

Performance and Analysis:

Vanilla Column Scan

Dataset Size	Time
1000	0.00106629 s
100,000	0.119536 s
1,000,000	1.13232 s

Dictionary Encoding

Dataset Size	Threads	Time
1000	1	.859583 ms
1000	4	.566125 ms
1000	8	.759333 ms
100,000	1	37.9302 ms
100,000	4	33.6858 ms
100,000	8	33.7831 ms
1,000,000	1	0.322139 s
1,000,000	4	0.311722 s
1,000,000	8	0.322885 s

In general, as the number of threads increases the time needed decreases. It appears that the optimal number of threads is 4.

Single Item Query

Dataset Size	SIMD	Time
1000	No	0 μ s
100,000	No	0 μ s
1,000,000	No	0 μ s
1000	Yes	2 μ s
100,000	Yes	302 μ s
1,000,000	Yes	2969 μ s

Unsure of why not using SIMD caused a return time of 0 seconds but as dataset size increases it takes more time to get through all the data

Prefix Scan

Dataset Size	SIMD	Time
1000	No	0 μ s
100,000	No	2 μ s
1,000,000	No	26 μ s
1000	Yes	8 μ s
100,000	Yes	881 μ s
1,000,000	Yes	12721 μ s

As dataset size increases it takes more time to get through all data

Data sorted by dataset size:

1000 elements:

Method used	Thread #	SIMD	Time
Vanilla Scan	N/A	N/A	0.00106629 s
Dictionary Enc	1	N/A	.859583 ms
Dictionary Enc	4	N/A	.566125 ms
Dictionary Enc	8	N/A	.759333 ms
Single Query	N/A	Yes	0 μ s
Single Query	N/A	No	2 μ s
Prefix Scan	N/A	Yes	0 μ s
Prefix Scan	N/A	No	8 μ s

Single query and prefix scans without SIMD appear much more efficient than the vanilla scans and dictionary encoding.