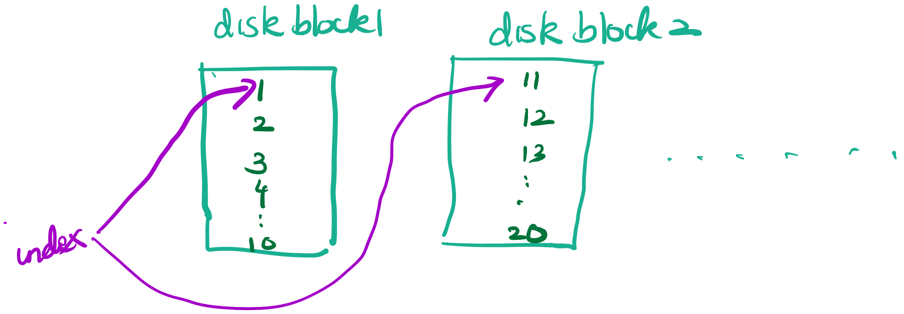
Types of indexes

1) Primary key
2) Clustering index
3) Secondary index
4) multiple attribute index

1) Primary key

- if table is stored in sorted order by primary key, then sparse index. (# of index entries < # of tuples)



- if not stored by primary key, then dense index.

2) Clustering index

- tuples are physically stored sorted on a non-key attribute

  e.g. dno in Employee table

- sparse index

3) Secondary index

- candidate keys 2

- foreign keys } dense index

4) multiple attribute index

Example: Select *
from Emp
where dno = 5 and salary > 50000;

- index on dno + salary

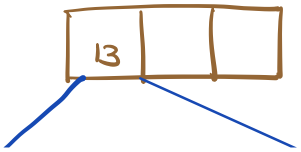
Example: Dept table

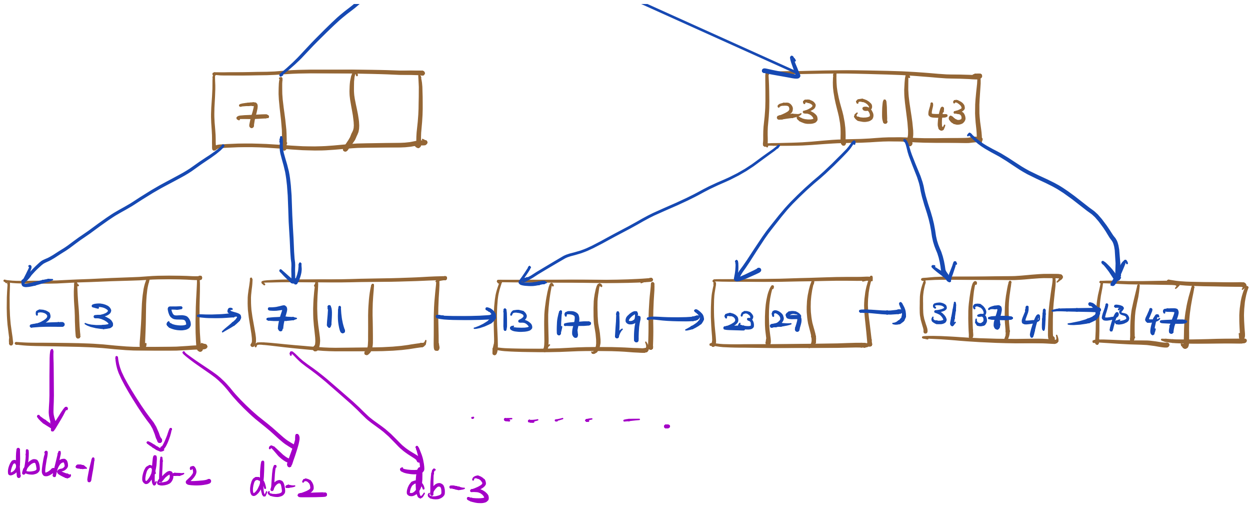
dno: 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47

Dept table on disk

db-1	db-2	db-3	db-4	db-5
2	3	17	47	41
13	11	7	19	37
43	5	21	29	23

B-tree index





```
Select * from Department
where dno ≤ 7;
```

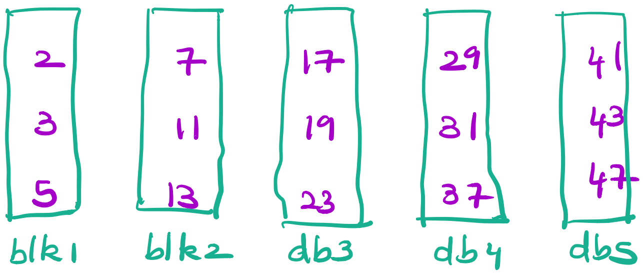
* No indexes : - access all 15 tuples

- access all 5 disk blocks

* index on dno : 2, 3, 5, 7

- access 4 tuples
- 3 disk accesses (blocks)

* Suppose department is stored in sorted order by dno.



- access 4 tuples
- 2 disk blocks accessed

Example:

→ 3000 tuples

Select * from Employee
where dno = 5 and Salary > 50000;

↓ 100 tuples

↓ 300 tuples

10 tuples

How many tuples accessed:

1) no indexes: 3000 tuples

2) index on dno: 100 tuples & then search for sal>50000

3) index on salary: 300 tuples & then search for dno=5

4) multi-attribute index on dno + salary: 10 tuples

5) how many disk blocks accessed: ✗

```
CREATE index <indname> ON TABLE <table> (attributes)
 USING {BTREE | HASH}

   CREATE index idno ON EMPLOYEE (dno);

SHOW INDEX FROM <table>

   SHOW index from EMPLOYEE;
```

* most DBMS build indexes (by default) on primary key.
* on UNIQUE sometimes
* on foreign keys sometimes

Example: JOIN query

```
Select pname, dname
from Department, Project
                                               ⋈
where mgr-SSN = '1234' and dept.dno = project.dno;
                   ↓
                1 tuple
```

- Dept has 1000 tuples
- Project has 5000 tuples
- mgr-SSN = 1234 manages dno = 5
  * 10 project tuples with dno = 5

How many tuples are accessed (in worst case):

1) No indexes 6000

$$\underset{\substack{\downarrow\\ \text{dept}}}{1000} + \underset{\substack{\downarrow\\ \text{project}}}{5000} = 6000$$

2) index on mgr-SSN 5001

1 tuple is Dept with mgr-SSN = 1234

5000 in Project to find dno = 5

3) index on project.dno 1010

1000 in Dept for mgr-SSN = 1234

10 project tuples with dno = 5

4) 2 indexes a) index on mgr-SSN in dept

b) index on project.dno

11

1 from dept for mgr SSN

10 from project for dno = 5

5) disk accesses × insufficient data