Druid\_Indexer was written in python for the purpose of performing following actions

**Actions Description**

launch Launch kafka indexing tasks

shutdown Shutdown kafka indexing tasks

display\_all Display all running kafka indexing tasks

reindex Reindex datasources

./druid\_indexer.py -h to see the help menu (lines above)

**launch**

./druid\_indexer.py launch -h to see the all the arguments and usage of action ‘launch’ below

usage: druidindexer launch [-h] --hostname\_port HOSTNAME\_PORT [--all]

[--file\_name\_list FILE\_NAME\_LIST [FILE\_NAME\_LIST ...]]

--wait WAIT\_TIME\_KAFKA\_INDEX

optional arguments:

-h, --help show this help message and exit

--hostname\_port HOSTNAME\_PORT Set the hostname and port for druidindexer (required)

--all Launch all kafka indexing tasks - associated json files locating under kafkatask directory

--file\_name\_list FILE\_NAME\_LIST [FILE\_NAME\_LIST …] list of kafka indexing task - json files under kafkatask

directory

--wait WAIT\_TIME\_KAFKA\_INDEX time interval in minutes between kafka indexing tasks

The sample usage of script with launch action is as follows

./druid\_indexer.py launch --hostname\_port localhost:8090  --wait 3   --all

In the first run of the script with the action of launch it creates druidindexer folder with the following subfolder structure if not exists, and then exits the script

/log

/config

/template

<config files>

/tmp

/kafkatask

/pid

After exiting, make sure that you put kafka indexing task json files under kafkatask directory then run the script again

**shutdown**

./druid\_indexer.py shutdown -h to see the all the arguments and usage of action ‘shutdown’ below

optional arguments:

-h, --help show this help message and exit

--hostname\_port HOSTNAME\_PORT Set the hostname and port for druidindexer (required)

--all Shutdown all running kafka indexing tasks

--data\_source\_list DATA\_SOURCE\_LIST [DATA\_SOURCE\_LIST …] Shutdown specified kafka indexing tasks

sample usage:

./druid\_indexer.py shutdown --hostname\_port localhost:8090 --all

**display\_all**

sample usage:

./druid\_indexer.py  display\_all --hostname\_port localhost:8090

**reindex**

./druid\_indexer.py reindex -h to see the all the arguments and usage of action ‘shutdown’ below

optional arguments:

-h, --help show this help message and exit

--input\_csv\_file CONFILE\_PATH config file path for reindexing (required)

--check\_period CHECK\_PERIOD time interval in minutes for checking # of running task (default 60 min)

--config\_item\_list CONF\_ITEM\_LIST [CONF\_ITEM\_LIST ...] item numbers of configFile

--max\_task MAX\_TASK\_REINDEX number of concurrent tasks (required)

sample usage:

./druid\_indexer.py reindex --input\_csv\_file /home/kafka/druidindexer/config/template/config\_file.csv --max\_task 2 --check\_period 3

In the first run of the script with the action of reindex, it creates druidindexer folder with the following subfolder structure given above if not exists, and then exits the script

After exiting, make sure that you first put the "reindex\_temp.json", "metricsSpec\_template.json" and "dimensionsSpec\_template.json" under config/template , then run the script again