

Indexing & Views

Monday, March 6, 2017

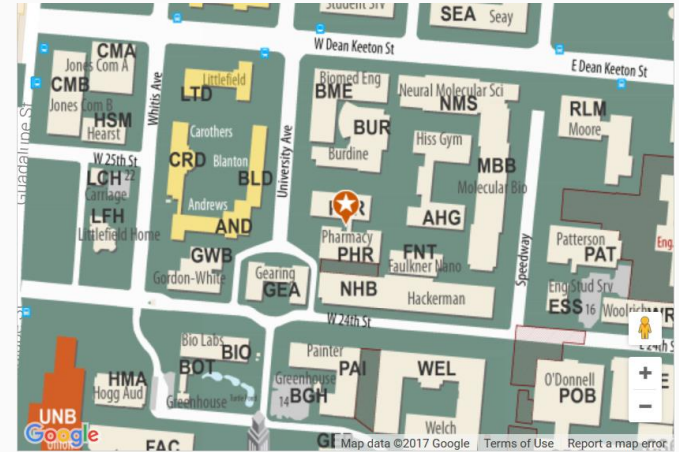


Agenda

- Announcements
- Reading Quiz
- Indexing
- Views
- Midterm details

Announcements

- Next class: **Midterm**
- Midterm location: **PHR 2.108**
- Review session: Wed 12-1pm @ GDC 2.210
- TAs' office hours this week:
 - Alan: Tues 10am-12pm and 5:30-7pm @ GDC 3rd Floor Lab
 - Sean: Wed 5-6pm @ GDC Basement TA Stations
- After Spring Break: **Final Project**



Q1: How do you create an index?

- a) CREATE INDEX table_name (column_name)
- b) CREATE INDEX index_name ON table_name (column_name)
- c) CREATE INDEX ON table_name (column_name)

Q2: What are the tradeoffs of indexing?

- a) Take up spaces in the database
- b) Slow down data modification operations
- c) Speed up access to data
- d) All of the above

Q3: What is not a motivation for views?

- a) Aggregation: to appear as though data is aggregated
- b) Complexity: making multiple tables appear to be a simple table
- c) Security: to avoid having to reveal individual data rows
- d) Space saving: to reduce the storage of database tables

Q4: Creating a view is giving a name to a _____ statement.

- a) CREATE TABLE
- b) ALTER TABLE
- c) SELECT
- d) UPDATE
- e) DROP TABLE

Q5: A data warehouse is a large repository of data that is designed for reporting, analysis, and planning.

a) True

b) False

Indexes: Key Concepts

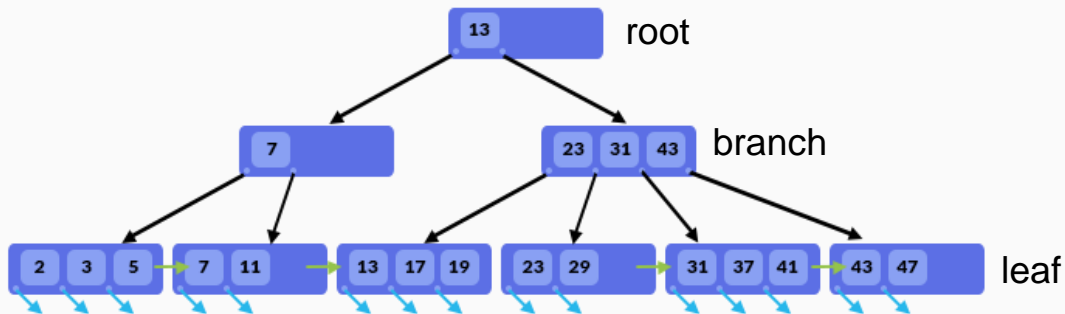
- **Critical** to database systems
- At least one index per table
- They work “behind the scenes”
- DBA looks at the workload and decides which indexes to create (no easy answers)
- Creating indexes can be an expensive operation
- Query optimizer decides which indexes to use during query execution

```
SELECT *  
FROM Employee  
WHERE empid = 37;  
  
SELECT *  
FROM Employee  
WHERE empid > 37;
```

Employee table

empid	firstname	lastname	depid
2	Michael	Dell	5
23	Betty	Jennings	
3	Bill	Gates	5
5	Fran	Bilas	8
...
37	Gordon	Moore	6

B+ tree on Employee.empid

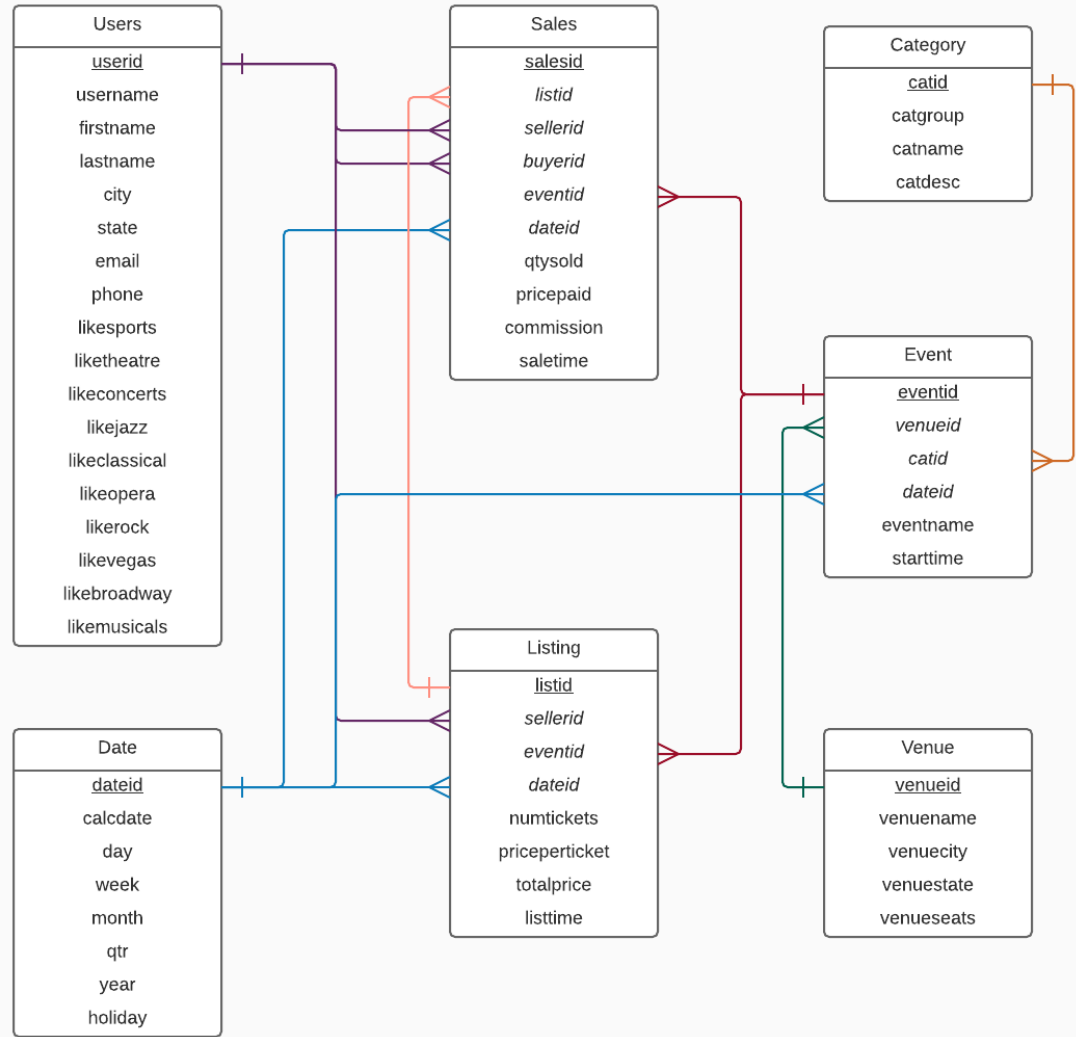


Practice Problem 1:
Consider a composite index on the fields (eventid, qtysold) of the Sales table.

Give 3 example queries that can be answered more efficiently using this index.

Notes:

- eventid range from 1 to 9798
- qtysold range from 1 to 8
- Assume a uniform distribution

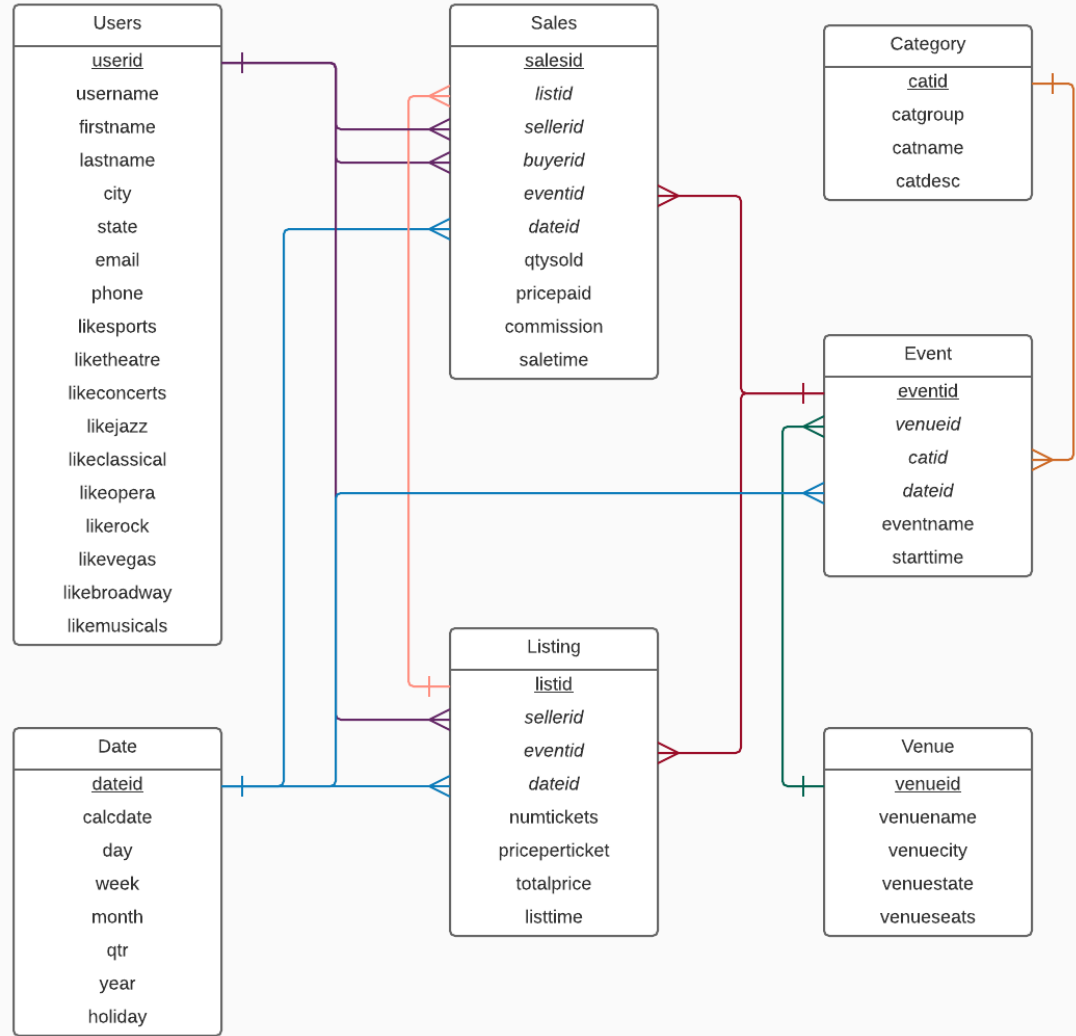


Practice Problem 1:
Consider a composite index on the fields (eventid, qtytsold) of the Sales table.

Does this index speed up the following query?

```
SELECT *  
FROM Sales  
WHERE qtytsold = 1;
```

- a) Yes
- b) No

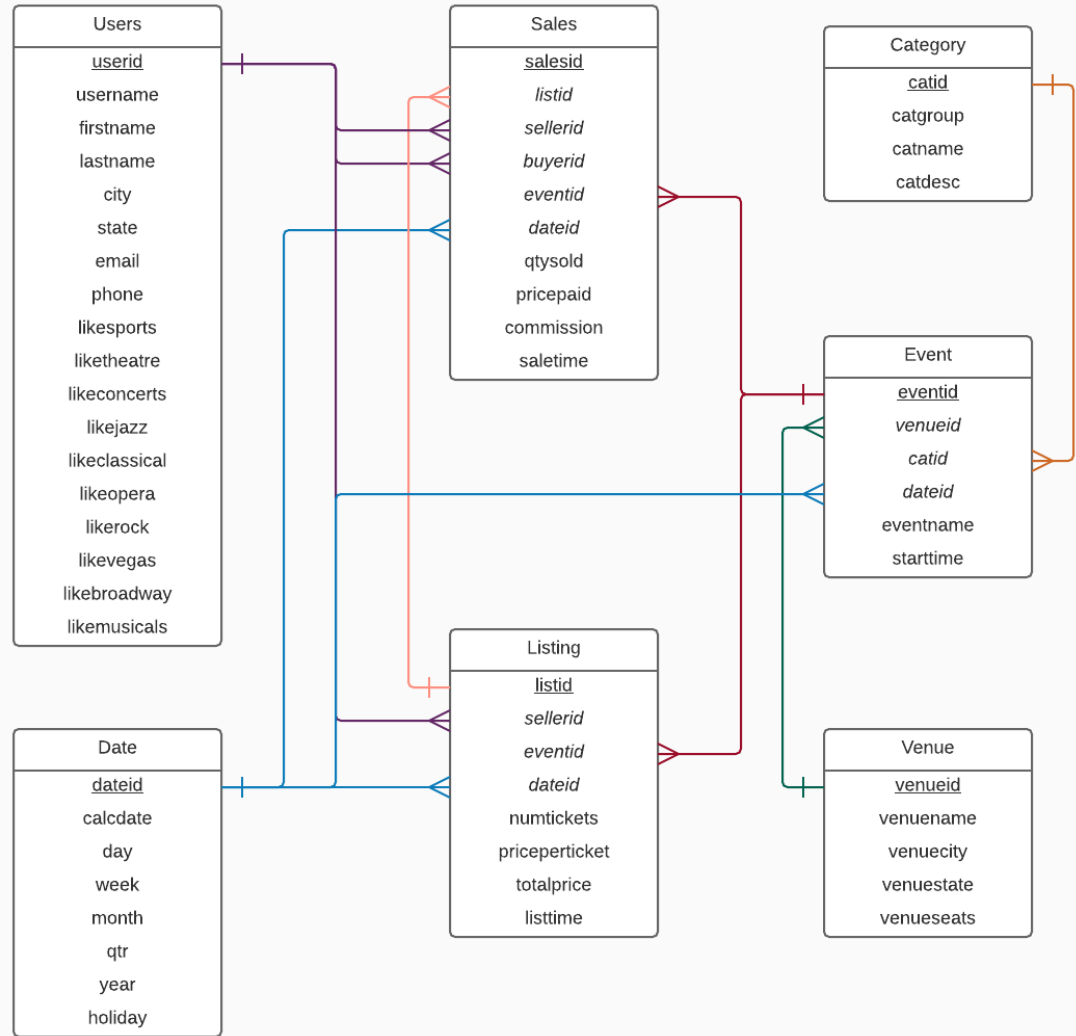


Practice Problem 2:

Consider a composite index on the fields (eventid, qtytsold) of the Sales table.

Give an example of an update that is definitely slowed down by this index.

Give an example of an update that is not slowed down by this index.

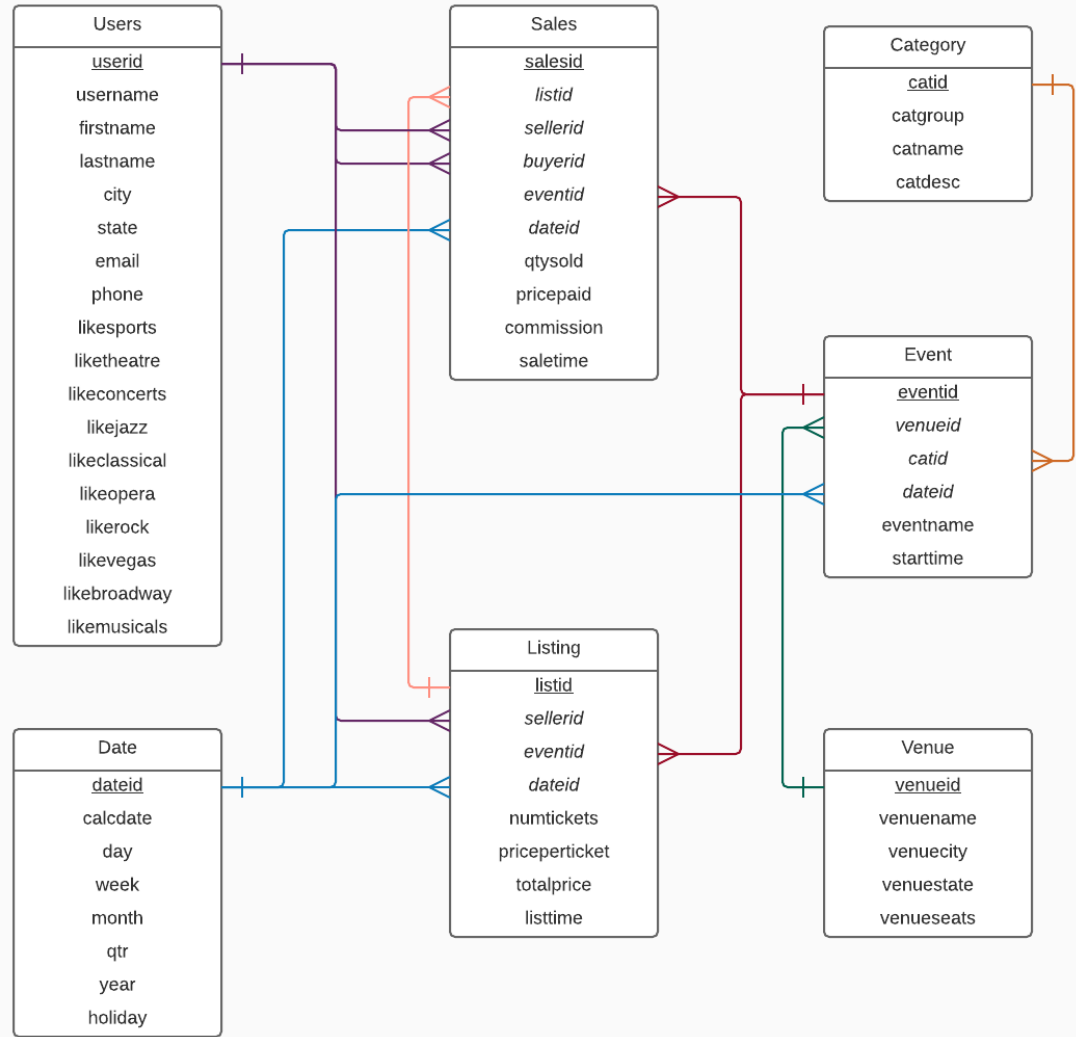


Practice Problem 2:
Consider a composite index on the fields (eventid, qtytsold) of the Sales table.

DELETE FROM Sales
WHERE salesid = 100;
is not slowed down by
this index.

a) True

b) False



Views: Key Concepts

- Defined by a SQL query
- Return a table of results from query
- Virtual views vs materialized views
- VV are used to hide sensitive columns and records (e.g. ssn, salary, etc.)
- MV are used to speed up long-running queries in a data warehouse

<u>empid</u>	firstname	lastname	depid
1	Michael	Dell	5
2	Betty	Jennings	
3	Bill	Gates	5
4	Fran	Bilas	8

Employee

<u>depid</u>	name
5	Executive
6	Operations
7	Sales
8	Product

Department

```
CREATE VIEW ExecutiveStaff AS
SELECT e.empid, e.firstname,
e.lastname, d.depid, d.name AS depname
FROM Employee e JOIN Department d
ON e.depid = d.depid
WHERE d.name = 'Executive';
```

```
Virtual table = ExecutiveStaff(empid,
firstname, lastname, depid, depname);
```

```
SELECT * FROM ExecutiveStaff;
```

Practice Problem 3:

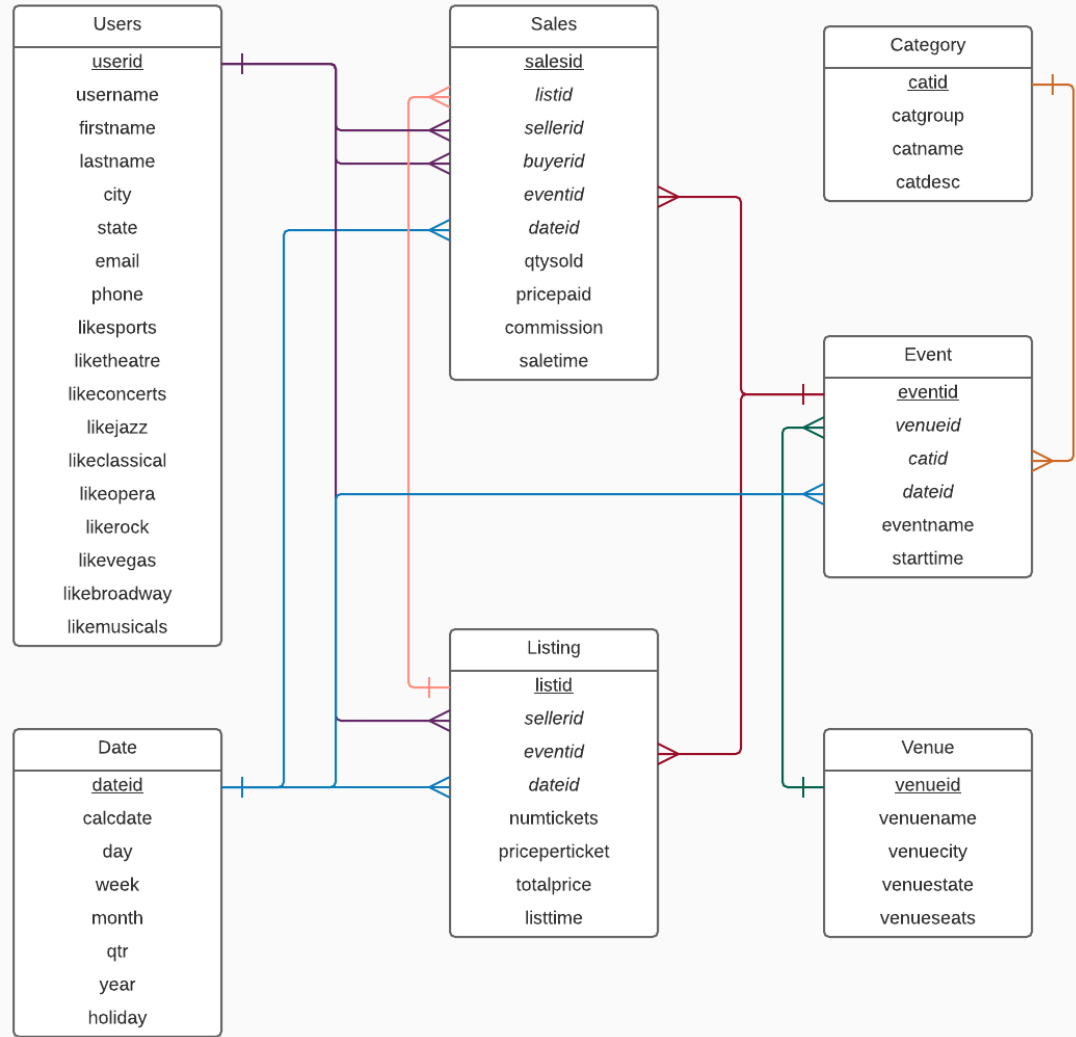
Consider the following workload:

- 1) find all concerts that are happening over next 30 days.
- 2) find all concerts that are happening near me.
- 3) find all concerts that are not sold out.

Construct a virtual view that can simplify the complex queries in this workload.

Notes:

- use catgroup = 'Concerts'



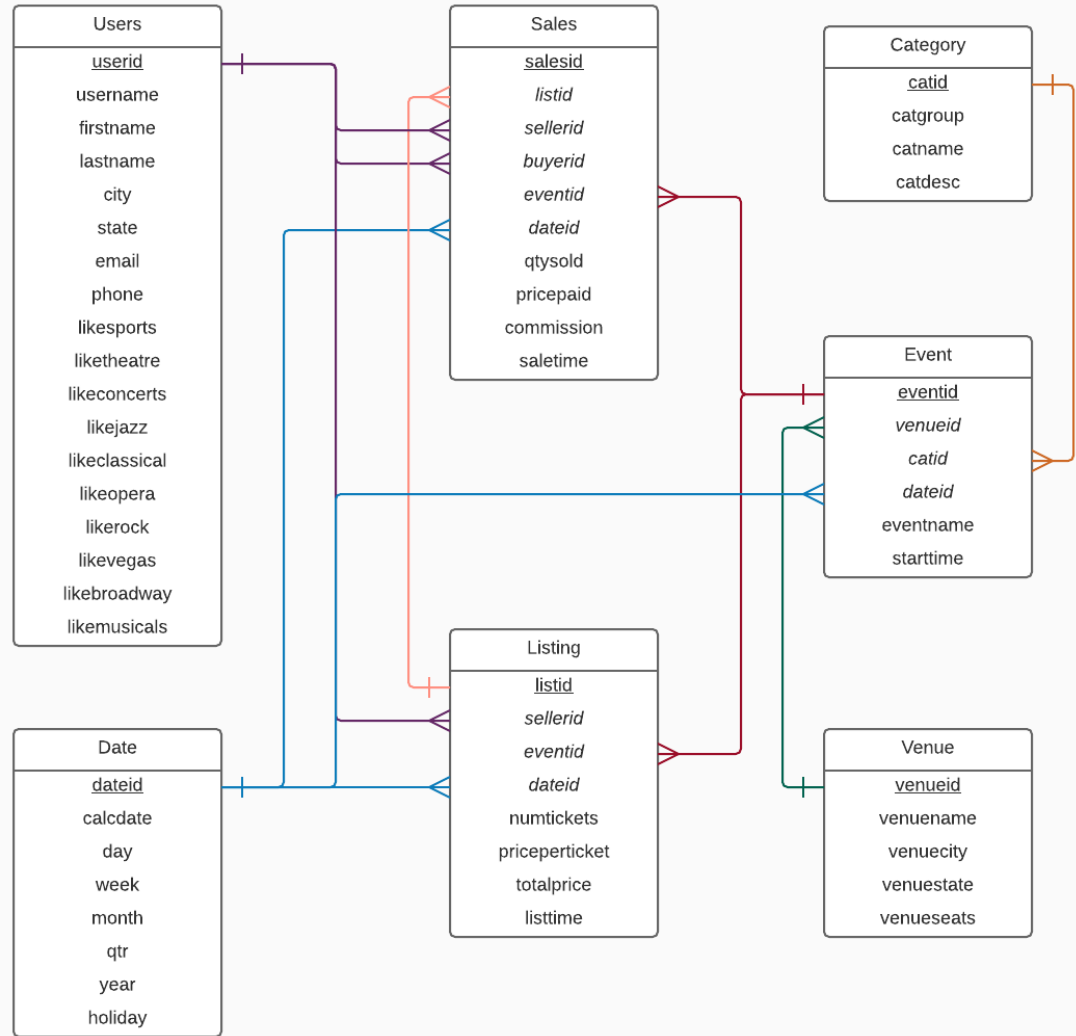
Practice Problem 3:

Consider the following workload:

- 1) find all concerts that are happening over next 30 days.
- 2) find all concerts that are happening near me.
- 3) find all concerts that are not sold out.

Which tables are in your view?

- a) {Event}
- b) {Event, Category}
- c) {Event, Date, Category}
- d) {Event, Sales, Users}
- e) All of the above



Midterm Logistics

Date: Wednesday, March 8, 2017

Time: 6:30pm - 8:00pm

Where: **PHR 2.108**

Closed book exam

No laptop, no phone, no cheat sheet

Midterm Format

2 Types of Questions:

1. Multiple choice & True/False
2. Free form and fill-in-the-blank

Data Modeling Topics

- Entities, attributes
- Relationships
- Conversion to relations (e.g. junction tables)
- SQL DDL:
 - creating tables
 - datatypes
 - constraints
- Data anomalies
- Normal forms (1NF – 3NF)

SQL Topics

- create table
- insert, update, delete
- select-from-where
- order by
- inner and outer joins
- table aliasing
- column renaming
- aggregate functions
- group by
- having
- limit
- create index
- create view
- answering queries using views

Study Materials

- Assigned Chapters
- Class Notes
- Reading Quizzes:
 - see class notes for answers to reading quizzes
- Practice Problems:
 - see snippets repo for TICKIT database
 - see snippets repo for solutions to practice problems