▼ Introduction to Data Warehousing and Business Intelligence Minitrack

Hugh J. Watson
Terry College of Business
The University of Georgia
hwatson@arches.uga.edu

Robert Winter
Institute of Information Mgt
University of St. Gallen
robert.winter@unisg.ch

Barbara H. Wixom
McIntire School of Commerce
The University of Virginia
bwixom@mindspring.com

Since its establishment in the late 1980s, data warehousing and business intelligence have become an integral part of business information management in most organizations. Data warehousing and business intelligence speed up decision processes and provide timely, accurate, integrated information for tactical decision making. In contrast to computer science where performance and architectural issues dominate, the information systems discipline focuses on the organizational and economical aspects of data warehousing and business intelligence. The data warehousing process is comprised of sub-processes which are development oriented, operations oriented or use oriented. Due to the business-to-IT character of these processes, the information systems perspective on data warehousing and business intelligence combines contributions from management and from computer science.

Regarding development, information requirements need to be elicited, consolidated, and aligned with available data. The economics of data integration and the tradeoff between different data quality properties must be understood for effective and efficient information logistics design. Regarding operations, effective processes for extracting, cleaning, and aggregating data from source systems must be implemented, and applications must be supported appropriately. Regarding use, the field comprises general communication oriented issues ("data marketing") as well as standard reporting, ad hoc analyses, and process support.

In particular process-centric business intelligence has challenged the traditional interpretation of data warehousing as a one-way data provider: New architectural and organizational challenges result from near-real-time demand for analytical data and from the need of business processes to update analytical data.

In addition, the various centralization and autonomy options in data warehousing and business intelligence need to be better understood in order to make situational recommendations.

In their paper "From Federated Databases to a Federated Data Warehouse System," Berger and Schrefl from the University of Linz, Austria, identify requirements on a federated data warehouse system and propose an architecture which supports the tightly coupled integration of heterogeneous data marts into a global, logical schema. In order to enable the processing of queries in the federation, their approach provides a dimension algebra and fact algebra to define the mappings between the global and local schemas.

Bucher and Dinter from the University of St. Gallen, Switzerland, address "Process Orientation of Information Logistics—An Empirical Analysis to Assess Benefits, Design Factors, and Realization Approaches." They describe empirical research providing substantial evidence that organizations can realize benefits by adopting process-centric business intelligence. The dominant design factors and distinct realization approaches of process-oriented information logistics are analyzed. Based on their findings, situational methods can be engineered which support the configuration of data warehousing / business intelligence reference solutions to individual company characteristics.

The paper "Measuring Data Believability: A Provenance Approach" by Prat (Essec, France) and Madnick (MIT, USA) proposes new ways to manage data quality. Being crucial for operational efficiency and sound decision making, data quality is becoming increasingly important as organizations strive to integrate an increasing quantity of data. They focus on believability, a major aspect of quality, which they propose to measure along three dimensions: trustworthiness of sources, reasonableness, and believability based on time. Their approach is grounded on provenance, i.e. the origin and subsequent processing history of data. They define metrics for assessing the believability of data sources, define metrics for assessing the believability of data resulting from one process run and finally assess believability based on all the sources and processing history of data.