ETL process

Authors

1. Export authors from mongo using the MongoDB Compass export function
   1. Open collection to be exported..in this case “authors”
   2. Top left - click on “collection”…”export collection”
   3. Click on green button “select fields” in bottom right
   4. Unclick “id”
   5. Click on csv box
   6. Select output location and name by clicking on “browse” button
   7. Click on export button….done!
2. Load into Postgres – getting & failure due to the registry error
   1. Used this:

COPY author(author\_name,born,description)

FROM 'C:\Users\edwar\Desktop\Projects\project\_gelato\Resources\authors.csv' DELIMITER ',' CSV HEADER;

* 1. Had to fix permissions with this:

<https://stackoverflow.com/questions/14083311/permission-denied-when-trying-to-import-a-csv-file-from-pgadmin>

Workaround: Had to build the database table author allowing “null” due to Madeliene L ‘Engle nulls

Author load completed successfully

Quotes

1. Followed the above and unloaded
2. Used the below to load

COPY quotes(author\_name,text)

FROM 'C:\Users\edwar\Desktop\Projects\project\_gelato\Resources\quotes.csv' DELIMITER ',' CSV HEADER;

1. Had to manually change the field to “text” due to quotes > 300 characters
2. Checked and the quote\_id is assigned sequentially

Quote\_tags - ETL

1. First Create initialization file: Quote\_tags\_initialize which has quote and tag only
2. Run to load:

COPY Quote\_tags\_initialize (text,tag)

FROM 'C:\Users\edwar\Desktop\Projects\project\_gelato\Resources\Mongo\_raw\_quote\_tags.csv' DELIMITER ',' CSV HEADER;

1. Join with the quotes database on the “quote” and insert the quote\_id and tag into the quote\_tag using sql: (this is the populate\_Quote\_Tags.sql file)

insert into quote\_tags

(select q.quote\_id, qti.tag

from quote\_tags\_initialize as qti,

quotes as q

where qti.text = q.text)

1. Export the data to csv

COPY quote\_tags TO 'C:\Users\edwar\Desktop\Projects\project\_gelato\Resources\quote\_tags.csv' WITH (FORMAT CSV, HEADER);

Quote\_tags – Load after csv built above (for Claire and Carter)

1. Create the table with the foreigh key
2. Run to load:

COPY Quote\_tags(quote\_id,tag)

FROM 'C:\Users\edwar\Desktop\Projects\project\_gelato\Resources\quote\_tags.csv' DELIMITER ',' CSV HEADER;

Special Note – manual cleanup done on the quote\_tags file:

* After discussion, the 3 “[]” entries were removed from the CSV file…

General link to postgres loads – code and manual

<https://www.postgresqltutorial.com/import-csv-file-into-posgresql-table/>