A Benchmarking Criteria for the Evaluation of OLAP Tools

Fiaz Majeed
Department of Information Technology, University of Gujrat,
Gujrat, Pakistan.
Email: fiaz.majeed@uog.edu.pk

***Abstract*— Generating queries on Online Analytical Processing (OLAP) tools for complex analysis is a difficult assignment for the novice users. To compose accurate OLAP queries for fulfilling demand, technical knowledge about schema and the data is required. This deficiency can be covered by providing an easy design for OLAP tool for the purpose of querying. In this paper, a scheme is proposed for comparison of the OLAP tools to identify easy and standardized aspects. For this purpose, seven parameters have been used which are interface, query, drill-down options, roll-up options, aggregation function support, data access and performance. For experimental analysis, two tools SQL Server and MicroStrategy Express have been evaluated based on the proposed parameters. The benefits and drawbacks of the standardization of tools for non-technical users have been identified.**

Keywords-OLAP tool; evaluation criteria; parameters; benchmark; standardize.

I. INTRODUCTION

The Online Analytical Processing (OLAP) comprises a relational or multidimensional database intended to deliver fast retrieval of multidimensional analysis and pre-summarized data. The OLAP contains three fundamental operations for results presentation including drill-down, slicing & dicing, and roll-up. The drill-down operation provides navigation towards details (upper to lower levels). For example, the user may be interested to view the detail of region's sale in the form of individual products. By contrast, roll-up operation consolidates the results (lower to upper levels). For instance, individual products can be roll-up to region's sale. Further, slice and dice operation is used to select data from the OLAP cube. The OLAP tools provide these and many other functionalities. Most popular tools include MicroStrategy Express, SQL Server, SAP business object, Oracle, QlickView and Pentaho Business Intelligence (BI).

To explore information by diverse angles, OLAP tools are utilized extensively and vendors related to them claim excellent performance. A number of tools are available which have distinct characteristics. All tools have a different way of responsiveness, the design of the interface, input query and performance. To the best of our knowledge, there is no benchmark to design an OLAP tool available. The users face a problem in the selection of an appropriate tool due to lacking standard. They cannot easily understand which OLAP tool is suitable for their requirement. In other words, a standard OLAP tool is needed to support novice users.

In this paper, a method has been formulated to identify standardized aspects for the comparison of the OLAP tools. For this purpose, seven parameters have been proposed which are the interface, query, drill-down, roll-up, aggregation function support, data access and performance. This is the list of parameters available in all of the existing tools. We have performed a comparative analysis of two OLAP tools: SQL Server and MicroStrategy Express. The experiments have been performed using the dataset AdventureWorksDW. The comparison exhibits basically the opportunities for non-technical users.

The paper is arranged as follows: detailed literature survey is furnished in Section II. The assessment parameters are introduced in Section III, while an experimental evaluation has been incorporated in Section IV. Finally, conclusions and future directions are given in Section V.

II. RELATED WORK

The OLAP supports for multi-dimensional complex analysis on data warehouses for decision making [1]. To compare OLAP tools, several features should be considered. The features to evaluate the usefulness of OLAP are ease of use, user-friendly, easy learning and easy to get information. Seven features are used to measure the OLAP tool, which is: visualizations, summarization, Navigation, query function, Sophisticated analysis, Dimensionality, and performance [2]. An important feature that makes the design of OLAP tool user friendly and easy to use is the interface. Visualization is an important aspect of interface design.

A. Interface

Visualizations of statistical data need to present relationship among data. Existing tools show isolated graphs and do not provide support for the relationship in different reports. A visual language named as CoDe is used to present relationship among data in tabular form. The visualization is performed in four phases: CoDe Modeling, OLAP operation pattern definition, OLAP Operation and Report Visualization [3]. Thus, graphs are an integral part of the interface so big process graph refers to process-related partly unstructured execution and heterogeneous data of large hybrid collections. A set of methods and a framework are given for initiating OLAP analytics which is called P-OLAP. The P-OLAP introduces analytics over process execution data based on the scalable graph. It is the extension of traditional OLAP analytics [4].

Key Performance Indicators (KPI) are manually integrated into scorecards and dashboards used by the decision makers. Due to this, KPIs are not related to their

 ISBN: 978-1-61208-590-6

business objectives and strategy. To make the KPIs dynamic, the modeling language Object Constraint Language (OCL) is used to represent OLAP actions, which are then translated to Multi-Dimensional eXpressions (MDX) query to be executed on OLAP engine [5]. As OLAP tools visualize results in the form of aggregations, drill-down up to maximum detail is also the important feature of the interface. Thus, the user may drill-down data to the maximum level where maximum measures are to be returned. In such a case, pivot table may not visualize data with full precision [6].

B. *Aggregations*

The aggregation operations are key to OLAP BI tools. The OLAP tools perform several operations including roll-up, drill-down slice and dice, ranking, selection and computed attributes [1]. The OLAP operation "shrink" balances data precision with the size of data cube via pivot table. It combines similar data in a single slice (f-slice) for the purpose of shrinking [6]. An aggregation operator has been built for the text embedded in the tweets content based on the Formal Concept Analysis (FCA) theory [7].

C. *Performance*

The selection of an appropriate server type plays a role in the performance of query processing. The OLAP servers generate aggregations for efficient query processing based on dimensions. The servers are categorized as ROLAP, MOLAP and Hybrid of ROLAP and MOLAP [1]. MOLAP presents data in multi-dimensional arrays format while ROLAP provides query processing on Relational databases [2]. Near real-time BI reduces the time of acquisition of data in operational sources and analyses on that data. Event processing on streaming data is an application of near real-time BI. Additionally, MapReduce paradigm can search in schema-less input files in comparison to parallel database approach. For enhanced BI performance, private clouds provide more security. Currently, BI is being switched to mobile devices as such devices are pervasive [8].

There are several advanced OLAP domains, which have emerged recently. The Skalla system has been built, which translates GMDJ operator into local site level plans. The Packet header, flow level traffic statistics, and router statistics can be analyzed effectively using OLAP. Heavy traffic cannot be loaded into a central data warehouse, thus local data warehouses on each site should be implemented to avoid loss of data and for efficient execution of OLAP queries. The Skalla performs optimization to minimize synchronization traffic and local level executions [9]. The tweet streams available in unstructured form are organized in an OLAP cube for analytics using Time-Aware Fuzzy Formal Concept Analysis theory. The microblog summarization algorithm is introduced and it provides the subset of the tweet that best represents the OLAP cube data for the analytic purpose. The definition of the multi-dimensional data model for the storage of tweet data streams for enabling OLAP analysis is performed [7]. The AOLAP maintains data stream's aggregations by providing OLAP queries in the form of approximate answers and maintains them in smaller space on the primary memory. In the OLAP cube, data summaries are stored related to each materialized node which is performed by the proposed Piece-wise Linear Approximation (PLA). To minimize overall querying error, lattice nodes based optimization technique is proposed [10].

Based on the literature, it is clear that there is no benchmark available to compare the existing tools. There is a need to develop such a benchmark to standardize the design and working of each tool to facilitate non-technical users.

III. ASSESSMENT PARAMETERS OF OLAP TOOLS

The evaluation criteria for OLAP tools contain seven parameters which include interface, query, drill-down options, roll-up options, aggregation function support, data access and performance. The parameters are depicted in Figure 1.

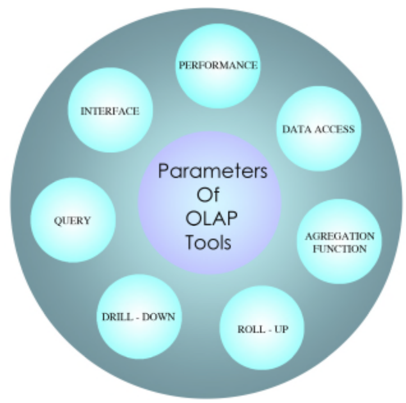

Figure 1. Evaluation criteria for the OLAP tools

For instance, User A wants to analyze his company's revenue with the revenue of his competitive companies. He wants to choose a BI tool for this purpose, which can provide him with a point and click environment to fulfill his demand. He selects SQL Server to input a query, but it requires training. Here the problem is that there should be a standard available based on which each tool can be designed. Each tool should provide similar easy query input mechanism for novice users.

The OLAP tools can be compared based on following seven parameters:

A. *Interface*

The interface is a fundamental source of communication between the user and the system. If interface design is adequate, the user can effortlessly interact with the system. The interface is evaluated based on following features:

Design: The aspects of design that are necessary to be considered are user understandability, first look impact of front page, the size of everything which is shown, colors effect, formal things to be used in designing, formatting of text and objects, number of formats to display the results, support of graphical results, and user understandability about results.

User-friendly: The user-friendly feature of the interface can be compared based on the sequence of steps understandable by the user, user understanding, how the response is shown at any step, and support with help and documentation. Understandability at each step during any task, make no confusion for the user, and user understandability at first look.

Type: The interface is graphical or query based. The graphical user interface is easy to use for novice users whereas a query based interface is preferable for expert users. Hybrid approach (a mixture of graphical and query based) provide an advantage to both types of the users.

User type: There are commonly two types of users for OLAP tools are available 1) novice user and 2) expert user. Interface comparison of different tools can be performed according to the facilities provided for each type of users such as the use of interface at first time and support available for the new user.

Visualizations: The interfaces are divided into three categories which are text-based, visualization-based and voice-based. The visualization-based interfaces provide analysis in the form of plots, trees, and charts. The comparison can be made based on easy to interpretable results, understanding at first glance, a large volume of data summarized in simple graphs.

Interactive: By this feature, user interaction at the front end of the tool is measured. At how many levels, the user can perform the task in an interactive way. The user finds a correct answer of the query and he may correct query on the basis of feedback provided by the system.

Complexity: The interface should have the capability to take input complex query in its simplest form. It should visualize complex OLAP analytics results in easy formats which may be the mixture of text, visualizations and voice. Many OLAP tools display simplified results on dashboards.

B. Query

The query is evaluated based on following features:

Format: It defines the format of query writing. There are four common formats of query input that are Command-based, Menu-based, Natural language-based, and Wizard-based. The easiest formats for query input considered for novice users are Menu-based, Natural language-based, and Wizard-based.

Procedure: It is categorized as query input procedure and results display procedure. The steps of query input are simple and natural as commonly input on web or other parallel interfaces. It should be according to previous familiar interfaces. The display of results is converged to the ease of understanding of results. The results must be focused for the less technical users.

Drill-down options: The comparison can be made based on summarization level of drill-down e.g. detailed level. How drill-down facility is provided either point and clicking way or another indirect way. If a user reaches to the detailed level of drill-down, how much a tool provides backtracking to roll-up levels?

Expertise required: This feature defines the type of users utilizing the tool. How much time is required for a particular user to achieve expertise of using the tool? For instance, the novice user may get expertise after 10 sessions span one-hour long.

Training required: Did any type of training requirements to use the tool? Further, if training is required then what level of training is demanded to learn and use the tool proficiently. Which training options are available? Few tools provide a manual for training while a professional training is necessary to utilize other tools.

C. Drill-down options

Support: According to support feature, the tools are compared based on whether drill-down support is available to navigate the results. Most of the tools provide this feature available in their interface whereas few of the tools do not provide drill-down options.

Options: How much depth and breadth level of navigation are provided? The tool provides drill-down support in one of the visualizations such as graphs, charts, tables or a mixture of them. Does drill-down performed on pre-built aggregations or run-time aggregations may be generated upon requirement?

Point and click: How drill-down facility is provided both point and clicking way or another indirect way? Does the tool expand a summarized value in the tree format or show detailed data in another format? If the user reaches to the detailed level of drill-down, how much a tool provides backtracking to roll-up levels?

Grouping different way: Is the grouping of data provided at run-time? How many attributes can be added in a group? Which grouping functions are supported by the tool?

Complexity: How drill-down results are presented to the user? Whether complex results are presented in simplest graphical formats? Does track of drill-down is given in some tree-like format to memorize the forward and backward tracking and switching to any other level of hierarchy?

D. Roll-up options

Support: In this feature, the tools are compared based on whether roll-up support is available to navigate the results. Most of the tools provide this feature available in their interface whereas few of the tools do not provide roll-up options.

Options: How much depth and breadth level of navigation are provided? The tool provides roll-up support in one of the visualizations such as graphs, charts and tables or the mixture of them. Does roll-up performed on pre-built aggregations or run-time aggregations may be generated upon requirement.

Point and click: How roll-up facility is provided either point and clicking way or another indirect way? Does the tool collapse a detailed value in the tree format or show summarized data in another format? If a user reaches the upper level of roll-up, how much a tool provides backtracking to drill-down levels?

Grouping different way: Does grouping of data is provided at run-time? How many attributes can be added in a group? Which grouping functions are supported by the tool?

Dimensions selection: How many dimensions can be selected at a time for grouping the data? How many dimensions can be used by default?

E. Aggregation function support

The OLAP servers use Materialized Views (MVs). If a particular MV is not found, then minimal MVs are further roll-up to generate required summarization level. The OLAP servers maintain data aggregated with several aggregation functions. At query processing time, the system selects desired MVs, aggregated data computed with the use of particular aggregation function. Furthermore, user query specifies the aggregation level with grouping attributes. These maintain aggregated data in specific structures to efficiently retrieve the results. The structures are relational and multi-dimensional etc. The tools can be compared based on aggregation functions supported by these tools.

F. Data access

In this, tools are analyzed by data type supported by them. The support of a number of fact tables, dimensions and measures are also verified.

G. Performance

The performance is measured by analyzing the response time by input queries on the tools.

IV. EXPERIMENTAL EVALUATION

For experimental analysis, two OLAP tools SQL Server and MicroStrategy Express have been taken. These tools are evaluated with respect to interface, query, drill-down options, roll-up options, aggregation function support, data access and performance.

SQL Server helps to build secure, reliable and scalable enterprise applications for the organizations. It also supports to deploy and maintain the applications. It provides analytical services to build data warehouses and the OLAP applications. It supports relational, multi-dimensional and hybrid data manipulations and provides facilities for complex analysis. The OLAP analytics has been successfully provided to the organizations using the SQL Server.

MicroStrategy Express facilitates secure and twenty times faster access to business data. There is no expert help needed, no data modeling, and no SQL scripts required. Get business insights quickly with interactive dashboards, pixel-perfect documents, and data visualizations.

A. Comparison of Tools

We analyzed these tools with respect to an interface, query writing, drill-down options, roll-up options, aggregation function support, data access and performance point of view. A survey has been conducted for the evaluation of both tools from 150 users. We are able to get different ideas of users about query-ability, performance, and interface point of view based on the questionnaire. The survey is comprised of three types of users which are categorized based on their level of expertise:

- Novice user
- Average user
- Professional user

Different type of user gave response according to their understanding. The professional user analyzes the tools according to their own needs. Average users analyze the tools according to their views and novice user analyzes the tools according to their understanding. The results gained for interface parameter are given in Table 1. The score is calculated in the range between 0 and 1. The response of each user is taken and the score is normalized within the specified range for each feature.

TABLE 1. RESULTS FOR THE INTERFACE PARAMETER

Ser. #	Features	SQL Server	Micro Strategy Express
01	Design	Good(0.64)	Very Good(0.81)
02	User-friendly	Yes(0.63)	Yes(0.83)
03	Type	GUI (Desktop)(0.9)	GUI (Web)(0.7)
04	User type	Known User(0.50)	Novice(0.77)
05	Graph support	Yes(0.5)	Yes(0.83)
06	Understanding	Yes(0.64)	Yes(0.82)
07	Interactive	Yes(0.73)	Yes(0.88)
08	Complexity	Yes(0.83)	No(0.63)
09	Ease of query input	Yes(0.73)	Yes(0.73)
10	Format of result	Grid(0.8)	Multiple format(0.9)

Similarly, the survey results have been calculated for each parameter. Based on the average score for both OLAP tools, the comparative analysis is depicted in Figure 2. It presents the comparative analysis of both tools based on seven parameters. According to this, both tools show following results:

Interface design: There are better features in MicroStrategy Express with respect to the interface. The design is more understandable, user friendly and simple of the MicroStrategy Express. Further, it is easy for novice users and having availability of additional options for visualizations.

Query: According to query parameter, SQL Server outperforms but requires training and expertise to use it. Whereas, from the structural point of view MicroStrategy Express is easier in query input.

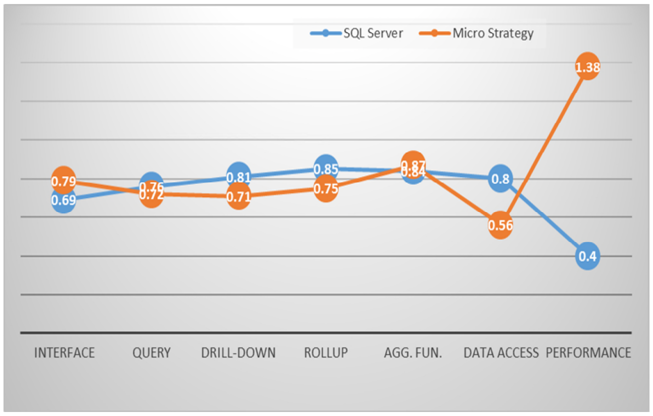

Figure 2. Comparison results of SQL Server and MicroStrategy Express

Drill-Down: SQL Server performs well for drill-down parameter and provides better options in comparison to the MicroStrategy Express. The SQL Server has point and click environment having descriptions in detailed level while MicroStrategy Express gives initial level drill-down options.
Roll-up: SQL Server provides additional roll-up options. For further aggregations, it delivers complete choices. Overall, SQL Server is best with respect to the roll-up options.
Aggregation function support: The average score of the MicroStrategy Express is 0.87 which is 0.03 better than the SQL Server. It replies quickly in the calculation of the aggregation functions.
Data access: From all point of views, this parameter is considered best in SQL Server. SQL Server supports the additional quantity of measures, data volume and dimensions. One of the deficiencies of the MicroStrategy Express is that it supports only excel-based datasets.
Performance: Performance is calculated based on response time in seconds. The MicroStrategy Express outperforms in comparison to the SQL Server.

V. Conclusions and Future Work

In this paper, criteria for the evaluation of tools have been proposed. Seven parameters include an interface, query writing, drill-down options, roll-up options, aggregation function support, data access and performance. For experimental evaluation, comparative analysis of two tools i.e., MicroStrategy Express and SQL Server has been performed. The results show that MicroStrategy Express outperforms in interface and aggregation method whereas input query is easy in comparison to the SQL Server. The SQL Server is more attractive along drill-down options, roll-up options and data access. The MicroStrategy Express only supports to excel-based datasets and does not compatible with the large databases. Similarly, any tool can be assessed based on seven parameters and variation in them can be eliminated for standardization purpose. As future work, it is required to implement a standardized tool for non-technical users for training purpose. After getting training of such a tool, the user will be able to use any OLAP tool.

References

[1] S. Chaudhuri and U. Dayal, “An Overview of Data Warehouse and OLAP Technology,” *Sigmod Rec.*, vol. 26, no. 1, pp. 65–74, 1997.

[2] N. Gorla, “Features to consider in a data warehousing system,” *Communications of the ACM*, vol. 46, no. 11, pp. 111–115, 2003.

[3] M. Risi, M. I. Sessa, M. Tucci, and G. Tortora, “CoDe modeling of graph composition for data warehouse report visualization,” *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 3, pp. 563–576, 2014.

[4] S. B. B. Benatallah and H. R. Motahari-nezhad, “Scalable graph-based OLAP analytics over process,” *Distributed and Parallel Databases,* pp. 379–423, 2014.

[5] A. Maté, J. Trujillo, and J. Mylopoulos, “Specification and derivation of key performance indicators for business analytics: A semantic approach,” Data Knowl. Eng., pp. 30–49, 2016.

[6] M. Golfarelli, S. Graziani, and S. Rizzi, “Data & Knowledge Engineering Shrink : An OLAP operation for balancing precision and size of pivot tables,” *DATAK*, vol. 93, pp. 19–41, 2014.

[7] A. Cuzzocrea, C. De Maio, and G. Fenza, “OLAP Analysis of Multidimensional Tweet Streams for Supporting Advanced Analytics,” pp. 992–999, 2016.

[8] S. Chaudhuri, U. Dayal, and V. Narasayya, “An overview of business intelligence technology,” *Commun. ACM*, vol. 54, no. 8, p. 88-98, 2011.

[9] M. Akinde and T. Johnson, “Efficient OLAP Query Processing in Distributed Data Warehouses Michael B ¨,” *Data Eng.*, vol. 32, no. 4, pp. 6382–6382, 2002.

[10] S. A. S. B and H. Kitagawa, “Approximate OLAP on Sustained Data Streams,” vol. 1, pp. 102–118, 2017.