Ceng352 Written Assignment 2

Bilge Eren 2171587

April 5, 2020

1

a)
We can evaluate this query with index only plan because the search key is (age,grade) so we have indexing on age and key. In this query we are looking for age and grade values since we have all information in index key we don't have to access the actual file, we can use B+ tree to get results.

b)
In this query we need "gender" information which is not a part of our search key. So we should access the actual records to get "gender" information. Therefore, index only plan evaluation is not possible on this query.

2

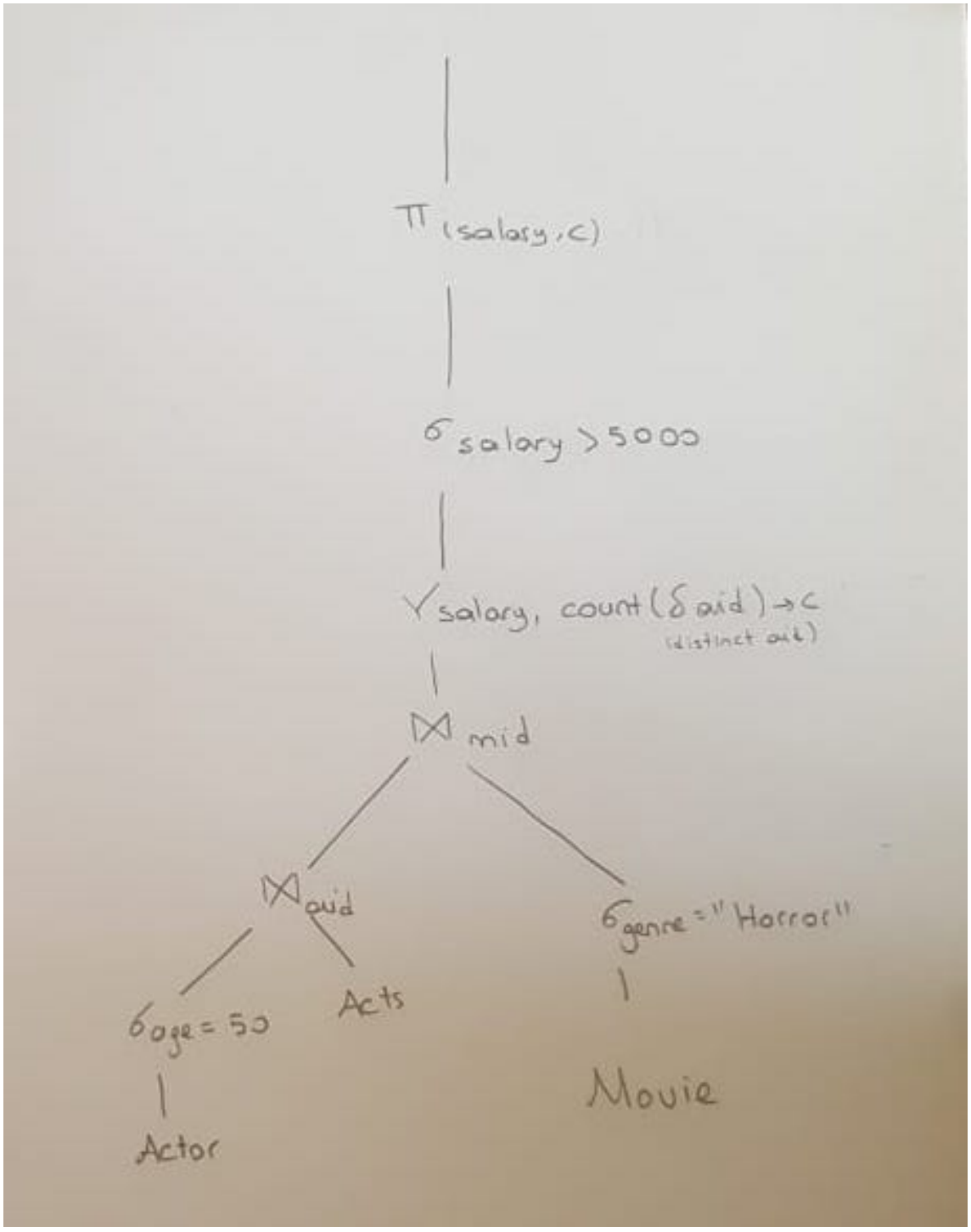
For a: Equality search;
Use a hash index on attribute R.A.

For b and c:. Range query
If we have B+tree index on R.A;
Use an unclustered B+ tree index on attribute R.A. and
Use a (clustered) B+ tree index on attribute R.A
works equally sufficient.
But if we don't have B+tree index on R.A;
Use a (clustered) B+ tree index on attribute R.A

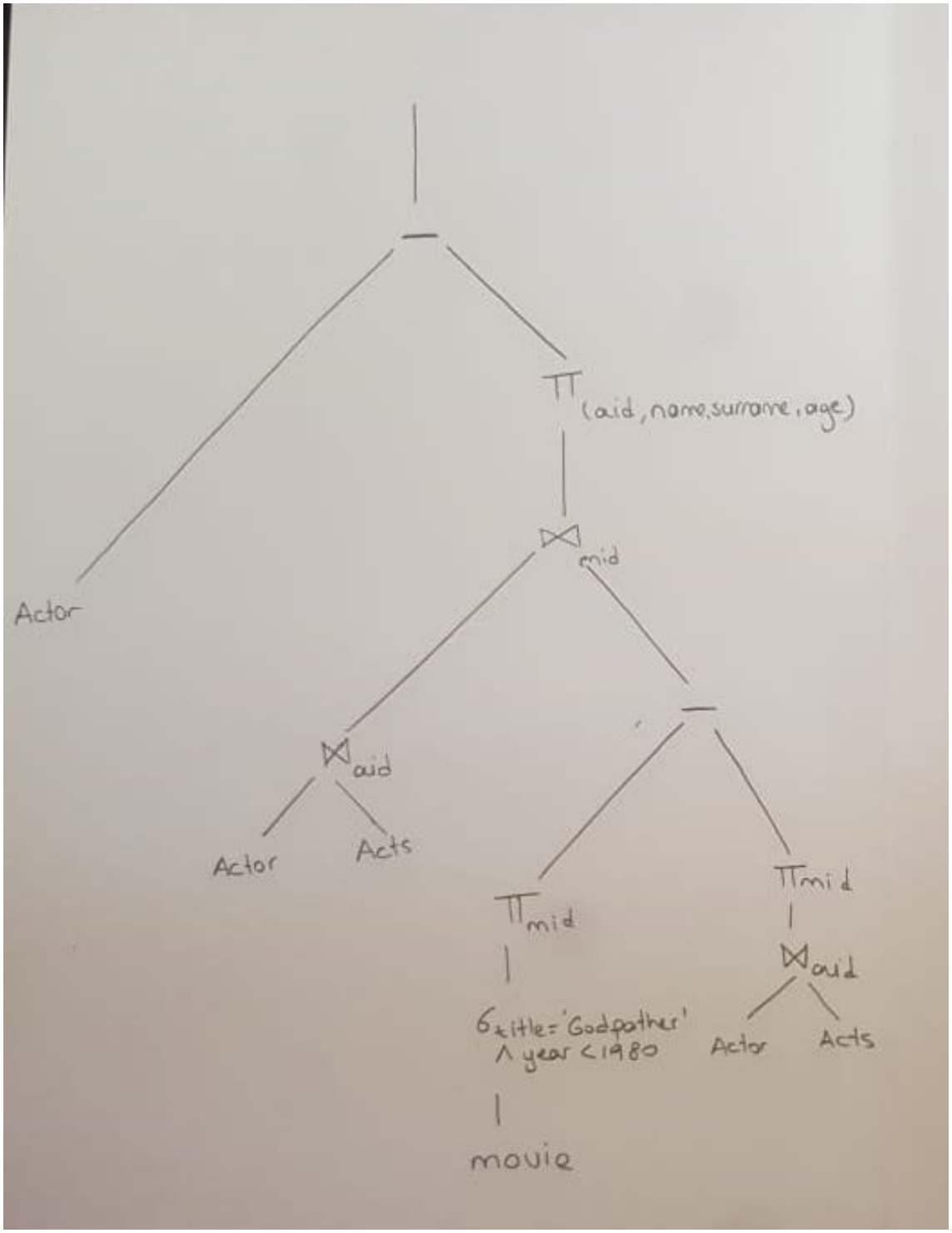
For d: All records must be traversed.
Doesnt matter which one but to avoid indexing cost, Use a heap file (i.e. an unsorted file) storing relation R.

3

a)


π (salary, c)
σ salary > 5000
γ salary, count(δ aid) → c
(distinct aid)
⋈ mid
⋈ aid
σ genre = "Horror"
σ age = 50
Acts
Actor
Movie

b)


Π (aid, name, surname, age)
⋈ mid
Actor
⋈ aid
Actor
Acts
Π mid
Π mid
⋈ aid
σ title='Godfather' ∧ year < 1980
Actor
Acts
movie

4

T(R) = 20.000 .... B(R) = 2.000
T(S) = 5.000 ..... B(S) = 500
M = 42
V(S,b) = # of distinct values of attribute b in S. = 5000 since b is primary key

4.1

Cost = B(R) + B(R)B(S)/(M-2)
Cost = 2000 + 2000.500/40
Cost = **27000**

4.2

Cost = B(S) + B(S)B(R)/(M-2)
Cost = 500 + 2000.500/40
Cost = **25500**

4.3

Cost = 7B(R) + 3B(S)
Cost = 7.2000 + 3.500
Cost = **15500**
Steps:

Since B(R) > $M2$ condition this process does not fit 2 pass. We should make 3 pass to sort R. So step (1) runs 3 times.

Total Sorting cost for R = 6B(R)
6B(R) = (Read R + Write sorted chunks 2B(R)) + (Read sorted chunks + Write bigger sorted chunks 2B(R))+(Read bigger sorted chunks+sort + Write 2B(R))

(1) Sorting R with **42** memory page.
Read from disk into memory → sort in memory → write disk from memory. $Ceil(2000/42)$ runs.
(2) Sorting S with **42** memory page.
Read from into memory disk → sort in memory → write disk from memory. $Ceil(500/42)$ runs.
(3) Merge R with S **41** memory page input, **1** memory l page output.
Read from disk into memory → merge in memory → write disk from memory.

4.4

Cost = 3B(R) + 3B(S)
Cost = 3.2000 + 3.500
Cost = **7500**
Steps:
* Hash R into **41**buckets.
Read from disk into memory → hash in memory → write disk from memory.

* Hash S into **41** buckets.
Read from into memory disk → hash in memory → write disk from memory.
* Join every pair of buckets **1** memory page input, **1** memory l page output, **1** memory page for R.
Read from disk into memory → merge in memory → write disk from memory.

4.5

Steps:
* Iterate over R
* For each tuple in R fetch corresponding tuple(s) from S.

Clustered Cost = B(R) + T(R)B(S)/V(S,b)
Clustered Cost = **2000 + 20000*500/5000 = 4000**

Unclustered Cost = B(R) + T(R)T(S)/V(S,b)
Unclustered Cost = **2000 + 20000*5000/5000 = 22000**

5

5.1

$(m_{10} + m_{11} + m_{12}).\frac{t_{swim}}{N}$

5.2

In this estimation we assume that there is a uniform distribution on records.
Since in real life we can not be sure about whether the records have uniform distribution or not, this estimation may be incorrect.