DEVELOPMENT OF WEB-BASED DISTRIBUTED DATABASE SYSTEM FOR GEOLOGIC DATA

Tatsuya Nemoto[1], Shinji Masumoto[1], Susumu Nonogaki[2] and Venkatesh Raghavan[3]

[1] Graduate School of Science, Osaka City University
3-3-138 Sugimoto, Sumiyoshi-ku, Osaka 558-8585, Japan
Email: tnemoto@sci.osaka-cu.ac.jp
[2] National Institute of Advanced Industrial Science and Technology (AIST)
1-1-1 Higashi, Tsukuba, Ibaraki 305-8567, Japan
[3] Graduate School for Creative Cities, Osaka City University
3-3-138 Sugimoto, Sumiyoshi-ku, Osaka 558-8585, Japan

ABSTRACT

A web-based distributed database system has been developed with an aim of continuing delivery of geologic data even if a large-scale disaster occurs. The distributed database system has an advantage that the system remains available if an individual database is down. The software which composes the system is FOSS (Free and Open Source Software). The distributed database is constructed by combination of PostgreSQL and pgpool-II. PostgreSQL is a powerful relational database management system. Pgpool-II which is utilized to manage distributed databases is a middleware for PostgreSQL. The prototype distributed database system for management and sharing of borehole data has been successfully implemented by using Replication mode and Parallel query mode of pgpool-II.

1. INTRODUCTION

Geologic information is one of the important spatial information supporting construction of secure and safe society. Because Japan is a country with many natural disasters, geologic information is essential as basic elements for measures against them. Recently, geologic data, such as borehole data and geological maps, and seismic, volcanic or landslide hazard maps are published via the internet by the national government, local governments, and research institutes in Japan. Many of the data is managed in a centralized database which are located and maintained in one location. It is easier to manage the centralized database system because all data resides in a single location. However, if a database fails, an entire system will be down. Actually, some of systems which deliver geologic data could not be available for several days or weeks after the 2011 off the Pacific coast of Tohoku Earthquake because the centralized database was down.

In the present study, a distributed database system has been developed to continue delivering geologic data even if a large-scale disaster occurs. The distributed database system has an advantage that the system remains available although an individual database is down. PostgreSQL and pgpool-II which can be downloaded over the internet for free are utilized to construct a distributed database system. PostgreSQL is a powerful relational database management system. Pgpool-II which is a middleware for PostgreSQL has a function for management of multiple PostgreSQL servers.

2. CENTRALIZED DATABASE AND DISTRIBUTED DATABASE

In a centralized database, all the data is stored in a single location. It is easier to manage data in the centralized database system because all data resides in a single computer. However, the centralized database system has a disadvantage that the system cannot be available if the database is down. A distributed database can solve the problem. The distributed database system remains available even if an individual database is down because the data is stored in some databases which are located in different locations.

In the present study, the distributed database system has been developed with an aim to continuing delivery of geologic data even when an individual database is down. The advantages of the distributed database system are shown below.

- Increase of reliability and availability: the system remains available even if an individual database is down.
- Protection of data: all of the data will not be lost at once because database servers are located in multiple locations.
- Improvement of performance: load is balanced because a query is executed in parallel.
- Reflection of organizational structure: each distributed database is located in the departments it relates to.

3. FUNCTIONS OF PGPOOL-II

Pgpool-II is a middleware that works between PostgreSQL servers and a PostgreSQL database client (Figure 1). In the present study, two functions shown below are used to manage the distributed database although pgpool-II has some functions for boosting reliability and performance.

Replication

It is possible to create a real-time backup on 2 or more PostgreSQL clusters by the replication function. The synchronization of data is performed by sending a same INSERT/UPDATE/DELATE query to all clusters at the same time, so that the service can continue if one of those clusters fails (Figure 2). A SELECT query used to select data from database is distributed to each cluster by activating load balancing (Figure 3).

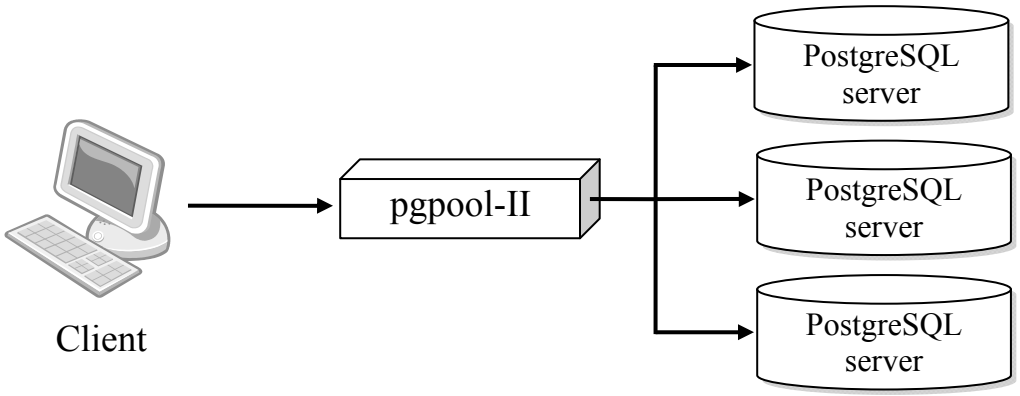

Figure 1. Management of distributed database by pgpool-II.

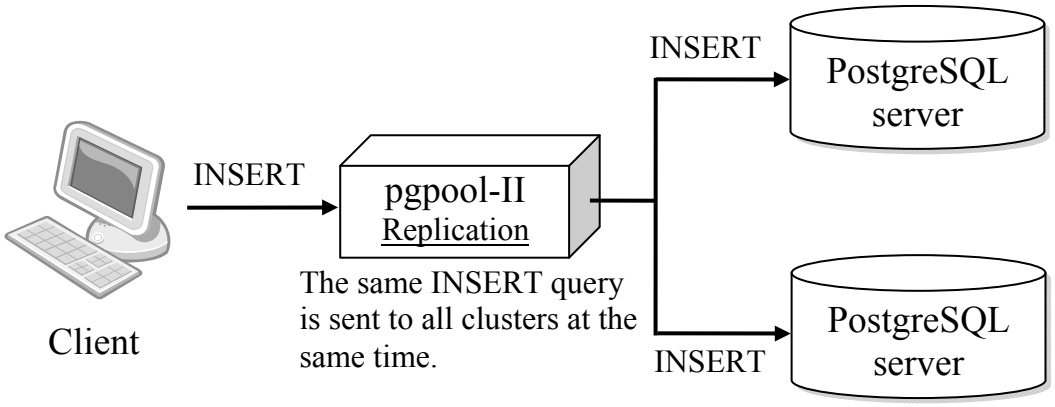

Figure 2. The flow of INSERT query in Replication mode.

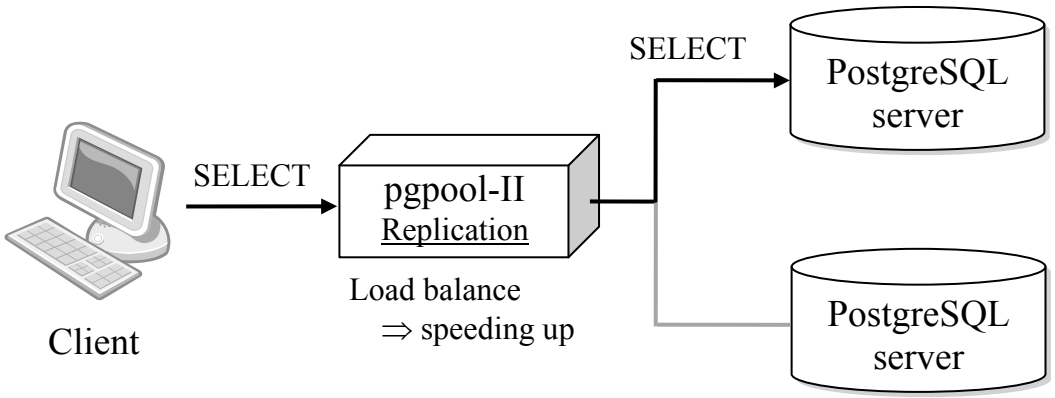

Figure 3. The flow of SELECT query in Replication mode.

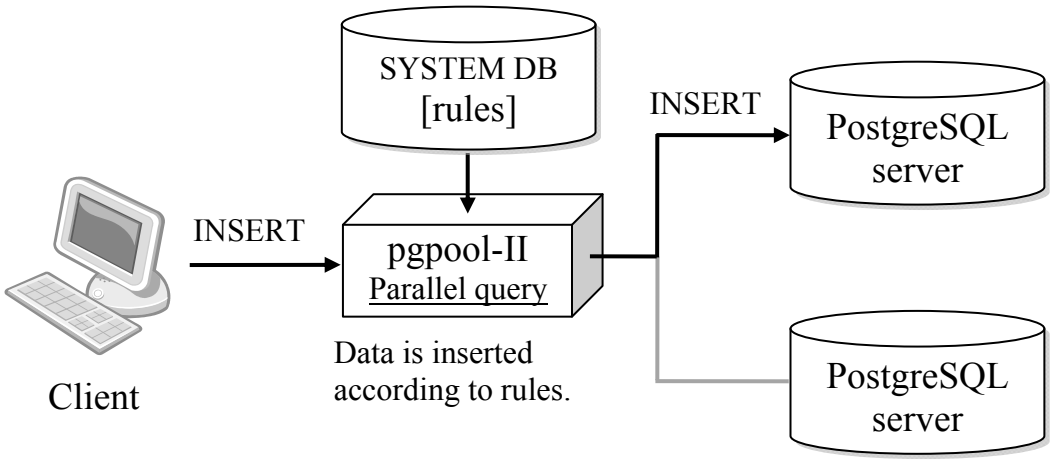

Figure 4. The flow of INSERT query in Parallel query mode.

Parallel query

Data can be split among multiple servers by using the parallel query mode. The System Database is required to use parallel query. It contains rules to send partitioned data to an appropriate cluster. An INSERT query is executed according to the rules (Figure 4). A SELECT query is executed on every cluster, and returns an integrated result to a client (Figure 5). If large-scale data is searched, the overall execution time will be reduced.

PgpoolAdmin is a management tool for pgpool-II. It is possible to monitor, start, stop and change settings of pgpool-II. An image of pgpoolAdmin is shown in Figure 6.

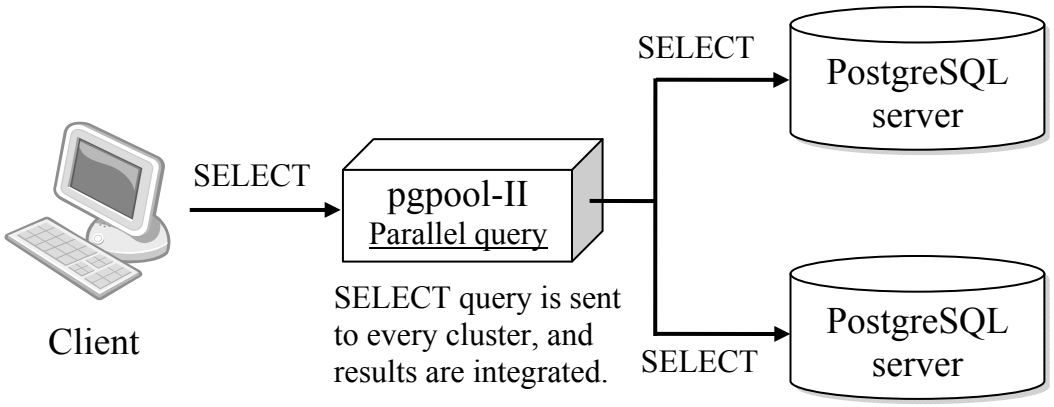

Figure 5. The flow of SELECT query in Parallel query mode.

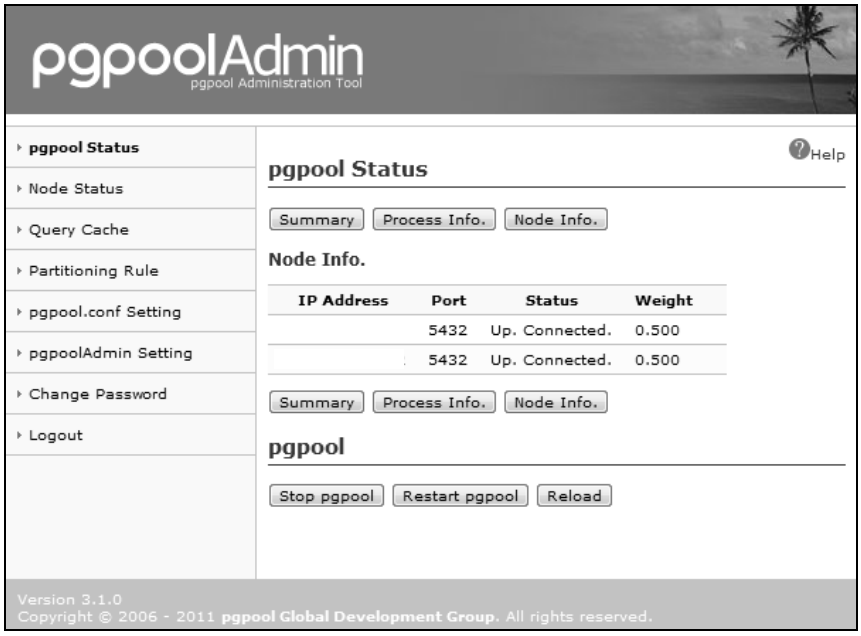

Figure 6. An screen of pgpoolAdmin .

Table 1. Software configuration.

Software		Download Site
Relational database management system	PostgreSQL	http://www.postgresql.org/
Distributed database management tool	pgpool-II	http://pgfoundry.org/projects/pgpool/
Mapping engine	MapServer	http://www.mapserver.org/
Web-GIS client	OpenLayers	http://openlayers.org/

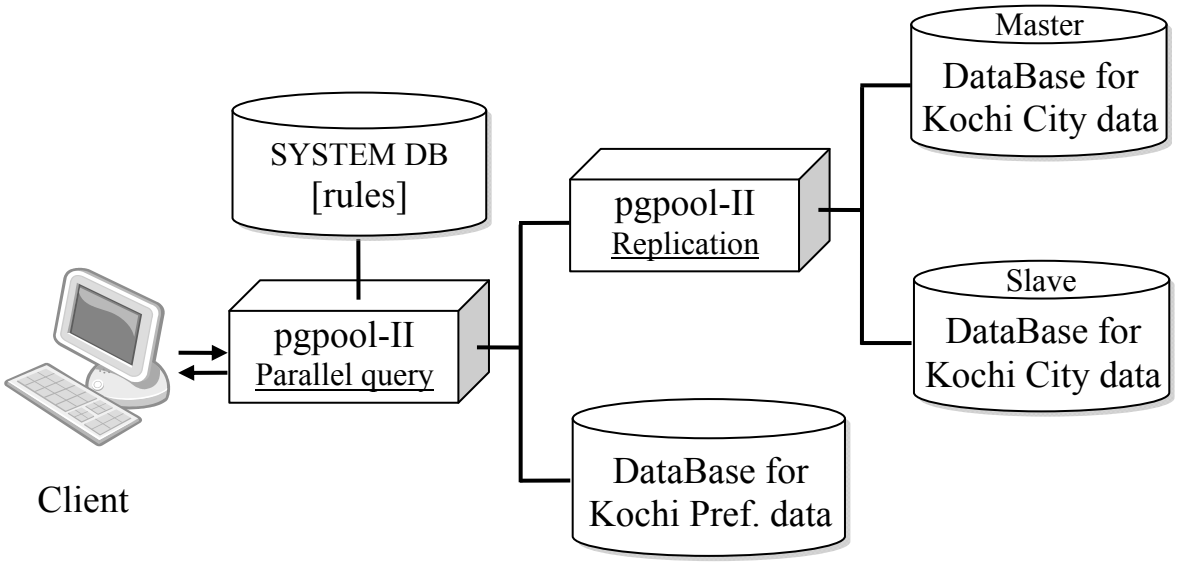

Figure 7. An outline of prototype system.

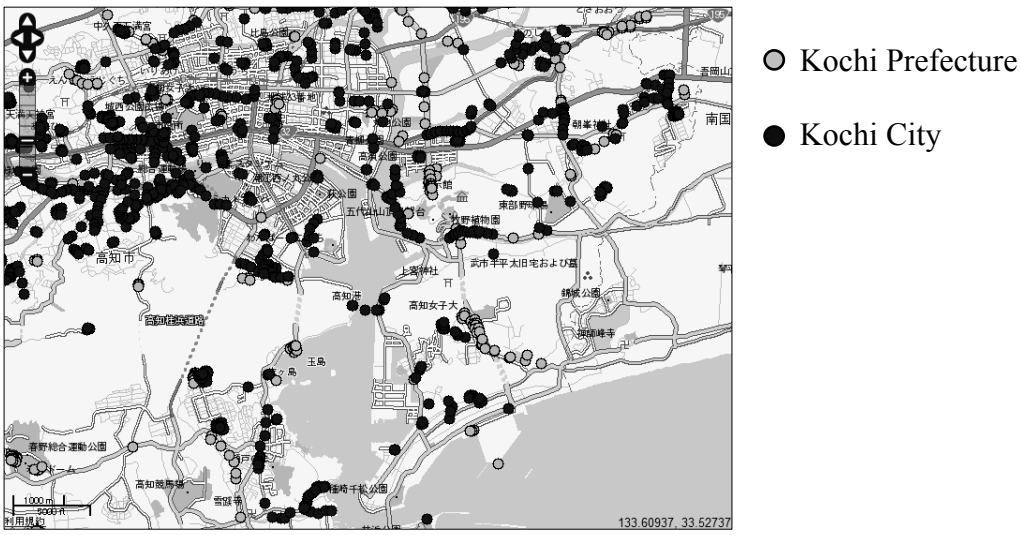

Figure 8. An image of the map which plotted the position of borehole data.

4. PROTOTYPE SYSTEM FOR BOREHOLE DATA

The prototype of distributed database system has been developed by extending Web-GIS system for management of geologic data developed by Nemoto *et al*. (2010). Pgpool-II was added to manage distributed database. The software configuration of the prototype is shown in Table 1. All software is FOSS that can be downloaded over the Internet.

400 borehole data owned by Kochi Prefecture and 1000 borehole data owned by Kochi City were used for validation of the system. The borehole data is published through Web-GIS developed in Kochi ubiquitous demonstration project for disaster prevention (Nakada *et al*., 2011; http://www.geonews.jp/kochi/). The format of borehole data is an XML based on the specification of the Ministry of Land, Infrastructure, Transport and Tourism of Japan. The data of file name, survey title, longitude and latitude extracted from XML file was inserted into the distributed database.

An outline of the system is shown in Figure 7. The Kochi Prefecture data and Kochi City data are inserted into deferent database by Parallel query. Moreover, Kochi City data is registered in two databases by Replication. The system continues to deliver borehole data even if an individual database is down. The format of search result by SELECT query is CSV, KML or JSON. An image of the map which plotted the position of borehole data obtained by SELECT query is shown in Figure 8.

5. CONCLUTIONS

The prototype of distributed database system has been developed with an aim of continuing delivery of geologic data even if a large-scale disaster occurs. The prototype for borehole data has been successfully implemented by combination of PostgreSQL and pgpool-II. Further development and improvement of the system are necessary for practical use. The inclusion of PostGIS, which is spatial database extension for PostgreSQL, in this distributed database system will be effective to manage and analyze various spatial data in addition to borehole data.

6. ACKNOWLEDGEMENT

This study was supported by KAKENHI (24700096; Grant-in-Aid for Scientific Research by Japan Society for Promotion of Science).

7. REFERENCES

Nakada, F., 2011. A Proposal on Sharing and Reusing Geotechnical Information Based on "Kochi Ubiquitous Demonstration Projects for Disaster Prevention". *Geoinformatics*, vol.22, 4, 189-200.

Nemoto, T. and Kimura, K., 2010. Development of Web-GIS System for Sharing and Visualizing Geologic Data. *Proc. GIS-IDEAS2010*, 119-124.