

Publicación y uso de scripts SQL en servidores WPS transaccionales

V. González¹, B. Schäffer² y F. González³.

¹Universidad Politécnica de Valencia
Camino de Vera, s/n 46022 Valencia
vgonzalez@itaca.upv.es

²Geoprocessing Community
52° North Initiative for Geospatial Open Source Software GmbH
Martin-Luther-King-Weg 24, 48155 Münster
schaeffer@uni-muenster.de

³Ingeniero Informático
Autónomo
fernando@fergonco.es

Resumen

Este documento trata sobre la extensión transaccional de los servicios WPS y la creación y ejecución de scripts SQL con la ayuda de GearScape. Detalla además el trabajo realizado en el contexto del 52° North Student Innovation Prize for Geoinformatics, consistente en la integración, mediante el uso de perfiles, de los scripts SQL de GearScape en la implementación realizada por 52° North de dichos servicios transaccionales.

Palabras clave: SQL, WPS, transaccional, script, geoprocésamiento.

1 Introducción

En los últimos años se han dedicado muchos esfuerzos al desarrollo de Infraestructuras de Datos Espaciales (IDE) a distintos niveles (local, nacional, etc.) [1]. Todos estos esfuerzos tienen los mismos objetivos: maximizar el acceso a los datos y reducir la inversión y la redundancia de esfuerzos. Sin embargo, los esfuerzos actuales en materia de IDE se han centrado únicamente en la obtención y visualización de datos [2]. La directiva europea INSPIRE ha añadido servicios web para procesar datos en un entorno distribuido. Esto resulta especialmente necesario

debido a que el número y tamaño de conjuntos de datos espaciales se está incrementando rápidamente. Este incremento es debido a la disponibilidad de técnicas de adquisición con más nivel de detalle, tales como escaneo láser aerotransportado o la adquisición de datos por medios colaborativos (crowd-sourcing). La complejidad de las operaciones aplicadas y del análisis de estos datos está creciendo debido a la mayor complejidad de los nuevos modelos.

Para tratar estas cuestiones, el OGC (Open Geospatial Consortium) especificó la interfaz OGC Web Processing Service (WPS) [3]. Dicha especificación describe un método estandarizado para publicar y ejecutar cualquier tipo de geoproceto. Ya existen algunas implementaciones de este estándar, tales como el 52° North WPS. El trabajo que se presenta aquí se basa en esta implementación libre del estándar WPS desarrollada por la iniciativa 52° North y consiste en desarrollar un perfil para la implementación SQL de GearScape que permita a los usuarios incluir procesos definidos por ellos mismos en un servicio WPS y publicar dicho proceso de manera estandarizada.

2 GearScape. Scripts SQL

GearScape es un Sistema de Información Geográfica de escritorio que proporciona un lenguaje para definir procesos completamente independiente del formato y origen de datos. Así, un proceso definido con este lenguaje puede tomar como entradas (y producir como salidas) ficheros, tablas de bases de datos, etc. sin que esto se refleje en el código. En el ejemplo siguiente podemos ver la creación de un resultado llamado “res” que contendrá un buffer sobre cada geometría de una fuente de datos llamada “src”, independientemente de dónde y cómo estén almacenados los datos de origen:

```
CREATE TABLE res AS SELECT Buffer(the_geom, 10) FROM src;
```

El lenguaje proporcionado por GearScape implementa un subconjunto de las características del estándar SQL92 [4] y se extiende espacialmente mediante la especificación OGC para fenómenos simples [5][6]. De esta manera es posible la manipulación de datos geográficos almacenados en ficheros, bases de datos remotas, etc. de la misma manera que serían tratados en un único Sistema de Gestión de Bases de Datos (SGBD) habilitado espacialmente.

El uso del estándar SQL como lenguaje de procesamiento presenta diversas ventajas sobre el uso de lenguajes imperativos. En primer lugar, SQL es más simple y, por tanto, más fácil de aprender que aquellos. Esta misma simplicidad

hace que sea menos probable introducir un error en SQL que en código imperativo. Además, los motores SQL encuentran una estrategia de procesamiento de manera automática (uso de índices, orden de las operaciones), mientras que la programación imperativa consiste precisamente en la programación de dicha estrategia.

Entre las limitaciones, podemos destacar el hecho de que un lenguaje imperativo es siempre más potente que el SQL, que está limitado por los operadores de álgebra relacional en los que se basa. Estos son muy adecuados para resolver determinados problemas pero ineficientes o insuficientes para resolver otros. GearScape soluciona esto permitiendo la extensión del lenguaje mediante la programación de nuevos operadores.

Otro de los inconvenientes es que las instrucciones SQL hacen referencia directamente a las fuentes de datos sobre las que operan, por lo que no es posible reutilizar una instrucción con dos fuentes de datos distintas sin adaptarla previamente. GearScape da solución a esto posibilitando la definición de parámetros de entrada. En el siguiente ejemplo podemos ver la instrucción vista anteriormente precedida por una sección de declaración de parámetros en la que aparece el parámetro “src”. Esta precisa declaración hace que la operación no se lleve a cabo directamente contra una hipotética fuente de datos llamada “src” sino contra cualquier fuente de datos que se pase como parámetro al *script* :

```
DOC('Source to create the buffer with');
DECLARE source:TABLE(the_geom:GEOMETRY);
DOC('Buffer size');
DECLARE size:INT;
-- End of parameter section.
CREATE TABLE res AS SELECT Buffer(the_geom, size) FROM src;
```

En cuanto a los resultados producidos, el motor SQL de GearScape es capaz de analizar el *script*, sin ejecutar ninguna de las instrucciones, y obtener la estructura de los campos de cada uno de los resultados. Esto posibilita la reutilización del script y su inclusión en cadenas de procesos más complejas sin necesidad de visualizar ni modificar su código.

De esta manera, los geoprosesos pueden ser creados y encapsulados por usuarios con conocimiento de SQL y reutilizados por usuarios sin conocimiento alguno de dicho lenguaje. Con el fin de facilitar dicha encapsulación y reutilización, GearScape proporciona un editor con validación sintáctica y semántica de los scripts, asistentes para su ejecución y un modelador para componerlos en cadenas más complejas.

3 Web Processing Service

La especificación de la interfaz OGC Web Processing Service (WPS) [3] describe un conjunto de operaciones estándar para la publicación y la ejecución de cualquier tipo de geoproceto en la web. De acuerdo con dicha especificación, un proceso se define como un algoritmo, cálculo o modelo que opera con datos geográficamente referenciados.

En concreto, la especificación WPS describe tres operaciones que pueden ser invocadas de manera independiente: *GetCapabilities*, *DescribeProcess* and *Execute*. *GetCapabilities* es común a todos los servicios web definidos por el OGC y devuelve metadatos sobre el servicio. En el caso de WPS devuelve, además, una breve descripción de los procesos ofrecidos por la instancia concreta de WPS. Para obtener mayor información sobre los procesos albergados, el WPS es capaz de devolver metadatos sobre procesos concretos mediante la operación *DescribeProcess*. Esta operación, proporciona la descripción de todos los parámetros que son requeridos para la ejecución del proceso. Basándose en esta información, el cliente puede realizar una operación *Execute* sobre el proceso deseado. De la misma manera que cualquier otro servicio web OGC, el WPS se comunica mediante peticiones HTTP-GET y HTTP-POST usando mensajes basados en esquemas XML específicos del OGC.

Adicionalmente, la especificación WPS describe mecanismos para procesamiento asíncrono, procesamiento de datos referenciados por URLs y almacenamiento de los resultados de los procesos. Esta última característica es especialmente interesante en el contexto descrito a continuación, ya que permite el acceso a los resultados de un proceso en el servidor mediante el uso de una URL y sin necesidad de ejecutarlo de nuevo. También es posible pedir los resultados directamente desde el cliente.

Existen implementaciones de la interfaz WPS que han sido aplicadas de manera satisfactoria en diferentes proyectos, tales como análisis de vulnerabilidad de acuíferos [2], escenarios de detección de amenazas de bomba [7] y generalización de mapas [8]. En el lado opuesto, una extensa discusión sobre la aplicabilidad de los WPS y sus inconvenientes actuales puede encontrarse en [9] y las direcciones futuras han sido analizadas en [10].

Sin embargo, la especificación WPS proporciona un esqueleto vacío que debe ser rellenado con implementaciones concretas de los algoritmos. Esto, unido a que la interfaz WPS no proporciona ningún medio para desplegar procesos al vuelo en tiempo de ejecución ha llevado a la propuesta de una extensión transaccional (WPS-T) [11]. Esta extensión permite la definición de perfiles de despliegue que definen la forma en la que un tipo de proceso específico debe de ser codificado

para ser entendido y desplegado mediante la interfaz WPS-T. De esta manera, modelos que previamente sólo eran accesibles en un dominio específico y difícilmente compartibles fuera de él están al alcance de cualquiera que tenga acceso a Internet, lo cual mejora considerablemente la accesibilidad. La promesa de compartir modelos geoespaciales [12] se convierte así en realidad de una forma estandarizada para poder ser incluida en las infraestructuras de datos espaciales.

4 Integración

Con el fin de integrar la potencia de GearScape en un servicio WPS se han realizado varios pasos. En primer lugar, se ha creado un perfil específico para GearScape, representado por un esquema XML, que define de manera precisa la forma en que un script SQL se debe codificar para que sea entendido por el servidor. Todos los perfiles soportados se pueden observar en la respuesta GetCapabilities del servicio WPS tal y como se muestra en [11]. Posteriormente, cuando se trata de incluir un script SQL concreto, en el perfil de la petición de inclusión se deberá proporcionar únicamente el código SQL del script en texto plano, tal y como aparece en el editor de GearScape, de forma que satisfaga la definición del perfil representada por el esquema XML:

```
<DeployProcessRequest service="WPS" ...>
  <wps:ProcessDescriptions ...>
    ...
  </wps:ProcessDescriptions>
  <wps:DeploymentProfile>
    <wps:Schema href="sqlScripts-script.xsd"/>
    <sql:script>
-- Creates a buffer
DOC('Source to create the buffer with');
DECLARE source:TABLE(the_geom:GEOMETRY);
DOC('Buffer size');
DECLARE size:INT;
-- End of parameter section.
CREATE TABLE res AS SELECT Buffer(the_geom, size) FROM src;
    </sql:script>
  </wps:DeploymentProfile>
</DeployProcessRequest>
```

Nótese que no es necesario especificar más que el código SQL ya que, como se ha comentado, es posible analizarlo para obtener las entradas y salidas de los procesos. Así, en el perfil no es necesario incluir ninguna otra información

adicional sobre las entradas y salidas del proceso que la incluya en el código del mismo. Sin embargo, cuando GearScape realiza el despliegue sí que incluye ciertos metadatos sobre los parámetros en la descripción del proceso (*wps:ProcessDescriptions*) con el fin de facilitar el uso por parte del cliente.

Una vez la petición ha llegado al servidor, ésta es comprobada y, si es validada por el perfil SQL de GearScape, se envía a su motor SQL. Entonces, el motor valida sintácticamente y semánticamente el código SQL enviado, obtiene la descripción del proceso a partir del documento XML y valida dicha descripción contra la definición de parámetros y salidas del script SQL especificado en el perfil. En este punto, si el script es válido desde el punto de vista sintáctico y semántico, y la descripción XML del proceso es coherente con la definición de los parámetros y las salidas en el código SQL el script se incluye en el servidor; si alguna de estas validaciones falla, el script no se incluye y se devuelve un mensaje de error significativo.

En lo que respecta a la ejecución, una vez los procesos están incluidos en el servidor, no existe ninguna diferencia entre ellos, ya que todos se comportan como procesos WPS independientemente del motor de ejecución interno. Así, cuando una petición de ejecución llega al servidor, éste obtiene el proceso requerido, le proporciona los datos de entrada y obtiene los resultados.

De esta manera, todas las ventajas del lenguaje SQL de GearScape se unen a las de la implementación WPS de 52° North dando lugar a una herramienta que posibilita una alta productividad tanto en la publicación de geoprocursos como en su reutilización.

Para la explotación de procesos publicados, se ha desarrollado un cliente para GearScape y el existente en uDig se ha mejorado por lo que los nuevos procesos pueden ser ejecutados desde varios clientes independientes. Cabe destacar que el cliente de GearScape hace uso de los metadatos mencionados anteriormente para ejecutar procesos remotos con la misma interfaz gráfica que un script en local.

Como prueba de concepto, se ha creado un escenario que pone de manifiesto la reusabilidad e interoperabilidad del sistema. En primer lugar, un usuario crea un script de geoprocuremento para un problema concreto con (o sin) la ayuda del editor de GearScape. En segundo lugar, se contacta con una instancia de la implementación WPS de 52° North para incluir el proceso. Finalmente, otro usuario puede ejecutar el proceso, ya sea desde GearScape, desde uDig o desde otro cliente capaz de acceder a WPS.

5 Trabajo futuro

Se está terminando la integración de datos ráster en el motor SQL de GearScape así como la integración de algoritmos de Sextante en el mismo. Por último, se considera interesante la computación paralela en el servidor así como el despliegue de modelos generados con GearScape.

Agradecimientos. Víctor González agradece a todo el equipo de 52° North Initiative for Geospatial Open Source Software GmbH y a los compañeros del Instituto para la Geoinformática de la Universidad de Münster por la oportunidad brindada para realizar el trabajo aquí expuesto, especialmente a Bastian Schäffer por su tutela a lo largo de todo el trabajo, a Christoph Stasch por su impagable labor humana y a Ann Hitchcock por la imprescindible labor logística y organizativa.

Referencias

- [1] R. Groot and J. McLauhlin (eds), *Geospatial data infrastructure. Concepts, cases and good practice*. Oxford University Press, New York: 2000.
- [2] C. Kiehle, “Business logic for geoprocessing of distributed data”, *Computers & Geosciences*, vol. 32, 2006, pp. 1746–1757.
- [3] OGC, *OpenGIS Web Processing Service*, Open Geospatial Consortium, 2007.
- [4] ISO, *Structured Query Language – 92*, International Organization for Standardization, 1992.
- [5] OGC, *OpenGIS Implementation Specification for Geographic information – Simple feature access – Part 1: Common architecture*, Open Geospatial Consortium, 2006.
- [6] OGC, *OpenGIS Implementation Specification for Geographic information – Simple feature access – Part 2: SQL option*, Open Geospatial Consortium, 2006.
- [7] B. Stollberg and A. Zipf, “OGC Web Processing Service Interface for Web Service Orchestration – Aggregating Geo-processing Services in a Bomb Threat Scenario,” *Proceedings of Web and Wireless Geographical Information Systems*, Heidelberg: Springer-Verlag, pp. 239–251.

- [8] T. Foerster and J.E. Stoter, "Establishing an OGC Web Processing Service for generalization processes", *ICA workshop on Generalization and Multiple Representation*, Portland, Oregon, USA: 2006.
- [9] A. Friis-Christensen, N. Ostlander, M. Lutz, and L. Bernard, "Designing Service Architectures for Distributed Geoprocessing: Challenges and Future Directions", *Transactions in GIS*, vol. 11, Dec. 2007, p. 799–818.
- [10] J. Brauner, T. Foerster, B. Schäffer, B. Baranski, Towards a research agenda for Geoprocessing services. Processing of the AGILE Conference 2009, Hannover.
- [11] B. Schäffer, "Towards a transactional Web Processing Service", Proceedings of the GI-Days, Münster. 2008. Münster.
- [12] Gehlot, S. & Verbree, E. 2006, "Web-based sharing of a geo-processing chain: combination and dissemination of data and services", Proceedings of the International Symposium on Geospatial databases for Sustainable Development, eds. S. Nayak & S.K. Pathan, ISPRS TC Committee IV Geo-databases and Digital Mapping, Goa, India.