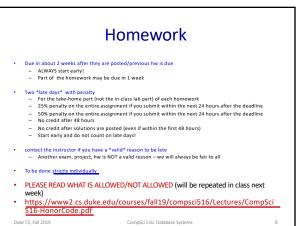


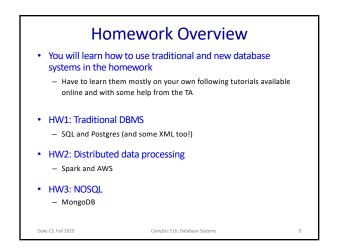
## **Grading Strategy**

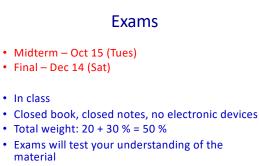
- Relative grading
  - The actual grade distribution at the end will depend on the performance of the entire class on all the components.
  - Topper of the class gets A+ irrespective of the number, and all and only "above expectation" performances get A+.
  - No fixed lowest grade or grade distribution.
  - Everyone can get good grade by working hard!

Duke CS, Fall 2019

CompSci 516: Database Systems







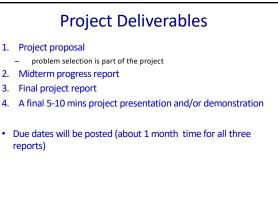
CompSci 516: Database Systems

Both exams are comprehensive

 would include every lecture up to the exams

Duke CS, Fall 2019

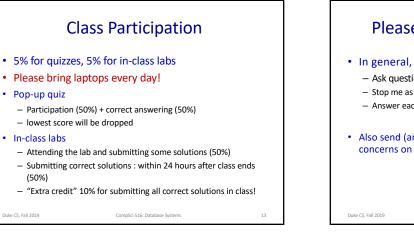
# <section-header> Projects 10% weight 10 groups of 3-4 20 cops of smaller and larger sizes need instructor's permission 20 cops of smaller and larger sizes need instructor's permission 20 cops of smaller and larger sizes need instructor's permission 20 cops of smaller and larger sizes need instructor's permission 20 cops of smaller and larger sizes need instructor's permission 20 cops of smaller and larger sizes need instructor's permission 20 cops of smaller and larger sizes need instructor's permission 20 cops of smaller and index for projects will be posted larger



Duke CS, Fall 2019

11

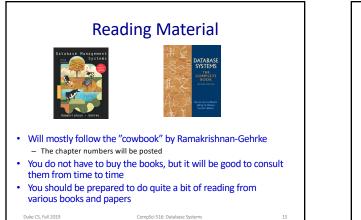
14

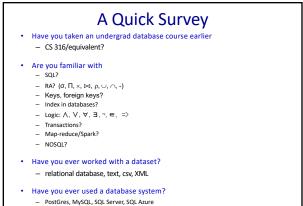


## Please ask questions in class!

- In general, actively participate in the class!
  - Ask questions in class and on piazza
  - Stop me as many times as you need to understand the lectures
  - Answer each other's questions on piazza
- Also send (anonymous or not) feedback, suggestions, or concerns on Piazza or by email

CompSci 516: Database System





CompSci 516: Database Svs

Duke CS, Fall 2019

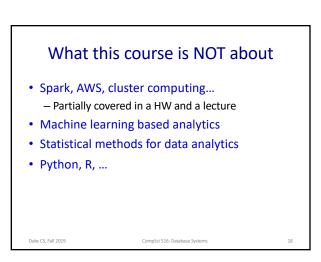
## What is this course about?

- This is a graduate-level database course in CS
  - We will cover principles, internals, and applications of database systems in depth
- Database concepts
  - Data Models, SQL, Views, Constraints, RA, Normalization
- Principles and internals of database management systems (DBMS)
   Indexing, Query Execution-Algorithms-Optimization, Transactions, Parallel and Distributed Query Processing, Map Reduce
- Advanced and research topics in databases
  - e.g. Datalog, NOSQL, Data mining, ...

Duke CS, Fall 2019

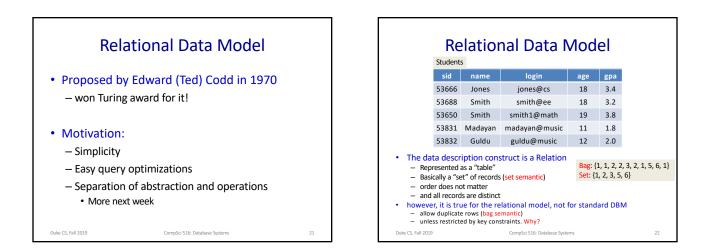
CompSci 516: Database Systems

17

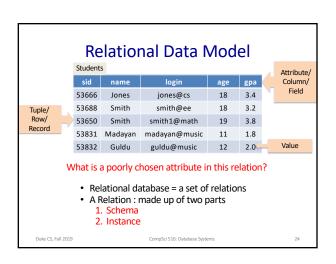




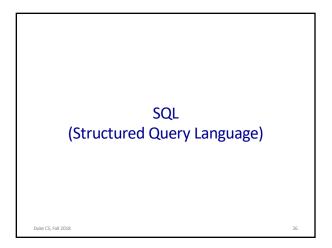




	Student		ag vs. Set			
	sid	name	login	age	gpa	
	53666	Jones	jones@cs	18	3.4	
	53688	Smith	smith@ee	18	3.2	
	53650	Smith	smith1@math	19	3.8	
	53831	Madayan	madayan@music	11	1.8	
	53832	Guldu	guldu@music	12	2.0	
DBMSs? – Prima – Duplic	rily perf cate elin operati	formance r nination is	d not "set sema easons expensive (require rojection"s are mu	es sorting	g)	
Duke CS, Fall 2019			CompSci 516: Database Syste	ems		23



## Schema and Instance One schema can have multiple instances Schema: A template for describing an entity/relationship (e.g. students) specifies name of relation + name and type of each column e.g. Students(sid: string, name: string, login: string, age: integer, gpa: real). Instance: - When we fill in actual data values in a schema a table, has rows and columns - each row/tuple follows the schema and domain constraints #Rows = cardinality, #fields = degree or arity example below sid name login age gpa Cardinality = 3, degree = 5 53666 Jones jones@cs 18 3.4 53688 Smith smith@ee 18 3.2 53650 Smith smith1@math 19 3.8 Duke CS, Fall 2019 CompSci 516: Database Sy



## Relational Query Languages • A major strength of the relational model: supports simple, powerful <u>querying of data</u>. • Queries can be written intuitively, and the DBMS is responsible for an efficient evaluation • The key: precise semantics for relational queries • Based on a sound theory! • Allows the optimizer to extensively re-order operations, and still ensure that the answer does not change.



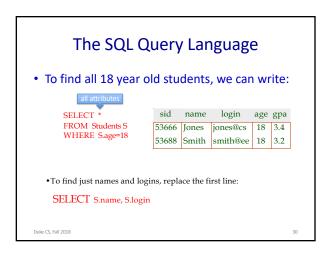
- Developed by IBM (systemR) in the 1970s based on Ted Codd's relational model

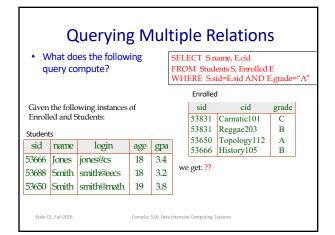
   First called "SEQUEL" (Structured English Query Language)
- First commercialized by Oracle (then Relational Software)in 1979
- Standards by ANSI and ISO since it is used by many vendors
  - SQL-86, -89 (minor revision), -92 (major revision), -96, -99 (major extensions), -03, -06, -08, -11, -16

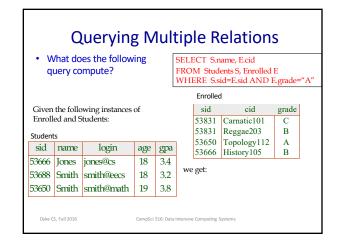
Duke CS, Fall 2018

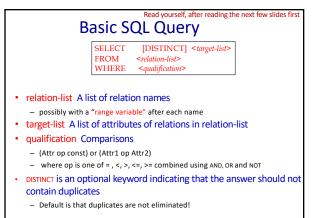
## Purposes of SQL • Data Manipulation Language (DML) – Querying: SELECT-FROM-WHERE – Modifying: INSERT/DELETE/UPDATE (next week) • Data Definition Language (DDL) – CREATE/ALTER/DROP (next week)

Duke CS, Fall 2018



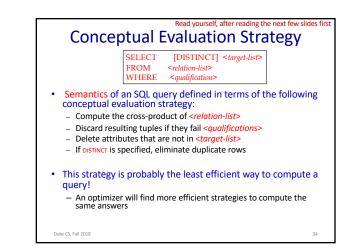


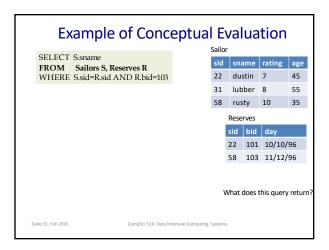


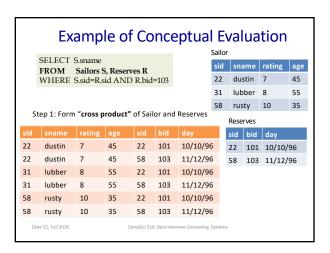


33

Duke CS, Fall 2018

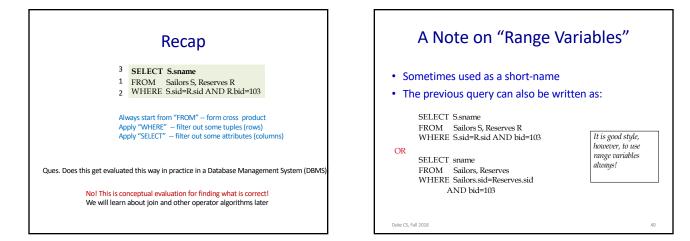


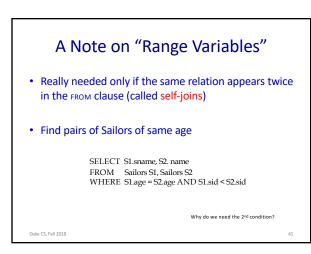


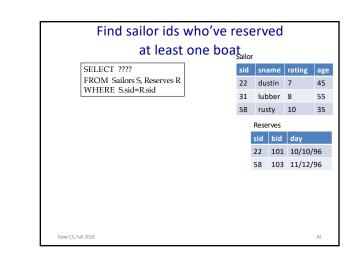


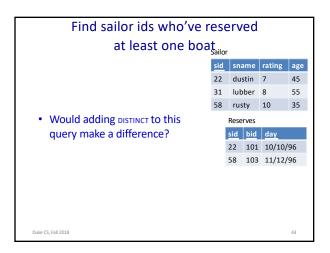
	EX	amp	le of	Co	nce	ptua	Sailo		ua	tion	
	SELECT		n	Б			sid	sn	ame	rating	age
	FROM WHERE	Sailors S, S.sid=R.		d=103		22	du	stin	7	45	
	31 lubber 8 55										
	58 rusty 10 35										35
Step 2	2: Discard	tuples th	at do no	t satisf	y <qua< th=""><th>lification&gt;</th><th></th><th>Rese</th><th>rves</th><th></th><th></th></qua<>	lification>		Rese	rves		
sid	sname	rating	age	sid	bid	day		sid	bid	day	
22	dustin	7	<del>45</del>	22	<del>101</del>	<del>10/10/90</del>	•	22	101	10/10/	96
<del>22</del>	<del>dustin</del>	7	<del>45</del>	<del>58</del>	<del>103</del>	<del>11/12/90</del>	•	58	103	11/12/	96
<del>31</del>	lubber	8	<del>55</del>	22	<del>101</del>	10/10/96	•				
<del>31</del>	lubber	8	<del>55</del>	<del>58</del>	<del>103</del>	<del>11/12/90</del>	•				
58	rusty	<del>10</del>	<del>35</del>	22	<del>101</del>	10/10/96	•				
58				58	103	11/12/96					

### **Example of Conceptual Evaluation** Sailor SELECT S.sname sid sname rating age FROM Sailors S, Reserves R 22 dustin 7 45 WHERE S.sid=R.sid AND R.bid=103 31 lubber 8 55 58 rusty 10 35 Step 3: Select the specified attribute(s) Reserves sid bid day 22 7 45 22 101 10/10/96 22 101 10/10/96 7 <del>45</del> <del>58</del> <del>11/12/96</del> 22 <del>103</del> 58 103 11/12/96 22 21 0 55 101 10/10/06 58 31 8 55 103 11/12/06 <del>58</del> <del>10</del> 35 22 101 10/10/96 rusty 58 rusty 10 35 58 103 11/12/96 Duke CS, Fall 2016 CompSci 516: Data In

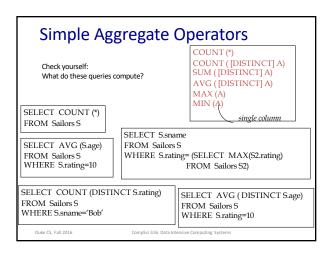


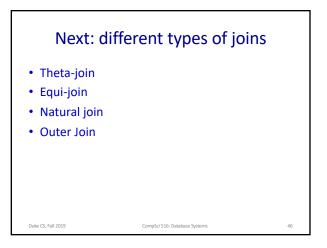








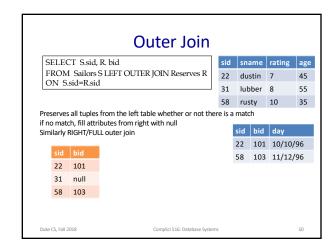


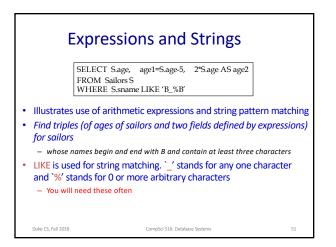


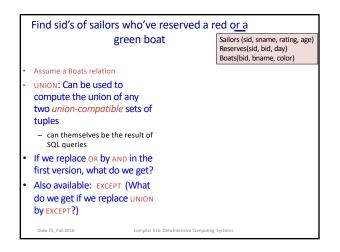
	Condition/Theta Join											
											age	
FROM Sailors S, Reserves R 22 dustin 7 45												
WHERE S.sid=R.sid and age >= 40 31 lubber 8 55												
							58	rus	sty	10	35	
Fo	rm cross pr	oduct, dis	card rows	that d	lo not sa	tisfy the co	nditio	on				
sid	sname	rating	age	sid	bid	day		sid	bid	day		
22	dustin	7	45	22	101	10/10/9	5	22	101	10/10/	96	
22	dustin	7	45	58	103	11/12/9	5	58	103	11/12/	96	
31	lubber	8	55	22	101	10/10/9	6					
31	iubber	8	55	58	103	11/12/9	6—					
58	rusty	10	35	22	101	10/10/9	5					
58	rusty	10	35	58	103	11/12/9	5					
Duk	e CS, Fall 2018			Cor	mpSci 516: C	)atabase System	5				47	

				Equ	ui Jo	oin	Equi Join											
	SELECT * sid sname rating age																	
FROM Sailors S, Reserves R 22 dustin 7 45																		
WHERE S.sid=R.sid and age = 45 31 lubber 8 55																		
A special case of theta join 58 rusty 10 35																		
Join condition only has equality predicate =																		
sid	sname	rating	age	sid	bid	day		sid	bid	day								
22	dustin	7	45	22	101	10/10/96	5	22	101	10/10/	96							
22	dustin	7	45	58	103	11/12/90	5	58	103	11/12/	96							
31	lubber	8	55	22	101	10/10/96	;											
31	lubber	8	55	58	103	11/12/96	;											
58	rusty	10	35	22	101	10/10/96	5											
58	rusty	10	35	58	103	11/12/96	;											
SX         TUST         LU         SS         S8         LU3         11/12/196           Duke CS, Fall 2018         Comp5ci 516: Database Systems         48																		

	Natural Join										
S	SELI	ECT *					si	d si	name	rating	age
F	FROM Sailors S NATURAL JOIN Reserves R 22 dustin 7 49										45
	31 lubber 8 55										
E	qual	cial case of ity conditic	on on ALL	5	3 ri	usty	10	35			
	Dupli	cate colum	ns are elir	ninated							
sic		sname	ns are elir rating	age	bid	day		sid	bid	day	
	d				bid 101	day 10/10/90	6	sid 22	bid 101		96
si	d 2	sname	rating	age					101	10/10/	
sid 22	d 2 2	sname dustin	rating 7	age 45	101	10/10/9	6	22	101	10/10/	
sic 22 22	d 2 2	sname dustin dustin	rating 7 7	age 45 45	101 103	10/10/90 11/12/90	6	22	101	10/10/	
22 22 31	d 2 2 1	sname dustin dustin lubber	rating 7 7 8	age 45 45 55	101 103 101	10/10/90 11/12/90 10/10/90	6 6 6	22	101	10/10/	







Sailors (sid, sname, rating, age) Reserves(sid, bid, day) Boats(bid, bname, color)
ting Systems

