

hive性能优化

使用分区剪裁、列剪裁

正确的写法是条件写在ON后面 或者直接写成子查询

合理使用Union All

对同一张表的union all 要比多重insert快的多，原因是hive本身对这种union all做过优化，即只扫描一次源表；

并行执行Job

```
set hive.exec.parallel=true; //打开任务并行执行
set hive.exec.parallel.thread.number=16; //同一个sql允许最大并行度，默认为8。
```

使用本地MR

```
set hive.exec.mode.local.auto=true; //开启本地mr
//设置local mr的最大输入数据量,当输入数据量小于这个值的时候会采用local mr的方式
set hive.exec.mode.local.auto.inputbytes.max=50000000;
//设置local mr的最大输入文件个数,当输入文件个数小于这个值的时候会采用local mr的方式
set hive.exec.mode.local.auto.tasks.max=10;
```