

1. 分区不能超过1w多
2. 要执行compute stats xxx 表
3. join时，把小表写前面，会把小表广播到其他节点。
4. 选择parquert 格式存储。

```
8 show table stats baofoo_cutpayment.protocol_payment_order
```

*	pk_year	pk_month	pk_day	#Rows	#Files	Size	Bytes Cached	Cache Replication	Format	Incremental stats	Location
1	2020	2020-04	2020-04-01	-1	1	1174.67MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
2	2020	2020-04	2020-04-02	-1	1	1164.33MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
3	2020	2020-04	2020-04-03	-1	1	1165.66MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
4	2020	2020-04	2020-04-04	-1	1	1151.52MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
5	2020	2020-04	2020-04-05	-1	1	1162.67MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
6	2020	2020-04	2020-04-06	-1	1	1162.36MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
7	2020	2020-04	2020-04-07	3845220	1	1166.47MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
8	2020	2020-03	2020-03-01	-1	1	1180.69MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
9	2020	2020-03	2020-03-02	-1	1	1180.69MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
10	2020	2020-03	2020-03-03	-1	1	1182.72MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
11	2020	2020-03	2020-03-04	-1	1	1182.20MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
12	2020	2020-03	2020-03-05	-1	1	1180.75MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
13	2020	2020-03	2020-03-06	-1	1	1172.68MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
14	2020	2020-03	2020-03-07	-1	1	1162.80MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
15	2020	2020-03	2020-03-08	-1	1	1164.89MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
16	2020	2020-03	2020-03-09	-1	1	1175.33MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
17	2020	2020-03	2020-03-10	-1	1	1186.52MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
18	2020	2020-03	2020-03-11	-1	1	1181.10MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
19	2020	2020-03	2020-03-12	-1	1	1172.62MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
20	2020	2020-03	2020-03-13	-1	1	1169.92MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
21	2020	2020-03	2020-03-14	-1	1	1161.06MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war
22	2020	2020-03	2020-03-15	-1	1	1165.19MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/war

执行 compute stats baofoo_cutpayment.protocol_payment_order 之后

```
13 show table stats baofoo_cutpayment.protocol_payment_order
```

*	pk_year	pk_month	pk_day	#Rows	#files	Size	Bytes Cached	Cache Replication	Format	Incremental stats	Location
1	2020	2020-04	2020-04-07	3845220	1	1166.47MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
2	2020	2020-04	2020-04-06	3793741	1	1162.36MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
3	2020	2020-04	2020-04-05	3800366	1	1162.67MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
4	2020	2020-04	2020-04-04	3539825	1	1151.52MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
5	2020	2020-04	2020-04-03	3832089	1	1165.66MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
6	2020	2020-04	2020-04-02	3799837	1	1164.33MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
7	2020	2020-04	2020-04-01	4073132	1	1174.67MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
8	2020	2020-03	2020-03-31	3399034	1	1143.93MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
9	2020	2020-03	2020-03-30	3372694	1	1143.11MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
10	2020	2020-03	2020-03-29	3286525	1	1138.56MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
11	2020	2020-03	2020-03-28	4348374	1	1187.12MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
12	2020	2020-03	2020-03-27	3872339	1	1168.37MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
13	2020	2020-03	2020-03-26	3802159	1	1164.55MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
14	2020	2020-03	2020-03-25	3867709	1	1168.68MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
15	2020	2020-03	2020-03-24	3748248	1	1163.09MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
16	2020	2020-03	2020-03-23	3858897	1	1166.83MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
17	2020	2020-03	2020-03-22	3653437	1	1156.87MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
18	2020	2020-03	2020-03-21	3702528	1	1160.26MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
19	2020	2020-03	2020-03-20	4193968	1	1183.25MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
20	2020	2020-03	2020-03-19	3888061	1	1169.47MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
21	2020	2020-03	2020-03-18	3977800	1	1172.24MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
22	2020	2020-03	2020-03-17	3755175	1	1164.06MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
23	2020	2020-03	2020-03-16	3855346	1	1167.95MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
24	2020	2020-03	2020-03-15	3828048	1	1165.19MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
25	2020	2020-03	2020-03-14	3710593	1	1161.06MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
26	2020	2020-03	2020-03-13	3998934	1	1169.92MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof
27	2020	2020-03	2020-03-12	3988934	1	1172.62MB	NOT CACHED	NOT CACHED	RC_FILE	false	hdfs://ns1/user/hive/warehouse/baof

「COMPUTE STATS」的作用就是得出Impala原先不知道的值（-1） 并让 Impala 基于每一个表的大小、每一个列不同值的个数、等等信息自动的优化查询

5. 用 explain , summary 分析

```
[cdh81-39:21000] > SUMMARY
> ;
```

Operator	#Hosts	Avg Time	Max Time	#Rows	Est. #Rows	Peak Mem	Est. Peak Mem	Detail
11:EXCHANGE	1	8.47s	8.47s	9	324	0 B	0 B	UNPARTITIONED
06:HASH JOIN	3	19.75ms	37.79ms	9	324	10.10 MB	1.94 MB	LEFT OUTER JOIN, BROADCAST
--10:EXCHANGE	3	10.82us	11.83us	274	274	0 B	0 B	BROADCAST
03:SCAN HDFS	1	55.32ms	55.32ms	274	274	144.00 KB	32.00 MB	baofoo_cutpayment.cutpayment_error_code c
05:HASH JOIN	3	112.59ms	160.56ms	9	324	10.13 MB	1.94 MB	RIGHT OUTER JOIN, PARTITIONED
--04:HASH JOIN	3	49.96ms	58.90ms	9	324	2.16 MB	1.94 MB	RIGHT OUTER JOIN, PARTITIONED
--08:EXCHANGE	3	2.51s	2.96s	9	324	0 B	0 B	HASH(bo.business_no)
00:SCAN HDFS	2	1.96s	3.20s	9	324	42.62 MB	176.00 MB	baofoo_cutpayment.protocol_payment_order bo
07:EXCHANGE	3	1.45s	1.48s	4.22M	1.76M	0 B	0 B	HASH(be.business_no)
01:SCAN HDFS	3	394.65ms	591.18ms	4.22M	1.76M	58.12 MB	176.00 MB	baofoo_cutpayment.protocol_payment_business_orde
09:EXCHANGE	3	2.45s	2.46s	4.22M	4.22M	0 B	0 B	HASH(pp.business_no)
02:SCAN HDFS	4	1.66s	3.32s	4.22M	4.22M	66.52 MB	176.00 MB	baofoo_cutpayment.protocol_payment_business_orde

```
[cdh81-39:21000] >
```