Ex 1.a

Patient			
Col Name	Available Size	Assumed Ratio	Actual Size Used
hcnum	12	1	12
Last Name	30	2/3	20
First Name	30	2/3	20
Email	42	2/3	28
phonenum	11	1	11
Bytes per row			91
Number of rows			30,000
Number of bytes			2,730,000
Average Data Page Size Bytes (D)			4,000
Fill Factor (E)			75%
Available Data page size (D * E)			3,000
Number of Data Pages			910
Number of Buffer Pages			100
Total Number of Data Pages			**1,010**

Doctor			
Col Name	Size	Assumed Ratio	Actual Size Used
practiceid	10	1	10
name	48	2/3	32
Startdate	10	1	10
Speciality	33	2/3	22
Bytes per row			74
Number of rows			200
Number of bytes			14,800
Average Data Page Size Bytes (D)			4,000
Fill Factor (E)			75%
Available Data page size (D * E)			3,000
Number of Data Pages			5
Number of Buffer Pages			100
Total Number of Data Pages			**105**

Diagnosis			
Col Name	Size	Assumed Ratio	Actual Size Used
digidid	12	1	8
notes	300	2/3	200
digdate		1	10
followup	30	2/3	1
Followup Date	30	2/3	8
Hcnum	12	2/3	12
practiceid	10	1	10
Bytes per row			249
Number of rows			300,000
Number of bytes			74,700,000
Average Data Page Size Bytes (D)			4,000
Fill Factor (E)			75%
Available Data page size (D * E)			3,000
Number of Data Pages			24,900
Number of Buffer Pages			100
Total Number of Data Pages			**25,000**

Ex 1.b (i)

25,000

Ex 1.b (ii)

Diagnosis date between 1 Jan 2019 to 31 Dec 2020 for practice ID d

If index is non clustered (Type 1) then we will use following formula :

T(R) * 1/V(R,a) where

– **B(R)** = # of pages
– **T(R)** = # of tuples (rows)
– **V(R, a)** = # of distinct values of attribute a
• When **a** is a key, **V(R,a) = T(R)**
• When **a** is not a key, **V(R,a)** can be anything < **T(R)**

1,500 * 1 / 1 = 1,500 I/Os

Ex 1.b (iii)

Diagnosis dates between 1 Jan 2019 to 31 Dec 2020 → 730

Clustured Type 2 on Digidate	
Index Table (Diagnosis)	
Rid	10
Pointer	8
Bytes per row	18
Number of rows	730
Number of bytes	13,140
Average Index Page Size Bytes (D)	4,000
Fill Factor (E)	75%
Available Index page size	3,000
Number of Data Pages	4
Number of Buffer Pages	100
Total Number of Index Pages	**104**

log(104) (base 2) = 7

Number of I/O for clustered type 2 index = 7 + 1 + 1 = 9

1 is for (if single entry may be spread over next page)

Ex 3(a)

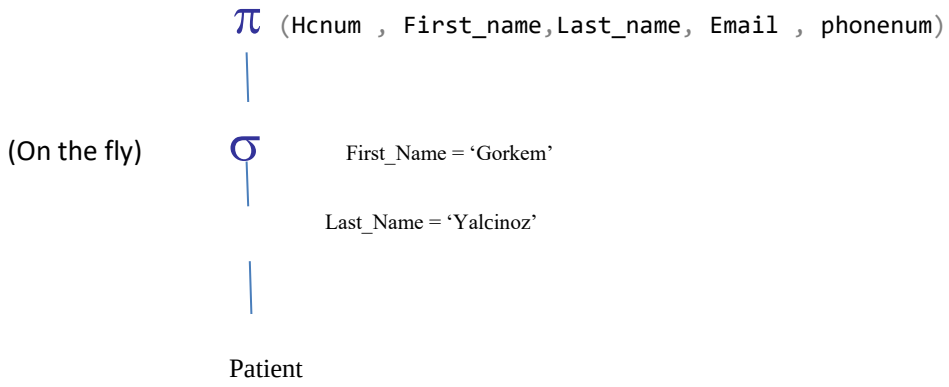
π P.hcnum, P.phonenum, D.practicid, D.diagdate, D.followupdate, Doc.practiceid, Doc.speciality, Doc.name (P ⋈ D ⋈ Doc)

σ P.hcnum = P.hcnum ^ and D.practicid = Doc.practiceid

^ diagdate >= '2020-01-01' and speciality = 'Oncology'

Ex 3 (b)

Rough estimation of I/O cost



Index Table	
Rid	10
Pointer	8
Bytes per row	18
Total Number of rows (Diagnosis)	30,000
Number of Index Bytes	540,000
Average Index Page Size Bytes (D)	4,000
Fill Factor (E)	75%
Available Index page size	3,000
Number of Index Pages	180
Number of Buffer Pages	100
Total Number of Index Pages	**280**

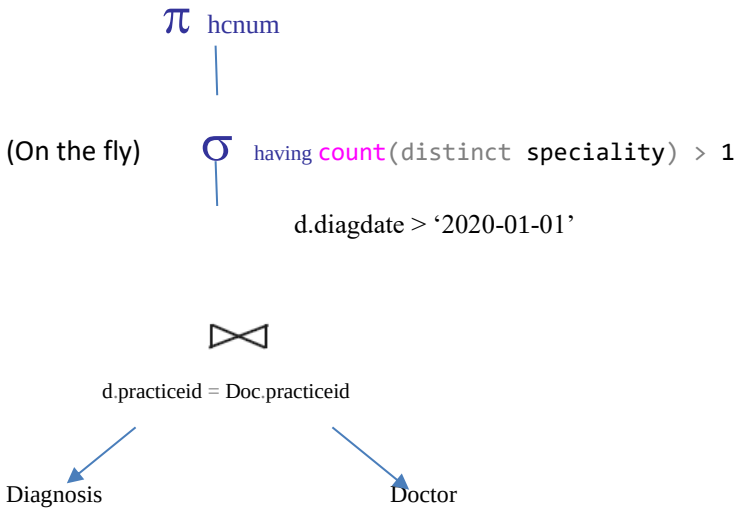
Log (280) = 8

I/O Cost = 8 + 1 = 9

Ex 4

```
select d.hcnum , count(distinct speciality)
from Diagnosis d, Doctor Doc
WHERE d.practiceid = Doc.practiceid
and d.diagdate >= '2020-01-01'
group by d.hcnum
having count(distinct speciality) > 1
```

Plan A



Plan B