A CHANGE OF COURSE

WITH JUST A FEW CHANGES TO OGC STANDARDS, CLOUD SERVICES COULD OFFER FAST, CONSISTENT INFORMATION TO LOCATION-AWARE APPS, ARGUES NORMAN BARKER

The market for database services that can handle geospatial data is on the rise, fuelled in no small part by the rapid growth of mobile-device applications that are increasingly using and requiring location-specific data services, be that for the delivery of advertising, mapping and guidance, delivering augmented reality annotation data – indeed, anything that needs to retrieve data from a central storage repository to deliver a location-based response to the user.

Several companies, including my own, Cloudant, are developing geospatial database services using a variety of underlying database and hosting technologies, to service this demand for cloud-based database services that can underpin location-aware applications.

For example, we have built a data layer using a combination of Apache CouchDB, an Amazon Dynamo-like elastic clustering platform, and callable functions such as full-text indexing and search. All of these act in concert to store and retrieve documents, index and serve particular views of the stored data, and serve Couch-compatible applications. Solutions such as this are then provided to customers – usually developers – as a database-as-a-service (DBaaS), either as a public cloud solution or as a hosted private cloud implementation in which data security, compliance or other customer requirements make shared hosting impractical or unacceptable.

In all cases, DBaaS solutions should provide a scalable database that is distributed and fault tolerant, so that if it fails, it gracefully recovers with no data loss to the customer or their end-user.

The heart of the application

When we add to DBaaS solutions the dynamic of location-based services, these solutions stop being just repositories in the cloud for datasets and become the heart of many applications: they
provide not only the data that they deliver to the user, but to some extent the underlying logic (the location-aware context) that makes the application function and useful to the end user.

Location-based services are a method of delivering relevant data to a user based on his or her current position. Part of a service guarantee is delivering the requested data with minimal latency. A geospatial cloud that is providing services beyond a localised area typically spans multiple data centres and routing users to the nearest data centre is essential for providing a usable service. This can be provided by using access control lists or by using the IP address of the user to locate the request.

A geospatial system that needs to distribute can only provide two of the following three guarantees according to the well-known ‘CAP’ theorem:

- Consistency (all database clients see the same data, even with concurrent updates).
- Availability (all database clients are able to access some version of the data).
- Partition tolerance (the database can be split over multiple servers).

Not using transactions is a cost-effective model for a large distributed system of networks and devices since this provides the guarantees of availability and partition tolerance that are essential when delivering location-based services in a distributed system.

**The cost of consistency**

If consistency is a requirement then the cost of transactional updates to implement the service will require more resources to provide this with either availability or partition tolerance. Multi-version concurrency control (MVCC) and over the wire replication is a method to get an eventually consistent system. Eventual consistency with online networks is typically measured in seconds, so that users receive relevant information quickly. An advantage of using MVCC and replication is that a system or device can be disconnected and reconnected to a network at any time and then synchronised with the rest of the cluster. This scalability feature for servers is also relevant for geospatial location-based services and data collection. An ad hoc network of devices can be formed and data replicated between the members. A user with a device can receive a filtered subset of the geospatial data from the larger system, go offline and work locally to return later and synchronise.

Data layers leverage a model of eventual consistency using an over-the-wire, data replication protocol. This replication protocol can be used to filter and transfer data from servers to the edge and back again. It is this feature that can be used for location-based services and data collection to provide applications that are not supported by the always-online services supported by geospatial standards.

**The latest standards**

The OGC have two specifications that are of interest: web feature service (WFS) and the proposed GeoPackage specification. The latter defines a geospatial database format for importing data onto devices and other systems, and the WFS specification defines methods for querying and updating data between systems and devices. Traditionally, implementations of WFS include ‘LockFeature’ to ensure the serialisation of the GML feature. However, extending the WFS response to indicate resource conflicts rather than just the identifiers for newly created or updated resources would extend the applicability of WFS to include the MVCC model and hence allow implementations to pick the goals of availability and partition tolerance that can be implemented at a total lower cost than a system that has consistency guarantees.

The proposed GeoPackage specification does not include an over-the-wire protocol to transfer data to mobile devices or between systems, but does target the disconnected or limited connected network environments. With the open replication protocol, it is possible to implement a model of eventual consistency using WFS and GeoPackage, if the update results response includes indicating resource conflicts.

Eventual consistency has been proven to work in other sectors and the application of this approach to location-based services and data collection is a model that fits geospatial in a domain that has typically embraced transactional updates. Frequently disconnected and ad hoc networks fit eventual consistency. It is exciting that with minor specification changes that a MVCC model can be added to existing WFS clients and servers that will enable vendors to be interoperable and allow users to pick services that fit their requirements. Moving from an enterprise server environment to a distributed enterprise server environment to handle capacity and to improve the performance of the location-based service requires an evaluation of the guarantees of consistency, availability and partition tolerance.

Current standards have laid good groundwork for the establishment of working protocols, developer best practice and for compatibility of location-based services going forward, but there remains scope for improvement and expansion of current ratified standards. Doing so will address both the evolving nature of end-user devices, as well as the demands of applications wishing to mine geospatial data or, indeed, harvest it. For now, the OGC continues to evaluate specifications, and listen to the input of members of the organisation as to how to augment and ratify these as workable industry standards in the medium-term.

**IT IS EXCITING THAT WITH MINOR SPECIFICATION CHANGES, VENDORS WILL BE ABLE TO BE INTEROPERABLE AND USERS ABLE TO PICK SERVICES THAT FIT THEIR REQUIREMENTS**

Norman Barker is director of geospatial at Cloudant (www.cloudant.com)