A Model For A Spatial Data Marketplace

Mahmoud Sakr and Esteban Zimányi
mahmoud.sakr, ezimanyi@ulb.ac.be

Université Libre de Bruxelles

September 28, 2016

1 Extended Abstract

A data marketplace is a platform that facilitates online data trading. It gives sellers the capabilities of publishing data, and defining the price function, the authorization rules, and the data contract. It gives customers the capabilities to search for data services, to query for their price, to invoke them, and to pay for the usage. As data is more and more emphasized as a payable asset, enterprises are increasingly seeking solutions for trading their own data. Recently many examples exist for data sellers whom main business is not about selling data. For such sellers, who typically do not possess the infrastructure for selling data, a marketplace is the suitable platform for publishing their offering. There is a lack of marketplace proposals for spatial data. This research is trying to fill-in this gap.

This work is done within the scope of the SeCloud project[1]. The goal is to develop a marketplace platform for spatial data services. This platform, besides being a contribution itself, shall help prototyping the techniques of Data as a Service (DaaS), such as: data service description, mashups, integrating data contracts, pricing functions, data service privacy, and other DaaS-related proposals.

A simple, yet rich, data model is being developed for representing the market data. We define an $n \times m$ uniform grid that covers the whole spatial space of the marketplace, with an arbitrary resolution (e.g., $1m \times 1m$). This grid discretizes the space into a finite set of cells. Every cell has a unique identifier, which acts as its key. This grid structure acts as the container of the spatial data in the market place. When publishing a new layer to the marketplace, it is projected on the grid. Each grid cell stores the identifiers of the objects from the new layer that intersects the spatial extent of the cell. We think this model is powerful for two reasons: (1) it can be materialized in a key-value store or a document store, and (2) the same model can represent the spatial data, the pricing function, and the authorization rights.

Having a key-value or a document store in the back-end, instead of a spatial DB, is highly advantageous, because the former are cloud native, offering the favorable features of scalability, high availability, etc. It remains challenging to establish an efficient bidirectional mapping between the two dimensional spatial space and the one-dimensional (sorted-by-key) store. This challenge revives the old discussion of space-filling curves and their alternatives, such as multidimensional indexes,yet in the scope of a state-of-art cloud platform.

Thanks to the simplicity and the clean design of the grid model, it can also accommodate the pricing and the authorization information. A price cell P[key] simply contains the price associated with the corresponding data cell D[key]. And an authorization cell A[key, role] contains the access rights for a given role for D[key]. The logic is Result $= \{D[\text{key}] | \text{key} \in \text{user-query} \wedge \text{A[key, user role]}\}$, and Amount $= \sum_{\text{key}} \{P[\text{key}] | D[\text{key}] \in \text{Result}\}$. We discuss two possible optimizations when the number of grid cells is big (i.e., a fine resolution). The first optimization is to cache certain aggregates to speed up both insertion and query. The second optimization is to use a language for linear ranges to compress the grid representation.

During the presentation we will describe the requirements for a data marketplace. We will illustrate the proposed model for the spatial data marketplace, and the parts that have been prototyped so far. The involved problems approached by us will be illustrated, and those that are still open will be motivated.

[1] http://www.securit-brussels.be/project/secloud/