

Design for the Distributed Data Locator Service for Multi-site Data Repositories

Wednesday, May 15, 2019 9:40 AM (20 minutes)

In modern fusion experiments, the remote data access has already come into wide use in both domestic and international research collaborations. SNET mutual data exchanging platform in Japan interconnects four fusion experimental sites, LHD, QUEST, GAMMA10, and TST-2, over 1 000 km distance and enables the remote collaborators access each site's data seamlessly as if they were in local site. Also, ITER plans to replicate the full dataset to the REC, over 10 000 km distance, which will equip massive computer resources to analyze the ITER physics data off-site. The outcome results would possibly be shared with international collaborators of other sites. In such multi-site data repository environments, the data location informing service will be essential to find the nearest repository from which users can retrieve the data most efficiently. Considering the latency time becomes more than 100 milli-seconds for inter-continental network transactions, it is more preferable to distribute not only the data repositories but also the locator servers to multiple sites in the world. Since the data locations will be served by a relational database (RDB), such as PostgreSQL, real-time data synchronization between multiple locator RDBs will be necessary to provide a consistent service over the world.

PostgreSQL 9.0 and higher has a mechanism to replicate the database from a single master to multiple slaves, which is called as "streaming replication." Since PostgreSQL 9.4, it has been equipped with the multi-master bi-directional replication (BDR) capability. To register the analyzed outcome results at each site, BDR of data indexes will be necessary for the worldwide distributed data indexing services.

From the viewpoint of the data consuming clients, the most important thing is to find out the best data locator and repository site with which the clients can communicate at the highest speed through the network. There are some methods to find the neighbor place on network; typically, the DNS top domain shows the country region but have some exceptions like "iter.org." GeoIP is another network service to know where the site IP address exists geographically, however, it is also reported that the precision is not necessarily so high. It must be also noted that geographical distance does not corresponds to the network neighborhood.

Therefore, this study proposed to use the ICMP echo reply (ping) or TCP SYN+ACK response (tcpping) to measure the network distance to each data replication site and then decide which site the client can reach with the minimum latency. The latency measurement should be once made before requesting the data, and the values can be stored for some while in each client. The combination uses of the BDR of locator RDBs and the site pre-selection by latency measurement have been tentatively implemented and verified. The test proved that the client can automatically choose the nearest data site, and the newly inserted record was replicated to the other locators within a definite delay. We can conclude that this method is effective for the multi-site data environments, such as SNET in Japan and also ITER with the REC.

Author: NAKANISHI, Hideya (National Institute for Fusion Science)

Co-authors: Dr YAMANAKA, Kenjiro (National Institute for Informatics); TOKUNAGA, Shinsuke (National Institutes for Quantum and Radiological Science and Technology (QST)); OZEKI, Takahisa (National Institute for Quantum and Radiological Science and Technology (QST)); ISHII, Yasutomo (National Institute for Quantum and Radiological Science and Technology (QST)); OHTSU, Hideo (National Institutes for Quantum and Radiological Science and Technology (QST)); SUGIE, Yoshihiko (National Institutes for Quantum and Radiological Science and Technology (QST)); NAKAJIMA, Noriyoshi (National Institute for Fusion Science); EMOTO, Masahiko (National Institute for Fusion Science); Dr YAMAMOTO, Takashi (National Institute for Fusion Science); Mr OHSUNA, Masaki (National Institute for Fusion Science); Mr ITO, Tatsuki (National Institute for Fusion Science); Mr IMAZU, Setsuo (National Institute for Fusion Science); Ms NONOMURA, Miki (National Institute for Fusion Science); Mr YOKOTA, Mitsuhiro (National Institute for Fusion Science); Mr OGAWA, Hideki (National Institute for Fusion Science); Mr MAENO, Hiroya (National Institute for Fusion Science); Ms AOYAGI, Miwa (National Institute for Fusion Science); Mr YOSHIDA, Masanobu (National Institute for Fusion Science); Mr INOUE, Tomoyuki (National

Institute for Fusion Science); Mr NAKAMURA, Osamu (National Institute for Fusion Science); Dr ABE, Shunji (National Institute for Informatics); Prof. URUSHIDANI, Shigeo (National Institute for Informatics)

Presenter: NAKANISHI, Hideya (National Institute for Fusion Science)

Session Classification: Plenary Oral

Track Classification: Database Techniques for Information Storage and Retrieval