

Optimisation of a Linear Algebra Approach to OLAP

FILIFE OLIVEIRA - A57816
a57816@alunos.uminho.pt

SÉRGIO CALDAS - A57779
a57779@alunos.uminho.pt

July 17, 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

The design and development of systems that generate, collect, store, process, analyze, and query large sets of data is filled with significant challenges both hardware and software. Combined, these challenges represent a difficult landscape for software engineers.

The relation database is the current solution for big data storage. Prior efforts have been made [2] [1] in the research project "Linear Algebra approach to OLAP", in order to fully represent relational algebra in terms of linear algebra operators.

Regarding the challenge of High Performance Computing, we implemented from the start a typed linear algebra solution given special importance to the data modelling which, if done poorly, limits the attainable efficiency in data-intensive systems like OLAP.

With respect to performance evaluation and results validation in a real work scenario, the used datasets were produced with TPC-H Benchmark, in which is workload consists of multiple query runs. In order to obtain realistic and meaningful results large datasets were considered, ranging from 1 to 64GB. In order to infer conclusions and compare relational and linear algebra the object-relational database management system PostgreSQL version 9.6, with roots in open source community, was chosen in order to represent the relational algebra approach.

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 evaluation platform.

III. STAREGY

I. Towards a linear algebra semantics for SQL

Aggregation Performance. Aggregations occur in all TPC-H queries, hence performance of group-by and aggregation is quite important.

Database, application, and storage servers ship with a large number of configuration parameters like buffer cache sizes,

number of I/O daemons, and parameters input to the database query optimizer's cost model. Finding good settings for these parameters is a challenging task because of the complex ways in which parameter settings can affect performance

IV. 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 ² Peak Memory BW	58.5GB/s

Table 2: Architectural characteristics of the two evaluation platforms.

V. 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 ³ Peak Memory BW	58.5GB/s

Table 3: Architectural characteristics of the two evaluation platforms.

VI. 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 ⁴ Peak Memory BW	58.5GB/s

Table 4: Architectural characteristics of the two evaluation platforms.

VII. 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 ⁵ Peak Memory BW	58.5GB/s

Table 5: Architectural characteristics of the two evaluation platforms.

VIII. 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 ⁶ Peak Memory BW	58.5GB/s

Table 6: 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.