Exercise 5.5: Table Optimizations

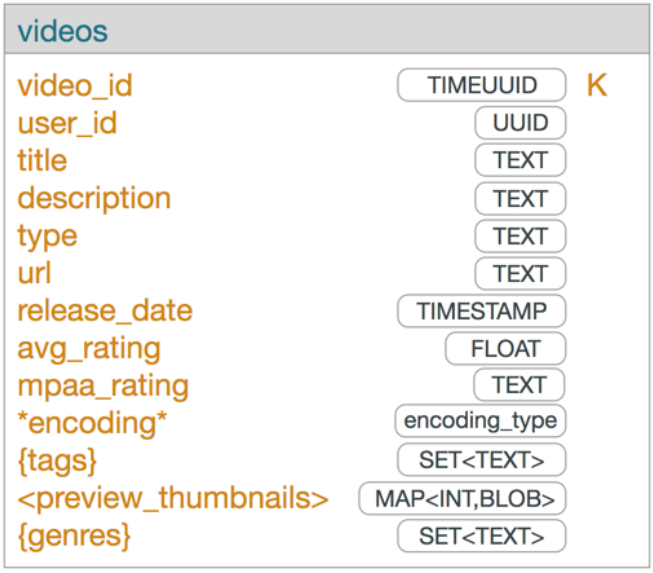
In this exercise, you will:

- Determine what tables might benefit from splitting partitions
- Determine what tables might benefit from splitting tables

Background

Even though you've already created your data model, there is some room for improvement. You have been assigned to take a look at some of the table diagrams and to analyze whether any further optimization is necessary.

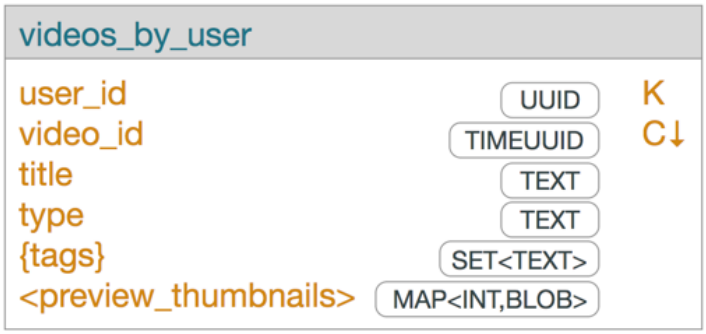
Steps



In the `videos` table, there is some concern that the `preview_thumbnails` column might make partitions too large to be manageable. The column contains screen capture images from videos, with one screenshot for every 20 seconds in the video. The average screenshot is about 20 KB and the longest expected video would be 6 hours.

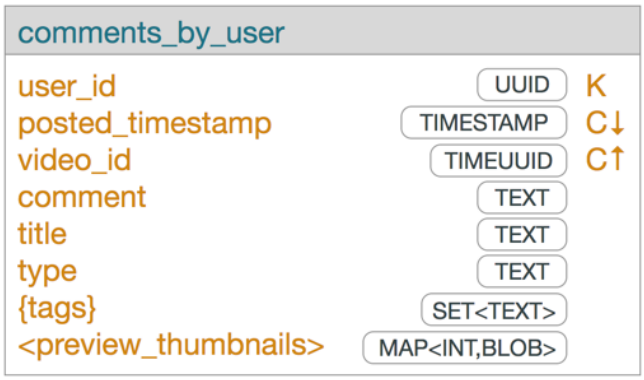
- What would be the estimated size of the column for the worst case of having a 6 hour video?
- Is this table a good candidate for splitting partitions?

- If the table is a candidate, what would be a viable way to split the partitions?



The `videos_by_user` table also has the same `preview_thumbnails` column, but in this case there may be multiple videos nested within a user partition. A user may upload several to hundreds of videos, which means that the thumbnail images would all need to be stored in that one partition.

- Assume that in the worst case, any individual user may upload 500 videos. What would the total size of the `preview_thumbnails` column by for that partition?
- Is this table a good candidate for splitting partitions?
- If the table is a candidate, what would be a viable way to split the partitions?



The last table to consider is the `comments_by_user` table. As before, there is a `preview_thumbnails` column. However this table stores all of the comment and video information for a user in a partition. A user may foreseeably comment on hundreds of videos, which gets stored in one partition along with those preview thumbnail images.

- The partitions will probably be too large, we will consider splitting the data up in some way.
- Are there any viable ways to split the partitions in the `comments_by_user` table?

One other possibility may be to split the table itself.

- What columns would remain in the `comments_by_user` table, and what columns will move to a different table?
- What would be the queries that can be used to access the columns in each of the tables?