Large Scale Distributed Database Systems Based on Semantic P2P Networks

Lican HUANG

Zhejiang Sci-Tech University, CN

There are the trends that the scientific projects are involved with large scale cooperation among lots of researchers who join the virtual organizations with different domains such as biology and computer science. The virtual organizations usually have their own database systems. However, the researchers hope to retrieve relative information among several virtual organizations. To connect all these individual database systems into a virtual database system is a significant issue.

This paper presents VIRGO_DDBMS, a framework of distributed Database system based on virtual hierarchical tree Grid organizations (VIRGO) p2p network. The virtual organizations are classified as hierarchical domains like DNS. The servers hosting database systems of virtual organizations construct n-tuple virtual hierarchical overlay network according to the virtual organizations’ semantic domains. VIRGO_DDBMS is effective for data update and retrieve. We use the policies caching the host machines’ addresses of database systems to avoid the overload of traffic of root nodes in tree structure. The proposal presented here can construct a virtual distributed database system managing huge volumes of data by connecting all database systems which belong to individual virtual organizations. The database language used in the proposal is similar to SQL which is extended as the following:

query::=SELECT{*[exp{[AS]c_alias}
{.exp{[AS]c_alias}[...]}]
FROM table@domainref
[WHERE search_condition];

The only difference between SQL and extended SQL is that the table name is defined by table@domainref, in which domainref is for the hierarchical virtual organizations’ domain names such as software-tools@bioinformatics.science. In this paper, we also describe the protocols of VIRGO_DDBMS and the primary implementation of the prototype.