

hive关联es , json嵌套, struct, array 等特殊类型数据处理

使用Hive读写ElasticSearch中的数据 <http://blog.ask3.cn/2253.html>

嵌套的数据从官网和google都没有找到答案，现在终于解决了。

json嵌套需要安装

[Hive-JSON-Serde](#)

举例：

es数据

```
{
  "_index": "rule_task_param",
  "_type": "fqd_ascore",
  "_id": "AVtXXR29bAbb6cOHQQE4",
  "_score": 1,
  "_source": {
    "appName": "rca",
    "appRequestId": "1f11a99b_af84_4d23_b870_92228fa01216",
    "curStepBean": "4cc503d8_a730_42dc_a850_eec805372293",
    "curStepId": 1820290,
    "curStepName": "脚本：分期AScore_总分",
    "curTime": 1491818645234,
    "execMode": "NO_TEST",
    "inputParam": {
      "rec_fst_call": 95,
      "borrow_nid": "201704101701021818328178",
      "name_length": 3,
      "company_address": "鄂尔多斯工务段",
      "job_office": "线路工",
      "id_card_6": "152222",
      "job_city": "鄂尔多斯",
      "phone_district": "",
      "diploma": "-9999",
      "job_district": "东胜区",
      "all_record_reg_num": 1,
      "test": "SUCCESS",
      "education_degree_w": "大专",
    }
  }
}
```

```
    "link_cus_same_prov_num": 59,
    "job_province": "内蒙古自治区",
    "graduate_school_level": 0,
    "record_in_ratio": 0.33,
    "user_id": "12964128",
    "company_name": "呼和浩特铁路局",
    "id_card_province": "内蒙古自治区",
    "credit_succ_cards": -1,
    "income": 8000,
    "id_card_district": "兴安盟",
    "user_id_num": "8",
    "gender": 1,
    "id_card_city": "兴安盟",
    "job_tel": "04716968741",
    "is_full_time": 0,
    "lxr_nounique_num": 0,
    "current_borrow_time_1": 1491818328,
    "current_borrow_time_2": 1491818328,
    "phone_city": "包头",
    "sex": 1,
    "id_prov_district": "华北",
    "job_tel_same_number_cnt": 1,
    "phone_province": "内蒙古",
    "job_tel_big_result": 0,
    "max_login_gap": 34511,
    "age": 20
  },
  "ipAddress": "10.15.12.13",
  "outputParam": {
    "fqd_ascore": 399,
    "fqd_ascore_detail": [
      -9
    ],
    40
  },
  53
  },
  43
  },
  22
  },
  41
  },
  49
  },
  41
```

```

        ,
        60
        ,
        27
        ,
        32
    ]
},
"productType": "fqd_ascore",
"ruleVersion": "20161230181143",
"taskId": 1133976,
"taskTime": "20170410180405"
}
}

```

hive关联es

```

add jar hdfs://boss/user/admin/jar/json-serde-1.3.7-jar-with-dependencies.jar;
add jar hdfs://boss/user/admin/jar/json-udf-1.3.7-jar-with-dependencies.jar;
drop TABLE if exists rule_task_param;
CREATE EXTERNAL TABLE rule_task_param (
appName string,
appRequestId string,
curStepBean string,
curStepId bigint,
curStepName string,
curTime bigint,
execMode string,
inputParam struct<
rec_fst_call:string,
borrow_nid:string,
name_length:string,
company_address:string,
job_office:string,
id_card_6:string,
job_city:string,
phone_district:string,
diploma:string,
job_district:string,
all_record_reg_num:string,
test:string,
education_degree_w:string,
link_cus_same_prov_num:string,
job_province:string,

```

```

graduate_school_level:string,
record_in_ratio:string,
user_id:string,
company_name:string,
id_card_province:string,
credit_succ_cards:string,
income:string,
id_card_district:string,
user_id_num:string,
gender:string,
id_card_city:string,
job_tel:string,
is_full_time:string,
lxr_nounique_num:string,
current_borrow_time_1:string,
current_borrow_time_2:string,
phone_city:string,
sex:string,
id_prov_district:string,
job_tel_same_number_cnt:string,
phone_province:string,
job_tel_big_result:string,
max_login_gap:string,
age:string
>,
borrow_nid string,
ipAddress string,
outputParam struct<
fqd_ascore:string,
fqd_ascore_detail:array<string>
>,
productType string,
ruleVersion string,
taskId bigint,
taskTime string
)
ROW FORMAT SERDE 'org.openx.data.jsonserde.JsonSerDe'
STORED BY 'org.elasticsearch.hadoop.hive.EsStorageHandler'
TBLPROPERTIES(
'es.nodes' =
'10.15.46.111:9200,10.15.182.137:9200,10.15.192.236:9200,10.15.148.40:9200,10.15.127.143:9200',
'es.index.auto.create' = 'false',
'es.resource' = 'rule_task_param',
'es.output.json' = 'true',
'es.mapping.names' = 'cookieid:_metadata._id, appName:appName, appRequestId:appRequestId,
curStepBean:curStepBean, curStepId:curStepId ,curStepName:curStepName ,curTime:curTime

```

```
, execMode:execMode, inputParam:inputParam, borrow_nid:borrow_nid, ipAddress:ipAddress, outputParam:outputPa  
select inputParam.borrow_nid ,outputParam.fqd_ascore_detail[0],* from rule_task_param where  
inputParam.rec_fst_call=-1 limit 10;
```

使用:

```
select inputParam.borrow_nid ,outputParam.fqd_ascore_detail[0],* from rule_task_param where  
inputParam.rec_fst_call=-1 limit 10;
```