{"device":"laptop","type":[{"name":"dell"},{"name":"lenovo"},{"name":"app
le"}],"browser":{"name":"mozila","version":5}}
{"device":"mobile","type":[{"name":"moto"},{"name":"samsung"}],"browser":
{"name":"chrome","version":2}}
{"device":"laptop","type":[{"name":"samsung"},{"name":"hp"},{"name":"sufa
ce"}],"browser":{"name":"mozila","version":5}}
{"device":"mobile","type":[{"name":"mi"},{"name":"vivo"}],"browser":{"nam
e":"chrome","version":3}}

Question: Read this json using pig on hdfs and create external table on stored data using hive. Create final table in hive as clickdata(device,cont()) and insert data from above external table. Use oozie to schedule these workflow at every 30 mnts from 10pm to 8am.

Actions: 1. Pig- Load and store json data on hdfs, 2. Hive-External table on this data
"3. Create final table in hive, 4. insert required data from ext table to final table
"5. Drop that external table"

Solution

```
REGISTER /home/umesh/newhadoop/NotesForBatch/OtherFiles&Jars/piggybank-
0.16.0.jar;
```

{"recipe":"Tacos","ingredients":[{"name":"Beef"},{"name":"Lettuce"},{"nam
e":"Cheese"}],"inventor":{"name":"Alex","age":25}}
{"recipe":"TomatoSoup","ingredients":[{"name":"Tomatoes"},{"name":"Milk"}
],"inventor":{"name":"Steve","age":23}}

2. Reading Nested JSON Data:

```
> table2 = LOAD 'second.json' USING
JsonLoader('recipe:chararray,ingredients: {(name:chararray)}, inventor:
name:chararray, age:int)');

myjson = LOAD '/home/umesh/Desktop/sample.json' USING
JsonLoader('device:chararray,type:{(name:chararray)},brower:(name:chararr
ay,version:int)');


table2 = LOAD
'/home/umesh/NotesForBatch_Updated/Pig/PigAdvance/second.json' USING
JsonLoader('recipe:chararray,ingredients: {(name:chararray)},
inventor:name:chararray,age:int)');
```