

Virtual Database for E-Agent

Megha Sawlani¹ Anita Nagdev² Monica Khubani³ Divya Paryani⁴ Dimple Bohra⁵

^{1,2,3,4,5}Department of Computer Engineering

^{1,2,3,4,5}Vivekanand Education Society Institute of Technology Mumbai, India

Abstract— This proposed paper describes the complete architecture of Virtual Database Technology (VDB)^[1] in detail. VDB can be created and managed using Virtual Database Management System (VDBMS). The description of VDBMS components are explained in this paper. This technology can be used further for implementation of distributed database applications as well. The proposed system i.e App which is about comparing different products from different (Four) Websites. Basically the App uses the technology of Virtual Database(VDB).

Key words: Virtual Database, Wrapper[1][2], RDBMS

I. INTRODUCTION

Virtual database technology makes the Internet and other data sources behave as an extension of an RDBMS system. According to some estimates, 85 % of the world's data is outside of relational database systems. Important data is fling across web sites and database systems. These data sources organize the data sources in different ways, in the vocabulary they use, and in their data-access mechanisms. Writing applications that combine data from these sources is a complex task as heterogeneity is involved.

II. TECHNOLOGY ARCHITECTURE

A. The Virtual Database:

A simple Virtual Database (VDB) run-time view is as shown in Figure 1^{[1][2]}. In this VDB the contents of two bookstores are integrated and a unified schema with two tables, books and reviews is presented. The database application operates on this unified schema, issuing SQL queries through the JDBC or ODBC API; the application is built using standard RAD tools such as Power Builder ,Visual Basic, Delphi, or similar Java toolkits.

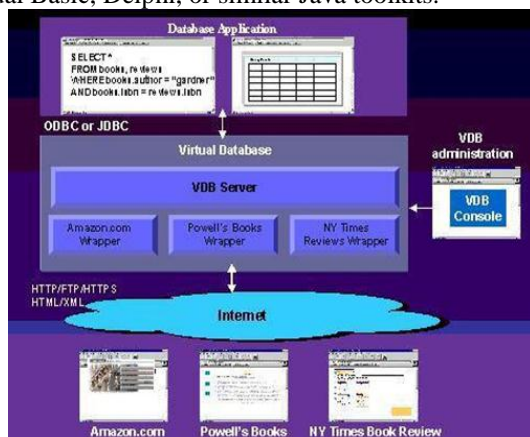


Fig. 1: The Books Virtual Database^[1]

The VDB is approached through the VDB Server, and is managed through the browser-based VDB Console. The VDB also consists each external data source, a *wrapper* that interfaces the data source to the VDB server. A wrapper makes an random external data source, such as a web site, behave like an RDBMS, while the VDB Server merges these separate relational databases into a one single Virtual Database (VDB).

An individual wrapper in action is shown in Figure 2^[2]. The wrapper interfaces with the web site, using HTTP and HTML. It manages HTTP protocol-related issues like forms, authentication and cookies. The wrapper is accessed via the JDBC API, through which clients can issue SQL queries. A SQL query provided to the wrapper in this case would result in the wrapper filling out a HTML form on the web site, parsing and navigating the resulting HTML pages, and altering the data into rows of a relational table. Extraction rules that apply enlightened linguistic processing to obtain attributes from the web pages, data transformation rules to map and format the data to fit the schema, and data validation rules to protect data integrity used by the wrapper.

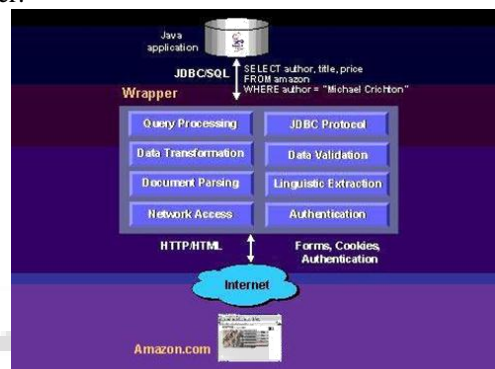


Fig. 2: A standalone wrapper^[1]

The application perceives each data source as a separate JDBC source with its own schema, and connect to each source separately. Then combine the data as required.

The full functionality of the VDBMS is used by sophisticated applications more than a few data sources , as shown in Figure 3^[2]. The VDBMS reveals tables in many data sources as virtual tables in a single Virtual Database (VDB), and supports full RDBMS functionality over virtual tables containing view definitions and query processing across sources. In the example of Figure 3, the VDB explains the view books as the merger of the amazon and powell's virtual tables. When the VDBMS accepts the query, the query processor decomposes the query, decides the fragments to be sent down to the individual data sources, and integrate their results. The results from data sources for performance is cached by query result cache.

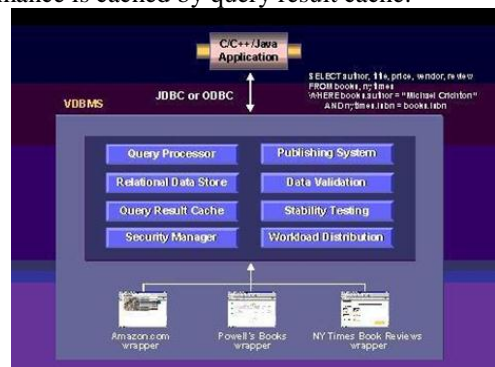


Fig. 3: From standalone wrappers to an integrated Virtual Database^[2]

The publishing system can be set up to periodically create physical snapshots of virtual tables in a local relational data store. The VDBMS can perform data validation tests. Those tests are more enlightened than those at individual wrappers; an example is stability testing, which differentiates data against historical statistical trends and elevate an alert if there is a large alteration.

B. The Virtual Database Management System

The creation and controlling of Virtual Databases is carried by Virtual Database Management System (VDBMS). Figure 4^[1] shows the components of the VDBMS.

1) Wrapper Development Kit (WDK):

The rapid creation of wrappers for web sites, file systems, and other network data sources is enabled by the Wrapper Development Kit. The WDK is built around the notion of wrapper frameworks. A wrapper framework is a collection of classes and a programming idiom that together ease the formation of wrappers for a family of data sources. For example, the Navigator framework makes it easy to create wrappers for web sites: a few simple lines of Java code can seize the structure of a complex web site, including sites whose pages are dynamically generated in acknowledgement to fill out forms. These frameworks also seize relationships between hyperlinked pages in a web site and return the relationships in the contents of the virtual database tables compatible to the site.



Fig. 4: VDBMS components^[2]

2) Extractor Development Kit

Data integration often demands withdrawing structure from unstructured textual data. For example, consider a newspaper web site that lists apartments for rent. The application requires a table with columns for attributes such as number of bedrooms, number of bathrooms, location, and rent. However, each apartment categorized listing is a block of identical text. Extraction rules relates how to extract the required features from the text.

The Extractor Engine is an interpreter for Jel (Junglee Extraction Language), a language outlined to express sophisticated text processing rules. Extraction rules depend on dictionaries of words and phrases with characteristics. For example, a location extraction rule may use a dictionary that lists the names of cities and states in the US, together with common acronyms for the names. The EDK covers Dictionary Management Utilities to create and control such dictionaries.

III. APPLICATION OF VIRTUAL DATABASE IN E-AGENT

The need to compare the prices of products crave to be purchased is absolutely common in all individuals. However clarifying this slow task of comparison will not only save the users time, it can also act advantageous to the builder of such a system.

As the demand for such comparators is high and acknowledging the booming exposure of online shopping in all sections from stationery to electronic goods, our project motivates to bridge the gap between the product and the user by providing him various prices of the product from various websites.

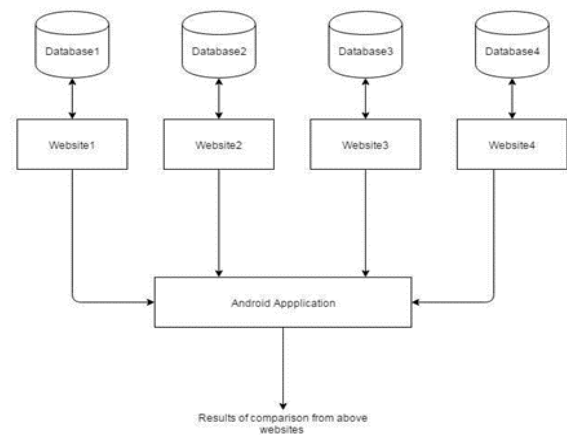


Fig. 5: Block Diagram of Proposed System.

The proposed design is as follows.

An app that searches the product with respect to keywords entered. These are then compared with the app's database and a match of the products are displayed.

The product once specified is searched over various sites (all of which are open to sharing the information associated to the product – which in our case are the sites we develop) A general review of the product and the price range in the divergent sites is presented leaving it to the user to use his fine perception of judgment to select the best.

IV. ABBREVIATION AND ACRONYM

Abbreviations and acronyms used in paper are:

- VDBMS: Virtual Database Management System
- VDB: Virtual Database
- WDK: Wrapper Development Kit
- ODBC: Oracle Database Connectivity
- JDBC: Java Database Connectivity
- SQL: Structured Query Language
- OLAP: Online Analytical Processing
- WWW: World Wide Web
- HTML: Hyper Text Markup Language
- HTTP: Hyper Text Transfer Protocol
- RDBMS: Relational Database Management System

V. CONCLUSION

VDB technology enables rapid deployment of applications. The Internet, and most Intranets have all of the following features and can benefit from Virtual Database technology:

Enormous numbers of data sources.
Data sources arc self-governing there is no centralized control

REFERENCES

- [1] Anand Rajaraman Peter Norvig: Virtual Database Technology-Transforming Internet into a database,1998
- [2] <https://who.rocq.inria.fr/Anne-Marie.Vercoustre/REUSE/pos8/>