

Making RDBMSs Efficient on Graph Workloads Through Predefined Joins

Guodong Jin
jinguodong@ruc.edu.cn
Renmin University of China

Semih Salihoglu
semih.salihoglu@uwaterloo.ca
University of Waterloo, Canada

ABSTRACT

Joins in native graph database management systems (GDBMSs) are predefined to the system as edges, which are indexed in adjacency list indices and serve as pointers. This contrasts with and can be more performant than value-based joins in RDBMSs and has lead researchers to investigate ways to integrate predefined joins directly into RDBMSs. Existing approaches adopt a strict separation of graph and relational data and processors, where a graph-specific processor uses left-deep and index nested loop joins for a subset of joins. This may be suboptimal, and may lead to non-sequential scans of data in some queries. We propose a purely relational approach to integrate predefined joins in columnar RDBMSs that uses row IDs (RIDs) of tuples as pointers. Users can predefine equality joins between any two tables, which leads to materializing RIDs in extended tables and optionally in RID indices. Instead of using the RID index to perform the join directly, we use it primarily in hash joins to generate semi-join filters that can be passed to scans using sideways information passing, ensuring sequential scans. In some settings, we also use RID indices to reduce the number of joins in query plans. Our approach does not introduce any graph-specific system components, can execute predefined joins on any join plan, and can improve performance on any workload that contains equality joins that can be predefined. We integrated our approach to DuckDB and call the resulting system *GRainDB*. We demonstrate that *GRainDB* far improves the performance of DuckDB on relational and graph workloads with large many-to-many joins, making it competitive with a state-of-the-art GDBMS, and incurs no major overheads otherwise.

1 INTRODUCTION

Perhaps the two most commonly used data structures to model data in enterprise database applications are tables, which are the core structures of relational database management systems (RDBMSs), and graphs, which are the core structures of several classes of systems, most recently of property graph database management systems (GDBMSs for short), such as Neo4j [6], TigerGraph [7], DGraph [2], and GraphflowDB [22, 28, 34–36]. Aside from developer preference for using a graph-specific data model and query language, GDBMSs target what are colloquially referred to as *graph workloads*, which refer to workloads that contain large many-to-many joins. For example, these workloads appear in social networking applications for finding long paths between two people

over many-to-many friendship relationships or in financial fraud detection applications for finding fraudulent patterns across many-to-many money transfers across bank accounts.

At the same time, several economic and technical factors have lead researchers to investigate techniques to support efficient graph querying natively inside RDBMSs. For example, it is recognized that the data stored in many specialized GDBMSs are extracted from RDBMSs [12, 45, 48, 51]. In many enterprises, users replicate parts of the tabular data stored in RDBMSs to a GDBMS because their applications require the fast join capabilities of GDBMSs. In addition, many applications require other processing on their graph workloads beyond evaluating large many-to-many joins, such as running predicates on node and edge properties or grouping and aggregations, for which RDBMSs already employ efficient techniques. Therefore leveraging mature RDBMS technology to support graph workloads natively is highly appealing to both users and vendors: users can avoid the challenges of duplicating data and keeping multiple systems in sync, while vendors can avoid the efforts to develop a new system from scratch. We revisit this goal and research challenge in the context of columnar RDBMSs, which are similar to GDBMSs in that they also target read-heavy analytical workloads. Our specific goal is to extend a columnar RDBMS natively with the fast join capabilities of GDBMSs.

Several prior approaches leverage RDBMS technology to evaluate graph workloads. One approach simply exposes a separate graph querying layer to users and implements a translation component that outputs SQL versions of queries, with no or minimal modifications to the query processor of the RDBMS. This approach is not focused on performance and is commonly employed in commercial products, such as IBM DB2 Graph [48], SQLGraph [47], SAP Hana’s graph database extension [44].

A second approach introduces a new graph-specific query processor that co-exists with the existing processor of the RDBMS. This has been most recently adopted by the GR-Fusion system [23, 24]. Specifically, SQL is extended to contain graph-specific constructs, using which users create graphs. The topologies of these views, i.e., the vertices and edges without properties, are stored in native adjacency list indexes, which are used during query processing for graph traversals/many-to-many joins, using new graph-specific operators, such as EdgeScan and PathScan. Parts of queries that refer to graph-specific constructs compile to these specialized graph operators, while the non-graph parts of queries compile to existing operators of the RDBMS. GQ-Fast [32] is another system that develops a separate query processor and storage sub-system specialized for graphs. GQ-Fast is not integrated into an RDBMS but the authors’ envisioned integration [32] is similar to GR-Fusion’s dual processor approach. Aside from being heavy-weight integration approaches that develop separate graph-specific components within

This work is licensed under the Creative Commons BY-NC-ND 4.0 International License. Visit <https://creativecommons.org/licenses/by-nc-nd/4.0/> to view a copy of this license. For any use beyond those covered by this license, obtain permission by emailing info@vldb.org. Copyright is held by the owner/author(s). Publication rights licensed to the VLDB Endowment.

Proceedings of the VLDB Endowment, Vol. 14, No. 1 ISSN 2150-8097.
doi:XX.XX/XXX.XX

an RDBMS, the strict separation of graph and non-graph data and operators can lead to inefficient accesses when a query accesses graph data and properties or fail to apply efficient optimizations to the entire query. For example, graph traversals in these systems, which are implemented in specialized traversal operators, are effectively left deep join plans that use index nested loop joins, so avoid using efficient bushy plans. For many queries, these approaches can be amenable to significant performance improvements.

To motivate the key approaches of our solution, we begin by analyzing the primary differences between the join evaluation techniques in RDBMSs and GDBMSs. This question has been discussed since the birth of data management between proponents of DBMSs that adopted graph-based models, adopting Charles Bachmann’s IDS system [14], and those that adopted Ted Codd’s relational model [17], such as System R [13]. Perhaps Codd himself has best articulated the primary differences in his Turing Award lecture [18]. As a primary difference, Codd notes, joins in GDBMSs happen along *predefined* access paths, i.e., between existing records through predeclared *pointers* (or links). In contrast, joins in RDBMSs are *value-based*, so arbitrary tables can be joined on arbitrary columns as long as those columns have the same data types. Although much has changed since IDS and System R, this characterization is still accurate for contemporary GDBMSs and RDBMSs. Contemporary GDBMSs are indeed optimized to perform joins between node records along predefined edges and use two common techniques to perform these joins efficiently: (i) dense integer ID-based joins, which serve as pointers to directly look up records; and (ii) an adjacency list index that is used to quickly find joining edge and node records with a given node record during join evaluation.

Motivated by these observations, our approach integrates predefined joins into a columnar RDBMS by extending two components of the system: (i) the physical storage and query processor; and (ii) the indexing sub-system, where each integration progressively yields more performance benefits. Users perform two actions, the second of which is optional, to benefit from predefined joins: (i) predefinition of a primary-foreign key join to the system; and (ii) an index creation on these tables:

- **Physical Storage and Query Processor:** When a user predefines a primary-foreign key join from table F to table P , where a column of F has a foreign key to a column of P , this performs an ALTER TABLE command that inserts an additional RID_p column to F that contains for each row r_f in F the row ID (RID) of row r_p in P that r_f points to. RIDs are dense integer-based system-level IDs in columnar RDBMS that are used to identify the physical locations of the column values of each row. They are therefore system-level pointers, similar to node IDs in GDBMSs.

In order to use these pointers to perform the primary-foreign key joins more efficiently, we rewrite queries to replace primary-foreign key equalities with RID equalities. Equality predicates in many columnar RDBMSs are primarily evaluated with hash-joins. To exploit the pointer-nature of predefined joins, we employ *sideways information passing* (sip) to speed up scans and indirectly other joins in query plans. Specifically, the hash join operator keeps the RIDs from the build side in compact bitmaps and passes them to the relevant scans on the probe side to perform semi-joins. Because joins in RDBMSs are value-based, existing applications of sip pass

information in probabilistic filters, often a bloom filter [27, 38, 40]. This requires running hash functions both when creating the filter in the joins as well as performing the semi-joins in scans. Since RIDs are dense integer-based IDs, we directly pass a compact bitmap filter and avoid any hash computations.

- **Indexing Sub-system:** A common way to represent many-to-many relationships between two sets of entities in relational databases is to have a table F that contains two foreign keys on two other (not necessarily different) tables P_1 and P_2 . For simplicity of terminology, we refer to such F as a *relationship* table and P_i as *entity* tables. If the joins with both entity tables have been predefined to the system, users can additionally build an index on table F on the two extended RID columns RID_{p1} and RID_{p2} . This index is stored in adjacency list format and serves two purposes. First, it is used to generate further information to pass when a query joins P_1 , F , and P_2 and when a hash join operator builds a table of P_1 or P_2 . Second, when a query refers to F only to facilitate the join of tuples in P_1 and P_2 , so contains no predicates on F and projects out F ’s columns, the index allows us to reduce the number of joins in query plans.

We integrated our techniques into DuckDB [41, 42], a new columnar RDBMS that is actively being developed at Centrum Wiskunde & Informatica [1], and call the extended system *GRainDB*. Unlike systems such as SQLGraph and IBM DB2 Graph, we modify the internals of the RDBMS to improve the performance on many-to-many joins. Unlike GR-Fusion and the envisioned GQ-Fast integration, our approach is purely relational and does not require a separate graph-specific query processing codeline. As a result: (i) our approach directly leverages DuckDB’s core components: the optimizer to generate efficient plans for the entire query, vector-based query processor, and bushy join plans; and (ii) any database in the RDBMS can predefine a set of joins and build a RID index to improve performance. We demonstrate that GRainDB improves the median query execution time of DuckDB by 3.6x on the relational JOB benchmark which contains many-to-many joins, and by 22.5x on the LDBC SNB graph benchmark, making a columnar RDBMS competitive with the state-of-the-art GraphflowDB GDBMS [22]. Because our approach is purely relational, GRainDB improves DuckDB even on some traditional relational analytics queries from TPC-H. In our detailed analysis, we show: (i) that our possibly bushy and sip- and hash-join-based plans can be more efficient than left-deep index nested loop join plans on many queries, such as those with selective predicates on tables that represent edges/relationships; and (ii) our use of sip makes the optimizer of a system more robust because its semi-join computations can mitigate a poor join order selection of the optimizer. Our code, queries, and data are available here [3].

2 RELATED WORK

There are many native GDBMSs [2, 5–7, 10, 36] that employ many read-optimized techniques, such as specialized indices [34], factorization [10], or worst-case optimal join algorithms [20, 35], to be very efficient on analytical queries that contain large joins over many-to-many relationships between entities. However, two of the core techniques that appear in every GDBMS we are aware of are native graph storage in adjacency list indices and predefined pointer-based joins, where node IDs serve as pointers, i.e., positional offsets, into these indices. Integration of these two core techniques

into RDBMSs is the focus of this paper. Below, we review prior work that leverage RDBMSs for supporting graph applications and the literature on sip and join indexes.

GR-Fusion [23, 24] is designed to perform graph querying natively inside an RDBMS. Users define graphs as views over tables. The topology of graph views are stored natively in an adjacency list index. In contrast, the node and edge properties are stored as pointers to the underlying tables. Users refer to the paths in a graph view as if they are a separate table using a new `Paths` construct in the `FROM` clause of SQL. Then, part of the query that enumerates paths and their constraints are evaluated with special operators, such as `VertexScan` or `PathScan`, whose results are tuples that can be input to further relational operators. Therefore this approach creates dual query processing pipelines inside the system. One advantage of this approach is that the original relational operators remain unchanged because outputs from the graph pipeline are regular tuples. However, GR-Fusion also has several shortcomings. First, `PathScan` enumerates only paths, so some other patterns, such as stars, need to be evaluated by the vanilla relational query processor, so do not benefit from the native graph storage or fast join algorithms unless users manually decompose these queries into paths. Instead, our approach is purely relational and can improve equality joins on arbitrary queries, including queries from traditional benchmarks. Second, paths are enumerated through DFS or BFS algorithms, which are akin to left-deep plans that use index nested loop join operators. These plans can be suboptimal compared to bushy join plans, which can be generated in GRainDB. This is further exacerbated if vertex and edge properties need to be scanned during DFS of BFS by following pointers to the tables, which can lead to many random accesses. In contrast, GRainDB uses adjacency list indices to generate information to pass to scan operators (for semijoins) and in some cases to reduce the number of join operators (See Section 5.2) but performs scans always sequentially. We intended to compare our solution against GR-Fusion but the publicly available code has several errors and is not maintained.

GQ-Fast [32] supports a restricted subset of SQL called “relationship queries” which contain joins of tables that are similar to path queries, followed with aggregations. Similar to GRainDB and GR-Fusion, GQ-Fast stores relationship tables in CSR-like indices. Unlike GRainDB and GR-Fusion, these indices also contain properties, i.e., non ID columns, of relationship tables and employ heavy-weight compression schemes. In addition, the system has a fully pipelined query processor that uses query compilation, which gives it performance advantages. However, similar to GR-Fusion, the joins are limited to paths and evaluated with left-deep index nested loop join operators that are equivalent to DFS traversals. In addition, unlike GRainDB and GR-Fusion, GQ-Fast is implemented as a standalone system from scratch and does not integrate these techniques into an underlying RDBMS to support more general queries, which is left as future work [32]. However, even this envisioned integration is similar to GR-Fusion, where the GQ-Fast layer is a separate query processor whose outputs are given to the query processor of the RDBMS. We intended to but could not compare against GQ-Fast because the system supports a very limited set of queries (e.g., none of the LDBC queries are supported) and the publicly available version is no longer maintained and has errors.

Another popular approach is to develop a translation layer between a graph data model and query language to the relational model and SQL and leverage the underlying RDBMS without any modifications. Systems such as IBM DB2 Graph [48], SQLGraph [47], and SAP Hana’s graph database extension [44] primarily provide a translation layer between the property graph data model and a query language or API, such as Gremlin [43], and convert the modeled graph into relational tables and queries into SQL. This is a very attractive approach for commercial vendors because it is lightweight and it requires almost no changes to the underlying RDBMS. Work in this space focus on optimizing the translation layer to minimize joins or how to utilize existing indexes of RDBMSs to speed up processing. This approach is not performance focused and is limited by the underlying RDBMS’s baseline performance. In contrast, our approach modifies the underlying RDBMS to improve its performance on some joins. Similar approaches have also been taken by several systems, such as Grail [19] and graph layers above the Vertica column store [26] or the Aster system [46], that translate batch iterative graph computations, such as computing PageRank or finding connected components, into recursive SQL procedures.

SIP is a technique that is used in RDBMSs to avoid scanning large tables or indices or data from remote compute nodes [15, 21, 25, 37, 38, 52]. The use of sip closest to our work has been proposed by Neumann et al. [38] inside the RDF-3X system that manages RDF databases. This work has proposed using sip to avoid scans of large fractions of indices that store RDF triples. This work specifically targets queries with large joins but small outputs that contain sub-queries with non-selective filters. Evaluation of these sub-queries in regular execution requires large index scans, but by passing information from other sub-queries, the system can avoid scanning parts of the index. Zhu et al. [52] have used sip in a similar fashion to avoid large table scans in in-memory star schema data warehouses when using left-deep query plans in queries. This paper has demonstrated that the difference between the best and worst performing left-deep plans shrink significantly when using sip in contrast to without using sip, which makes the optimizer more robust. Our use of sip to integrate predefined joins is similar to the use of sip in these works with several differences. In these systems and in RDBMSs in general, joins are value-based so passing the values, which may be of arbitrary data types, requires compacting the keys in probabilistic data structures, specifically bloom filters. This requires running hash functions both when creating the filter as well as performing the semi-joins in scans. Since our pointer-based predefined joins are over dense integer-based ID, we can directly compact the keys in a deterministic bitmap filter and avoid any hash computations. Similar to reference [52], we also demonstrate that using sip in graph or relational workloads with large many-to-many joins makes the system more robust by analyzing GRainDB’s plan space.

The analogue of adjacency list indices in our solution are the *RID indices* (Section 5) that we use to index tables that are part of predefined joins. Our RID indices can be seen as a form of join index. Valduriez originally introduced join indices [50] to index results of arbitrary join queries, e.g., consisting of equality or inequality predicates, and index for each RID of one table, the list of matching RIDs from one or more other tables. Join indices are therefore simple materialized views. Similar to join indices, our RID

Person		Follows		
ID	name	ID1	ID2	year
101	Mahinda	101	202	2021
303	Karim	303	404	2019
202	Karim	101	303	2021
303	Carmen	202	303	2020
404	Zhang	101	404	2021

(a) Person table.

(b) Follows table.

Table 1: Input tables for our running example.

indices store RID keys and list of RID values but are over base tables instead of results of join queries. Prior work on join indices have focused on their efficient use within RDBMSs, such as extending them to support multi-table joins [39], designing fast disk-based join algorithms [29, 31] or their application to top-k query processing [49]. In our context, we use RID indices primarily to generate information to pass to scan operators and when a scan can be completely avoided, to reduce the number of joins in the query plan.

3 RID MATERIALIZATION

We start by describing the changes at the physical data storage layer of the system. Users predefine their joins using a PREDEFINE JOIN command that we added to the SQL dialect in DuckDB. In this command users specify an equality join from a table $F(A_{f1}, \dots, A_{fk_f})$ to $P(A_{p1}, \dots, A_{pk_p})$ on attributes $A_{f_{t_1}} = A_{p_{z_1}}, \dots, A_{f_{t_\ell}} = A_{p_{z_\ell}}$, such that $A_{f_{t_1}}, \dots, A_{f_{t_\ell}}$ forms a foreign key to P . Upon executing this command, the system adds a new column $RID(A_{f_{t_1}}, \dots, A_{f_{t_\ell}})$ to F that contains for each row $r_f \in F$, the RID of the row $r_p \in P$ to which r_f has the foreign key. This column is visible only to the system and not to users. RIDs in columnar RDBMSs serve as system-level pointers and can be directly used to compute the locations of rows in storage. So the $RID(A_{f_{t_1}}, \dots, A_{f_{t_\ell}})$ column stores for each r_f the pointer to the matching r_p , similar to how edges in GDBMSs point to their source or destination node records. If F contains foreign keys to multiple tables, multiple joins on F can be predefined. This is common for relationship tables that represents many-to-many joins, such as the Follows table in the next example.

EXAMPLE 1. Table 1 shows a simple database with two tables, a Person(ID, name) table and a Follows(ID1, ID2, year) table, that will serve as our running example. The ID1 and ID2 columns in Follows are both foreign keys to the ID column of Person. Table 2 shows the extended Follows table (as Follows') when a user predefines the Person.ID = Follows.ID1 and Person.ID = Follows.ID2 joins. The Follows table is extended with $RID(ID1)$ and $RID(ID2)$ columns (abbreviated as RID1 and RID2) that contain the RIDs of the rows in Person that match the values in the ID1 and ID2 columns, respectively. Both Person and Follows tables also have RID columns (abbreviated as R) that show the contiguous RIDs of the rows in these tables. These are shown in gray to indicate that unlike $RID(ID1)$ and $RID(ID2)$ columns, they are not materialized in storage.

4 SJOIN: SIP OF RIDS

Our implementation of predefined joins consists of two steps:

Step 1: Rule-based query optimization. We use the system's default optimizer to generate a regular logical plan for the query. We recursively traverse this plan and find each join operator that evaluates a predefined join from F to P , e.g., the Person.ID=Follows.ID1.

Person			Follows'					
R	ID	name	R	R1	ID1	R2	ID2	year
0	101	Mahinda	0	0	101	1	202	2021
1	202	Karim	1	2	303	3	404	2019
2	303	Carmen	2	0	101	2	303	2021
3	404	Zhang	3	1	202	2	303	2020
4			4	0	101	3	404	2021

(a) Extended Person table.

(b) Extended Follows table.

Table 2: Extended tables. RID columns are abbreviated as R and are in gray to indicate that they are not materialized. The $RID(ID_i)$ columns of Follows' are abbreviated as R_i .

In our implementation, these are HashJoin operators because DuckDB evaluates equality joins with HashJoin. Upon finding these HashJoins, we perform one of two sets of actions:

Case 1: F is the build and P is the probe side. In this case we make the following changes to the operators in the plan tree:

- HashJoin is replaced with a new join operator we call SJoin (explained momentarily in Step 2).
- Scan(F) operator (on the build side sub-tree) is modified to (i) scan the materialized RID column of F ; and (ii) if any of the original join attribute $F.A_{iz}$ is projected out later in the query, we remove the scan of A_{iz} from the scanned columns of F .

- Scan(P) operator (on the probe side) is replaced with a modified scan operator, which we refer to as ScanSJ, for **scan semijoin**.

As we discuss momentarily below in Step 2, if we are in Case 1, we will perform sip to pass information from F to P during evaluation.

Case 2: F is the probe and P is the build side. Now the changes are:

- HashJoin is now not replaced but we replace the join condition to be $P.RID = F.RID(A_{it_1}, \dots, A_{it_\ell})$. Note that because RIDs are integers and always form a single join attribute, this is more performant if the original join predicate contains multiple columns or non-integer data types, e.g., strings.
- Scan(F) operator (on the probe side sub-tree) is modified in exactly the same way as before.

If the plan is in Case 2, we cannot perform sip because we can pass information from F to P only if F is on the build side. This is because we need to read the information to pass from F to P before P is scanned. Alternatively, we can swap the build and probe sides, but we chose not to overwrite the optimizer's choice here. This is because as we next explain SJoin is a modified hash join operator and the optimizer optimizes to put the smaller table on the build side to keep the constructed hash table small.

Step 2: Sideways information passing during query evaluation:

If we are in Case 1, we then use sip during query evaluation from SJoin operators to the ScanSJ(P) operators. SJoin is a specialized hash join operator. SJoin performs the join on the replaced RID equality predicates instead of the original join columns in the query. In addition, SJoin passes the materialized RID values from scanned F tuples, which are pointers to P , to ScanSJ(P) operators in its probe side. Similar to standard hash join, SJoin first reads all of the tuples from its build side. These tuples contain materialized RID values that are scanned from F and point to the tuples in P . Using these RIDs, SJoin constructs two bitmask filters:

- **Zone bitmask:** For each zone of P , i.e., a block of tuples on disk, indicates whether the zone has any matching tuples joining with

F . This bitmask contains 1 bit for each zone and is constructed by taking the modulo of the RIDs with the zone size.

- **Row bitmask:** Indicates whether each row r_p of P joins with an F tuple. This bitmask contains $|P|$ many bits and is constructed by directly setting the positions of the seen RIDs to 1.

Note that unlike existing applications of sip in DBMSs [21, 25, 27, 38], the information we pass to scans do not need to be probabilistic filters, such as bloom filters. This is because RID values are dense integers and their exact domain, which is 0 to the number of tuples in P , is known by the system and can be compressed into a single bit. If P is very large, the row bitmask can be large, in which case a system can resort to even smaller filters at a granularity level between zones and individual tuples. Once SJoin receives all of the build side tuples, it passes both of these bitmasks to all of the ScanSJ(P) operators in its probe side recursively. These filters are used to perform the semijoin $P \ltimes F$ in the ScanSJ operators as follows. Zone bitmask is used to skip over scanning zones of P whose bits are 0. For zones with matching tuples, ScanSJ operator scans the zone into vectors as regular scan operator and adds a new RID vector to the intermediate tuples that store the RIDs of the scanned tuples. This does not require any actual I/O because RIDs of P are virtual positional offsets of the tuples, which can directly be written into the RID vector. For example, if zones are of size 1024 and the second zone has been read, then this vector contains values 1024 to 2047. Finally, to perform the semi-join, ScanSJ attaches the row bitmask of this zone as a selector vector to the intermediate tuples. This filters out the P tuples without matching F tuples.

EXAMPLE 2. Consider a query that finds two hop friends of Karim:

```
SELECT *
FROM Pers. P1, Follows F1, Pers. P2, Follows F2, Pers. P3
WHERE P1.ID=F1.RID1 AND F1.RID2=P2.ID AND P2.ID=F2.RID1
      AND F2.RID2=P3.ID AND P1.name = Karim
```

Figure 1a shows an example plan for this query that has: (i) replaced two HashJoins with SJoin operators; (ii) replaced two Scan Person table operators (for P2 and P3) with ScanSJ; and (iii) modified the Scan Follows operators to read the materialized RID columns. HashJoin₁ and HashJoin₂ operators are not replaced with SJoin because the Scans of F1 and F2 are on their probe sides. Instead, we only modify their join predicates to be over RIDs. The information passed from SJoin operators are in the form of two bitmasks, which can be seen at the ScanSJ P2 and ScanSJ P3 operators. The top one is the zone bitmask and the bottom one the row bitmask. The figure assumes zones of size 2. In our running example, HashJoin₁ joins the (1, 202, Karim) and (1, 202, P2.RID=2, 303, 2020) tuples, which is given to SJoin₁. Because the only matching P2 in this tuple has RID 2, the row bitmask passed to ScanSJ P2 is [0, 0, 1, 0] and the zone bitmask is [0, 1] because RID 2 is in the second zone. Therefore ScanSJ P2 only scans the second zone and puts the [1, 0] selector vector to the two tuples in this zone (filtering out the tuple with RID 3). The output of SJoin₁ is (1, 202, Karim, 2, 303, Carmen, 2020) and the following HashJoin₂ produces (1, 202, Karim, 2, 303, Carmen, 2020, P3.RID=3, 404, 2019). This is given to SJoin₂ (during build), which passes the [0, 1] zone bitmask and the [0,0,0,1] row bitmask to ScanSJ P3. The final output is (1, 202, Karim, 2, 303, Carmen, 2020, 3, 404, Zhang, 2019).

5 RID INDEX AND ITS APPLICATIONS

Next, we describe two applications of indexing the RID values in table F that contains materialized RIDs. We call this index the *RID index*. Section 5.1 describes our RID index and our first application, which is performing reverse semijoin of $F \ltimes P$ through sip. Section 5.2 considers the case when F is a relationship table, so contains two predefined joins and describes an optimization that merges two consecutive joins in a plan to avoid the scan of F completely.

5.1 Reverse Semijoins

In our approach of evaluating predefined joins so far, we can pass RID values only from F to scans of P and not vice versa, so we can only perform $P \ltimes F$ through sip. In many settings, F is a much larger table than P , and the ability to perform $F \ltimes P$ is very beneficial. For example, in LDBC benchmark with scale 30, Knows table is 41x larger than Person. However, given a row $r_p \in P$, we cannot directly find from the RID value of r_p the RIDs of rows $r_{f1}, \dots, r_{fp} \in F$ that join with r_p , as this list is not materialized in P . In order to perform this reverse semijoin, we need an index on F that for each r_p returns this list. We call this index the *RID index*. In our implementation, users can construct RID index on any table F on which at least one join has been predefined (say to a table P). Therefore, F already has a materialized RID($A_{it_1}, \dots, A_{it_r}$) column and its own virtual RID column. The RID index stores for each value in the RID($A_{it_1}, \dots, A_{it_r}$) column the RIDs of $r_{f1}, \dots, r_{fp} \in F$ that join with r_p . RID index is the analogue of adjacency list indices in GDBMSs and similar to many GDBMSs we store them in memory using a compressed sparse row data structure [16].

Recall that in absence of a RID index, we could not replace the join operators in the system's original plan if F was in the probe side of the join (Case 2 in Section 4). When there is a RID index, we replace such join operators with a modified SJoin operator we call SJoinIdxR and all of the Scan(F) operators on the probe side with ScanSJ(F) operators. The Idx suffix is for using the RID index and R suffix is for reverse. Similar to SJoin, SJoinIdxR builds a hash table, now of tuples from P and constructs the bitmasks for sip as follows: For each tuple r_p from the build side, SJoinIdxR consults the RID index on F to find the RIDs of the F tuples that join with r_p and sets the bits corresponding to these RIDs. Then, similar to SJoin, these bitmasks are passed to the ScanSJ(F) operators, which perform $F \ltimes P$ semijoins.

EXAMPLE 3. Figure 2 shows the RID index that indexes the (RID1, RID) columns of the Follows table, such that for each RID of a row r_p from the Person table, we have a list of RIDs of matching Follows tuples. Ignore the Follows(RID2) values in the figure for now. Figure 1b shows the plan we now generate in presence of this RID index. The two HashJoin operators from the plan in Figure 1a are replaced with SJoinIdxR operators and the previous Scan operators of the Follows table are replaced with ScanSJ operators. The figure also shows bitmasks that the new ScanSJ operators take. For example, the ScanSJ F1 operator takes a tuple bitmask with only the index 3 set to 1 and zone bitmask with only index 2 set to 1. This is because the RID of the (1, 202, Karim) tuple is 1 and 1's list of matching RIDs contains only the RID 3 of Follows, because 202 joins with (3, 1, 202, 2, 303, 2020) (see Table 2). This can also be seen from the RID index for Person.RID=1 in Figure 2.

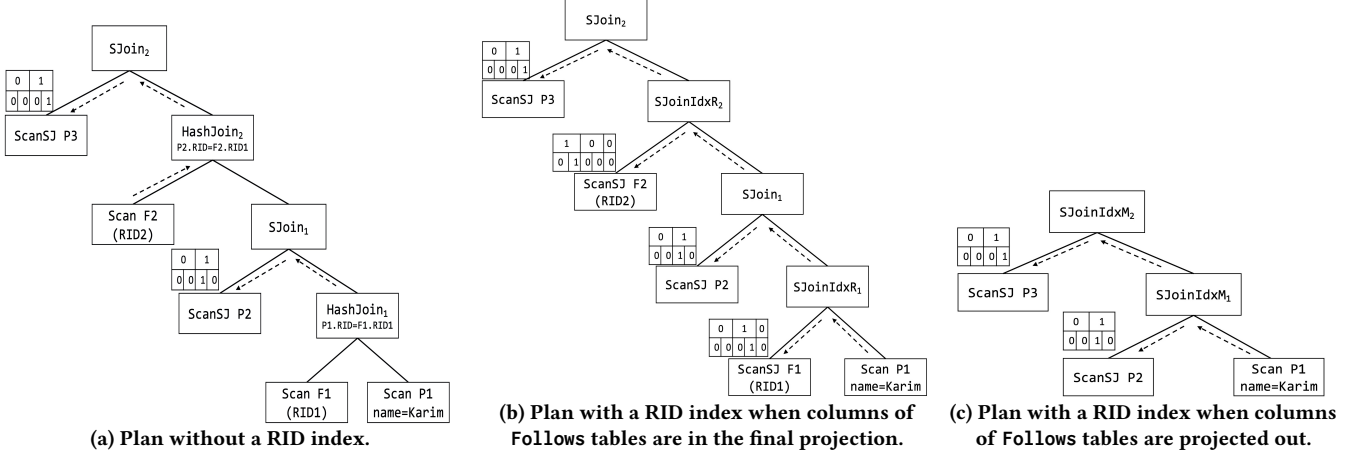


Figure 1: Example plans for our running example queries with different system configurations.

Follows(RID1)	0	1	2
Follows(RID)	0	2	4
Follows(RID2)	1	2	3

Figure 2: An example RID index on Follows in the CSR format. In graph terms, for each RID of a source entity Person (RID1) stores a set of edges and destination entities (Follows RID, Person RID) pairs, where destination Person RIDs are stored in Follows(RID2) column.

5.2 Extended RID Index and Join Merging

Many-to-many joins between two tables P_1 and P_2 that represent two (possibly same) sets of entities are often facilitated through a third relationship table F . In these settings, many queries use the table F to join P_1 and P_2 . This is, for example, the case in our running example, where each Follows table is joined with two Person tables. Therefore, it can be beneficial to predefine two joins on F . In this case, each row r_f of F would contain the virtual RID of F and two materialized RIDs, one for row $r_{p_1} \in P_1$ and the other $r_{p_2} \in P_2$ that r_f joins with. Consider building a RID index from the RIDs of P_1 to lists of RIDs of F tuples. So for each RID of P_1 , say i_1 , we store a list $L_{i_1} = \{r_{f_1}, \dots, r_{f_k}\}$ of RIDs of F rows that have i_1 in their materialized RID column for P_1 . We can also extend L_i to store the RIDs of P_2 tuples along with the RIDs of F as follows: $\{(r_{f_1}, r_{p_{2_1}}), \dots, (r_{f_k}, r_{p_{2_k}})\}$. This is similar to how GDBMSs store both the edge IDs and neighbor node IDs in their adjacency lists. Analogous to *forward* and *backward* adjacency list indices in GDBMSs, one can similarly build a second RID index that now stores for each RID of P_2 a list of RIDs of joining F and P_1 tuples. Figure 2 is an example “forward” extended RID index for the Follows table, that stores for each “source” Person tuple r_p , the list of RIDs of the joining Follows tuples, shown as Follows(RID) values, as well as the RIDs of the “destination” Person tuples that these Follows tuples point to, shown as Follows(RID2) values.

Consider a query that performs a join of $P_1 \bowtie F \bowtie P_2$ with the predefined conditions, but F is only used to facilitate the join, so: (i) there are no filters, group by and aggregations, or others joins on F ; and (ii) the final projection does not contain any columns of F . Then

we can use an extended RID index to directly join the P_1 tuples with P_2 tuples, without ever scanning the F table and joining it with P_1 or P_2 . We call this the *join merging optimization*. Specifically, in our query optimization step, we look for two consecutive join operators J_1 , evaluating the predefined $P_1 \bowtie F$, and J_1 ’s parent J_2 , evaluating $P_2 \bowtie F$ such that conditions (i) and (ii) above are satisfied. Note that if the query satisfies condition (i), i.e., F is not involved in any other joins, the only operator on the probe side of J_1 must be the scan of F . We replace J_1 and J_2 with a new SJoinIdxM operator S , where M stands for *merged*. S takes as its build side J_1 ’s build side and as its probe side J_2 ’s probe side, and we drop the scan of F , i.e., the probe side of J_1 . During evaluation for each P_1 tuple r_p , S looks for the RIDs of joining P_2 tuples directly from the RID index, and passes these RIDs as bitmasks to the ScanSJ(P_2) operators on its probe side, without ever scanning F . The join with F happens implicitly while accessing the RID index to read the RIDs of P_2 tuples.

Note that if a query needs to scan F , the extended index is not directly useful as the scanned F rows already materialize the P_2 RIDs which can directly be used to construct the bitmasks for sip.

EXAMPLE 4. Figure 1c shows our plan in the presence of an extended RID index from RID1 to RID2 columns of Follows. Observe that compared to the plan in Figure 1b, we have merged SJoinIdxR₁ and SJoin₁ into a new SJoinIdxM₁ operator and SJoinIdxR₂ and SJoin₂ into a new SJoinIdxM₂ operator.

6 IMPLEMENTATION CONSIDERATIONS

We next elaborate on two system components, optimizer and update handling, under our proposed integration of predefined joins. In our proposed solution, we have chosen to use the default join optimizer of DuckDB to generate an initial plan P_d and then replace some of the hash joins in a rule-based approach with our S-Join variants to obtain P_d^* . Even if P_d is the best default join order of DuckDB, in principle, modifying another plan P with predefined joins can outperform P_d^* . Therefore, one can extend our integration to develop a sip-aware optimizer to generate such plans. This opportunity arises for example on a query $T_1 \bowtie \dots \bowtie T_k$, where assume table T_k contains a very selective predicate and is very small after the predicate. Suppose further that the database has RID indices to pass

information from T_k to each T_i . Without sip, a left deep plan that first joins the small T_k with T_{k-1} and iteratively with T_{k-2}, T_{k-3} until T_1 would be more efficient than a left-deep plan with the reverse join order, i.e., one that first joins large tables and finally the smallest T_k . However, under sip, the reverse plan can be more efficient because it can pass information from T_k to each T_i . In Section 7, we present a plan spectrum study on a subset of the JOB benchmark queries to study the effects of predefined joins on the plan space of DuckDB and how much room for improvement we found for adopting a sip-aware optimizer. Although we found several queries for which we could obtain non-negligible improvements, broadly we found P_d^* and P_s^* competitive, which is why we chose not to modify the optimizer of the system within the scope of this work.

Second, although we do not focus on updates within the scope of this work, our integration requires further considerations for handling updates. Insertions of a tuple t to a table F that has a predefined join to P requires finding the RID of the tuple in P that t refers to and inserting it in the system-visible RID($A_{it_1}, \dots, A_{it_r}$) column F . If there is a RID index on F , possibly an extended one, we need to further update the index. Deletions also require additional handling. Observe that because we use RIDs as pointers, we materialize them in system-visible RID columns or a RID index. Therefore once a tuple t is assigned a RID, it needs to remain fixed. Suppose a tuple t with RID k is deleted. In addition to removing any references to k in system-visible RID columns or RID indices, the system also needs to keep track of the gap in k and assign it to the next inserted table. That is the system cannot shift tuples with RIDs $k + 1$, which would change a large number of RIDs and require updating the references to these RIDs. Reusing gaps or IDs is common practice both in RDBMSs and GDBMSs, e.g., MySQL reuses gaps left by deleted tuples to use for new insertions [9], and Neo4j keeps deleted nodeIDs in a separate file and reassigns them to later inserted nodes [4]. DuckDB, on which we base our work, currently leaves these gaps as they are there is an open issue in the system for handling them as the system is further developed [8].

7 EVALUATION

We next evaluate our proposed predefined join support that is implemented inside DuckDB. We call this version of DuckDB as GRainDB. Our goal is to demonstrate and validate several behaviors of our implementation. First, we aim to demonstrate that predefined joins provide fast join capabilities on several relational and graph workloads, improving the performance of vanilla DuckDB significantly as well as being competitive with a state-of-the-art specialized GDBMSs on many queries. We also demonstrate that our approach does not incur major overheads on workloads that are not amenable to benefiting from predefinition of joins. Second, we aim to perform an ablation study to demonstrate that each of our optimizations that facilitated different levels of integration has additional benefits. Third, we aim to compare the performance characteristics of our approach against index nested loop join-based implementations that are prevalent in GDBMSs and prior approaches that integrate predefined joins into RDBMSs [23, 32]. Fourth, we demonstrate that sip makes DuckDB’s optimizer more robust by analyzing the plan spectrum of DuckDB and GRainDB on a suite of queries.

7.1 Setup

Baseline Systems: We compare GRainDB against vanilla DuckDB and GraphflowDB, a state-of-art academic graph database system, as demonstrated in several prior publications [22, 28, 34, 35]. We use the most performant version of GraphflowDB from reference [22]. We also performed preliminary experiments with Neo4j’s community edition, but as with several prior work [22, 28, 34] did not find it competitive with GraphflowDB (or GRainDB on many queries) and omit these experiments. We emphasize that our goal in this paper is not to argue that an RDBMS can be more efficient, even after integrating these approaches, than an efficient GDBMS because specialized systems should be expected to be more performant on the workloads they optimize for. However, in our evaluations we will demonstrate that our approach can be competitive with a state-of-the-art GDBMS on many queries from graph workloads.

We also intended to compare against GQ-Fast and GR-Fusion. GQ-Fast and GR-Fusion are both academic prototype systems. However, the publicly available versions of both systems have errors on sample queries and are out of maintenance and we failed to setup these systems on our desired benchmarks. One of our goals in the GQ-Fast and GR-Fusion comparisons were to show that the pure left-deep and index nested loop join-based plans used by these approaches can be suboptimal to bushy and hash join-based plans of GRainDB. Instead, we will perform this comparison against similar plans from Neo4j and GraphflowDB. The GraphflowDB version we use [22] also only supports such plans.

Benchmarks: We expect predefined joins to provide performance improvements on queries with the following properties:

- (i) *Existence of predefined joins:* As a necessary condition, the query must contain at least one predefined join, so that we can replace a value-based hash join operator with our S-Join variants.
- (ii) *Existence of selective predicates on F and/or P :* This is critical because when F (or P) has a selective predicate, the semi-joins used by sip more effectively reduces the scan of P (or F) and probes in the following hash join.
- (iii) *Existence of one/many-to-many joins:* We also expect to see performance improvements when queries contain one/many-to-many joins for two reasons. First, predefined joins primarily improves join performance (as opposed to say aggregations), and the join performance is often an important runtime factor in queries with one/many-to-many joins, which are challenging as they lead to growing intermediate results. Second, the reverse semijoins and join-merging optimizations can benefit primarily queries with one/many-to-many joins. This is because these optimizations require a RID index, which is generally built on tables that represent one/many-to-many relationships between two other tables that represent entities, such as Follows in our running example.

In light of these, we used one relational and one graph benchmark that contain queries that satisfy these properties and for a more complete evaluation, a second relational workload that does not.

- **Join order benchmark (JOB)** on the IMDB dataset [30], which contains more than 2.5 M movie titles produced by 235K different companies with over 4 M actors. When using GRainDB, we predefine every one-to-many primary foreign key relationship in the database and for tables that represent many-to-many relationships, such as movie-companies, we build a RID index.

- LDBC Social Network Benchmark [11] (SNB) benchmark at scale factor 10 and 30, which is a commonly used graph benchmark that models a social networking application with users, forums, and posts. In relational format, LDBC10 dataset contains 8 entity (i.e., node) and 10 relationship (i.e., edge) tables, with a total number of 36.5M and 123.6M tuples, respectively. LDBC30 contains 106.8M entity and 385.2M relationship tuples. We use SNB primarily to compare against GraphflowDB (and Neo4j, which was not competitive). GraphflowDB is an academic prototype system that does not implement several language features, such as recursive queries. Therefore, we slightly modified the benchmark and refer to it as *SNB-M*, for **m**odified. We removed queries involving shortest paths and decomposed queries with variable-length joins into multiple queries, each of which has a fixed path join (we denote each version with a suffix $-\ell$, where ℓ denotes the length). SNB-M contains variants of 18 out of 21 queries from the original SNB interactive simple (IS) and interactive complex (IC) benchmarks. Our full SNB-M queries are listed in Appendix A. SNB is generated in both relational and property graph formats. We use the relational format in DuckDB and GRainDB. For every edge type in the graph format of SNB, e.g., Knows edges, we build a RID index over the corresponding table in GRainDB.
- TPC-H benchmark at scale factor 10. We include TPC-H to perform a sanity check that making the primary-foreign key joins on such traditional workloads does not hurt performance. We do not expect GRainDB to provide meaningful improvements on TPC-H as it does not contain selective many-to-many joins. We predefined every one-to-many primary foreign key relationships in GRainDB, such as customer and orders. Although we did not expect performance speedups, we still found several queries on which we obtained non-negligible runtime improvements.

In our detailed evaluation in Section 7.3, we also use modifications of some of the join queries from JOB and SNB-M to create microbenchmarks.

System Configurations and Hardware: We set DuckDB to the in-memory mode. GraphflowDB is already an in memory system. DuckDB is still in early stage and does not integrate full cardinality estimation. We observed that this limits its ability to choose good plans on many instances, especially in queries with large joins and selective predicates. To isolate the influence of join order selection, we injected true cardinalities into the system. In Appendix B, we present a demonstrative experiment on the JOB benchmark that this uniformly improves the performance of both DuckDB and GRainDB. The GraphflowDB version we use does not contain an optimizer, so does not need to estimate cardinalities. We manually picked the systems' best join order, which for many queries was obvious. For example, many of the queries in our benchmarks are path queries that have a highly selective predicate on the left-most node, in which case we picked the plan that evaluates the join from left to right.

All experiments were conducted on a machine with two Intel E5-2670 @2.6GHz CPUs and 256 GB of RAM, consisting of 16 physical cores and 32 logical cores. Because GraphflowDB runs only in serial mode, we set DuckDB to run in serial mode as well. All reported times are averages of five successive runs after a warm-up running.

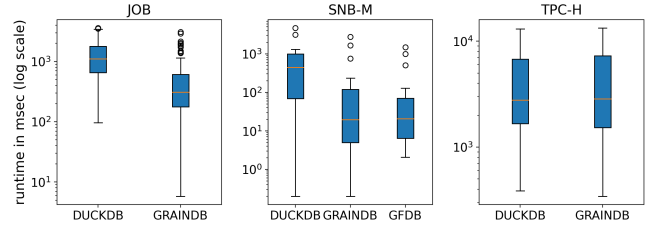


Figure 3: Runtimes (in ms) of DuckDB and GRainDB on JOB, SNB-M and TPC-H, and GraphflowDB on SNB-M.

Our measurements reflect the end-to-end query evaluation time, and a timeout of 10 minutes is imposed on each running.

7.2 End-To-End Benchmarks

We first present end-to-end evaluations on JOB, SNB-M, and TPC-H. We compare DuckDB and GRainDB on JOB and TPC-H, and DuckDB, GRainDB, and GraphflowDB on SNB-M. We expect to see large performance improvements of GRainDB over DuckDB on JOB and SNB-M because queries in JOB and SNB-M, with a few exceptions, satisfy the three properties we reviewed in Section 7.1. In contrast, we do not expect broad improvements in TPC-H, but expect minor overheads as well. For reference, Figure 3 presents box plot charts that show the performances of these systems on these workloads. Each boxplot shows the distribution of the runtimes of the queries in the workloads, specifying the 5th, 25th, 50th, 75th, and 95th percentiles of the distribution with marks.

7.2.1 JOB: Relational Workload with Selective Many-to-Many Joins. The box plots of DuckDB and GRainDB on JOB are shown in Figure 3. As we expect, we observe GRainDB outperforms DuckDB by large margins. JOB contains 113 queries. Table 3 lists detailed percentiles for query execution times of DuckDB and GRainDB of these queries. We see consistent large runtime reductions for each percentile. For example, the 25th percentile, median, and 75th percentile query execution times reduce respectively from 652.4ms to 176.4ms (3.7x), from 1110ms to 309ms (3.6x), and from 1797ms to 614.2ms (2.9x). For reference, Table 4 presents the execution times of a subset of the queries in JOB. Specifically, JOB queries contain between 2 to 6 variants and we present the first variant of each query in the table. The full table can be found in Appendix E. Importantly, we see consistent runtime improvements on all queries, with a few exceptions. Table 4 also presents the reduction on the amount of scanned tuples for each query in DuckDB and GRainDB. Although runtime reductions depend on many factors besides the reductions in scanned tuples, such as the actual outputs from the joins or how complex the predicate expressions are, this is still a good proxy for explaining when sip and predefined joins improve performance. For example, we observe that the queries in which we observe the largest improvement factors, such as Q6a, Q21a, Q27a and Q32a, also have large reductions in scanned tuples by 348.9x, 182.2x, 185.4x, and 53.8x. Similarly queries with negligible improvements, such as Q5a and Q20a also have respectively no or small (1.3x) reductions in scanned tuples.

7.2.2 SNB-M: Graph Workload with Selective Many-to-Many Joins. The box plots of DuckDB, GRainDB, and GraphflowDB on SNB-M are shown in Figure 3. Table 5 also lists detailed percentiles for query execution times of the systems. We see that GraphflowDB

	Min	5th	25th	50th	75th	95th	Max
DuckDB	96.0	203.6	652.4	1110.0	1797.0	2939.9	3584.6
GRainDB	5.8	27.4	176.4	309.0	614.2	1878.5	3104.4

Table 3: Detailed percentiles of runtimes (in ms) for DuckDB and GRainDB on JOB.

outperforms DuckDB by large margins on SNB-M. Specifically for the 25th percentile, median, and 75th percentile query execution times, GraphflowDB outperforms DuckDB respectively by 10.7x (68.4ms vs 6.4ms), 22.5x (441.8ms vs 20.8ms), and 14.1x (989.0ms vs 70.3ms). However, by predefining the joins in SNB-M, GRainDB closes this performance gap significantly, making DuckDB competitive with GraphflowDB on majority of the queries. Specifically for the 25th percentile, median, and 75th percentile query execution times, GRainDB and GraphflowDB compare as follows: 5.0ms vs 6.4ms (0.78x), 19.6ms vs 20.8ms (0.94x), and 119.4ms vs 70.3ms (1.7x). This shows that our implementation of predefined joins can make a columnar RDBMS competitive with a performant GDBMS on a workload that GDBMSs are optimized for.

Table 6 shows the detailed execution times of each query for all systems. We see that GRainDB outperforms DuckDB on almost all queries by up to 90x, except for IS1 and IS4, which are two small queries executed within 2ms. Similarly, GraphflowDB outperforms DuckDB in most queries by up to 426.6x. We observe that there are also 9 queries in which GRainDB outperforms GraphflowDB by large margins. We analyzed each of these queries to study GRainDB’s performance advantages over GraphflowDB. First are IS01 and IS04-IS07, which are point lookup queries over large base tables with inexpensive joins. Here, GraphflowDB resorts to sequential scans of these tables/nodes while GRainDB (and DuckDB) use a primary key index. This is not an inherent limitation of GraphflowDB plans and can be remedied if GraphflowDB also supports primary key indexes. On the remaining 4 queries, GRainDB plans have two separate advantages.

- *Bushy vs Left-deep Plans:* IC1-3, IC6-2, IC11-2 are three queries in which GRainDB outperforms DuckDB and uses a bushy plan. We describe IC6-2 as an example. IC6-2 is a complex query with 8 joins in SQL. In graph version, this is a 5-path query with selective predicates on both ends of the path. GraphflowDB does not implement bushy plans, so this query is implemented with a left-deep plan. This is less performant than the bushy plan that GRainDB uses that breaks the path into two parts. Figure 8 and 9 in Appendix D show the plans from both systems. This is an example of when the left-deep plan-based approaches to evaluate such path queries, which are used in systems like GR-Fusion and GQ-Fast, can be suboptimal to bushy plans.
- *Hash Join vs Index Nested Loop Joins and Scanning Edges Before vs After Joins:* The IC9-2 query is a smaller query with 4 joins in SQL. This is a 3-path query that also has filters on both ends. Now both systems use left-deep plans. The majority of the time in this query is spent in the very last join, which requires joining 7681 tuples from a Person table with a Comments table. 2.7M of these tuples successfully join with the 7681 tuples and 2.4M of these also pass a filter on the Comments table. In graph terms, 7681 Person nodes have 2.7M outgoing Comment edges (so an average degree of 351). As every GDBMS we are aware of, GraphflowDB follows these steps: (i) *joining nodes with edges*: looks up the edges

of each of 7681 keys in a large adjacency list index that point to Comments. This effectively performs 7681 random lookups into a hash table of size 26.5M, and then generates 2.7M intermediate tuples. (ii) *property scan and edge filtering*: reads the necessary properties of the 2.7M Comments and runs the filter predicates on these edges. In contrast, GRainDB, as typical of columnar RDBMSs for evaluating equality joins, follows these steps: (i) *hash table build*: creating a hash table of size 7681; (ii) *edge scanning and filtering*: sequentially scanning a large Comments table with 26.5M tuples and running a predicate on them which returns 2.4M tuples; (iii) *joining nodes with edges*: and finally doing 2.4M lookups into this very small hash table and performing the join. Now the joins happen after a sequential scan and filter of the “edge” table, leveraging columnar RDBMS techniques for highly optimized for sequential scans and filters of large columns. In addition, in the final join, now the lookups are into a very small hash-table instead of a large adjacency list index. This is more performant than performing the joins by lookups into a large index and non-sequentially scanning and filtering the joined edges. It is interesting to note that no GDBMS we are aware of (nor GRFusion or GQFast’s approaches to perform predefined joins) generates plans that can sequentially scan and filter all of the edges and then join these edges with their source nodes. GDBMSs typically joins nodes with their edges by performing lookups using the nodes as keys. For example, Neo4j’s plan on the same query is the same as GraphflowDBs. We will present a more controlled experiment to demonstrate this difference in Section 7.3.2.

Finally, this experiment gives us a point of comparison for the memory consumption of our RID indices. We profiled the memory consumption of GRainDB’s RID indices and GraphflowDB’s adjacency list indices. For each RID index we have on SNB-M, there is a corresponding adjacency list in GraphflowDB. In total, GRainDB’s RID indices take 5.9GB while GraphflowDB’s indices take 2.8GB. This is expected because GraphflowDB implements several compression techniques [22], such as compressing trailing 0s in IDs, which in SNB-M reduces 8 byte IDs to 4 bytes, and compressing empty/null adjacency lists. Instead, we store each RID in 8 bytes. As acknowledged in reference [22], the compression techniques in GraphflowDB are modifications of widely adopted techniques from columnar RDBMSs and can be integrated into our index implementation to close this gap.

7.2.3 TPC-H: Traditional OLAP Workloads. For completeness of our work and to verify that predefined joins have small overheads in a workload that does not contain many queries with selective many-to-many joins, we also compared the performances of DuckDB and GRainDB on TPC-H. The box plots of DuckDB and GRainDB are shown in Figure 3. Table 9 in the Appendix E also gives the detailed execution time of each query. As expected we do not see significant speedups or slowdowns on this benchmark. GRainDB replaces value-based hash joins with predefined joins in 13 of the 22 queries in TPC-H. The median runtime improvement out of these queries is 1.1x, with the maximum slow-down and speedup of 0.8x (so 1.2x slowdown) and 2.6x, respectively. Interestingly, even on a benchmark of traditional analytical queries, we found two queries with one/many-to-many joins on which replacing value-based joins

	1a	2a	3a	4a	5a	6a	7a	8a	9a	10a	11a	12a	13a	14a	15a	16a	17a
DuckDB	234.2	207	1491.4	216	96	885.4	879.8	1307.4	1933.2	1404.2	264.2	684	981.2	1644.2	987.8	652.4	1164.8
GRainDB	34.2	154.0	328.0	114.2	116.4	57.0	203.4	349.8	554.8	614.2	44.4	242.4	336.8	202.2	283.4	177.6	486.0
	6.8x	1.3x	4.5x	1.9x	0.8x	15.5x	4.3x	3.7x	3.5x	2.3x	6.0x	2.8x	2.9x	8.1x	3.5x	3.7x	2.4x
Scan Reduction	751.2x	23.9x	97.2x	58.7x	1x	348.9x	47.0x	9.2x	6.0x	13.0x	73.7x	44.8x	42.4x	142.1x	7.4x	14.4x	10.7x

	18a	19a	20a	21a	22a	23a	24a	25a	26a	27a	28a	29a	30a	31a	32a	33a
DuckDB	1797.0	2632.4	1118.4	1629.2	1471.8	866.6	2554.8	2318.8	1074.6	761.4	2068.4	2742.6	2198.0	2523.8	126.0	336.2
GRainDB	612.6	491.6	1071.6	54.2	864.0	296.8	788.8	1376.6	733.4	44.4	240.2	266.6	673.0	612.0	8.2	149.4
	2.9x	5.4x	1.0x	30.1x	1.7x	2.9x	3.2x	1.7x	1.5x	17.1x	8.6x	10.3x	3.3x	4.1x	15.4x	2.3x
Scan Reduction	13.8x	7.8x	1.3x	182.2x	7.6x	8.2x	8.7x	32.2x	15.7x	185.4x	156.9x	10.8x	282.2x	165.5x	53.8x	2.6x

Table 4: Runtimes (in ms) of DuckDB and GRainDB on each query in JOB.

	Min	5th	25th	50th	75th	95th	Max
DuckDB	0.2	1.5	68.4	441.8	989.0	2762.5	4647.0
GRainDB	0.2	0.7	5.0	19.6	119.4	1482.4	2768.0
GraphflowDB	2.1	2.5	6.4	20.8	70.3	888.2	1473.6

Table 5: Detailed percentiles of runtimes (in ms) for DuckDB, GRainDB, and GraphflowDB on SNB-M.

with predefined joins lead to visible speedups (2.6x for Q2 and 1.8x for Q3) and no queries visibly slowed down, indicating the low performance overheads of our implementation when queries are not suitable to benefitting from predefined joins.

7.3 Detailed Evaluation

We next provide a more detailed evaluation consisting of (i) an ablation study to verify that each of our optimizations on top of DuckDB leads to additional performance benefits; (ii) a controlled experiment comparing the performances of index nested loop join-based plans (adopted in GDBMSs and systems such as GR-Fusion and GQ-Fast) and our hash-join-based plans when joining records from relationship tables with entity tables under varying selectivities; and (iii) an analysis of the effects of our sip-based predefined joins in the plan space of DuckDB on a suite of queries.

7.3.1 Ablation Study. We performed an ablation study, to show the positive performance benefits of each of the optimizations we integrated into DuckDB: (i) RID materialization (Section 3); (ii) reverse semijoins (Section 5.1); and (iii) extended RID index and join merging (Section 5.2). Note that our optimizations are not independent of each other, e.g., without RID materialization we cannot perform either reverse semijoins or join merging. We therefore turned them off in a specific order and in growing sets. We first turned off extended RID index and join merging (-JM), then, we turned off reverse semijoins (-JM-RSJ), and finally we turned off all optimizations, which gives us vanilla DuckDB. Then we ran each version of the system on the SNB-M benchmark. Figure 4 shows the box plot charts of each version of the system. GR-FULL in the figure is the configuration with all optimizations on. We see that each optimization has a positive effect on performance, which can be seen by inspecting the median and 25 percentile lines, which consistently shift down as we add more optimizations. We see most impact from the reverse semijoin optimization, which is expected as it allows passing information from smaller entity tables (P in our notation) to much larger relationship tables (F in our notation). For reference, we show the runtime numbers of each query on each

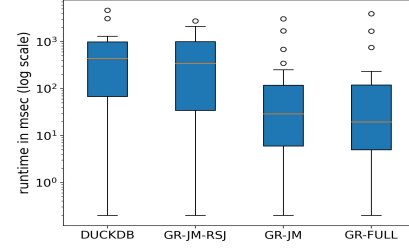
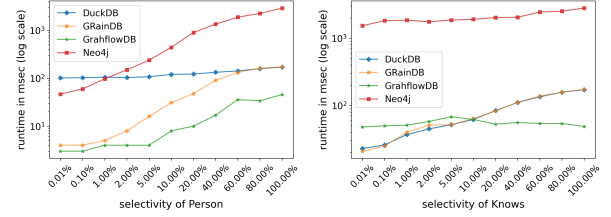


Figure 4: Ablation tests for different optimizations in GRainDB. DuckDB implements no optimizations. GRainDB-JM-RSJ only implements RID materialization. GRainDB-JM in addition implements reverse semijoins. GRainDB in addition implements join merging.



(a) MICRO-P.

(b) MICRO-K.

Figure 5: Runtimes (in ms) of DuckDB, GRainDB, GraphflowDB and Neo4j. The left and right figures show, respectively, the times on MICRO-P and MICRO-K queries, which vary the selectivity of Person and Knows, respectively.

system configuration in Appendix F. We observe queries where RID materialization leads up to 29.6x additional improvement (IS6), reverse semijoins up to 40.9x (IC7) additional improvements, and join merging up to 7.3x (IC2) additional improvements.

7.3.2 Performance of Predefined Joins Under Varying Entity vs Relationship Table Selectivity. We next do a controlled experiment to demonstrate the behavior of our sip- and hash-join based implementation of predefined joins under various selectivities on the P and F tables. Our goals are twofold: (i) to show the cases when sip yields performance improvements; and (ii) to demonstrate the different performance behaviors of index nested loop joins, which GDBMSs use, vs hash joins, which many RDBMSs, including DuckDB, use for equality joins. We take the LDBC30 dataset and the 1-hop $(p_1:Person) \xrightarrow{e:Knows} (p_2:Person)$ query, where the Person and Knows tables have 18.4K and 7.5M tuples, respectively. This

	IS1	IS2	IS3	IS4	IS5	IS6	IS7	IC1-1	IC1-2	IC1-3	IC2	IC3-1	IC3-2
DuckDB	0.8	524.8	36.6	0.2	4.5	148.0	989.0	38.0	72.0	110.5	926.0	1177.8	4647.0
GRainDB	1.2	19.6	3.4	0.2	0.6	5.0	11.0	4.0	6.4	38.2	134.8	119.4	1665.0
	0.7x	26.8x	10.8x	1.0x	7.5x	29.6x	90.0x	9.5x	11.2x	2.9x	6.9x	9.9x	2.8x
GraphflowDB	6.8	3.0	2.5	42.7	82.7	66.4	72.0	2.1	6.4	70.3	47.8	11.5	505.4
	0.1x	175.8x	14.5x	0.005x	0.05x	2.2x	13.7x	18.3x	11.3x	1.6x	19.4x	102.6x	9.2x

	IC4	IC5-1	IC5-2	IC6-1	IC6-2	IC7	IC8	IC9-1	IC9-2	IC11-1	IC11-2	IC12
DuckDB	402.0	636.0	3125.0	244.6	471.2	1186.8	1017.0	441.8	1312.6	35.8	68.4	788.4
GRainDB	54.0	174.0	2768.0	13.0	22.0	33.2	14.0	113.6	752.0	2.8	9.0	234.8
	7.4x	3.7x	1.1x	18.8x	21.4x	35.7x	72.6x	3.9x	1.8x	12.8x	7.6x	3.4x
GraphflowDB	12.3	20.8	984.0	8.1	127.2	2.8	2.8	55.6	1473.7	2.6	14.4	28.8
	32.6x	30.6x	3.2x	30.2x	3.7x	426.6x	359.5x	7.9x	0.9x	13.8x	4.8x	27.3x

Table 6: Runtimes (in ms) of DuckDB, GRainDB, and GraphflowDB on each query in SNB-M.

is representative of the general case when there are many more relationship/edge tuples than node/entity tuples in databases. We then run two sets of micro-benchmark queries: (1) MICRO-P: we fix a predicate with 99.9% selectivity on the `creationDate` property of `Knows` and vary the selectivity of a predicate on the `id` property of `Person` between 0.01% to 100%. (2) MICRO-K: we now fix a predicate with 99.9% selectivity on the `id` property of `Person` and vary the `creationDate` property of `Knows` between 0.01% to 100%. We run each set of queries on DuckDB, GRainDB, GraphflowDB and Neo4j. Our goal in including Neo4j in these experiments, which was omitted in our baselines, is to show that the two GDBMSs behave very similarly albeit in different performance levels. Before we discuss the results, we note that in both sets of queries GraphflowDB and Neo4j’s executions are always as follows: (i) scan the `Person` nodes and their `id` property and run the filter on `id`; (ii) join these nodes with their `Knows` edges by index nested loop join using the `Knows` adjacency list index; (iii) read the `creationDate` property of the joined edges and run a filter. This execution is standard in every GDBMS we are aware of, where joins are always from nodes to edges and we will momentarily show that this is in fact too rigid and can be suboptimal. This is also the execution in systems such as GR-Fusion and GQ-Fast.

Figure 5a shows the results for MICRO-P. First, we note that on all MICRO-P queries, DuckDB makes `Person` the build side as it is already much smaller than `Knows` and gets even smaller as we decrease the selectivity on the predicate on `Person`. Therefore, in GRainDB, as we decrease the selectivity, we can pass selective information to `Knows` table and decrease the amount of scanned `Knows` tuples and hash table probes. Therefore we see GRainDB to outperform DuckDB significantly and close the gap with GraphflowDB’s performance at these lower selectivities. Second, observe that both GDBMSs have consistent upward curves, indicating that their runtimes decrease as selectivity on the `Person` nodes decreases. This happens because the amount of join work that GDBMSs perform decreases proportionately as fewer `Person` nodes pass the filter. We cannot observe this desirable behavior with DuckDB because although its cost of hash table build decreases, its probe cost, which is the dominant cost here, does not. In fact, although broadly Neo4j is not competitive with other systems, it can still outperform DuckDB at lower selectivities on MICRO-P, because of its performance gains from decreasing selectivity on `Person`. Unlike DuckDB, GRainDB

however also behaves similarly to GDBMSs and obtains this desirable behavior because it can also decrease the amount of probes through sip. Finally, we note that it is expected in this microbenchmark that GraphflowDB is the most performant system at all selectivity levels because it can decrease its probes with decreasing selectivity and by default has several advantages over DuckDB and GRainDB, such as not incurring the cost of building hash tables or accessing data without going through a buffer manager because it is an in-memory system (which DuckDB does).

We next analyze the results of MICRO-K, shown in Figure 5b. First observe that now the GDBMSs do not react as positively to the decreasing selectivity on `Knows`. This is because now selectivity on `Person` is fixed, so the amount of probes GDBMSs perform is fixed. So both Neo4j and GraphflowDB curves are relatively straight (similar to the DuckDB curve in Figure 5a). Now note that DuckDB has a downward curve. This is because at all selectivity levels except 0.01% and 0.01%, DuckDB chooses `Person` as the build side (recall that the predicate on `Person` is not selective), so decreasing the selectivity proportionately decreases the probe amount. DuckDB can even outperform GraphflowDB when the selectivity is low enough. Note also that as expected GRainDB does not improve the performance of DuckDB now because although it passes information from `Person` to `Knows`, since `Person` does not have a selective predicate (it is fixed at 99.9%), this information is not useful. However, we also do not observe visible overheads. We see minor benefits at the lowest two selectivity levels, when DuckDB starts to choose `Knows` as the build side, and can pass selective information to scans of `Person`. Although we do not observe major improvements, this shows the flexibility of join processing in RDBMS, where there is no notion of node vs edge tables and for hash joins, systems can make any table the probe or build side. In contrast, in every GDBMS we are aware of, first node records are scanned and then the adjacency list indices are probed with the IDs of these nodes to perform the join (and not vice versa). As the MICRO-K benchmark demonstrates this can sometimes prevent them from benefiting from selective predicates on the edge records.

7.3.3 Effectiveness of Predefined Joins in the Plan Space and Room for Improvement for a Sip-aware Optimizer. Our next set of experiments analyze the effects predefined joins have on the plan space of DuckDB. Prior work has observed that sip-based [52] query processing makes systems broadly more robust to join order selection by decreasing the performance differences between different join orders and more importantly by increasing the number of join

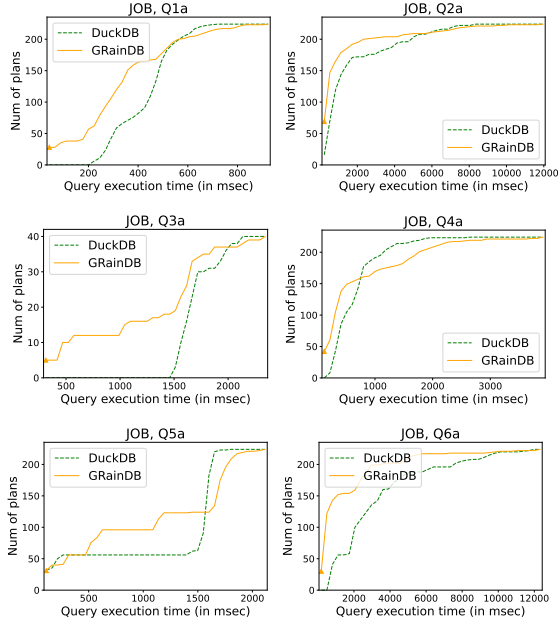


Figure 6: Cumulative distributions of the number of DuckDB and GRainDB plans (y-axis) that have runtimes below different thresholds (x-axis).

orders that perform competitive with the best default order. To demonstrate that our proposed solution also has similar effects on the plan space, we picked the first six query groups in JOB, and for the first two variants of each query (so a total of 12 queries) performed a plan spectrum analysis as follows. We take each plan P for each Q , corresponding to one join order, and execute both the default version of P (call it P_d) and the version where we apply our optimizations to change value-based joins with sip-based predefined joins (call it P_d^*). We then plotted two cumulative distribution lines for Q , one for the set of P_d and one for P_d^* plans, which show the number of plans (on y-axis) for different runtime value cutoffs, e.g., 100ms, 200ms, ..., 1000ms (on x-axis). Figure 6 shows the distributions we obtained for the first variants of these queries. The remaining charts are shown in Appendix G. Dashed and straight lines are the distributions of P_d and P_d^* plans, respectively. We observe on left sides of the curves, which summarize the best performing plans, the line showing P_d^* is consistently above the line for P_d . This shows that by predefined joins, we obtain larger sets of good plans. In many of these queries we also observe runtimes that were not achievable by any default plan. For example, on Q1a, while there are 60 plans with a runtime of ≤ 200 ms under predefined joins, there is no such plan with default value-based joins. In addition, there are now a set of 35 plans with runtime of ≤ 100 ms. Therefore not only are the best plans under predefined joins more performant, the optimizer is more robust to making mistakes when picking a plan, as there is a larger set of good performing plans. We also observe that on the right ends of many curves, which plot the set of worst-performing plans, the curve for P_d^* plans is now below the curve for P_d plans. This is also expected because we expect there to be some plans that do not benefit from predefined joins and instead incur minor overheads that S-Join operators incur, e.g., to prepare bitmasks.

	Q1a	Q1b	Q2a	Q2b	Q3a	Q3b	Q4a	Q4b	Q5a	Q5b	Q6a	Q6b
P_{Duck}^*	34	3	154	143	328	502	114	73	116	146	57	97
P_{opt}^*	31	3	77	67	287	135	72	47	110	112	32	82

Table 7: Runtimes (in ms) of P_{Duck}^* and P_{opt}^* on JOB queries.

Let us call the P_d^* of a plan P_d the *predefined version* of P_d . Recall that for a query Q , GRainDB’s plan, P_{Duck}^* is the predefined version of the plan P_{Duck} that DuckDB’s default optimizer picks for Q . Next, on these 12 queries, we analyzed the potential room for improvement on our rule-based approach if a system implements a sip-aware optimizer. We do a thought process and assume that an oracle sip-aware optimizer could pick the best GRainDB plan P_{opt}^* , i.e., the best performing P_d^* , and compare it against P_{Duck}^* . Table 7 shows this comparison. Although we did not find large rooms of improvement on most of these queries, we still found several queries, Q2a, Q2b, and Q3b with $>2x$ improvements. The largest improvement is on Q3b, from 502ms to 135ms (3.7x). Q3b is a join query with 4 tables, and contains a selective predicate on a table called keyword that returns only 30 of the 134K tuples in this table. P_{opt}^* is a left-deep plan where the last-join has keyword on its build side. Normally putting this table as the last join on a left-deep plan is not efficient because joining the smaller tables first and creating smaller intermediate results is more efficient. However, under sip, this is a good plan because this can lead to iterative information passing to reduce the amount of scans in other tables. Instead P_{Duck}^* uses a bushy plan. In principle, a sip-aware optimizer that can accurately estimate the selectivities of our zone and bit mask filters can further improve our performance by generating such plans that are normally not efficient but become efficient due to sip. However, our analysis indicates that our rule-based approach also generates competitive plans on almost all of these queries.

8 CONCLUSIONS

We described a novel approach to integrate predefined and pointer-based joins, which are prevalent in GDBMSs, into columnar RDBMSs. Our approach is based on materializing and optionally indexing RIDs similar to how edges are indexed in adjacency lists. In contrast to native GDBMSs and prior implementations of predefined joins in RDBMSs [23, 24, 32, 33] that use such indices in index nested loop joins, we use them primarily to generate semi-join filters that are passed from hash join operators to scans. This ensures sequential scans when results of joins need to access other properties of joined tuples for further processing, such as running predicates. Unlike prior approaches that propose a graph-specific optimizer and query processor that produce left-deep join plans, our approach directly leverages the default optimizer of the system to generate an arbitrary and possibly bushy plan, and transforms this plan by replacing some join and scan operators. We also described an optimization that can use an extended RID index to avoid scans of a relationship table entirely in some settings. We demonstrated the practicality of our approach by implementing it in the DuckDB system [41, 42] and demonstrated its performance benefits on both relational as well as graph workloads that contain queries with large selective many-to-many joins that can benefit from predefined, such as the LDBC SNB benchmark on which our approach makes DuckDB competitive with the state-of-the-art GraphflowDB GDBMS.

REFERENCES

- [1] 2021. CWI. <https://www.cwi.nl>
- [2] 2021. DGraph. <https://dgraph.io>
- [3] 2021. GRainDB. <https://github.com/graindb/graindb>
- [4] 2021. How deletes work in Neo4j. <https://neo4j.com/developer/kb/how-deletes-workin-neo4j/>
- [5] 2021. JanusGraph. <https://janusgraph.org>
- [6] 2021. Neo4j. <http://neo4j.com>
- [7] 2021. TigerGraph. <http://tigergraph.com>
- [8] 2021. Vacuum deletes in DuckDB. <https://github.com/duckdb/duckdb/issues/109>
- [9] 2021. Value reuse for auto-increment columns in MySQL. <https://dev.mysql.com/doc/refman/8.0/en/delete.html>
- [10] Zahid Abul-Basher, N. Yakovets, P. Godfrey, Stanley Clark, and M. Chignell. 2021. Answer Graph: Factorization Matters in Large Graphs. In *EDBT*.
- [11] Renzo Angles, János Benjamin Antal, Alex Averbuch, Peter Boncz, Orri Erling, Andrey Gubichev, Vlad Haprian, Moritz Kaufmann, Josep Lluís Larriba Pey, Norbert Martínez, et al. 2020. The LDBC social network benchmark. *CoRR* (2020).
- [12] Nafisa Anzum, Semih Salihoglu, and Daniel Vogel. 2019. GraphWrangler: An Interactive Graph View on Relational Data. In *ICDE*.
- [13] M. M. Astrahan, M. W. Blasgen, D. D. Chamberlin, K. P. Eswaran, J. N. Gray, P. P. Griffiths, W. F. King, R. A. Lorie, P. R. McJones, J. W. Mehl, G. R. Putzolu, I. L. Traiger, B. W. Wade, and V. Watson. 1976. System R: Relational Approach to Database Management. *TODS* 1, 2 (1976).
- [14] Charles W Bachman. 1973. The programmer as navigator. *Commun. ACM* 13, 6 (1973).
- [15] Francois Bancilhon, David Maier, Yehoshua Sagiv, and Jeffrey D Ullman. 1985. Magic Sets and Other Strange Ways to Implement Logic Programs (Extended Abstract). In *PODS*.
- [16] Angela Bonifati, George Fletcher, Hannes Voigt, and Nikolay Yakovets. 2018. *Querying Graphs*. Morgan & Claypool.
- [17] Edgar Frank Codd. 1970. A relational model of data for large shared data banks. *Commun. ACM* 26, 1 (1970).
- [18] Edgar F Codd. 1982. Relational database: a practical foundation for productivity. *Commun. ACM* 25, 2 (1982).
- [19] Jing Fan, Adalbert Gerald Soosai Raj, and Jignesh M Patel. 2015. The Case Against Specialized Graph Analytics Engines.. In *CIDR*.
- [20] Michael Freitag, Maximilian Bandle, Tobias Schmidt, Alfons Kemper, and Thomas Neumann. 2020. Adopting Worst-Case Optimal Joins in Relational Database Systems. *PVLDB* 13, 12 (2020).
- [21] Goetz Graefe. 1993. Query evaluation techniques for large databases. *ACM Computing Surveys (CSUR)* 25, 2 (1993).
- [22] Pranjal Gupta, Amine Mhedhbi, , and Semih Salihoglu. 2021. Integrating Column-Oriented Storage and Query Processing Techniques Into Graph Database Management Systems. *CoRR* (2021).
- [23] Mohamed S. Hassan, Tatiana Kuznetsova, Hyun Chai Jeong, Walid G. Aref, and Mohammad Sadoghi. 2018. Extending In-Memory Relational Database Engines with Native Graph Support. In *EDBT*.
- [24] Mohamed S Hassan, Tatiana Kuznetsova, Hyun Chai Jeong, Walid G Aref, and Mohammad Sadoghi. 2018. Gfusion: Graphs as first-class citizens in main-memory relational database systems. In *SIGMOD*.
- [25] Zachary G Ives and Nicholas E Taylor. 2008. Sideways information passing for push-style query processing. In *ICDE*.
- [26] Alekh Jindal, Samuel Madden, Malú Castellanos, and Meichun Hsu. 2015. Graph analytics using vertica relational database. In *Big Data*.
- [27] Srikanth Kandula, Laurel Orr, and Surajit Chaudhuri. 2019. Pushing data-induced predicates through joins in big-data clusters. *PVLDB* 13, 3 (2019).
- [28] Chathura Kankanamge, Siddhartha Sahu, Amine Mhedhbi, Jeremy Chen, and Semih Salihoglu. 2017. Graphflow: An active graph database. In *SIGMOD*.
- [29] Hui Lei and Kenneth A. Ross. 1999. Faster joins, self-joins and multi-way joins using join indices. *Data & Knowledge Engineering* 29, 2 (1999).
- [30] Viktor Leis, Andrey Gubichev, Atanas Mirchev, Peter Boncz, Alfons Kemper, and Thomas Neumann. 2015. How good are query optimizers, really? *PVLDB* 9, 3 (2015).
- [31] Zhe Li and Kenneth A. Ross. 1999. Fast Joins Using Join Indices. *VLDBJ* 8, 1 (1999).
- [32] Chunbin Lin, Benjamin Mandel, Yannis Papakonstantinou, and Matthias Springer. 2016. Fast In-Memory SQL Analytics on Typed Graphs. In *ICDE*.
- [33] Chunbin Lin, Jianguo Wang, and Yannis Papakonstantinou. 2017. GQFast: Fast graph exploration with context-aware autocompletion. In *ICDE*.
- [34] Amine Mhedhbi, Pranjal Gupta, Shahid Khaliq, and Semih Salihoglu. 2021. A-Indexes: Lightweight and Highly Flexible Adjacency Lists for Graph Database Management Systems. In *ICDE*.
- [35] Amine Mhedhbi, Chathura Kankanamge, and Semih Salihoglu. 2021. Optimizing One-time and Continuous Subgraph Queries using Worst-Case Optimal Joins. *TODS* (2021).
- [36] Amine Mhedhbi and Semih Salihoglu. 2019. Optimizing subgraph queries by combining binary and worst-case optimal joins. *PVLDB* 12, 11 (2019).
- [37] Inderpal Singh Mumick and Hamid Pirahesh. 1994. Implementation of Magic-Sets in a Relational Database System. In *SIGMOD*.
- [38] Thomas Neumann and Gerhard Weikum. 2009. Scalable join processing on very large RDF graphs. In *SIGMOD*.
- [39] Patrick O’Neil and Goetz Graefe. 1995. Multi-Table Joins through Bitmapped Join Indices. *SIGMOD Rec.* 24, 3 (1995).
- [40] Jignesh M Patel, Harshad Deshmukh, Jianqiao Zhu, Navneet Potti, Zuyu Zhang, Marc Spehlmann, Hakan Memisoglu, and Saket Saurabh. 2018. Quickstep: A data platform based on the scaling-up approach. *PVLDB* 11, 6 (2018).
- [41] Mark Raasveldt and Hannes Mühleisen. 2019. DuckDB: an embeddable analytical database. In *SIGMOD*.
- [42] Mark Raasveldt and Hannes Mühleisen. 2020. Data Management for Data Science - Towards Embedded Analytics. In *CIDR*.
- [43] Marko A Rodriguez. 2015. The gremlin graph traversal machine and language (invited talk). In *DBPL*.
- [44] Michael Rudolf, Marcus Paradies, Christof Bornhövd, and Wolfgang Lehner. 2013. The graph story of the SAP HANA database. *BTW*.
- [45] Siddhartha Sahu, Amine Mhedhbi, Semih Salihoglu, Jimmy Lin, and M Tamer Özsu. 2020. The ubiquity of large graphs and surprising challenges of graph processing: extended survey. *VLDBJ* 29, 2 (2020).
- [46] David Simmen, Karl Schnaitter, Jeff Davis, Yingjie He, Sangeet Lohariwala, Ajay Mysore, Vinayak Shenoi, Mingfeng Tan, and Yu Xiao. 2014. Large-scale graph analytics in aster 6: bringing context to big data discovery. *VLDBJ* 7, 13 (2014).
- [47] Wen Sun, Achille Fokoue, Kavitha Srinivas, Anastasios Kementsietsidis, Gang Hu, and Guotong Xie. 2015. Sqlgraph: An efficient relational-based property graph store. In *SIGMOD*.
- [48] Yuanyuan Tian, En Liang Xu, Wei Zhao, Mir Hamid Pirahesh, Sui Jun Tong, Wen Sun, Thomas Kolanko, Md Shahidul Haque Apu, and Huijuan Peng. 2020. IBM Db2 Graph: Supporting Synergistic and Retrofittable Graph Queries Inside IBM Db2. In *SIGMOD*.
- [49] P. Tsaparas, T. Palpanas, Y. Kotidis, N. Koudas, and Divesh Srivastava. 2003. Ranked Join Indices. In *ICDE*.
- [50] Patrick Valduriez. 1987. Join indices. *TODS* 12, 2 (1987).
- [51] Konstantinos Xirogiannopoulos, Virinchi Srinivas, and Amol Deshpande. 2017. GraphGen: Adaptive graph processing using relational databases. In *GRADES*.
- [52] Jianqiao Zhu, Navneet Potti, Saket Saurabh, and Jignesh M Patel. 2017. Looking ahead makes query plans robust: Making the initial case with in-memory star schema data warehouse workloads. *PVLDB* 10, 8 (2017).

A SNB-M QUERIES

For completeness, we include all modified LDBC SNB queries used in our evaluation.

1. IS1

```
SELECT p.firstname, p.lastname, p.birthday, p.locationip,
       p.browserused, pl.placeid, p.gender, p.creationdate
FROM person p, place pl
WHERE person.id=933 AND person.placeid=place.placeid;
```

2. IS2

```
SELECT m1.id, m1.creationdate, m2.id, p2.personid,
       p2.firstname, p2.lastname
FROM person p1, comment m1, post m2, person p2
WHERE m2.creatorid=p2.personid AND m1.replyof_post=m2.id
      AND m1.creatorid=p1.personid AND p1.id=933;
```

3. IS3

```
SELECT p2.personid, p2.firstname, p2.lastname, k.creationdate
FROM knows k, person p1, person p2
WHERE p1.id=933 AND p1.personid=k.person1id
      AND k.person2id=p2.personid;
```

4. IS4

```
SELECT c.content, c.creationdate
FROM comment c
WHERE id=4947802324993;
```

5. IS5

```
SELECT p.personid, p.firstname, p.lastname
FROM comment c, person p
WHERE c.id=4947802324993 AND c.creatorid=p.personid;
```

6. IS6

```
SELECT f.forumid, f.title, p.personid, p.firstname, p.lastname
FROM comment m1, post m2, person p, forum f
WHERE m1.id=4947802324993 AND m1.replyof_post=m2.id
      and m2.forumid=f.forumid AND f.moderatorid=p.personid;
```

7. IS7

```
SELECT m2.id, m2.content, m2.creationdate, p.personid,
      p.firstname, p.lastname
FROM comment m1, comment m2, person p
WHERE m1.id=8246337208329 AND m2.replyof_comment=m1.id
      AND m2.creatorid=p.personid;
```

8. IC1-1

```
SELECT p2.id, p2.lastname, p2.birthday, p2.creationdate,
      p2.gender, p2.browserused, p2.locationip, p1.name
FROM person p1, knows k, person p2, place p1
WHERE p1.personid=k.person1id AND k.person2id=p2.personid
      AND p2.placeid=p1.placeid AND p1.id=933 AND
      p2.firstname='Rahul';
```

9. IC1-2

```
SELECT p2.id, p2.lastname, p2.birthday, p2.creationdate,
      p2.gender, p2.browserused, p2.locationip, p1.name
FROM person p1, knows k1, knows k2, person p2, place p1
WHERE p1.id=933 and p2.firstname='Rahul'
      AND p1.personid=k1.person1id AND k1.person2id=k2.person1id
      AND k2.person2id=p2.personid AND p2.placeid=p1.placeid;
```

10. IC1-3

```
SELECT p2.id, p2.lastname, p2.birthday, p2.creationdate,
      p2.gender, p2.browserused, p2.locationip, p1.name
FROM person p1, knows k1, knows k2, knows k3,
      person p2, place p1
WHERE p1.id=933 and p2.firstname='Rahul'
      AND p1.personid=k1.person1id AND k1.person2id=k2.person1id
      AND k2.person2id=k3.person1id AND k3.person2id=p2.personid
      AND p2.placeid=p1.placeid;
```

11. IC2

```
SELECT p2.id, p2.firstname, p2.lastname, c.id,
      c.content, c.creationdate
FROM person p1, knows k, person p2, comment c
WHERE p2.personid=c.creatorid AND c_creationdate<1338552000
      AND k.person2id=p2.personid AND p1.personid=k.person1id
      AND p1.id=933;
```

12. IC3-1

```
SELECT p2.id, p2.firstname, p2.lastname
FROM person p1, knows k1, person p2, comment m1,
      place p11, comment m2, place p12
WHERE m1.creationdate>=1313591219 AND m1.creationdate<1513591219
      AND m2.creationdate>=1313591219 AND m2.creationdate<1513591219
      AND p1.personid=k1.person1id AND k1.person2id=p2.personid
```

```
AND m2.creatorid=p2.personid AND m1.locationid=p11.placeid
AND m1.creatorid=p2.personid AND m2.locationid=p12.placeid
AND p1.id=933 AND p11.name='India' AND p12.name='China';
```

13. IC3-2

```
SELECT p2.id, p2.firstname, p2.lastname
FROM person p1, knows k1, knows k2, person p2, comment m1,
      place p11, comment m2, place p12
WHERE m2.creationdate>=1313591219 AND m2.creationdate<1513591219
      AND m1.creationdate>=1313591219 AND m1.creationdate<1513591219
      AND p1.id=933 AND p11.name='India' AND p12.name='China'
      AND p1.personid=k1.person1id AND k2.person1id=k1.person2id
      AND k2.person2id=p2.personid AND m2.creatorid=p2.personid
      AND m1.locationid=p11.placeid AND m1.creatorid=p2.personid
      AND m2.locationid=p12.placeid;
```

14. IC4

```
SELECT t_name
FROM knows k1, person p1, knows k2, person p2, post ps,
      post_tag mt, tag t
WHERE mt.tagid=t.tagid AND ps.id=mt.messageid AND
      p2.personid=ps.creatorid AND k2.person2id=p2.personid
      AND p1.personid=k2.person1id AND p1.personid=k1.person1id
      AND p1.id=933 AND ps.creationdate>=1313591219
      AND ps.creationdate<1513591219;
```

15. IC5-1

```
SELECT f.title
FROM person p1, knows k1, person p2, forum_person fp, forum f,
      post m
WHERE f.forumid=m.forumid AND fp.forumid=f.forumid
      AND p2.personid=fp.personid AND k1.person2id=p2.personid
      AND p1.personid=k1.person1id AND p1.id=933
      AND fp.joindate>=1353819600;
```

16. IC5-2

```
SELECT f.f_title
FROM person p1, knows k1, knows k2, person p2, forum_person fp,
      forum f, post m
WHERE f.forumid=m.forumid AND fp.forumid=f.forumid
      AND p2.personid=fp.personid AND k2.person2id=p2.personid
      AND k1.person2id=k2.person1id
      AND p1.personid=k1.person1id AND p1.id=933
      AND fp.joindate>=1353819600;
```

17. IC6-1

```
SELECT t2.t_name
FROM person p1, knows k1, person p2, post m, post_tag mt1,
      tag t1, post_tag mt2, tag t2
WHERE mt1.tagid=t1.tagid AND m.id=mt1.messageid AND
      mt2.tagid=t2.tagid AND m.id=mt2.messageid AND
      m.creatorid=p2.personid AND k1.person2id=p2.personid
      AND p1.personid=k1.person1id AND p1.id=933 AND
      t1.t_name='Rumi' AND t2.t_name!='Rumi';
```

18. IC6-2

```
SELECT t2.t_name
FROM person p1, knows k1, knows k2, person p2, post m,
      post_tag mt1, tag t1, post_tag mt2, tag t2
WHERE mt1.tagid=t1.tagid AND m.id=mt1.messageid
```

```

AND mt2.tagid=t2.tagid AND m.id=mt2.messageid
AND m.creatorid=p2.personid AND k2.person2id=p2.personid
AND k1.person2id=k2.person1id AND p1.personid=k1.person1id
AND p1.id=933 AND t1.t_name='Rumi' AND t2.t_name!='Rumi';

```

19. IC7

```

SELECT p2.personid, p2.firstname, p2.lastname,
       l.creationdate, c.content
FROM person p1, comment c, likes_comment l, person p2
WHERE p2.personid=l.personid AND c.id=l.messageid
      AND c.creatorid=p1.personid AND p1.id=933;

```

20. IC8

```

SELECT c.creatorid, p2.firstname, p2.lastname, c.creationdate,
       c.id, c.content
FROM person p1, post ps, comment c, person p2
WHERE c.creatorid=p2.personid AND c.replyof_post=ps.id AND
      p1.personid=ps.creatorid AND p1.personid=933;

```

21. IC9-1

```

SELECT p2.firstname, p2.lastname, c.creationdate
FROM person p1, knows k1, person p2, comment c
WHERE p2.personid=c.creatorid AND k2.person2id=p2.personid
      AND p1.personid=k1.person1id AND p1.id=933
      AND c.creationdate<1342840042;

```

22. IC9-2

```

SELECT p2.firstname, p2.lastname, c.creationdate
FROM person p1, knows k1, knows k2, person p2, comment c
WHERE p2.personid=c.creatorid AND k2.person2id=p2.personid
      AND k1.person2id=k2.person1id AND p1.personid=k1.person1id
      AND p1.id=933 AND c.creationdate<1342840042;

```

23. IC11-1

```

SELECT p2.id, p2.firstname, p2.lastname, o.name, pc.workfrom
FROM person p1, knows k1, person p2, person_company pc,
     organisation o, place pl
WHERE o.placeid=p1.placeid AND pc.organisationid=o.organisationid
      AND p2.personid=pc.personid AND k1.person2id=p2.personid
      AND p1.personid=k1.person1id AND p1.id=933
      AND pc.workfrom<2016 AND pl.name='China';

```

24. IC11-2

```

SELECT p2.id, p2.firstname, p2.lastname, o.name, pc.workfrom
FROM person p1, knows k1, knows k2, person p2,
     person_company pc, organisation o, place pl
WHERE p1.id=933 AND pc.workfrom<2016 AND pl.name='China'
      AND p2.personid=pc.personid AND k1.person2id=k2.person1id
      AND p1.personid=k1.person1id AND k2.person2id=p2.personid
      AND o.placeid=p1.placeid AND pc.organisationid=o.organisationid;

```

25. IC12

```

SELECT f.personid, friend.p_firstname, friend.p_lastname
FROM person p1, knows k, person f, comment c, post ps,
     post_tag pt, tag t, tagclass tc1, tagclass tc2
WHERE tc1.subclassoftagclassid=tc2.tagclassid
      AND t.tagclassid=tc1.tagclassid AND mt.tagid=tag.tagid
      AND c.replyof_post=ps.id AND c.creatorid=f.personid
      AND ps.id=mt.messageid AND k.person2id=f.personid AND p1.id=933
      AND p1.personid=k.person1id AND tc2.tc_name='Person';

```

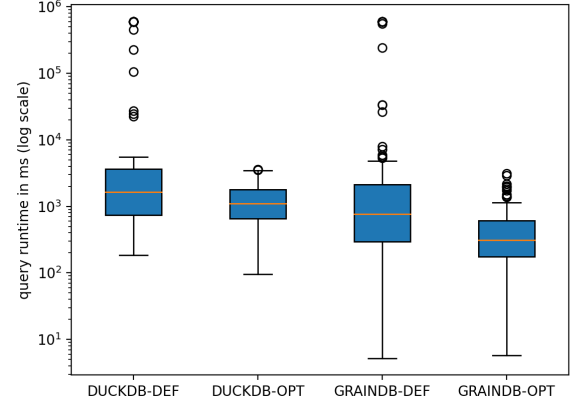


Figure 7: The query runtime (ms) of DuckDB and GRainDB on JOB with the system-default (DUCKDB-DEF, GRAINDB-DEF) and optimized (DUCKDB-OPT, GRAINDB-OPT) join orders. The boxplot shows the 5th, 25th, 50th, 75th, and 95th percentiles.

B DUCKDB AND GRAINDB WITH DEFAULT AND OPTIMIZED JOIN ORDERS

To isolate the influence of join order selection, instead of using DuckDB’s default join orders, we injected true cardinalities into the system to generate optimized join orders. Figure 7 shows the boxplots of running DuckDB on JOB with default and optimized join orders. We see with optimized join orders, the system reduces outlier runtimes largely, and improves the query performance in general. For 113 queries in JOB, DuckDB performs timeouts on 15 of them under default join orders, while 0 under optimized join orders. And the 25th percentile, median, and 75th percentile query execution times reduce respectively from 741.0ms to 652.4ms (1.4x), from 1656.4ms to 1110ms (1.5x), and from 3624.6 to 1797.0ms (2.0x). Recall that in GRainDB, we directly apply the same join order as DuckDB. And the performance improvement due to optimized join orders applies on GRainDB uniformly. The number of timeout queries in GRainDB reduces from 9 to 0 after replacing default join orders with optimized ones. And the 25th percentile, median, and 75th percentile query execution times reduce respectively from 2125.0ms to 614.2ms (3.5x), from 770.0ms to 309.0ms (2.5x), and from 295.6 to 176.4ms (1.7x). Moreover, noticeably in either default or optimized join orders, GRainDB shows better performance over DuckDB.

C QUERY EXECUTION TIME OF DUCKDB AND GRAINDB ON JOB

In Section 7.2.1, we presented box plots and detailed percentiles for query execution times of DuckDB and GRainDB on all JOB queries, and also execution times of the first variant of each query. In Table 8, we show detailed query execution time of DuckDB and GRainDB on all 113 queries on JOB.

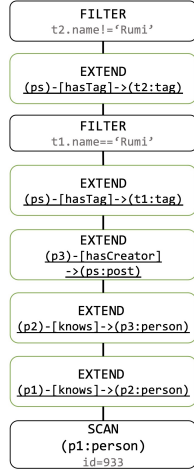


Figure 8: The Left-deep Plan in GraphflowDB. The final Projection is omitted.

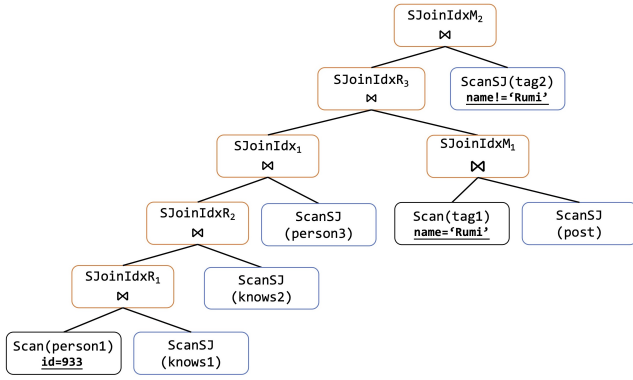


Figure 9: The Bushy Plan in GRainDB. The final Projection is omitted.

D BUSHY AND LEFT-DEEP PLANS FOR SNB-M IC6-2

Figure 8 presents the left-deep plan in GraphflowDB. The EXTEND operator is the join operator in GraphflowDB, which takes as input k tuples and extends each tuple t to one or more matches from adjacency list indexes. The join algorithm of EXTEND is essentially index nested-loop join. In this left-deep plan, GraphflowDB starts with scanning and filtering *Person* on $id = 933$, and extends to $p2$ and $p3$ through *knows* relationship, then extends to *post* created by $p3$, and further extends to all tags of *post* (*tag1* and *tag2*) and apply filters on tag names. Figure 9 demonstrates the bushy plan in GRainDB, in which hash joins are replaced by SJoin variants, and $SJoinIdxM_1$ and $SJoinIdxM_2$ merges two consecutive joins tag JOIN *post_tag* and post JOIN *post_tag* into one, respectively. Compared to the left-deep plan, the bushy plan takes advantage of both selective filters *person1.id = 933* and *tag1.name = 'Rumi'* to reduce intermediate result size.

E QUERY EXECUTION TIME OF DUCKDB AND GRAINDB ON TPC-H

In Section 7.2.3 we presented box plots of DuckDB and GRainDB on TPC-H, and showed no significant speedups or slowdowns on this benchmark. In Table 9, we list detailed query execution time of DuckDB and GRainDB on all 22 queries on TPC-H. In all queries, the biggest slowdown is from 6863.8ms to 8217.8ms (0.8x) on Q7. Though not expected, we still found performance speedups on Q2 and Q3. On Q2, we reduce runtime from 1324.0ms to 519.0s (2.6x), and on Q3, from 2757.0ms to 1515.0ms.

F QUERY EXECUTION TIME FOR ABLATION TESTS ON SNB-M

Section 7.3.1 showed the settings and box plots of our ablation study. In Table 10, we present detailed query execution time for each query in SNB-M under different system configurations, including GR-FULL, GR-JM, GR-JM-RSJ, and DuckDB (all optimizations turned off).

G PLAN SPECTRUM

Figure 10 shows the result of plan spectrum analysis over the second variant of Q1-Q6 in JOB. Dashed and straight lines in the figure are the distributions of P_d and P_d^* plans, respectively, which show the number of plans (on y-axis) for different runtime value cutoffs, e.g., 100ms, 200ms, ..., 1000ms (on x-axis).

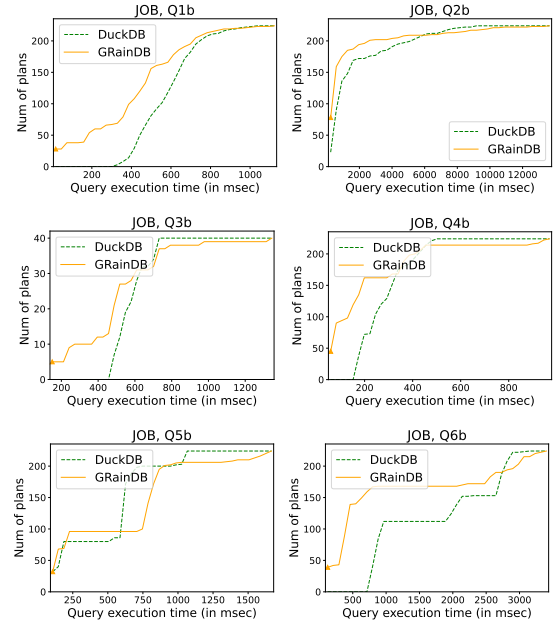


Figure 10: Cumulative distributions of the number of DuckDB and GRainDB plans (y-axis) that have runtimes below different thresholds (x-axis) on JOB Q1b-Q6b.

	Q1a	Q1b	Q1c	Q1d	Q2a	Q2b	Q2c	Q2d	Q3a	Q3b	Q3c	Q4a	Q4b	Q4c	Q5a	Q5b	Q5c	Q6a	Q6b
DuckDB	234.2	331.8	196.8	327.6	207.0	208.2	156.0	239.0	1491.4	491.8	1551.8	216.0	193.8	216.8	96.0	177.0	1628.4	885.4	878.4
GRainDB	34.2	3.0	13.8	6.0	154.0	143.0	17.2	227.2	328.0	502.0	473.6	114.2	73.6	144.4	116.4	146.0	425.4	57.0	97.0
	6.8x	110.6x	14.3x	54.6x	1.3x	1.5x	9.1x	1.1x	4.5x	1.0x	3.3x	1.9x	2.6x	1.5x	0.8x	1.2x	3.8x	15.5x	9.1x
	Q6c	Q6d	Q6e	Q6f	Q7a	Q7b	Q7c	Q8a	Q8b	Q8c	Q8d	Q9a	Q9b	Q9c	Q9d	Q10a	Q10b	Q10c	Q11a
DuckDB	857.2	863.4	849.4	1601.0	879.8	764.4	1383.6	1307.4	1267.4	1707.6	1097.6	1933.2	1538.4	1906.8	3538.4	1404.2	1248.4	1396.4	264.2
GRainDB	416.8	430.6	419.0	1727.6	203.4	166.8	864.6	349.8	208.6	1944.6	973.2	554.8	494.2	881.4	3104.4	614.2	526.2	1436.8	44.4
	2.1x	2.0x	2.0x	0.9x	4.3x	4.6x	1.6x	3.7x	6.1x	0.9x	1.1x	3.5x	3.1x	2.2x	1.1x	2.3x	2.4x	1.0x	6.0x
	Q11b	Q11c	Q11d	Q12a	Q12b	Q12c	Q13a	Q13b	Q13c	Q13d	Q14a	Q14b	Q14c	Q15a	Q15b	Q15c	Q15d	Q16a	Q16b
DuckDB	220.0	307.2	307.6	684.0	643.2	859.0	981.2	398.2	326.8	679.8	1644.2	1683.4	1695.2	987.8	984.4	903.0	612.4	652.4	2316.0
GRainDB	97.8	254.4	142.6	242.4	16.0	415.6	336.8	208.0	145.6	614.6	202.2	158.8	270.8	283.4	87.6	317.4	309.0	177.6	2025.6
	2.2x	1.2x	2.2x	2.8x	40.2x	2.1x	2.9x	1.9x	2.2x	1.1x	8.1x	10.6x	6.3x	3.5x	11.2x	2.8x	2.0x	3.7x	1.1x
	Q16c	Q16d	Q17a	Q17b	Q17c	Q17d	Q17e	Q17f	Q18a	Q18b	Q18c	Q19a	Q19b	Q19c	Q19d	Q20a	Q20b	Q20c	Q21a
DuckDB	1215.4	1110.0	1164.8	704.4	595.0	726.4	2134.2	1400.4	1797.0	2433.0	3565.8	2632.4	2395.4	2913.8	2979.0	1118.4	1399.2	1102.0	1629.2
GRainDB	1378.8	1026.8	486.0	309.2	206.4	325.6	2921.4	978.0	612.6	525.2	1952.2	419.6	226.0	857.0	2159.4	1071.6	993.6	1065.0	54.2
	0.9x	1.1x	2.4x	2.3x	2.9x	2.2x	0.7x	1.4x	2.9x	4.6x	1.8x	5.4x	10.6x	3.4x	1.4x	1.0x	1.4x	1.0x	30.1x
	Q21b	Q21c	Q22a	Q22b	Q22c	Q22d	Q23a	Q23b	Q23c	Q24a	Q24b	Q25a	Q25b	Q25c	Q26a	Q26b	Q26c	Q27a	Q27b
DuckDB	718.6	1653.0	1471.8	1425.0	1881.8	1710.2	866.6	838.6	859.6	2554.8	2551.4	2318.8	2138.0	3584.6	1074.6	816.4	1142.6	761.4	717.6
GRainDB	52.2	49.2	864.0	433.0	387.0	412.6	296.8	100.6	280.0	788.8	297.4	1376.6	176.4	1834.4	733.4	424.0	1145.0	44.4	36.6
	13.8x	33.6x	1.7x	3.3x	4.9x	4.1x	2.9x	8.3x	3.1x	3.2x	8.6x	1.7x	12.1x	2.0x	1.5x	1.9x	1.0x	17.1x	19.6x
	Q27c	Q28a	Q28b	Q28c	Q29a	Q29b	Q29c	Q30a	Q30b	Q30c	Q31a	Q31b	Q31c	Q32a	Q32b	Q32c	Q33a	Q33b	
DuckDB	1595.4	2068.4	993.0	1882.8	2742.6	2205.6	2506.0	2198.0	2287.6	3189.0	2523.8	2432.6	3457.8	126.0	275.2	350.0	336.2	574.6	
GRainDB	46.6	240.2	199.4	264.6	266.6	227.0	259.0	673.0	233.8	1501.2	612.0	296.0	817.8	8.2	173.6	178.6	149.4	223.8	
	34.2x	8.6x	5.0x	7.1x	10.3x	9.7x	9.7x	3.3x	9.8x	2.1x	4.1x	8.2x	4.2x	15.4x	1.6x	2.0x	2.3x	2.6x	

Table 8: Runtime (in ms) of DuckDB and GRainDB on all 113 queries in JOB.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22
DuckDB	8830.2	1324.0	2757.0	2817.2	3616.4	742.8	6863.8	2347.5	8633.0	6531.8	386.3	2639.8	6365.0	836.5	1856.6	1624.8	4684.8	13063.0	7279.6	1802.0	11446.6	1265.8
GRainDB	9450.6	519.0	1515.0	2830.8	2881.8	683.8	8217.8	1598.5	7492.5	4158.0	344.0	3247.8	6736.8	867.0	1832.8	1798.4	4219.4	13296.4	8418.0	1817.2	10236.8	1233.4
	0.9x	2.6x	1.8x	1.0x	1.3x	1.1x	0.8x	1.5x	1.2x	1.6x	1.1x	0.8x	0.9x	1.0x	1.0x	0.9x	1.1x	1.0x	0.9x	1.0x	1.1x	1.0x

Table 9: Runtime (in ms) of DuckDB and GRainDB on all 22 queries in TPC-H.

	IS1	IS2	IS3	IS4	IS5	IS6	IS7	IC1-1	IC1-2	IC1-3	IC2	IC3-1	IC3-2	IC4	IC5-1	IC5-2	IC6-1	IC6-2	IC7	IC8	IC9-1	IC9-2	IC11-1	IC11-2	IC12
DuckDB	0.8	524.8	36.6	0.2	4.5	148.0	989.0	38.0	72.0	110.5	926.0	1177.8	4647.0	402.0	636.0	3125.0	244.6	471.2	1186.8	1017.0	441.8	1312.6	35.8	68.4	788.4
GrainDB	1.2	19.6	3.4	0.2	0.6	5.0	11.0	4.0	6.4	38.2	134.8	119.4	1665.0	54.0	174.0	2768.0	13.0	22.0	33.2	14.0	113.6	752.0	2.8	9.0	234.8
	0.7x	26.8x	10.8x	1.0x	7.5x	29.6x	90.0x	9.5x	11.2x	2.9x	6.9x	9.9x	2.8x	7.4x	3.7x	1.1x	18.8x	21.4x	35.7x	72.6x	3.9x	1.8x	12.8x	7.6x	3.4x
GrainDB-JM	1.1	19	3.2	0.2	0.6	5.4	10.0	6.0	25.0	68.8	349.2	119.0	1706.3	54.0	198.2	3082.2	13.6	61.6	29.0	15.4	113.6	686.0	3.8	38.6	253.0
	0.7x	27.6x	11.4x	1.0x	7.5x	27.4x	98.9x	6.3x	2.9x	1.6x	2.7x	9.9x	2.7x	7.4x	3.2x	1.0x	18.0x	7.6x	40.9x	66.0x	3.9x	1.9x	9.4x	1.8x	3.1x
GrainDB-JM-RSJ	0.8	407.8	34.6	0.2	0.6	5.0	1020.8	33.2	62.8	107.0	1004.0	1195.0	2122.4	345.0	620.0	3195.2	244.2	299.2	1185.8	1016.0	430.2	804.4	34.2	68.0	640.2
	1.0x	1.3x	1.1x	1.0x	7.5x	29.6x	1.0x	1.1x	1.1x	1.0x	0.9x	1.0x	2.2x	1.2x	1.0x	1.0x	1.0x	1.6x	1.0x	1.0x	1.6x	1.0x	1.0x	1.2x	

Table 10: Runtime (in ms) of GRainDB on each query in SNB-M under different optimizations. DuckDB implements no optimizations. GRainDB-JM-RSJ only implements RID materialization. GRainDB-JM in addition implements reverse semijoins. GRainDB in addition implements join merging.