Chapter 14
Geospatial Service Composition in Grid Environments

Tino Fleuren
University of Kaiserslautern, Germany

Paul Müller
University of Kaiserslautern, Germany

ABSTRACT

The composition of geospatial services (like OGC Web services) allows defining powerful geospatial simulations. In order to benefit from the compute power, parallel execution capabilities, and data management possibilities of a Grid environment (like OGSA), the workflow designer has to meet several challenges including heterogeneous service technologies (like WSRF or REST) large amounts of geospatial data, parallel execution, and instable Grid resources.

The workflow management system enacting these workflows has to support parallelism in an efficient way; it must avoid becoming a bottleneck, which is often the case with classic centralized workflow engines, where all data has to pass through the engines.

This chapter presents a workflow enactment system that maintains the robustness of centralized control (using service orchestration), but is enhanced by distributed components called "proxy services" that can communicate with each other to allow for efficient coupling between parallel tasks and avoiding of unnecessary data transfers (using service choreography).

INTRODUCTION

This chapter discusses the challenges that have to be dealt with when combining existing services to create geospatial service compositions, also known as geospatial workflows that retrieve and process spatial data in a Grid or Cloud environment. Already existing services could be provided by third parties that define the interface, decide

DOI: 10.4018/978-1-60960-192-8.ch014

which service technology is used, own the sources of the service, and deploy the service on their computers. A workflow management system must be able to orchestrate such different services without requiring special service technologies or particular service interfaces. This chapter discusses a workflow management system addressing these challenges.

Web Services originating from Spatial Data Infrastructures (SDI) and Grid services are integrated into workflows to perform geospatial simulations that are large-scale, long-lived, and data-intensive. Since spatial data is highly complex and significantly large, especially if multi-dimensional data is involved, the complexity of algorithms is also affected. These geospatial workflows take advantage of being enacted in a Grid environment, where it is possible to execute tasks on several compute resources in parallel. These compute resources cannot only be the nodes of one compute cluster, but several clusters around the world that are integrated in the Grid environment.

While spatial technology is highly standardized through two international standardization bodies, i.e., the Open Geospatial Consortium (OGC, www.opengeospatial.org) and the ISO (International Standardization Organization), a vital market focusing on spatially aware software has been established within the past ten years. There is a noticeable tendency of most software vendors to integrate interfaces and Web service technologies defined by the OGC (Percivall, 2003).

The Open Grid Forum (OGF, www.ogf.org) is the driving force behind the developments in the field of Grid Computing. Grid computing environments offer impressive promises to overcome problems regarding management of distributed spatial data as well as storage capacity of hitherto unknown degree for spatial data dissemination. Grid infrastructures are typically complex and consist at least of data management services, resource allocation management services and mechanisms providing a high degree of security. Most Grid middleware implementations make use of Web services for communication between resources and use Transport Layer Security (TLS) and its predecessor, Secure Sockets Layer (SSL) for encryption on otherwise insecure communication channels, such as the Internet. Cloud computing is an evolution of Grid computing in that it hides the underlying Grid infrastructure and offers a simpler and more user-centric way of how a customer can access its services. Consequently, geospatial workflows can be made available in a Cloud environment for use as a service on demand (Armbrust et al., 2009).

Because Spatial Data Infrastructures and Grid environments have developed independently of each other, a way to integrate the different technologies needs to be found. OGC and OGF established different technologies and architectures: the abstract geospatial architecture (ISO19119) and the Open Grid Services Architecture (OGSA) (Foster et al., 2005). Both approaches use different types of Web service technology. Services in Grid environments show a great difference to standard Web services, as they are stateful in nature. Therefore, the envisaged geospatial workflows integrate different types of services that must be supported by the workflow engine.

In Fleuren & Müller (2008), we described the proxy-based workflow system we developed within the "Spatial Data Infrastructure Grid" project (GDI-Grid) (von Voigt, 2009). Proxy services were introduced to support the integration of different types of services in geospatial workflows, which we execute in the domain scenarios of the project. These types of services enable us to combine Web services, OGC Web services, Grid services, and Grid jobs directly in workflow descriptions.

In this chapter, we enhance the proxy services in order to improve the performance of parallel execution and to provide efficient data handling. It would be possible to design a new workflow engine that is optimized to cope with these issues. Nevertheless, using an industry standard workflow system as an integral component offers

Related Content