Tarefa avaliada por colega: Designing a Data Model for 'Catch the Pink Flamingo'

Você foi aprovado!
Congratulations. You earned 60 / 60 points. Review the feedback below and continue the course when you are ready. You can also help more classmates by reviewing their submissions.

Avaliar o trabalho de um colega

ⓘ Parece que esta é sua primeira tarefa com avaliação entre colegas. Saiba mais

Instruções

My submission

Discussões

Pink Flamingo Data Base

Enviado em 14 de Dezembro de 2016

Shareable Link

We will need a table to keep track of **user clicks**. Below is a rough structure of the table, but it is incomplete in the following ways:

1. There are no primary keys,
2. More columns may need to be added,
3. The data types are not all specified.

Copy the table below, paste it into your response, and complete the missing parts of the schema according to the list above. You may provide a brief explanation as needed.

userID: long	sessionID: _____	timestamp: dateTime	clickedPoint: coordinate	missionID: _____	isHit: _____	_____: _____
100	4356	10/12/2015::14:15:09	(4,8)	13	yes	_____
101	3241	10/23/2015::14:15:19	(20,5)	18	no	_____
102	4537	11/4/2015::14:15:20	(17,43)	21	no	_____

Data in CSV format separated by "|":
userID: long|**sessionID**: long|**timestamp**: dateTime|clickedPoint: coordinate|missionID: long|isHit: boolean|missionId: long|speed: long
100|4356|10/12/2015::14:15:09|(4,8)|13|yes|0|0.55
101|3241|10/23/2015::14:15:19|(20,5)|18|no|1|-1
102|4537|11/4/2015::14:15:20|(17,43)|21|no|0|-1
The primary key is the union of the userID, sessionID and timestamp, because should not exist two userID in the same sessionID at the same time.

I also added the missionId column to identificate the mission and speed column as the time spent by the user (in seconds) to execute the capture, set to -1 if isHit set to no.

Central de Ajuda

Did the student correctly identify an appropriate primary key for the userClicks table schema?

- 10 pts
  Yes
- 0 pts
  No

Did the student provide reasonable data types for the missing values?

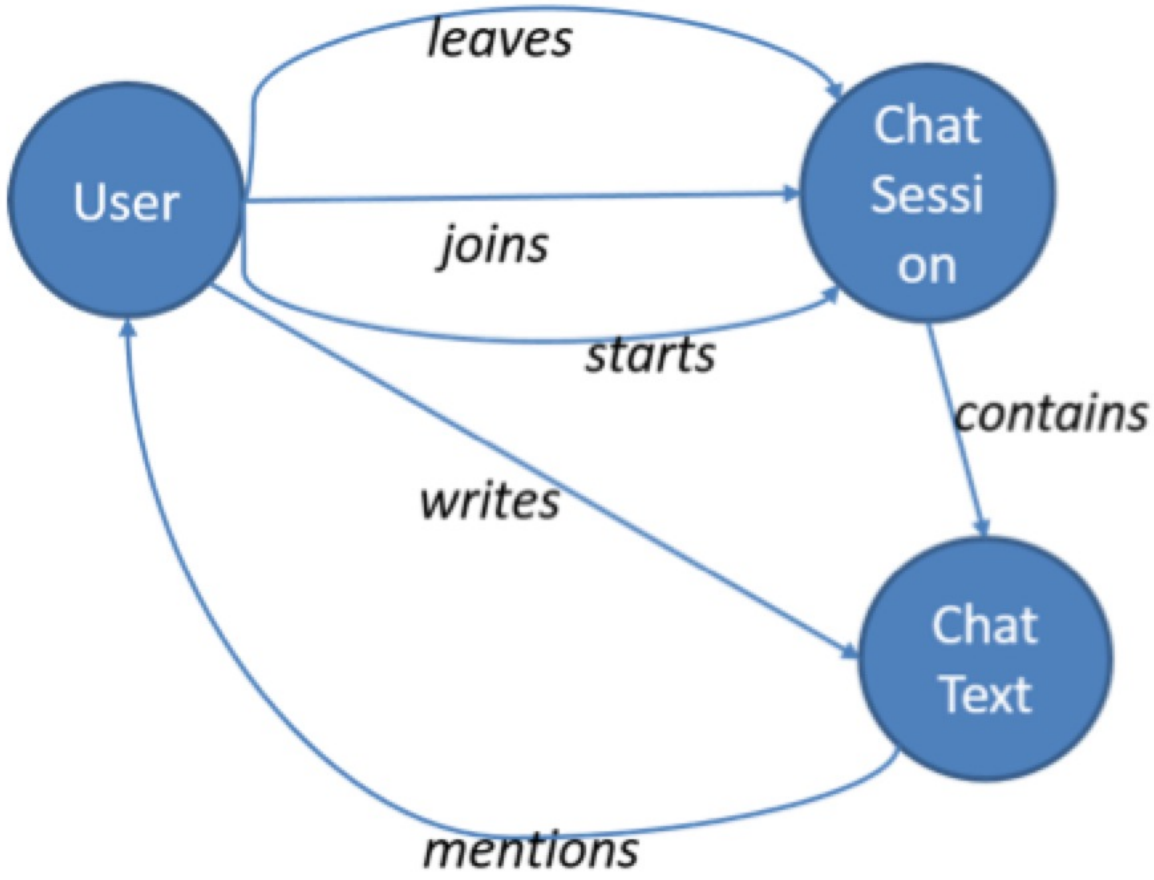
- 10 pts
  Yes
- 0 pts
  No

Did the student add a field to the table which made logical sense?

- 10 pts
  Yes
- 0 pts
  No

Next, we will look at the **chat data**. We know users can chat when they discuss their team's strategy or when they decide if a new player should join a team. This can be modeled as a graph, as shown below.

Central de Ajuda



The nodes of the graph are entities and the edges are actions they take and the content of the chat. When this graph is populated, we need to include a number of properties to the nodes and edges (recall the lecture covering the property graph model). For example, each action edge should have a timestamp attribute to indicate when the action occurred. Using this graph, we should be able to analyze different types of questions.

List all node and edge properties for a graph.

Briefly explain how you would use the graph to answer the following five questions?

1. Which teams are having more conversations?
2. Do users chat more (or less) before they leave a team?
3. What are the dominant terms (words) used in a chat session within a specific time period?
4. Which users are most active in a specific chat session?
5. How many chat sessions is a user participating in at the same time?

Vertex:
user, team, team_chat_session, chat_text

Edges (Action : fields):
Create chat: user, team, team_chat_session, timestamp
action - Join to chat: user, team, team_chat_session, timestamp
action - Leave chat: user, team_chat_session, timestamp
action - Create chat text: user, team_chat_session, chat_text, timestamp
action - Mention a text: user, chat_text_mentioned, timestamp
action - Respond a text: user, chat_text_mentioned, chat_text, timestamp

1 - For the teams with more conversation, just look for the file of edge "Create chat text" for the teams with more numbers

of the entries.

2 - Look in the edge "Create chat text" all the content of the user in the team_chat_session before the timestamp contained in the edge "Leave chat".

3 - You can open the edge "Leave chat", select only the entries of a time period and execute a map-reduce word counting.

4 - Just open the edge "Create chat text" and look for the users that have posted in that chat session and count the number of posts.

5 - Enter in the edge "Create chat text" and count the distinct values of "team_chat_session" for this user in a certain time window.

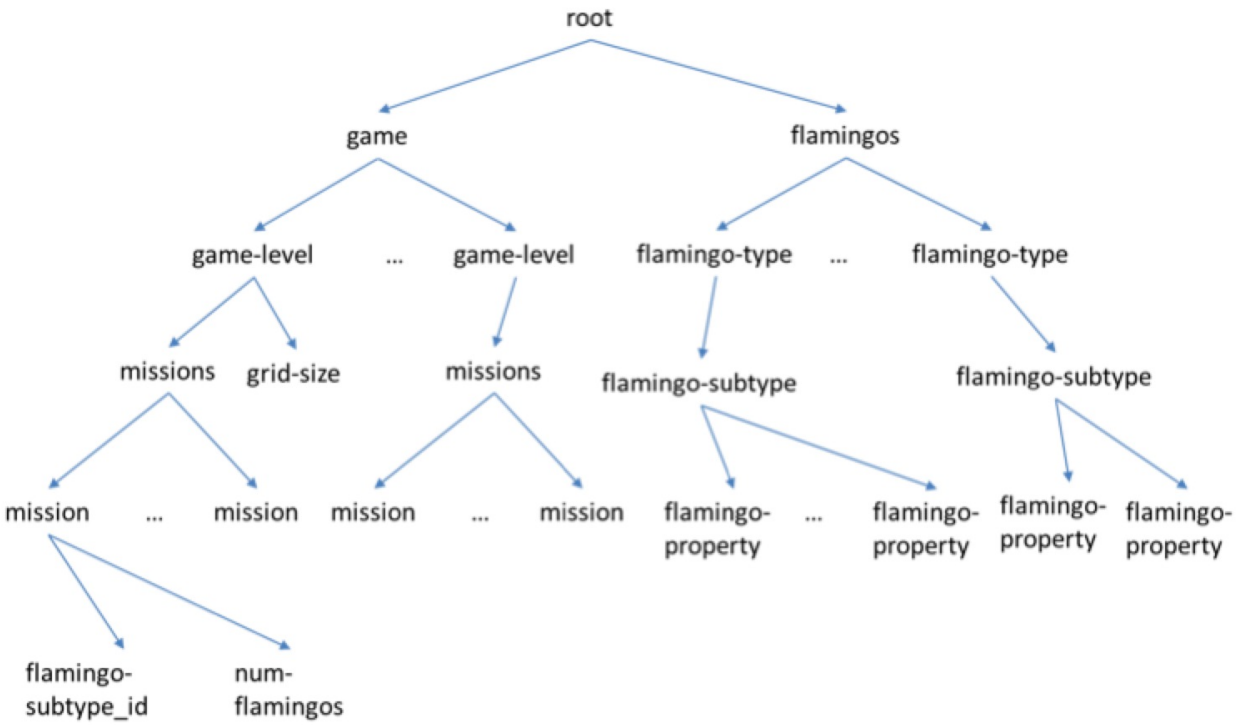
Did the student correctly list all of the node and edge properties?

- 10 pts
  Yes
- 0 pts
  No

Did the student provide a reasonable explanation of how they would use graphs to answer the five questions?

- 10 pts
  Yes
- 0 pts
  No

We will need to design the structure of the game itself. We will use the following partially specified semistructured data object.



Central de Ajuda

We have seen that both XML and JSON are essentially trees. The tree here says that the “missions” category can have several “mission” items. Each mission item will have an associated flamingo type (e.g., flamingos with stars on their backs) and the number of flamingos of that type released to the users to catch. We also have flamingos of different types and subtypes, where each subtype will have a list of properties.

Explain how you might extend the tree with at least five specific flamingo properties. For example, beak-color whose values might be “bright-red” or “pink”.

Your objective is to make sure the tree has enough properties to launch the game and connect with the three data structures we have used so far.

At the first it's possible to create theses flamingo's properties: size {values: tinny, medium, large), body-color (values: pink, red, white), beak-collor (values: yealow, pink, red), paw-style (values: 2 or 4 fingers), paw-color (values: pink, red, white), eye-color (values: black, brown, blue), feather-type (values: normal, bright). With this combination is possible to create 972 different types of flamingos.

Did the student provide a reasonable explanation of how they would extend the tree?

- 10 pts
  Yes
- 0 pts
  No

Editar envio

Comentários

Visível para os colegas de turma

compartilhe suas ideias...

Manuel Rodríguez Pascual 7 dias atrás

I think that in exercise 2 the user should list the properties of the nodes and edges, besides the provided "timestamp" example.

All the rest of the content is clear and well structured, a very good work.



Central de Ajuda