An Efficient Stochastic Update Propagation Method in Data Warehousing

Research-in-Progress

Bijoy Bordoloi  
Southern Illinois University Edwardsville  
bbordol@siue.edu

Bhushan Kapoor  
California State University Fullerton  
bkapoor@exchange.fullerton.edu

Tim Jacks  
Southern Illinois University Edwardsville  
tjacks@siue.edu

Abstract

This paper develops an approach to update propagation for data warehousing environments where data storage (and retrieval) is required as a sum of data at individual source nodes in an organization’s distributed information system network. Our procedure is expected to result in less network traffic (as compared with the real-time method) due to update propagation required because of changes in source data, in particular due to Big Data. The method allows system users to place limits on the discrepancy between the source data and the data in the data warehouse which could result due to a time lag between source data changes and the update operation. Finally, the pre-specified limits on the discrepancy are maintained while accounting for two crucial factors in distributed systems: the fact that some nodes are situated on more congested network links and that some of the links on the network are less reliable than others.

Keywords

Data Warehousing, Update Propagation, Data Communications, Simulation