Optimisation of a Linear Algebra Approach to OLAP

Filipe Oliveira - A57816
a57816@alunos.uminho.pt

Sérgio Caldas - A57779
a57779@alunos.uminho.pt

July 6, 2016

Abstract

Online Analytical processing (OLAP) systems, perform multidimensional analysis of business data and provides the capability for complex calculations, trend analysis, and sophisticated data modeling. All the referred analysis depends on Relational Algebra, which lack algebraic properties, and qualitative and quantitative proofs for all the relational operator. The proposed solution focus on a typed linear algebra approach, encoding OLAP functionality solely in terms of Linear Algebra operations (matrices).

It has been argued that linear algebra (LA) is better suited than standard relational algebra for formalizing and implementing queries in on-line multidimensional data analysis [2] [1]. This can be achieved over a small LA sparse matrix kernel which, further to multiplication and transposition, offers the Kronecker, Khatri-Rao and Hadamard products.

I. Introduction

In a initial fase we implemented from the start a typed linear algebra solution, and evaluate its performance in a real word scenario, using different datasets. The datasets was produced with TPC-H Benchmark, was produced multiples datasets with different data sizes (1, 2, 4, 8, 16, 32 Gb) in order to obtain a realistic results.

II. Hardware Characterization

The platform used by us for our study at Search6 is a dual-socket system equipped with two Intel® Ivy Bridge processors. The system, referenced as compute node 652-1, has two Intel® Xeon® E5-2670v2 (Ivy Bridge architecture) and features 64 GB of DDR3 RAM, supported at a frequency of XXXX MHz divided in 4 memory channels.

System	compute-652-1
# CPUs	2
CPU	Intel® Xeon® E5-2670v2
Architecture	Ivy Bridge
# Cores per CPU	10
# Threads per CPU	20
Clock Freq.	2.5 GHz
L1 Cache	320KB 32KB per core
L2 Cache	2560KB 256KB per core
L3 Cache	25600KB shared
Inst. Set Ext.	SSE4.2 & AVX
#Memory Channels	4
Vendors Announced Peak Memory BW	59.7 GB/s
Measured[1] Peak Memory BW	58.5GB/s

Table 1: *Architectural characteristics of the two evaluation platforms.*

References

[1] Rogério António da Costa Pontes. Benchmarking a linear algebra approach to olap master thesis. 2015.

[2] Hugo Daniel Macedo and José Nuno Oliveira. A linear algebra approach to olap. *Formal Aspects of Computing*, 27(2):283–307, 2015.