RBI Vector Retrieval Evaluation V2

Curator Research Private Limited. Final edition, 22 October 2026. Experiments run on 18 October 2026.

Prepared for technical review by Zilliz. The results describe the setups and data we tested. This is an author-prepared experimental report. It does not certify any service for production use, and it gives no service-level guarantee.

Summary in plain language

What we tested. Curator turns regulatory documents into lists of numbers called vectors. Similar text gets similar numbers. A vector database stores these vectors and finds the closest ones to a question very quickly. We tested five setups of this kind of database with the same documents from the Reserve Bank of India (RBI):

- our own Milvus server (local Milvus)
- Zilliz Cloud Dedicated
- Zilliz Cloud Serverless
- Zilliz Cloud Free
- Databricks Vector Search STANDARD, as an extra reference

Why it matters. Curator builds AI tools for banking, financial services, insurance and legal work. Professionals must be able to check every source. Search therefore has to be fast, and it has to find the right document.

What we found.

- Local Milvus, Zilliz Dedicated and Databricks returned almost the same results as a slow, exact search. Their agreement with exact search was between 98.9% and 99.5%.
- Zilliz Serverless and Free agreed with exact search on 85.7% to 85.9% of results. We tested them only at their default search setting. Zilliz offers a setting that trades speed for accuracy, and we have not yet tested it.
- Zilliz Dedicated in Pune answered a single search in 8.8 to 9.5 milliseconds (median) from our client in Mumbai. The services in Frankfurt and Ohio took longer. Network distance accounts for much of that extra time.

What it means for us. Zilliz Dedicated in Pune is a strong managed option for a pilot in India. Local Milvus stays as our cost and speed baseline. We would like to review the Serverless and Free settings with Zilliz before we decide how to use them.

What these tests do not show. We measured search only. We did not measure whether an AI answer to a regulatory question is correct. We also did not test our full data volume.

Key findings

Finding	View A / View B
Local Milvus, agreement with exact search (ANN recall@10)	99.2% / 99.0%
Zilliz Dedicated, agreement with exact search (ANN recall@10)	98.9% / 99.0%
Zilliz Serverless and Free, each, at default settings	85.7% / 85.9%
Zilliz Dedicated, single-search p95 time, Mumbai client to Pune	10.2 ms / 9.6 ms

View A has 196,158 vectors and 1,305 test questions. View B has 119,803 vectors and 756 test questions. View B leaves out documents that lack source metadata. Section 4 explains both views. [S1, S3]

1. **No single winner.** No database alone solves regulatory research. The results describe tested setups and network paths. They do not rank vendors in general.
2. **Dedicated is a credible managed candidate.** It combined high agreement with exact search and low response times for our India-based client. Choosing between Dedicated and local Milvus still needs a full operating-cost comparison. An existing local container adds no new subscription. It still has running and staffing costs. [S1]
3. **Serverless and Free need a tuning review.** We completed only their default search setting. Their Frankfurt location added a long network round trip from Mumbai. The tests do not show their best possible recall or their speed in the same region. [S1, S6]
4. **Source hit rates were closer together.** The share of questions where the right document appeared in the top ten was 87.4% to 91.0% across all backends. That gap is smaller than the gap in agreement with exact search. This is why we report the two measures separately. [S1, S3]
5. **Sizing.** The figure of about one million vectors that we discussed for the pilot covers RBI material only. It counts vectors from RBI documents. It does not count unique documents, and it sets no limit on Curator’s total need. In the proposed three-month period, Curator may add other regulators and an approved company knowledge base. The V2 tests themselves used smaller, fixed RBI subsets.

Next steps we request from Zilliz

1. A technical review of the recall gap at the default setting, and a tuning run that compares services at matched quality.
2. An India-suitable deployment and capacity path that can grow beyond the RBI starting point.
3. Written terms for three-month startup or proof-of-concept support. Please cover the credit amount, activation dates, eligible services, capacity changes and spending controls.

Please also name the technical owner on your side and tell us what information you need from Curator.

1. What the results mean for the product

Curator is developing evidence-linked AI for document-heavy work in banking, financial services, insurance and law. A useful search result must lead to a source that a professional can check, date and read in the right context. Text that only looks similar to the question is not enough. A vector database is one part of that workflow. Authoritative text, metadata, version records and links between documents are the other parts.

Three separate questions

We kept three questions apart. Each uses a different measure.

Question	Measure and limit
Did fast approximate search reproduce exact search?	ANN neighbour recall@k. ANN stands for approximate nearest neighbour. The measure is the overlap between the returned vector IDs and an exact top-k reference list.
Did we find the labelled source document?	Document hit@10. We take the first 100 vector results, reduce them to distinct documents and check whether the labelled (“gold”) document is among the first ten.
Was the final regulatory answer correct?	We did not test this. It needs source support, checks of effective date and applicability, interpretation and answer-level evaluation.

Document MRR@10 (mean reciprocal rank) scores the position of the gold document among those distinct documents. It is zero when the gold document is missing from the first ten. Section hit@10 checks for the exact gold vector IDs within the first ten vectors. These measures count different things, so one cannot replace another. These results establish neither precision@k nor full answer accuracy. [S3, S4]

2. Retrieval quality in detail

View A keeps documents whose scraper metadata is missing. View B removes those documents and the questions whose gold sources sit in them. The values are medians of three repeats, rounded as in the source ledger. They are descriptive. We make no claim of statistical equivalence or significance. [S1]

View A: 196,158 vectors, 1,305 queries

Backend	ANN R@10	ANN R@100	Doc hit@10	Doc MRR@10
Local Milvus	99.2%	98.3%	90.3%	0.654
Zilliz Dedicated	98.9%	98.4%	91.0%	0.657
Zilliz Serverless	85.7%	84.6%	87.4%	0.635
Zilliz Free	85.7%	84.6%	87.4%	0.635
Databricks STANDARD	99.4%	99.6%	90.7%	0.655

View B: 119,803 vectors, 756 queries

Backend	ANN R@10	ANN R@100	Doc hit@10	Doc MRR@10
Local Milvus	99.0%	98.2%	90.3%	0.700
Zilliz Dedicated	99.0%	98.4%	90.7%	0.703
Zilliz Serverless	85.9%	84.4%	88.1%	0.673
Zilliz Free	85.9%	84.4%	88.1%	0.673
Databricks STANDARD	99.5%	99.6%	90.7%	0.705

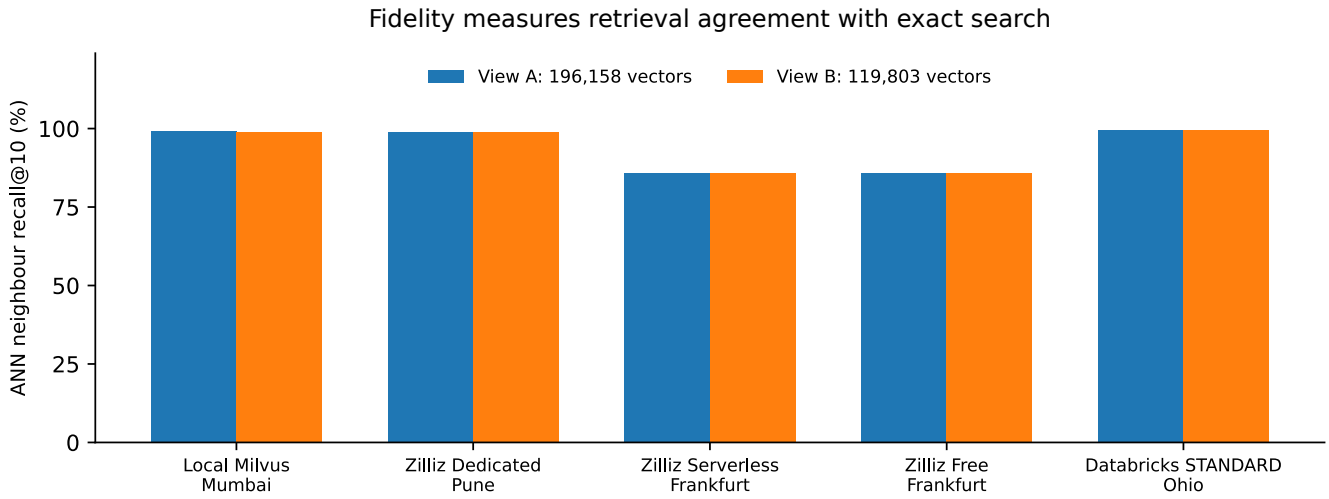

Figure 1. Agreement of each backend with exact search, measured as strict vector-ID overlap. The chart does not show the share of correct legal answers. Views A and B have different question sets. All Zilliz rows use the default search setting.

The small differences between local Milvus, Dedicated and Databricks do not show a product-quality advantage. The larger gap for Serverless and Free at the default setting is a useful question for vendor review. It does not show that tuning cannot close the gap. Duplicate passages across documents and tied scores also limit what strict-ID overlap can tell us. [S1, S3]

3. Speed measured at the client

Each cell shows p50 / p95 / p99 in milliseconds. These are the times within which 50%, 95% and 99% of searches finished. Each search sent one query vector and asked for ten results. The time includes the network path and client work. It excludes text embedding, graph traversal, reranking, text loading and answer generation. [S1, S4]

Backend	View A: p50 / p95 / p99	View B: p50 / p95 / p99
Local Milvus	2.1 / 3.5 / 6.0	2.5 / 3.9 / 4.8
Zilliz Dedicated	9.5 / 10.2 / 10.5	8.8 / 9.6 / 10.2
Zilliz Serverless	141.5 / 160.7 / 211.1	134.9 / 147.6 / 205.3
Zilliz Free	139.6 / 156.8 / 180.3	138.6 / 145.1 / 153.3
Databricks STANDARD	305.2 / 508.9 / 881.3	298.4 / 345.7 / 483.0

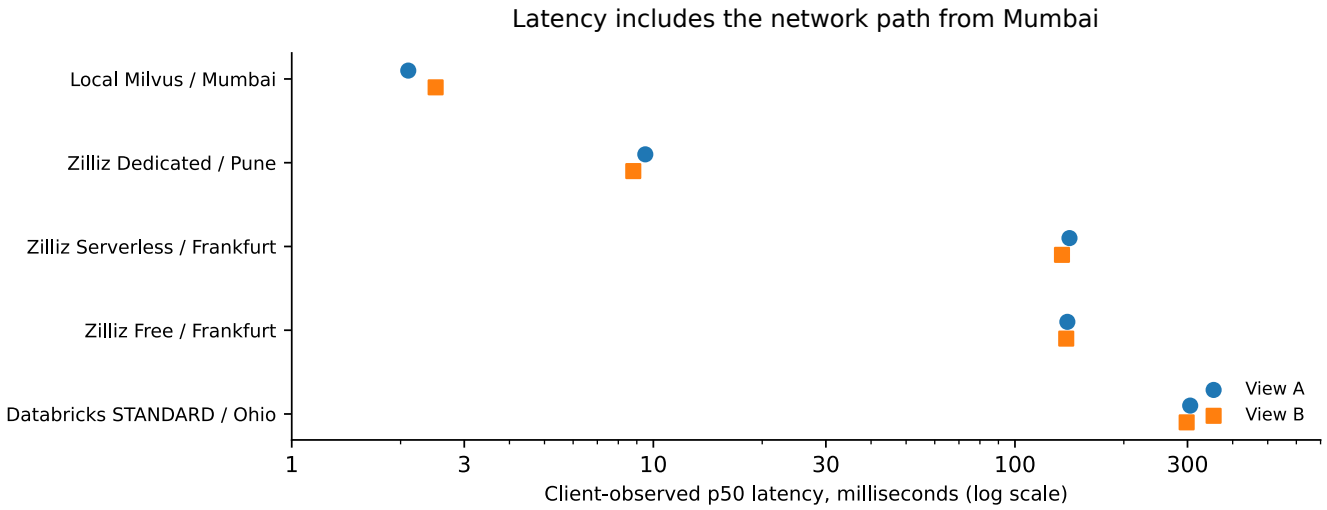

Figure 2. Median single-search time seen by the client, on a logarithmic axis. Local loopback, Pune, Frankfurt and Ohio are different network paths. The chart does not compare database engines alone under controlled conditions.

What we can conclude

For this Mumbai-based application path, Dedicated in Pune was much faster than the Frankfurt shared tiers and the Ohio Databricks endpoint. Local loopback was fastest. Simple round-trip probes to the control plane took roughly 8 ms to Pune, 132-135 ms to Frankfurt and 266-275 ms to Ohio. These probes help explain the effect of geography. Subtracting them from search time would still not isolate the exact server processing time. [S1]

Other work on the client host competed with many timing phases. Even the attempted quiet-window rerun of local Milvus stayed busy during parts of its measurement windows. We therefore label the values as observed and contended. They should not be read as mathematical upper or lower bounds. A fair comparison of service efficiency needs an isolated client and matched regions. [S1]

The quality test asked for 100 vectors per search. The speed table above asks for ten. Please do not combine these speed figures with document hit@10 to claim a full document-search response time. The source ledger keeps separate top-100 timings. [S3, S4]

4. Data used and where it came from

The parent V1 corpus has 795,654 cached vectors linked to 13,516 RBI document identities. The vectors cover document summaries, sections, chunks, generated rules and knowledge statements. So the corpus has far fewer documents than vectors. The V2 review selected 2,000 documents with seed 20261018. Of these, 800 were sampled around the pool of gold questions and 1,200 were random. This enriched sample does not represent all RBI material. [S2]

Data waterfall	Documents	Vectors	Queries
Reviewed base slice	2,000	216,246	1,310
Hard exclusions	1 document-level item	20,088	5 duplicate queries
View A: retained, metadata orphans included	1,999	196,158	1,305
Additional removal for View B	482	76,355	549
View B: metadata orphans excluded	1,517	119,803	756

Document totals follow the retained-document count in the frozen review record (1,999 and 1,517). The results ledger repeats document-list counts of 2,000 and 1,518. Vector and query counts are the same in both records. Document counts and vector counts must stay separate. [S2]

The 20,088 hard-excluded vectors were 16,749 duplicate texts within the same document, 3,212 texts shorter than 20 characters and 127 website navigation items. We flagged a further 25,796 duplicate passages across documents and kept them. They matter for strict-ID overlap and for any assumption of independent samples. [S2]

Every reviewed vector ID resolved to text and a document. There were zero ID-orphan rows. Missing scraper metadata is a separate issue. 482 documents were metadata orphans. They covered 77,046 vectors before hard exclusions and 76,355 retained vectors, which View B removes. Only 5,547 rows had exact text-joined section lineage. 208,699 non-document rows kept lineage at document level only. [S2]

View A has 999 primary and 306 section-linked questions. View B has 678 and 78. The questions and source labels come from the earlier enrichment and section-linked evaluation. A new, independent, human-reviewed test set was not completed. We report both views. Their hit rates should not be read as a before-and-after comparison. [S2, S5]

The review found 433 base-slice documents without a publication date. Version and as-of metadata was incomplete. We reviewed passage text, document identity and the link to cached vectors. We did not establish full legal-source completeness or answer correctness. [S2]

Embeddings

All tests used 1,024-dimensional vectors from the Qwen3-Embedding-8B cache. The 1,024 dimensions are a Matryoshka projection of the cached 4,096-dimensional vectors. The corpus vectors came from cached GPTQ (quantised) serving output. The query vectors came from a full-precision CPU run. A fresh re-embedding at matched precision was outside V2. The review flagged 1,897 cached vectors with an off-norm signature linked to earlier pooling. We kept and

labelled them. All backends share the same frozen arrays. Even so, the comparison cannot separate every effect of how the embeddings were made. [S2]

A separate 0.6B experiment covered document text only, cut to 1,200 characters. Its re-embedding link check was incomplete at the report date. It is outside the five-backend comparison. Lakehouse and V1 rows use different corpora or repeats, and they are also outside the main tables. [S1, S2]

5. Tested systems

Backend	Setup and path from the Mumbai client
Local Milvus	Milvus 2.6.11, 3 vCPU / 6 GB container, HNSW M=16, efConstruction=200, ef=128. Mumbai, loopback.
Zilliz Dedicated	Performance-optimized, 1 CU (compute unit), AUTOINDEX, default search level, Milvus 2.6-compatible response. Pune, Azure Central India.
Zilliz Serverless	AUTOINDEX, default search level, Milvus 3.0-compatible response. Frankfurt, AWS eu-central-1.
Zilliz Free	AUTOINDEX, default search level, Milvus 3.0-compatible response. Frankfurt, AWS eu-central-1.
Databricks STANDARD	Vector Search STANDARD, Delta Sync / TRIGGERED, L2 distance on the supplied normalized vectors. Ohio, AWS us-east-2.

All five backends used the same 1,024-dimensional Qwen3-Embedding-8B data view and the same query vectors within that view. The local container had resource limits. It ran on a shared host. Cloud resources are not hardware-equivalent to that container. The Dedicated and shared Zilliz tiers also reported different compatibility versions. [S1, S2]

Zilliz and the exact reference used inner-product scoring. The Databricks run used L2 distance on the normalized vectors. Anyone reproducing the test must match the vector normalization. The distance metrics are not interchangeable in general. [S1, S2]

6. How we ran the tests

The runs took place on 18 October 2026. Frozen manifests for each view identify the arrays, metadata and query files. We loaded the same view into each backend, and the source ledger records matching dataset hashes. We created managed clusters for the evaluation and removed them afterwards. These runs used public RBI material and no customer data. [S1, S2]

Step	Protocol and reproduction requirement
1. Freeze the input	Use the A or B manifest below. Keep corpus and query vectors paired with the 8B model. Keep string IDs intact end to end. Keep the separate 0.6B data out.
2. Set the reference	Compute exact inner-product neighbours over the same eligible corpus. For filtered tests, compute them only among rows that pass the same filter.
3. Match the setup	Record index and search parameters, field precision, backend response version, client version, region and resource limits. Use the setups in section 5.
4. Measure retrieval	For each query, request 100 vectors. Score strict ANN overlap, distinct-document hit and MRR, and section hits separately. Repeat three times and keep ranked IDs and scores.
5. Measure speed and throughput	Time SDK search calls at the stated nq (query vectors per request) and top-k (results per query). Sweep 1, 8, 32 and 64 clients for 30 seconds per step, three repeats. Keep errors, client load and warm-up traces.
6. Review deviations	Keep failed attempts and contention labels apart from successful medians. Never merge the best quality from one setting with the best speed from another.
7. Reproduce before scaling	Use isolated clients in matched regions. Tune to an agreed recall target. Add paired, family-aware uncertainty, current billing attribution and tests of source and version quality.

Latency sampling follows the base harness: 200 requests per nq/top-k case, three repeats. V2 added the concurrency sweep and filters for each view. For the nq=10 case, Databricks used concurrent single requests. Milvus and Zilliz used one batched request. So that case is outside the headline latency table. Exact warm-up and quiet-window deviations differ by backend. [S1, S4]

Frozen dataset fingerprints (SHA-256)

```
View A: c7aade3a4e832a1dd9301e9a9f88dd98
        a5b8e27aeebaf0e77f2645b868758389
View B: 33d57353e51473d233e39bf572150809
        3521583b90745c6adaf96f98fe110b23
```

Curator keeps the complete run receipts, per-query results and data manifests. A reproducible exchange with Zilliz should include only agreed, rights-cleared files and sanitized setups. This release supplies a machine-readable metrics ledger. It does not share the source corpus, credentials or private runtime records.

7. Throughput, errors and repeatability

The tables show successful searches per second (QPS) at 1, 8, 32 and 64 concurrent clients. Each search sent one query vector and asked for ten results. Each step ran for 30 seconds. Values are medians of three repeats. Throughput describes the whole tested path of client, network and service. It does not prove a server capacity limit. [S1]

View A

Backend	1 client	8 clients	32 clients	64 clients
Local Milvus	408	1,284	1,378	1,270
Zilliz Dedicated	24.5	635.9	1,617	1,466
Zilliz Serverless	6.9	55.1	200.1	189.4
Zilliz Free	7.1	56.9	200.6	187
Databricks STANDARD	3.2	25.3	71.8*	67.3*

View B

Backend	1 client	8 clients	32 clients	64 clients
Local Milvus	413	1,210	1,466	829
Zilliz Dedicated	106.2	857.5	1,550	1,542
Zilliz Serverless	7	56.5	199.4	189.6
Zilliz Free	7	53.7	200.1	188.9
Databricks STANDARD	3.2	24.6	64*	66.7*

* Databricks returned HTTP 429 (rate-limit) errors at 32 and 64 clients. Its best error-free steps were 25.3 QPS in View A and 24.6 QPS in View B, both at eight clients. The higher successful-request rates therefore do not show error-free operating capacity. [S1]

Read the anomalies as well as the peaks

The single-client step for Dedicated in View A was noisy. It gave 24.5 QPS with a 27.7 ms median during that throughput phase. The separate latency test gave 9.5 ms. The single-client step for Dedicated in View B gave 106.2 QPS at a 9.1 ms median, which is closer to its latency test. We keep both observations and treat neither as a universal figure. [S1]

Local Milvus and all Zilliz tiers reported zero errors within the measured windows. Three shared-tier runs logged closed-channel errors after their measured windows, while client work was still in flight. We do not count these as measured-window errors. They do mean that we cannot call every aspect of those runs error-free. [S1]

Local reruns moved a lot. In View A at eight clients, the median fell from 1,284 to 854 QPS. In View B it rose from 1,210 to 1,509. For this reason we claim no definitive throughput winner and no formal saturation ceiling. [S1]

8. Filtering, ingestion and index readiness

Agreement with exact search under a filter

The test filter selected rows with a recent publication date. A publication date is different from the date on which a provision takes legal effect. We recomputed the exact reference over the rows that pass the filter. The harness labelled the run mode “post”. That label does not prove how the cloud service executes the filter internally. We did not measure iterative-filter alternatives. [S1]

Backend	Filtered ANN R@10: A	Filtered ANN R@10: B
Local Milvus	98.6%	98.6%
Zilliz Dedicated	99.1%	99.3%
Zilliz Serverless	87.7%	88.1%
Zilliz Free	87.7%	88.1%
Databricks STANDARD	99.1%	98.7%

The local runs used date_days >= 16617 in View A and date_days >= 19443 in View B. 57,963 and 30,019 rows passed. The filter was meant to cover a recent quarter. Because of the date distribution, it admitted about 29.5% and 25.1% of rows. Missing dates and gold documents outside the window mean that filtered source-hit scores do not cleanly test legal applicability. [S1]

Ingestion and indexing

Backend	Ingest rows/s A / B	Build seconds A / B
Local Milvus	5,564 / 6,138	290 / 179
Zilliz Dedicated	4,357 / 4,453	41 / 151
Zilliz Serverless	946 / 1,201	9.4 / 16
Zilliz Free	502 / 679	9.7 / 10.2
Databricks STANDARD	5,721 / 3,641	1,603 / 1,334

Each backend measures these steps at different boundaries. Local Milvus and Zilliz ingest used SDK inserts. Databricks measured upload plus table creation before Delta Sync indexing. Build time is only part of total onboarding time, and load and readiness time is separate. Use these figures to plan a reproducible follow-up. They do not rank ingestion speed in general. [S1, S4]

Local Milvus ran explicit compaction and readiness steps. For the managed services we treated compaction as service-managed and did not observe it directly. A common production test should record, for every backend, the searchable row count, index completion, warm-up, background maintenance and restart behaviour. [S1]

For Curator the lesson is clear. Fast filtering cannot repair missing source metadata. A publication-date filter cannot replace logic for effective dates, amendments and supersession.

9. Cost and the next three months

Recorded cost assumptions (no current quotation)

Backend	Dated cost basis for the evaluation
Local Milvus	Existing shared infrastructure. No new managed-service fee. EC2, storage, backups and engineering effort still cost money.
Zilliz Dedicated	Recorded list rate on 16 October: USD 0.36 per CU-hour. 1 CU for 730 hours gives USD 262.80. The V2 Dedicated lane cost an estimated USD 0.63, including aborted attempts.
Zilliz Serverless	V2 lane estimates of at most about USD 5.4 for View A and USD 4.8 for View B, based on earlier vCU rates. These are estimates. They have not been reconciled with invoices.
Zilliz Free	No charge within the tier quota during these runs. Quota and commercial terms need confirmation for a sustained pilot.
Databricks STANDARD	The ledger assumes a dated USD 0.28 per hour endpoint rate, or USD 204.40 for 730 hours. Loading and SQL services and billing delays limit how fully we can attribute total cost.

These figures name no overall cost winner. A future proposal should compare equal data scope, recall settings, region, error rate and availability. It should include storage, updates, query volume, embedding inference, backup and recovery and operating effort. Charges covered by credits are still usage costs. Historical list prices, estimated usage, credits and invoice amounts are separate quantities. We assert no current quote, credit amount or permanently free self-hosted deployment. [S1, S4]

Sizing and growth

The starting figure of about one million vectors covers RBI material only. It counts vectors and sets no platform-wide limit. V1 evaluated 795,654 vectors derived from RBI documents. V2 used two smaller reviewed subsets to study data and backend behaviour. This report did not measure the historical 16.57-million-row environment or the larger long-term corpus scenarios. [S1, S2]

In the proposed three-month period, Curator may add other regulators and a company knowledge base. Growth should follow admitted documents, actual passage counts, measured usage and written spending controls. Company and internal material must use approved data-handling and residency arrangements. This report gives no permission to move such material to a shared service overseas.

10. Limitations and decision boundaries

Limitation	What we cannot claim
Default search settings only	The best recall and speed that Zilliz tuning could reach. A V2 attempt at a higher search level failed in the harness. Later fixes did not produce replacement measurements.
Shared client and different regions	A speed ranking of engines alone, efficiency per unit of hardware, or a definitive server saturation ceiling.
Limited data and cached embeddings	Results at one million vectors, 100 million vectors or the whole regulatory corpus. That 1,024 dimensions always keep the quality of 4,096 dimensions.
Strict-ID overlap and duplicates	Tie-aware semantic recall. Passages with identical content can carry different IDs and lower the overlap.
Incomplete metadata and version lineage	Legal applicability, correct handling of superseded rules, or accuracy of answers as of a given date.
No paired, family-aware statistics	Formal superiority or equivalence from small score gaps. Rounded medians are descriptive and prove no significance.
No answer-generation test	That a 90% document hit rate means 90% correct answers or professional-grade legal advice.

Decision supported now

Keep local Milvus as the baseline. Take Zilliz Dedicated forward as the managed candidate once its operating benefits justify a paid India-region service. Review the shared-tier recall settings and deployment region with Zilliz before deciding on interactive use. Treat the observed gap at the default setting as an open tuning question. The data does not show a defect or a final quality ceiling for those tiers.

This is the final write-up of the completed, scoped measurements. The broader V2 validation plan still has open tasks. These include human-reviewed hard questions, complete version and section lineage, paired family-aware statistics and sustained testing from an isolated client. Finishing this document does not complete those tasks. An independent scientific review has also not taken place. [S5]

11. Sources and reporting notes

We keep measurement records apart from interpretation. The technical report, website article and LinkedIn post take their shared values from curator-rbi-v2-results.json. Values keep the rounding of the source ledger. We substituted no new backend measurement during editing.

ID	Source and scope
S1	Curator Research, RESULTS_V2_2026-10-18.md. Five backend sections, two frozen views, local rerun and documented limitations. Experiments dated 18 October 2026.
S2	Curator Research, DATA_V2_2026-10-18.md and the frozen review_report.json. Cohort, exclusions, metadata and ID-orphan distinction, embedding provenance and source links.
S3	Curator Research, tools/eval/vector/rbi_vector_eval.py. quality_metrics, METRIC_LABELS and V2_NAMES define vector overlap and distinct-document ranking units.
S4	Curator Research, METHOD_V1.md, as amended by the V2 backend sections. The V2 dataset sizes and concurrency sweep replace the earlier V1 values.
S5	Curator Research, V2_INPUTS_2026-10-17.md and V2_PLAN.md. Audit inputs, planned controls and external reporting requirements. Section 10 lists the open validation work.
S6	Zilliz documentation, Tune Recall Rate, accessed 21 October 2026. It explains how the level setting balances recall and performance. It cannot show the result of a tuning run we did not perform.

Public technical reference: https://docs.zilliz.com/docs/tune-recall-rate

Source pin and integrity

Repository source snapshot: `205cf71d81d62a18e193f0065304c1000e56c622`

Frozen data-review record SHA-256:

```
d04f7a8ddd3c12a5f3b77162ba9e1b6f
547b9df1cf5ef287b69dbf98c47ad590
```

Corrections carried into this release

This report replaces earlier broad statements that all stores had equal fidelity. It now gives observations for each setup. It uses retained document counts throughout. It keeps strict neighbour recall apart from document quality and answer quality. It labels contention and claims no bounds that the data cannot support. It keeps historical list-rate estimates apart from billed cost. We changed no raw metric to favour any backend.

The report leaves out private cloud project and cluster identifiers, invoice identifiers, credential references and internal file locations. It also leaves out performance at the one-million-vector planning point (we did not measure it), current live index totals and the separate historical global-news (GDELT) benchmark.

Prepared by Curator Research Private Limited, 22 October 2026, from the 21 October 2026 release. This is an author-prepared experimental report. It does not provide independent certification or a production service-level guarantee.