

# json 数据导入hive 利用 get\_json\_object 和json\_tuple 函数

通过flume把kafka中的数据落到hdfs。但是数据是json格式。现在想办法把数据放到hive。

数据源:

```
hadoop fs -cat /datahouse/ods/topic/bigdata_rca_jsystem_message/2018-03-18/00.1521302400714 | head

{"appDomain":"pub-message-api","data":
{"batchNumber":"1670318000017179300nx","commitResult":"提交成功
","commitStatus":1000,"commitTime":1521302417689,"contents":"您的手机验证码为：328644，有效期
为30分钟。关注官方微信“手机贷”，了解更多详情！
","createTime":1521302417689,"department":"","flagId":"fb2d5f283d6f4472ae17e83c1562c67b","pl
{},"sendStatus":0,"sendTime":1521302417688,"sendType":"sms","serverIp":"10.15.141.96","ser
玄
","smsThirdCode":"nx","updateTime":1521302417689,"userId":4133736,"usetype":5},"domain":"jsy
```

--建立外部扩展hive表

```
create EXTERNAL table h_biz_data.MESSAGE_INFO_sms_flume (content string)
location 'hdfs:///datahouse/ods/topic/bigdata_rca_jsystem_message/2018-03-18/';
```

-- 创建分区表

```
CREATE TABLE h_biz_data.MESSAGE_INFO_sms_new(
key string,
flagId string,
sendStatus string,
sendTime string
)PARTITIONED BY(dt STRING);
```

---增量数据处理 利用 hive的 get\_json\_object 和json\_tuple 函数

```
insert OVERWRITE table h_biz_data.MESSAGE_INFO_sms_new PARTITION (dt)
select
get_json_object(tt.data,'$.batchNumber'),
get_json_object(tt.data,'$.flagId'),
get_json_object(tt.data,'$.sendStatus'),
get_json_object(tt.data,'$.sendTime'),
'2018-03-18'
from(select json_tuple(content, "data") as data from h_biz_data.MESSAGE_INFO_sms_flume) as tt
```

如果不知道是否能解析到数据，单条数据的验证可以用下面方法验证：

```
select get_json_object('{"batchNumber":"180100000011499572mandao","commitResult":"提交成功
","commitStatus":1000,"commitTime":1521306001084,"contents":"尊敬的卡宜贷用户，您有4000.0元的额度
需立即提现，赶紧猛戳https://m.91kavidai.com申请提现！
```

\\",\\"createTime\\":1521306001084,\\"department\\":\\"\\",\\"flagId\\":\\"ae5fd11f694f4ec6b8074790d1f92729\\",\\"phone  
{},\\"sendStatus\\":0,\\"sendTime\\":1521306001084,\\"sendType\\":\\"sms\\",\\"serverIp\\":\\"10.15.19.130\\",\\"service  
道  
\\",\\"smsThirdCode\\":\\"mandao\\",\\"updateTime\\":1521306001084,\\"userId\\":21231580,\\"usetype\\":1}','\$.sendTime