RAE

BENCHMARK REPORT

Raelog Reasoner

Speed, scale, correctness, reproducibility and verifiable evidence

5 November 2026

In brief

The Raelog Reasoner is the reasoning engine in Rae. It applies rules written in the logic language Raelog to the facts of a case and returns an answer that carries its certainty and a complete chain of reasoning.

We tested it on seven counts: speed as inference gets deeper, growth in the size of the model, growth in the volume of data, correctness against independently computed answers, reproducibility, verifiability of the evidence it gives, and behaviour under concurrent load. We also tested the operational behaviour a deployment depends on: saving and resuming sessions, asking for inputs as questions, and stopping runaway reasoning.

Every workload was generated, so every answer could be checked against a known result. All tests ran on a single machine. Unless stated otherwise, each timing is the median of 15 fresh sessions after a warm-up.

Table 1. Headline results

224 µs	to answer a question resting on 1.9×10^{15} derivation paths, across 32 levels of compounding inference
0	wrong answers out of 450,205, each checked against independently computed ground truth
100%	of 8,845 answers carry evidence that passes every proof check, and 100% of 5,970 deliberately corrupted evidence trees were caught
100%	identical answers, in identical order, across 181,896 repeated runs, whatever the session, input order, process, thread count or resume point
426.8 ms	to roll up every conclusion over 305,000 facts, and 50.5 ms over 15,001 rules
7,959/s	queries over HTTP with 256 concurrent clients, with no failed requests and 3,979 complete sessions a second

©2026 Rainbird Technologies Limited
Public

Summary of results

Table 2. Summary of results

Test	Measure	Result
Inference depth	query time, 32 compounding levels (1.9×10^{15} paths)	224 µs
	query time, 5,000-level rule chain	8.47 ms (1.69 µs per level)
	left-recursive closure, chain of 10,000	183.8 ms
Model complexity	rollup over 15,001 rules	50.5 ms
	one conclusion, any model size	29 µs to 40 µs
	compile 15,001 rules	138.3 ms
Data size	rollup over 305,000 facts	426.8 ms
	inject 305,000 facts	146.0 ms (2.1 million facts/s)
Correctness	answers checked against independent ground truth	450,205, 0 wrong
Reproducibility	repeated runs giving the identical answer, in identical order	100% of 181,896
	identical evidence, same inputs and questions	100% of 138,489
Evidence	answers whose evidence passes every proof check	100% (8,845 of 8,845)
	premises sufficient to re-derive the answer	100% (1,058 of 1,058)
	corrupted evidence detected	100% (5,970 of 5,970)
Concurrency	throughput at 256 concurrent clients, over HTTP	7,959 queries/s
	failed requests	0
	throughput, 1 to 16 cores	9.0×
Operations	resume a 305,000-fact session	1.81 s
	memory per live session (620 facts)	466.9 KB

HOW TO READ THIS REPORT

Each section sets out what we tested, the result and the evidence behind it.

Timing charts show the median at each measured point. The appendix lists every measured point.

©2026 Rainbird Technologies Limited
Public

Method

Most tests (depth, scale, correctness, evidence, reproducibility and operations) called the engine directly. Every timed sample used a new engine and a new session, so no sample benefited from another's work. Facts were injected first and the clock covered the query alone.

The concurrency test ran the engine as a separate server and drove it over HTTP, as a deployment would. Each client repeated a complete session: start, inject 620 facts, run two queries and fetch one evidence tree. Each load level ran for a fixed period after a warm-up, and the server restarted between levels.

Ground truth never came from the engine. Random models were evaluated by a separate, deliberately simple evaluator, reachability in graphs by breadth-first search, and expected certainties were computed independently. The remaining workloads have answers known by construction. Every workload was generated deterministically from a fixed seed, so each test can be repeated exactly.

Table 3. Workloads

Workload	Shape	What it tests
Compounding depth	each level counts a derived relationship of the level below; every item links to all three items of the next level	exponential derivation paths, aggregates over derived facts
Rule chain	a chain of distinct rules, each joining the level below and adding 1	long dependency chains
Recursion	transitive closure over a chain, in left- and right-recursive forms	recursive rules
Wide rule set	records with typed attributes; independent conclusions over the attribute values, three rules each; a rollup that collects every conclusion	model size, data size, string building
Random models	random weighted rules over random facts, stratified or recursive	joins, certainty, recursion
Reachability	random directed graphs with cycles	recursion through cyclic data
Certainty chains	weighted two-condition rules with rule certainties from 50 to 100	certainty arithmetic, minimum-certainty cut-off
Interactive	an eligibility decision whose inputs are asked as questions	question and answer loop
Runaway	a rule that derives the next number forever	resource limits

©2026 Rainbird Technologies Limited
Public

Speed and depth of inference

Deeper inference does not slow the engine down. A rule-based answer often rests on rules that rest on other rules. In the compounding test every item links to three items at the next level, so an answer d levels deep has 3^d derivation paths, and an evaluator that follows every path does work in proportion to 3^d.

At 32 levels there are 1.9×10^{15} derivation paths, and the engine answers in 224 µs. From one level to 32, the number of paths grows by a factor of 6.2×10^{14} and query time by a factor of 12. All 13 depth points returned the expected answer.

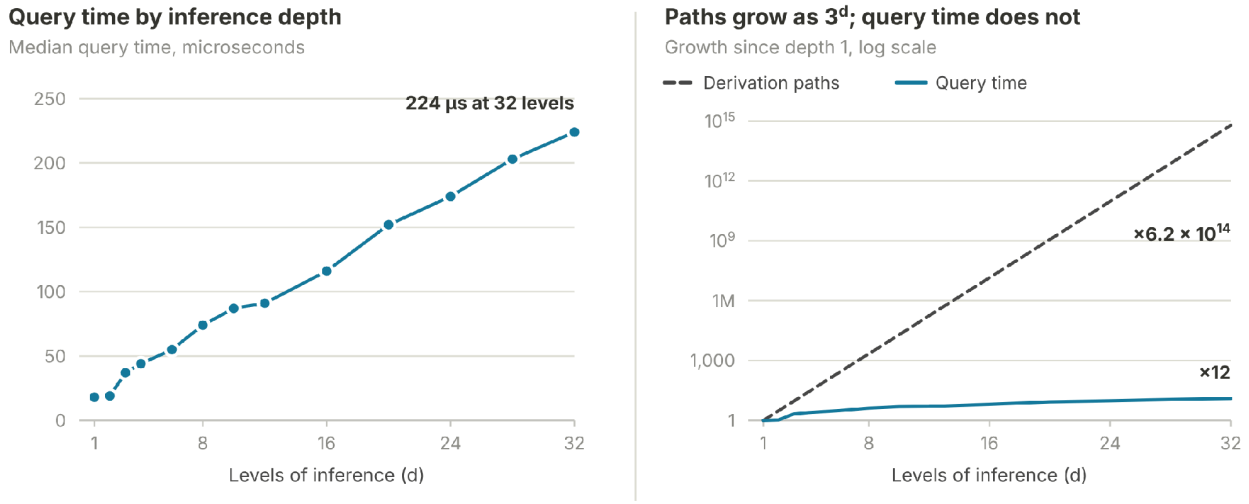


Fig. 1. From one level of inference to 32, derivation paths grow by a factor of 6.2×10^{14} while query time grows by a factor of 12.

THE FIT

Query time is linear in depth.

A least-squares fit of median query time over the 13 depths gives $t(d) \approx 14 + 6.66d$ µs, with $R^2 = 0.996$. Exponential growth would show as a straight line on a log scale; in Figure 1, query time stays flat beside it.

Rule chains. A chain of 5,000 distinct rules, each joining the level below and doing arithmetic, answers in 8.47 ms, or 1.69 µs per level of inference. Every answer was correct.

Recursion. The left-recursive closure of a 10,000-link chain takes 183.8 ms. The right-recursive form of the same question derives the closure of every node on the chain, about $n^2/2$ facts (2,001,000 at n = 2,000), in 1.10 s, which is 1.8 million derived facts a second. Every closure was exact.

©2026 Rainbird Technologies Limited
Public

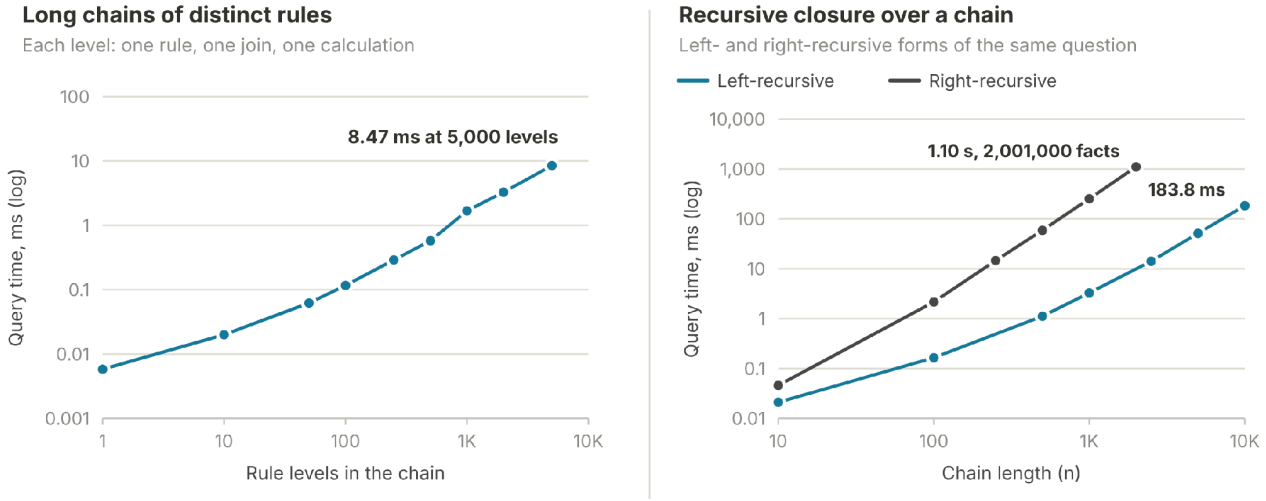


Fig. 2. Long rule chains cost about 1.7 µs per level of inference, and both recursive forms of the closure scale smoothly. Every answer was correct.

Growth of model complexity

A query pays only for the rules it uses. We grew the wide rule set from 10 to 5,000 independent conclusions, each derived by three rules, while holding the facts fixed. Three queries separate the effects of model size: one conclusion, whose work does not depend on the size of the model; a rollup that collects every conclusion, whose work grows with the model; and a lookup on a relationship with no rules, which shows the fixed overhead of a query.

One conclusion stays between 29 µs and 40 µs across the whole range, and the lookup stays below 12 µs. The rollup over 15,001 rules returns 5,875 results in 50.5 ms. The model compiles once, in 138.3 ms for 15,001 rules (9.2 µs per rule, 32.1 MB in memory), and every session then shares the compiled model. Every answer matched the expected results.

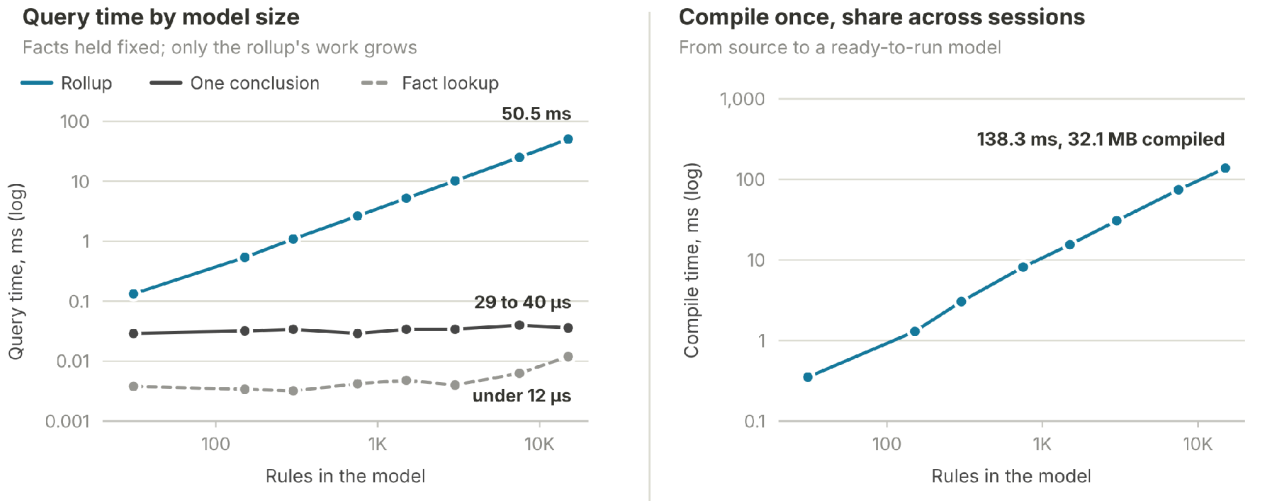


Fig. 3. One conclusion costs the same whatever the size of the model; only the rollup, which collects every conclusion, grows with it. The model compiles once and every session shares it.

©2026 Rainbird Technologies Limited
Public

Growth of data size

Query time grows in line with the data. The same rule set, with ten conclusions, ran over sessions of up to 305,000 facts, each record carrying 20 typed attributes.

The rollup over 305,000 facts takes 426.8 ms. Rollup time grows in proportion to the data, with a log-log slope of 1.08 against 1.0 for perfectly linear growth. Each fact costs about 1.40 µs, and the answer at 305,000 facts lists 43,500 results. Facts inject at 2.1 million a second (305,000 in 146.0 ms), and a session holds about 387 bytes per injected fact. Every rollup returned exactly the results computed from the data.

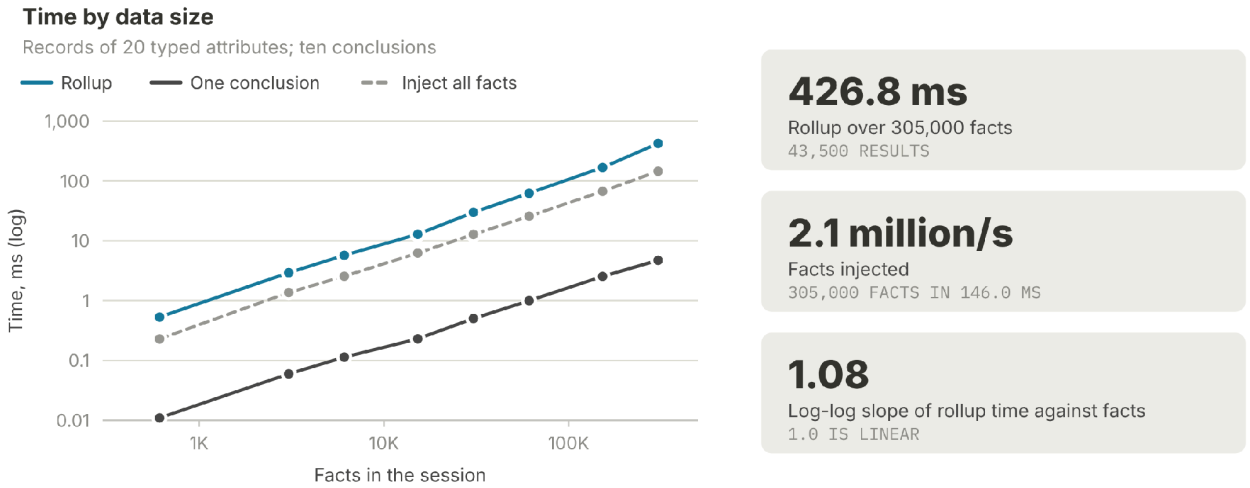


Fig. 4. Rollup time, single-conclusion time and injection time all grow in proportion to the data.

Correctness

Speed only matters if the answers are right, and the engine gave no wrong answers. Every answer below was compared with ground truth computed without the engine. A wrong answer means any missing fact, any extra fact, or any fact with a different certainty.

Table 4. Answers checked against independent ground truth

Workload	Size	Queries	Answers expected	Wrong
Random models, stratified	400 models, 5,600 rules	38,400	35,976	0
Random models, recursive	400 models, 5,600 rules	38,400	15,934	0
Reachability, cyclic graphs (both recursive forms)	5 graphs, up to 400 nodes	1,540	396,552	0
Certainty propagation chains	200 chains	1,743	1,743	0

Certainty. We evaluated 200 chains of rules with mixed condition weights and rule certainties, and compared every derived certainty with an independently computed value. All matched exactly, including 287 facts that fall below the minimum certainty and are correctly not derived.

©2026 Rainbird Technologies Limited
Public

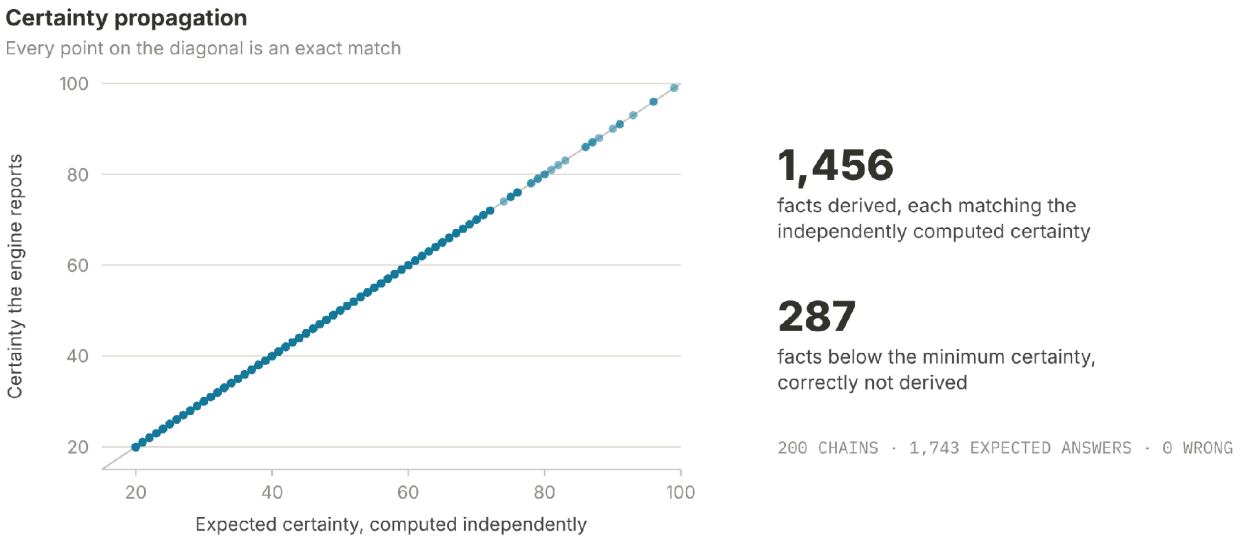


Fig. 5. Every derived certainty matched the independently computed value exactly.

Known answers. The performance tests also checked every answer: compounding depth (13 points), rule chains (9), recursive closure (13), model size (24) and data size (16). All were correct.

Reproducibility and stability

The same facts give the same answer, whatever else is going on. Each workload ran under nine variations, and each run was compared with a baseline on three counts: the answer (the facts and their certainties), the order of the results, and the evidence behind every result, compared node by node with session-specific identifiers removed.

Table 5. Same inputs and questions

Variation	Comparisons	Same answer	Same order	Same evidence
same session, repeated	20,670	100%	100%	100%
fresh sessions	41,340	100%	100%	100%
concurrent sessions	66,144	100%	100%	100%
resumed from a saved record	2,067	100%	100%	100%
separate processes	6,201	100%	100%	100%
process limited to a single thread	2,067	100%	100%	100%

Table 6. Inputs or questions in a different order

Variation	Comparisons	Same answer	Same order	Proof checks	Passed
inject order and batching	20,670	100%	100%	3,138,498	100%
query order	20,670	100%	100%	3,138,876	100%
resumed mid-session	2,067	100%	100%	296,844	100%

©2026 Rainbird Technologies Limited
Public

The answer and its order were identical in all 181,896 comparisons. With the same inputs and questions the evidence was identical too, in all 138,489 comparisons. When facts arrive in a different order or in batches, when questions come in a different order, or when a session is saved and resumed part-way, we require the evidence to be valid rather than identical, because a conclusion can have more than one proof. All 6,574,218 proof checks on those evidence trees passed.

Verifiability of evidence

Every answer carries evidence that can be checked independently. The evidence holds the rule that produced the answer, the variable bindings, each condition with the fact it matched or the expression it evaluated, and the evidence for those facts in turn, all the way down to stated facts. An independent checker walked each evidence tree in the public API format and checked every node against the model's source. It shares no code with the engine except the expression evaluator, which it uses to re-evaluate expressions.

Table 7. Proof checks

Proof check	What it confirms	Applied	Failed
`resolved`	every cited fact ID resolves	39,854	0
`grounded`	every leaf is a stated fact, with its stated certainty	6,008	0
`acyclic`	no fact rests on itself	79,708	0
`rule`	the cited rule exists for the relationship, and its conditions and weights match	33,846	0
`head`	the rule's head under the bindings is the fact	33,846	0
`conditions`	each condition under the bindings is the cited fact	65,628	0
`expressions`	re-evaluated from the bindings and cited facts, with the same result and outcome	4,334	0
`aggregates`	a list function cites exactly the facts that answer its sub-question	528	0
`certainty`	the fact's certainty, recomputed from its conditions, matches	33,846	0
`premise certainty`	each condition's certainty is the cited fact's certainty	65,628	0

The checker ran 363,226 proof checks on the evidence of 8,845 answers (39,854 distinct evidence nodes, up to 1,000 levels deep) across eight workload families. Every answer's evidence passed every check: 8,845 of 8,845. Evidence is shared wherever the reasoning is shared, so the largest tree, which would have 3.1×10^{15} nodes written out in full, is stored as a graph of a few hundred. Fetching evidence costs about 8.4 µs per node.

Are the premises the right ones?

The premises are sufficient. A fresh session given only the stated facts at the leaves of an answer's evidence derives the same answer with the same certainty, in all 1,058 cases tested.

The premises are necessary. We removed one stated fact cited by the evidence and asked again in a fresh session. Either the conclusion was lost or weakened, or the engine proved it again without that fact through a second, independent proof that itself passed every check. This held for all 1,637 facts cited by a rule step, and for all 2,891 leaves of an answer's evidence (2,703 lost or weakened, 188 re-proved independently).

©2026 Rainbird Technologies Limited
Public

Does the checker catch bad evidence?

A checker that never fails proves nothing on its own, so we corrupted one evidence node in each sampled tree, in one of eight ways, and checked the tree again. The checker rejected all 5,970 corrupted trees. Evidence that misstates a number, a binding, a premise or an aggregate is caught.

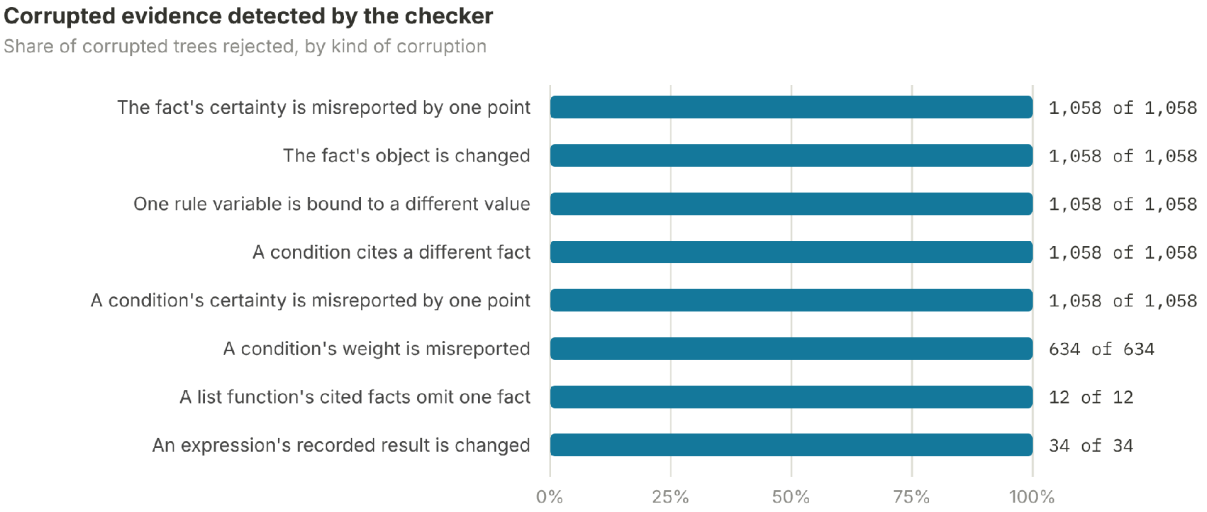


Fig. 6. The checker rejected every corrupted evidence tree, across all eight kinds of corruption.

Concurrency

Throughput holds under heavy load, and no request failed. The load test drove the HTTP server as a deployment would. Each client repeated a complete session as fast as it could: start, inject 620 facts, run two queries and fetch one evidence tree. Sessions are independent and run in parallel across every available core.

Table 8. Concurrency at a glance

7,959	queries a second with 256 concurrent clients (3,979 complete sessions and 23,876 HTTP requests a second)
0	failed requests, across every load level and core count
8.7×	the throughput of a single client (968 rising to 8,408 queries a second)

Throughput rises with concurrency until the cores are busy and then holds. It peaks at 8,408 queries a second at 32 clients, and sits at 7,959 at 256 clients and 8,209 at 512. Far more clients than cores do not reduce throughput.

©2026 Rainbird Technologies Limited
Public

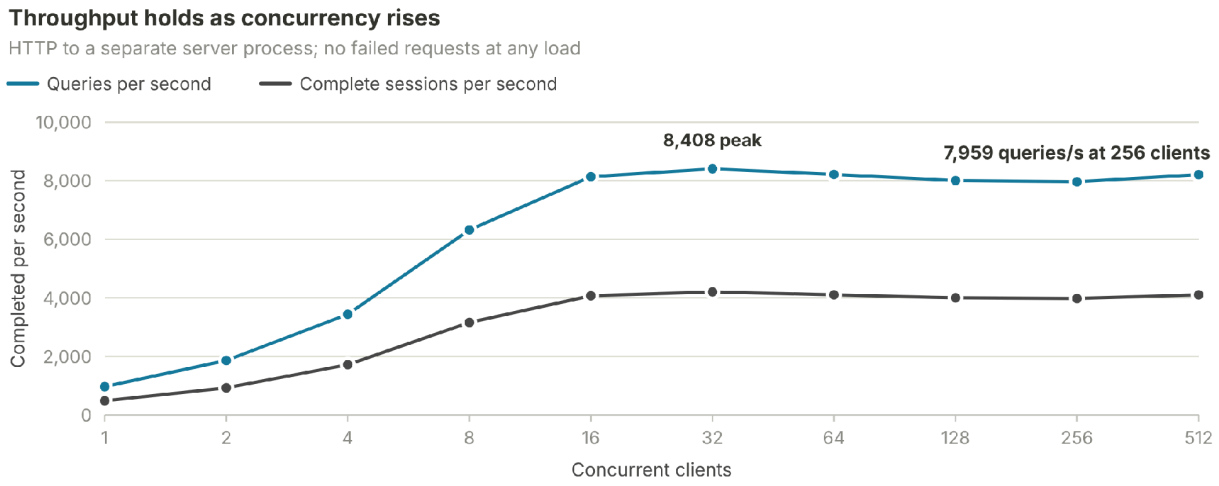


Fig. 7. Throughput rises until the cores are busy and then holds, from 32 clients to 512.

LITTLE'S LAW

Response time is queueing, not slowdown.

With N clients in a closed loop and throughput X, Little's law gives the mean time per session as W = N/X. Once the cores are saturated X is constant, so W grows in proportion to N: the extra time is spent waiting for a free core, not doing more work. The measured mean follows N/X to within 3% at every load. At 256 clients N/X is 64.3 ms and the measured mean is 63.3 ms. A single client's complete session takes 2.06 ms.

Scaling across cores. Throughput rises from 1,000 queries a second on one core to 9,014 on sixteen, a 9.0× speed-up. The clients ran on the same cores as the server, so this understates how far the server itself scales.

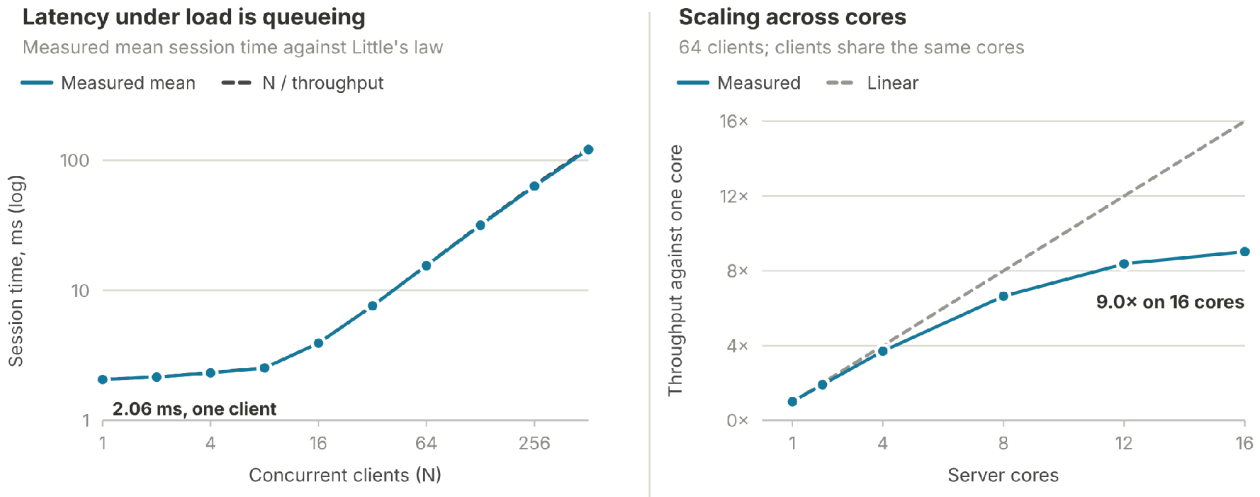


Fig. 8. Session time tracks N/X at every load, so added latency is queueing for a free core. Throughput on 16 cores is 9.0 times that on one.

©2026 Rainbird Technologies Limited
Public

Operations

Sessions are cheap to start, cheap to hold and quick to resume.

Persistence. A session can be saved and resumed in another engine instance. A 305,000-fact session snapshots in 42.3 ms and resumes in 1.81 s, and the resumed session's answers matched the original.

Sessions. Starting a session takes 0.8 µs (median). A live session holding 620 facts and its query results uses 466.9 KB, so 10,000 sessions fit in 4.5 GB.

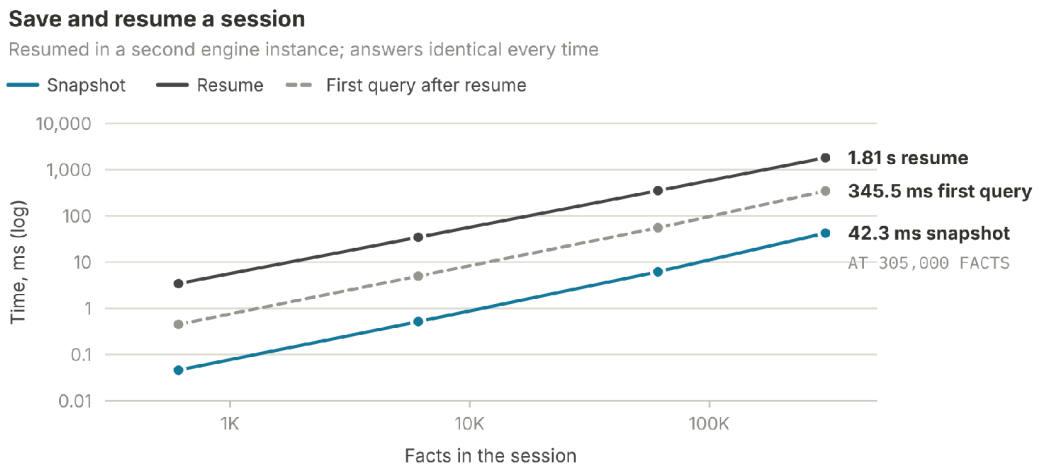


Fig. 9. Snapshot, resume and first-query times all grow in line with the size of the session, and the resumed answers were identical every time.

Interactive reasoning

Inputs can be asked for as clarifying questions rather than injected up front, and the engine asks only what the reasoning needs, when it needs it. Each answer is processed in 6.4 µs (median), and the decision reached through conversation is the same as the decision reached with every fact given up front.

Table 9. Interactive reasoning

Inputs	Questions asked	Answer p50	Answer p90	Whole conversation	Same as batch
1	1	4.2 µs	4.7 µs	6.5 µs	yes
2	2	4.3 µs	4.7 µs	11 µs	yes
5	5	5.3 µs	8.5 µs	31 µs	yes
10	10	7.4 µs	14 µs	84 µs	yes
20	20	13 µs	22 µs	273 µs	yes
50	50	30 µs	51 µs	1.57 ms	yes

©2026 Rainbird Technologies Limited
Public

Bounded resources

A model can contain rules that never stop deriving new facts. Every query runs under a configurable deadline and work budget, and stops with a clear error when it reaches either. The test model derives the next number forever. Every runaway query stopped, at most 50.8 ms after its deadline.

Table 10. Runaway queries stopped at their deadline

Deadline	Stopped after	Overshoot	Error returned
10.0 ms	14.1 ms	4.08 ms	`query timed out after 8192 steps`
100.0 ms	110.3 ms	10.3 ms	`query timed out after 24576 steps`
1.00 s	1.05 s	50.8 ms	`query timed out after 77824 steps`

Scope and limitations

All workloads are synthetic so that they can be published, reproduced and checked against known answers. They exercise the shapes found in production models (deep derivation, aggregates over derived facts, string building and recursion), but they are not customer models.

All tests ran on a single Apple M3 Max, 16 cores, 128 GB, and in the concurrency test the clients shared that machine with the server. Throughput over HTTP is the figure most sensitive to the machine's thermal state: earlier runs on the same day, under sustained load, measured about half the throughput shown here. Latencies of single queries varied far less.

Correctness is checked on the fragment each workload covers: joins, condition weights, rule certainty, several rules deriving one fact, recursion, counting and arithmetic. Questions are covered by the interactive test. Models in which negation runs through recursion (non-stratified models) are outside these tests.

©2026 Rainbird Technologies Limited
Public

Appendix: measured points

Table A1. Compounding depth

Depth	Paths	p50	p90	Correct
1	3	18 µs	30 µs	yes
2	9	19 µs	23 µs	yes
3	27	37 µs	43 µs	yes
4	81	44 µs	46 µs	yes
6	729	55 µs	57 µs	yes
8	6,561	74 µs	76 µs	yes
10	59,049	87 µs	91 µs	yes
12	531,441	91 µs	94 µs	yes
16	4.3×10^{7}	116 µs	120 µs	yes
20	3.5×10^{9}	152 µs	157 µs	yes
24	2.8×10^{11}	174 µs	178 µs	yes
28	2.3×10^{13}	203 µs	205 µs	yes
32	1.9×10^{15}	224 µs	232 µs	yes

Table A2. Rule chains

Levels	p50	Per level	Correct
1	5.8 µs	5.79 µs	yes
10	20 µs	1.98 µs	yes
50	62 µs	1.23 µs	yes
100	116 µs	1.16 µs	yes
250	290 µs	1.16 µs	yes
500	574 µs	1.15 µs	yes
1,000	1.68 ms	1.68 µs	yes
2,000	3.28 ms	1.64 µs	yes
5,000	8.47 ms	1.69 µs	yes

Table A3. Recursion

Form	Chain	Derived facts	p50	Correct
left	10	10	21 µs	yes
left	100	100	165 µs	yes
left	500	500	1.12 ms	yes
left	1,000	1,000	3.29 ms	yes

©2026 Rainbird Technologies Limited
Public

Form	Chain	Derived facts	p50	Correct
left	2,500	2,500	14.2 ms	yes
left	5,000	5,000	51.2 ms	yes
left	10,000	10,000	183.8 ms	yes
right	10	55	46 µs	yes
right	100	5,050	2.16 ms	yes
right	250	31,375	14.6 ms	yes
right	500	125,250	59.2 ms	yes
right	1,000	500,500	253.6 ms	yes
right	2,000	2,001,000	1.10 s	yes

Table A4. Model size

Rules	Query	Results	p50	p90	Correct
31	one conclusion	0	29 µs	45 µs	yes
31	rollup (all conclusions)	18	133 µs	141 µs	yes
31	fact lookup	1	3.8 µs	4.6 µs	yes
151	one conclusion	0	32 µs	38 µs	yes
151	rollup (all conclusions)	65	538 µs	579 µs	yes
151	fact lookup	1	3.4 µs	4.6 µs	yes
301	one conclusion	0	34 µs	44 µs	yes
301	rollup (all conclusions)	126	1.09 ms	1.13 ms	yes
301	fact lookup	1	3.2 µs	6.7 µs	yes
751	one conclusion	0	29 µs	34 µs	yes
751	rollup (all conclusions)	300	2.65 ms	2.80 ms	yes
751	fact lookup	1	4.2 µs	8.6 µs	yes
1,501	one conclusion	0	34 µs	42 µs	yes
1,501	rollup (all conclusions)	596	5.22 ms	5.50 ms	yes
1,501	fact lookup	1	4.8 µs	9.6 µs	yes
3,001	one conclusion	0	34 µs	37 µs	yes
3,001	rollup (all conclusions)	1,175	10.2 ms	10.5 ms	yes
3,001	fact lookup	1	4.0 µs	5.8 µs	yes
7,501	one conclusion	0	40 µs	43 µs	yes
7,501	rollup (all conclusions)	2,946	25.2 ms	26.6 ms	yes
7,501	fact lookup	1	6.3 µs	9.1 µs	yes
15,001	one conclusion	0	36 µs	44 µs	yes

©2026 Rainbird Technologies Limited
Public

Rules	Query	Results	p50	p90	Correct
15,001	rollup (all conclusions)	5,875	50.5 ms	53.8 ms	yes
15,001	fact lookup	1	12 µs	14 µs	yes

Table A5. Compiling the model

Rules	Source	Compile p50	Compiled size
31	13.4 KB	352 µs	74.0 KB
151	60.5 KB	1.30 ms	335.6 KB
301	119.4 KB	3.05 ms	664.1 KB
751	297.2 KB	8.17 ms	1.7 MB
1,501	593.6 KB	15.5 ms	3.3 MB
3,001	1.2 MB	30.7 ms	6.6 MB
7,501	2.9 MB	74.6 ms	15.9 MB
15,001	5.8 MB	138.3 ms	32.1 MB

Table A6. Data size

Facts	Query	Results	p50	p90	Correct
610	rollup (all conclusions)	88	532 µs	569 µs	yes
610	one conclusion	8	11 µs	16 µs	yes
3,050	rollup (all conclusions)	433	2.94 ms	3.07 ms	yes
3,050	one conclusion	45	60 µs	68 µs	yes
6,100	rollup (all conclusions)	870	5.73 ms	6.02 ms	yes
6,100	one conclusion	96	114 µs	127 µs	yes
15,250	rollup (all conclusions)	2,173	12.9 ms	13.3 ms	yes
15,250	one conclusion	237	231 µs	261 µs	yes
30,500	rollup (all conclusions)	4,350	29.9 ms	30.8 ms	yes
30,500	one conclusion	480	503 µs	534 µs	yes
61,000	rollup (all conclusions)	8,700	62.4 ms	62.9 ms	yes
61,000	one conclusion	960	999 µs	1.06 ms	yes
152,500	rollup (all conclusions)	21,750	167.8 ms	179.8 ms	yes
152,500	one conclusion	2,400	2.54 ms	2.59 ms	yes
305,000	rollup (all conclusions)	43,500	426.8 ms	449.2 ms	yes
305,000	one conclusion	4,800	4.74 ms	5.18 ms	yes

©2026 Rainbird Technologies Limited
Public

Table A7. Injecting facts

Facts	Inject p50	Facts/s	Session memory	Per fact
610	229 µs	2,660,360	259.6 KB	436 B
3,050	1.37 ms	2,230,010	1.3 MB	437 B
6,100	2.57 ms	2,369,890	2.4 MB	415 B
15,250	6.24 ms	2,445,640	6.4 MB	439 B
30,500	12.8 ms	2,380,970	12.7 MB	436 B
61,000	25.8 ms	2,365,450	25.3 MB	435 B
152,500	67.2 ms	2,268,140	56.6 MB	389 B
305,000	146.0 ms	2,088,850	112.6 MB	387 B

Table A8. Evidence by workload

Workload	Answers	Valid	Nodes	Depth	Full tree	Proof checks	Failed	Sufficient	Necessary
compounding	15	100%	1,086	33	3.1×10^{15}	8,100	0	100%	100%
rule chain	3	100%	1,113	1,000	1,001	10,002	0	100%	100%
recursion	500	100%	21,875	200	400	216,746	0	100%	100%
wide rule set	1,084	100%	5,211	2	6	38,188	0	100%	100%
reachability	1,776	100%	2,819	19	38	27,096	0	100%	100%
random stratified	3,381	100%	4,662	6	29	38,222	0	100%	100%
random recursive	1,994	100%	2,976	32	3,726	23,872	0	100%	100%
certainty chain	92	100%	112	13	27	1,000	0	100%	100%

Table A9. Concurrency

Clients	Queries/s	Sessions/s	Requests/s	Mean session	N/X	Errors
1	968	484	2,905	2.06 ms	2.07 ms	0
2	1,860	930	5,580	2.15 ms	2.15 ms	0
4	3,440	1,720	10,319	2.32 ms	2.33 ms	0
8	6,321	3,160	18,962	2.53 ms	2.53 ms	0
16	8,139	4,069	24,416	3.93 ms	3.93 ms	0
32	8,408	4,204	25,223	7.60 ms	7.61 ms	0
64	8,213	4,106	24,638	15.5 ms	15.6 ms	0
128	8,004	4,002	24,011	31.7 ms	32.0 ms	0
256	7,959	3,979	23,876	63.3 ms	64.3 ms	0
512	8,209	4,105	24,628	121.1 ms	124.7 ms	0

©2026 Rainbird Technologies Limited
Public

Table A10. Scaling across cores

Cores	Queries/s	Sessions/s	Speed-up
1	1,000	500	1.00×
2	1,909	954	1.91×
4	3,702	1,851	3.70×
8	6,634	3,317	6.64×
12	8,364	4,182	8.37×
16	9,014	4,507	9.02×

Table A11. Latency of each request at 16 concurrent clients

Request	Count	p50	p90	p99	p99.9
start	20,347	103 µs	322 µs	1.43 ms	4.02 ms
inject 620 facts	20,347	1.40 ms	2.52 ms	6.11 ms	20.9 ms
query: rollup	20,347	1.08 ms	2.11 ms	4.38 ms	15.5 ms
query: one conclusion	20,347	147 µs	415 µs	1.57 ms	4.69 ms
evidence	20,347	114 µs	344 µs	1.44 ms	3.87 ms
session	20,347	3.18 ms	6.10 ms	12.5 ms	37.8 ms

Table A12. Persistence

Facts	Record	Snapshot	Encode	Resume	First query	Identical
610	412.6 KB	46 µs	917 µs	3.43 ms	450 µs	yes
6,100	4.0 MB	517 µs	9.45 ms	34.4 ms	4.97 ms	yes
61,000	40.5 MB	6.18 ms	97.1 ms	350.9 ms	55.5 ms	yes
305,000	204.0 MB	42.3 ms	519.0 ms	1.81 s	345.5 ms	yes

Table A13. Memory per live session

Live sessions	Facts each	Memory	Per session
100	620	45.6 MB	466.9 KB
1,000	620	455.9 MB	466.9 KB
10,000	620	4.5 GB	466.9 KB

©2026 Rainbird Technologies Limited
Public