Chapter 18

Mining Environmental Data in the ADMIRE Project Using New Advanced Methods and Tools

Ondrej Habala
The Slovak Academy of Sciences, Slovakia

Martin Šeleng
The Slovak Academy of Sciences, Slovakia

Viet Tran
The Slovak Academy of Sciences, Slovakia

Branislav Šimo
The Slovak Academy of Sciences, Slovakia

Ladislav Hluchý
The Slovak Academy of Sciences, Slovakia

ABSTRACT

The project Advanced Data Mining and Integration Research for Europe (ADMIRE) is designing new methods and tools for comfortable mining and integration of large, distributed data sets. One of the prospective application domains for such methods and tools is the environmental applications domain, which often uses various data sets from different vendors where data mining is becoming increasingly popular and more computer power becomes available. The authors present a set of experimental environmental scenarios, and the application of ADMIRE technology in these scenarios. The scenarios try to predict meteorological and hydrological phenomena which currently cannot or are not predicted by using data mining of distributed data sets from several providers in Slovakia. The scenarios have been designed by environmental experts and apart from being used as the testing grounds for the ADMIRE technology; results are of particular interest to experts who have designed them.

DOI: 10.4018/978-1-4666-0906-8.ch018

Copyright © 2012, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
INTRODUCTION

We present our work in the project ADMIRE, where we use advanced data mining and data integration technologies to run an environmental application, which uses data mining instead of standard physical modeling to perform experiments and obtain environmental predictions.

The chapter starts with a description of the project ADMIRE, its vision and goals, and motivation and goals specific to our work in the project. Then we describe the history and current status of the environmental application. The core of the chapter then presents our approach to the integration of data from distributed resources. We have developed a prototype of a data integration engine that allows users to specify data integration process in the form of a workflow of reusable processing elements.

THE EU ICT PROJECT ADMIRE

The project ADMIRE (Advanced Data Mining and Integration Research for Europe (ADMIRE)) is a 7th FP EU ICT project which aims to deliver a consistent and easy-to-use technology for extracting information and knowledge from distributed data sources. The project is motivated by the difficulty of extracting meaningful information by mining combinations of data from multiple heterogeneous and distributed resources. It will also provide an abstract view of data mining and integration, which will give users and developers the power to cope with complexity and heterogeneity of services, data and processes. One of main goals of the project is to develop a language that serves as a canonical representation of the data integration and mining processes.

The work presented in this chapter concentrates on using the ADMIRE technology and platform in order to bring new methods of data integration, exchange, and use to the environmental management domain, specifically in Slovakia. In the current situation, we often see environmental applications that are either using only in-house data, or use external data obtained as a static package, which over time loses accuracy simply by not being updated with new information, measurements, and corrections. An establishment of a new data-transfer contract between a provider and a user of an environmental data set is usually a cumbersome process, taking sometimes months, making some environmental data use scenarios outright impossible, since in the time it would take to obtain the data, the scenario would lose its purpose. While this situation may be alleviated by a rapid move of the data providers and data consumers in Slovakia towards the INSPIRE-prescribed web services technology (INSPIRE, 2007), we currently see almost no such efforts (and this situation is not specific to Slovakia). Only at the end of 2009 a law actually prescribing the creation of INSPIRE-compliant national infrastructure for environmental data has been passed, and it is probable that its implementation will take more than two years.

While the ADMIRE platform does not follow the INSPIRE-related directives in existence so far, the components of the applications are web services, and so surely follow the general spirit of INSPIRE. By engaging Slovak environmental management institutions (Slovak Hydro-meteorological Institute, Slovak Water Enterprise, Institute of Hydrology of the Slovak Academy of Sciences, and others) in preparing and evaluating experimental scenarios for environmental data integration and data mining, we try to achieve two goals. The first is to test the ADMIRE platform on a real application, with real and critical end-users. This goal serves the project itself. The other goal is to introduce the end-users to state-of-the-art data integration technologies, capable of negotiating and delivering environmental data on-line, without the usual delays and tedious preparation. This effort may prepare the foundation for later deployment of production web services for environmental data access, transfer, processing and visualization, as is foreseen in INSPIRE.
Related Content

Reliability and Performance Models for Grid Computing
[www.igi-global.com/chapter/reliability-performance-models-grid-computing/64481?camid=4v1a](www.igi-global.com/chapter/reliability-performance-models-grid-computing/64481?camid=4v1a)

Exploring Video Sharing Websites Content with Machine Learning
[www.igi-global.com/article/exploring-video-sharing-websites-content-with-machine-learning/119192?camid=4v1a](www.igi-global.com/article/exploring-video-sharing-websites-content-with-machine-learning/119192?camid=4v1a)

Using Simulation Systems for Decision Support
[www.igi-global.com/chapter/using-simulation-systems-decision-support/38267?camid=4v1a](www.igi-global.com/chapter/using-simulation-systems-decision-support/38267?camid=4v1a)

Towards a Generic Security Framework for Cloud Data Management Environments
[www.igi-global.com/article/towards-generic-security-framework-cloud/63633?camid=4v1a](www.igi-global.com/article/towards-generic-security-framework-cloud/63633?camid=4v1a)