

1 TB TPC-DS Performance comparison (Hive Parquet): GCP, 2 workers n2-custom (14 vCPU / 112 GB) + 1 driver

TabbyDB 4.1.3-tabbydb.2 vs Apache Spark 4.1.3

Summary

Table format	Stock Spark total (s)	TabbyDB total (s)	Less total time	Geo-mean speedup	Queries faster on TabbyDB
Hive Parquet	7,884.7	6,253.7	20.7%	1.49x	88 of 103
Iceberg Parquet	9,034.4	7,066.7	21.8%	1.40x	85 of 103

Benchmark setup

Cloud / region	Google Cloud, us-west4-a
Driver node	1 x e2-highmem-4 (4 vCPU, 32 GB) - Spark master + driver (24 GB heap)
Worker nodes	2 x n2-custom-14-114688 (14 vCPU, 112 GB RAM), 2 x 375 GB local NVMe SSD each for shuffle/spill
Executors	4 executors (2 per worker) x 7 cores x 40 GB heap (+1 GB overhead) = 28 cores, 160 GB
OS / JVM	Ubuntu 24.04 LTS (amd64), OpenJDK 21
Engines	Apache Spark 4.1.3 (stock) vs TabbyDB 4.1.3-tabbydb.2
Iceberg	Stock: iceberg-spark-runtime-4.1_2.13-1.11.0 TabbyDB: iceberg-tabbydb-runtime-4.1.3_2.13-1.11.0-tabbydb.2
Data	TPC-DS scale factor 1000 (1 TB), Parquet on Google Cloud Storage, non-partitioned, date-sorted on locally written partition; DECIMAL generated as DOUBLE
Table formats	Hive external tables (with ANALYZE statistics) and Iceberg tables over the same data
Workload	spark-sql-perf, TPC-DS v2.4 queries: 103 queries incl. ss_max (q84 excluded); 1 iteration; OS page caches dropped on workers before each run
Common Spark settings	autoBroadcastJoinThreshold=100MB, shuffle.partitions=448, preferSortMergeJoin=false, maxShuffledHashJoinLocalMapThreshold=10MB
TabbyDB-only settings	See tabbydb-spark-defaults.conf (e.g. autoBroadcastJoinThresholdInnerJoin=125MB, BHJ key pushdown flags)
Timing	Per-query time = compile (parse/analyze/optimize/plan) + execution, as reported by spark-sql-perf

Result correctness

Every query's full result set was compared with stock Spark on Hive. On both formats 93 queries match exactly and 8 match as unordered sets (ORDER BY ties). q65 and q67 differ in values; they also differ between stock Spark on Iceberg and stock Spark on Hive, because of ORDER BY ties and DOUBLE rounding. Row counts match for all 103 queries.

Per-query results (Hive Parquet)

Query	Stock Spark (s)	Result rows Stock Spark	TabbyDB (s)	Result rows TabbyDB	Time diff (s)	Improvement	Speedup	Row count match	Result check
q1	32.54	100	19.58	100	-12.96	39.8%	1.66x	TRUE	MATCH
q2	51.83	2513	33.38	2513	-18.45	35.6%	1.55x	TRUE	MATCH
q3	22.06	100	16.45	100	-5.61	25.4%	1.34x	TRUE	MATCH
q4	340.10	100	249.85	100	-90.25	26.5%	1.36x	TRUE	MATCH
q5	126.64	100	70.45	100	-56.19	44.4%	1.80x	TRUE	MATCH
q6	23.56	52	13.60	52	-9.96	42.3%	1.73x	TRUE	MATCH
q7	40.05	100	20.55	100	-19.50	48.7%	1.95x	TRUE	MATCH
q8	29.33	10	16.74	10	-12.59	42.9%	1.75x	TRUE	MATCH
q9	114.49	1	111.90	1	-2.59	2.3%	1.02x	TRUE	MATCH
q10	36.61	100	25.83	100	-10.78	29.4%	1.42x	TRUE	MATCH
q11	187.93	100	152.32	100	-35.61	18.9%	1.23x	TRUE	MATCH
q12	11.33	100	7.98	100	-3.35	29.5%	1.42x	TRUE	MATCH
q13	54.47	1	32.94	1	-21.53	39.5%	1.65x	TRUE	MATCH
q14a	253.50	100	224.10	100	-29.40	11.6%	1.13x	TRUE	MATCH
q14b	228.53	100	196.62	100	-31.91	14.0%	1.16x	TRUE	MATCH
q15	51.47	100	20.69	100	-30.78	59.8%	2.49x	TRUE	MATCH
q16	104.72	1	81.77	1	-22.95	21.9%	1.28x	TRUE	MATCH
q17	66.17	100	38.12	100	-28.05	42.4%	1.74x	TRUE	MATCH
q18	47.53	100	22.04	100	-25.49	53.6%	2.16x	TRUE	MATCH(unordered)
q19	41.71	100	22.09	100	-19.62	47.0%	1.89x	TRUE	MATCH
q20	17.03	100	5.86	100	-11.17	65.6%	2.91x	TRUE	MATCH
q21	8.60	100	3.18	100	-5.42	63.0%	2.70x	TRUE	MATCH
q22	33.38	100	31.03	100	-2.35	7.0%	1.08x	TRUE	MATCH
q23a	338.35	1	315.56	1	-22.79	6.7%	1.07x	TRUE	MATCH
q23b	346.72	100	324.43	100	-22.29	6.4%	1.07x	TRUE	MATCH
q24a	175.38	2919	176.14	2919	+0.76	-0.4%	1.00x	TRUE	MATCH(unordered)

Query	Stock Spark (s)	Result rows Stock Spark	TabbyDB (s)	Result rows TabbyDB	Time diff (s)	Improvement	Speedup	Row count match	Result check
q24b	170.02	322	170.68	322	+0.66	-0.4%	1.00x	TRUE	MATCH(unordered)
q25	64.25	100	30.76	100	-33.49	52.1%	2.09x	TRUE	MATCH
q26	26.58	100	11.91	100	-14.67	55.2%	2.23x	TRUE	MATCH
q27	36.08	100	17.52	100	-18.56	51.5%	2.06x	TRUE	MATCH
q28	131.29	1	132.45	1	+1.16	-0.9%	0.99x	TRUE	MATCH
q29	87.52	100	67.91	100	-19.61	22.4%	1.29x	TRUE	MATCH
q30	23.05	100	20.48	100	-2.57	11.1%	1.13x	TRUE	MATCH
q31	87.75	278	45.96	278	-41.79	47.6%	1.91x	TRUE	MATCH
q32	31.28	1	9.35	1	-21.93	70.1%	3.34x	TRUE	MATCH
q33	52.56	100	21.63	100	-30.93	58.8%	2.43x	TRUE	MATCH
q34	26.27	8489	24.20	8489	-2.07	7.9%	1.09x	TRUE	MATCH
q35	50.13	100	40.41	100	-9.72	19.4%	1.24x	TRUE	MATCH
q36	32.89	100	16.62	100	-16.27	49.5%	1.98x	TRUE	MATCH
q37	17.28	12	13.59	12	-3.69	21.4%	1.27x	TRUE	MATCH
q38	101.92	1	97.88	1	-4.04	4.0%	1.04x	TRUE	MATCH
q39a	15.10	14551	8.29	14551	-6.81	45.1%	1.82x	TRUE	MATCH
q39b	14.77	594	7.84	594	-6.93	46.9%	1.88x	TRUE	MATCH
q40	28.16	100	9.92	100	-18.24	64.8%	2.84x	TRUE	MATCH
q41	1.18	100	1.14	100	-0.04	3.5%	1.04x	TRUE	MATCH
q42	19.16	13	6.86	13	-12.30	64.2%	2.79x	TRUE	MATCH
q43	20.36	100	12.87	100	-7.49	36.8%	1.58x	TRUE	MATCH
q44	40.14	10	41.08	10	+0.94	-2.3%	0.98x	TRUE	MATCH
q45	25.70	100	24.95	100	-0.75	2.9%	1.03x	TRUE	MATCH
q46	46.69	100	46.59	100	-0.10	0.2%	1.00x	TRUE	MATCH
q47	56.36	100	46.71	100	-9.65	17.1%	1.21x	TRUE	MATCH
q48	39.71	1	19.68	1	-20.03	50.4%	2.02x	TRUE	MATCH
q49	89.08	100	37.96	100	-51.12	57.4%	2.35x	TRUE	MATCH(unordered)

Query	Stock Spark (s)	Result rows Stock Spark	TabbyDB (s)	Result rows TabbyDB	Time diff (s)	Improvement	Speedup	Row count match	Result check
q50	148.83	100	176.40	100	+27.57	-18.5%	0.84x	TRUE	MATCH
q51	98.00	100	89.83	100	-8.17	8.3%	1.09x	TRUE	MATCH
q52	19.40	100	7.26	100	-12.14	62.6%	2.67x	TRUE	MATCH
q53	22.81	100	10.88	100	-11.93	52.3%	2.10x	TRUE	MATCH
q54	53.42	100	28.12	100	-25.30	47.4%	1.90x	TRUE	MATCH
q55	19.88	100	8.38	100	-11.50	57.9%	2.37x	TRUE	MATCH
q56	50.81	100	23.11	100	-27.70	54.5%	2.20x	TRUE	MATCH
q57	30.42	100	22.53	100	-7.89	25.9%	1.35x	TRUE	MATCH
q58	46.93	6	47.76	6	+0.83	-1.8%	0.98x	TRUE	MATCH
q59	43.99	100	31.86	100	-12.13	27.6%	1.38x	TRUE	MATCH
q60	50.30	100	22.53	100	-27.77	55.2%	2.23x	TRUE	MATCH
q61	67.80	1	26.72	1	-41.08	60.6%	2.54x	TRUE	MATCH
q62	12.53	100	11.53	100	-1.00	8.0%	1.09x	TRUE	MATCH
q63	22.15	100	10.26	100	-11.89	53.7%	2.16x	TRUE	MATCH
q64	225.22	12185	74.66	12185	-150.56	66.8%	3.02x	TRUE	MATCH(unordered)
q65	120.05	100	98.60	100	-21.45	17.9%	1.22x	TRUE	DIFF values (first at row 1)
q66	40.98	20	19.08	20	-21.90	53.4%	2.15x	TRUE	MATCH
q67	366.06	100	358.58	100	-7.48	2.0%	1.02x	TRUE	DIFF values (first at row 23)
q68	54.95	100	49.42	100	-5.53	10.1%	1.11x	TRUE	MATCH
q69	33.43	100	24.03	100	-9.40	28.1%	1.39x	TRUE	MATCH
q70	53.39	49	30.85	49	-22.54	42.2%	1.73x	TRUE	MATCH
q71	44.53	207916	19.29	207916	-25.24	56.7%	2.31x	TRUE	MATCH(unordered)
q72	136.38	100	125.48	100	-10.90	8.0%	1.09x	TRUE	MATCH
q73	26.00	253	23.44	253	-2.56	9.8%	1.11x	TRUE	MATCH(unordered)
q74	156.65	100	131.61	100	-25.04	16.0%	1.19x	TRUE	MATCH
q75	167.34	29	111.63	29	-55.71	33.3%	1.50x	TRUE	MATCH
q76	50.15	100	49.81	100	-0.34	0.7%	1.01x	TRUE	MATCH

Query	Stock Spark (s)	Result rows Stock Spark	TabbyDB (s)	Result rows TabbyDB	Time diff (s)	Improvement	Speedup	Row count match	Result check
q77	63.88	100	25.06	100	-38.82	60.8%	2.55x	TRUE	MATCH
q78	248.26	100	224.40	100	-23.86	9.6%	1.11x	TRUE	MATCH
q79	43.62	100	43.72	100	+0.10	-0.2%	1.00x	TRUE	MATCH(unordered)
q80	106.95	100	39.74	100	-67.21	62.8%	2.69x	TRUE	MATCH
q81	29.32	100	29.74	100	+0.42	-1.4%	0.99x	TRUE	MATCH
q82	24.51	19	21.08	19	-3.43	14.0%	1.16x	TRUE	MATCH
q83	12.59	100	12.90	100	+0.31	-2.4%	0.98x	TRUE	MATCH
q85	22.75	64	25.46	64	+2.71	-11.9%	0.89x	TRUE	MATCH
q86	13.54	100	11.11	100	-2.43	17.9%	1.22x	TRUE	MATCH
q87	105.58	1	93.68	1	-11.90	11.3%	1.13x	TRUE	MATCH
q88	130.63	1	146.79	1	+16.16	-12.4%	0.89x	TRUE	MATCH
q89	25.78	100	15.21	100	-10.57	41.0%	1.69x	TRUE	MATCH
q90	12.87	1	17.83	1	+4.96	-38.5%	0.72x	TRUE	MATCH
q91	29.05	42	5.87	42	-23.18	79.8%	4.95x	TRUE	MATCH
q92	17.05	1	10.43	1	-6.62	38.8%	1.63x	TRUE	MATCH
q93	166.19	100	166.58	100	+0.39	-0.2%	1.00x	TRUE	MATCH
q94	56.68	1	54.18	1	-2.50	4.4%	1.05x	TRUE	MATCH
q95	259.02	1	255.23	1	-3.79	1.5%	1.01x	TRUE	MATCH
q96	16.70	1	18.68	1	+1.98	-11.9%	0.89x	TRUE	MATCH
q97	125.65	1	113.79	1	-11.86	9.4%	1.10x	TRUE	MATCH
q98	25.01	44858	9.27	44858	-15.74	63.0%	2.70x	TRUE	MATCH
q99	23.58	100	16.12	100	-7.46	31.6%	1.46x	TRUE	MATCH
ss_max	43.74	1	44.72	1	+0.98	-2.2%	0.98x	TRUE	MATCH
Total	7,884.66		6,253.68		-1,630.98	20.7%	1.26x		