FedVS: Towards Federated Vector Similarity Search with Filters

Zeheng Fan, Yuxiang Zeng, Zhuanglin Zheng, Binhan Yang, Yongxin Tong

August 3-7, 2025

State Key Laboratory of Complex & Critical Software Environment
Beihang University

1 Introduction

Vector similarity search is a new search paradigm inspired by a hybrid data type that integrates both high-dimensional embeddings and structured attributes. Given a query vector and a filter constraint on structured attributes, it identifies k objects from large-scale datasets based on two criteria: (1) their attributes must match the filter and (2) they are the k nearest neighbors (kNNs) to the query vector within the set of filtered data objects. While both industry and academia have developed efficient solutions to vector similarity search, they cannot address the challenge involved in **searching across multi-sourced datasets**, which is widely applied in scenarios like collaborative pharmaceutical development. It serves as the core problem in our research.

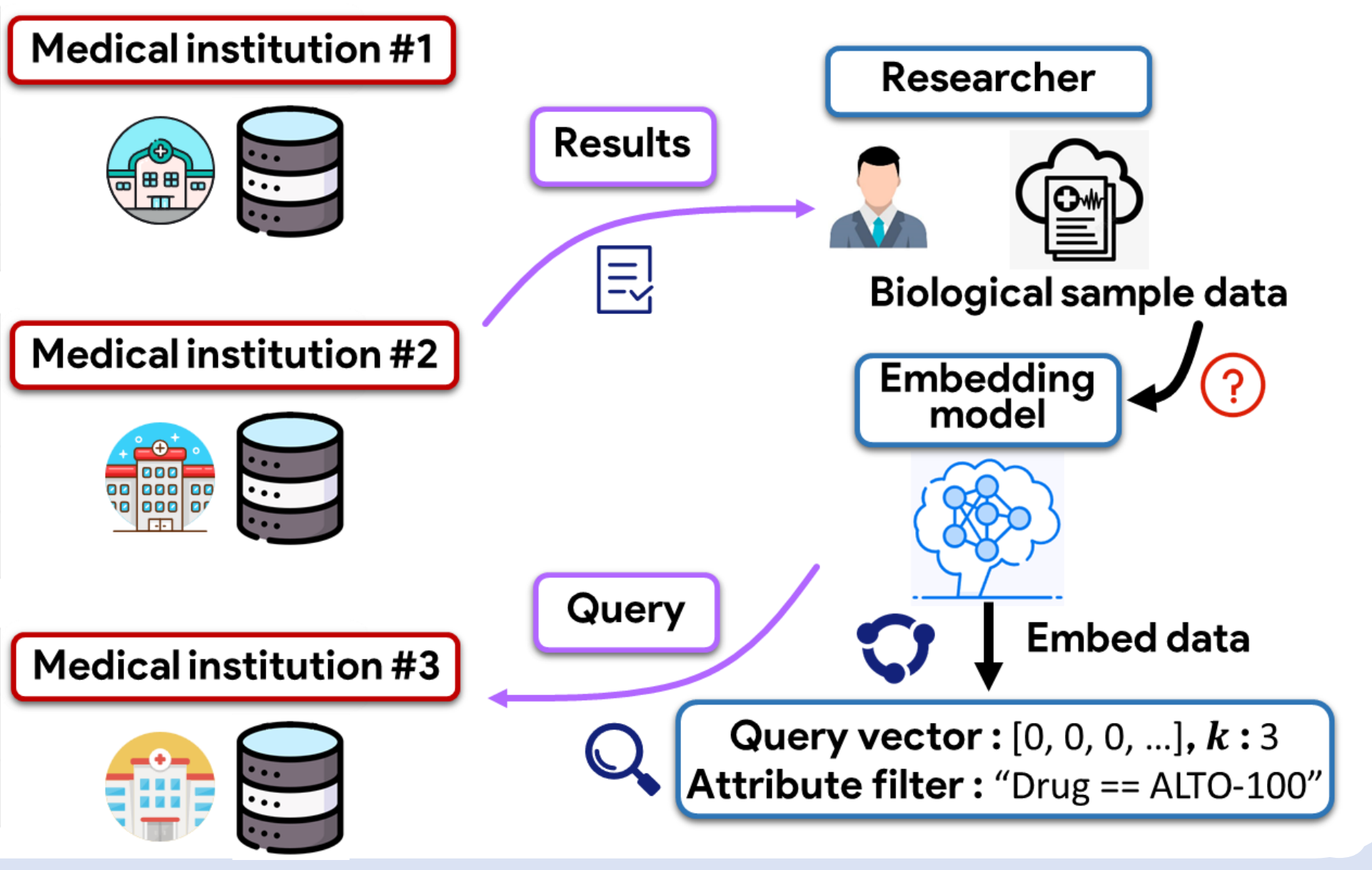
Existing methods for federated kNN search can be extended to solve this challenge. These methods adopt either encryption [1] or secure multi-party computation [2] to securely find kNNs to a given query object. However, encryption-based methods are computationally expensive. The other methods [3], which were originally designed for 2D locations or sequence data, exhibit inefficiency or low recall when handling high-dimensional vectors. Thus, our core challenge can be summarized as:

strike a balance between effectiveness and efficiency while ensuring privacy preservation

ID	Embedding	Drug
6	[0.2, 0, 0.1, ...]	ALTO-100
7	[0.1, 0, 0.1, ...]	5-MAPB
8	[0.2, 0.3, 0.3, ...]	ALTO-100

ID	Embedding	Drug
3	[0.1, 0.1, 0.1, …]	ALTO-100
4	[0, 0.1, 0.1, …]	ALTO-100
5	[0.2, 0.1, 0.2, …]	ALTO-100

ID	Embedding	Drug
1	[0.1, 0.2, 0, ...]	5-MAPB
2	[0, 0, 0.1, ...]	ALTO-100



Query vector : [0, 0, 0, ...], k : 3
Attribute filter : "Drug == ALTO-100"

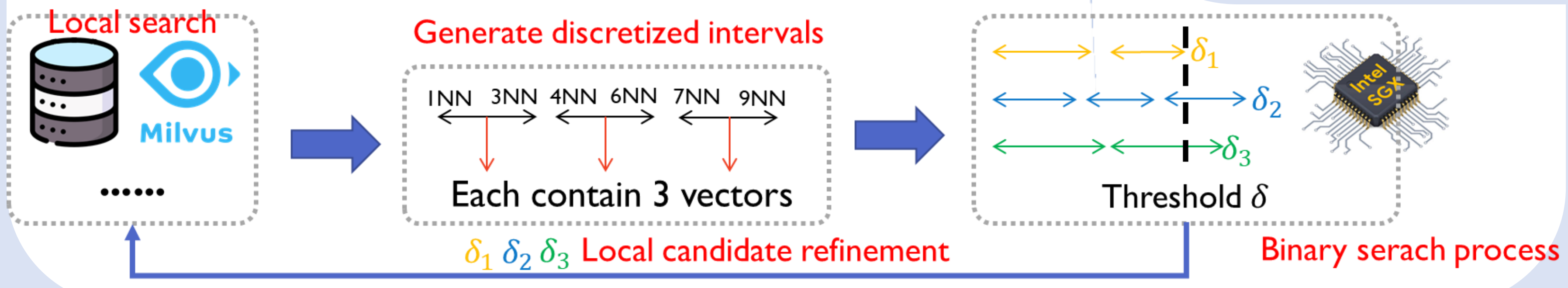
2 Algorithm

Our framework is structured into two phases:

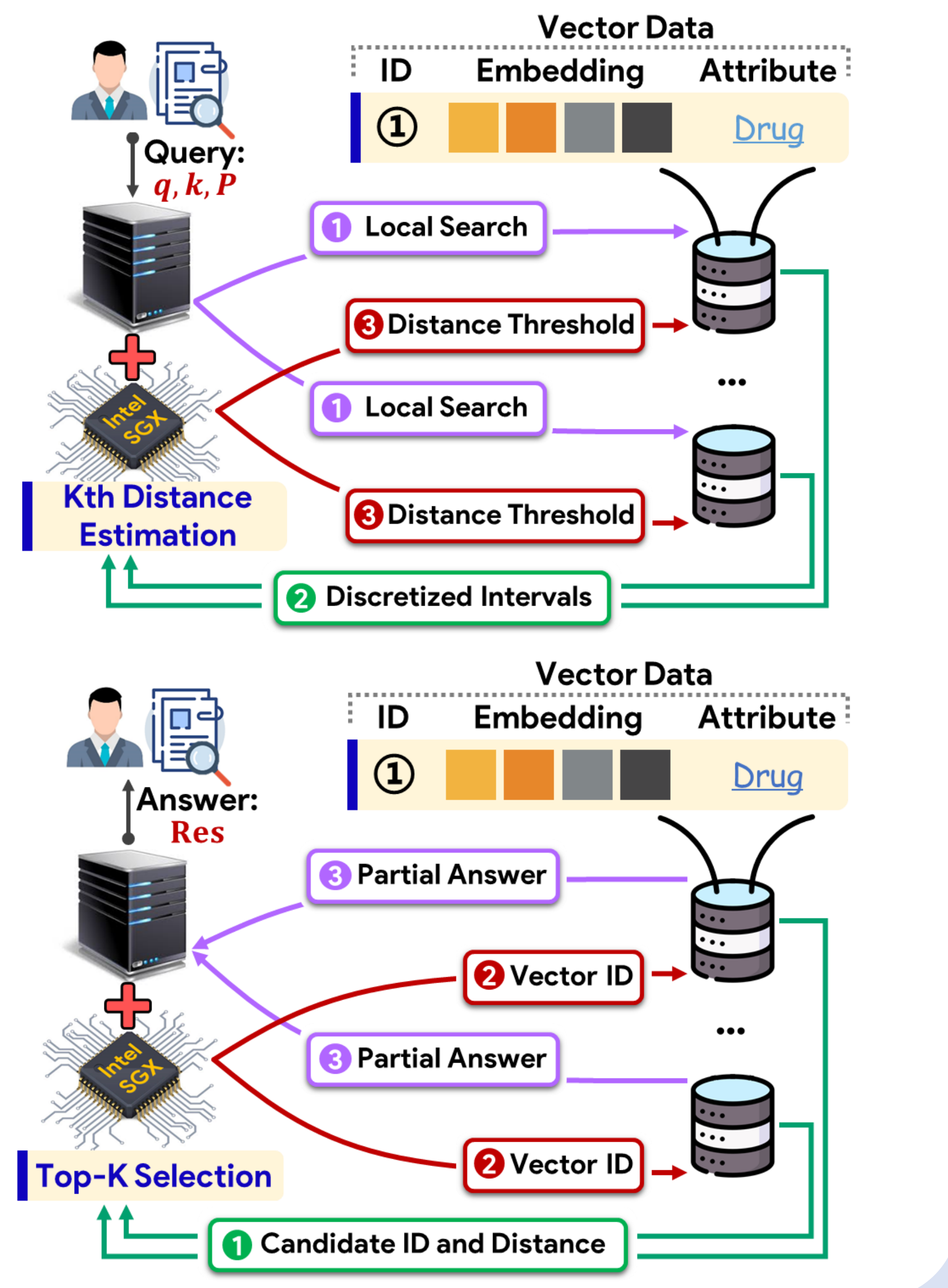
(i) **Federated Candidate Refinement**.

(ii) **Federated Top-K Selection**.

- □ Phase I: derive a threshold for upper bound of the Kth nearest distance
  - □ Data Provider: Local search and discretize candidates into $\sqrt{k}$ intervals
  - □ Central Server (TEE): calculate threshold with binary search based on intervals
- □ Phase II: select top-k from refined candidates
  - □ Data Provider: eliminate candidates with distance larger than threshold
  - □ Central Server (TEE): top-k selection with an (oblivious) priority queue

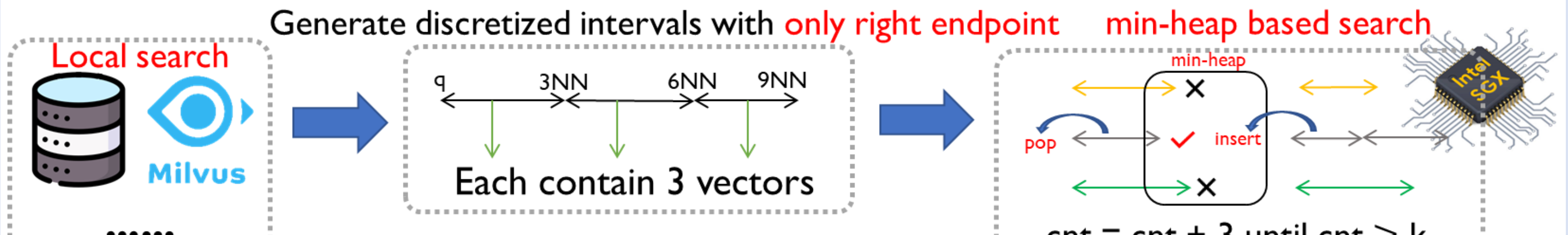


Two-phase framework



3 Optimization

□ Optimization #1 : **Reducing communication cost**

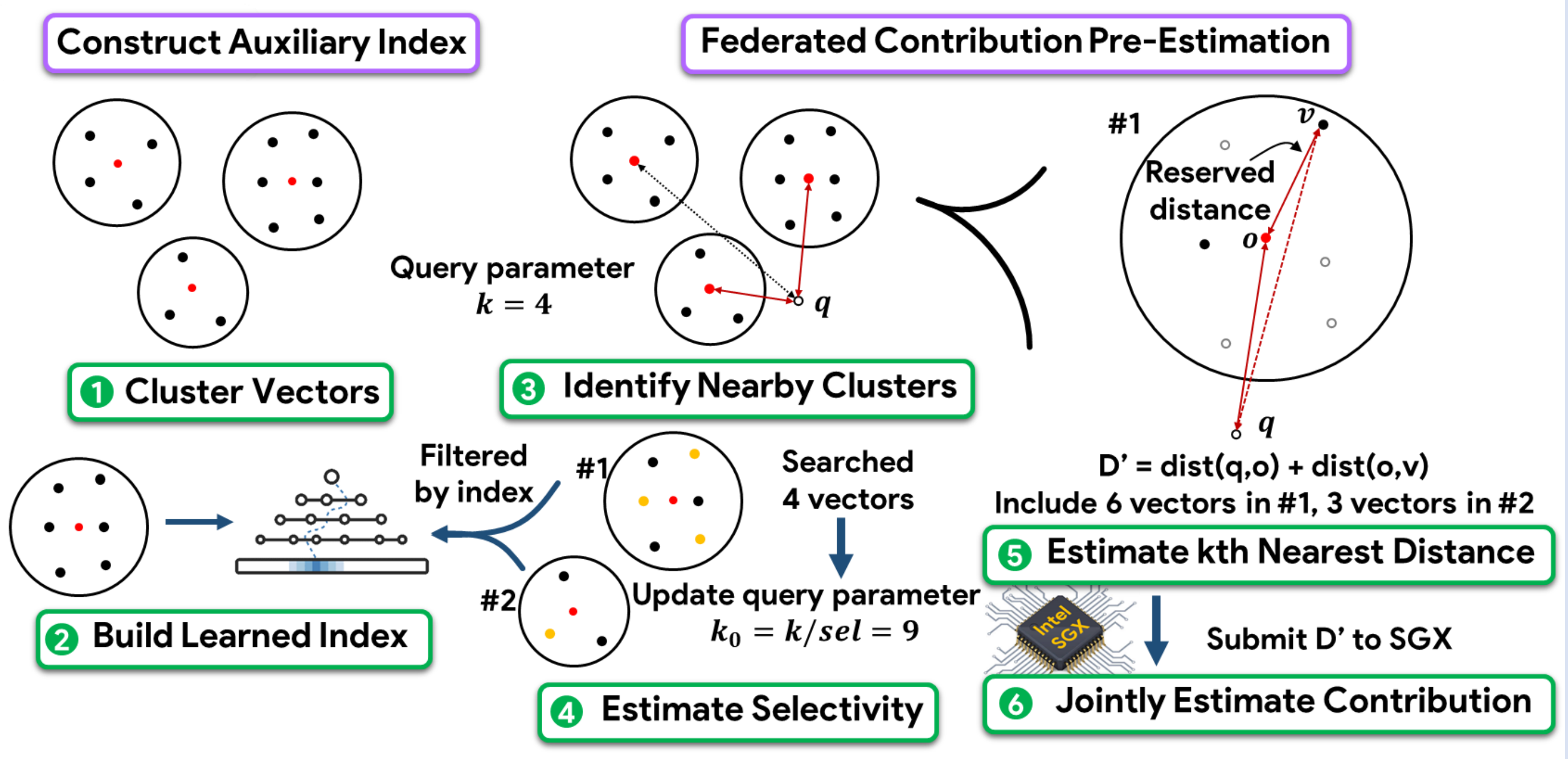


Optimization over phase I

□ Optimization #2 : **Pruning via contribution estimation**

Optimization over phase II

Contribution estimation through **Cluster-based Learned Index**

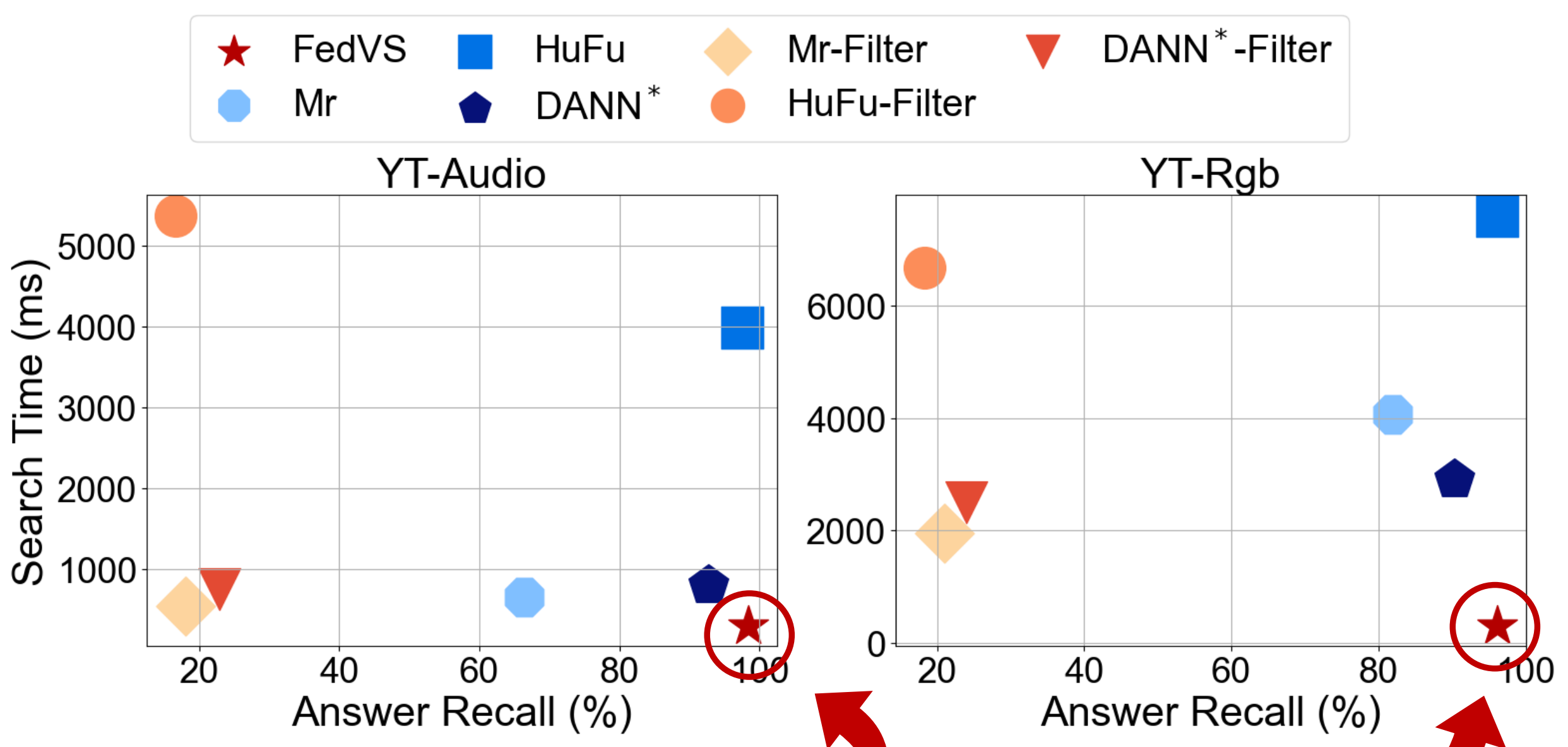

Construct Auxiliary Index
Federated Contribution Pre-Estimation
1 Cluster Vectors
2 Build Learned Index
Query parameter
$k = 4$
q
3 Identify Nearby Clusters
Filtered by index
#1
#2
Searched 4 vectors
Update query parameter
$k_0 = k/sel = 9$
4 Estimate Selectivity
#1
Reserved distance
v
o
q
D' = dist(q,o) + dist(o,v)
Include 6 vectors in #1, 3 vectors in #2
5 Estimate kth Nearest Distance
Intel SGX
Submit D' to SGX
6 Jointly Estimate Contribution

4 Experiment

We implement our algorithms with **industrial vector database Milvus** [4] and compare query performance against six baselines extended from state-of-the-art methods on four datasets.

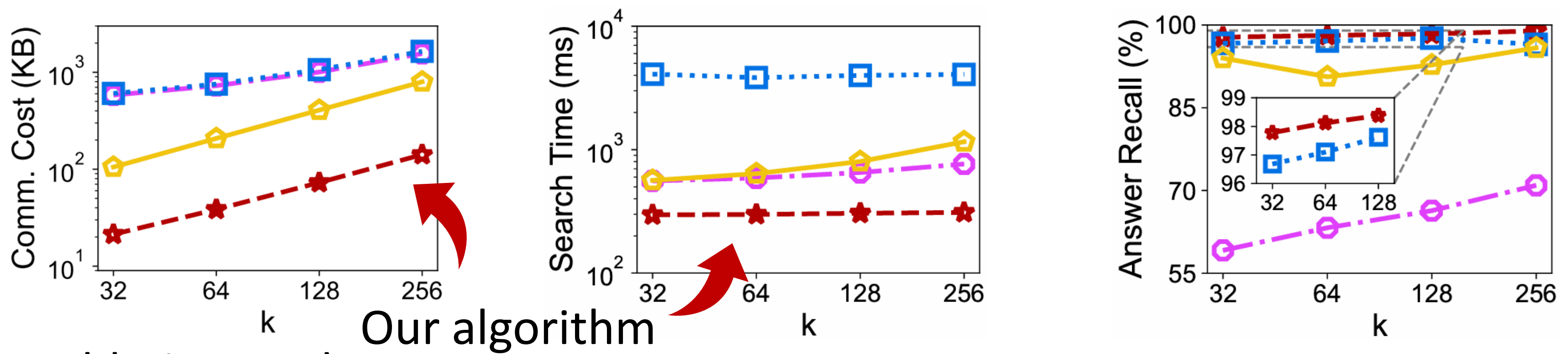
Dataset	Card.	Dim.	Embedding	Attribute	Partition
WIT	5×10^4	2048	Image	Image Size	IID
YT-Audio	10^6	128	Audio	Category	Dirichlet
YT-Rgb	10^6	1024	Video	Category	Dirichlet
DEEP	10^7	96	Image	Synthetic	Quantity

□ Overall Performance



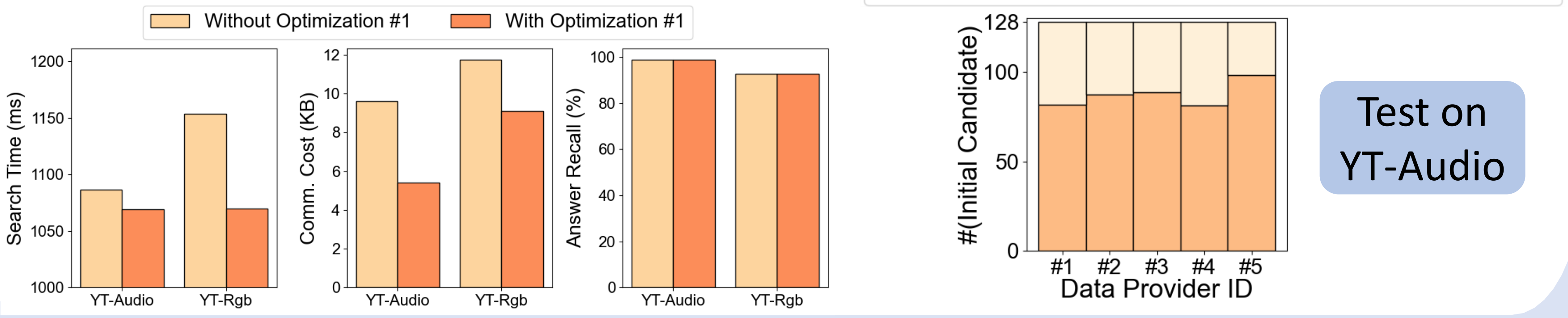
Our algorithm

□ Impact of query parameter k on YT-Audio



Our algorithm

□ Ablation study



Test on YT-Audio

5 References

[1] Kesarwani M, Kaul A, Naldurg P, et al. Efficient Secure k-Nearest Neighbours over Encrypted Data[C]//EDBT. 2018: 564-575.

[2] Yongxin Tong, Xuchen Pan, Yuxiang Zeng, et al. 2022. Hu-Fu: Efficient and Secure Spatial Queries over Data Federation. PVLDB 15, 6 (2022), 1159–1172.

[3] Kaining Zhang, Yongxin Tong, Yexuan Shi, et al. 2023. Approximate k-Nearest Neighbor Query over Spatial Data Federation. In DASFAA. 351–368.

[4] 2025. Milvus. https://milvus.io/

6 Acknowledgment

*This work was partially supported by National Key Research and Development Program of China under Grant No. 2023YFF0725103,
National Science Foundation of China (NSFC) (Grant Nos. 62425202, U21A20516, 62336003),
the Beijing Natural Science Foundation (Z230001),
the Fundamental Research Funds for the Central Universities No. JK2024-03,
the Didi Collaborative Research Program and
the State Key Laboratory of Complex & Critical Software Environment (SKLCCSE)*

KDD2025