

One Step Further: Pull-Process-Push Data Processing

Authors : Romeo Botes, Imelda Smit

Abstract : In today's modern age of technology vast amounts of data needs to be processed in real-time to keep users satisfied. This data comes from various sources and in many formats, including electronic and mobile devices such as GPRS modems and GPS devices. They make use of different protocols including TCP, UDP, and HTTP/s for data communication to web servers and eventually to users. The data obtained from these devices may provide valuable information to users, but are mostly in an unreadable format which needs to be processed to provide information and business intelligence. This data is not always current, it is mostly historical data. The data is not subject to implementation of consistency and redundancy measures as most other data usually is. Most important to the users is that the data are to be pre-processed in a readable format when it is entered into the database. To accomplish this, programmers build processing programs and scripts to decode and process the information stored in databases. Programmers make use of various techniques in such programs to accomplish this, but sometimes neglect the effect some of these techniques may have on database performance. One of the techniques generally used, is to pull data from the database server, process it and push it back to the database server in one single step. Since the processing of the data usually takes some time, it keeps the database busy and locked for the period of time that the processing takes place. Because of this, it decreases the overall performance of the database server and therefore the system's performance. This paper follows on a paper discussing the performance increase that may be achieved by utilizing array lists along with a pull-process-push data processing technique split in three steps. The purpose of this paper is to expand the number of clients when comparing the two techniques to establish the impact it may have on performance of the CPU storage and processing time.

Keywords : performance measures, algorithm techniques, data processing, push data, process data, array list

Conference Title : ICHPC 2015 : International Conference on High Performance Computing

Conference Location : Madrid, Spain

Conference Dates : March 26-27, 2015