ASSIGNMENT 1 - 601.315/415/615 - Databases

Due date: Tuesday, October 10, 2017 in class

Part 1: Database Schema Design

Your task is to design a database to support the needs of a new internet auction startup company called *JBay*. Its business model is quite close to EBay's, and thus you can assume that most of the same goals and functions that EBay's website supports will also be necessary for JBay. If you are not familiar with EBay or Amazon auctions you should briefly explore their websites.

But the companies are different, and JBay has several specific requirements and general goals that are outlined for you below. You are also expected to create a vision of new database functions and capabilities that JBay may wish to support, *beyond* those current found in EBay/AmazonAuctions or specified below. Part of your grade will be based on your creativity and design principles.

At a minimum your database design should cleanly support several types of auctions, including single-item auctions as well as cases where there are numerous equivalent copies of the same item (e.g. a BlueRay), and these will either be sold in multiple auctions or a single dutch auctions. There is also the need to support auction relists when an item doesn't sell, in all cases avoiding unnecessary repetition of information and expensive-to-store images.

Also, at a minimum, your database needs to support different types of JBay users (e.g. types of buyers, sellers, etc.) and support their various common types of interactions (e.g. feedback, communications, payment arrangements, etc.). You should also support some appropriate payment mechanism, such as a PayPal, that is closely linked with the auction process for efficiency.

Please design a database using the E-R and relational table representations that is capable of supporting these above goals and providing answers to the following questions. You *should* include other attributes and relationships that are not explicitly required by the questions below if you think that they are appropriate for a good and complete database design. Completeness, creativity and good design principles are all very important to the evaluation of this assignment.

These following queries are only samples of the minimal functionality that needs to be supported:

(a) List the usernames of all individuals with a feedback score less than -20 who have joined *JBay* since 7-1-2017.

(b) List the seller's location for all auctions won by a bidder located in Towson, Maryland.

(c) List all items that *ImTooTimid* was "watching" but didn't bid on and remained unsold (received no bids in a finished auction).

(d) List the item names, item types and their closing prices and winners for all auctions that *jjcool* has bid on since 12-31-2016.

(e) List all negative or neutral feedback listings that *IlikePonies* has received from a seller who has any negative feedback, including both the type of the feedback, the date/auction/price in question and the feedback text.

(f) List the usernames of all bidders who have bid on every auction by *IlikePonies.*

(g) List the usernames of all bidders who have won every auction by *IlikePonies* between 8-31-2016 and 8-31-2017.

(h) List all the auctions for *SuzieLoozie* that are scheduled to close on 10-30-2017 without any bids on them.

(i) List all the usernames and cities/countries of origin of the users who have bid on *all* the auctions that *MileyC* has bid on previously (i.e. she had a lower bid on the same auction).

(j) List all the recipients and content of negative feedback left since 12-31-2017 by a bidder with a feedback score less than *-50* (now).

(k) List all completed auctions bid on by *ICantStop* that had a closing price that was *either* less than half the highest (hidden) proxy bid price he was willing to spend on the auction, or more than twice his highest proxy bid price for the auction.

(l) List all previously held userid's of *PerlJam26*, including the dates that the userid was started and terminated and the home cities/countries listed for each previous userid.

(m) (In order to identify possible cases of bid "shielding"), list all possible pairs of userid's of bidders who have retracted a bid in an auction within 10 minutes of auction closing, and all bidders on the same auctions who have placed a bid lower than the retracted bid.

(n) List the userID's of all users with a tier ranking of SuperShiningBlazingMeteors or above (based on their feedback score as described in the Assumptions below).

(o) For every tier ranking **above** a *jjcool*'s current tier ranking, list both the name of the ranking, the feedback score needed to achieve it, and the number of additional feedback points that *jjcool* currently needs to achieve this ranking.

(p) List the current bidders and their feedback ratings for *toyz2you*'s most recent dutch auction of *Attak Trak*'s.

(q) List the userid's of all users who have only bid on auctions of type *clothing*, subtype *women's* or *children's.*

(r) List the userid's of all users who have never bid on auctions of type *Collectibles* subtype *BeanieBabies* or any subtype of *VideoGame.*

(s) (The top bidder on auction number 88114123 has refused to pay.) List the userid, highest bid price and feedback rating of the next highest bidder.

(t) (In order to help identify collaborators of a suspected shill bidder) List the usernames of any bidders who have placed more than 80% of their total bids on an auction hosted either by *TooHotToHandle* or *WayTooHot.*

(u) List the userIDs of users who have bid within 1 minute of auction opening for at least 100 times.

(v) List the items that have the most canceled top bidders (top bidders who refused to pay).

(w) List the userIDs of users who have bought every item they bid on, but have not bid on items of which they are the only bidder.

(x) List items selling for a loss, which is defined as the final sale price being lower than the (non-public) cost for the item.

Assumptions:

Auctions must have exactly one seller, but may have 0 or more bidders at any given time. Bids are stored in sequential order, and later bids must be higher than previous bids.

"Standard" auctions have exactly one item for sale. "Dutch" auctions allow multiple identical copies of the same item to be sold. Auctions can be relisted if not sold. Multiple standard auctions all of the same item from the same seller (but at different times) should be linked to avoid needless repetition of photos and descriptions.

Bidders may bid multiple times on the same auction (e.g. when they have been outbid at a lower price). Keep a record of the date, time and amount of each of these previous bids.

Auctions have one of a set list of item types and subtypes (such as Collectibles – MoshiMonsters). Optionally, you may allow an entire hierarchy/tree of item types and subtypes with a minimum depth of 2.

Biders may use proxy bidding, in which they enter a maximum price they are willing to bid for an auction along with their bid. This information is hidden and distinct from the current bid price and available only to the bidder.

Bidders may watch auctions that they do not bid on. You should allow users to provide a maximum bid price they would be willing to spend, and then notify them if the current bid price is below this maximum 1 hour before auction closing.

You should assume that a single userid can be used both to sell items and bid on items. Users should receive a single stream of feedback, but each feedback message should distinguish whether the user was a bidder or seller on the auction.

Assume that users are assigned to tier rankings (e.g. Bronze, Gold, ShiningStar, SuperShiningBlazingMeteors) based on their feedback scores. Each ranking has a cutoff point (e.g. 100+ for silver, 200+ for gold) that should be readily changeable. Also, each tier ranking comes with certain benefits, such as a discount on seller's commissions and frequent-bidder perks.

In general, guidance and explanations of typical auction concepts can be found from the help sections of `www.ebay.com`.

Your auction design should at a minimum support *all* the queries above and all the assumptions above, but you are encouraged to add other useful and/or original attributes, entities and relationships to this space to support other functionalities.

To Do:

1.1 (*22 points*) Design the database using the entity-relationship database model and draw it. Your design should minimize repetitions of information. Be sure to mark the mapping constraints ($\leftrightarrow$, $\rightarrow$, $\leftarrow$), participation constraints ($\rightarrow$ or $\Rightarrow$) and underline primary keys.

You should *very* briefly justify any unusual or potentially controversial design decisions you make. Do *not* spend much time on such notes.

This section will be graded on aesthetics and completeness as well as correctness.

1.2 (*10 points*) Represent this database design using the relational model. You should use a tabular notation and include at least one row of sample values for each relation.

You should *very* briefly justify any unusual or potentially controversial decisions you make in the conversion process. Do *not* spend much time on such notes.

1.3 (*18 points*) Write relational algebra expressions to answer some of the English queries (a) through (t) given on page 1-2 of this assignment.

- Students in 601.315 must answer **all** 4 of the queries (b), (c), (d) and (j).
- Students in 601.415/615 must answer **all** 7 of the queries (d), (f), (i), (j), (k), (r), (s).

You may use assignment to intermediate relations to make your expressions clearer and simpler.

Part 2: Relational Algebra and Relational Calculus (50 points)

Consider the following relational database schema (with sample instantiations):

TV_SHOW:

ShowID	*ShowTitle*	*StudioName*
S1131	Game of Thrones	HBO
S2287	Saturday Night Live	NBC
S8714	Star Trek - Original Series	Desilu

EPISODE:

EpisodeID	*ShowID*	*EpisodeTitle*	*CompletionDate*
E4411	S1131	Dragonstone	4-16-2017
E4412	S1131	Beyond the Wall	5-20-2017
E8472	S2287	SNL hosted by Melissa McCarthy	5-13-2017

ARTIST:

ArtistID	*ArtistName*	*Sex*	*BirthLocationID*
A8720	Peter Dinklage	M	L2276
A7329	Emilia Clarke	F	L6453
A2912	Jennifer Aniston	F	L3911

WORKED_ON:

ArtistID	*EpisodeID*	*Job*	*Role*
A8720	E4411	Lead Actor	Tyrion Lannister
A7329	E4411	Lead Actor	Daenerys Targaryen
A8745	E5529	Director	Himself

FILMED_IN:

EpisodeID	*LocationID*	*DateFilmed*
E4411	L3918	1-14-2017
E8472	L2276	2-24-2017

BROADCAST_IN:

EpisodeID	*LocationID*	*Date*	*Station*	*Rating*
E8720	L2276	7-16-2017	KLWX	38.0
E4411	L2276	7-04-2017	KGAN	12.5

LOCATION:

LocationID	*City*	*State*	*Country*
L2276	New York	New York	USA
L3918	Sevilla	Andalucia	Spain

EXPENDITURE:

ExpenseID	*ExpenseType*	*BroadCategory*
EX381	Actors	Labor
EX382	Musicians	Labor
EX384	Stunt Staff	Labor
EX386	Costume Staff	Labor
EX162	Computers	IT

SPENT_ON:

EpisodeID	ExpenseID	*Amount*	*Date*
E4411	EX381	200000	2-01-2017
E2042	EX162	40000	1-04-2016

REVIEW_SITE:

SiteID	*SiteName*
R1234	IMDb
R2345	Rotten Tomatoes
R9841	IGN

EPISODE_REVIEW:

SiteID	*EpisodeID*	*Score*
R1234	E4412	9.3
R2345	E4411	9.6
R9841	E4412	6.9

QUESTIONS:

601.415/615 students should give *relational algebra* expressions in 14 of the the queries 1-15, **plus** 2.16, 2.18, 2.19 and 2.20, also in the relational algebra.

601.315 students should give *relational algebra* expressions in 11 of the the queries 1-15, **plus** 2.16, 2.17 and 2.20, also in the relational algebra.

Give expressions in the tuple relational calculus for queries 2.1, 2.2, 2.3, 2.6, 2.8, 2.9, 2.10, 2.12, `(Students in 601.315 should do any 5 of these 8.)` All 601.315/415/615 students should also answer 2.18 in the relatinal algebra.

2.1 List the name and birth city+state of all artists who have worked on a TVShow episode that was filmed in the city and state of their birth.

2.2 List the name and AID of all artists who have never worked on a TVShow episode that was filmed in the state of their birth **and** have never worked on an episode that was broadcast in the state of their birth.

2.3 List the name of all artists who have worked on every TVShow produced by Sundance studios.

2.4 List the name of all TV shows that have been broadcast in every city in California (doesn't have to be the same episode in all cities).

2.5 List the names and AID's of all artists who have been the lead actor in an episode that they have also directed.

2.6 List the names and AID's of all artists who have worked on an episode with someone who worked on an episode with 'Clint Eastwood' as director.

2.7 List the names and AID's of all artists who have worked on an TVShow with someone who worked on TVShow that 'Seth Meyers' also worked on. (It is possible for two people to work on the same TV show but not ever work on the same episode)

2.8 List the name of all TV shows that did *not* earn over a 35.0 rating on every broadcast of the show in Baltimore in 2016.

2.9 List the names of all studios that have *never* produced an episode of a TVshow that has earned more than a 20.0 rating *and* has never produced an episode that Jennifer Lopez has worked on.

2.10 List the name of the producing studio (and the episode name) of episodes that have never been broadcast in any location where they were filmed.

2.11 List all pairs of artists who have worked together on all episodes produced by HBO after 01-01-2017.

2.12 List the TV show name, episode title and episode ID of all episodes in which Martin Sheen was both the director and producer in the same episode.

2.13 List the TV show name and show ID of all TV shows in which the same person has been simultaneously both the director and producer of some episode (i.e. for at least one episode of the show the same person both was director and producer of the same episode).

2.14 List the names and show IDs of all TV shows broadcast on 10-04-2017 in Baltimore which have a higher rating than the 10-04-2017 broadcast of “Game of Throne” in Baltimore (you can assume that there was only one broadcast of “Game of Thrones” in Baltimore on that date).

2.15 List the show name, episode ID and episode title of the highest rated broadcast ever by any TV station in Baltimore. (You cannot use an aggregation operator such as max for this query. Simple relational algebra operators are sufficient).

2.16 List the most expensive “Labor” category (e.g. Stunt Staff) for any episode of Game of Thrones, along with the title of the episode and total amount spent on that category for that episode.

2.17 List the title and date of all episodes that spent more on Musicians than Actors.

2.18 List the title and date of all episodes that spent more on Musicians than Actors, along with the title of the TV show containing that episode and the name of the producer of that TV show.

2.19 List the total amount spent on all types of “Labor” (i.e. a single sum) for each episode of Game of Thrones, along with the title and episode ID of that episode.

2.20 List the name of the review sites that have rated every episode of Game of Thrones above 9.0 (¿= 9.0)