

Database Systems

14 Q&A and Exam Preparation

Matthias Boehm

Graz University of Technology, Austria
Computer Science and Biomedical Engineering
Institute of Interactive Systems and Data Science
BMVIT endowed chair for Data Management

Announcements/Org

#1 Video Recording

- Since lecture 03, video/audio recording
- Link in [TeachCenter](#) & [TUbe](#)

#2 Exercises

- Exercise 1 graded, feedback in TC, office hours
- [Exercise 2,3](#) in progress of being graded
- [Exercise 4 due Jun 25, 11.59pm](#)

#3 Course Evaluation

- Evaluation period: [Jun 18 – Aug 13](#)
- Please, participate w/ honest feedback (pos/neg)

#4 CS Talks x4 ([today 5pm](#), Aula Alte Technik)

- [Claudia Wagner](#) (University Koblenz-Landau, Leibnitz Institute for the Social Sciences)
- Title: [Birds of a Feather Flock Together: The Influence of Social Homophily on the Visibility of Minorities Online](#)



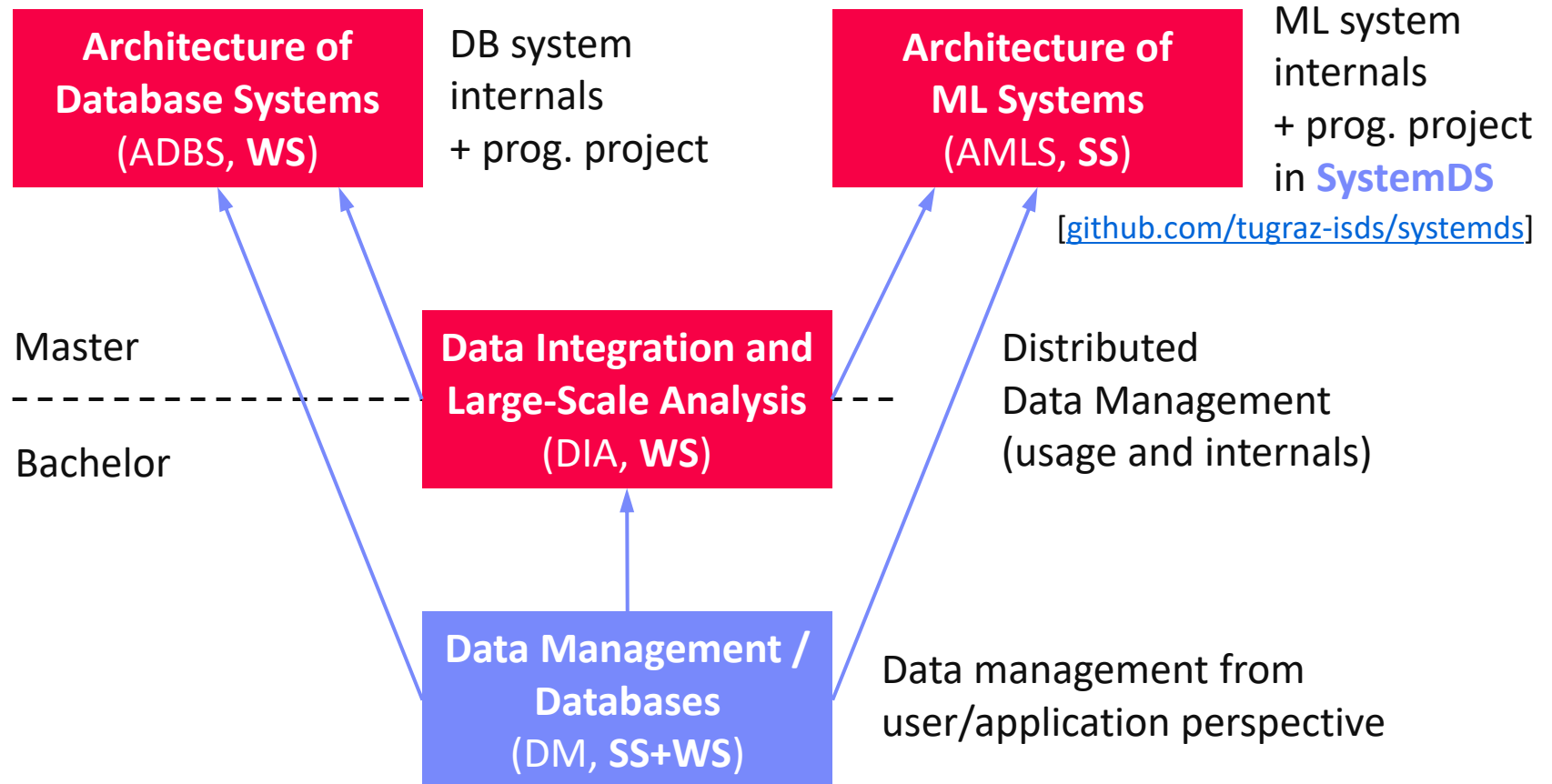
77.4%

60.4%

47.2%



#5 Data Management Courses



Summary Exam

- **Alternative Exam Dates**

- Jun 24, 4pm, HS i13 (41/160)
- Jun 27, 4pm, HS i13 (113/160)
- Jun 27, 7.30pm, HS i13 (24/160)
- Don't forget to register 2 days before the exam

**We'll create 6
different exams
(3 x DB/DB1)**

- **Exam Logistics**

- **Start 10min after, 90min working time** → more than enough
- Bring **identification and student card**, paper, and pen (not red)
- No lecture materials or mobile devices allowed
- Different exams for databases and databases 1 → **don't pick wrong one**

- **Exam Preparation**

- Exercises 1-4!
- Read through the lecture slides

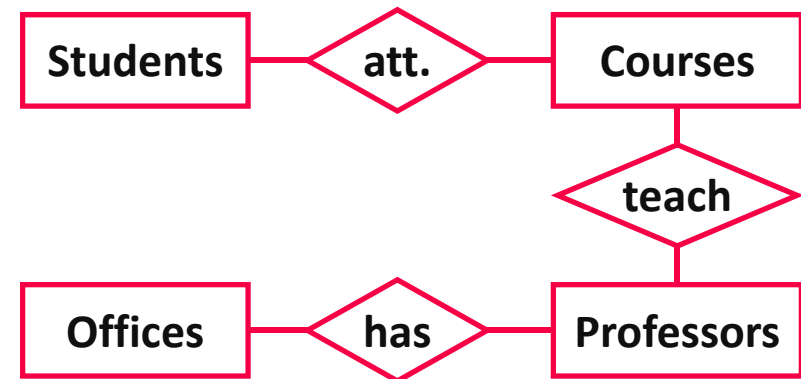
Exam Preparation

Basic focus: fundamental concepts and ability to apply learned techniques to given problems

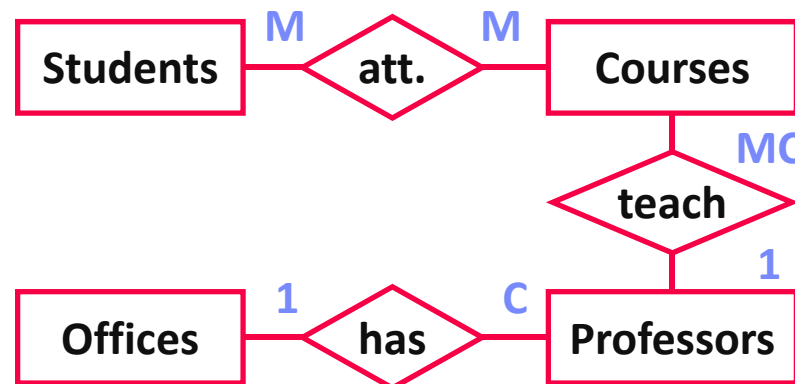
#1 Entity-Relationship Diagrams

Task 1a: Fill in the cardinalities in Modified Chen notation

- Students need to attend at least 5 courses
- Course is taught by 1 professor, at least 3 students must attend
- Only single-room offices, every person has one and only one office



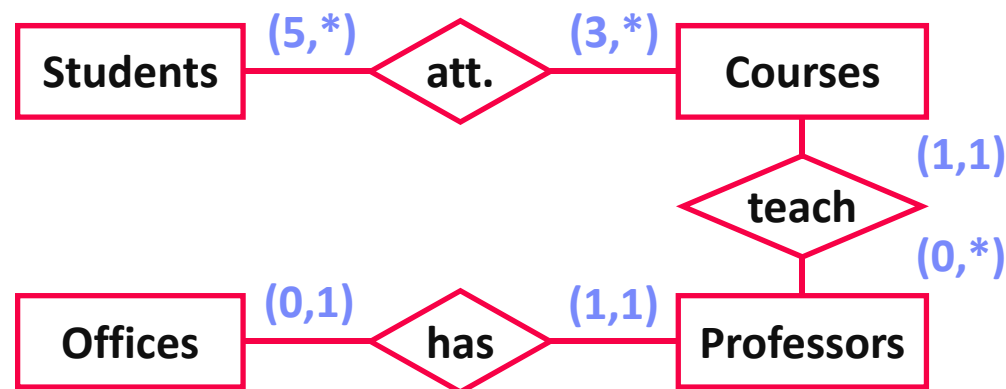
Solution



#1 Entity-Relationship Diagrams, cont.

- **Task 1b: Fill in the cardinalities in (min,max)-notation**
 - Students need to attend at least 5 courses
 - Course is taught by 1 professor, at least 3 students must attend
 - Only single-room offices, every person has one and only one office

- **Solution**



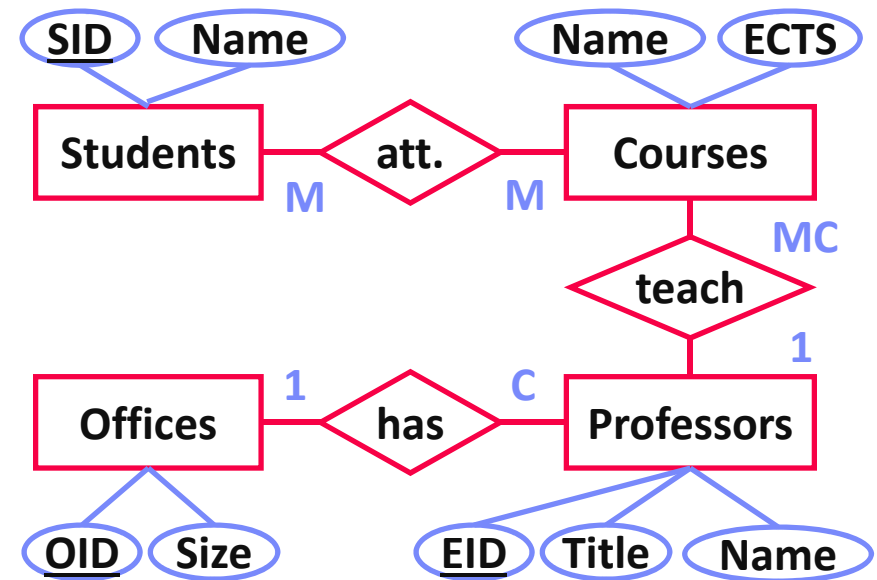
- **Remark on Perspectives**

- Chen Notation focuses on entities
- Min/Max notation focuses on relationships



#2 Relational Schema

- **Task 2: Create a normalized relational schema**
 - Given ER diagram
 - Map into relational model
 - Simplify and normalize if necessary
 - Describe why normalized



- **Solution**

- Students(SID:int, Name:varchar(128))
 - **StudentCourse**(SID:int, CID:int)
 - Courses(CID:int, Name:varchar(256), ECTS:int, EID:int)
 - Professors(EID:int, Title:varchar(128), Name:varchar(256), OID:int)
 - Offices(OID:int, Size:int)
- + Explanation Normal Forms**

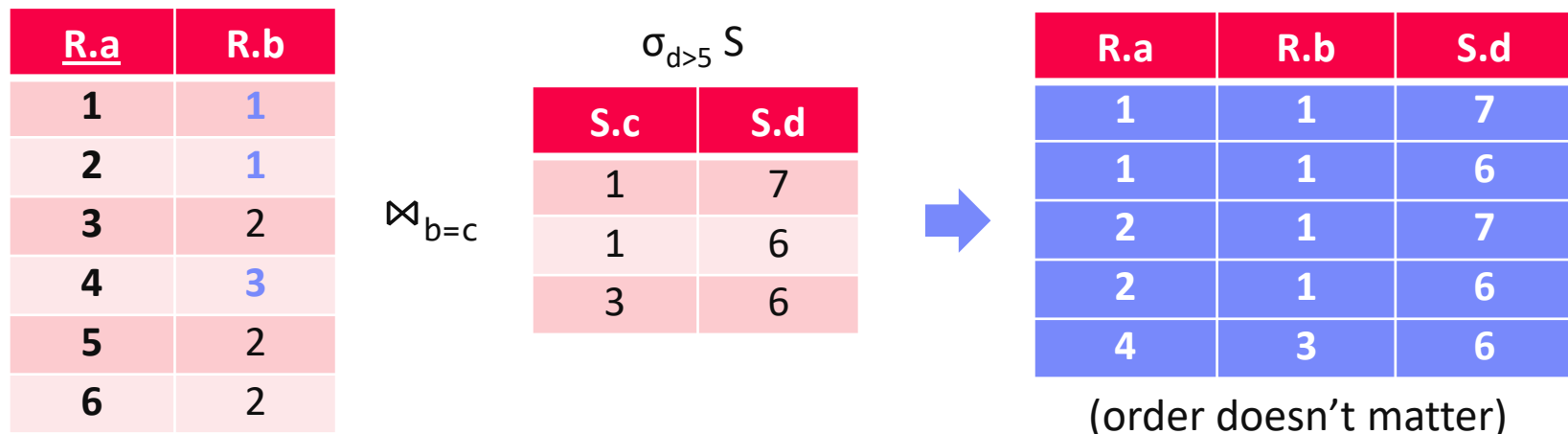
#3 SQL and Relational Algebra

- Task 3a: What are the results for the given query and data?

```
SELECT R.a, R.b, S.d
FROM R JOIN S ON R.b = S.c
WHERE S.d > 5
```

R.a	R.b	S.c	S.d
1	1	1	7
2	1	1	6
3	2	2	5
4	3	3	6
5	2	4	3
6	2	4	2

- Solution



#3 SQL and Relational Algebra, cont.

- Task 3b: What are the results for the given query and data?

```
SELECT R.b, S.b, count(*)
FROM R, S
WHERE R.c = S.a
AND S.c != 7
GROUP BY R.b, S.b
ORDER BY (R.b, S.b) ASC
```

R.a	R.b	R.c	S.a	S.b	S.c
1	X	3	1	A	6
2	X	2	2	A	5
3	Y	7	3	B	4
4	X	2	4	C	3
5	Y	6	5	B	2
6	Z	5	6	B	7
			7	A	1

- Solution

R.a	R.b	R.c	S.b	S.C
1	X	3	B	4
2	X	2	A	5
3	Y	7	A	1
4	X	2	A	5
6	Z	5	B	2



R.b	S.b	count
X	A	2
X	B	1
Y	A	1
Z	B	1

#4 Query Languages (SQL)

- **Task 4a: Write a SQL query to find the name(s) of professors with the largest office**

- Return professor name

Students(SID:int, Name:varchar(128))
StudentCourse(SID:int, CID:int)
Courses(CID:int, Name:varchar(256),
ECTS:int, EID:int)
Professors(EID:int, Title:varchar(128),
Name:varchar(256), OID:int)
Offices(OID:int, Size:int)

- **Solution**

```
SELECT P.Name
FROM Professors P, Offices O
WHERE P.OID = O.OID
AND O.Size = (SELECT max(Size) FROM Offices)
```

#4 Query Languages (SQL), cont.

- **Task 4b: Write a SQL query to find the courses attended by more than 100 students**

- Return course name and number of students

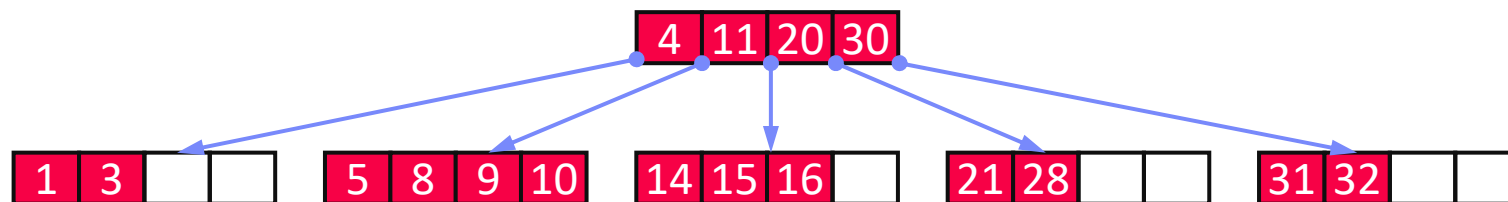
Students(SID:int, Name:varchar(128))
StudentCourse(SID:int, CID:int)
Courses(CID:int, Name:varchar(256),
 ECTS:int, EID:int)
Professors(EID:int, Title:varchar(128),
 Name:varchar(256), OID:int)
Offices(OID:int, Size:int)

- **Solution**

```
SELECT C.Name, count(*)
FROM StudentCourse SC, Courses C
WHERE SC.CID = C.CID
GROUP BY C.Name
HAVING count(*) > 100
```

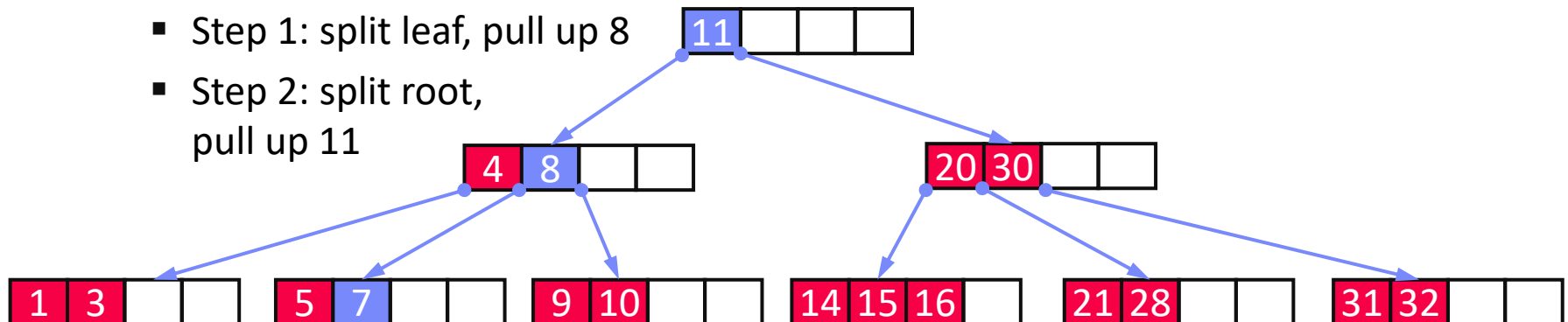
#5 Physical Design – B-Trees

- **Task 5a: Insert 7 and draw the resulting B-Tree**
 - Degree $k=2$



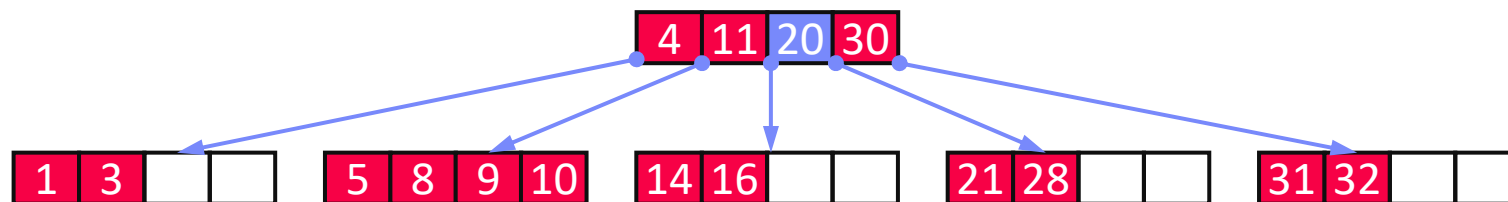
- **Solution**

- Step 1: split leaf, pull up 8
- Step 2: split root, pull up 11



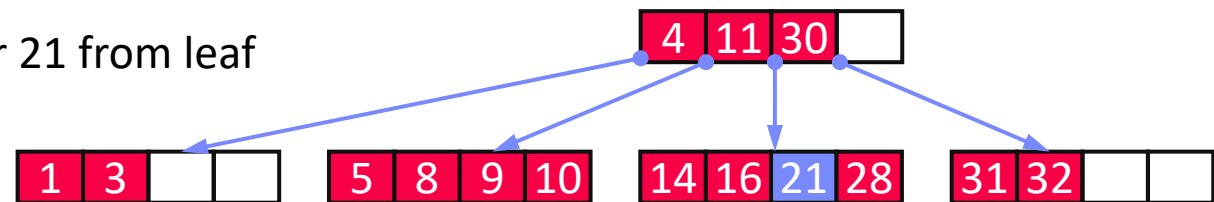
#5 Physical Design – B-Trees, cont.

- **Task 5b: Delete 20 and draw the resulting B-tree**
 - Degree $k=2$

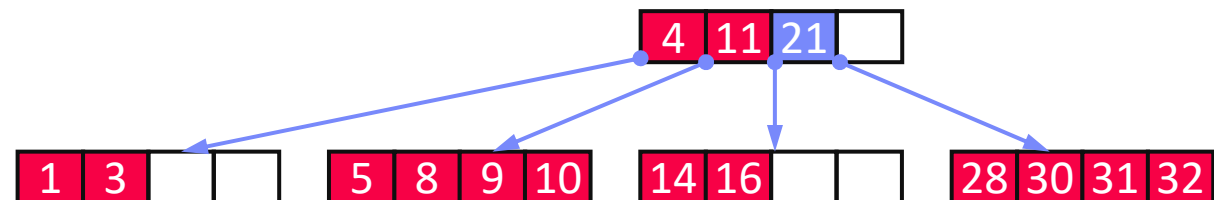


▪ Solution

- Step 1: Pull 16 or 21 from leaf
- Step 2: Merge siblings



Valid
alternative



#6 Query Processing – Query Trees

- Task 6a: Draw the logical query tree for the given query

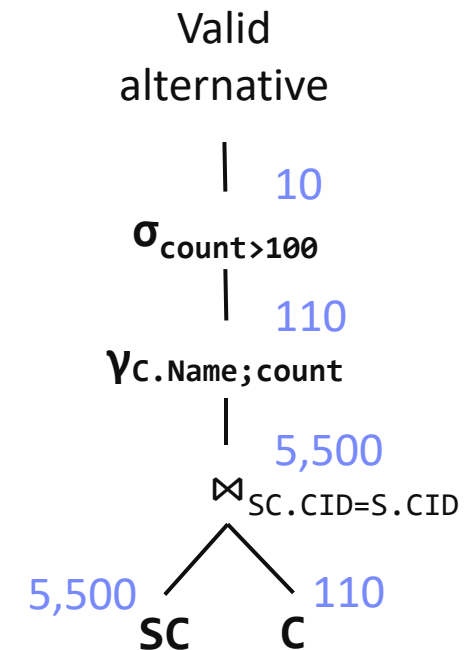
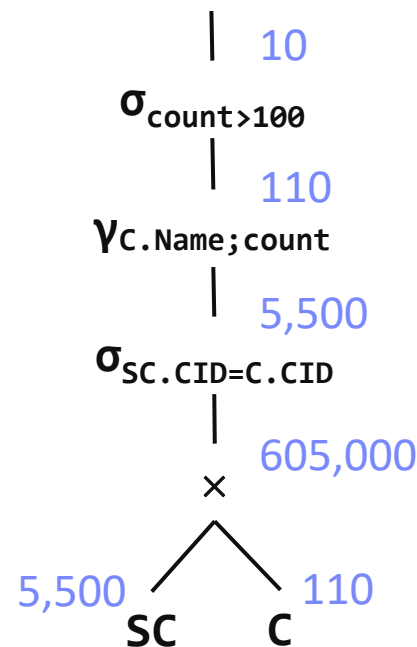
- In relational algebra

- Task 6b: Annotate the cardinality of intermediates

- $| \text{Courses} | = 110$
- $| \text{StudentCourses} | = 5500$
- $| \text{Course w/ } >100 | = 10$

- Solutions

```
SELECT C.Name, count(*)
FROM StudentCourse SC, Courses C
WHERE SC.CID = C.CID
GROUP BY C.Name
HAVING count(*) > 100
```



#7 Query Processing – Operators

- **Task 7a: Discuss the space and time complexity of nested loop and hash joins in Big-O notation**
 - Input relations of size M and N with $M \leq N$

→ Solution

- Nested loop join: space $O(1)$, time $O(M * N)$
 - Hash join: space $O(M)$, time $O(M + N)$
-
- **Task 7b: Why are nested loop joins still relevant despite their poor time complexity?**

→ Solution

- Arbitrary join predicates (**theta joins**), not just equality predicates

#8 Transaction Processing – Isolation Levels

- **Task 8: Which guarantees does the isolation level READ COMMITTED provide and what problems does it solve?**
- **Solution**
 - Transactions read only modifications of committed transactions
 - Inconsistent intermediates of failed/aborted transactions do not leak into updates of successful transactions

```
UPDATE Students SET Pts=100  
WHERE Sid=789;
```

ROLLBACK TRANSACTION;

**Student received 124
instead of 24 points**



```
SELECT Pts INTO :points  
FROM Students WHERE Sid=789;  
points += 24.0;  
UPDATE Students SET Pts=:points  
WHERE Sid=789;  
COMMIT TRANSACTION;
```

#9 NoSQL: Data Partitioning

Task 9: Horizontal Data Partitioning

- Assume a key-value collection $R(k,v)$ stored in a distributed key-value store
- a) Partition this collection horizontally into three fragments (via relational algebra expressions) that are **complete**, **reconstructable**, and **disjoint**
- b) Reconstruct the collection from the fragments via a relational algebra expression

R

<u>k</u>	v
1	Blob1
2	Blob2
4	Blob4
7	Blob7
15	Blob15
9	Blob9
14	Blob14
8	Blob8

Solution

- a) Partitioning

$$R1 = \sigma_{k \leq 5}(R)$$

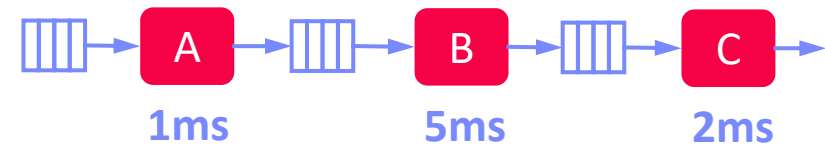
$$R2 = \sigma_{k > 5 \wedge k \leq 10}(R)$$

$$R3 = \sigma_{k > 10 \wedge k \leq 15}(R)$$
- b) Reconstruction

$$R = (R1 \cup R2) \cup R3$$

#10 Stream Processing – Overload Handling

- **Task 10: Given the continuous query A-B-C, what are the resulting characteristics?**



- Bounded queues of size $b=4$, parallel operators
 - Max throughput (tuples per second)?
 - Min and max tuple latency (system time, started on entering queue)?
 - What are means of handling overload?
-
- **Solution**
 - **Max throughput:** $1/\max(C(op_i)) = 1/5 \text{ tuples/ms} = 200 \text{ tuples/s}$
 - **Min tuple latency:** $\sum(C(op_i)) = 1\text{ms} + 5\text{ms} + 2\text{ms} = 8\text{ms}$
 - **Max tuple latency:** $2 \cdot (b+1) \cdot 5\text{ms} + 2\text{ms} = 52\text{ms}$
 - **Overload handling:** back pressure (slow down sources), load shedding (e.g., random dropping of tuples $> 200 \text{ tuples/s}$), distributed stream processing (e.g., key range partitioning)

Remaining Questions & Answers

Conclusions and Q&A

- **Summary 14 Q&A and Exam Preparation**
 - Exam Preparation → different questions but same flavor
 - Remaining Questions & Answers

- **Exercise 4: due Jun 25, 11.59pm**
 - News group questions answered until June 25, 5pm

- **Next Lectures/Exams**
 - **Jun 24, 4pm Exam** DB / DB1, HS i13
 - **Jun 27, 4pm Exam** DB / DB1, HS i13
 - **Jun 27, 7.30pm Exam** DB / DB1, HS i13