

SQL+Vector: Empower GenAI Apps With Relational Vector Database

Cheng Tai | CEO of MyScale

Chengt@myscale.com

LLM + Vector Databases is Key Stack to Building GenAI Applications

LLMs



ANTHROPIC



+ Vector Stores



Search



Generation



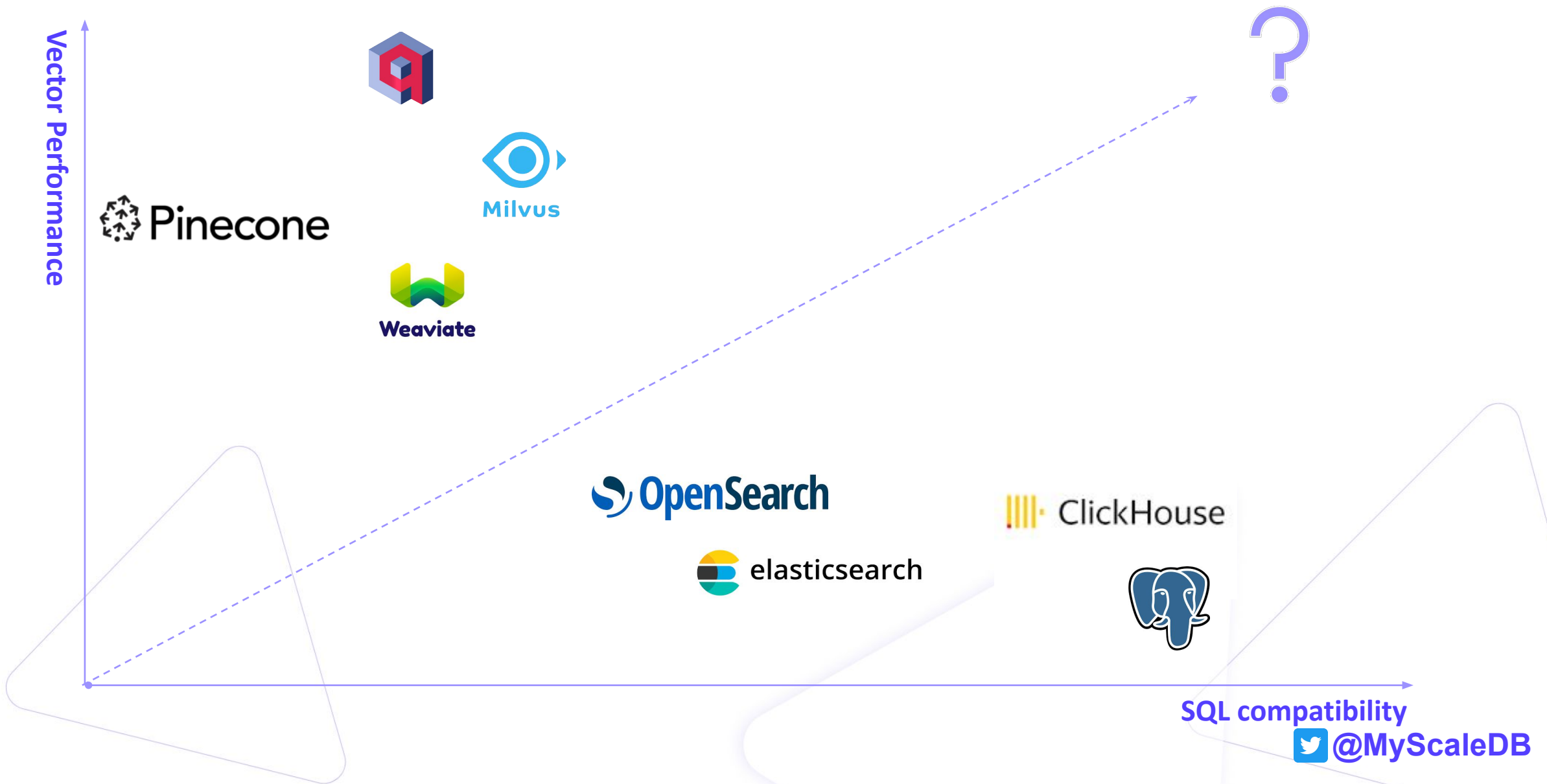
Data Management



Automation



The Dilemma: Convenience vs Vector Performance

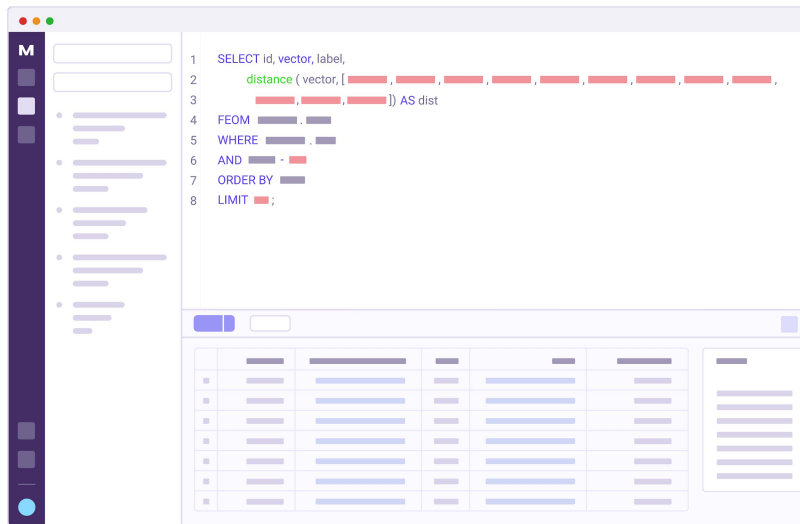


MyScale: Run Vector Search With SQL

Streamlined and fully managed

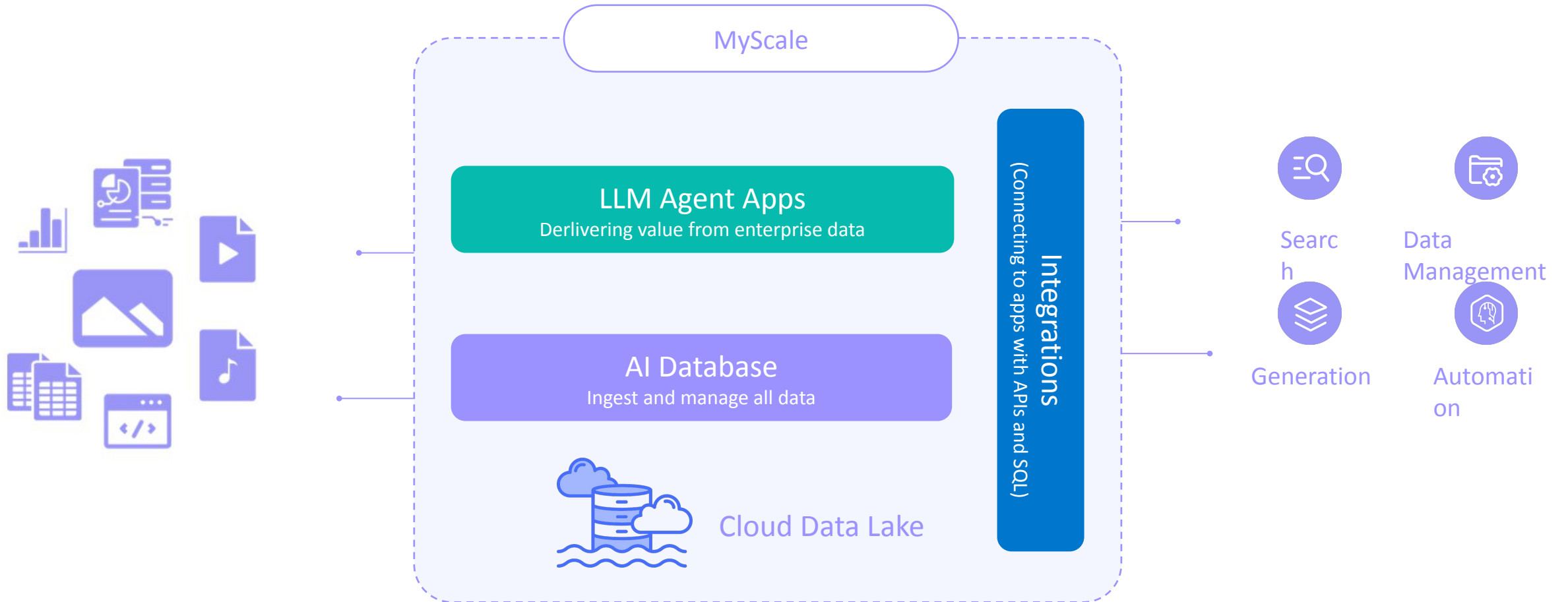
Combining the power of SQL and vector for your AI data workloads

Use the full library of SQL's intuitive syntax in obtaining fast answers to even the most complex SQL + vector queries on your vectorized and structured data.



```
SELECT
  myscale_movie.id,
  myscale_movie.title,
  myscale_rating.rating,
  distance(
    myscale_movie.embedding,
    [0.00005786368, 0, ..., 0.000098422155]
  ) AS dist
FROM
  myscale_movie
  INNER JOIN myscale_rating
  ON myscale_rating.movieId = myscale_movie.id
WHERE
  myscale_rating.rating > 4
ORDER BY
  dist DESC
LIMIT
  10;
```

MyScale Platform Outperforms Specialized Vector Databases



MyScale simplifies the complexity, cuts the costs, and maximizes the value of customers' AI initiatives.

User Story 1: MyScale Allows Users to Run Complex Queries



As a prominent **digital asset management** company, BitCap conducts extensive daily data analysis. Given the abundance of unstructured data in the financial market, MyScale's comprehensive support empowers BitCap to obtain real-time market signals and valuable decision-making insights.

- **LIKE** syntax for performing filter searches.
- Various data types, including **dates** and **strings**.
- Integration with **Langchain & Self-query**.

```
CREATE TABLE db.message_app
(
  user_id      FixedString(16),
  timestamp    DateTime,
  message_id   FixedString(16),
  message_embedding Array(Float32),
  CONSTRAINT check_length CHECK length(message_embedding) = 768
) ENGINE = MergeTree
ORDER BY (user_id, message_id)
PARTITION BY sipHash64(user_id) % 10
```

In this case, MyScale is the best/only choice.

User Story 2: MyScale Helps Academic Users Achieves Best Cost-Effectiveness

Enable academic customer to build a **research assistant** that is able to chat with millions of papers

ChatData

We provides you metadata columns below for query. Please choose a natural expression to describe filters on those columns.

For example:

If you want to search papers with complex filters:

- What is a Bayesian network? Please use articles published later than Feb 2018 and with more than 2 categories and whose title like computer and must have cs.cv in its category.

If you want to ask questions based on papers in database:

- What is PageRank?
- Did Geoffrey Hinton wrote paper about Capsule Neural Networks?
- Introduce some applications of GANs published around 2019.
- 请根据 2019 年左右的文章介绍一下 GAN 的应用都有哪些
- Veuillez présenter les applications du GAN sur la base des articles autour de 2019 ?
- Is it possible to synthesize room temperature super conductive material?

VectorSQL Self-Query Retrievers

You can retrieve papers with button Query or ask questions based on retrieved papers with button Ask.

```
CREATE TABLE default.ChatArXiv (
  `abstract` String,
  `id` String,
  `vector` Array(Float32),
  `metadata` Object('JSON'),
  `pubdate` DateTime,
  `title` String,
  `categories` Array(String),
  `authors` Array(String),
  `comment` String,
  `primary_category` String,
  VECTOR INDEX vec_idx vector TYPE MSTG('metric_type=Cosine'),
  CONSTRAINT vec_len CHECK length(vector) = 768
) ENGINE = ReplacingMergeTree ORDER BY id
```

Ask a question:

What is PageRank?

Query Ask

Query Log

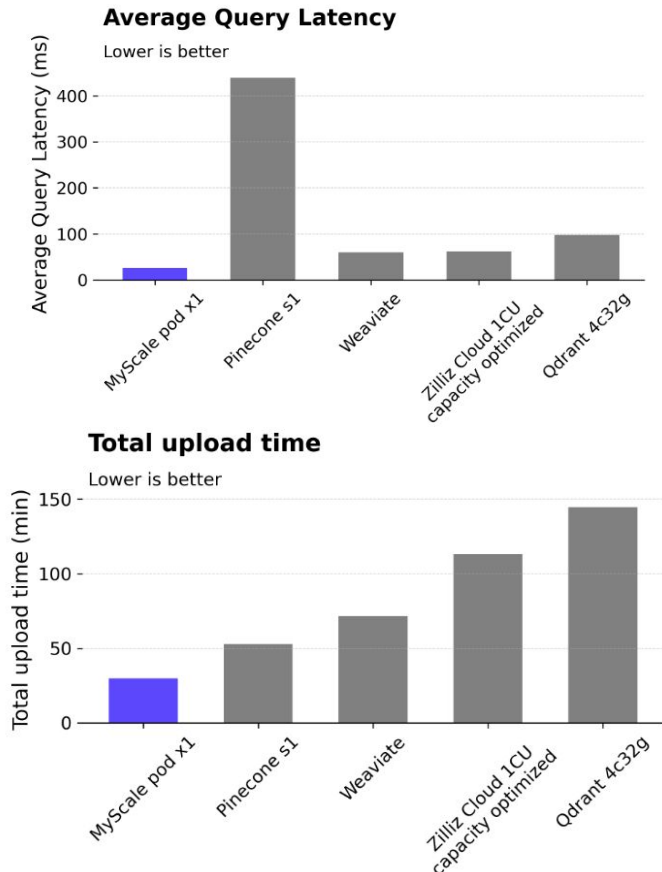
Done!

We generated Vector SQL for you:

```
SELECT title,
  abstract,
  authors,
  pubdate,
  categories,
  id
FROM chatarxiv
ORDER BY distance(vector, neuralarray(pagerank)) limit 10
```

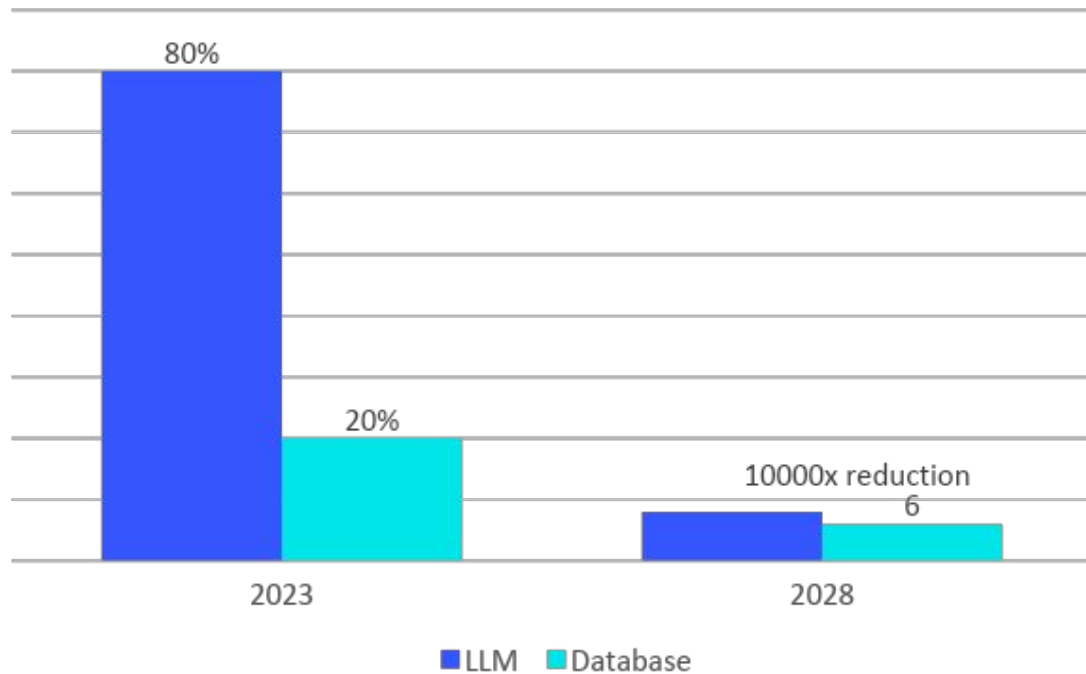
	title	abstract
0	PageRank: Standing on the shoulders of giants	PageRank is a Web page r
1	The PageRank Problem, Multi-Agent Consensus and Web Aggregation -- A	PageRank is an algorithm
2	Advanced Page Rank Algorithm with Semantics, In Links, Out Links and	In this paper we have moc
3	PageRank for evolving link structures	In this article we will look
4	Distributed Randomized Algorithms for the PageRank Computation	In the search engine of Go
5	Multiple equilibria of nonhomogeneous Markov chains and self-validating	PageRank is a ranking of ti
6	Discovering and Leveraging the Most Valuable Links for Ranking	On the Web, visits of a pag
7	PageRank beyond the Web	Google's PageRank metho
8	How to make the top ten: Approximating PageRank from in-degree	PageRank has become a k
9	A Web Aggregation Approach for Distributed Randomized PageRank	The PageRank algorithm e

Latency reduced by **4x**
Load time for 100k papers reduced to **30 mins**
3.7x savings compared with best alternative

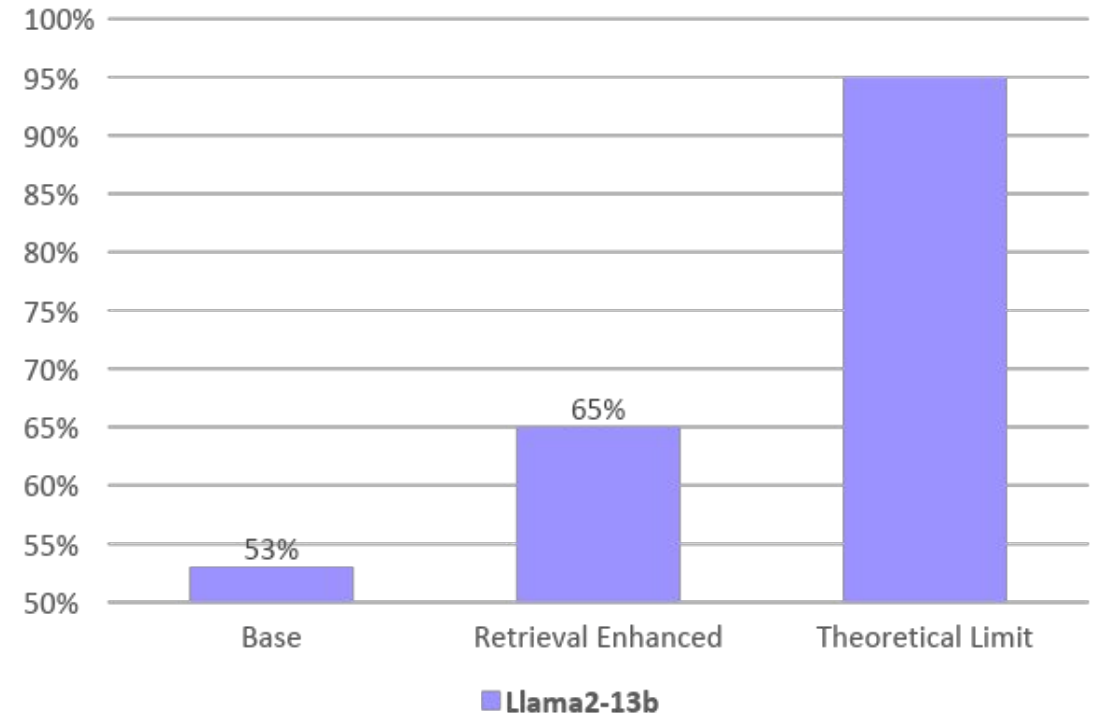


There is Huge Room to Optimize Performance, Cost and Quality

Cost of building GenAI apps will go down



Accuracy will increase



Contact Us -- Booth Myscale



Twitter



Discord



Try MyScale Now