

# 1 select表忘记选择分片

后果：会对所有分片进行操作，如果是进行join等计算密集任务，会导致集群阻塞  
hive表不同于mysql中的表，如果是全量表，一张表通常由很多分片组成。如果不指定分片，默认对所有分片进行操作。

以表data\_center.dim\_merge\_showing\_house\_day为例

```
1. hive> show partitions data_center.dim_merge_showing_house_day;
2. pt=20160418000000
3. pt=20160419000000
4. pt=20160420000000
5. pt=20160421000000
6. pt=20160422000000
7. pt=20160423000000
8. pt=20160424000000
9. Time taken: 0.095 seconds, Fetched: 116 row(s)
```

解决方法：对于一张从没用过的新表，拿到表名之后先 `desc table_name;` 查看各表字段，然后 `show partitions table_name;` 查看表是否有分片。

## 2 如何结束掉一个hadoop job

在本地开启作业之后，如果作业超时或者中途发现错误，想要停掉进程。可以使用 `kill pid` 命令。切记不可使用 `kill -9 pid` 命令。

两者的主要区别是 `kill pid` 发送的是SIGTERM信号，而 `kill -9 pid` 发送的是SIGKILL信号。SIGKILL信号信号不会被系统阻塞，迫使进程在运行时突然终止，进程在结束后不能自我清理。一个形象的比喻是如果 `kill pid`，则进程还可以挣扎一下,给家人打个电话,告诉他们他就要死了,不能回家吃晚饭了. 而 `kill -9 pid`，就是进程突然死在密室里,外面等候的人也不知道他死了。

因此，`kill -9 pid` 虽然会杀掉本地进程，但是并不会杀死hadoop job，`kill pid` 会将hadoop job一并杀死。

最后，不管是使用 `kill pid` 还是 `kill -9 pid` 都建议杀掉进程之后到hadoop作业监控平台 <http://jx-bd-hadoop00.lianjia.com:8088/cluster/apps/RUNNING>上实际查看下，对应作业是否被杀死。如果对应作业进程还未被杀死，可以使用 `hadoop job -kill job_id` 手动杀死对应作业。

另外，本地提交hive查询，生成的log里含有kill集群任务和查看集群任务状态的入口，从这里也可以找到杀死hadoop作业的入口。

```
* hive -e 'use data_center; select A.house_id,A.total_price,A.created_time,A.resblock_id,A.state,A.build_area from dim_merge_showing_house_day as A where pt="20160423000000" and (house_usage_code="107500000001" or house_usage_code="3000000003" or house_usage_code="107500000004" or house_usage_code="107500000005" or house_usage_code="107500000006" or house_usage_code="107500000007" or house_usage_code="107500000008" or house_usage_code="107500000009" or house_usage_code="107500000010" or house_usage_code="107500000011" or house_usage_code="107500000012" or house_usage_code="107500000013" or house_usage_code="107500000014" or house_usage_code="107500000015") and biz_type="200200000001" and state="200100000009";'
Logging initialized using configuration in jar:file:/home/work/local/apache-hive-1.2.1-bin/lib/hive-common-1.2.1.jar!/hive-log4j.properties
OK
Time taken: 2.122 seconds
Query ID = work_20160425104923_c1b6755b-376d-45c9-b26c-9be15f44b3e6
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1459253925558_173170, Tracking URL = http://jx-bd-hadoop00.lianjia.com:8088/proxy/application_1459253925558_173170/
Kill Command = /home/work/bin/hadoop/bin/hadoop job -kill job_1459253925558_173170
Hadoop job information for Cluster: number of reducers: 0, number of reducers: 0
```