CS 4750 Assignment 4

Aleksander Schultz (aps4yj), Kavin Bapat (kb7pcr), Justin Kaplan (jrk5ak)

PART 1

1.

1. $\pi_{sname, email}$ (σ_cID='CS4750'(Student ⋈ Grades))

2. π_{sID} (Student) - (π_sID(σ_grade='D'(Grades)))

3. π_{cID}, cname (Course ⋈ (ρ_instructorCount(π_cID, count(insID)(Course)) ⋈ instructorCount>1))

4. $\pi_{cID, cname}$ (σ_{num_ins>=2} (_{count(*)->num_ins}G_{cID}(π_{cID, cname}(Course) ⋈ π_{cID, insID}(Course))))

5. $\pi_{sID, sname}$ (Student) ∪ π_insID, iname (Instructor)

2.

1. π_{sname} ((π_{sid}(π_{pid}(σ_{color='red'}(Part))⋈_{Catalog.pid=Part.pid} Catalog))⋈_{Supplier.sid=Catalog.sid} Supplier)

2. π_{sid}((σ_{color='red'}(Part))⋈_{Catalog.pid=Part.pid} Catalog)- π_{sid}(σ_{color≠'red'}(Catalog⋈_{Part.pid=Catalog.pid} Part))

3. $(\pi_{sid,pid}(Catalog)) ÷ (\pi_{pid}(\sigma_{cost < 20}(Catalog)))$

PART 2

1. $\pi_{fname,lname,number\_movies}$ (_{did}G_{count(*)->number_movies}(Directors ⋈_{Directors.did=Movie_Directors.did} Movie_Directors))

2. The following is the RA tree:

$$\Pi_{fname, lname, number\_movies}$$

$$|$$

$$_{did}G_{count(*) \rightarrow number\_movies}$$

$$|$$

$$\bowtie_{did}$$

Directors   Movie_Directors
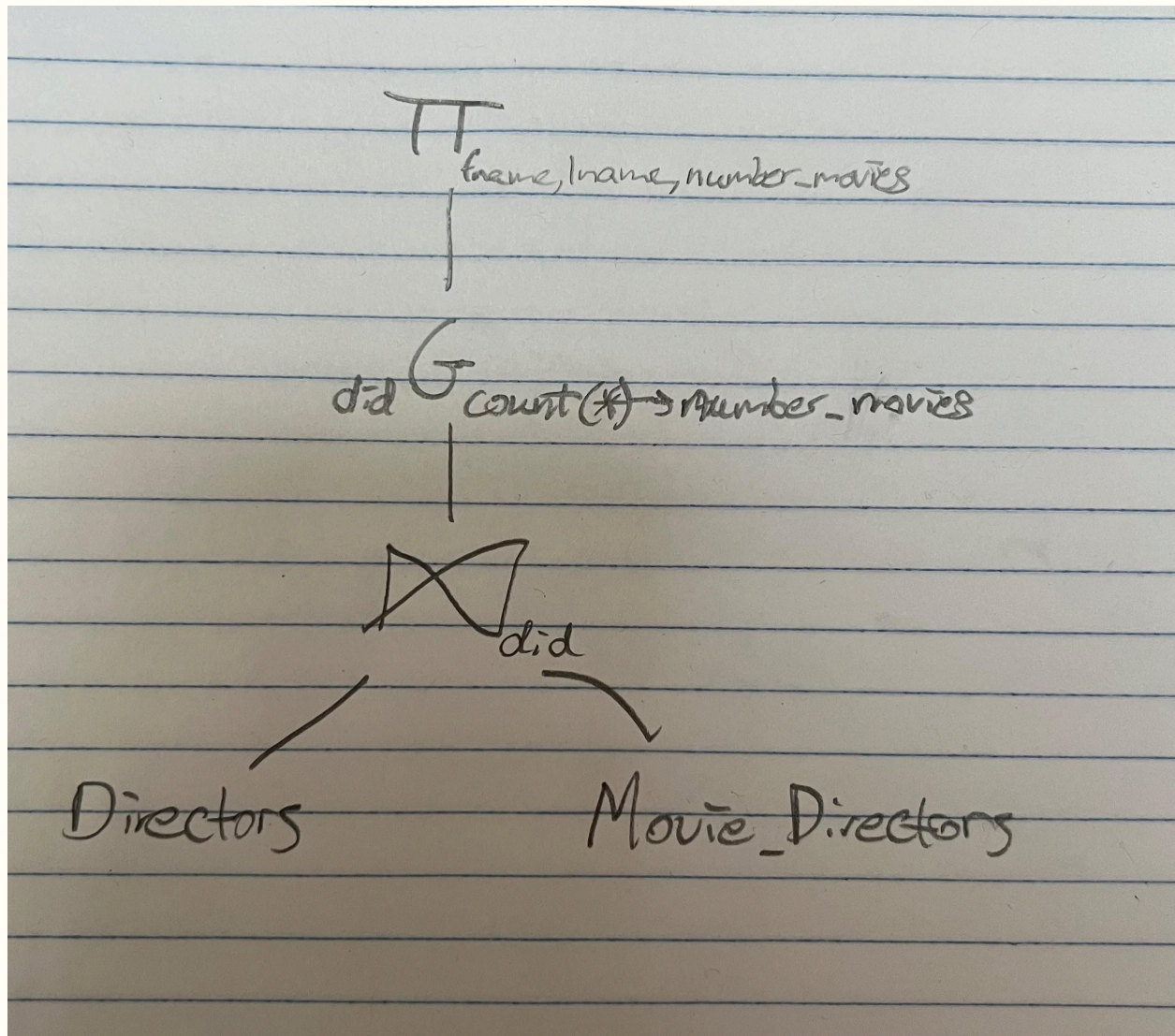
Next, we can estimate the number of queries.
We know we can compute the number of queries after the
join as T(Directors)*T(Movie_Directors) /
max{V(Directors, did), V(Movie_Directors, did)} = 480
* 600 / max(480, 450) = 288000/480 = 600.
However, this doesn't matter because next, we are
grouping by did, of which we know there are 450
distinct values (V(Movie_directors, did) = 450). Thus,
there are **450** expected output tuples.

3. The range of number_movies is hard to estimate without knowing more about the table and the distribution of movies. The minimum value is 1, because a director must direct at least 1 movie to appear in the joined table. The theoretical maximum for number_movies would occur if 449 of the directors directed exactly 1 movie, and the 450th director directed all the other 600-449=151 movies. In this case, the range is **1 to 151**.