Johny Chen Jy - 2020 Poster Contest Resources

Johny Chen Jy is a student at Federal University of Santa Catarina, Florianopolis-SC in Brazil.

Johny's university course covers a broad range of subjects including Object Oriented Programming, Data Structures, Database (Relational and up to newSQL), as well as some marketing and business modules.

He is developing his undergraduate thesis with the supervision of Artur Baruchi and Hugo Watanuki from the LexisNexis Risk Solutions Group Brazil office.

Poster Abstract

In the big data processing and querying world, in order to provision more accurate and up to date information to end users, a relatively common technological approach is to combine an Online Analytical Processing (OLAP) solution with an Online Transactional Processing (OLTP) solution. Whereas the OLAP component is responsible for performing most of the big data processing based on read-only data extracted from different sources, the OLTP component can be used to provide write access to the result data and complement the query results with real time data.

Similarly, in big data parallel processing and querying environments supported by the HPCC (High Performance Computing Cluster) Systems platform, ROXIE query results based on data that was first extracted, transformed, and loaded by the Thor cluster, can be complemented with additional data coming from an external online database. This external component, frequently referred to as Deltabase, corresponds to a OLTP database that can be used to provide real time data and complement query results with data that eventually has not yet been processed by Thor. Despite the obvious benefit of providing more accurate and up to date information to end users; the Deltabase, in case of its failure or because of its own OLTP nature not optimized for data reading, can become a bottleneck in terms of performance and availability of the entire querying system. In such contexts, the utilization of a caching solution can become attractive.

Recently, NoSQL databases have been leveraged as a caching mechanism in hybrid OLAP/OLTP solutions by avoiding that queries recently executed are once again processed by the OLTP system, which are usually slower for providing read access to stored data in comparison to NoSQL databases. By providing an additional optimized component for real time data access and an additional layer of resilience against failures, the inclusion of a NoSQL database as a caching solution can potentially increase the performance and availability of hybrid big data processing and querying solutions.

The overall objective of this in progress study is to explore the usage of a NoSQL database as a potential caching solution to the Deltabase component of the HPCC Systems platform. To this end, an experimental approach will be leveraged. The alternative database architectures and caching algorithms will be evaluated and compared, both from a ROXIE query response time and from an overall system availability.

Presentation

In this Video Recording, Johny provides a tour and explanation of his poster content.

Poster Title: Deltabase caching solutions: an exploratory analysis for increasing ROXIE queries performance

Click on the poster for a larger image.

Deltabase caching solutions: an exploratory analysis for increasing ROXIE queries' performance

Johny Chen Jy

Problem:
OLAP + OLTP solutions provide more accurate and up to date information to end users, but it may run into performance problems

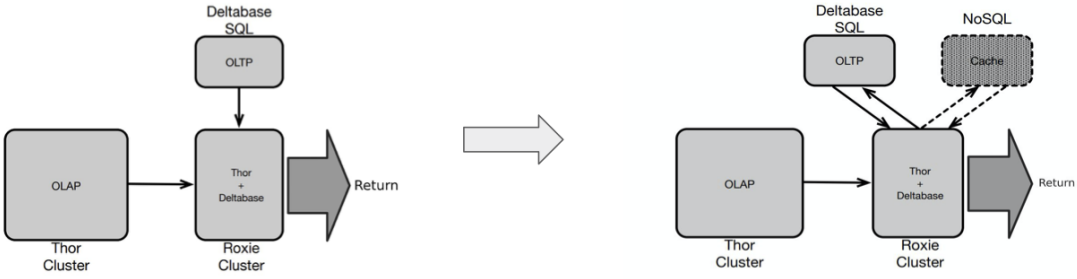
Possible solution:
Add a NoSQL database as a caching solution. It can potentially increase performance and availability

Benefit 1:
Avoids that recently executed are once again processed by the OLTP system

OLTP - Online Transaction Processing databases may fail or become a bottleneck because it's not optimized for data reading

Benefit 2:
Faster read access to data when compared to OLTP databases

Benefit 3:
Optimized component for real time data access and additional layer of resilience against failures



UNIVERSIDADE FEDERAL DE SANTA CATARINA