

## HIVE中map , array和structs使用

笔记本： 大数据

创建时间： 2015/8/12 9:16

更新时间： 2015/8/12 10:12

URL： <http://stackoverflow.com/questions/11373543/explode-the-array-of-struct-in-...>

---

1: 怎样导入文本文件（格式是怎样的？），2: 怎样查询数据，已经能否在join中使用？在子查询中使用？等等

知道怎么在hive中导入数组不？

例如：我想把 数组[1, 2, 3] 和 数组 ["a", "b", "c"]

导入到table1中

```
create table table2 ( a array<int> , b array<string> );
```

那么 我如何 导入呢？使得

```
select * from table1;
```

结果为：

```
[1, 2, 3] ["a", "b", "c"]
```

同样 在 hive 中 对于 map

怎样 查询呢？

例如

```
create table table2 ( a MAP<STRING, ARRAY<STRING>>);
```

select \* from table2 结果为：

```
{"d01":["d011","d012"], "d02":["d021","d022"]}
```

```
{"d01":["d011","d012"], "d02":null}
```

```
{"d01":null, "d012","d02":["d021","d022"]}
```

那么 我想获得 key 为 d01的value值

该怎么操作呢

关于数组的操作说明：

```
drop table table2;
```

```
create table table2 (a array<string>, b array<string>)
```

```
ROW FORMAT DELIMITED
```

```
FIELDS TERMINATED BY '\t'
```

```
COLLECTION ITEMS TERMINATED BY ',';
```

```
load data local inpath "../hive/examples/files/arraytest.txt" overwrite into table table2;
```

arraytest.txt中的数据形式为：（不同数组间用\t分割，同一数组内不同元素用逗号分割）

```
b00, b01
```

```
b00, b01
```

```
b00, b01      b00, b01
b00, b01      b00, b01
b00, b01      b00, b01
```

```
hive> select * from table2;
OK
["b00", "b01"]  ["b00", "b01"]
["b00", "b01"]  ["b00", "b01"]
["b00", "b01"]  ["b00", "b01"]
["b00", "b01"]  ["b00", "b01"]
Time taken: 0.056 seconds
```

```
hive> select a from table2;
OK
["b00", "b01"]
["b00", "b01"]
["b00", "b01"]
["b00", "b01"]
Time taken: 15.903 seconds
```

```
hive> select a[0] from table2;
OK
b00
b00
b00
b00
Time taken: 12.913 seconds
```

```
hive> select * from table2 where a[0] = b[0];
OK
["b00", "b01"]  ["b00", "b01"]
["b00", "b01"]  ["b00", "b01"]
["b00", "b01"]  ["b00", "b01"]
["b00", "b01"]  ["b00", "b01"]
Time taken: 11.803 seconds
```

关于map的操作说明：  
drop table table2;

```
hive> CREATE TABLE table2 (foo STRING , bar MAP<STRING, STRING>)
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY '\t'
> COLLECTION ITEMS TERMINATED BY ','
> MAP KEYS TERMINATED BY ':'
> STORED AS TEXTFILE;
```

```
hive> load data local inpath "../hive/examples/files/maptest.txt" overwrite into table
table2;
```

maptest.txt中的文件格式为：（不同列之间用一个tab分割，map中key和value用冒号分割，不同K/V间用逗号分割）

```
a00      b0:b01,b1:b11
a01      b1:b11,b2:b12
a02      b2:b12,b3:b13
a03      b3:b13,b4:b14
```

```
hive> select bar from table2;
```

OK

```
{"b0":"b01","b1":"b11"}
```

```
{"b1":"b11","b2":"b12"}
```

```
{"b2":"b12","b3":"b13"}
```

```
{"b3":"b13","b4":"b14"}
```

Time taken: 19.237 seconds

怎么根据 key来查询value呢？

```
hive> select bar['b1'] from table2;
```

OK

b11

b11

NULL

NULL

Time taken: 11.65 seconds

查看map中的键值对个数：

```
hive> select size(bar) from table2;
```

OK

2

2

2

2

Time taken: 12.137 seconds

ref: <http://www.cnblogs.com/subsir/articles/2606764.html>  
<http://blog.csdn.net/yfkiss/article/details/7842014>

## Array of Struct in Hive

<http://stackoverflow.com/questions/25088488/use-inlinearraystruct-struct-in-hive>  
<http://stackoverflow.com/questions/11373543/explode-the-array-of-struct-in-hive>