCLUSION

This paper presents a rate based come close to reduce the number of refreshes mandatory to implement an incoherency delimited nonstop query. For best effecting we split the question into sub-queries and assess each sub-query at a selected aggregator. Show results that by our method the query can be executed using a reduced amount of one third the communication required for presented schemes. Further we showed that by executing queries such that more active information things are part of a better sub-query we can advance performance. Our system of query execution can be implemented using schemes like used in CDNs. Our query rate replica can also be used for other purposes such as load balancing a variety of aggregators, best query finishing plan at an aggregator node, etc. Using the cost replica for other applications and mounting the cost model for more composite queries is our upcoming work.

11. REFERENCES

[1] Rajeev Guptha and Kirthi Ramamrithm (2004)."Query planning for continuous aggregation queries over a network of data aggregators", The real time systems Journel, and the VLDB Journal.

[2]Devis.A, Parikh.J and Weihi.W (2004),"Edge computing: Extending Enterprises Applications to the Edge of the Internet, www

[3]Dilley.J, Maggs.B, Parikh.Z, Prokop.H,Sitaraman.R, and Weihl.B, J. Sept. 2002 "Globally Distributed Content Delivery," IEEE Internet Computing, vol. 6, no. 5, pp. 50-58.

[4]Gupta.R,Puri.A,andramamritham.K(2005), "Executing Incoherency Bounded Continuous Queries at Web Data Aggregators".

[5]Jain.N,Kit.D,Yalangandula.P,Dahlin.M,Zhang.Y(2007),"ST AR: Self-Tuning Aggregation for Scalable Monitoring",VLDB.Beginning ASP.NET 4: in C# and VB by Imar Spaanjaars.

[6]Olston.C, Jiang.J, and Widom.J(2003),"Adaptive Filter for Continuous Queries over Distributed Data Streams". SIGMOD

[7]Rangarajan.S, Mukerjee.S, and Rodriguez.P(2002), "User Specific Request Redirection in a Content Delivery Network," Proc. Eighth Int'l Workshop Web Content Caching and Distribution (IWCW).