

# Index

---

# Overview over this video

---


This video will discuss the topic of indices

# Reminder: Computing $\sigma_{\text{condition}}(R)$

Basic procedure:

R:

tuple 1
tuple 2
tuple 3
tuple 4
...



```
for each tuple t in R:  
    if t satisfies condition:  
        output t
```

Needs to read the entire relation

Yes, sometimes!

Can this be done faster?

# Example

$\sigma_{\text{programme}='G401'}(\text{Students})$

Students

id	name	programme
...	...	...
1234	Anna	G401
2345	Ben	G701
3456	Chloe	G401
4567	Dave	G401
...	...	...

Selection can be performed faster if we know  
**where to find the rows for 'G401'**

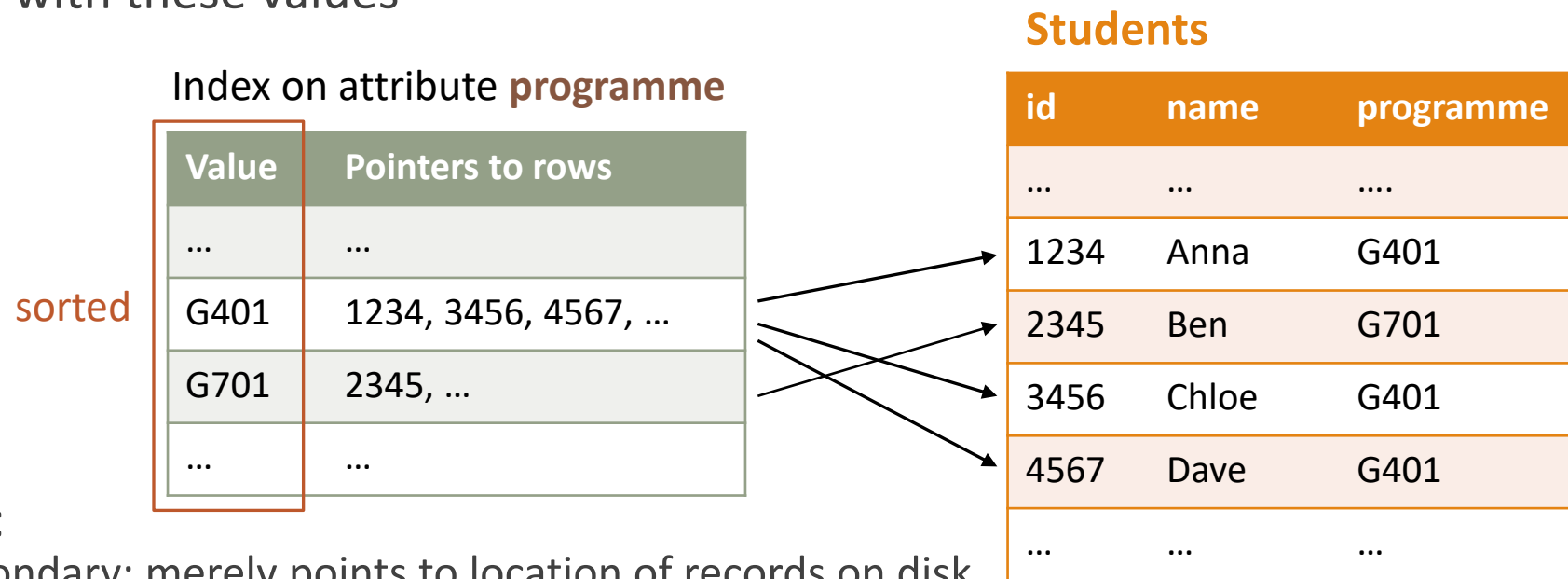
Two solutions: **sorting & index**

See last video

This video

# Index

Given the values for one or more attributes of a relation **R**, provides quick access to the tuples with these values



Types:

- Secondary: merely points to location of records on disk
- Primary: in addition, defines how data is sorted on disk

Always dense

Good when attributes involve primary key

# Example Revisited

$\sigma_{\text{programme}='G401'}(\text{Students})$

Index on attribute **programme**

Value	Pointers to rows
...	...
G401	1234, 3456, 4567, ...
G701	2345
...	...

**Students**

id	name	programme
...	...	...
1234	Anna	G401
2345	Ben	G701
3456	Chloe	G401
4567	Dave	G401
...	...	...

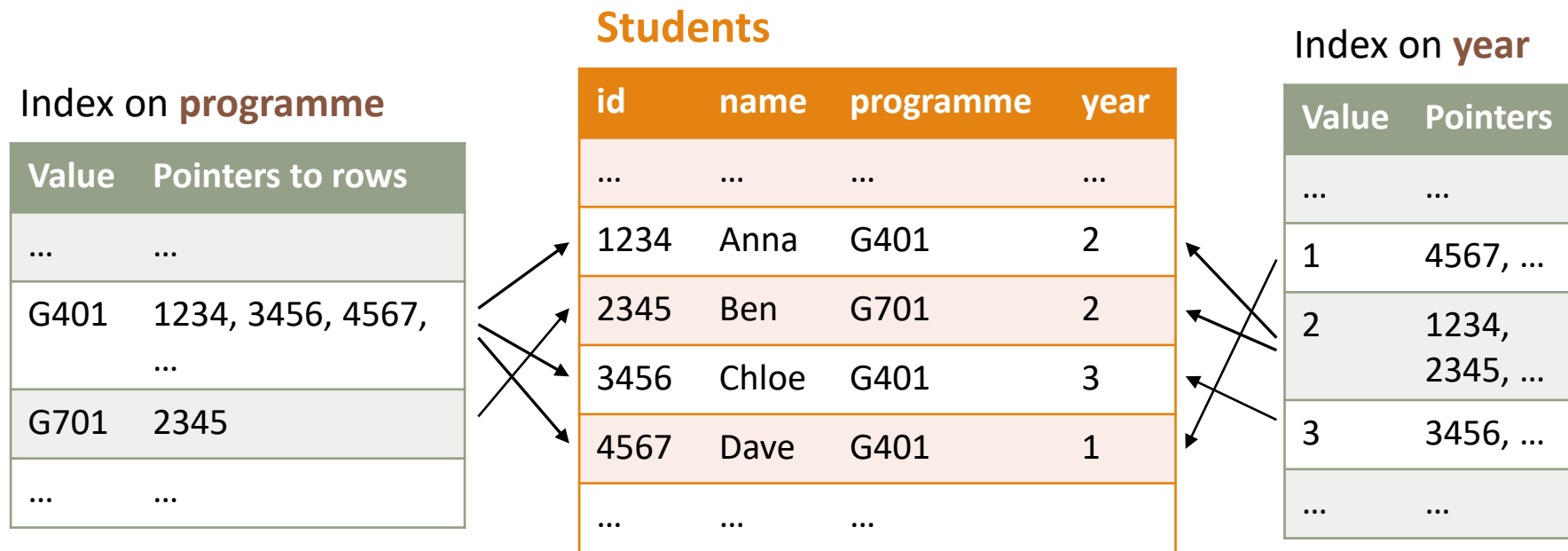
Selection with index:

- Find entry for '**G401**' in index
- Visit all rows in **Students** whose ids occur in the index entry for '**G401**'

Running time?  $O(\log k + \text{size of output})$

# Example 2

$\sigma_{\text{programme}='G401' \text{ AND } \text{year}=2}$  (**Students**)



Selection with two indexes:

- Find entries for '**G401**' & **2** in indexes for **programme** & **year**
- Visit all rows in **Students** whose ids occur in both index entries

# Example 3

$\sigma_{\text{programme}='G401' \text{ AND } \text{year}=2}$  (**Students**)

Index on **programme, year**

Value		Pointers to rows
...	...	...
G401	1	4567
G401	2	1234
G401	3	3456
G701	2	2345
...	...	...

**Students**

id	name	programme	year
...	...	...	...
1234	Anna	G401	2
2345	Ben	G701	2
3456	Chloe	G401	3
4567	Dave	G401	1
...	...	...	...

Selection with two indexes:

- Find entries for '**G401**', **2** in index for **programme, year**
- Visit all rows in **Students** whose ids occur in that index



# Forms of Indexes

## B+ Trees

- Good if selection condition specifies a range
- E.g.,  $\sigma_{\text{programme}='G401' \text{ AND year} > 1}$
- Most widely used

```
CREATE INDEX
  ON Students USING btree
  (programme, year);
```

default

## Hash tables

- Good if selection involves equality only
- E.g.,  $\sigma_{\text{programme}='G401'}$

```
CREATE INDEX
  ON Students USING hash
  (programme);
```

```
CREATE INDEX
  ON Students USING hash
  (lower(name));
```

Many more...

# Summary

---

There are a variety of different indices

In particular:

- Hash-index: good for equality
- B+-Tree: good for ranges