SQL\*LOADER TUTORIAL

1. Environment Setup:

* Go to <https://github.com/Yelp/dataset-examples>
* Download zip file of dataset-examples directory and unzip it under Documents directory.
* Open a terminal window and install all package dependencies:

$ cd Documents

$ pip install –e .

To test:

$ tox

Note: if you are using a Python version below 2.7.9, you will need to

install pip before you can run the above pip command.

2. Download the latest yelp dataset from: <https://www.yelp.com/dataset_challenge/dataset> or alternatively use the sample dataset yelp\_academic\_dataset\_business.json that’s included with this tutorial.

3. Use json\_to\_csv\_converter\_clean.py to create a clean csv file:

$ python json\_to\_csv\_converter\_clean.py yelp\_academic\_dataset\_business\_clean.json

4. Rename column names in the generated yelp\_academic\_dataset\_business\_clean.csv file as follows:

* open 🡪 open\_bool
* name 🡪 name\_desc
* type 🡪 type\_desc

5. Create an Oracle schema to load the data into. In this example, we are using ‘yelp’ as the schema name, but feel free to use a different name.

* Connect to yelp as SYS or SYSTEM
* sql> create user yelp identified by yelp;
* sql> grant connect, resource, dba to yelp;
* sql> grant create session to yelp;
* sql> grant unlimited tablespace to yelp;

1. Connect to the yelp schema and create a business table using create\_business\_table.sql
2. Optional step: load the dataset using SQL Developer:

* Right-click on the business table in the yelp schema and choose “import data” and select yelp\_academic\_dataset\_business\_clean.csv
* Click the next button and accept the defaults. It should take around 22 seconds to load 60,000 rows.

Note: This is an optional step. If you don’t want to use SQL

Developer, feel free to skip directly to step 8.

1. Load the dataset using SQL Loader:

* Make sure that $ORACLE\_HOME/bin is in your $PATH:

echo $ORACLE\_HOME/bin

* Open yelp.ctl and edit the infile path value. Replace the path '/home/yuming/Downloads/yelp\_academic\_dataset\_business\_clean.csv' with the actual path where this csv file is stored on your machine
* Use the edited yelp.ctl file to load the dataset using the SQL Loader executable called sqlldr:

$ cd /home/yuming/Downloads

$ sqlldr yelp/yelp control=yelp.ctl

SQL Loader will create a log file summarizing the load status and it will also generate an error file which contains any errors that occurred during the load.