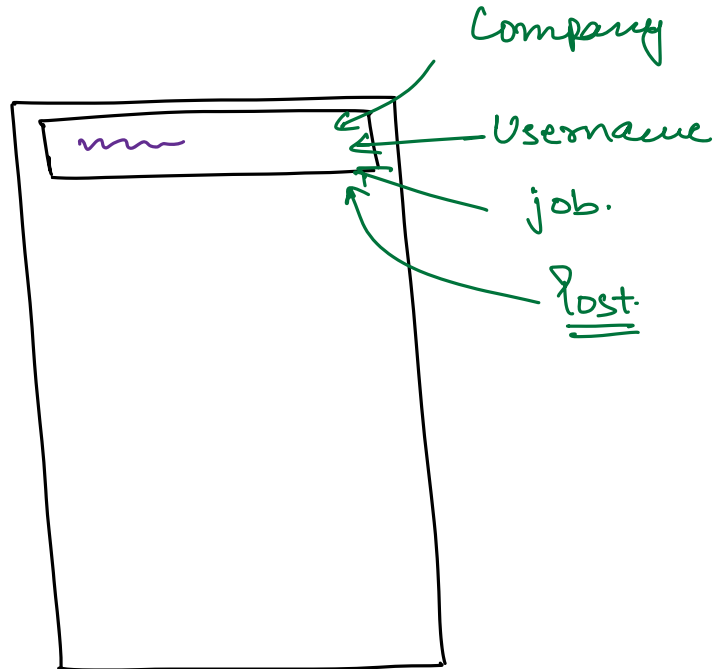


## Agenda.

→ ElasticSearch.

↳ Full Text Search

⇒ LinkedIn.



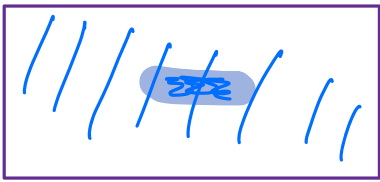
I) Attribute based Search.

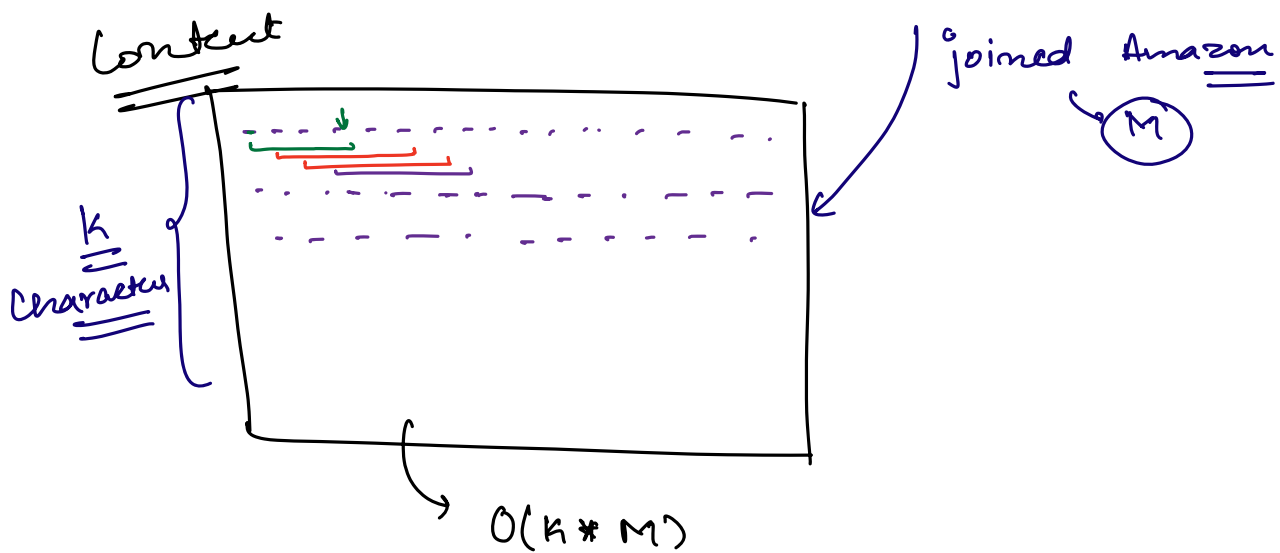
⇒ Select \* from users  
where username LIKE "%shaked%".

II) Full Text Search.

Select \* from posts  
where content like "%joined Amazon%".

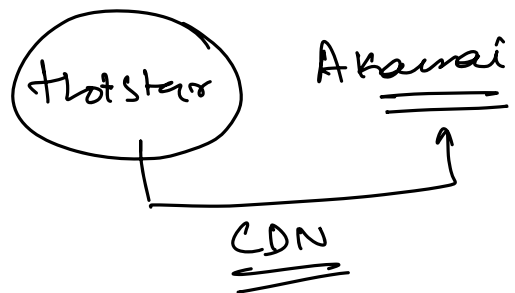
posts.

id	author-id	time	Content	- - -
<u>z</u>	<u>z</u>	<u>z</u>		

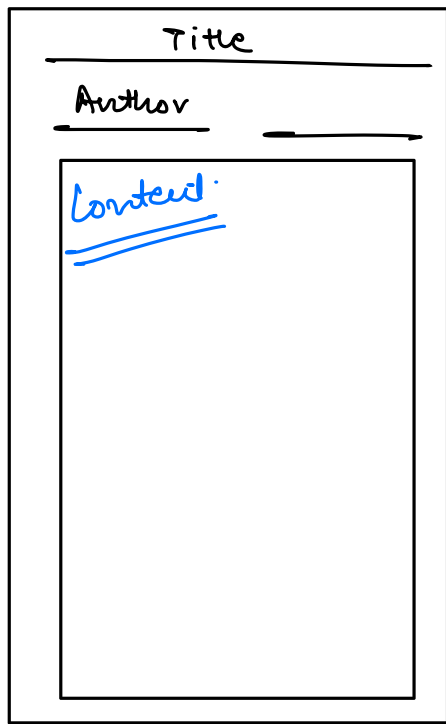


TC:  $O(N * k * M)$

⇒ Google Search



"% hotstar %" "% Akamai %"

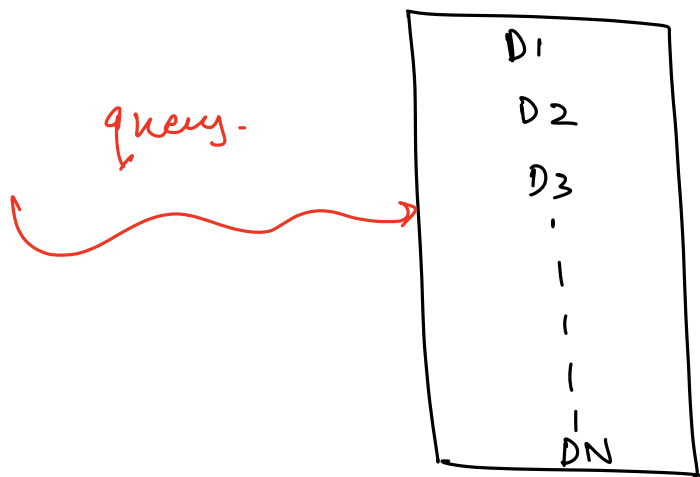


"operates latest version Algorithm"

SQL X

NOSQL

⇒ Document Based DB.



Brute force of Full Text Searching : Iterate all the documents a check if the keyword is present or not.

⇒ Millions of Documents

## # ElasticSearch.

- Document Based DB.
- \* Inverted Index.
- full Text Search Query.

#

- D1: India won T20 world cup after 17 Yrs.
- D2: General elections in India happened in 2024 and Modi came to power again.
- D3: India got smashed badly by Australia Test cricket in 2024.
- D4: Rohit Sharma and Virat Kohli planning to retire after a very successful career.
- D5: Indian cricket team to get a new Captain soon.

⇒

<u>India</u>	0.89 D <sub>1</sub>	0.75 D <sub>2</sub>	0.71 D <sub>3</sub>	0.65 D <sub>5</sub>
Modi		D <sub>2</sub>		
Cricket		D <sub>3</sub>	D <sub>5</sub>	

Inverted  
Index

Indian Cricket  
India Cricket  
Cricket India  
Cricket  
≡



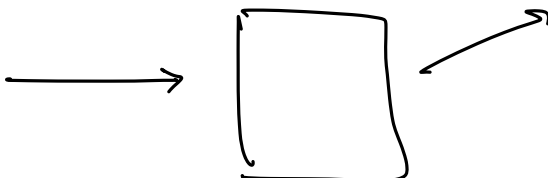
India : Top 20 docs.

Cricket: Top 20 docs.



Intersection.

OTC

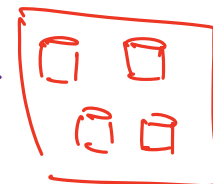


SI



DB

(MongoDB / SQL)



Elastic  
Search

⇒ Elasticsearch is used as a secondary DB to make our searches faster using Inverted Indexing.

⇒ Availability >> Consistency.

⇒ We need to pre process all the documents to convert into Inverted Index, to make our searches optimized.

# Indian Cricket team world cup.

Unigram

↓  
[Indian] [Cricket] [team] [World] [Cup]

Bigram

[Indian Cricket] [Cricket team]

[World Cup] - - - -

Trigram

[Indian Cricket team] - - -

↓  
India

Inverted Index.

India	D <sub>5</sub> D <sub>7</sub> D <sub>10</sub>
Cricnet	D _____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____

Ind (vs) Pak.

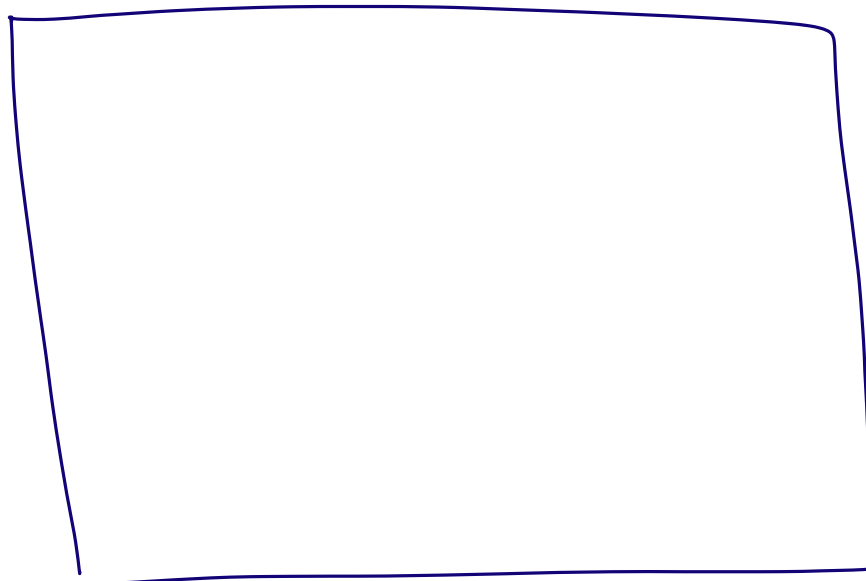
⇒ English language.



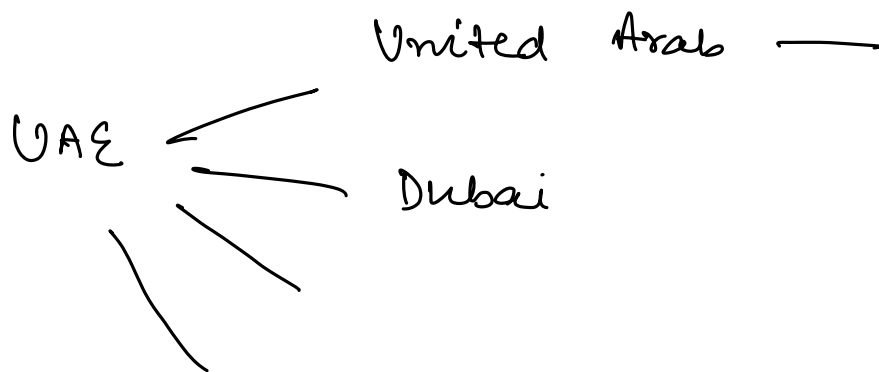
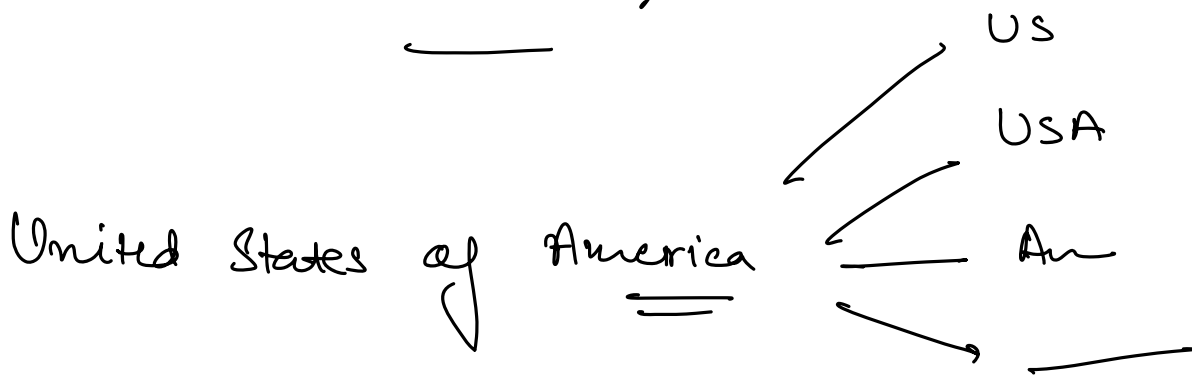
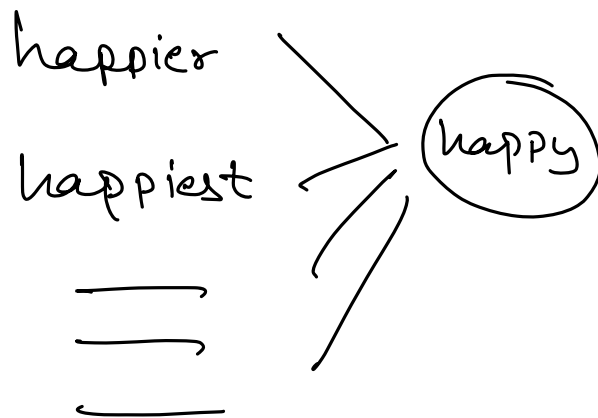
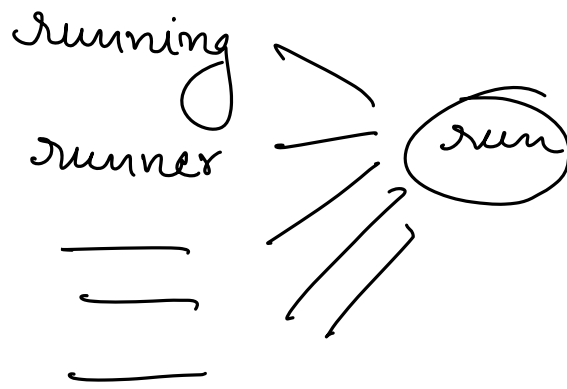
Word Corpus.

→ SL Unique Words.

(LM)

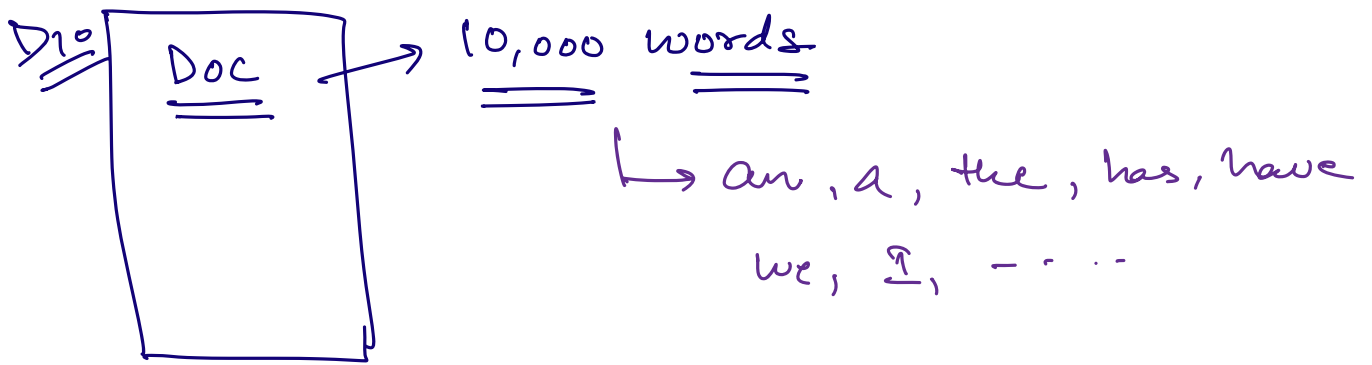


← (LM)



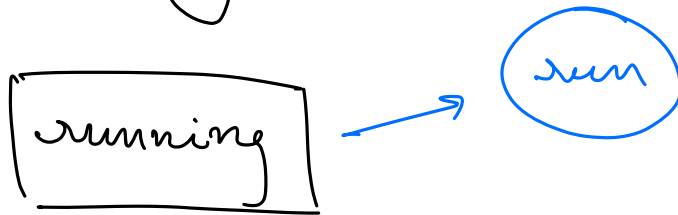


⇒

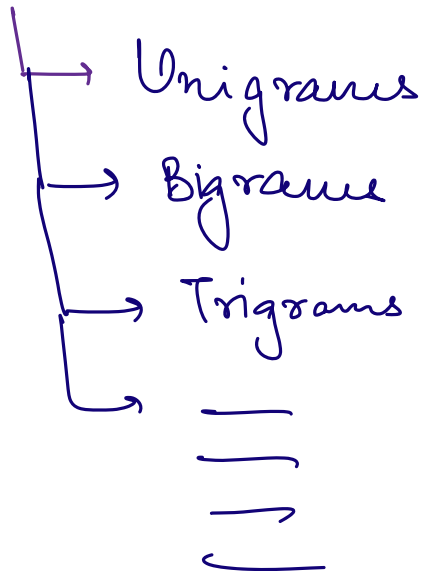


I) Removing the stop words.

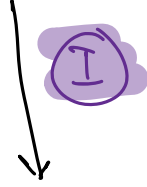
II) Stemming | Root Word | -



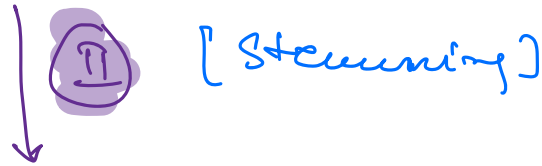
III) Generate Tokens



D: Indian cricket team is one of the  
greatest teams in world sports.



D: Indian cricket team  
world sports.



India cricket team world  
sports.



India  
Cricket  
Team

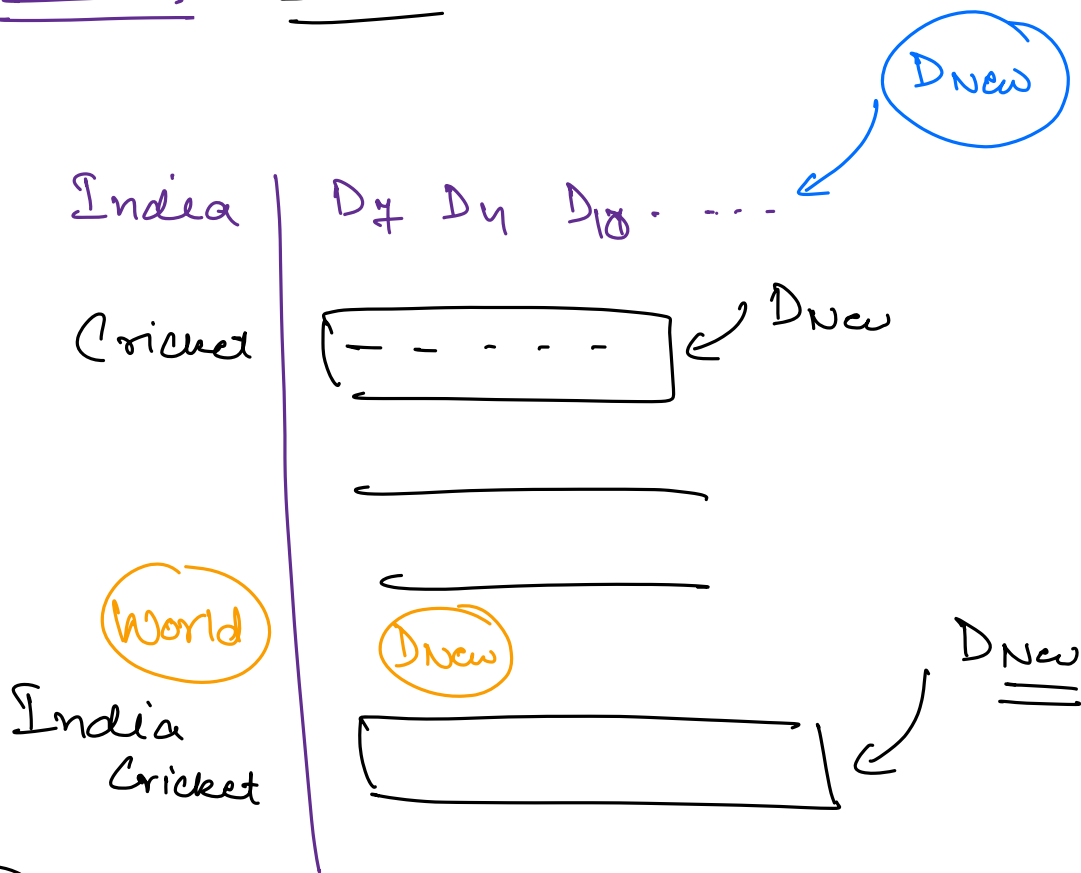
World  
Sports.

[India Cricket]

[Cricket Team]

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

# IV) Inverted Index.



⇒ Tf Idf

frequency / Tf - Idf / . . . / . . .

India | D<sub>7</sub> : 0.87 | D<sub>10</sub> . . . . .

✓

---

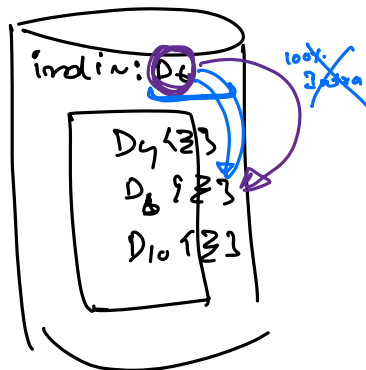
---

## Documents.

↓

Doc id  
(SHARDING KEY)

~~Docid~~



~~100% India~~

↓  
Cache for Document  
with high frequencies.

⇒ ES