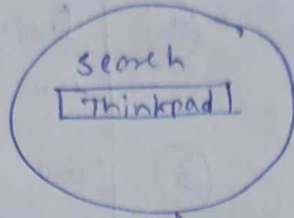
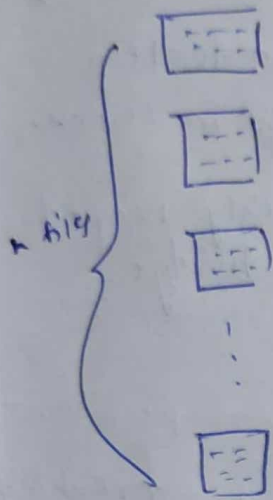


Google Search Engine

- ① Elastic Search } are the tools to make search engine.
② Apache solr }



Iterate in n times and search for that keyword 'Thinkpad'

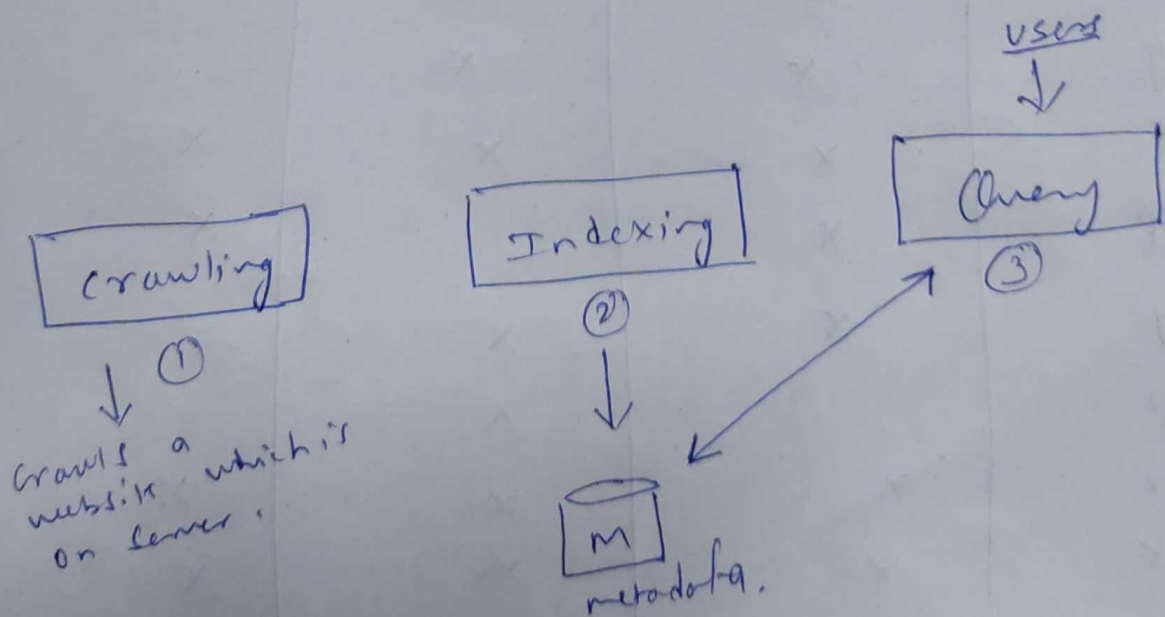
Search fast
low Latency $\downarrow\downarrow$
Throughput $\uparrow\uparrow$
Answer all the query.

Indexing

Xing helps to identify the resource based on some identity.

example $\therefore X = \{1, 2, 3, 4, 5\}$ ($O(n)$ operations)

B-Tree, etc



Inverted Index

Inverted Index

→ stopwards

I, me, my, must,
you etc.

→ noise removal

tag, url

→ ~~stemming~~ stemming : trouble, troubling, troubled, ~~trouble~~ troubl

→ Lemmatization : trouble

Term	Doc1	Doc2	Doc3
Quick		X	X
Brown	X	X	
dog	X	X	
fox	X	X	X
Jump	X		X
lazy	X	X	
leap		X	
over	X	X	X
quick	X	X	X
summer		X	
brige			X

→ It will take much time and space to store the data so we can use B-Tree or Binary digits to represent table.

Limitations:-

- ① consumes more space.
- ② No word position information is available.

Term	freq	Occurrences Inverted Index
Quick	3	[1, (2)] [2, (1)] [3, (3)]
Brown	2	[1, (3)] [2, (2)]
Bridge	1	[3, (7)]

↑
This is the process for one word retrieval of data.

Note:- It stores the cache of memory

Query/search quick brown fox.

quick \rightarrow [1, [2]]

[2, [1]]

[3, [3]]

brown \rightarrow [1, [3]]

[2, [2]]

fox \rightarrow [1, [4]]

[2, [3]]

[3, [2]]

① conjunctive AND

② disjunctive OR. \rightarrow
Compare all the documents.

Intersect \rightarrow result.

If we need to find out the sentence from the DB then we need to perform conjunctive (AND) operation \rightarrow Intersect the indexes and search for data.