Chapter 4

Storage of Simulation Result Data: A Database Perspective

François Pinet
Irstea – Clermont-Ferrand, France

Bruno Cheviron
Irstea – Montpellier, France

Nadia Carluer
Irstea – Lyon, France

Sandro Bimonte
Irstea – Clermont-Ferrand, France

Claire Lauvernet
Irstea – Lyon, France

André Miralles
Irstea – Montpellier, France

ABSTRACT

DBMS is a traditional technology for the storage of business application data. In this chapter, we show that this technology can be of interest in scientific fields. We present a survey of the emergence of the concept of simulation result database. Scientific simulation models become more complex, use more data and produce more outputs. Stochastic models can also be simulated. In this case, numerous simulations are run in order to discover a general trend in results. The replications of simulation increase again the amount of produced data, which makes exploration and analysis difficult. It is also often useful to compare the results obtained in using different model versions, scenarios or assumptions (for example different weather forecasts). In this chapter, we provide several examples of projects of simulation result databases. We show that the database technology can help to manage the large volumes of simulation outputs. We also illustrate this type of projects on environmental databases storing pesticide transfer simulation results. We conclude in highlighting some trends and future works.

1. INTRODUCTION

Simulation process is used to assess and analyse phenomena, for example related to the behaviour and the change of our environment (Hirabayashi, Kroll, & Nowak, 2011; Li & Mao, 2011; Pogson, Hastings, & Smith, 2012; Trolle, Hamilton, Pilditch, Duggan, & Jeppesen, 2011). The goal of the modelling is to understand, formalize or reproduce phenomena. Simulation models can be used to predict their future

DOI: 10.4018/978-1-4666-8841-4.ch004
evolutions or assess non-measurable values. The development of models is usually a multi-disciplinary task. Experts in different fields are often required to formalize and implement models.

The amount of data available to be analysed for reaching conclusions is constantly increasing and so does the computer performance increases over the years. Consequently, models become more complex, use more input data and produce more output results (Fernández-Quiruelas, Fernández, Cofiño, Fita, & Gutiérrez, 2011; Nakano & Higuchi, 2014; Pijanowski, et al., 2014; Steed, et al., 2013). Stochastic behaviours can also be simulated. In this case, several simulation runs can be required to discover a general trend in results. These replications increase the volume of result data, which makes exploration and analysis difficult (Boulil, et al., 2013). Scientists can also calculate aggregated results from large amounts of data produced by simulation runs (Boulil, et al., 2013; Mahboubi, Bimonte, Faure, & Pinet, 2010). It is also often useful to compare the results obtained in using different model versions, scenarios or assumptions (for example different weather forecasts).

Simulation results can have different formats. Different simulation runs can produce multiple text files, shape files, Excel files, images, etc. The format heterogeneity and the volume of produced files make it difficult the analysis and the tractability of results. For example, text-based storage makes it difficult to explore, select, and visualize the data (Boulil, et al., 2013). Spreadsheet tools (for example, OpenOffice, MS Excel) can help to display and extract the simulation results, but they are not suitable for large volumes of information (Boulil, et al., 2013).

To overcome these drawbacks, some scientists have proposed to use databases to store and extract the simulation data. Database-oriented solutions facilitate the management of simulation results. Database provides technical solutions for the storage of large volume of information and for their accesses via local network or Internet. Relational databases provide efficient methods for user authentication, integrity constraint controls, data backup, data insertion and data selection (Boulil, et al., 2013; Pokorný, 2006; Basta & Zgola, 2011).

In this chapter, we show how Database Management Systems (DBMS), a classical technology used in business applications, can be also used for simulation model applications. In Section 2, we present an overview of different examples of projects of simulation result databases. In these projects, the features of traditional database management system have been sometimes extended and adapted to the specificities of simulation results. Section 3 illustrates the design and the implementation of a simulation database on a project of environmental databases storing pesticide transfer simulation results. Section 4 concludes and provides some trends and future works in the field of simulation result database.

2. USING DATABASES FOR SIMULATION DATA STORAGE

The authors of (Pfaltz & Orlandic, 1999) propose an object-oriented DBMS called ADAMS (Advanced Data Management System) especially dedicated for scientific simulations. According to the authors, the system is scalable to manage large volume of scientific data and provides a variety of data organization capabilities, including aggregation, linear ordering and multi-dimensional clustering. To overcome the issue related to the large volume of scientific data, the ADAMS system described in (Pfaltz & Orlandic, 1999) provides the possibility to distribute the objects on different sites. Gradual data migrations are possible when the quantity of stored data increases. ADAMS can also manage spatial data which are used in models.