Astronomical Data Analysis Software and Systems XII
ASP Conference Series, Vol. 295, 2003
H. E. Payne, R. I. Jedrzejewski, and R. N. Hook, eds.

A Collaborative Extension to the Solar Web Tool

Romain Linsolas, Isabelle Scholl & Eric Legay

Institut d'Astrophysique Spatiale - CNRS, Bât 121 - Université Paris-Sud, F-91405 Orsay Cedex, France

Abstract. The Solar Web Project is a tool that permits the consultation of distributed and heterogeneous solar observations databases. It is currently based on a 3-tiered architecture. Its main evolution is to adapt the Solar Web server to a collaborative model. Two solutions are considered: the peer-to-peer architecture and the GRID model.

1. Introduction

The number of archives of solar observations is continuously growing and the location of their storage is getting more and more scattered. Consequently, the number of existing tools or web sites for consulting these observation catalogs (i.e., metadata) is equally in augmentation. The interest of a single program capable of accessing distributed and heterogeneous archives is therefore obvious. The Solar Web Project, developed by the MEDOC IAS team, is designed to come up to these expectations, as a first step toward a Virtual Solar Observatory.

2. The Current State of the Solar Web Architecture

2.1. Principles

The architecture of the current version of Solar Web[1] (Scholl, Linsolas & Legay 2001) is based on the 3-tiered model. Clients are connected to a single server that provides them with results obtained by querying archives. Since they are heterogeneous and distributed, the Solar Web server acquired the capabilities to access distant SQL databases, FTP server (where files are stored) and web-based portals, built on PHP or CGI scripts (see Figure 1, left).

2.2. Query Processing

Solar Web uses XML for communication between clients and server, as well as for the description of accessible databases (i.e., metadata). The processing of an XML query by the server can be broken down into four steps (see Figure 1):

1. The server receives an XML query from the client. This query is parsed by the *QueryProcessor* module, which creates a subquery for each database involved. This operation is performed using an XML document describing

[1] `http://www.medoc-ias.u-psud.fr/archive/solar_web`

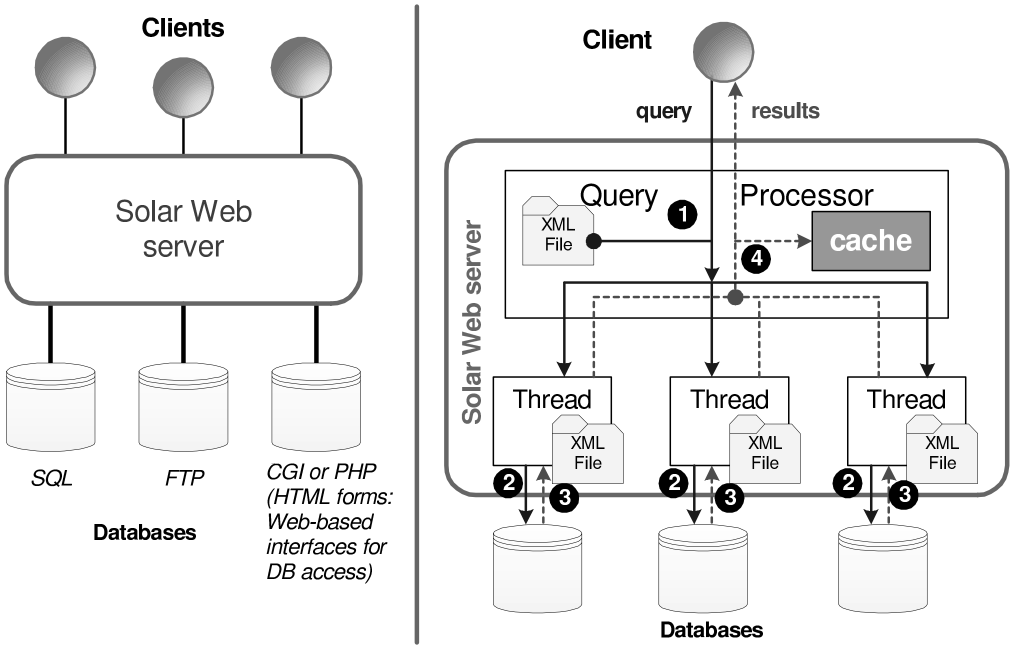

Figure 1. Current Solar Web Architecture (left) and Query Processing (right).

each accessible database. Subqueries are sent to threads dedicated in each databases access (one thread per accessible database).

2. Each thread translates an XML subquery into an query understandable by the database (e.g., a SQL query). This is made using an XML document, which describes the structure of the corresponding database.
3. The reverse processing, i.e., the translation from the database query language into the XML language, is accomplished by these threads. The XML document is also used here by the thread that receives results from database.
4. The last step of the query processing consists in merging the results obtained from accessed databases, and, if needed, in processing them (e.g., sort). Final results are then sent to the client through the server cache.

2.3. Characteristics

The centralized architecture of the Solar Web tool can have a significant impact on security, performance, and flexibility of the policy management.

First of all, since there is only one instance of the server, it may be overloaded. The outcome is that the query processing will be slowed down, especially if the number of connected clients is important. Secondly, the presence of only one server is a single point of failure from a reliability point of view. Thirdly, some data centers would prefer to have an instance of a Solar Web server at their disposal, instead of letting a distant server access directly their databases.

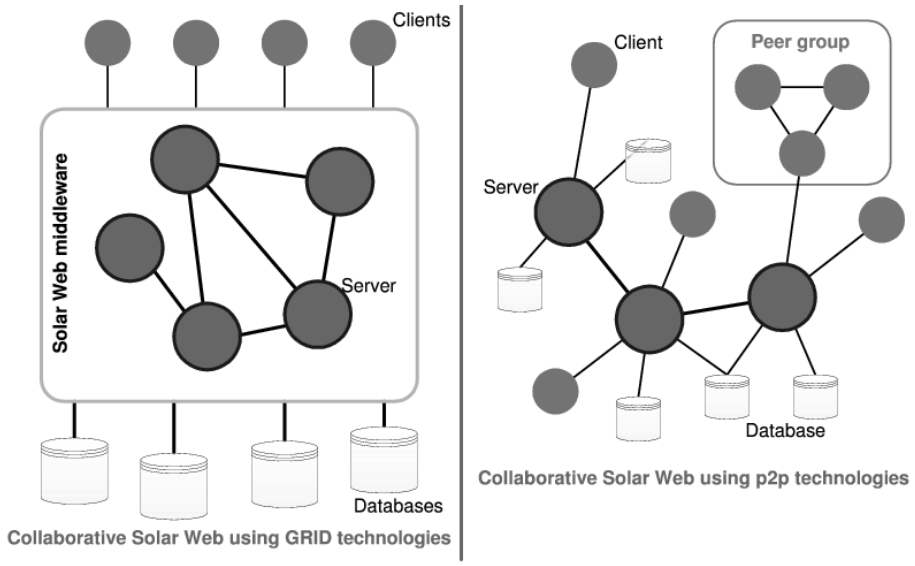

Figure 2. Future Architectures for Solar Web: GRID (left) vs. Peer-to-Peer (right).

3. Future Version of Solar Web

3.1. Objectives

Characteristics previously described can be seen as drawbacks, especially its oneness and encourage us to improve the current version of the Solar Web system. The main evolution consists in moving to a distributed architecture. Two solutions are currently under consideration:

1. One solution is to create collaborative servers, where all instances of the Solar Web server are networked together. The purpose is to distribute to all servers a query sent by one client. This architecture is inspired by GRID technologies (OGSA[2]).
2. Another solution is built on top of the peer-to-peer networks technology. It consists in redesigning the network level by using JXTA infrastructure.[3] This solution can easily provide new features such as the dynamic creation of groups of users around their domain of interest.

3.2. GRID Solution

According to Foster & Iamnitchi (2002), GRID consists of "sharing environments implemented via the deployment of a persistent, standard-based service infrastructure that supports the creation of, and resource—computers, storage space, sensors, software applications, and data—sharing within distributed computing." The concept of the GRID can be applied to Solar Web servers in order

[2]Open GRID Services Architecture

[3]Sun's peer-to-peer technology: http://www.jxta.org/

to create a network of servers, also called middleware. Clients are connected to this middleware which will provide them with different kinds of services, from data providing to remote data processing.

The GRID solution brings reliability to the Solar Web, thanks to the multiplicity of servers and therefore to the absence of a single point of failure. GRID also improves the system's performances by allowing collaborative works between servers.

3.3. Peer-to-Peer Solution

Peer-to-peer terminology refers to networks where members (also called peers) may act as client and server. As a server, it offers services such as data providers, while it takes advantage of these services as a client. The Sun's JXTA project is a platform designed for the development of applications based on this kind of technology. It provides basic mechanisms for communications, security and resources discovery for such applications.

The JXTA platform can be used for the evolution of Solar Web, as a foundation for servers and clients' applications. The Sun's platform takes care of security and communications (composed of XML and binary messages), as well as connection processing (which also includes resources and servers discovery). The creation and distribution of subqueries are based on the same mechanism as for the current Solar Web architecture. The main difference, due to the scattering of servers, is that servers have to communicate to update their metadata: since they describe the databases accessible by a server, they must be up-to-date.

Peer-to-peer technologies may also be exploited to introduce new capabilities to clients, such as the creation of users groups. These groups are built to suit to user expectations. They can be compared to collaborative workspaces that allow distributed work on data, or even discussions between scientists.

4. Conclusion

These possible evolutions for Solar Web architecture will improve the reliability and the performance of the system. Nevertheless, the principle of Solar Web (based on the use of XML files for database descriptions) will remain similar to the current implementation. In conclusion, the next version of the Solar Web tool is designed to be a powerful tool to access distributed and heterogeneous archives of solar observations. It will also provide an environment to introduce collaborative works.

References

Scholl, I., Linsolas, R., & Legay, E. 2001, in ASP Conf. Ser., Vol. 281, Astronomical Data Analysis Software and Systems XI, ed. D. A. Bohlender, D. Durand, & T. H. Handley (San Francisco: ASP), 307

Foster, I. & Iamnitchi, A. 2002, On Death, Taxes, and the Convergence of Peer-to-Peer and Grid Computing, draft