**Steps to reload scraped job data**

1. **Delete existing screen scraped jobs:**

Before reloading the new data, user can run below MySQL query to clear the existing job data in the database from screen scraping. In the below query, job data is deleted for Encore Organization.

use encore\_dev;

delete from lt\_data;

set @encore\_org\_id = (select OrganizationId from Organization where OrganizationName = 'Encore');

set @encore\_emp\_id = (select UserId from Users where OrganizationId = @encore\_org\_id LIMIT 1);

delete from MedicalJobSeekerSavedJobs

where MedicalEmployerJobPostingId

in (select MedicalEmployerJobPostingId from MedicalEmployerJobPosting where MedicalEmployerUserId = @encore\_emp\_id);

delete from MedicalJobSeekerJobApplicationInfo

where MedicalEmployerJobPostingId

in (select MedicalEmployerJobPostingId from MedicalEmployerJobPosting where MedicalEmployerUserId = @encore\_emp\_id);

delete from MedicalEmployerJobPosting where MedicalEmployerUserId = @encore\_emp\_id;

1. **Generate json data via screenscraping:**

There exist lt\_data table in database which stores json format screen scraped data from locum tenens site. EncoreMiscellaneous project contains java helper classes which can be used for screen scrapping the data and loading the data to database.

By running main method of java class com.encore.datapull.LT.java with program arguments - <URL> <StartLink> <EndLink> will save all the job data in json format from <StartLink> to <EndLink> for the <ULR> and populate the json data in lt\_data table.

Example to pull links 3, 4, 5 from url - https://www.locumtenens.com/anesthesiology-jobs/crna/ohio/ 3 5

1. **Load json data to DB:**

Job data from lt\_data table need to be parsed and stored to MedicalEmployerJobPosting table and later jobs from the table MedicalEmployerJobPosting are used in the web api calls.

By running the main method of java class com.encore.dataload.JobPostingDataLoad will parse the json data from lt\_data and save it in MedicalEmployerJobPosting. Below are some of the defaults used while parsing data:

* If compensation rate is not specified, then 100/hr will be used as default.
* Years of experience is set as 3 years by default.
* MalpracticeAcceptable, Supervision and PrescriptionAuthorityNeeded will be set as false by default if the values are not present in job information.
* If only month is specified for assignment from or to date, then 1st day of the month will be used for assignment from date and last day of the month will be used as assignment to date. For eg. Aug – Oct will be assumed as AssignmentFromDate 1st Aug 2021 and AssignmentToDate 31st Oct 2021.
* If AssignmentFromDate and AssignmentToDate both are not present, then first day of next month will be used as AssignmentFromDate and AssignmentToDate will be next 6 months end date from AssignmentFromDate.
* If AssignmentToDate falls behind today’s date then the job will be not saved to database.
* Last date of the current year will be used as job ExpiryDate.
* If shift type is not found in job, then by default ShiftType Day will be used.