FMQ: Fast Matrix Sub-Pattern Querying

Matrix sub-pattern search for querying of single-cell data

Alperen Keles, Ataberk Donmez — CMSC858D Term Project Presentation

Problem

Problem

	G0	G1	G2	G3	G4
1	0	0	1	0	0
2	0	0	0	0	0
3	1	0	1	0	1
4	0	0	0	1	0
5	0	0	0	0	1
6	0	0	0	0	1
7	1	0	0	0	0
8	1	0	0	0	0
9	0	1	0	1	0
10	0	1	0	0	0
11	0	0	1	0	1
12	0	0	0	0	0
13	0	0	1	0	0
14	1	1	0	0	0
15	1	1	1	0	0
16	0	1	0	0	0
17	0	0	1	0	1

Problem

	G0	G1	G2	G3	G4
1	0	0	1	0	0
2	0	0	0	0	0
3	1	0	1	0	1
4	0	0	0	1	0
5	0	0	0	0	1
6	0	0	0	0	1
7	1	0	0	0	0
8	1	0	0	0	0
9	0	1	0	1	0
10	0	1	0	0	0
11	0	0	1	0	1
12	0	0	0	0	0
13	0	0	1	0	0
14	1	1	0	0	0
15	1	1	1	0	0
16	0	1	0	0	0
17	0	0	1	0	1

$$G0 \wedge G1 \wedge (\neg G2 \vee G4)$$

Problem

	G0	G1	G2	G3	G4
1	0	0	1	0	0
2	0	0	0	0	0
3	1	0	1	0	1
4	0	0	0	1	0
5	0	0	0	0	1
6	0	0	0	0	1
7	1	0	0	0	0
8	1	0	0	0	0
9	0	1	0	1	0
10	0	1	0	0	0
11	0	0	1	0	1
12	0	0	0	0	0
13	0	0	1	0	0
14	1	1	0	0	0
15	1	1	1	0	0
16	0	1	0	0	0
17	0	0	1	0	1

$$G0 \wedge G1 \wedge (\neg G2 \vee G4)$$

{3, 7, 8, 14, 15}

Problem

	G0	G1	G2	G3	G4
1	0	0	1	0	0
2	0	0	0	0	0
3	1	0	1	0	1
4	0	0	0	1	0
5	0	0	0	0	1
6	0	0	0	0	1
7	1	0	0	0	0
8	1	0	0	0	0
9	0	1	0	1	0
10	0	1	0	0	0
11	0	0	1	0	1
12	0	0	0	0	0
13	0	0	1	0	0
14	1	1	0	0	0
15	1	1	1	0	0
16	0	1	0	0	0
17	0	0	1	0	1

$$G0 \wedge G1 \wedge (\neg G2 \vee G4)$$

{~~3, 7, 8,~~ 14, 15}

Problem

	G0	G1	G2	G3	G4
1	0	0	1	0	0
2	0	0	0	0	0
3	1	0	1	0	1
4	0	0	0	1	0
5	0	0	0	0	1
6	0	0	0	0	1
7	1	0	0	0	0
8	1	0	0	0	0
9	0	1	0	1	0
10	0	1	0	0	0
11	0	0	1	0	1
12	0	0	0	0	0
13	0	0	1	0	0
14	1	1	0	0	0
15	1	1	1	0	0
16	0	1	0	0	0
17	0	0	1	0	1

$$G0 \wedge G1 \wedge (\neg G2 \vee G4)$$

~~{3, 7, 8,~~ 14, ~~15}~~

Problem

	G0	G1	G2	G3	G4
1	0	0	1	0	0
2	0	0	0	0	0
3	1	0	1	0	1
4	0	0	0	1	0
5	0	0	0	0	1
6	0	0	0	0	1
7	1	0	0	0	0
8	1	0	0	0	0
9	0	1	0	1	0
10	0	1	0	0	0
11	0	0	1	0	1
12	0	0	0	0	0
13	0	0	1	0	0
14	1	1	0	0	0
15	1	1	1	0	0
16	0	1	0	0	0
17	0	0	1	0	1

$G0 \wedge G1 \wedge (\neg G2 \vee G4)$

Efficiency

~~{3, 7, 8,~~ 14, ~~15}~~

Solution — Pattern Table

	G0	G1	G2	G3	G4
1	0	0	1	0	0
2	0	0	0	0	0
3	1	0	1	0	1
4	0	0	0	1	0
5	0	0	0	0	1
6	0	0	0	0	1
7	1	0	0	0	0
8	1	0	0	0	0
9	0	1	0	1	0
10	0	1	0	0	0
11	0	0	1	0	1
12	0	0	0	0	0
13	0	0	1	0	0
14	1	1	0	0	0
15	1	1	1	0	0
16	0	1	0	0	0
17	0	0	1	0	1

	G0	G1	G2	G3	G4
1	0	0	1	0	0
2	0	0	0	0	0
3	1	0	1	0	1
4	0	0	0	1	0
5	0	0	0	0	1
6	0	0	0	0	1
7	1	0	0	0	0
8	1	0	0	0	0
9	0	1	0	1	0
10	0	1	0	0	0
11	0	0	1	0	1
12	0	0	0	0	0
13	0	0	1	0	0
14	1	1	0	0	0
15	1	1	1	0	0
16	0	1	0	0	0
17	0	0	1	0	1

Solution — Pattern Table

	G0	G1	G2	G3	G4
1	0	0	1	0	0
2	0	0	0	0	0
3	1	0	1	0	1
4	0	0	0	1	0
5	0	0	0	0	1
6	0	0	0	0	1
7	1	0	0	0	0
8	1	0	0	0	0
9	0	1	0	1	0
10	0	1	0	0	0
11	0	0	1	0	1
12	0	0	0	0	0
13	0	0	1	0	0
14	1	1	0	0	0
15	1	1	1	0	0
16	0	1	0	0	0
17	0	0	1	0	1

	G0	G1	G2	G3	G4
1	0	0	1	0	0
2	0	0	0	0	0
3	1	0	1	0	1
4	0	0	0	1	0
5	0	0	0	0	1
6	0	0	0	0	1
7	1	0	0	0	0
8	1	0	0	0	0
9	0	1	0	1	0
10	0	1	0	0	0
11	0	0	1	0	1
12	0	0	0	0	0
13	0	0	1	0	0
14	1	1	0	0	0
15	1	1	1	0	0
16	0	1	0	0	0
17	0	0	1	0	1

Pattern	Row List
00100	1
00000	2
10101	3
00010	4
00001	5, 6
10000	7, 8
01010	9
01000	10
00101	11
00000	12
00100	13
11000	14
11100	15
01000	16

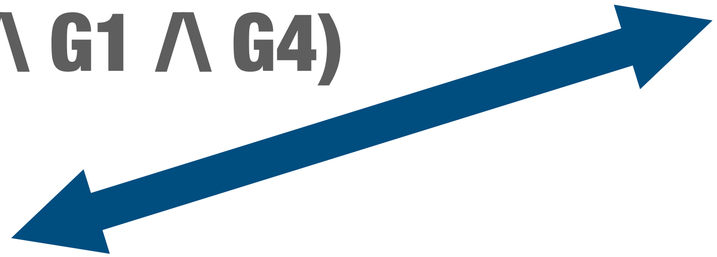
Solution — Query Expansion

$G0 \wedge G1 \wedge (\neg G2 \vee G4)$

G0	G1	G2	G3	G4
1	1	?	X	?

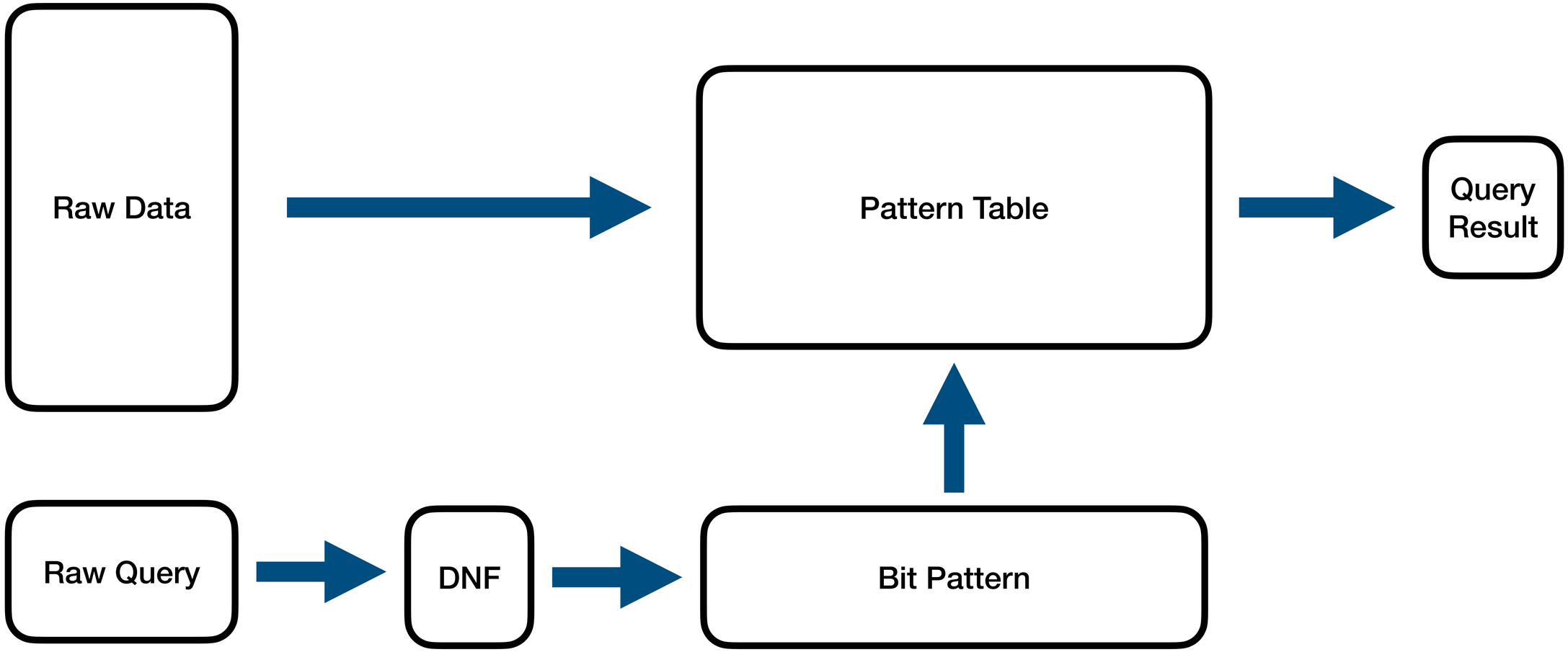
$(G0 \wedge G1 \wedge \neg G2) \vee (G0 \wedge G1 \wedge G4)$

G0	G1	G2	G3	G4
1	1	0	X	X
1	1	X	X	1

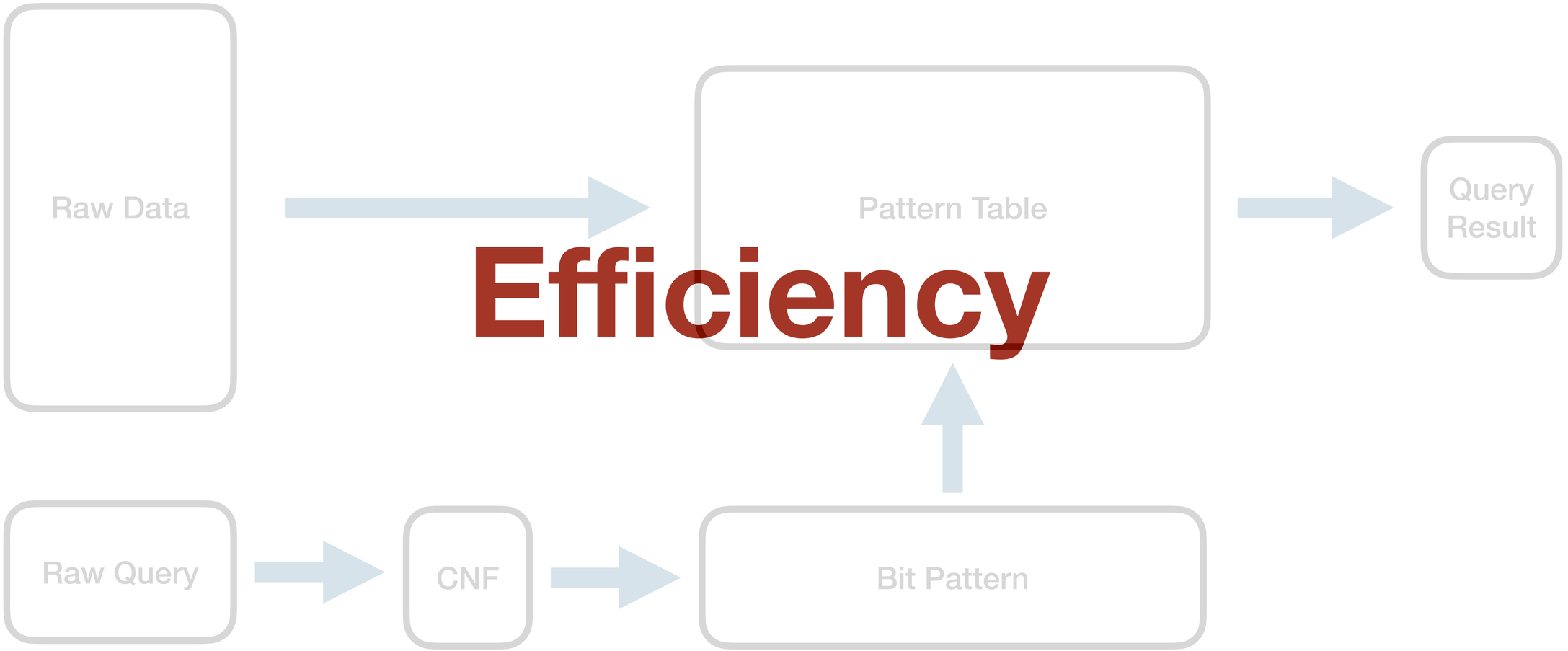


G0	G1	G2	G3	G4
1	1	0	0	0
1	1	0	0	1
1	1	0	1	0
1	1	0	1	1
1	1	1	0	0
1	1	1	0	1
1	1	1	1	0
1	1	1	1	1

Solution


Raw Data
Pattern Table
Query Result
Raw Query
DNF
Bit Pattern

Solution


Raw Data
Pattern Table
Query Result
Efficiency
Raw Query
CNF
Bit Pattern

Pros

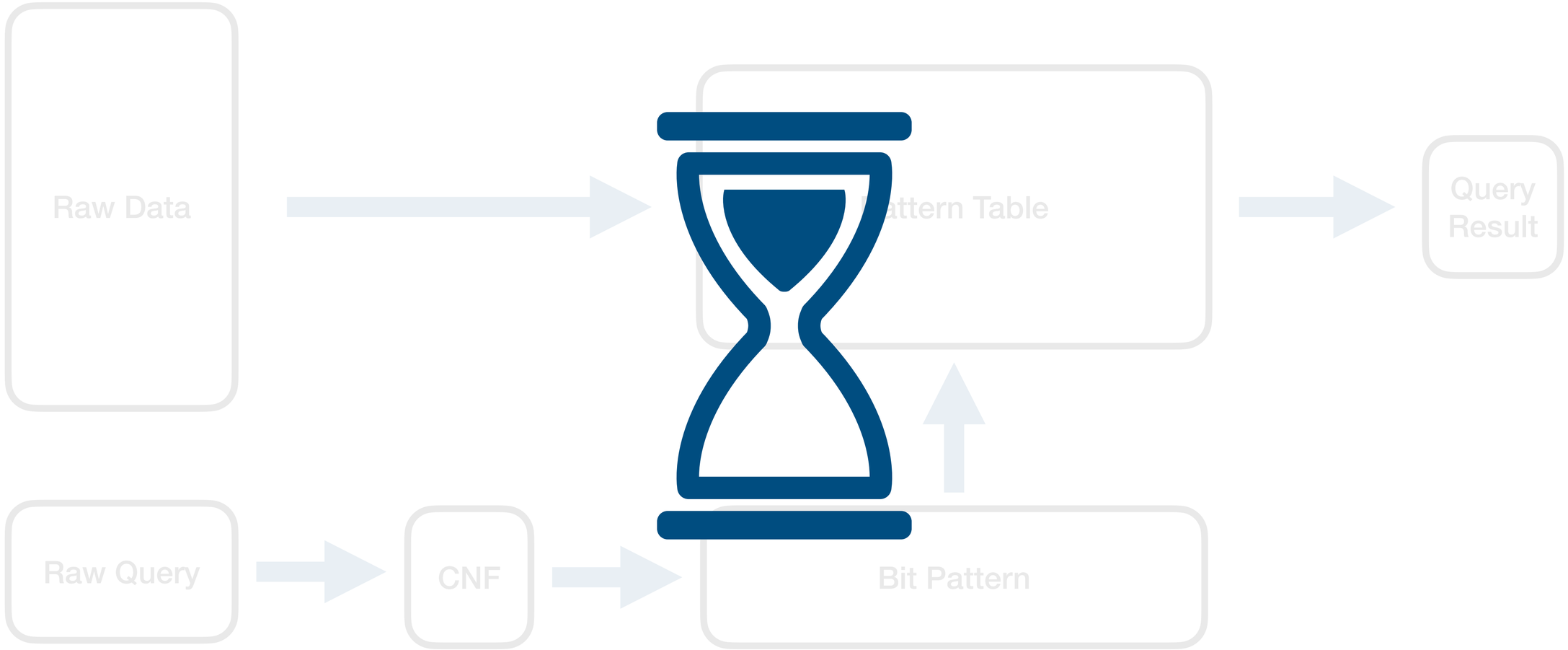
- Dense Queries
- Biased Distributed Rows
- Scaling

Cons

- Sparse Queries
- Uniform Distributed Rows
- Redundant Expansions

Future Possibilities

- Higher Level Indexes
- Efficient Expansion
- Massively Parallel Queries


Raw Data
Pattern Table
Query Result
Raw Query
CNF
Bit Pattern