

Многомерный подход к анализу экспериментальных данных исследований

Чередниченко А.В., Концевич В.Г.
СумГУ, maluy4ok@ukr.net.

The work is considered of aspects in using the multivariate analysis data. It means the ability to identify new relationships and prediction of the trial's course.

ВВЕДЕНИЕ

При обработке больших потоков информации, особенно результатам испытаний изделий машиностроения следует особое внимание уделять средствам реализации и концепции построения информационных систем, ориентированных на аналитическую обработку данных. В первую очередь это касается систем управления базами данных, основанными на многомерном подходе.

МЕТОДОЛОГИЯ МНОГОМЕРНОГО ПОДХОДА

При работе с многомерными базами данных, созданных по результатам испытаний, следует применять математические, статистические и другие методы, основанные на формальной логике, в свою очередь применяемых для построения или отбора оптимальных методов планирования экспериментов и обработки результатов, а также для извлечения наиболее важной информации при анализе экспериментальных данных. В результате построения многомерной модели данных исследований можно обнаружить новые явления или взаимозависимости, так как появляется возможность выявить скрытые объективно существующие связи между данными испытаний. Существует два типа обработки данных:

- OLTP (Online Transaction Processing-обработка транзакций в реальном времени) предназначенного на эффективный сбор данных в реальном времени;
- OLAP (Online Analytical Processing-аналитическая обработка в реальном времени) предназначенная для осуществления выборки и обработки данных эффективными способами. [1].

ЭФФЕКТИВНОСТЬ ИСПОЛЬЗОВАНИЯ OLAP-ПОДХОДА

Многомерное представление данных и OLAP-подход являются распространенными методами построения хранилищ данных (Data Warehouse) и аналитических систем различного назначения. Для анализа экспериментальных данных эффективным является использование OLAP, с дальнейшим построением OLAP-кубов. Основными преимуществами данного типа построений данных являются: [2]

- Быстрый доступ к данным не зависимо от размера массива и является преимуществом OLAP-систем.
- Преагрегация, которая представляет собой возможность выбрать наиболее вероятно получаемые результаты исследований.
- Иерархия. OLAP технология изначально рассматривает данные с точки зрения иерархий и взаимоотношений с другими параметрами одной и той же сущности

- Работа со временем. Так как анализ данных происходит на временных участках, OLAP технология экономит время анализа данных
- Возможность применения Data Mining как средства интеллектуального анализа данных с целью выявления скрытых закономерностей или взаимосвязей между переменными в больших массивах данных. [3].

Информация в OLAP технологии представляется в виде многомерного куба(OLAP-куба) с возможностью произвольного манипулирования ею. Данные рассматриваются либо как факты (суждение (предложение), в котором упоминается, какой либо субъект или именованный объект) с численными параметрами, либо как текстовые измерения, характеризующие данные, что является ключевым при построении базы данных основанной на результатах экспериментов.

Множество всех измерений куба (до 20) образует систему координат представляемого пространства данных. Значения измерений играют роль индексов, находящихся в ячейках куба. Особенностью измерений является их иерархическая структура, которая используется для агрегации и детализации значений показателей. Для обработки результатов экспериментальных исследований следует рекомендовать следующие два вида иерархий [4]:

- Сбалансированные – иерархии, в которых число уровней неизменно, т.к. определено её структурой, и все ветви имеют одинаковую глубину. Для формирования сбалансированной иерархии необходимо наличие связи «один-ко-многим» между объектами менее детального уровня по отношению к объектам более детального уровня.
- Несбалансированные – иерархии, в которых число уровней может быть изменено, некоторые уровни логически не одинаковы, и глубина ветвей иерархии может быть разной. Все объекты несбалансированной иерархии принадлежат одному типу.

OLAP технология не является противопоставлением существующей методам обработки данных, она лишь дополняет и расширяет ее. По-прежнему можно обращаться к существующим бумажным хранилищам данных, электронным таблицам и плоским файлам. Однако, применение OLAP-подхода для выполнения аналитических запросов определяет возможность создания OLAP серверов.

ВИСНОВКИ

Многомерная обработка данных экспериментальных исследований становится необходимым компонентом любого хранилища данных. Применение OLAP-систем обеспечивает возможность работать с данными без знания архитектуры хранения информации [5]. Результаты обработки данных исследования машиностроительных изделий могут быть представлены в обычных 2х-мерных таблицах, что позволяет создать удобное для работы хранилище данных.

ЛІТЕРАТУРА

Літературне джерело 1 Nigel Pendse. What is OLAP? [Электронный ресурс] – URL: <http://www.olapreport.com/fasmi.htm>

[1] Thilo Maier. A Formal Model of the ETL Process for OLAP-Based Web Usage Analysis // Proceedings of the sixth WEBKDD workshop: Webmining and Web Usage Analysis (WEBKDD'04), in conjunction with the 10th ACM SIGKDD conference (KDD'04). Seattle, Washington, USA, August 22, 2004.

Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP. – Спб.: БХВ_Петербург, 2007. – 384 с

Malinowski E., Zimanyi E. Hierarchies in a Multidimensional Model: From Conceptual modeling to Logical Representation// Data&Knowledge Engineering,2006. Vol. 59. № 2. P. 348 – 377.

Keith Laker «OLAP Workshop 1: Basic OLAP Concepts»[Электронный ресурс] –
URL:http://oracleolap.blogspot.com/2007/12/olap_workshop_1_basic_olap_concepts.htm