

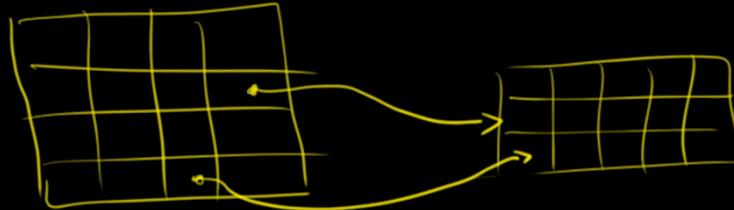
Page 1

Different Databases:

SQL db ↗ MySQL
↗ Oracle
↗ Postgres

- data representation
- data traversal
- data storage way
- distributed things

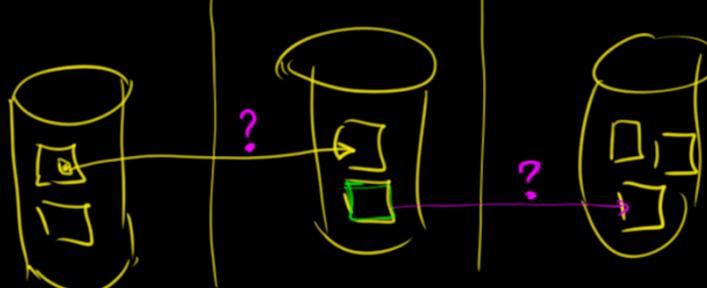
Relational Database (SQL)



computer#1

#2

#3

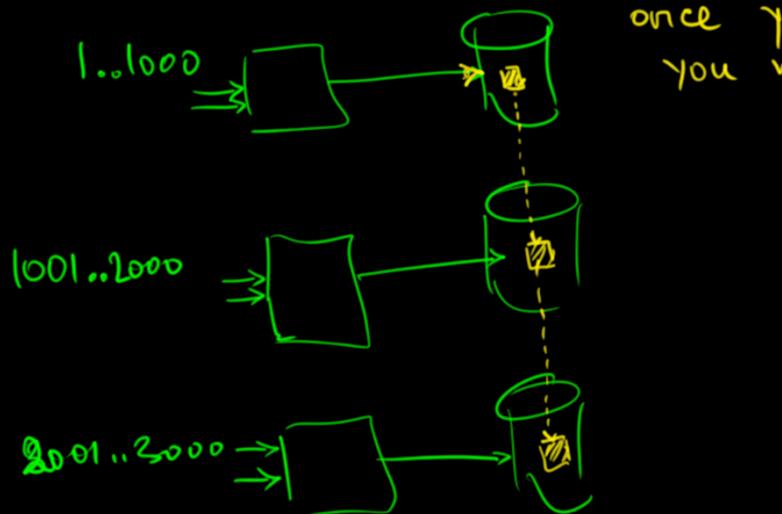
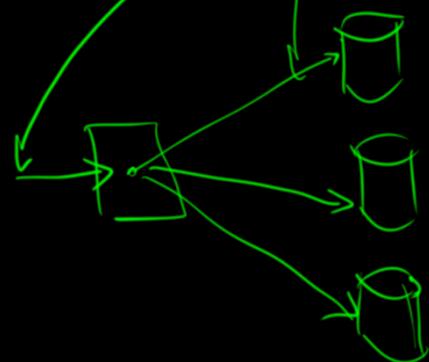


many computers
data doesn't fit into one

Δt for query 1 → 2 → 3

Page 2

Huge amount of data doesn't fit into 1 computer
Huge amount of requests can't be handled by 1 computer



once you write
you need to sync



Distribution.
more than
one
instance

CAP theorem



availability
we need to serve
many requests

don't have FK
non-Rel DB
noSQL

Partition Tolerance

at 1ms / 10ms / 100 / 1000 ...
it's never \emptyset
because of connection overhead

Relational DB, SQL

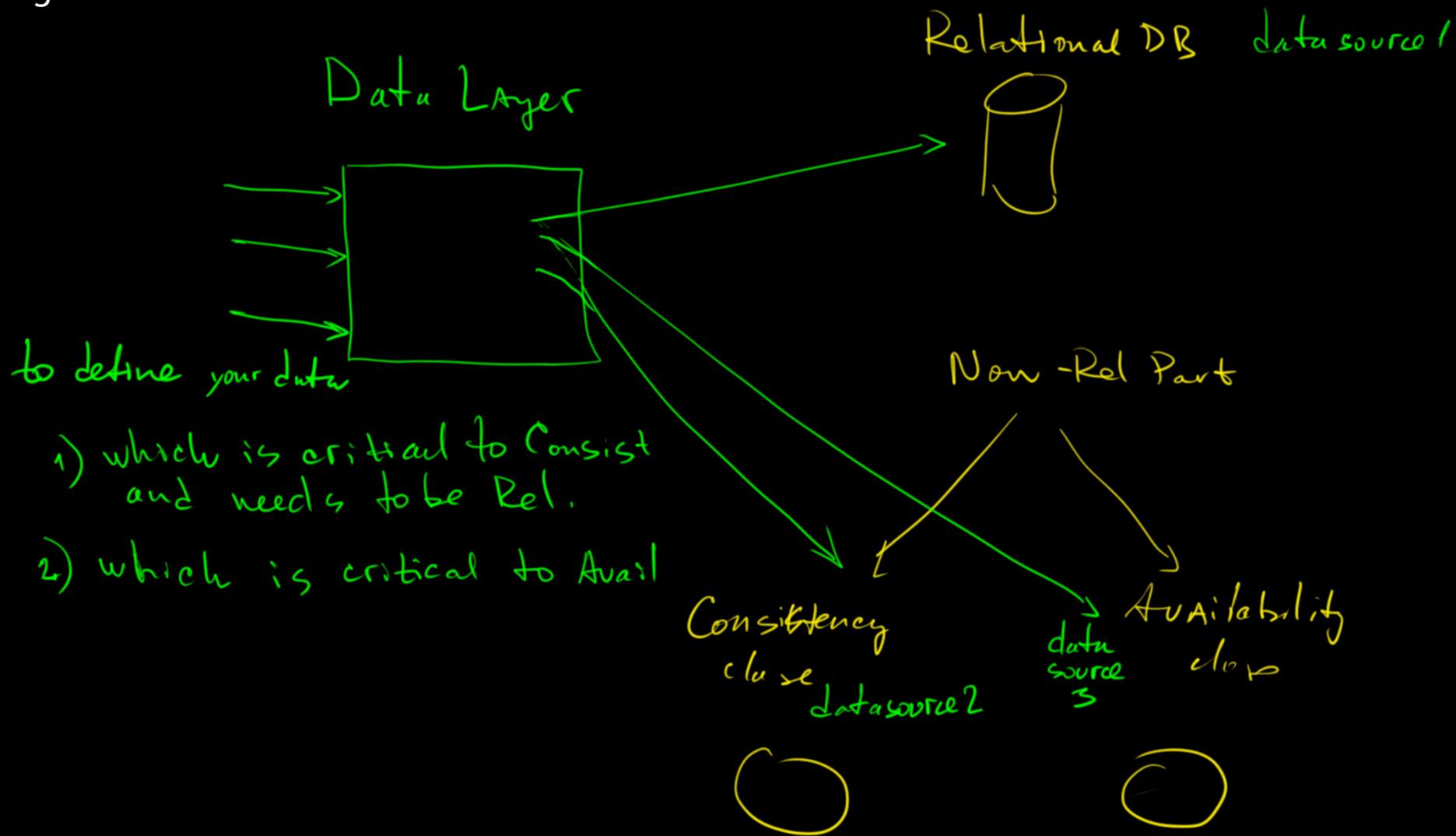
we need to maintain
consistency

1. we need to sacrifice Availability in favor of Consistency

2. we sacrifice Consistency in favor of Availability

Social Net
Video
Audio Stream.

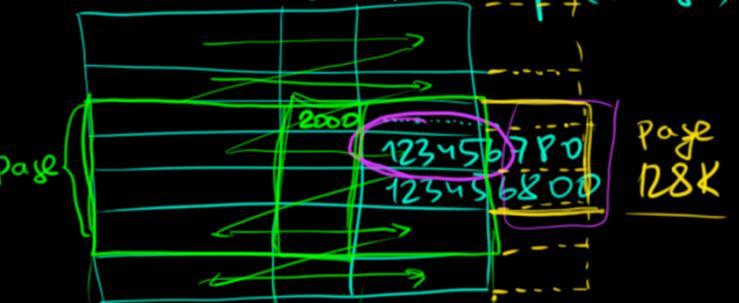




Page 5

Representation

Year Timestamp (long) + 1 at 0.001s



Timestamp

123456 789

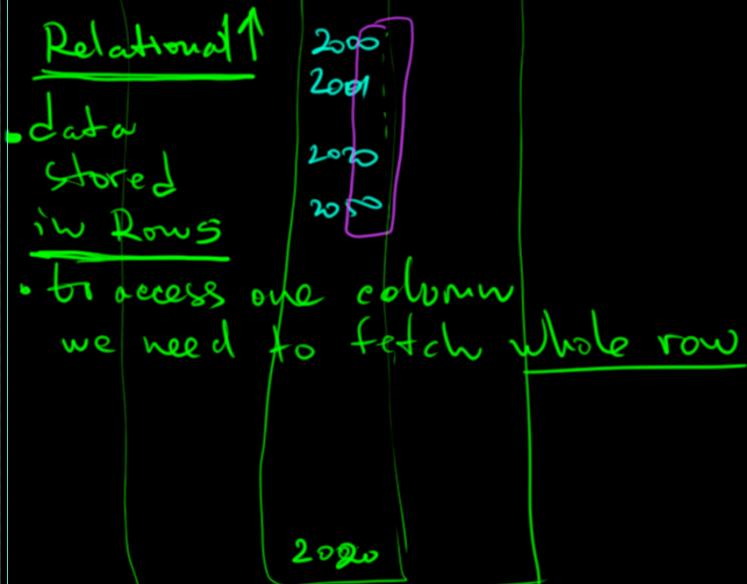
790

800

805

Row 16kb

1page → 8rows



⇒ reduce the size

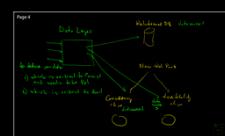
⇒ reduce the response time

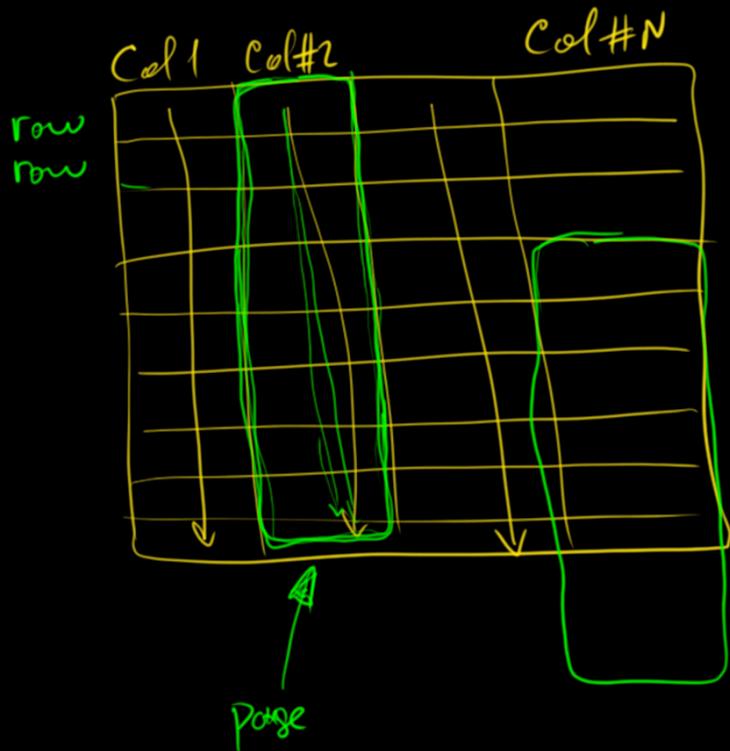
Col → Row → Page

• adding a column can take A LOT of time
because we need to rewrite EVERYTHING

to read Year only we need to
read all the data (Column)

real data which can be
combined
with prefix





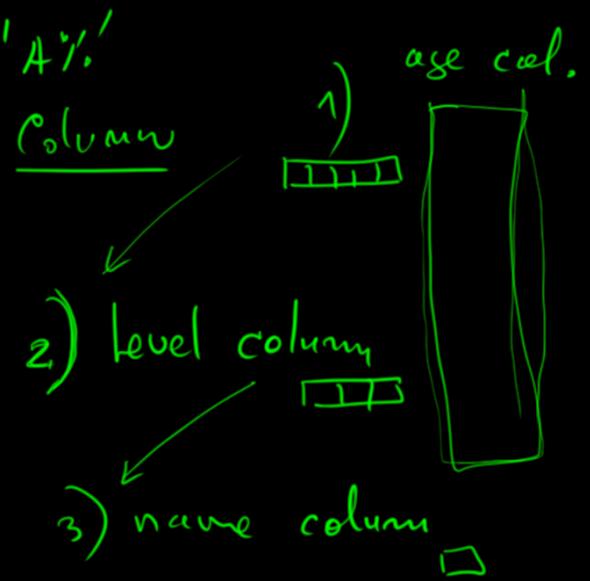
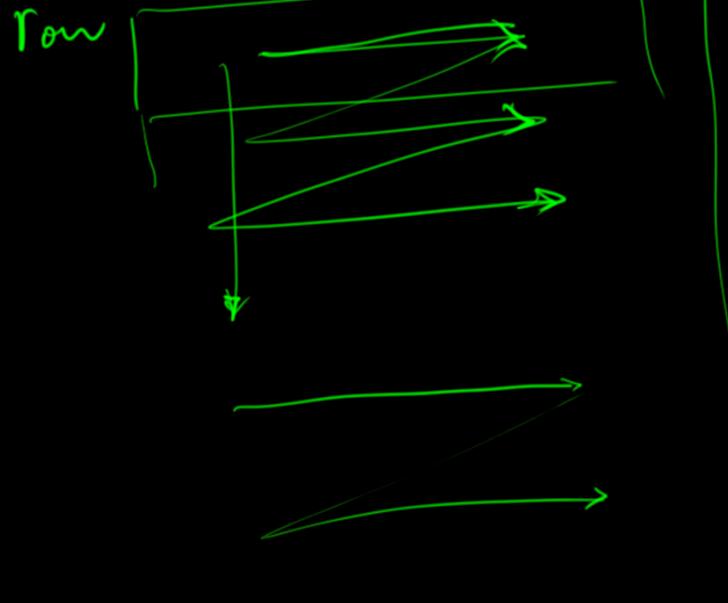
Columnar Databases

`select * from users`

`where age > 18`

`and level = 'admin'`

`and name like 'A%'`



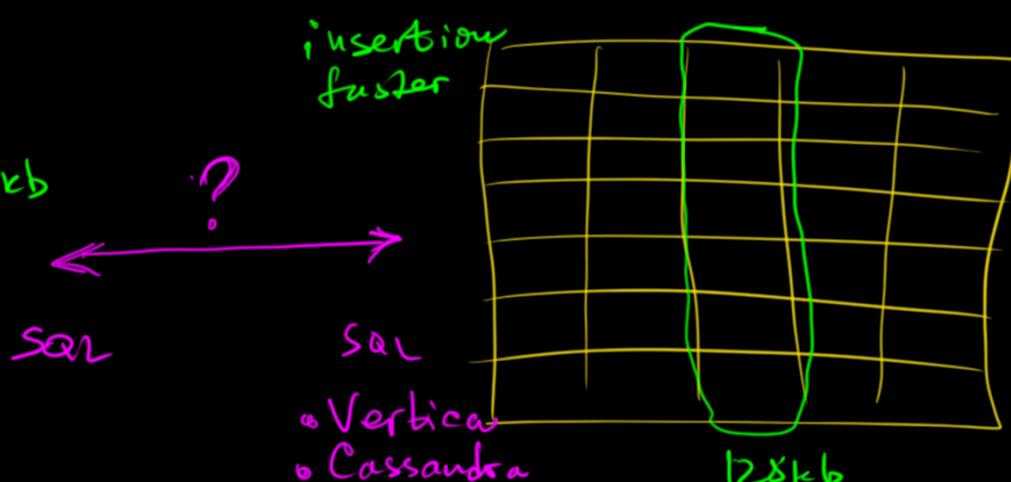
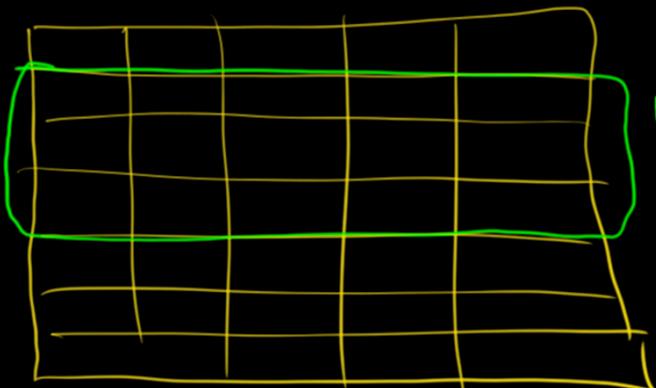
Is Columnar db better than std row db?

→ it depends

```
select * from users
where age > 18
and level = 'admin'
and name like 'A%'
```

anyway we need to read all

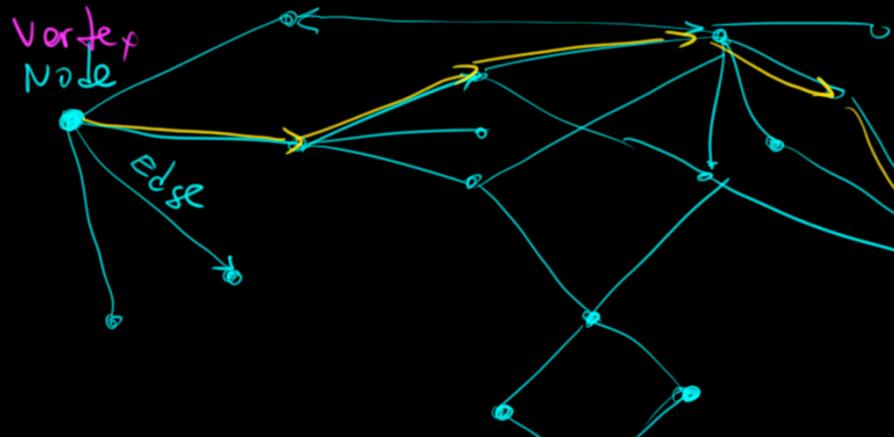
It doesn't matter



- easy to scale
- each col can be on different inst. in query in parallel => faster.

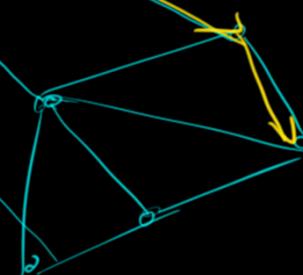
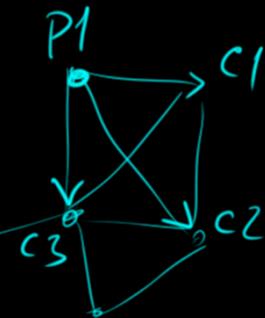
Graph Databases

Vertex
Node

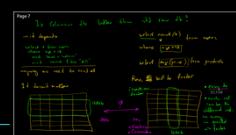


Parent Id | Children Ids[]

Parent Id	Children Ids[]
P1	c1, c2, c3



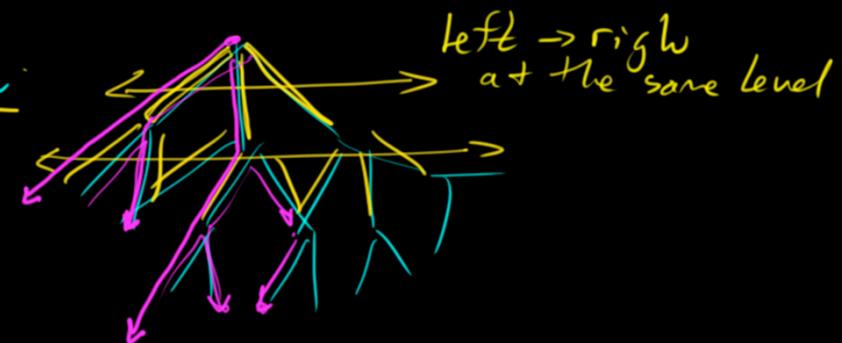
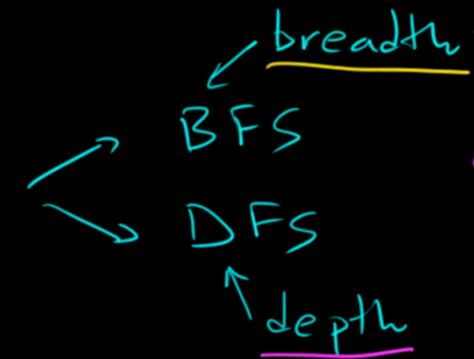
- People Relations
 - Road Tasks (Maps)
 - Science
- disease spreading
- electricity through grid
- medicine



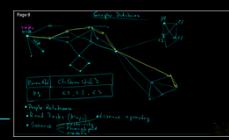
Tasks:

1. is there a relation $A \rightarrow B$
2. all the children w/ subchildren (whole subgraph)
3. all possible ways $A \rightarrow B$
 - shortest path $A \rightarrow B$
 - longest path $A \rightarrow B$
 - number of ways $A \rightarrow B$

graph traversal ways



we always reach end of the graph
or already visited vertex



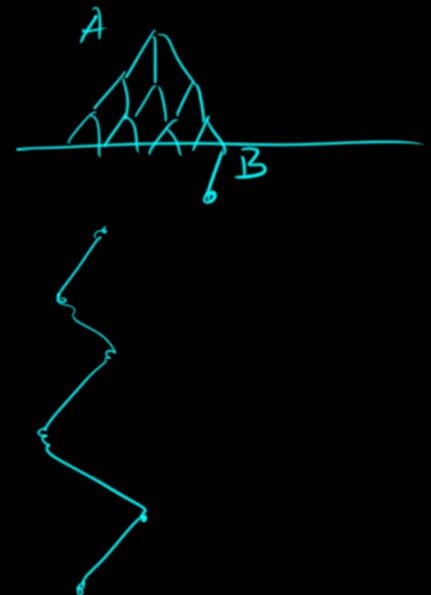
Page 19

is there path $A \rightarrow B$ shorter than N BFS

is there path $A \rightarrow B$ DFS

BFS memory $\max(\text{width})$

DFS memory $\max(\text{depth})$.



Database Migrations

- we focus on the DB structure

tableA	
c1	c2

→ information about the structure

- database metadata as a code

V001.sql create table T1 (c1 int, c2 string)

V002.sql alter table T1 add column c3 double

-

-

-