

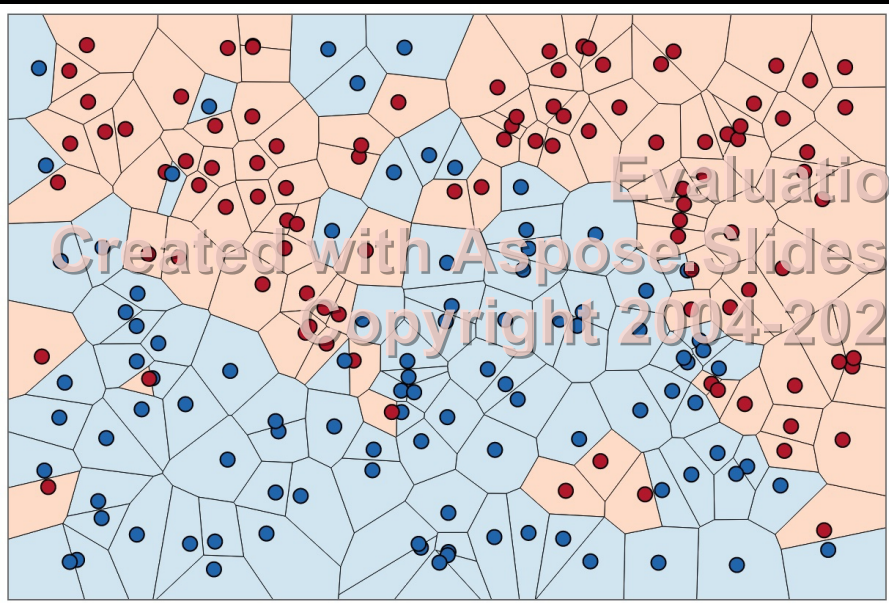
Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.

Copyright 2004-2025 Aspose Pty Ltd.



SMILE IDENTITY



Evaluation only.

Created with Aspose Slides for Python via .NET 25.2.

Copyright 2004-2025 Aspose Pty Ltd.



SMILE IDENTITY



Microsoft



Meta

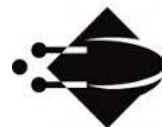
CMU

Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.

Copyright 2004-2025 Aspose Pty Ltd.

Yandex



IT University
of Copenhagen



SMILE IDENTITY

Evaluation only.
Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025Aspose Pty Ltd.



SMILE IDENTITY

erikbern / ann-benchmarks Public

Watch 96 Fork 456 Star 2.8k

<> Code Issues 39 Pull requests 8 Actions Projects Wiki Security Insights

master 6 branches 0 tags

maumuelle Add export to csv (closes #256). (#269) 2 months ago

.github/workflows Add export to csv (closes #256). (#269) 2 months ago

ann_benchmarks Add export to csv (closes #256). (#269) 2 months ago

install Add Vearch to ANN Benchmarks 3 months ago

protocol Fix some flake8 issues 3 years ago

results rebased 4 months ago

templates fix website building 17 months ago

test Make Puffinn algorithm compatible with sparse dataset format 11 months ago

About

Benchmarks of approximate nearest neighbor libraries in Python

ann-benchmarks.com

docker benchmark nearest-neighbors

Readme

MIT License

2.8k stars

96 watching

456 forks

<http://github.com/erikbern/ann-benchmarks>



SMILE IDENTITY

Evaluated

- Annoy
- FLANN
- scikit-learn: LSHForest, KDTree, BallTree
- PANNNS
- NearPy
- KGraph
- NMSLIB (Non-Metric Space Library): SWGraph, HNSW, BallTree, MPLSH
- hnswlib (a part of nmslib project)
- RPF
- FAISS
- DolphinnPy
- Datasketch
- PyNNDescend
- MRPT
- NGT: ONNG, PANNG, QG
- SPTAG
- PUFFINN
- N2
- ScaNN
- Elastiknn
- OpenSearch KNN
- DiskANN: Vamana, Vamana-PQ
- Vespa
- scipy: cKDTree
- vald

Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025 Aspose Pty Ltd.

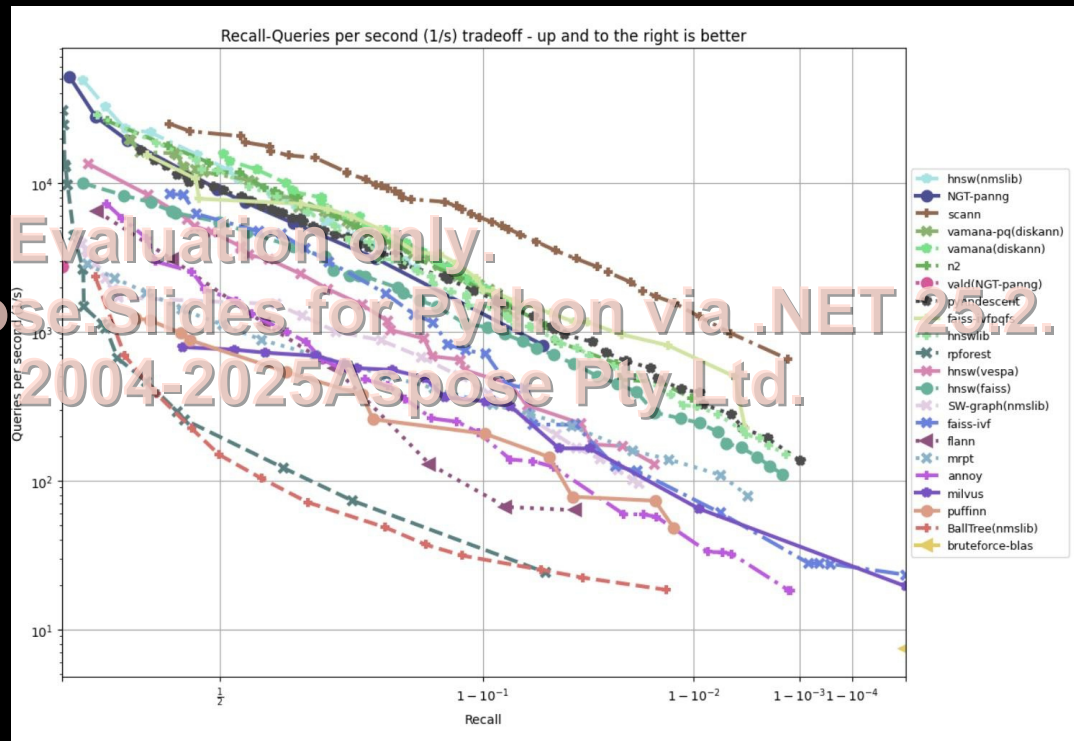
<http://github.com/erikbern/ann-benchmarks>



SMILE IDENTITY

Evaluated

- Annoy
- FLANN
- scikit-learn: LSHForest, KDTree, BallTree
- PANNNS
- NearPy
- KGraph
- NMSLIB (Non-Metric Space Library): SWGraph, HNSW, BallTree, MPLSH
- hnswlib (a part of nmslib project)
- RPF (res)
- FAISS
- DolphinnPy
- Datasketch
- PyNNDescend
- MRPT
- NGT: ONNG, PANNG, QG
- SPTAG
- PUFFINN
- N2
- ScaNN
- Elastiknn
- OpenSearch KNN
- DiskANN: Vamana, Vamana-PQ
- Vespa
- scipy: cKDTree
- vald



Data sets

We have a number of precomputed data sets for this. All data sets are pre-split into train/test and come with ground truth data in the form of the top 100 neighbors. We store them in a HDF5 format:

Dataset	Dimensions	Train size	Test size	Neighbors	Distance	Download
DEEP1B	96	9,991,000	10,000	100	Angular	HDF5 (3.6GB)
Fashion-MNIST	784	60,000	10,000	100	Euclidean	HDF5 (217MB)
Fast	256	1,000,000	10,000	100	Euclidean	HDF5 (3.3GB)
GloVe	25	1,183,514	10,000	100	Angular	HDF5 (121MB)
GloVe	50	1,183,514	10,000	100	Angular	HDF5 (235MB)
GloVe	100	1,183,514	10,000	100	Angular	HDF5 (463MB)
GloVe	200	1,183,514	10,000	100	Angular	HDF5 (918MB)
Kosarak	27983	74,962	500	100	Jaccard	HDF5 (2.0GB)
MNIST	784	60,000	10,000	100	Euclidean	HDF5 (217MB)
NYTimes	256	290,000	10,000	100	Angular	HDF5 (301MB)
SIFT	128	1,000,000	10,000	100	Euclidean	HDF5 (501MB)
Last.fm	65	292,385	50,000	100	Angular	HDF5 (135MB)

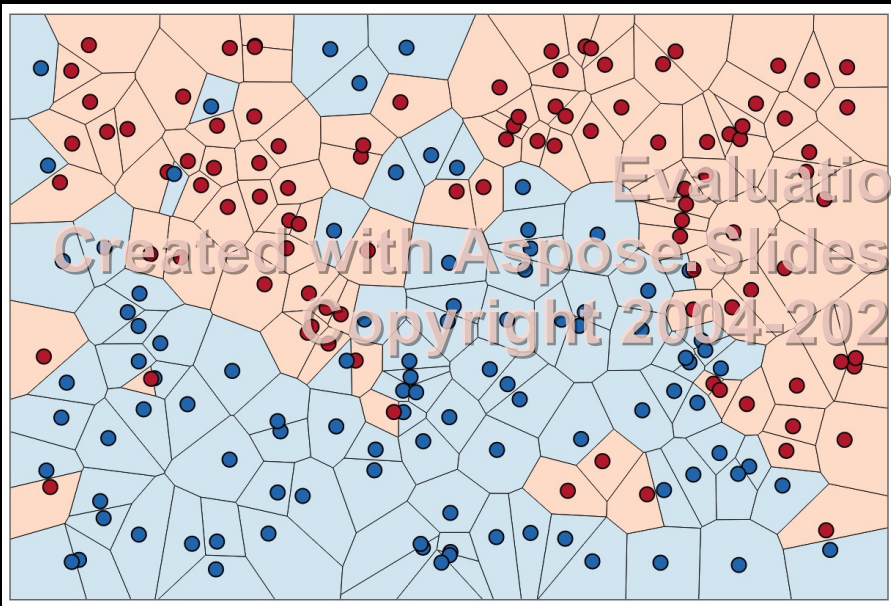


Data sets

We have a number of precomputed data sets for this. All data sets are pre-split into train/test and come with ground truth data in the form of the top 100 neighbors. We store them in a HDF5 format:

Dataset	Dimensions	Train size	Test size	Neighbors	Distance	Download
DEEP1B	96	9,991,000	10,000	100	Angular	HDF5 (3.6GB)
Fashion-MNIST	784	60,000	10,000	100	Euclidean	HDF5 (217MB)
Fast	280	1,000,000	10,000	100	Euclidean	HDF5 (3.3GB)
GloVe	25	1,183,514	10,000	100	Angular	HDF5 (121MB)
GloVe	50	1,183,514	10,000	100	Angular	HDF5 (235MB)
GloVe	100	1,183,514	10,000	100	Angular	HDF5 (463MB)
GloVe	200	1,183,514	10,000	100	Angular	HDF5 (918MB)
Kosarak	27983	74,962	500	100	Jaccard	HDF5 (2.0GB)
MNIST	784	60,000	10,000	100	Euclidean	HDF5 (217MB)
NYTimes	256	290,000	10,000	100	Angular	HDF5 (301MB)
SIFT	128	1,000,000	10,000	100	Euclidean	HDF5 (501MB)
Last.fm	65	292,385	50,000	100	Angular	HDF5 (135MB)





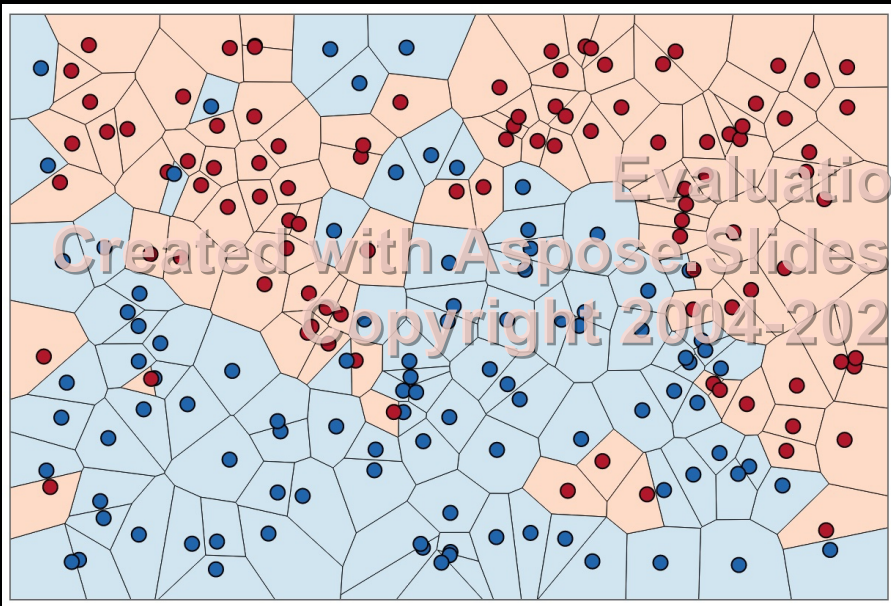
Evaluation only.

Created with Aspose Slides for Python via .NET 25.2.

Copyright 2004-2025 Aspose Pty Ltd.



SMILE IDENTITY



Evaluation only.

Created with Aspose Slides for Python via .NET 25.2.

Copyright 2004-2025 Aspose Pty Ltd.



SMILE IDENTITY

Evaluation only.
Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025Aspose Pty Ltd.



SMILE IDENTITY

Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.

BigMig

2011

**Jegou, Tavenard,
Douze, Laurent (INRIA)**

Deep1B

2016

Babenko, Lempitsky (CVPR)

BigANN
2011

Deep1B
2016

Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025 Aspose Pty Ltd.

SSNPP
2021

MSTuring
2021

MSSpaceV
2021

Text2Image
2021



SMILE IDENTITY

BigANN
2011

Deer1B
2016

Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025 Aspose Pty Ltd.

SSNPP
2021



MSTuring
2021



MSSpaceV
2021



Text2Image
2021



SMILE IDENTITY

	Dataset	Type	d	Dist.	Query Type
	BIGANN	uint8	128	L2	k-NN
Facebook	SimSearchNet++	uint8	256	L2	Range
Microsoft	Turing ANNS	float32	100	L2	k-NN
	Microsoft SpaceV	int8	100	L2	k-NN
	DEEP	float32	96	L2	k-NN
Yandex	Text-to-Image	float32	200	IP	k-NN

Evaluation only.
Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025Aspose Pty Ltd.



SMILE IDENTITY

Track #1

Datcenter Server

64 GB RAM



Track #2
Evaluation only.

Track #3



SMILE IDENTITY

Track #1

Datcenter Server
64 GB RAM

Track #2

Datcenter Server
64 GB RAM
2TB HTD SSD

Track #3



SMILE IDENTITY

Track #1

Datacenter Server

64 GB RAM

Track #2

Datacenter Server

64 GB RAM

2TB HTD SSD

Track #3

Any Hardware

Evidence of Cost

Power Monitoring



SMILE IDENTITY

Track #1

Datcenter Server
64 GB RAM

Track #2

Datcenter Server
64 GB RAM
2TB HTD SSD

Track #3

Any Hardware
Evidence of Cost
Power Monitoring

- Limited Hardware Resource Budget
- Ranked By ***Recall***



Track #1

Datcenter Server
64 GB RAM



- Limited Hardware Resource Budget
- Ranked By **Recall**

Track #2

Datcenter Server
64 GB RAM
2TB HTD SSD



Track #3

Any Hardware
Evidence of Cost
Power Monitoring



- **Recall**, QPS,
Power, Cost



Evaluation only.
Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025Aspose Pty Ltd.



SMILE IDENTITY



Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025 Aspose Pty Ltd.

Size	vCPU's	Memory: GiB	Temp storage (SSD) GiB	Max data disks	Max cached and temp storage throughput: IOPS/MBps (cache size in GiB)	Max uncached disk throughput: IOPS/MBps	Max burst uncached disk throughput: IOPS/MBps
Standard_F4s_v2	4	8	32	4	4000/31 (32)	3200/47	4000/200
Standard_F8s_v2	8	16	64	8	8000/62 (64)	6400/94	8000/400
Standard_F16s_v2	16	32	128	16	16000/127 (128)	12800/190	16000/400
Standard_F32s_v2	32	64	256	32	32000/255 (256)	25600/380	32000/800
Standard_F48s_v2	48	96	384	48	48000/384 (384)	38400/512	48000/1000
Standard_F64s_v2	64	128	512	64	64000/512 (512)	51200/750	64000/1600
Standard_F72s_v2	72	144	576	72	72000/576 (576)	57600/750	72000/1600

FSv2 Compute Optimized Series



SMILE IDENTITY



Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025 Aspose Pty Ltd.

Size	vCPU's	Memory: GiB	Temp storage (SSD) GiB	Max data disks	Max cached and temp storage throughput: IOPS/MBps (cache size in GiB)	Max uncached disk throughput: IOPS/MBps	Max burst uncached disk throughput: IOPS/MBps
Standard_F4s_v2	4	8	32	4	4000/31 (32)	3200/47	4000/200
Standard_F8s_v2	8	16	64	8	8000/62 (64)	6400/94	8000/400
Standard_F16s_v2	16	32	128	16	16000/127 (128)	12800/190	16000/400
Standard_F32s_v2	32	64	256	32	32000/255 (256)	25600/380	32000/800
Standard_F48s_v2	48	96	384	48	48000/384 (384)	38400/512	48000/1000
Standard_F64s_v2	64	128	512	64	64000/512 (512)	51200/750	64000/1600
Standard_F72s_v2	72	144	576	72	72000/576 (576)	57600/750	72000/1600
Standard_F96s_v2	96	192	768	96	96000/768 (768)	76800/1100	80000/2000
Standard_F128s_v2	128	256	1024	128	128000/1024 (1024)	80000/1100	80000/2000
Standard_F160s_v2	160	320	1280	160	160000/1280 (1280)	80000/1100	80000/2000
Standard_F224s_v2	224	448	1792	224	224000/1792 (1792)	80000/1100	80000/2000

FSv2 Compute Optimized Series



SMILE IDENTITY



Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.

Copyright 2004-2025 Aspose Pty Ltd.

Size	vCPU	Memory (GiB)	Temp disk ¹ (GiB)	NVMe Disks ²	NVMe Disk throughput ³ (Read IOPS/MBps)	Uncached data disk throughput (IOPS/MBps) ⁴	Max burst uncached data disk throughput (IOPS/MBps)
Standard_L32s_v2	32	256	320	4x1.92 TB	1.5M/8000	32000/640	32000/128
Standard_L16s_v2	16	128	160	2x1.92 TB	800000/4000	16000/320	16000/128
Standard_L32s_v2	32	256	320	4x1.92 TB	1.5M/8000	32000/640	32000/128
Standard_L48s_v2	48	384	480	6x1.92 TB	2.2M/14000	48000/960	48000/200
Standard_L64s_v2	64	512	640	8x1.92 TB	2.9M/16000	64000/1280	64000/200
Standard_L80s_v2 ⁶	80	640	800	10x1.92TB	3.8M/20000	80000/1400	80000/200

Lsv2 Storage Optimized w/High Throughput NVM SSD



SMILE IDENTITY



Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025 Aspose Pty Ltd.



SMILE IDENTITY



Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.

Copyright 2004-2025 Aspose Pty Ltd.



SMILE IDENTITY



Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025 Aspose Pty Ltd.



SMILE IDENTITY



Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025 Aspose Pty Ltd.



SMILE IDENTITY



Evaluation only.
Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025Aspose Pty Ltd.

“This Track Is Wild!”



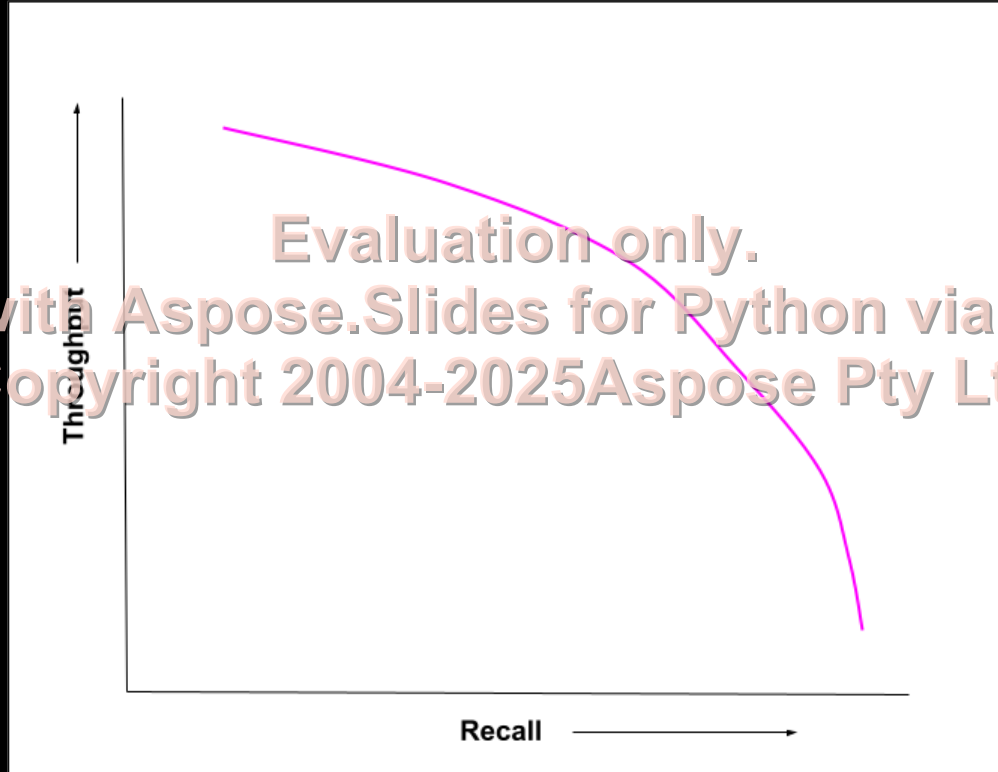
SMILE IDENTITY

Evaluation only.
Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025Aspose Pty Ltd.



SMILE IDENTITY

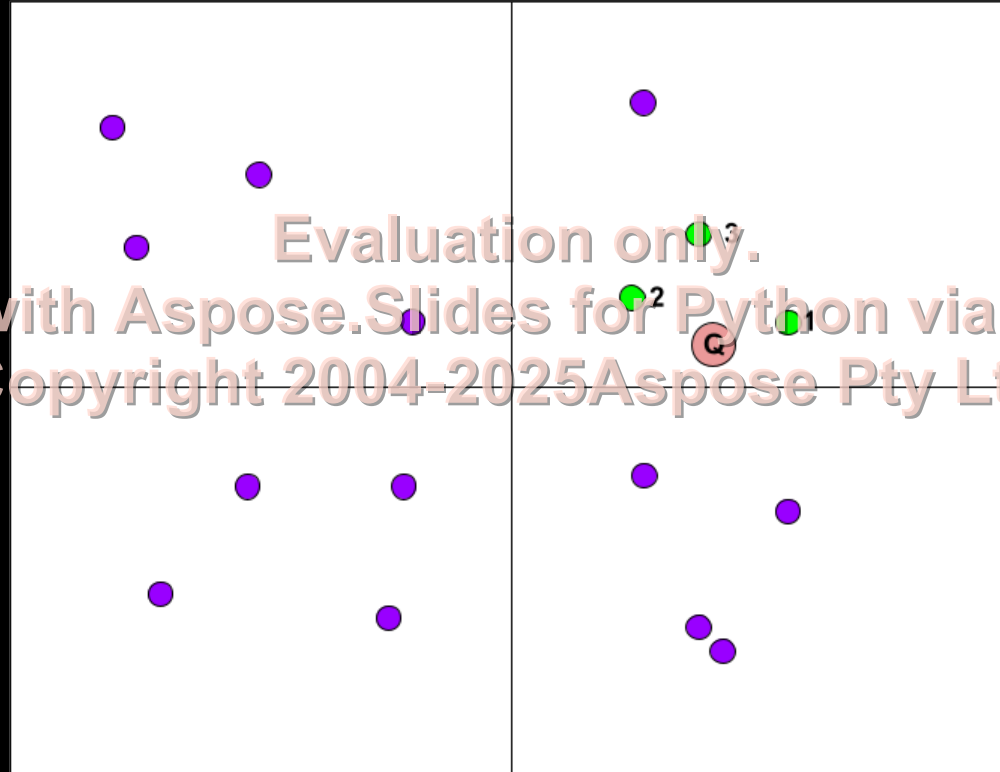
Evaluation only.
Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025 Aspose Pty Ltd.



SMILE IDENTITY





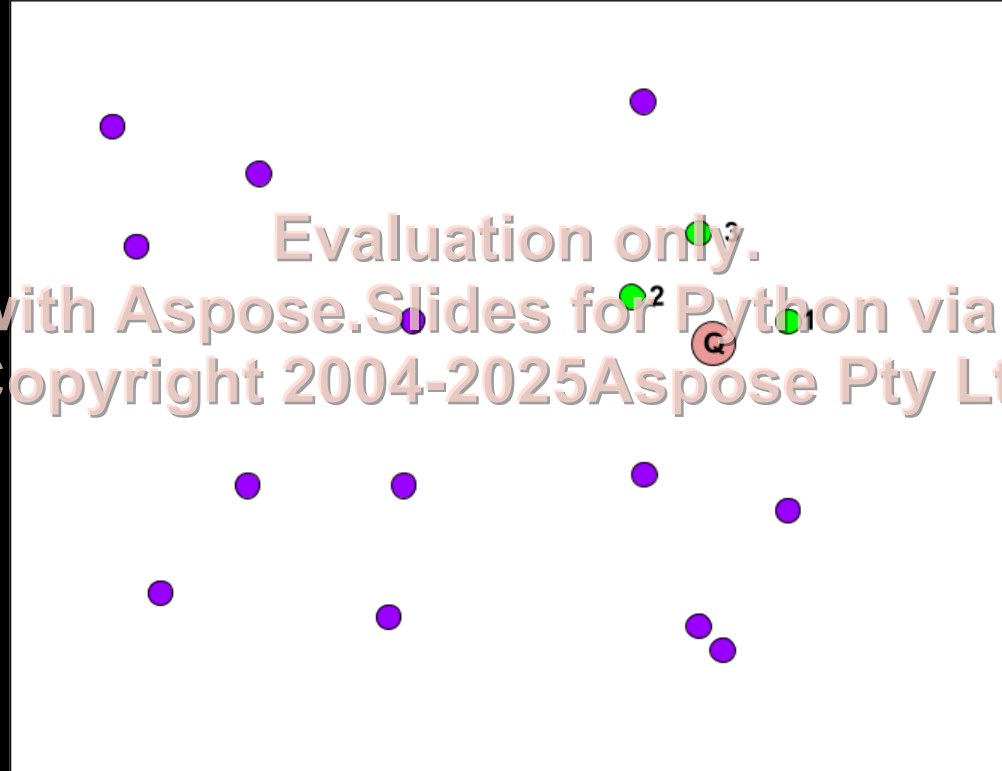


Evaluation only.
Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025 Aspose Pty Ltd.

$N = 3$, Recall = 3/3



SMILE IDENTITY



Evaluation only.
Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025 Aspose Pty Ltd.

N = 3, Recall = 3/3

N = 3, Recall = 3/3



SMILE IDENTITY

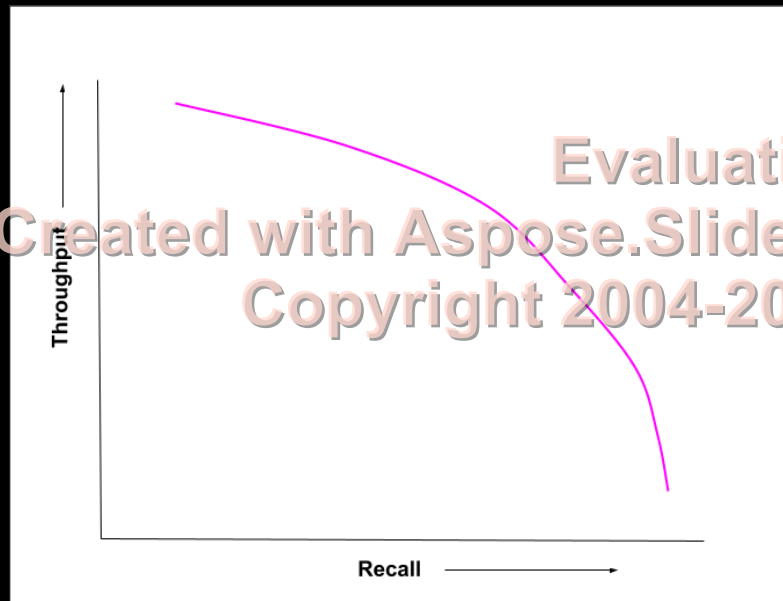


N = 3, Recall = 2/3

N = 3, Recall = 2/3



SMILE IDENTITY



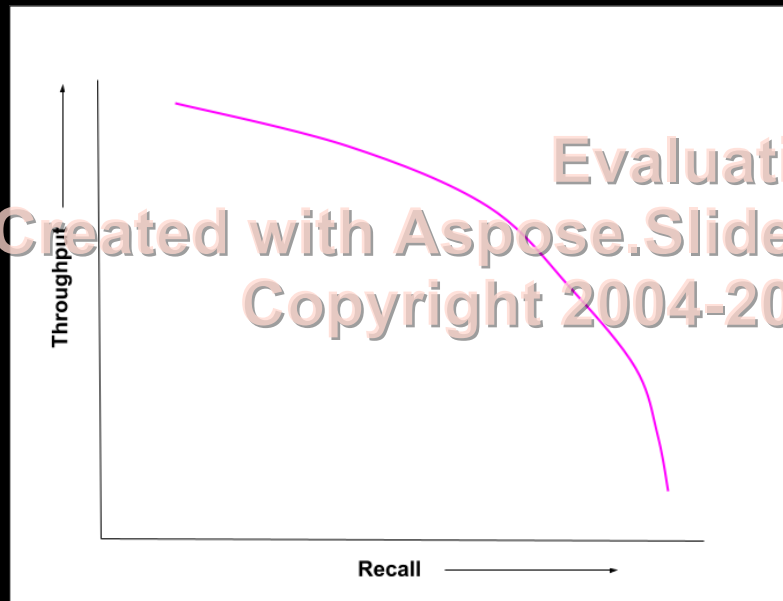
Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.

Copyright 2004-2025 Aspose Pty Ltd.



SMILE IDENTITY



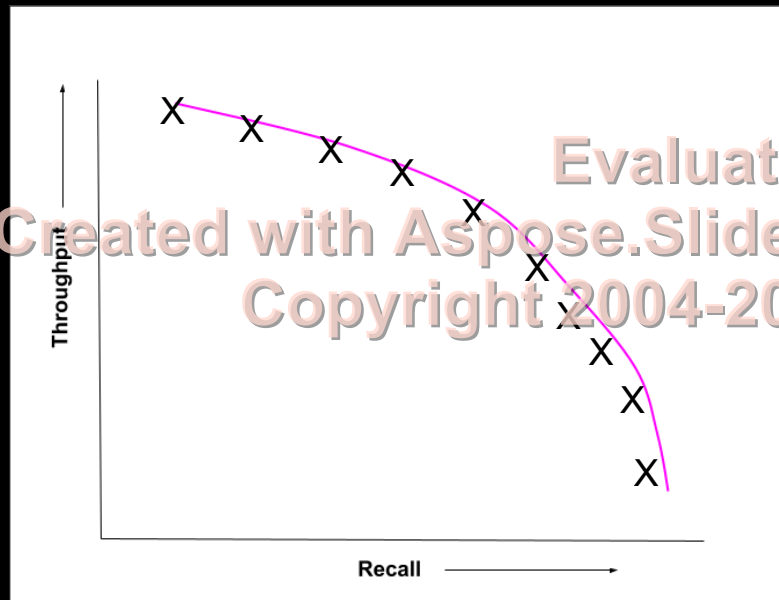
Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.

Copyright 2004-2025 Aspose Pty Ltd.



SMILE IDENTITY



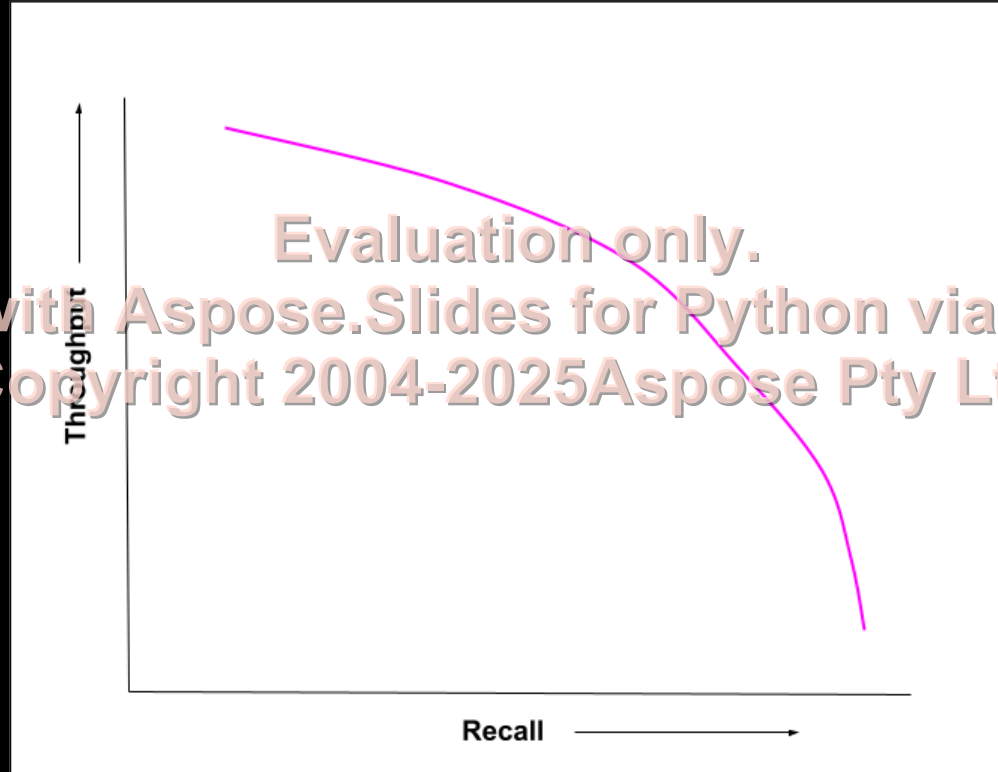
Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.

Copyright 2004-2025 Aspose Pty Ltd.



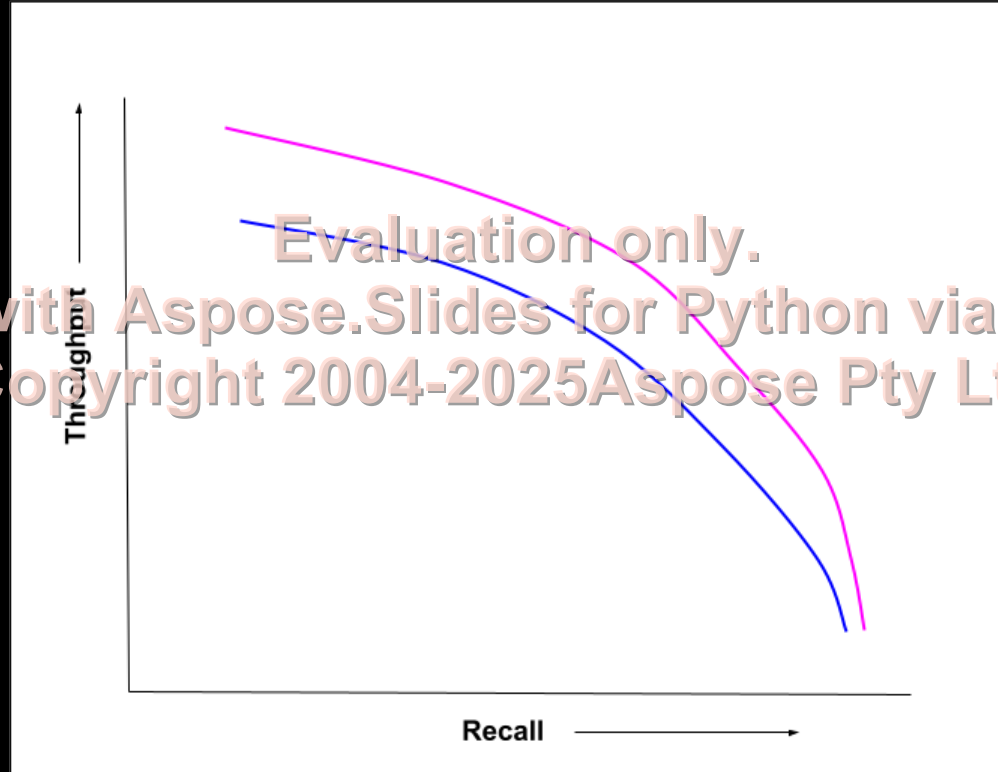
SMILE IDENTITY

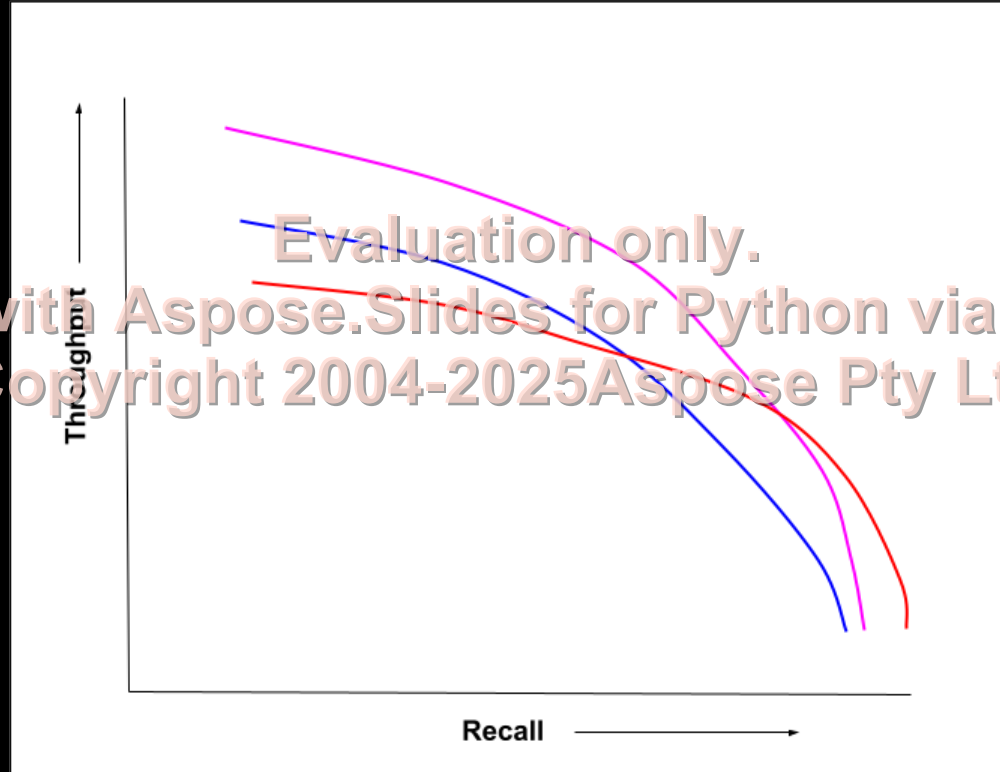


Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025 Aspose Pty Ltd.



SMILE IDENTITY





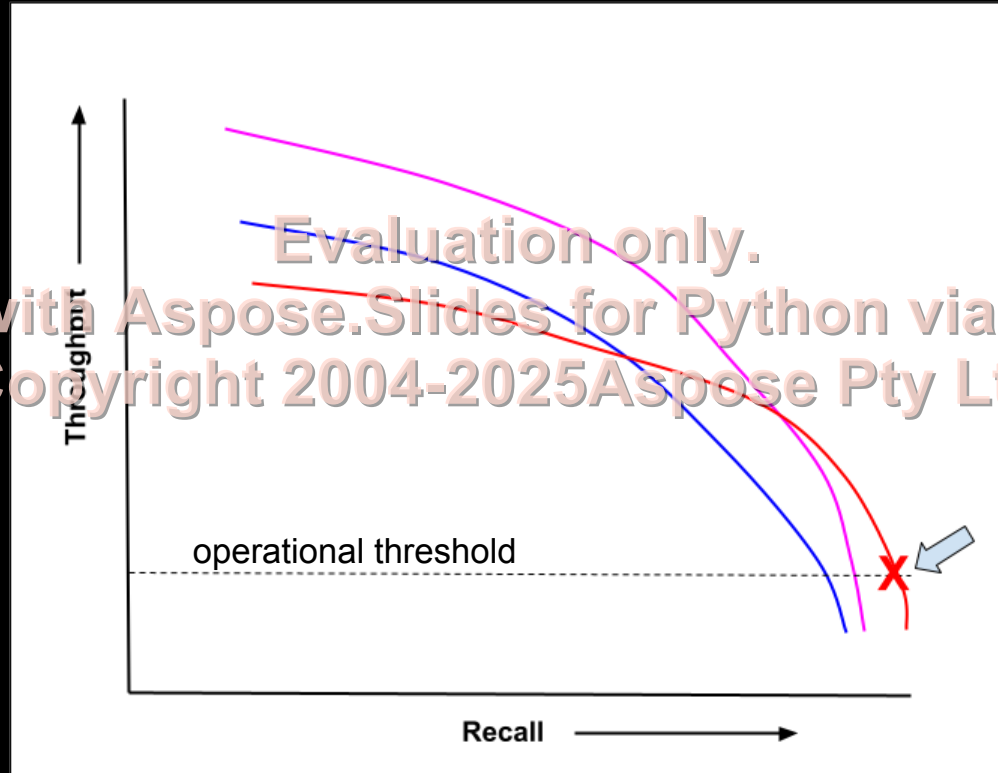
Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025 Aspose Pty Ltd.



SMILE IDENTITY

Evaluation only.
Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025 Aspose Pty Ltd.



Track	Operational Threshold
#1	10,000 QPS
#2	1,500 QPS
#3	2,000 QPS



Evaluation only.
Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025Aspose Pty Ltd.



SMILE IDENTITY

Track	Baseline Algorithm
#1	Evaluation Only.
#2	MS-DiskANN
#3	Faiss-IVFPQ + 1 GPU

Created with Aspose.Slides for Python via .NET 25.2.

Copyright 2004-2025 Aspose Pty Ltd.

- * Baseline algorithms provided by organizers
- * Final ranking based on cumulative improvements over baseline



SMILE IDENTITY

Evaluation only.
Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025Aspose Pty Ltd.



SMILE IDENTITY

Track	Benchmark	Winners
#1	Recall@10	Kuo's Python+Tring+NET
#2	Recall@10	Zilliz+SUST
#3	Recall@10	Intel on Optane



Winner: Kuaishou+Tsinghua U (15 participants)

Team	BigANN	Deep	MSpaceV	MSJuring	Text2Image	SSNPP
K+T U	0.711	0.723	0.7645	0.756	-	-

* threshold = 10,000 QPS



SMILE IDENTITY

Winner: Kuaishou+Tsinghua U (15 participants)

Team	BigANN	Deep	MSpaceV	MSJuring	Text2Image	SSNPP
K+T U	0.711	0.723	0.7645	0.756	-	-
Faiss-IVFPQ	0.635	0.650	0.729	0.704	0.069	0.754

* threshold = 10,000 QPS



Winner: Zilliz+SUST (5 Participants)

Algo	BigANN	Deep	MSpaceV	MSJuring	Text2Image	SSNPP
Zilliz+SUST			0.760		0.495	0.886

* threshold = 1,500 QPS



SMILE IDENTITY

Winner: Zilliz+SUST (5 Participants)

Algo	BigANN	Deep	MSpaceV	MSJuring	Text2Image	SSNPP
Zilliz+SUST	0.950	0.937	0.900	0.936	0.489	0.886
MS-DISKANN	0.950	0.937	0.900	0.936	0.489	0.163

* threshold = 1,500 QPS



SMILE IDENTITY

Winner: Intel Optane (8 Participants)

Algo	BigANN	Deep	MSpaceV	MSJuring	Text2Image	SSNPP
Intel on Optane	0.9998	0.9998	0.9994	0.9957	0.9973	-

* threshold = 2,000 QPS



SMILE IDENTITY

Winner: Intel Optane (8 Participants)

Algo	BigANN	Deep	MSpaceV	MSJuring	Text2Image	SSNPP
Intel on Optane	0.9998	0.9998	0.9994	0.9957	0.9734	-
Faiss-IVFPQ on GPU	0.9326	0.9428	0.9085	0.9132	0.8603	0.9786

* threshold = 2,000 QPS



SMILE IDENTITY

Track	Benchmark	Winners
#1	Recall@10	Kuaishou+Tsinghua U
#2	Recall@10	Zilliz+USTC
#3	Recall@10	Intel on Optane
#3	Throughput (qps)	NVidia on 8 GPUs
#3	Power (Kwh / q)	Intel on Optane
#3	Cost (\$ / 100K qps * 4yr)	Intel on Optane

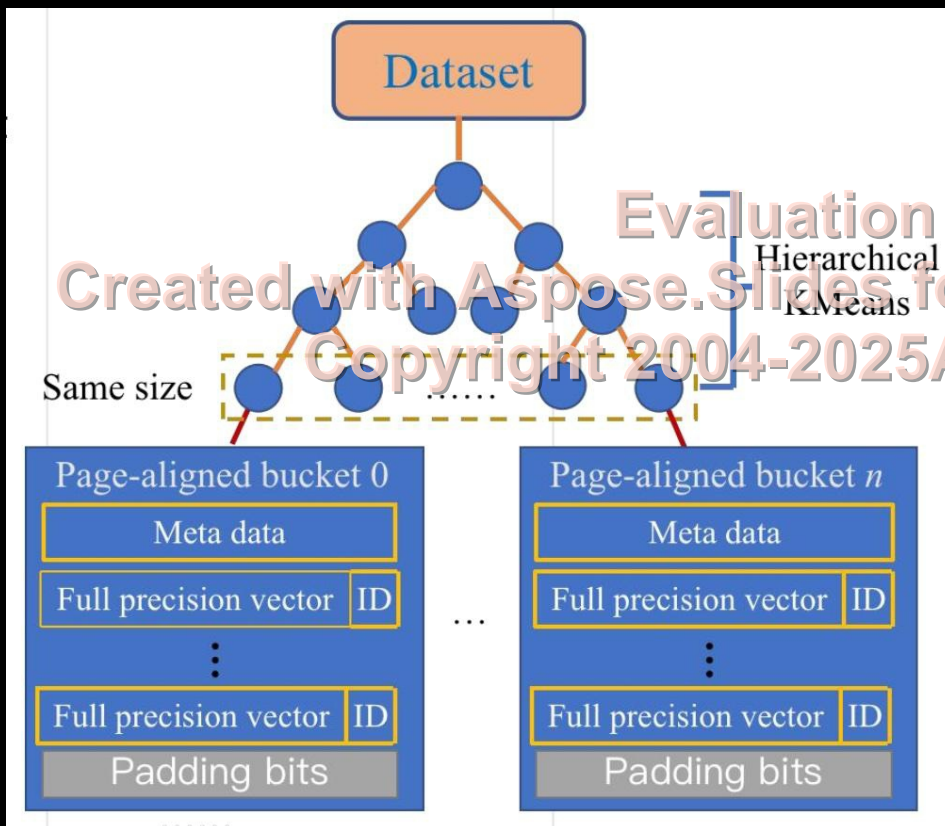
Evaluation only.
 Created with Aspose.Slides for Python via .NET 25.2.
 Copyright 2004-2025 Aspose Pty Ltd.



Evaluation only.
Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025Aspose Pty Ltd.



SMILE IDENTITY



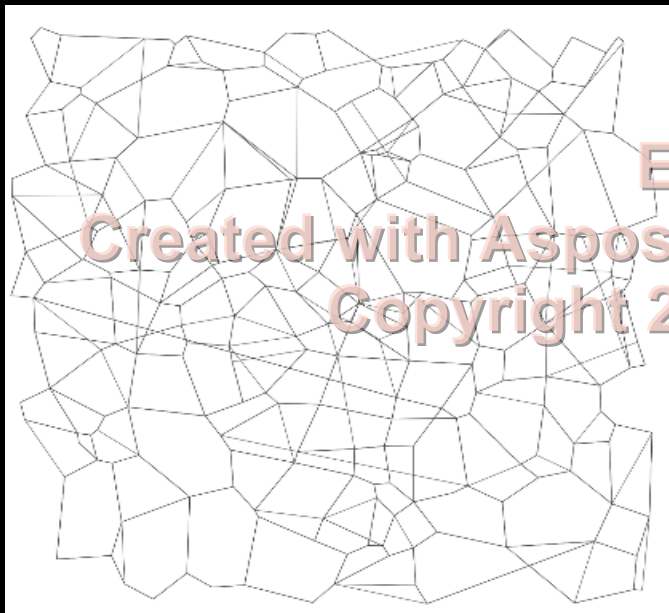
Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.

Copyright 2004-2025 Aspose Pty Ltd.



SMILE IDENTITY



graph construction similar to HNSW

DiskANN Design

Evaluation only.

Compressed vectors
(50-60GB/billion vectors)

Low-diameter graph
+full-precision vectors

DRAM

SSD



SMILE IDENTITY



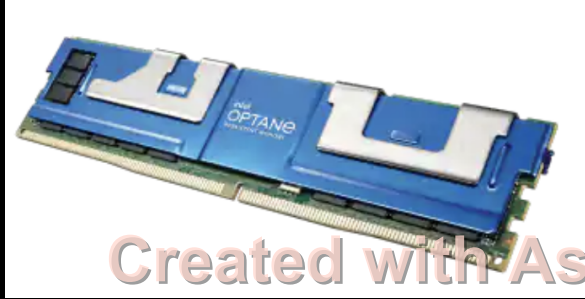
Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.

Copyright 2004-2025Aspose Pty Ltd.



SMILE IDENTITY



Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.

Copyright 2004-2025Aspose Pty Ltd.

	Team	BigANN	Deep	MSSpaceV	MSTuring	Text2Image	SSNPP
1	Intel	0.9998	0.9998	0.9984	0.9957	0.9973	-
2	Nvidia	0.9988	0.9954	0.9899	0.99443	0.9469	-



SMILE IDENTITY



Evaluation only.

**Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025Aspose Pty Ltd.**



SMILE IDENTITY



Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025Aspose Pty Ltd.

	Team	BigANN	Deep	MSSpaceV	MSTuring	Text2Image	SSNPP
1	Nvidia	747K	802K	840K	584K	-	-
2	Intel	336K	197K	158K	161K	17K	-



SMILE IDENTITY



Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.

Copyright 2004-2025Aspose Pty Ltd.



SMILE IDENTITY



Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.

Copyright 2004-2025Aspose Pty Ltd.

	Team	BigANN	Deep	MSSpaceV	MSTuring	Text2Image	SSNPP
1	Intel	0.0022	0.0041	0.0049	0.0048	0.0446	-
2	Nvidia	0.0119	0.0112	0.0090	0.0090	0.0480	-

* (watt-sec) / q shown here



SMILE IDENTITY



Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.

Copyright 2004-2025Aspose Pty Ltd.



SMILE IDENTITY



Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.

Copyright 2004-2025Aspose Pty Ltd.

	Team	BigANN	Deep	MSSpaceV	MSTuring	Text2Image	SSNPP
1	Intel	\$15K	\$16K	\$16K	\$16K	\$104K	-
-	Nvidia	\$304K	\$304K	\$153K	\$153K	\$917K	-

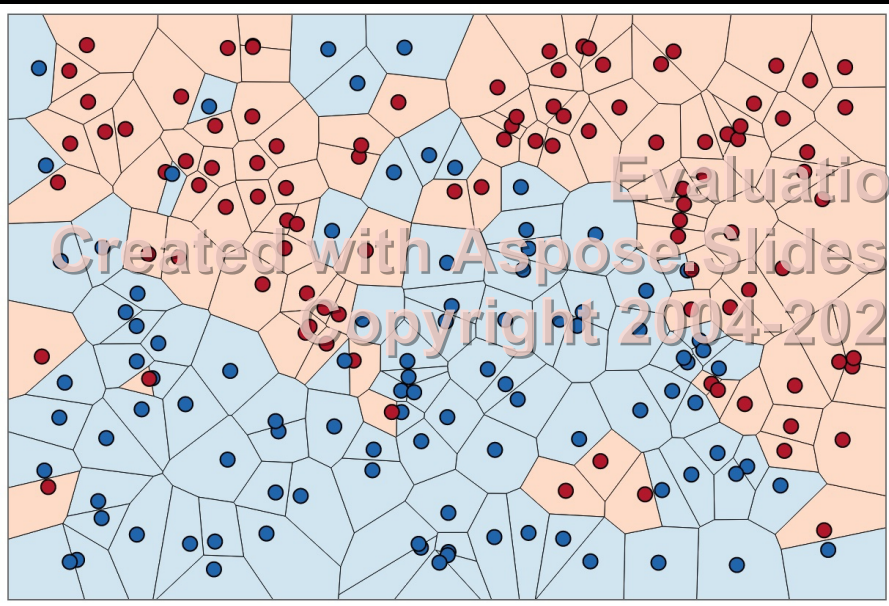


SMILE IDENTITY

Evaluation only.
Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025Aspose Pty Ltd.



SMILE IDENTITY



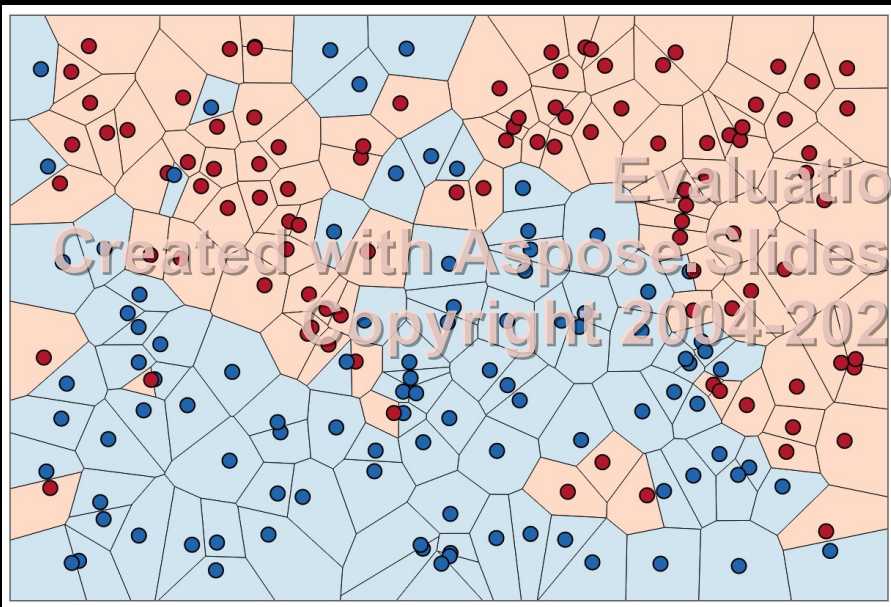
Evaluation only.

Created with Aspose Slides for Python via .NET 25.2.

Copyright 2004-2025 Aspose Pty Ltd.



SMILE IDENTITY



Evaluation only.

Created with Aspose Slides for Python via .NET 25.2.

Copyright 2004-2025 Aspose Pty Ltd.



SMILE IDENTITY

Evaluation only.
Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025Aspose Pty Ltd.



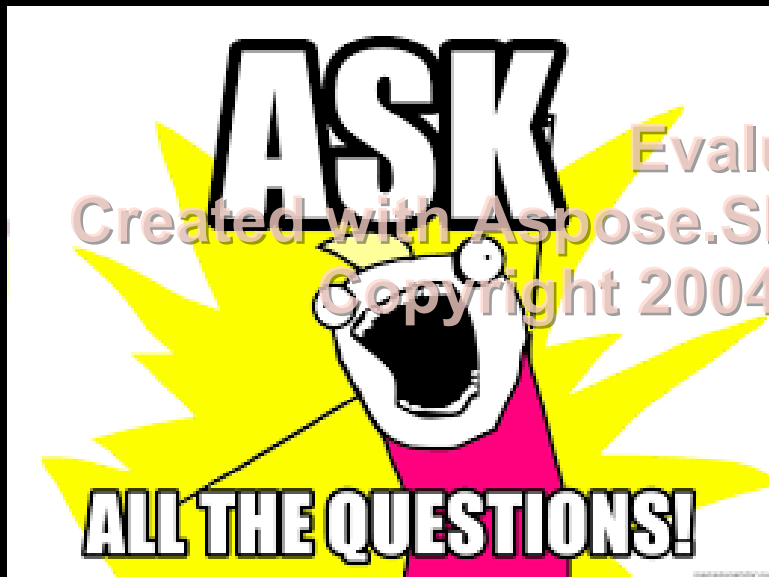
SMILE IDENTITY



Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025 Aspose Pty Ltd.





Evaluation only.

Created with Aspose.Slides for Python via .NET 25.2.

Copyright 2004-2025 Aspose Pty Ltd.

Evaluation only.
Created with Aspose.Slides for Python via .NET 25.2.
Copyright 2004-2025Aspose Pty Ltd.



SMILE IDENTITY