Department of Telecommunication

Analyzing the Web Site Traffic Using Data Warehouse Tools

Branimir Putniković, Boris Vrdoljak
University of Zagreb
Faculty of Electrical Engineering and Computing
Address: Unska 3, HR-10000 Zagreb, Croatia
E-mail: branimir.putnikovic@fina.hr, boris.vrdoljak@fer.hr

Introduction

- Data warehousing tools – for a flexible interactive analysis of the web site traffic.
- Data warehousing – the process of planning, building, using and maintaining a database, where

  data is collected for the purpose of being analyzed.
- We analyze the Web site of the Department of Telecommunication (www.tel.fer.hr).
- To improve the organization of the site,

  enable better presentation of subject materials and information about classes.
- Access log files – data source.

Motivation

- Existing web site traffic reporting tools:
  - fixed reports with simple statistical analysis
  - don't allow ad-hoc analytical queries
  - they give too little details
  - cannot separate data about the particular part of the web site
  - cannot easily change their query scripts

- Using data warehouse concept – everything is possible and more!

What do we get?

Using data warehouse concept for analyzing the web site traffic we get:

- More flexible and interactive analysis.
- Ad-hoc analytical queries become reality.
- Users can dynamically compose or change their own queries.
- The data stored in the warehouse can be “drilled up and down” (by changing the hierarchy level) in order to get more or less details.
- Users can get more details about one particular part of the web site.

The data available from the log files provides statistics such as:

- the subject with top access in selected academic year,
- accesses within selected date or month by hours during the day,
- type of contents that selected users were accessing.

Data warehouse tools

- Data warehouse system:

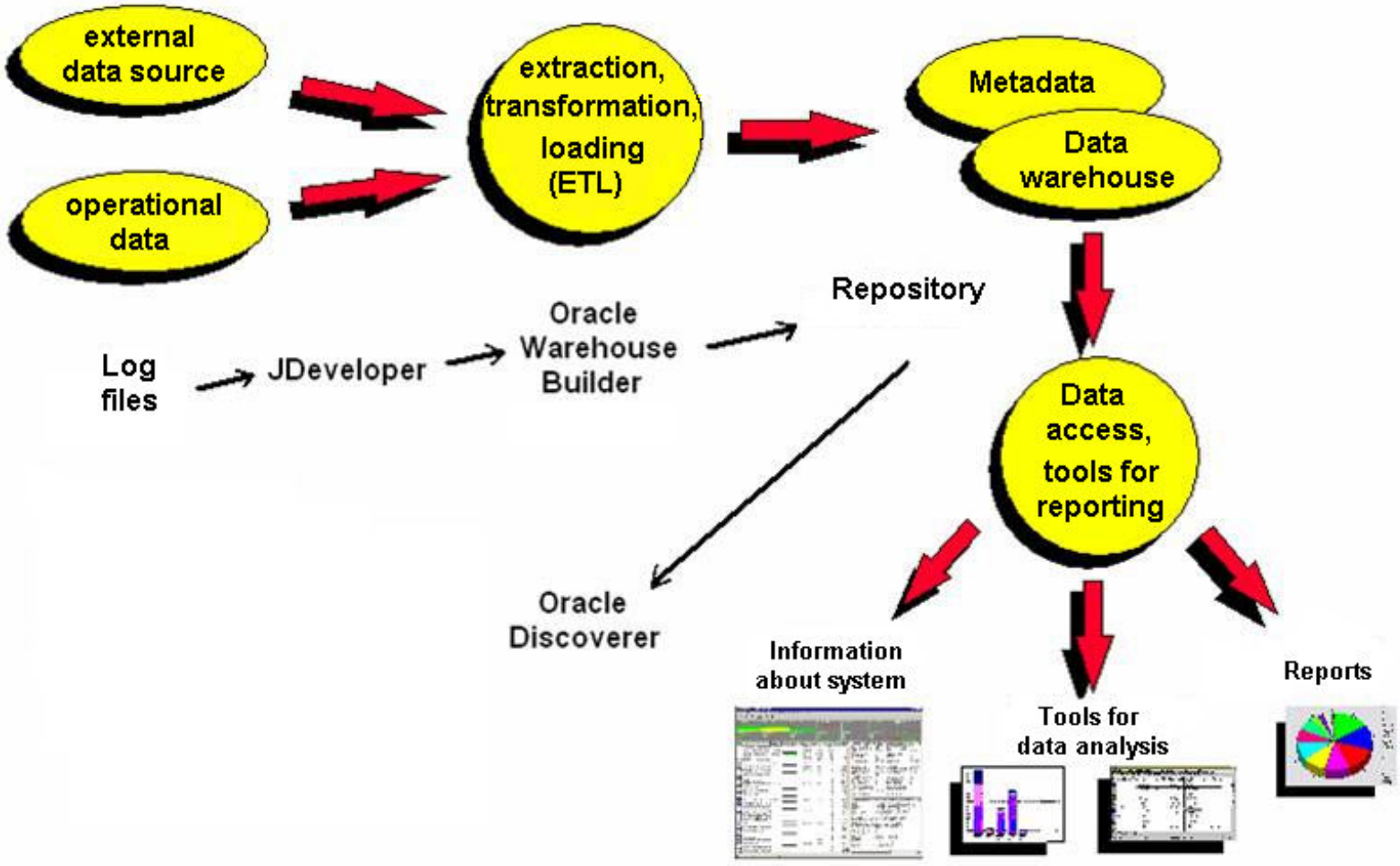
  Oracle9i database, Oracle JDeveloper, Warehouse Builder and Discoverer.

- Program “Logs” – analyzes log files and converts them in a form favorable for loading in the warehouse.

- Warehouse Builder – a central tool for modeling and designing data warehouse, definition of the ETL process.

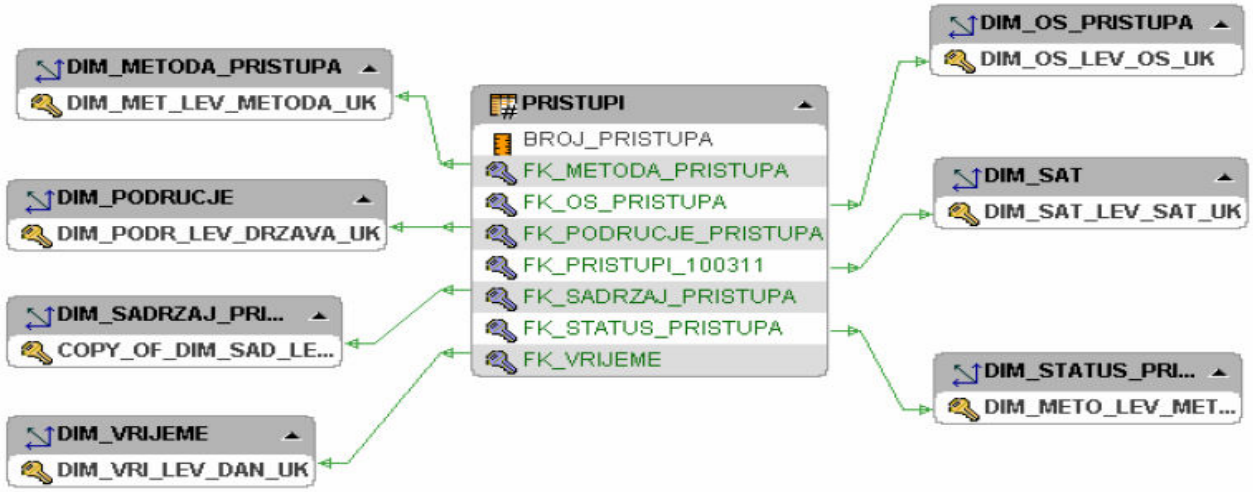
- Discoverer – a tool for making reports and viewing data from the data warehouse.

Data extraction and transformation



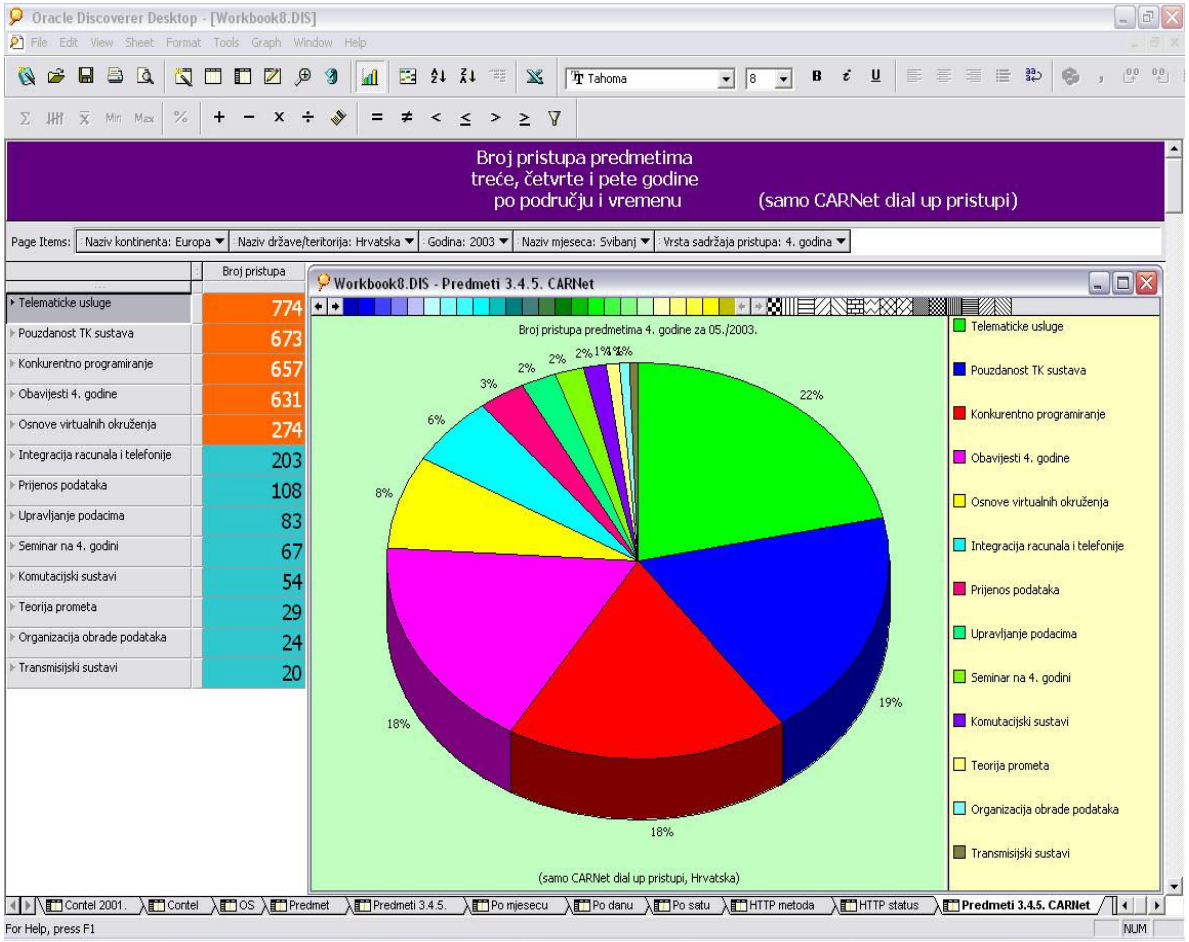
Data storage

- Data – stored in a relational database
- Tables – organized in a “star schema”
- The star schema – composed of a dimensional tables (territory, time, operating system, content, HTTP status code, HTTP method and hours) with single-part key and one fact table with a multi-part key.
- Hierarchy – expressed explicitly in the dimension tables where hierarchical levels are shown as attributes.



Reports

- Number of accesses to subjects from the 3rd, 4th and 5th academic year by chosen territory and time period.
- Selected – the 4th year, accesses only from CARNet (dial up users), May 2003, Croatia.
- Subjects are sorted by their number of accesses and top 5 are colored orange .
- More details – drilling through hierarchies .



Conclusion

- We proposed the usage of data warehouse concept for analyzing the web site traffic.
- Web server access log files – a data source.
- The Logs program – analyzes log files and prepare the data for loading into the data warehouse.
- Data is viewed and reports are made using Discoverer.
- Discoverer – multidimensional view of data, a flexible and interactive access to data.
- Users can navigate through hierarchy of data and customize the view in the way they want.

References

- [1] R. Kimball, *The Soul of the Data Warehouse*, Wiley Computer Publishing 2003.
- [2] V. R. Gupta, *An Introduction to Data Warehousing*, System Services Corporation, Chicago, Illinois, 1997.
- [3] B. Vrdoljak, G. Gledec, Z. Skočir, *An Application for Multidimensional Analysis of the Web Site Traffic*, Proc. of the 10th IEEE Mediterranean Electrotechnical Conference MELECON 2000, Limassol, Cyprus 2000.
- [4] Oracle Corporation, *Oracle9i Developer Suite Documentation,* Oracle Corporation 2003.
- [5] Oracle Corporation, *Oracle9i Warehouse Builder 9.2. User's Guide,* Oracle Corporation 2003.