Abstract

Real time OLAP, or RTOLAP, is the capability to quickly retrieve, aggregate, analyze and present multidimensional data for cubes whenever there are changes to the data in the relational data sources, without having to run heavy processing on the cube. A big advantage of real time OLAP is that it calculates all relevant data and provides immediate output. One of the main roles of an RTOLAP system is that data is stored directly in main memory, or in an in-memory database, enabling quicker access to the data. Another factor affecting the speed of calculation is compression data is compressed, in such a way that it can be accessed much faster in its compressed form. Additionally, pre-calculated values are not stored, therefore avoiding “data explosion”. In contrast to queries for online transaction processing (OLTP) system that typically access only a small portion of a database, OLAP queries may need to aggregate large portion of a database which often leads to performance issues. In this paper introduced CR-OLAP, a cloud based Real Time OLAP system based on a new distributed index structure for OLAP, the distributed PDCR tree, that utilizes a cloud infrastructure consisting of (m+1) multicore processors. With increasing database size, CROLAP dynamically increases m to maintain performance. The distributed PDCR tree data
structure supports multiple dimension hierarchies and efficient query processing on the sophisticated dimension hierarchies which are so central to OLAP system. It is particularly efficient for complex OLAP queries that need to aggregate large portions of the data warehouses. The static data cube approach proposed by Gray et. al. and materialize all or a subset of the cuboids of the data cube in order to ensure adequate query performance. Practitioners have called for some time for a real-time OLAP approach where the OLAP system gets updated instantaneously as new data arrives and always provides an up-to-date data warehouse for the decision support process. However, a major problem for real-time OLAP is the significant performance issues with large scale data warehouses. The main aim of our research is to address these problems through the use of efficient parallel computing methods. In this paper proposed a distributed data structure for real time OLAP. To our knowledge, the real-time OLAP system that has been parallelized and optimized for contemporary multi-core architectures allows for multiple insert and multiple query transactions to be executed in parallel and in real-time.

References


- Amazon elastic compute cloud (amazon ec2). http://aws.amazon.com/ec2/.

Index Terms

Computer Science Distributed System
Keywords
RTOLAP  CROLAP  OLTP  MOLAP  PDCR  performance latency