Importing JSON Data in HIVE using Native Functions

Input Data is in JSON format

[{"id":1,"first_name":"Alexandrina","email":"abalding0@goo.ne.jp","date":"9/11/2016","country":"Nigeria"},
{"id":2,"first_name":"Craggie","email":"ctremain1@photobucket.com","date":"12/7/2016","country":"China"},
{"id":3,"first_name":"Allie","email":"adearan2@ycombinator.com","date":"10/31/2016","country":"Russia"},
{"id":4,"first_name":"Poul","email":"psondland3@mlb.com","date":"9/11/2016","country":"Nigeria"},
{"id":5,"first_name":"Delmor","email":"dbrion4@behance.net","date":"9/6/2016","country":"Philippines"},
{"id":6,"first_name":"Verine","email":"vmcquorkell5@upenn.edu","date":"9/22/2016","country":"China"}]

1. Creating intermediate table to store the data

```
CREATE TABLE format.json_table (json string);
LOAD DATA INPATH  '/user/cloudera/JSON/MOCK_DATA.json' INTO TABLE json_table;
```

Since there are no delimiters, we leave off the ROW FORMAT section of the table DDL

Built in function #1: **get_json_object**

The get_json_object takes two arguments: tablename.fieldname and the JSON field to parse, where '$' represents the root of the document.

```
1.select get_json_object(json_table.json, '$') from json_table;
```
Returns the full JSON document.

2. Prepare final table and load the data within

```
SET hive.exec.dynamic.partition= true;
SET hive.exec.dynamic.partition.mode= nonstrict;

Create table format.prtn_json (id int, name string, email string)
partitioned by (date string, country string)
row format delimited fields terminated by '\t'
stored as textfile;

INSERT OVERWRITE TABLE format.prtn_json
partition (date,country)
select get_json_object(json_table.json, '$.id') as id,
get_json_object(json_table.json, '$.first_name') as name,
get_json_object(json_table.json, '$.email') as email,
get_json_object(json_table.json, '$.date') as date,
get_json_object(json_table.json, '$.country') as country from
format.json_table;
```

Output:
hive (format)> dfs -ls /user/hive/warehouse/format.db/prtn_json;

```
Found 30 items
drwxrwxrwx   - cloudera supergroup          0 2017-04-25 17:06
/user/hive/warehouse/format.db/prtn_json/date=10%2F22%2F2016
drwxrwxrwx   - cloudera supergroup          0 2017-04-25 17:06
/user/hive/warehouse/format.db/prtn_json/date=10%2F31%2F2016
drwxrwxrwx   - cloudera supergroup          0 2017-04-25 17:06
/user/hive/warehouse/format.db/prtn_json/date=10%2F9%2F2016
drwxrwxrwx   - cloudera supergroup          0 2017-04-25 17:06
/user/hive/warehouse/format.db/prtn_json/date=11%2F3%2F2016
..etc
```

For nested inputs like structs,tuples use SerDes