Chapter XII

Modeling Data-Intensive Web Sites for Personalization, Integrity and Performance

Christos Bouras
University of Patras, Greece
Computer Technology Institute-CTI, Greece

Agisilaos Konidaris
University of Patras, Greece
Computer Technology Institute-CTI, Greece

ABSTRACT

This chapter presents a step-by-step approach to the design, implementation and management of a Data-Intensive Web Site (DIWS). The approach introduces five data formulation and manipulation graphs that are presented analytically. The core concept behind the modeling approach is that of “Web fragments,” that is an information decomposition technique that aids design, implementation and management of DIWS. We then present the steps that must be followed in order to “build” a DIWS based on Web fragments. Finally, we show how our approach can be used to ensure the basic DIWS user requirements of personalization, integrity and performance.
INTRODUCTION

The Web is made up of millions of Web sites, which in turn are made up of millions of interconnected Web pages. A Web site can be seen as a set of Web pages (files in general) that share a common domain name (assigned to a Web server), are linked to each other (and to external Web pages) and can usually be accessed through a gateway Web page called the “home page.” Clients request information from Web sites (through HTTP requests) and Web servers respond, by sending this information over the Internet (through HTTP responses). Some of the issues that play a central role in the formulation of the request pattern of a Web site are information popularity and scope, the language used, response times, appealing design and easy navigation. It is clear that certain Web sites are accessed more (or much more) than others, continuously or over a certain period of time. Large volumes of data are requested of these Web sites during the peak periods of user demand. These Web sites are generally called Data-Intensive Web Sites (DIWS henceforward). In this chapter, we give the following definition to DIWS:

A DIWS is a Web site that uses a database as a back-end for the storage of large volumes of frequently updated data. DIWS show high demand for data accessed through their Web pages, continuously or in specific peak periods.

One must always keep in mind that a Web site can be characterized as “Data-intensive” only after its usage. The metric of intensiveness is always vague when designing a Web site. That is why, even when one can predict that a Web site will be data-intensive, the “intensiveness” metric cannot be absolutely defined. Thus, a lot of attention must be given to request prediction and scalability, in order to answer to potentially higher demand than predicted.

This chapter will take an in-depth look at DIWS and provide a uniform solution to the following scenario:

Given a large volume of available and frequently updated data, how can a content provider utilize the Web/database paradigm to model, manipulate, present, provide, manage and maintain the data, in a way that answers to the basic user requirements of personalization, integrity and performance.

RELATED WORK

A lot of work has been done in research fields related to Data-intensive Web sites. In this section, we will review work that focuses on three basic DIWS fields. These fields are:
Related Content

Current Trends in Integrating the Internet of Things Into Software Engineering Practices

Intellectual Property and the Internet
[www.igi-global.com/chapter/intellectual-property-internet/16857?camid=4v1a](www.igi-global.com/chapter/intellectual-property-internet/16857?camid=4v1a)

Property-Based Object Management and Security
[www.igi-global.com/chapter/property-based-object-management-security/49522?camid=4v1a](www.igi-global.com/chapter/property-based-object-management-security/49522?camid=4v1a)
Load Balancing Aware Multiparty Secure Group Communication for Online Services in Wireless Mesh Networks
www.igi-global.com/chapter/load-balancing-aware-multiparty-secure/77419?camid=4v1a