

```
1 #!/bin/bash -l
2
3 # Use the stacclass partition. Only applies if you are in STA141C
4 #SBATCH --partition stacclass
5
6 # Give the job a name
7 #SBATCH --job-name subset_sqlite
8
9 # Send me an email when the job is completed!
10 #SBATCH --mail-type=ALL
11 #SBATCH --mail-user=ebatzner@ucdavis.edu
12
13 # Loading in bio module for sqlite3
14 module load bio
15
16 # Setting location of SQL file
17 DATAFILE="/scratch/usaspending.sqlite"
18
19 # Generating SQL command to select relevant rows, filter data
20 printf "SELECT DISTINCT name, subtier_agency_id
21 FROM subtier_agency
22 ORDER BY name ASC;\n" > unique_subtiers.sql
23
24 # Join on meaningful columns
25 printf "SELECT award_id, action_date, fiscal_year, total_obligation,
26 generated_pragmatic_obligation, transaction_description, award_category,
27 pop_country_name, pop_state_code, pop_county_name, pop_zip5, naics_code,
28 awarding_subtier_agency_name, funding_subtier_agency_name, business_categories
29 FROM universal_transaction_matview
30 LEFT JOIN (SELECT DISTINCT name, subtier_agency_id FROM subtier_agency)
31 ON awarding_subtier_agency_name=name
32 WHERE subtier_agency_id = 778
33 OR subtier_agency_id = 776
34 OR subtier_agency_id = 257" > filter_rows_awarding.sql
35
36 # Join on meaningful columns
37 printf "SELECT award_id, action_date, fiscal_year, total_obligation,
38 generated_pragmatic_obligation, transaction_description, award_category,
39 pop_country_name, pop_state_code, pop_county_name, pop_zip5, naics_code,
40 awarding_subtier_agency_name, funding_subtier_agency_name, business_categories
41 FROM universal_transaction_matview
42 LEFT JOIN (SELECT DISTINCT name, subtier_agency_id FROM subtier_agency)
43 ON funding_subtier_agency_name=name
44 WHERE subtier_agency_id = 778
45 OR subtier_agency_id = 776
46 OR subtier_agency_id = 257" > filter_rows_funding.sql
47
48 # Find rows that match our subtier agency criteria
49 cat unique_subtiers.sql | sqlite3 -header -csv ${DATAFILE} > subtier_names.csv
50
51 # Run SQL command and produce subsetted dataset
52 cat filter_rows_awarding.sql | sqlite3 -header -csv ${DATAFILE} > filtered_data_awarding.csv
53
54 # Run SQL command and produce subsetted dataset
55 cat filter_rows_funding.sql | sqlite3 -header -csv ${DATAFILE} > filtered_data_funding.csv
56
```