INTRODUCTION

Data warehousing is an emerging technology that greatly extends the capabilities of relational databases specifically in the analysis of very large sets of time-oriented data. The emergence of data warehousing has been somewhat eclipsed over the past decade by the simultaneous emergence of Web technologies. However, Web technologies and data warehousing have some natural synergies that are not immediately obvious. First, Web technologies make data warehouse data more easily available to a much wider variety of users. Second, data warehouse technologies can be used to analyze traffic to a Web site in order to gain a much better understanding of the visitors to the Web site. It is this second synergy that is the focus of this article.

DATA WAREHOUSE

A data warehouse is a repository of nonvolatile temporal data used in the analysis and tracking of key business processes. Temporal or time varying is the most important characteristic that distinguishes a data warehouse from a traditional relational database, which represents the state of an organization at a point in time. A relational database is a snapshot of the organization, whereas the data warehouse is a collection of longitudinal data.

One could argue that it should be possible to store longitudinal data in a relational database, and this claim is true. However, relational databases, which model data as entities, create severe limitations in data exploitation.

First, although standard SQL does provide a DateTime data type, it is very limited in its handling of dates and times. If an analyst wanted to compare summary data on weekends versus weekdays or holidays versus non-holidays, it would be difficult if not impossible using standard SQL. Second, analysis involving drill down or roll up operations becomes extremely awkward using standard SQL against entities as represented in relational tables.

Data warehousing technology overcomes these deficiencies in the relational model by representing data in a dimensional model. A dimensional model consists of a fact table (see Figure 1) and the associated dimensions. The fact table contains measures of the business process being tracked and the dimensional tables contain information on factors that may influence those measures. More specifically, the fact table contains dependent variables while the dimension tables contain independent variables. Online analytical processing (OLAP) tools provide a means of summarizing the measures in the fact table according to the dimensions provided in the dimension table toward the end of determining what factors influence the business process being modeled. Typically OLAP tools provide a means of easily producing higher levels of summary (roll-up) or greater levels of detail (drill-down).

WEB LOG

A visitor to a Web site requests a page by typing in the address of the page in a Web browser, or by clicking on a link that automatically requests that page. A message is
Related Content

Information Technology Adoption and the Role of Organizational Readiness: The Case of an Indian Bank
www.igi-global.com/article/information-technology-adoption-role-organizational/3205?camid=4v1a

Cloud Computing for Secure Services in E-Government Architecture
www.igi-global.com/article/cloud-computing-for-secure-services-in-e-government-architecture/127049?camid=4v1a

A Crisis at Hafford Furniture: Cloud Computing Case Study
Keith Levine and Bruce A. White (2011). Teaching Cases Collection (pp. 57-71).
www.igi-global.com/article/crisis-hafford-furniture/53556?camid=4v1a

ICT Policies in Africa
www.igi-global.com/chapter/ict-policies-africa/13383?camid=4v1a