# JsonSerde - a read/write SerDe for JSON Data

Features:

* Read data stored in JSON format
* Convert data to JSON format when INSERT INTO table
* arrays and maps are supported
* nested data structures are also supported.
* modular to support multiple versions of CDH

IMPORTANT!!! READ THIS BELOW!! Json records must be *one per line*, that is, the serde WILL NOT WORK with multiline Json. Why ? Because the way hadoop works with files, they have to be *splittable*, for instance, hadoop will split text files at end of line..but in order to split a text file with json at a certain point, we would have to parse everything up to that point. See below

Input data:

{"field1":"data1","field2":100,"field3":"more data1","field4":123.001}

{"field1":"data2","field2":200,"field3":"more data2","field4":123.002}

{"field1":"data3","field2":300,"field3":"more data3","field4":123.003}

{"field1":"data4","field2":400,"field3":"more data4","field4":123.004}

add jar /home/MyJars/json-serde-1.3.6-SNAPSHOT-jar-with-dependencies.jar;

CREATE EXTERNAL TABLE IF NOT EXISTS my\_table (

field1 string, field2 int, field3 string, field4 double

)

ROW FORMAT SERDE 'org.openx.data.jsonserde.JsonSerDe';

LOAD DATA INPATH '/user/hue/Hiveinput/jsoninput.txt'

OVERWRITE INTO TABLE my\_table;

hive> select \*from my\_table;

OK

data1 100 more data1 123.001

data2 200 more data2 123.002

data3 300 more data3 123.003

data4 400 more data4 123.004

Time taken: 0.163 seconds, Fetched: 4 row(s)