NullDB: Instantaneously Answering Any OLAP Query

Andrew Crotty
Brown University
crottyan@cs.brown.edu

ABSTRACT

This paper describes NullDB, a new system that can instantaneously answer any OLAP query. With this important problem solved, the data management community is now free to work on problems other than improving TPC-H benchmarking results. You are welcome.

1. INTRODUCTION

Despite the growing prevalence of machine learning (ML), online analytical processing (OLAP) is and will continue to be an essential part of the data-driven decision-making process for the foreseeable future [7]. In fact, this assertion was even anecdotally confirmed by a mid-level employee of a major tech company during a forced conversation in line during the lunch rush at an In-N-Out in Santa Clara.

At the same time, big data is growing faster than ever. According to a recent Gartner report [1], for example, the combined amount of data in the universe might even exceed a trillion yottabytes in size at some nonspecific point in the future. Moreover, according to another recent report [3], our capacity to store and analyze data is barely improving, due equally in part to the decline of Moore's law and increase in frequency of chemtrails.

We therefore present NullDB, a new database management system (DBMS) capable of instantaneously answering any OLAP query. Users pose queries to NullDB using the well-known Structured Query Language (SQL), and query results are returned subject to the Atomicity, Consistency, Isolation, and Durability (ACID) gaurantees. To demonstrate the effectiveness of NullDB, we provide a detailed and thorough experimental analysis using one TPC-H query.

2. NULLDB

As mentioned, NullDB is an OLAP-optimized DBMS that can instantaneously answer SQL queries with full ACID guarantees. Although this claim might seem too good to be true at first glance, NullDB actually has a rigorous theoretical foundation, although a full formal proof could not be included due to space constraints.

The key innovation of NullDB is the novel timestamp-based query scheduler, which leverages the happened-before relation [5] to instantaneously return the null set as the answer to any query. More formally, a DBMS is a thing that you put stuff in, and, before you put any stuff in it, it is empty. Conceptually, all queries in NullDB are executed at this initial empty state.

3. EVALUATION

This section presents our extensive evaluation of NullDB. We compare against outdated versions of at least one state-of-the-art DBMS on a machine for which we do not provide the specifications to better impede reproducibility. Due to space constraints, we were unable to include the results, but we can assure the reader that, as is not shown in Figure 1, NullDB achieves nearly infinite speedup.

The question naturally arises: how did we select which TPC-H query to (not) include in the paper? As is common practice, we ran experiments for all 22 TPC-H queries and then included only the results for those queries on which our techniques exhibited the greatest speedup.

4. CONCLUSION & FUTURE WORK

This paper presented NullDB, which solves OLAP. Now, we as a community can move on to other problems that are almost as important as improving TPC-H benchmarking results. We present three such problems.

Paint-Drying Problem: The current best known algorithm for watching paint dry is O(N); that is, for a paint-drying problem (PDP) that takes N time-steps, the watcher must wait for the full N time-steps before completing the watching task. It remains unclear whether a sublinear PDP solution exists, but this important problem deserves significantly more attention.

Automatic Acronym Verification: Crafting a clever acronym for a research paper is known to be an NP-hard problem [4], but the related problem of automatic acronym verification (AAV) has received far less attention. The goal of AAV is to automatically verify the safety property of a research paper acronym, which is easy for humans but difficult for machines. For example, the *Fair and Reciprocal Tarriff Act* was recently lampooned by the fake news for accidentally forming the acronym FART. Another recent example was the controversy surrounding the VLDB 2019 submission *[REDACTED]: A New System for Human-in-the-loop Entity Resolution*, which resulted in a great deal of embarrassment for the German-born authors. Even a partial solution to the AAV problem would be a tremendous achievement.

DEEPavlo: Noted databaseologist Andy Pavlo has contributed a great deal to the community over the years. A detailed breakdown of how Andy spends his time reveals that his greatest bottleneck is the act of composing and not proofreading bizarre Tweets [6]. We believe that the application of deep learning techniques, similar to the DeepDrumpf project [2], can effectively automate Andy's Twitter feed. In fact, if we as a community could solve this problem, Andy would personally have one to two orders of magnitude more free time to pursue his other great passion: turntablism.

5. REFERENCES

[1] Gartner. A Gartner Report that No One Will Verify, 2017.
[2] B. Hayes. https://twitter.com/deepdrumpf, 2018.
[3] D. J. Khaled. Another One. https://www.youtube.com/watch?v=FSswEzWJdWo, 2015.
[4] D. E. Knuth. Postscript About NP-hard Problems. *SIGACT News*, 6(2):15–16, Apr. 1974.
[5] L. Lamport. Time, Clocks, and the Ordering of Events in a Distributed System. *Commun. ACM*, 21(7):558–565, 1978.
[6] A. Pavlo. https://twitter.com/andy_pavlo, 2018.
[7] L. Wrinkles. On the Importance of OLAP. *SIGMOD Record*, 43(3):55–60, 2014.