

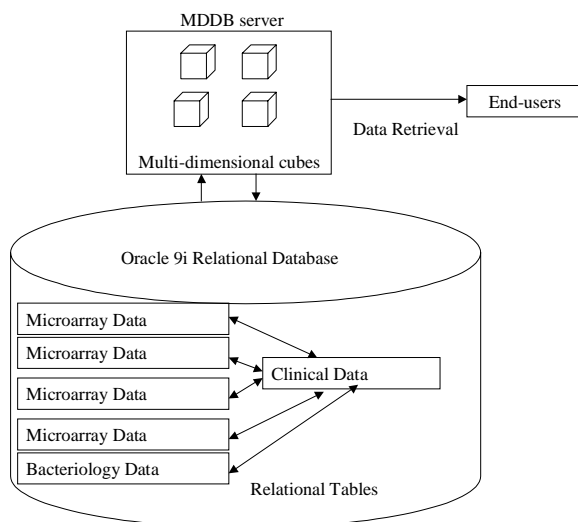
Architecture of a Gene Expression Data Warehouse

Aryya Gangopadhyay
Information Systems
UMBC
1000 Hilltop Circle
Baltimore, MD 21250
Phone: (410)455-2620
Email: gangopad@umbc.edu

INTRODUCTION

Biomedical researchers are rapidly generating vast amounts of genomic data that need to be modeled, organized, stored, and shared among researchers and practitioners. Genomic data is quite different than the data types ordinarily found in relational databases. The raw data consists of long strings of letters (A, T, G, C) representing the four nucleotide bases that constitute the gene. However, the raw data is not the only data type used in biomedical research. Technological advances such as microarrays have made it possible for researchers to study large numbers of gene expressions, which was previously been an intractable task. With this new technology comes the challenge of being able to store, retrieve, and analyze gene expression data stored in microarrays and linking them with other data such as clinical records and even other microarray data. This poses an interesting challenge for data management, warehousing, and mining. In this paper we develop an architecture for a data warehouse that would allow researchers to store microarray data using multidimensional cubes.

Figure 1. Data warehouse Architecture



METHODOLOGY

The datawarehouse design is shown in Figure 1. It consists of a relational database that stores microarray data, bacteriological data, and clinical data in relational tables. These datasets are then fed into a multi-dimensional database and converted into data cubes. The end-users query and visualize the data by interacting with a multi-dimensional (MDDB) server.

The microarray data comes from gene expression data of affected cells and no-affected cells. The bacteriology data represents DNA sequences of bacteria that are found in the infected host. Clinical data includes relational records describing patient characteristics.

Issues that are addressed in this paper are as follows:

1. How to store microarray and bacteriological data in a relational database.
2. How to provide linkage among datasets such that they can be compared and combined to generate meaningful information.
3. How to develop aggregate measures from relational tables into multi-dimensional cubes.
4. How to visualize data from multi-dimensional cubes in response to user requests.

REFERENCES

- Artale, A. and Franconi, E. "A Temporal Description Logic for Reasoning about Actions and Plans," *Journal of Artificial Intelligence Research (JAIR)*, Vol. 9, pages 463-506, 1998.
- Calvanese, D., Lenzerini, M., and Nardi, D. "Descriptoin Logic for Conceptual Data Modeling," in J. Chomicki and G. Saake (Eds.) *Logics for Databases and Information Systems*, Kluwer, 1998.
- Gangopadhyay, A. "Conceptual Modeling from Natural Language Functional Specifications," *AI in Engineering*, 15(2): 207-218, 2001.
- Gangopadhyay, A. "A Methodology for Designing Data Warehouses from Functional Semantics," *Journal of Applied Systems Studies*, (to appear in 2003).
- Hüsemann, B., Lechtenbörger, J., and Vossen, G. "Conceptual Data Warehouse Design," *Proceedings of the International Workshop on Design and management of Data Warehouses*, Stockholm, Sweden, 2000.
- Jarke, M., Lenzerini, M., Vassiliou, Y., and Vassiliadis, P. *Fundamentals of Data Warehouses*. Springer-Verlag 2000.
- Jeusfield, M. A, Quix, C., and Jarke, M. "Design and Analysis of Quality Information for Data Warehouses," in *Proceedings of the 17th International Conference on Conceptual Modeling*, Singapore, Nov. 16-19, 1998.

0 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/proceeding-paper/architecture-gene-expression-data-warehouse/32015

Related Content

Design of Healthcare Lighting in Medical Centers Based on Power Carrier Communication

Yan Huang and Yongfeng Zhang (2023). *International Journal of Information Technologies and Systems Approach* (pp. 1-14).

www.irma-international.org/article/design-of-healthcare-lighting-in-medical-centers-based-on-power-carrier-communication/324748

Profit Maximizing Network Modeling With Inventory and Capacity Considerations

Tan Miller and Renato de Matta (2018). *Encyclopedia of Information Science and Technology, Fourth Edition* (pp. 5503-5515).

www.irma-international.org/chapter/profit-maximizing-network-modeling-with-inventory-and-capacity-considerations/184252

A Semiosis Model of the Natures and Relationships among Categories of Information in IS

Tuan M. Nguyen and Huy V. Vo (2013). *International Journal of Information Technologies and Systems Approach* (pp. 35-52).

www.irma-international.org/article/a-semiosis-model-of-the-natures-and-relationships-among-categories-of-information-in-is/78906

Determinants of the Use of Knowledge Sources in the Adoption of Open Source Server Software

Kris Ven and Jan Verelst (2012). *Knowledge and Technology Adoption, Diffusion, and Transfer: International Perspectives* (pp. 287-304).

www.irma-international.org/chapter/determinants-use-knowledge-sources-adoption/66951

Fault Analysis Method of Active Distribution Network Under Cloud Edge Architecture

Bo Dong, Ting-jin Sha, Hou-ying Song, Hou-kai Zhao and Jian Shang (2023). *International Journal of Information Technologies and Systems Approach* (pp. 1-16).

www.irma-international.org/article/fault-analysis-method-of-active-distribution-network-under-cloud-edge-architecture/321738