HybridPerfopticon: Query Visualization for Hybrid Distributed Databases

Brandon Haynes **Shrainik Jain**
University of Washington
{bhaynes, shrainik}@cs.washington.edu

ABSTRACT
Hybrid databases have recently gained a lot of attention because of the promise they show in several prominent use cases in scientific data management. A hybrid distributed database system usually consists of multiple underlying databases with different data models. In order to better understand such systems we present a query profiling and visualization tool for queries authored over these hybrid systems. Our approach is to extend an existing visualization tool for distributed databases by adding support for multiple underlying database systems.

Author Keywords
Hybrid Databases, Query Visualization

ACM Classification Keywords
H.5.2 Graphical user interfaces (GUI): Miscellaneous

INTRODUCTION
Hybrid databases are an emerging area in database research. A hybrid database comprises of dissimilar systems fused together in various ways. Stonebraker *et al.* . [9] recently observed that the "one size fits all" era is over, and that the demands of big data management warrant a more federated approach to database design. Lim *et al.* in Cyclops [7] offered a top-down approach that combined compatible systems into a monolithic whole. Polybase [5], on the other hand, took a bottom-up approach where disparate systems were more loosely connected. To aid understanding, development and debugging of these new systems, we need better visualization tools as well.

Many profiling systems exists that target queries in a single database system. For example, SQL Server, Postgres, MySQL all have a basic "'explain'" functionality, which displays a query plan in a human-readable (but often unfriendly) format. Most 'explain'-type tools lack deeper analysis features such as the ability to view performance over the duration of the query. Further compounding this problem is the limited availability of tools available in a distributed database context, where operator-worker assignments, data skew, and similar problems are much more important.

Some tools, however, have been recently introduced that attempt to fill this need in a distributed database environment. Gprof [2], is a basic visualization tool for system development; however, it tends to be oriented toward developers. Dapper [1] is Googles production distributed systems tracing tool designed to aid in distributed system development, but it lacks support for distributed query visualization. Finally, Ambrose [4] is a framework designed to visualize map-reduce jobs. While it offers many of the features that one would like to see in a robust distribute database visualization tool it currently targets only the map-reduce space.

All of these tools lack a complete visualization story (*i.e.* where the tool is useful across both users and system developers) which has resulted in a scarcity of debugging and visualization in the hybrid distributed database space. Perfopticon [8] seems to be the most advanced system that addresses the profiling requirements for a distributed database system and is targeted to both audiences.

This paper presents an extension to Perfopticon to support these emerging use cases, where users require robust profiling of plans that co-occur across multiple database systems. These present a number of unique challenges, such as how to display non-relational features (e.g., array chunks) in a framework that has heretofore targeted only tuple-based relational systems. Additionally, a number of corollary challenges arise including coordinating the display of multiple sub-plans, visualizing data transfer between systems, and gathering profiling data from multiple sources.

For the purpose of demonstration, we chose a hybrid distributed database with SciDB [3] and Myria [6] as its component systems. SciDB's data-model is completely different from Myria, making it critical to generate semantics for communication between these systems. HybridPerfopticon, our extension to Perfopticon, highlights this communication by helping visualize the overall hybrid query plan, tuple flow among operators, skew in distribution of workloads among system and internal query time statistics while being completely compatible with the existing design.

In the next section, we present the architecture of the system, which details the steps necessary to orchestration hybrid plan execution, creation, and serving logs(for visualization purposes). This is followed by some results from the

use case analysis so far. We conclude by looking at some future extensions to HybridPerfopticon.

HYBRIDPERFOPTICON ARCHITECTURE

Our approach to extending Perfopticon to support hybrid queries involved three major additions:

- Generating HybridPerfopticon-compatible SciDB plans
- Creating HybridPerfopticon by extending Perfopticon to recognize components from varying underlying DBMSs
- Creating hybrid plans and exposing them to HybridPerfopticon

In this section we describe each of these changes in detail and highlight the trade-offs in our design decisions.

Generating SciDB Plans & Metrics

In the Myria DBMS, each operator is instrumented to store profiling data. When Perfopticon needs this profiling data, a HTTP request is generated, a query is performed over these data, and the aggregated results are returned.

While we could take a similar route in SciDB by instrumenting each operator and responding to aggregation requests, this would be a time-consuming task and would decrease the performance of SciDB itself. Instead, we elected an alternative approach: since the relevant profiling metadata is already present in the SciDB logs, we generated our performance metrics by scraping these log files. While this resulted in decreased performance on startup (approximately ten seconds per 100MB of logging data), our approach has the added benefit of requiring no changes to the SciDB source code. We expect that this approach would allow for a low-barrier introduction of similar systems, assuming that the salient metrics exist in the log files.

The architecture of our log-scraping process is illustrated in Figure 2. The entry point to this process is the `SciDBParser` class, which is responsible (via the `SciDBLog` class) for converting raw log entries into sequences of statements. Statements associated with a given query are aggregated into a physical plan, and the operators for that plan are identified and hierarchically assembled within that plan. During the log interpretation process we also identify performance metrics associated with each executed query and associate those results with the plan.

An illustration of the log-scraping process is listed in Figure 1. In this log fragment (abbreviated for readability), the `SciDBLogParser` identifies the beginning of the query execution process, identifies (and interprets) the associated physical plan, associates a scatter/gather operation with an underlying operator, and notes the time at which the query is committed. Each of these entries are converted into an equivalent statement, and the useful metadata for each statement is embedded therein.

Extending Perfopticon to Support Additional DBMSs

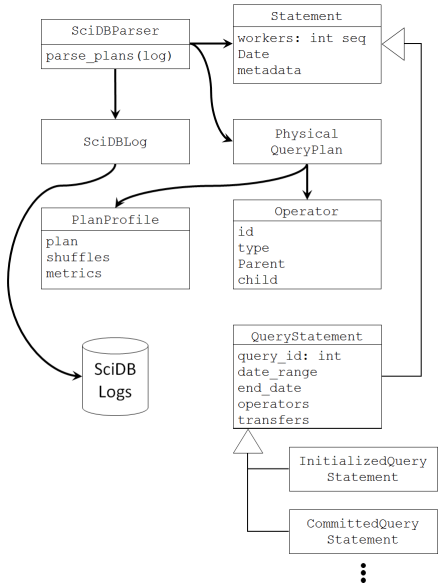

Figure 2. Architecture of the SciDB log parsing components.

In order to extend Perfopticon to support queries across multiple DBMSs, we first explored the use of existing extension points. Since Perfopticon is not designed to be directly extendable (e.g., via a plugin-style architecture), extensive modification to the framework would be required to support the kinds of hybrid plans we intended to visualize. Instead, we elected an alternative approach where a substantial portion of the necessary logic was hidden behind the HTTP server (a "hybrid HTTP server", described in detail below) to which Perfopticon connects for plan and profiling metadata.

While this approach greatly reduced the modifications required to realize our desired functionality, some changes were nonetheless required to Perfopticon itself. For example, it is necessary to be able to identify the system of origin for each plan fragment; this is important not only for visualization purposes (e.g., to color fragments by system type) but also functionally. For example, the hybrid HTTP server must know from which system a fragment originated in order to retrieve profiling metrics.

A number of similar minor modifications were required throughout the Perfopticon framework. By way of example, SciDB fragments required a SciDB-specific query identifier be attached to each fragment; we envision that additional metadata would be required in order to support other systems.

Hybrid HTTP server

Given that HybridPerfopticon continues to expect HTTP responses in the same form as Perfopticon, our hybrid HTTP server needed to do a substantial amount of manipulation to compose hybrid plans without violating this requirement. A high-level architectural overview of our approach is illustrated in Figure 3.

For each incoming request, our server first distinguished between three cases:

SciDB Log Entries

```
2015-05-17 03:13:37,310: Initialized query (123)

2015-05-17 03:13:37,310: Parsing query(123): save_myria(redimension(...))

2015-05-17 03:13:37,365: The physical plan is detected as DML [pPlan]:
  >[pNode] impl_save agg 0 ddl 0 tile 0 children 1
    schema transform_1_1<value:int64 NULL DEFAULT null>
                        [id=0:599,256,0,bin=0:9,*,0]
    output full chunks: yes
    changes dstribution: no
    props sgm 1 sgo 1
    diout hash
    bound start {0, 0} end {599, 9} cells 6000 chunks 3 est_bytes 294006
  ...

2015-05-17 03:13:38,057: requestNextChunk: attId=0, stream=1,
                                           numSent=172, numRecvd=173



2015-05-17 03:13:37,836: Query (123) is being committed
```

HybridPerfopticon Internal Representation

```
InitializedQueryStatement(query_id=123)

ParsedQueryStatement(
        statement="save_myriad(redimension(...)")
PhysicalQueryPlanStatement(
        operators={
                Operator(
                        type="impl_save",
                        array="transform_1_1",
                        ...),
                ...})

ScatterGatherChunkStatement(
        array="transform_1_1",
        sent=172,
        received=173)

CommittedQueryStatement(123)
```

Figure 1. SciDB log fragment (abbreviated for readability) and mapped equivalents in the HybridPerfopticon HTTP server

- **Hybrid requests**. When a request for a hybrid plans was received, copies of the associated Myria and SciDB plans were retrieved (the former via the Myria REST endpoint). These plans were then combined into a Perfopticon-compatible format using the algorithm identified in Algorithm 1. Note that this process (i) attaches a system of origin to each fragment, (ii) concatenates all fragments into a single list, and (iii) ensures that parent/child relationships are established between fragment pairs having different origin systems.

- **Myria requests**. When we receive a request for Myria-oriented, non-hybrid resources (e.g., metadata about a relation that exists on the Myria DBMS). In this case we issued a 301 redirect to the Myria REST server. This enabled the hybrid HTTP server to be the sole communication endpoint for Perfopticon without our needing to multiplex across two URLs on the client side.

- **SciDB requests**. Requests for SciDB profiling data (e.g., the execution time offsets for a SciDB subquery) directly leveraged the profiling logic described above; responses for these requests were in a schema that conformed precisely to their Myria counterparts.

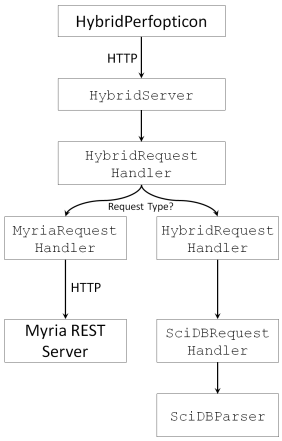

Figure 3. Architecture of the HTTP server that aggregates plans from each database system and returns a (combined) hybrid plan.

RESULTS

We now present an overview of our visual encoding choices and visualization extensions to Perfopticon by presenting a sample use case. The user first inputs the hybrid query as shown in figure 5. This query is first performs a redimension in SciDB and then loads the results in Myria to perform a join operation. Figure 6 shows the hybrid visualization for the same. The encoding choice to distinguish fragment of constituent database system is color *i.e.* we color the fragments from SciDB differently than those from Myria.

Since SciDB is not multithreaded, part of the hybrid query which is run in SciDB is represented as a single fragment. If the original hybrid query had multiple such SciDB parts interleaved, we would have seen multiple SciDB fragments in the query plan. As with perfopticon, the user can select individual fragments and get detailed profiling views as shown in figure 4. The hybrid-server which serves the plans, also serves the profiling API requests and automatically handles differences between SciDB and Myria fragments.

Our next extension to perfopticon is to animate the query plan by showing tuple-flow across different operators at various points in query runtime. This is shown to the user by animating the size of the connecting arrows.

Final extension in Hybridperfopticon is to show inter-database transfer skew. Skew is binned, color-encoded and displayed as a (normalized) mean over the previous two seconds (in order to avoid distracting rapid fluctuations).

We would like to point out that animation for very large

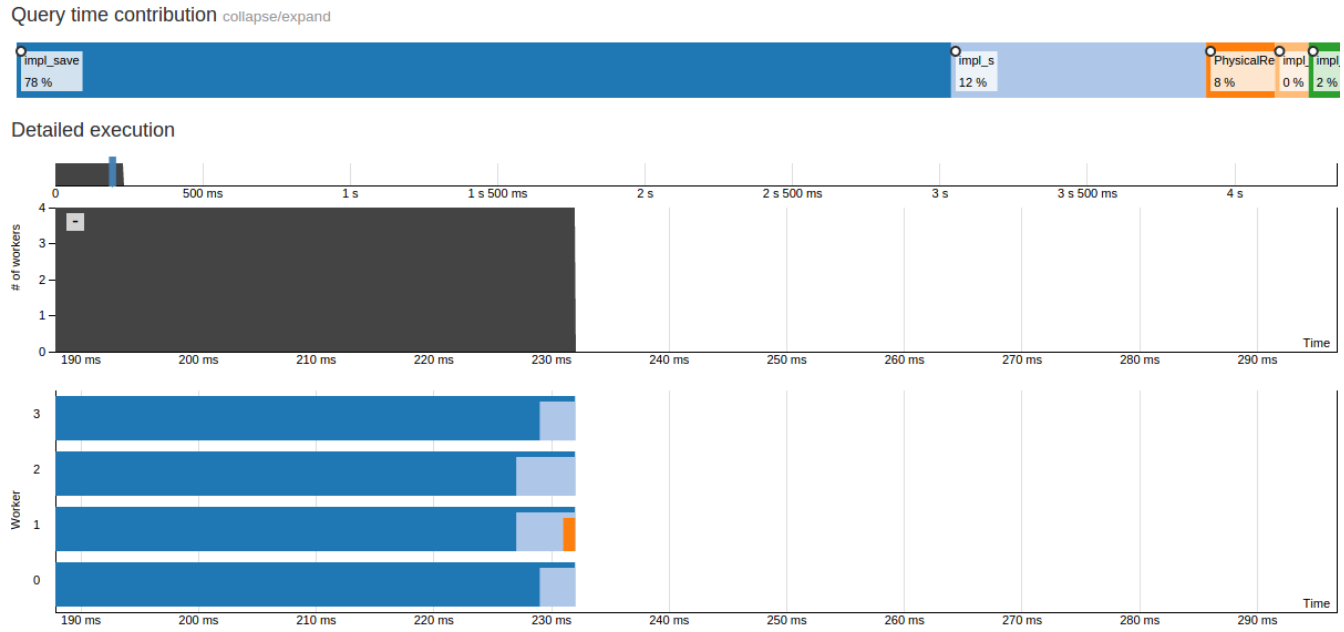

Figure 4. Detailed profiling results for the SciDB fragment in Figure 6.

Algorithm 1 Combining Myria and SciDB plans

function CREATE-HYBRID-PLAN(id, url_{MYRIA})
1: plan = `HTTP-GET`(id, url_{MYRIA})
2: **for each** $\text{fragment}_{\text{MYRIA}} \in$ plan **do**
3: $\text{fragment}_{\text{MYRIA}}$.origin = "Myria"
4: **for each** $\text{op}_{\text{MYRIA}} \in \text{fragment}_{\text{MYRIA}}$ **do**
5: **if** op_{MYRIA}.type is `SciDBScan` **then**
6: $\text{plan}_{\text{SCIDB}}$ = load(op_{MYRIA}.parent.id)
7: **for each** $\text{fragment}_{\text{SCIDB}} \in \text{plan}_{\text{SCIDB}}$ **do**
8: $\text{fragment}_{\text{SCIDB}}$.origin = "SciDB"
9: plan.fragments += $\text{fragment}_{\text{SCIDB}}$
10: **for each** $\text{op}_{\text{SCIDB}} \in \text{fragment}_{\text{SCIDB}}$ **do**
11: **if** op_{SCIDB}.type is `MyriaSave` **then**
12: op_{SCIDB}.child.id = op_{MYRIA}.id
13: return plan

SciDB

```
save(
  redimension(
    load('auto-mpg.csv', -2, 'csv'),
    <mpg: double, type:string>[id=0:*, 1, 0]),
  'intermediate', -1, 'csv+');
```

Myria

```
mpgs = load("intermediate",
            csv(schema(id:int,
                       mpg:float,
                       type:string)));
prices = load("auto-prices.csv",
              csv(schema(type:string,
                         price:int)));
dollars_per_mpg = [from mpgs, prices
                   where mpgs.type = prices.type
                   emit id, type, price / mpg];

store(dollars_per_mpg, result);
```

Figure 5. A hybrid query used as input to HybridPerfopticon. See Figure 6 for generated hybrid plan.

queries lags a bit due to many API calls to the hybrid plan server. We are exploring possible ways to fix this.

CONCLUSIONS & FUTURE WORK

In this work we described HybridPerfopticon, a visualization tool for hybrid queries that cross DBMS boundaries. A tool that does so is critically needed, given that the performance profiles of this class of queries are not currently well-understood. Accordingly, we expect that this tool will serve as a useful means by which the hybrid DBMS space may be explored.

While our approach of using existing log files to generate performance metrics is certainly not novel, the use of log file parsing to quickly add a second DBMS to the visualization framework represents an opportunity to extend the reach of HybridPerfopticon to a wide-ranging set of systems. Future work should evaluate the applicability of this approach, and if generalizable, additional attention should be paid to identifying the optimal interface by which logging metadata may be exposed to HybridPerfopticon.

However, while our log-parsing approach allowed us to quickly expose profiling data for a second DBMS inside HybridPerfopticon, more work remains to be done. In particular, the hybrid HTTP server needs a far more robust caching sub-

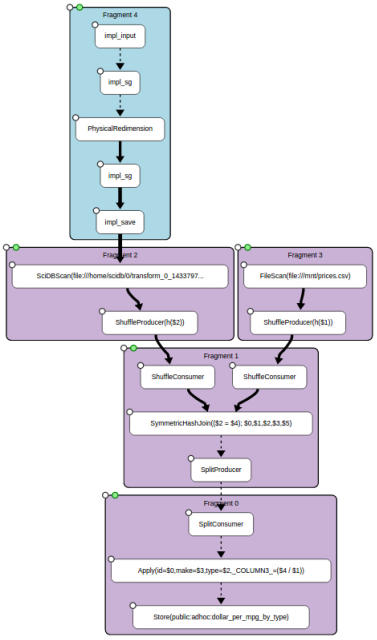

Figure 6. A hybrid plan as visualized by HybridPerfopticon. This hybrid query begins with a SciDB scan/redimension, after which the resulting array is transferred to Myria for joining on a second dataset.

system; in particular, increasingly large logging files will quickly lead to an intractable startup time. This would be easily remedied by ensuring that the results of the parsing operation were themselves logged to a DBMS for quick subsequent retrieval.

Finally, future work should evaluate how to best degrade in the face of differing levels of profiling granularity. For example, we produced SciDB profiling metrics at the millisecond level, whereas Myria profiles at nanosecond granularity. A hybrid visualization framework should be able to gracefully reconcile both; similarly, we expect that some systems will not have support for some profiling areas and that the framework should evolve to handle these cases.

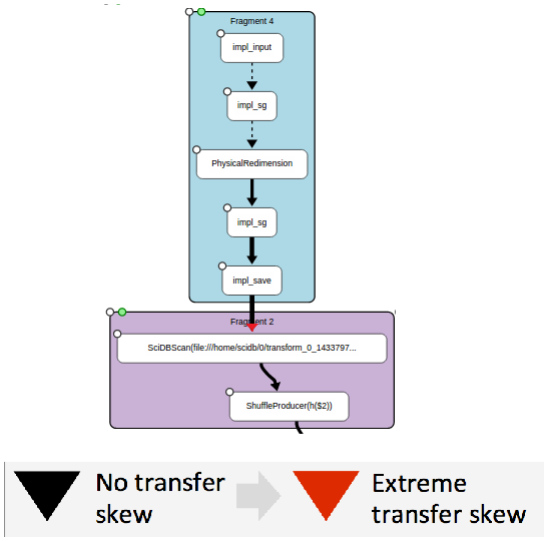

Figure 7. Visualizing skew in inter-database transfers

REFERENCES

1. Google dapper. `http://research.google.com/pubs/pub36356.html`.

2. Gprof. `https://sourceware.org/binutils/docs/gprof/`.

3. Scidb by paradigm4. `http://www.paradigm4.com/`.

4. Twitter ambrose. `https://github.com/twitter/ambrose`.

5. D. J. DeWitt, A. Halverson, R. Nehme, S. Shankar, J. Aguilar-Saborit, A. Avanes, M. Flasza, and J. Gramling. Split query processing in polybase. In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, pages 1255–1266. ACM, 2013.

6. D. Halperin, V. Teixeira de Almeida, L. L. Choo, S. Chu, P. Koutris, D. Moritz, J. Ortiz, V. Ruamviboonsuk, J. Wang, A. Whitaker, et al. Demonstration of the myria big data management service. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, pages 881–884. ACM, 2014.

7. H. Lim, Y. Han, and S. Babu. How to fit when no one size fits. In *CIDR*, volume 4, page 35, 2013.

8. D. Moritz, D. Halperin, B. Howe, and J. Heer. Perfopticon: Visual query analysis for distributed databases. In *Computer Graphics Forum (EuroVis), Cagliari, Italy*, volume 34, 2015.

9. M. Stonebraker, S. Madden, D. J. Abadi, S. Harizopoulos, N. Hachem, and P. Helland. The end of an architectural era:(it's time for a complete rewrite). In *Proceedings of the 33rd international conference on Very large data bases*, pages 1150–1160. VLDB Endowment, 2007.