

CS 327E Project 3, due Thursday, 09/23.

This assignment builds upon the `spanner.ipynb` notebook we worked on in class.

Make a copy of the `spanner.ipynb` notebook and rename it to `project3.ipynb`. Implement the following logic in the `project3.ipynb` notebook:

- Change the appropriate variables for your environment and run through all the commands in the notebook in the proper sequence.
- Add a markdown comment on the last cell with the heading “begin project 3 work”.
- Populate the `reviews` table in the shopify database from `reviews.csv`.
- Run a count on the `reviews` table.
- Describe the `reviews` table using the `show create table` command.
- Insert 100,000 fake reviews into the `reviews` table. Reviews should be inserted in batches of 1000 using the `batch_insert` method, similar to how we loaded the previous tables in the shopify database. Apply the following logic to generate the reviews records:
 - `app_id` should be assigned the constant value `'6286d4cd-3fcb-4ee3-bb3c-de051c28b83c'`.
 - `posted_at` should be assigned a constant date value (e.g. “2021-09-20”).
 - `author` should be assigned a constant value (e.g. “Homer Simpson”) followed by a random number to ensure uniqueness.
 - `rating` should be assigned a random integer value between 1 - 5.
- Update the corresponding `apps` record in the `apps` table with an average rating and review count based on your assigned values. Wrap this logic into a transaction.
- From the Spanner UI, run the following query and review the execution plan by looking at the explanation tab. Make note of the run time for the query (total elapsed time). Download the execution plan by clicking on the Download JSON button and name the file `query_text1.json`.

```
select * from apps a join reviews r on a.id = r.app_id
where a.rating >= 5.0 and r.author = 'Funky Moose Records';
```

- Think of an index to speed up the above query and create the index from your notebook.
- Go back to the Spanner UI, re-run the same query and review the execution plan. Were you successful in speeding up the query? Add a Markdown comment in your notebook

(right below the index creation statement) that briefly describes your results. Also, download the latest execution plan (name the file `query_text2.json`).

- Look for any missing foreign key constraints between the shopify tables and add them using the `ALTER TABLE` command.
- Describe each table which you modified in the previous step.

CS 327E Project 3 Rubric

Due Date: 09/23/21

<p>Create a copy of the <code>spanner.ipynb</code> notebook and rename it as <code>project3.ipynb</code></p> <ul style="list-style-type: none"> -5 incorrect file name -5 copied incorrect notebook 	5
<p>Run through all the commands in the <code>project3.ipynb</code> notebook (which were copied from <code>spanner.ipynb</code>).</p> <ul style="list-style-type: none"> -3 for each command not run -3 for each command which produced an error 	12
<p>Populate the <code>reviews</code> table in the shopify database and run a count on this table.</p> <ul style="list-style-type: none"> -10 <code>reviews</code> table is empty or incorrect CSV file was used -3 count is incorrect 	10
<p>Describe the <code>reviews</code> table using the <code>show create table</code> command.</p> <ul style="list-style-type: none"> -3 output is incorrect 	3
<p>Insert batches 1000 reviews for a total of 100,000 review records and one updated app record.</p> <ul style="list-style-type: none"> -30 review records missing or inserted incorrectly -10 incorrect update statement or not wrapped in transaction 	40
<p>Create an index for the mentioned query. Run query before and after index creation and describe results in Markdown. Include execution plans in files <code>query_text1.json</code>, <code>query_text2.json</code>.</p> <ul style="list-style-type: none"> -10 invalid or irrelevant index created. -5 missing brief description of results. -5 missing one or both query execution plans (<code>query_text1.json</code>, <code>query_text2.json</code>). 	20
<p>Find and add missing foreign key constraints.</p> <ul style="list-style-type: none"> -5 for each missing foreign key constraint -5 for each incorrect foreign key constraint -2 for each missing describe table 	10
<p><code>project3.ipynb</code> pushed to your group's private repo on GitHub. Your project will not be graded without this submission. Please also remember to include <code>query_text1.json</code> and <code>query_text2.json</code> in your repo.</p>	Required
<p><code>submission.json</code> submitted into Canvas. Your project will not be graded without this submission. The file should have the following schema:</p> <pre>{ "commit-id": "your most recent commit ID from GitHub", "project-id": "your project ID from GCP" }</pre> <p>Example:</p>	Required

<pre>{ "commit-id": "dab96492ac7d906368ac9c7a17cb0dbd670923d9", "project-id": "some-project-id" }</pre>	
Total Credit:	100