Semi-Operational Data Reductions for Query Processing in Highly Distributed Data Environments

Semi-Operational Data Reductions for Query Processing in Highly Distributed Data Environments*

Akira Kawaguchi[1], Abbe Mowshowitz[1], and Masahiro Shibata[2]
[1]Department of Computer Science, The City College of New York, U.S.A.
[2]Department of Computer Science and Electronics, Kyushu Institute of Technology, Japan

ABSTRACT

Characteristics of network topology can have significant effects on network performance. This is especially true for a highly distributed data network such as a 5G platform. Distance between nodes, for example, can affect latency, and may also play a role in traffic congestion if large amounts of data must be moved in response to queries. In previous research, as participants in the International Technology Alliance 2006-2016 (sponsored by the U.S. Army Research Lab and the U.K. Ministry of Defence), the authors have demonstrated that traffic congestion in overlay networks can be reduced by engineering the network to conform to a hypercube structure. Research currently underway funded by NSF (*Resilient Edge Cloud Designed Network*) is directed toward implementing the hypercube and other network overlays using software defined networking to support query optimization in distributed database systems, making use of network distance information and data quantity.

This study will demonstrate a novel method for consolidating data in an engineered network for the purpose of optimizing query processing. The method exploits so-called semi-operational data reductions to reduce data transfer, measured as the product of data quantity and network distance. Query processing typically calls for merging data collected from a small subset of server nodes in a network. This poses the problem of managing efficiently the exchange of data between processing nodes to complete some relational data operation. One way to do this is to delegate the processing to a node that is relatively central to the subset so as to minimize data transfer between nodes. Assuming that network distance between nodes can be computed easily (as in the engineered hypercube), it is possible to find a node to serve as the center for any consolidation operation. We will show how the consolidation process can be performed by selecting a subgraph of a complex network designed to simplify the selection of a central node and thus facilitate the computations required. The process is built on a two-phase exchange method, adapted from well-known semi-join operations. Data transfer between this central node and the data server nodes ensures a substantial reduction in the overall data transfer required for query processing.

This approach to query processing can be applied in any setting in which mobile devices connected to an appropriate substrate need to share information.

* This material is based upon work supported by the National Science Foundation under Grant No. 1818884.