Performance Evaluation of Storing Inhomogeneous Descriptive Data of Digital Maps

Roberto Giachetta and Zsigmond Máriás

In contemporary Geographical Information Systems (GIS) the large variety of digital map sources requires the handling and storing of various descriptive data in a single storage facility. When even the number and type of attributes can vary map by map, the storage of such inhomogeneous data in a single database is difficult, as querying is an essential task and requires fast retrieval of data based on any present attribute. Examples of such databases can be found under the domain of the Institute of Geodesy, Cartography and Remote Sensing.

This paper presents a performance analysis study regarding the storage of descriptive data of digital map in database systems. Map description data can vary from simple numbers and text to complex data structures, like coordinates and timelines. Even with creating different categories for maps, it cannot be guaranteed that all items in the category have the same types of attributes stored. In traditional relational databases, tables must have a predefined structure according to the database schema, therefore simply providing columns for each descriptive data this kind of flexibility cannot be met, and data must be grouped and stored in given columns. Therefore fast querying features of relational databases (including indexing) cannot be used, which results in significant performance loss. The authors have studied several solutions addressing this problem.

The first solution is the usage of a semi-structured document-oriented database system called MongoDB, which defines database items as documents that can contain any kind and any number of attributes (even other documents) regardless of any predefined structure. These items are indexed according to their attribute types and values. These documents can be stored in collections, and there is no restriction on the structure of documents within a collection, therefore no restriction is needed on descriptive data.

The second solution is based on the creating an object-oriented environment inside a relational data model (based on our previous studies) and storing digital maps as objects in the database. This structure relies on creating inheritance taxonomy within the database and providing facilities to create classes and objects that can be altered in run-time. We have two approaches for this solution. In the first one, attributes' descriptions and classes are described in separate tables with a third table providing the link between a class and its attributes. Objects are stored in one table, and different types of attribute data are stored in different tables. This model has the advantage of being very flexible, when changing class structures with no need to redefine any database schema, however, queries must fetch object data using different tables, and inheritance structure is hard to reconstruct. Several improvements have been introduced for these problems (denormalization, inherited field storage, attribute table contraction). In the second approach class and attribute schemas are stored in the same way, but a separate table is created for each class to store the instances. Therefore creating, altering or even dropping tables is needed when changing schemas, but querying will be much faster.

In this study, the authors compare the original rational database solutions with these three approaches by implementing and testing with massive inhomogeneous and altering descriptive data. The article will conclude by presenting which method is best with digital map data at different search/insert/update command ratios. Since this problem is not only typical in the field of GIS, the solution can be applied generally to any domain using inhomogeneous data, like e-commerce systems and document warehouses.

Keywords: geographical information systems, large-scale spatial data storage, document-oriented databases, object-oriented databases, performace analysis.

Acknowledgements

The research is supported by the TÁMOP-4.2.1/B-09/1/KMR-2010-003 project.

References

[1] Csornai, G., László, et al., The Integrated Utilization of Satellite Images in Hungary: Operational Applications from Crop Monitoring to Ragweed Control, *Proceedings of the 26th Symposium of EARSeL, Millpress*, Rotterdam, The Netherlands, 2007.

[2] Chodorow, C., Introduction to MongoDB, *Free and Open Source Software Developers' European Meeting (FOSDEM)*, Brussels, Belgium, 2010.

[3] Giachetta, R., Elek., I., Developing an Advanced Document Based Map Server, *8th International Conference on Applied Informatics (ICAI)*, Eger, Hungary, 2010.

[4] Máriás, Zs., Design and Performance Analysis of Hierarchical Large-scale Inhomogeneous Databases, *8th International Conference on Applied Informatics (ICAI)*, Eger, Hungary, 2010.

[5] Nadkarni, P. M., et al., Organization of Heterogeneous Scientific Data Using the EAV/CR Representation *The Journal of the American Medical Informatics Association (JAMIA)*, vol. 6, issue 6, 1999.