SQL Part 2

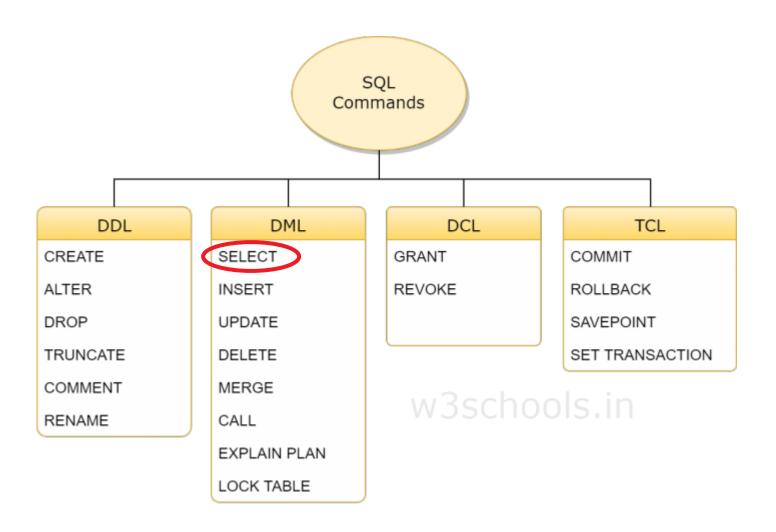
UNIVERSITY OF MANNHEIM Data and Web Science Group

CS460 Databases for Data Scientists



Parts of SQL: The Big Picture





Outline



Last week

- Overview of The SQL Query Language
- Basic Query Structure
- Join Operators
- Set Operations
- Aggregate Functions
- Null Values
- Subqueries
 - In WHERE part
 - In FROM part
 - In SELECT part

Today

- Data Definition
- Data Types in SQL
- Modifications of the database
- Views
- Integrity Constraints
- Roles & Rights

SQL Data Definition Language (DDL)



- Allows the specification of information about relations, including
 - The schema for each relation
 - The domain of values associated with each attribute
 - Integrity constraints
- And as we will see later, also other information such as
 - The set of indices to be maintained for each relations
 - Security and authorization information for each relation
 - The physical storage structure of each relation on disk

Recap: Domain of an Attribute



- The set of allowed values for an attribute
 - Programmers: think datatype

ID	name	dept_name	salary
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	<i>7</i> 5000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

Simple Domains in SQL



- char(n) Fixed length character string, with user-specified length n.
- varchar(n) Variable length character strings, with user-specified maximum length n.
- Int Integer (a finite subset of the integers that is machine-dependent).
- **smallint** Small integer (a machine-dependent subset of the integer domain type).
- numeric(p,d) Fixed point number, with user-specified precision of p digits, with d digits to the right of decimal point. (ex., numeric(3,1), allows 44.5 to be stores exactly, but not 444.5 or 0.32)
- real, double precision
 Floating point and double-precision floating point numbers, with machine-dependent precision.
- **float(n)** Floating point number, with user-specified precision of at least *n* digits.

Date and Time Data Types in SQL



- We have already encountered characters and numbers
- date: Dates, containing a (4 digit) year, month and date
 - Example: date '2005-7-27'
- time: Time of day, in hours, minutes and seconds.
 - Example: time '09:00:30' time '09:00:30.75'
- timestamp: date plus time of day
 - Example: timestamp '2005-7-27 09:00:30.75'
- interval: period of time
 - Example: interval '1' day
 - Subtracting a date/time/timestamp value from another gives an interval value
 - Interval values can be added to date/time/timestamp values

Arithmetics with Dates



- Dates can be compared
 - i.e., < or >
 - e.g., select employees who started before January 1st, 2017
 - Special function: NOW() (in MariaDB; name may differ for other DBMS)
- Dates can be added to / substracted from intervals and other dates
 - e.g., select students who have been enrolled for more than five years
- Implementation often not standardized
 - Details differ from DBMS to DBMS!

User Defined Types



CREATE TYPE construct in SQL creates user-defined type

CREATE TYPE Dollars AS NUMERIC (12,2) FINAL

```
CREATE TABLE department(

dept_name VARCHAR (20),
building VARCHAR (15),
budget Dollars
);
```

required due to SQL standard; not really meaningful

User-defined Domains



CREATE DOMAIN construct creates user-defined domain types

CREATE DOMAIN person_name CHAR(20) NOT NULL

- Types and domains are similar
 - Domains can have constraints, such as NOT NULL, specified on them

```
CREATE DOMAIN degree_level VARCHAR(10)
```

CONSTRAINT *degree_level_test*

CHECK (VALUE IN ('Bachelors', 'Masters', 'Doctorate'));

Domain Constraints vs. Table Constraints



- Some checks may reoccur over different relations
 - e.g., degrees for students or instructors
 - e.g., salutations
 - e.g., valid ranges for ZIP codes
- Binding them to a domain is preferred
 - Central definition
 - Consistent usage





Large Object Types



- Large objects (photos, videos, CAD files, etc.) are stored as a large object:
 - BLOB: binary large object -- object is a large collection of uninterpreted binary data (whose interpretation is left to an application outside of the database system)
 - CLOB: character large object -- object is a large collection of character data
- When a query returns a large object, a pointer is returned rather than the large object itself

Creating Relations



An SQL relation is defined using the CREATE TABLE command:

```
CREATE TABLE r (A_1 D_1, A_2 D_2, ..., A_n D_n, (integrity-constraint<sub>1</sub>), ..., (integrity-constraint<sub>k</sub>))
```

- r is the name of the relation
- each A_i is an attribute name in the schema of relation r
- D_i is the datatype/domain of values in the domain of attribute A_i
- Example:

```
CREATE TABLE instructor (

ID CHAR(5),

name VARCHAR(20),

dept_name VARCHAR(20),

salary NUMERIC(8,2))
```

Recap: Keys



- Primary keys identify a unique tuple of each possible relation r(R)
 - Typical examples: IDs, Social Security Number, car license plate
- Primary keys can consist of multiple attributes
 - e.g.: course ID plus semester (CS 460, FSS 2019)
 - Must be minimal (ID, semester, instructor) would work as well
- Foreign keys refer to other tables



Defining Keys



- PRIMARY KEY $(A_1, ..., A_n)$
- FOREIGN KEY $(A_m, ..., A_n)$ REFERENCES r
- Example:

```
CREATE TABLE instructor (

ID CHAR(5),

name VARCHAR(20),

dept_name VARCHAR(20),

salary NUMERIC(8,2))

PRIMARY KEY (ID),

FOREIGN KEY (dept_name)

REFERENCES department(dept_name)

);
```

Removing and Altering Relations



- Removing relations
 - DROP TABLE r

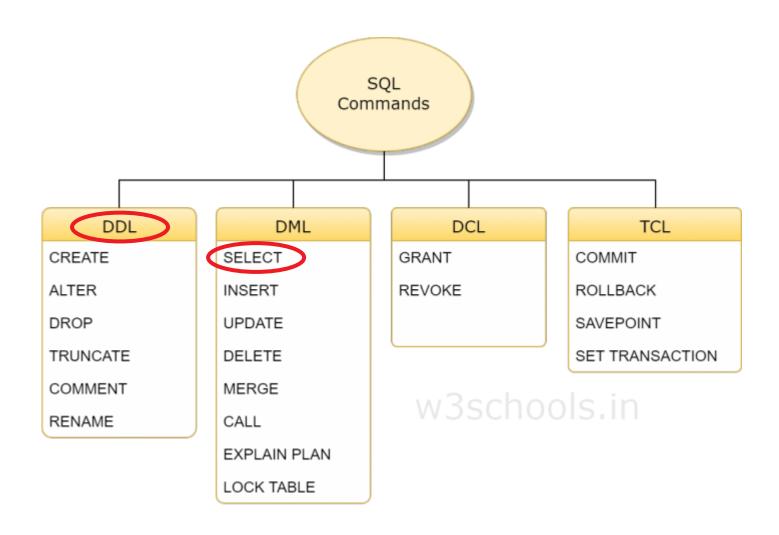


- ALTER TABLE r ADD A D
 - where A is the name of the attribute to be added to relation r, and D is the domain of A
 - all existing tuples in the relation are assigned NULL as the value for the new attribute
- ALTER TABLE r DROP A
 - where A is the name of an attribute of relation r
 - not supported by many databases



Parts of SQL: The Big Picture





Insertion into a Relation



Add a new tuple to course

INSERT INTO course

Note: there is an inherent ordering in the columns (the one we defined when we entered CREATE TABLE ...)

course id title dept_name credits

VALUES ('CS-437', 'Database Systems', 'Comp. Sci.', 4);

or equivalently

INSERT INTO course (course_id, title, dept_name, credits)
VALUES ('CS-437', 'Database Systems', 'Comp. Sci.', 4);

Add a new tuple to student with tot_creds set to null

INSERT INTO student
VALUES ('3003', 'Green', 'Finance', null);

Insertion of Data from Other Tables



 Add all instructors to the student relation with tot_creds set to 0

INSERT INTO student
SELECT ID, name, dept_name, 0
FROM instructor

 Note: the SELECT FROM WHERE statement is evaluated fully before any of its results are inserted into the relation Otherwise queries like

INSERT INTO table 1 **SELECT** * **FROM** table 1

would cause problems

Inserting Data into Relations with Constraints



- Effect of primary key constraints:
 - INSERT INTO instructor VALUES ('10211', 'Smith', 'Biology', 66000);
 - INSERT INTO instructor VALUES ('10211', 'Einstein', 'Physics', 95000);
 - ...and we defined ID the primary key!
- Effect of **NOT NULL** constraints
 - INSERT INTO instructor VALUES ('10211', NULL, 'Biology', 66000);
- Recap: DBMS takes care of data integrity

Caveats with NOT NULL Constraints

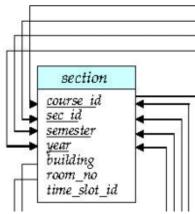


Rationale:

- Each course takes place at a specific room and time slot
- We'll create a NOT NULL constraint on those fields
- Note: no online courses here

Use case:

- First: enter all courses in the system
- Second: run clever time and room allocation algorithm
 - Which will then fill all the buildings and time slots







Example: every employee needs a substitute

```
— CREATE TABLE employee (
ID VARCHAR(5),
name VARCHAR(20) not null,
substitute VARCHAR(5) not null,
PRIMARY KEY (ID),
FOREIGN KEY (substitute) REFERENCES employee(ID)
);
```

What do you think?



Updating Data



Example: update the salary of a single person

UPDATE *employee*

SET *salary* = 80000

WHERE *person_id* = 43743

Example: update all salaries by 5%

UPDATE *employee*

SET salary = salary * 1.05

Example: moving all people from a department to a new building

UPDATE *employee*

SET building = 'Taylor' **WHERE** dept_name = 'Biology'

- Anatomy of an **UPDATE** query
 - SET defines which updates to carry out
 - WHERE defines which records to update (omitted = all records)

Updating Data



- Cut salaries above 100,000 by 5%, below 100,000 by 3%
- Write two **UPDATE** statements:

Thought experiment: Tom's salary is 102,000

```
UPDATE instructor
SET salary = salary * 0.95
WHERE salary > 100000;
UPDATE instructor
SET salary = salary * 0.97
WHERE salary <= 100000;</pre>
```

• Should rather be done using the CASE statement (next slide)





Cut salaries above 100,000 by 5%, below 100,000 by 3%

```
UPDATE instructor

SET salary = CASE

WHEN salary > 100000 THEN salary * 0.95

ELSE salary * 0.97

END
```

Updates with Subqueries



PREcompute and update tot_creds value for all students

UPDATE student S

SET tot_cred = (SELECT SUM(credits)

FROM takes, course

WHERE takes.course_id = course.course_id AND

S.ID= takes.ID AND takes.grade <> 'F' AND

takes.grade IS NOT NULL);

- Sets tot_creds to null for students who have not taken or passed any course
- Instead of SUM(credits), use:

```
CASE
WHEN SUM(credits) IS NOT NULL THEN SUM(credits)
ELSE 0
END
```



Delete

Remove all tuples from the *student* relation
 DELETE FROM *instructor*





Delete all instructors from the Finance department

DELETE FROM *instructor* **WHERE** *dept_name=* 'Finance';

 Delete all tuples in the *instructor* relation for those instructors associated with a department located in the Watson building

DELETE FROM instructor

WHERE dept_name IN (SELECT dept_name

FROM department

WHERE *building* = 'Watson');

where clause may contain everything also usable for select



Delete all instructors whose salary is less than the average salary of instructors

DELETE FROM instructor

WHERE salary < (SELECT AVG (salary) FROM instructor);

- This would delete five tuples
 - But then, the average changes!
- How does the query behave if the deletion is processed one by one?

ID	name	dept_name	salary
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
7/7//	Crick	D' 1	72000
10101	CITCK	Diology	(5000
10101	Jimiivasan	Comp. Sci.	05000
FOEO2	C-1::::-:::	T Ti a La mar	(2000
30303	Canner	Thistory	02000
83821	Brandt	Comp. Sci.	92000
15151	Mozort	Music	40000
10101	Cala	TVIUDIC DI A	10000
33456	Gold	Physics	87000
76543	Singh	Finance	80000



 Delete all instructors whose salary is less than the average salary of instructors

```
WHERE salary < (SELECT AVG (salary) FROM instructor);
```

- Processing this query in SQL
 - First, the **select** query is evaluated
 - i.e., the result is now treated as a constant
 - Then, the delete statement is executed

DELETE vs. TRUNCATE



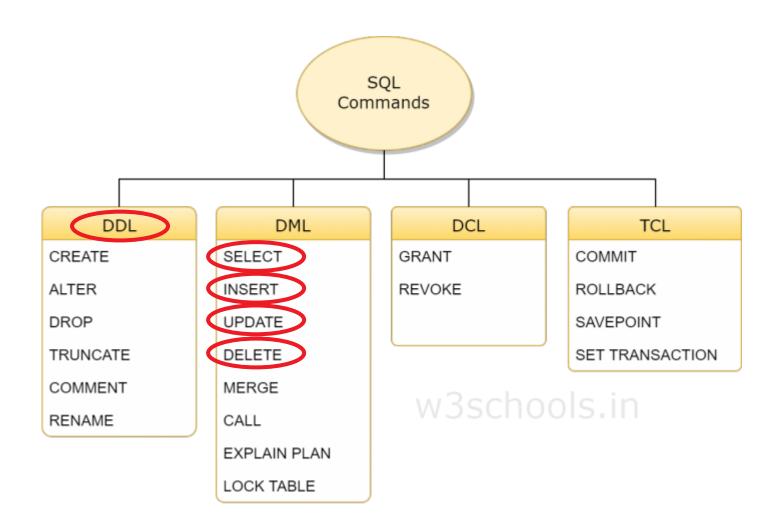
- All records from a table can also be removed using TRUNCATE TABLE instructor;
- Difference to
 DELETE FROM instructor;
- ?
- DELETE keeps the table and deletes only the data
- TRUNCATE drops and re-creates the table
 - much faster
 - but cannot be undone
- DELETE is DML, TRUNCATE is DDL
 - Different rights may be necessary (see later!)

Description

TRUNCATE TABLE empties a table completely. It requires the DROP privilege (before 5.1.16, it required the DELETE privilege.) See GRANT.

Parts of SQL: The Big Picture





Views



- Recap: logical database model
 - The relations in the database and their attributes
- Views:
 - Virtual relations
 - Different from those in the database
 - But with the same data
 - ...hide data from users
- Example: instructors' names and departments without salaries, i.e.,
 SELECT ID, name, dept_name
 FROM instructor

Views



Example: some users may see employees with salaries,

others only without salary

- How about two tables
 - One with salaries
 - One without salaries
- ?



Defining Views



- A view is defined using the CREATE VIEW statement:
 CREATE VIEW v AS < query expression >
 - <query expression> is any legal SQL expression
 - the view name is represented by v
- Once the view has been created
 - it can be addressed as v as any other relations
 - it will always contain the data read by the SQL expression
 - live, not at the time of definition!



Example Views



Instructors without their salary

```
CREATE VIEW faculty AS

SELECT ID, name, dept_name
FROM instructor

Using the view: find all instructors in the Biology department
SELECT name
FROM faculty
WHERE dept_name = 'Biology';
```

Create a view of department salary totals

```
CREATE VIEW departments_total_salary(dept_name, total_salary) AS

SELECT dept_name, SUM(salary)

FROM instructor

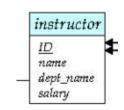
GROUP BY dept_name;
```



 Definition of a simple view (recap: instructors without salaries):

CREATE VIEW faculty AS

SELECT ID, name, dept_name
FROM instructor



- Add a new tuple to faculty view which we defined earlier INSERT INTO faculty VALUES ('30765', 'Green', 'Music');
- This insertion must be represented by the insertion of the tuple

('30765', 'Green', 'Music', NULL) \circ cinto the *instructor* relation

This can only work if salary is not defined as **NOT NULL!**



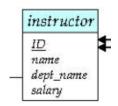
Consider the view

```
CREATE VIEW biology_faculty AS

SELECT ID,name

FROM faculty

WHERE dept_name = 'Biology';
```



and

```
INSERT INTO biology_faculty
   VALUES (43278,'Smith');
```

 Would this lead to INSERT INTO instructor VALUES (43278,'Smith','Biology',NULL);



- Most WHERE constraints cannot be translated into a value to insert
- ConsiderWHERE dept_name = 'Biology' or dept_name = 'Physics'
- or
 WHERE salary > 50000
- Hence, WHERE clauses are typically not translated into a value



Other example used before

```
CREATE VIEW departments_total_salary(dept_name, total_salary) AS

SELECT dept_name, SUM (salary)

FROM instructor

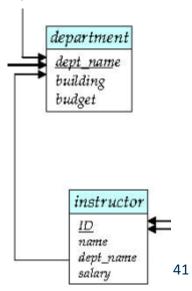
GROUP BY dept_name;
```

What should happen upon

```
UPDATE departments_total_salary
SET total_salary = total_salary * 1.05
WHERE dept_name = "Comp. Sci.";
```



- CREATE VIEW instructor_info AS
 SELECT ID, name, building
 FROM instructor, department
 WHERE instructor.dept_name= department.dept_name;
- INSERT INTO instructor_info VALUES ('69987', 'White', 'Taylor');
 - which department, if multiple departments are in Taylor?
 - what if no department is in Taylor?



Updateable Views



- A view is called updateable if
 - The FROM clause has only one database relation
 - The SELECT clause contains only attribute names of the relation, and does not have any expressions, aggregates, or DISTINCT specification
 - Any attribute not listed in the select clause can be set to null
 - The query does not have a group by or having clause
- Most DMBS only allow updates on such views!

Materialized vs. Non-Materialized Views



- Normal views are not materialized
 - When issuing a **SELECT** against a view, the underlying data is created on the fly
 - Pro: guarantees recent and non-redundant data, saves space
 - Con: some views may be expensive to compute (e.g., extensive use of aggregates)
- Materializing a view: create a physical table containing all the tuples in the result of the query defining the view
 - If relations used in the query are updated, the materialized view result becomes out of date
 - Need to maintain the view, by updating the view whenever the underlying relations are updated

Integrity Constraints



- Data errors may occur due to, e.g.,
 - Accidental wrong entries in form fields
 - Faulty application program code
 - Deliberate attacks
- Integrity constraints
 - guard against damage to the database
 - ensuring that authorized changes to the database do not result in a loss of data consistency
- Examples
 - A checking account must have a balance greater than \$10,000.00
 - A salary of a bank employee must be at least \$20.00 an hour
 - A customer must have a (non-null) phone number





- We have already encountered
 - NOT NULL
 - PRIMARY and FOREIGN KEY
- We will get to know
 - UNIQUE
 - CHECK (P), where P is a predicate

NOT NULL and UNIQUE Constraints



NOT NULL

Declare name and budget to be NOT NULL

name VARCHAR(20) NOT NULL budget NUMERIC(12,2) NOT NULL

- **UNQIUE** $(A_1, A_2, ..., A_m)$
 - The unique specification states that the attributes A_1 , A_2 , ... A_m form a candidate key
 - Candidate keys are permitted to be null (in contrast to primary keys)

The CHECK Constraint



- CHECK (P)
 - where P is a predicate
- Example: ensure that semester is either fall or spring

```
CREATE TABLE section (
    course_id VARCHAR (8),
    sec_id VARCHAR (8),
    semester VARCHAR (6),
    year NUMERIC (4,0),
    building VARCHAR (15),
    room_number VARCHAR (7),
    time slot id VARCHAR (4),
    PRIMARY KEY (course_id, sec_id, semester, year),
    CHECK (semester IN ('Fall', 'Spring'))
);
```

Foreign Keys and Referential Integrity

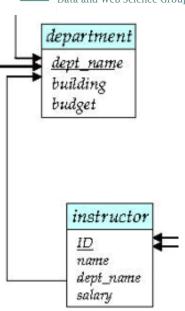


Example:

- instructors have a department
- each department should also appear in the *department* relation

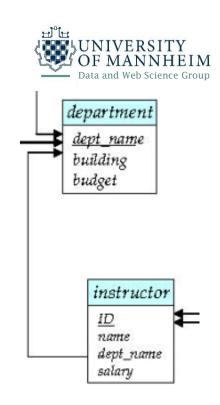
Definition:

- Let A be a set of attributes
- Let R and S be two relations that contain attributes A and where A is the primary key of S
- A is said to be a **FOREIGN KEY** of R if for any values of A appearing in R these values also appear in S



Cascading Actions in Referential Integrity

- Example:
 - instructors have a department
 - each department should also appear in the *department* relation
- How to ensure referential integrity?
 - i.e., what happens if a department is deleted from the *department* relation
- Possible approaches
 - Reject the deletion default action
 - Delete all instructors as well
 - Set the department of those instructors to null



Cascading Actions in Referential Integrity

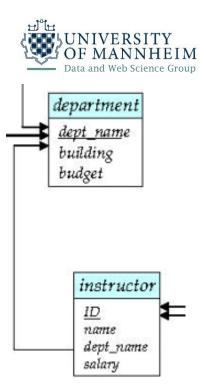
- Cascading updates
 - If a foreign key is changed (e.g., renaming a department)
 - ...then rename in all referring relations
- Cascading deletions
 - If a foreign key is deleted (e.g., deleting a department)
 - ...then delete all rows in referring relations
- CREATE TABLE instructor (

```
dept_name VARCHAR(20),

FOREIGN KEY (dept_name) REFERENCES department

ON DELETE CASCADE

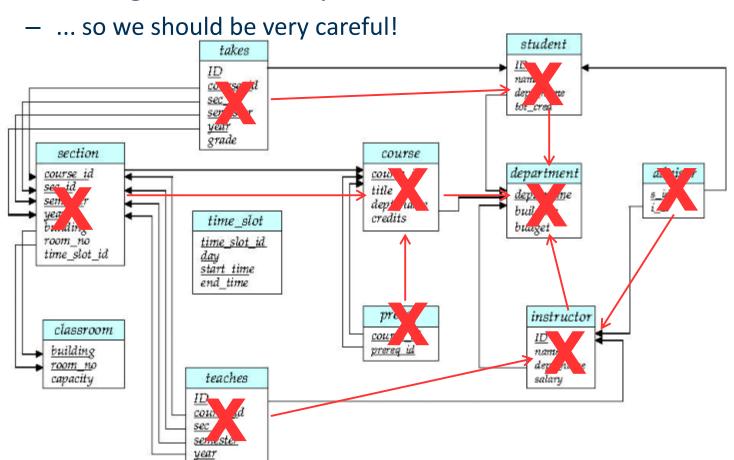
ON UPDATE CASCADE,
....
```







Cascading deletions may run over several tables



Cascading Actions in Referential Integrity



- SET NULL for updates
 - If a foreign key is changed (e.g., renaming a department)
 - ...then set null for all referring relations
- SET NULL for deletions
 - If a foreign key is deleted (e.g., deleting a department)
 - ...then set null in referring relations
- CREATE TABLE instructor (

```
dept_name VARCHAR(20),

FOREIGN KEY (dept_name) REFERENCES department

ON DELETE SET NULL

Different behavior for update and delete is also possible
```

Authorization



- Rights for accessing a database may differ
 - Only administrators may change the schema
- Rights for accessing a database can be very fine grained
 - Not everybody may see a persons' salary
 - Not everybody may alter a persons' salary
 - Nobody may alter their own salary
 - Special restrictions may apply for entering salaries over a certain upper bound

– ...

Authorization



- Forms of authorization on parts of the database:
 - Read allows reading, but not modification of data
 - Insert allows insertion of new data, but not modification of existing data
 - Update allows modification, but not deletion of data
 - Delete allows deletion of data

- Forms of authorization to modify the database schema
 - Index allows creation and deletion of indices
 - Resources allows creation of new relations
 - Alteration allows addition or deletion of attributes in a relation
 - Drop, Truncate allows deletion of relations

Authorization Specification in SQL



- The GRANT statement is used to confer authorization
 - **GRANT** <privilege list>
 - **ON** <relation name or view name> **TO** <user list>
- <user list> is:
 - a user-id
 - PUBLIC, which allows all valid users the privilege granted
 - A role (more on this later)
- Granting a privilege on a view does not imply granting any privileges on the underlying relations
- The grantor of the privilege must already hold the privilege on the specified item (or be the database administrator)

Privilege Definition in SQL



- SELECT: allows read access to relation, or the ability to query using the view
 - Example: grant users Stephen and Mary select authorization on the instructor relation:

grant select on instructor to Stephen, Mary

- INSERT: the ability to insert tuples
- UPDATE: the ability to update using the SQL update statement
- DELETE: the ability to delete tuples.
- ALL PRIVILEGES: used as a short form for all the allowable privileges

Revoking Privileges



- The REVOKE statement is used to revoke authorization.
 - **REVOKE** <privilege list>
 - **ON** <relation name or view name> **FROM** <user list>
- Example:

REVOKE SELECT ON branch FROM Stephen, Mary

- <privilege-list> may be ALL to revoke all privileges the revokee may hold
- If <user list> includes PUBLIC, all users lose the privilege except those granted it explicitly
- If the same privilege was granted twice to the same user by different grantees, the user may retain the privilege after the revocation
- All privileges that depend on the privilege being revoked are also revoked

Revoking Privileges



- Scenario 1:
 - GRANT select ON instructor TO John, Mary
 - REVOKE select ON instructor FROM John
 - → Mary retains right
- Scenario 2:
 - GRANT select ON instructor TO PUBLIC
 - GRANT ALL ON instructor TO John
 - REVOKE ALL ON instructor from PUBLIC
 - → John retains right, since he has been granted the right explicitly

Roles



- Databases may have many users
 - e.g., all students and employees of a university
- Managing privileges for all those individually can be difficult
 - User groups (also called: roles) are more handy
 - Example roles
 - Student
 - Instructor
 - Secretary
 - Dean
 - ...

Roles

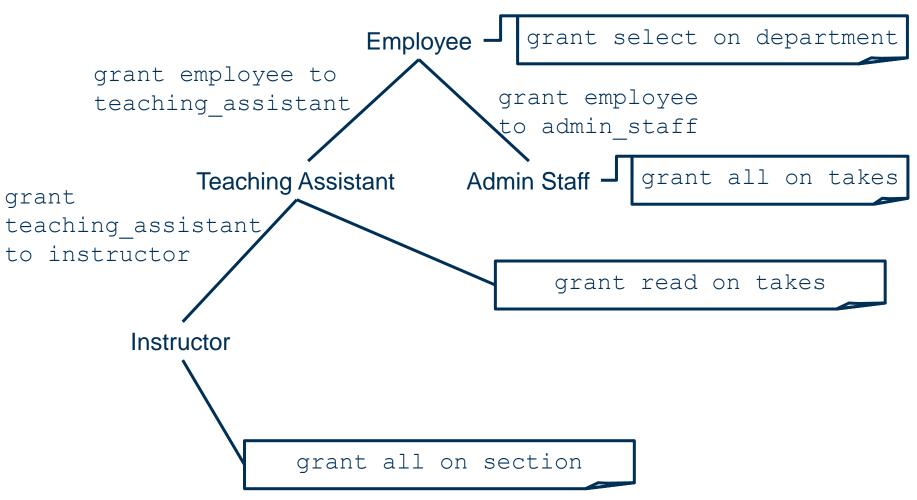


- Creating roles and assigning them to individual users
 - CREATE ROLE instructor;
 - GRANT instructor TO Amit
- Granting privileges to roles
 - GRANT select ON takes TO instructor

- Roles can form hierarchies
 - i.e., a role inherits from other roles
 - CREATE ROLE teaching_assistant
 - GRANT teaching_assistant TO instructor
 - Instructor inherits all privileges of teaching_assistant

Roles: Example





Roles on Views



 Example: Geology department members can administrate their own staff, but not others

```
CREATE VIEW geo_instructor AS (
    SELECT *
    FROM instructor
    WHERE dept_name = 'Geology');
GRANT select ON geo_instructor TO geo_staff
```

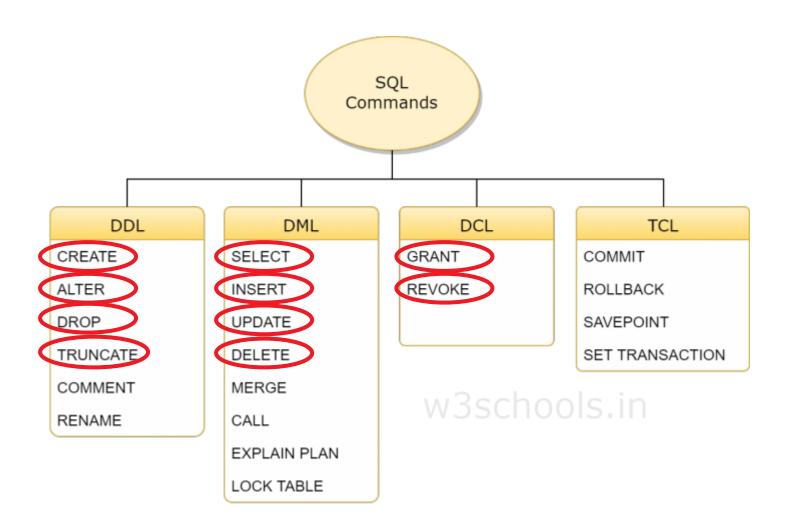
Suppose that a geo_staff member issues

```
SELECT * FROM geo_instructor;
```

- What if
 - geo_staff does not have permissions on instructor?
 - creator of view did not have some permissions on instructor?

Parts of SQL: The Big Picture





Wrap-up



- Today, we have seen
 - How to manipulate data in databases
 - i.e., INSERT, UPDATE, andDELETE statements

Views

- are used to provide different subsets and/or aggregations of data
- updateable views
- materialized views



Wrap-up



- Integrity constraints
 - unique and not null constraints
 - cascading updates and deletions
- Access rights
 - can be fine grained
 - can be bound to user groups and roles
 - roles may inherit from each other



Questions?



