DATA IN THE CLOUD?

Kevin MacDonald
Web Mapping Architect
loxcel.cluster@gmail.com

Cloud-based data management services appear to be exciting and useful, but it is for the user to decide as to what data he will keep in the cloud, as the user may not have complete control over access to his data once it steps into the cloud

The "map + cloud" equation has nicely served over 600,000 MapQuest, Bing and Google maps. But is the cloud the right place to store the data for these maps – the pushpins that mark restaurant locations or the polygons that demarcate public parks?

I recently had to ask myself this question. An interloper recently copied a table of 18,000 Starbucks store locations I kept in a cloud-based database service – and then shared it with the world (This table supported the operation of a Starbucks store locator map). For me, this cloud hovered too close to the ground, and I got lost in its fog.

Amazon SimpleDB, Microsoft SQL Azure, Google Fusion Tables and many others offer free or low-cost services that take care of your data management needs, and let you focus on delivering quality maps.

It's easy to get excited by these useful and free services, yet ignore the consequences of trusting your precious spatial data to a service that in many ways is out of your control. Who owns the data? Who can access or modify the data? What happens if the service is discontinued? Is there a risk of vendor lock-in?

In the cloud or not, it's your data. By copying it to the cloud, however, you grant the vendor rights to use your data for certain purposes. As these vendors operate from the United States, the US government also gains these rights, which, for one vendor alone, it has exercised some 6,000 times over the past six months (a 30 per cent increase from the previous six months). You might own the data you put in the cloud, but you lose some control what happens to it after it gets there.

Each vendor also has its own idea of what it means to be a database.

SQL or NoSQL? Transactional consistency, eventual consistency or none?

If spatial objects and queries are supported (eg. SQL Azure and Fusion Tables), how does it work? Differences among services and between releases may vary. Feature stability is at the discretion of the vendor – not you. If you use only features common to all, if your vendor cancels its service (this happens from time-to-time) or alters the terms (e.g. institutes usage fees), you can switch to another. For now, common features include bulk import and simple key/value queries, which is quite limiting, but it gives you options.

Starbucks stores come and go on a daily basis, so the public copy of my table will quickly lose relevance and value to its audience. From now on, I will asses the business value of future data, and use it to decide if it will float in the cloud or stay in-house. There is no free ride.

A final note for all Google Map API users out there. Google just announced that they have begun metering usage of their API. Maps that consistently generate more than 25,000 normal map loads (or 2,500 stylised map loads) daily will need to pay for this service. Overage charges begin at $4 per 1,000 map loads. There are methods to reduce your map load count, depending on how your map interacts with the surrounding web page. For more details, visit the Google Map API blog.