# Spark SQL在不同存储格式下的性能对比

本文测试的目的是用来对比 Spark 三种存储格式txt、parquet、ya100的性能差异。

因机器环境以及配置的不同，测试结果可能略有差异，该测试报告仅对笔者的软硬件环境负责。

# 测试结论

## txt格式

太慢了,与后两者无可比性，故后面的对比仅针对parquet与ya100.

## 检索过滤性能

## 排序性能

### 1~2个列的TOP N排序

### 10~12个列的TOP N排序

## 统计性能

### 1个列的统计

### 1~2个列的Group by

### 10~12个列的统计与group by

## 高纬值列的性能

### 高纬值列统计与排序性能

### 高纬值group by性能

## 占用存储空间大小

## 导入速度对比

## 测试结果数据总览

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 测试指标 | | 测试方法 | | 12亿条测试数据 | | |
| txt | parquet | ya100 |
| 检索过滤性能 (只返回少量行数的数据) | | count(\*) | 无过滤条件 | 377 | 17.7 | 2.6 |
| 等值匹配 | 492 | 70 | 2.5 |
| range范围查找 | 504 | 41 | 1.4 |
| 模糊匹配(全文检索) | 473 | 43 | 13.9 |
| 多条件组合 | 844 | 111 | 30 |
| 检索后返回明细 | 等值匹配 | 1952 | 484 | 1.7 |
| range范围查找 | 1903 | 477 | 10 |
| 模糊匹配(全文检索) | 1930 | 529 | 20.7 |
| 多条件组合 | 1995 | 521 | 54 |
| 统计性能 | 10~12个 低纬值列 | 小范围 | 只sum,avg,max,min统计 | 1656 | 199 | 13.3 |
| Top n排序 | 1807 | 290 | 5.5 |
| group by加max与sum |  | 238 | 9.1 |
| 中等范围 | 11个中低等维值列sum,avg,max,min统计 | 1653 | 181 | 52.5 |
| 较多列值的 Top n排序 | 1858 | 240 | 13.6 |
| 3个列group by ,8个列max与sum | 1709 | 186 | 76 |
| 全表扫描 | 11个中低等维值列sum,avg,max,min统计 | 1878 | 594 | 116 |
| 返回较多列值的 Top n排序 | 1647 | 448 | 70.2 |
| 3个列group by ,8个列max与sum | 3323 | 2102 | 331 |
| 1~2个 低纬值列 | 小范围 | amtlong的sum,avg,max,min统计 | 489 | 38 | 9.7 |
| ydb\_sex的max,min统计 | 599 | 50 | 9.2 |
| ydb\_province的max,min统计 | 592 | 52 | 7.9 |
| ydb\_age Top n排序 | 716 | 66.6 | 9.1 |
| ydb\_age,ydb\_sex Top n排序 | 706 | 65 | 7 |
| ydb\_sex的单列group by count(\*) | 607 | 55 | 8,7 |
| ydb\_province的单列group by max(ydb\_age) | 737 | 73 | 10.5 |
| ydb\_sex,ydb\_province的多列group by count(\*)与排序 | 717 | 72 | 6.6 |
| ydb\_sex,ydb\_province与低维值列amtlong的多列group by sum | 723 | 96 | 8.8 |
| 中等范围 | amtlong的sum,avg,max,min统计 | 378 | 17 | 20 |
| ydb\_sex的max,min统计 | 485 | 30 | 3.86 |
| ydb\_province的max,min统计 | 490 | 33.6 | 8.5 |
| ydb\_age Top n排序 | 585 | 46.9 | 6.9 |
| ydb\_age,ydb\_sex Top n排序 | 595 | 46 | 4.4 |
| ydb\_sex的单列group by count(\*) | 495 | 36 | 7 |
| ydb\_province的单列group by max(ydb\_age) | 608 | 53 | 15 |
| ydb\_sex,ydb\_province的多列group by count(\*)与排序 | 728 | 53 | 13 |
| ydb\_sex,ydb\_province与低维值列amtlong的多列group by sum | 718 | 65 | 22 |
| 全表扫描 | amtlong的sum,avg,max,min统计 | 405 | 49.7 | 33.7 |
| ydb\_sex的max,min统计 | 473 | 72 | 10 |
| ydb\_province的max,min统计 | 472 | 70.7 | 13.4 |
| ydb\_age Top n排序 | 544 | 85.3 | 30 |
| ydb\_age,ydb\_sex Top n排序 | 538 | 82 | 63 |
| ydb\_sex的单列group by count(\*) | 473 | 54.6 | 12.7 |
| ydb\_province的单列group by max(ydb\_age) | 789 | 555 | 32.8 |
| ydb\_sex,ydb\_province的多列group by count(\*)与排序 | 565 | 83 | 54 |
| ydb\_sex,ydb\_province与低维值列amtlong的多列group by sum | 722 | 94 | 66 |
| 1~2个 高纬值列 | 小范围 | amtdouble的sum,avg,max,min统计 | 491 | 36 | 9.5 |
| usernick的max,min统计 | 606 | 84 | 9.5 |
| phonenum的max,min统计 | 511 | 61 | 8.7 |
| phonenum, usernick Top n排序 | 641 | 101 | 6.3 |
| usernick的单列group by count(\*) | 731 | 82 | 10 |
| phonenum的单列group by count(\*) | 518 | 60 | 9.3 |
| ydb\_sex,ydb\_province与高纬值列amtdouble的多列group by sum | 717 | 72 | 10.2 |
|  |  |  |  |
| 中等范围 | amtdouble的sum,avg,max,min统计 | 491 | 38.9 | 54 |
| usernick的max,min统计 | 485 | 56 | 16.8 |
| phonenum的max,min统计 | 405 | 39 | 32 |
| phonenum, usernick Top n排序 | 521 | 78 | 158 |
| usernick的单列group by count(\*) | 495 | 59 | 25 |
| phonenum的单列group by count(\*) | 403 | 39 | 35 |
| ydb\_sex,ydb\_province与高纬值列amtdouble的多列group by sum | 881 | 93 | 66 |
| 全表扫描 | amtdouble的sum,avg,max,min统计 | 495 | 58 | 61 |
| usernick的max,min统计 | 476 | 87 | 30.4 |
| phonenum的max,min统计 | 388 | 37.8 | 41.2 |
| phonenum, usernick Top n排序 | 505 | 103 | 172 |
| usernick的单列group by count(\*) | 556 | 153 | 219 |
| phonenum的单列group by count(\*) | 585 | 751 | 1000 |
| ydb\_sex,ydb\_province与高纬值列amtdouble的多列group by sum | 875 | 103 | 93 |
| 存储空间大小 | |  |  | 184218770684.00 | 46813665946.00 | 107042091971.00 |
| 导入性能 | |  |  | - | 4825 | 29749 |

Spark SQL在不同存储格式下的性能对比 1

一、 测试结论 1

(一) txt格式 1

(二) 检索过滤性能 1

(三) 排序性能 2

(四) 统计性能 3

(五) 高纬值列的性能 6

(六) 占用存储空间大小 8

(七) 导入速度对比 8

(八) 测试结果数据总览 9

二、 数据以及硬件条件 16

1. 数据条数: 16

2. 硬件条件 16

3. 软件版本 16

4. spark启动 16

5. 配置ya100的必备函数 16

6. 配置parquet的按照参数（spark没有像hive那样对第三方存储开放此参数，故ya100只能通过set方式传参） 17

三、 数据表结构 17

四、 性能对比-导入速度 18

五、 检索过滤性能测试 21

1. count(\*) 21

2. 等值匹配 23

3. 范围查找 24

4. 模糊匹配 26

5. 多条件组合count(\*) 28

6. 等值匹配+返回15列的 order by phonenum desc 30

7. 范围查找+返回15列的order by phonenum desc 32

8. 模糊匹配+返回15列的order by phonenum desc 38

9. 多条件组合+返回15列的order by phonenum desc 44

统计性能测试-小范围统计 51

1. 中等纬值列amtlong的sum,avg,max,min统计 51

2. 高纬值列amtdouble的sum,avg,max,min统计 53

3. 低纬值列ydb\_sex的max,min统计 55

4. 低纬值列ydb\_province的max,min统计 57

5. 高纬值列usernick的max,min统计 59

6. 高纬值列phonenum的max,min统计 61

7. 11个中低等维值列sum,avg,max,min统计 63

8. 低维值列ydb\_age Top n排序 66

9. 低维值列ydb\_age,ydb\_sex Top n排序 69

10. 高纬值列phonenum, usernick Top n排序 72

11. 返回较多列值的 Top n排序 75

12. 低维值列ydb\_sex的单列group by count(\*) 81

13. 低维值列ydb\_province的单列group by max(ydb\_age) 83

14. 高维值列usernick的单列group by count(\*) 86

15. 高维值列phonenum的单列group by count(\*) 90

16. 低维值列ydb\_sex,ydb\_province的多列group by count(\*)与排序 92

17. 低维值列ydb\_sex,ydb\_province与低维值列amtlong的多列group by sum 95

18. 低维值列ydb\_sex,ydb\_province与高纬值列amtdouble的多列group by sum 98

19. 3个列group by ,8个列max与sum 101

统计性能测试-中等范围统计 104

1. 中等纬值列amtlong的sum,avg,max,min统计 105

2. 高纬值列amtdouble的sum,avg,max,min统计 107

3. 低纬值列ydb\_sex的max,min统计 109

4. 低纬值列ydb\_province的max,min统计 110

5. 高纬值列usernick的max,min统计 112

6. 高纬值列phonenum的max,min统计 114

7. 11个中低等维值列sum,avg,max,min统计 116

8. 低维值列ydb\_age Top n排序 120

9. 低维值列ydb\_age,ydb\_sex Top n排序 123

10. 高纬值列phonenum, usernick Top n排序 127

11. 返回较多列值的 Top n排序 129

12. 低维值列ydb\_sex的单列group by count(\*) 135

13. 低维值列ydb\_province的单列group by max(ydb\_age) 137

14. 高维值列usernick的单列group by count(\*) 140

15. 高维值列phonenum的单列group by count(\*) 144

16. 低维值列ydb\_sex,ydb\_province的多列group by count(\*)与排序 146

17. 低维值列ydb\_sex,ydb\_province与低维值列amtlong的多列group by sum 150

18. 低维值列ydb\_sex,ydb\_province与高纬值列amtdouble的多列group by sum 152

19. 3个列group by ,8个列max与sum 155

统计性能测试-全表统计 162

1. 中等纬值列amtlong的sum,avg,max,min统计 162

2. 高纬值列amtdouble的sum,avg,max,min统计 164

3. 低纬值列ydb\_sex的max,min统计 166

4. 低纬值列ydb\_province的max,min统计 167

5. 高纬值列usernick的max,min统计 169

6. 高纬值列phonenum的max,min统计 171

7. 11个中低等维值列sum,avg,max,min统计 173

8. 低维值列ydb\_age Top n排序 177

9. 低维值列ydb\_age,ydb\_sex Top n排序 180

10. 高纬值列phonenum, usernick Top n排序 183

11. 返回较多列值的 Top n排序 186

12. 低维值列ydb\_sex的单列group by count(\*) 192

13. 低维值列ydb\_province的单列group by max(ydb\_age) 194

14. 高维值列usernick的单列group by count(\*) 197

15. 高维值列phonenum的单列group by count(\*) 201

16. 低维值列ydb\_sex,ydb\_province的多列group by count(\*)与排序 204

17. 低维值列ydb\_sex,ydb\_province与低维值列amtlong的多列group by sum 207

18. 低维值列ydb\_sex,ydb\_province与高纬值列amtdouble的多列group by sum 210

19. 3个列group by ,8个列max与sum 213

# 数据以及硬件条件

## 数据条数:

1亿对应分区名为dt='100million' 原始数据大小为13G

12亿对应分区名为dt='1200million' 原始数据大小为158G

测试数据为伪造数据，伪造数据的程序源码放在example目录下，文件名称为：MrMakeShuData.java

## 硬件条件

一共4台阿里云虚拟机,每台的磁盘均是100G\*3

其中 3台 CPU 1核，内存8G

其中1台 CPU 2核，内存16G

## 软件版本

hadoop 2.6.2

spark 1.6

ya100 1.0.8

## spark启动

./start-thriftserver.sh --master yarn-client --executor-memory 990m --executor-cores 2 --num-executors 16

./beeline -u jdbc:hive2://localhost:10000

## 配置ya100的必备函数

create function Yfilter as 'cn.net.ycloud.ydb.handle.fun.Yfilter';

create function Ytop10000 as 'cn.net.ycloud.ydb.handle.fun.Ytop10000';

create function Ycombine as 'cn.net.ycloud.ydb.handle.fun.Ycombine';

create function Ycount as 'cn.net.ycloud.ydb.handle.fun.Ycount';

create function Ymax as 'cn.net.ycloud.ydb.handle.fun.Ymax';

create function Ymin as 'cn.net.ycloud.ydb.handle.fun.Ymin';

create function Yavg as 'cn.net.ycloud.ydb.handle.fun.Yavg';

create function Ysum as 'cn.net.ycloud.ydb.handle.fun.Ysum';

create function Ymaxstring as 'cn.net.ycloud.ydb.handle.fun.Ymaxstring';

create function Yminstring as 'cn.net.ycloud.ydb.handle.fun.Yminstring';

## 配置parquet的按照参数（spark没有像hive那样对第三方存储开放此参数，故ya100只能通过set方式传参）

set spark.sql.parquet.filterPushdown=true;

该参数1.5版本以后默认开启，其他版本请设置，否则影响检索与过滤的速度

# 数据表结构

|  |  |
| --- | --- |
| **type** | **memo** |
| **text** | CREATE external table spark\_txt(  phonenum bigint, usernick string, ydb\_sex string, ydb\_province string, ydb\_grade string, ydb\_age string, ydb\_blood string, ydb\_zhiye string, ydb\_earn string, ydb\_prefer string, ydb\_consume string, ydb\_day string, amtdouble double,amtlong int,content string  ) partitioned by (dt string) row format delimited fields terminated by ',' location '/data/ydb/shu'; |
| **paraquet** | CREATE external TABLE spark\_parquet (  phonenum bigint, usernick string, ydb\_sex string, ydb\_province string, ydb\_grade string, ydb\_age string, ydb\_blood string, ydb\_zhiye string, ydb\_earn string, ydb\_prefer string, ydb\_consume string, ydb\_day string, amtdouble double,amtlong int,content string  ) partitioned by (dt string) STORED AS PARQUET  location '/data/ydb/shu\_parquet'; |
| **ya100** | **//注，**ydbpartion为ydb的分区字段与ya100\_pipe 为ya100的管道字段，必须要都要创建，否则有些功能会被限制使用。  CREATE external table spark\_ya100(  phonenum bigint, usernick string, ydb\_sex string, ydb\_province string, ydb\_grade string, ydb\_age string, ydb\_blood string, ydb\_zhiye string, ydb\_earn string, ydb\_prefer string, ydb\_consume string, ydb\_day string, amtdouble double,amtlong int, content string, ydbpartion string, ya100\_pipe string  ) partitioned by (dt string)  STORED BY 'cn.net.ycloud.ydb.handle.Ya100StorageHandler'  LOCATION '/data/ydb/shu\_ya100'  TBLPROPERTIES(  "ya100.handler.table.name"="ydb\_example\_shu",  "ya100.handler.schema"="phonenum long,usernick string,ydb\_sex string,ydb\_province string,ydb\_grade string,ydb\_age string,ydb\_blood string,ydb\_zhiye string,ydb\_earn string,ydb\_prefer string,ydb\_consume string,ydb\_day string, amtdouble double,amtlong int,content textcjk,ydbpartion string,ya100\_pipe string"  ) |
| **备用语句** |  |

# 性能对比-导入速度

|  |  |
| --- | --- |
| **type** | **memo** |
| **txt** | load data inpath '/data/ydb/shu\_bak' into table spark\_txt partition (dt='100million');  load data inpath '/data/ydb/shu\_bak\_bak1' into table spark\_txt partition (dt='1200million'); |
| **paraquet** | insert into table spark\_parquet partition (dt='100million') select phonenum,usernick,ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtdouble,amtlong,content from spark\_txt where dt='100million';    insert into table spark\_parquet partition (dt='1200million') select phonenum,usernick,ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtdouble,amtlong,content from spark\_txt where dt='1200million'; |
| **ya100** | ya100因为生成的最终不是一个文件，而是一个目录，这个目录下会有索引，所以要通设置hive.mapred.supports.subdirectories 让spark支持子目录，否则向表里导入数据会报错失败。  生成的索引文件个数最好不要太多，每个executer能处理1~8个为佳，不要生成太多数量的索引。  set hive.mapred.supports.subdirectories=true;  set mapred.min.split.size=536870912;  insert into table spark\_ya100 partition (dt='100million') select phonenum,usernick,ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtdouble,amtlong,content,2,1 from spark\_txt where dt='100million';    set hive.mapred.supports.subdirectories=true;  set mapred.min.split.size=2147483648;  insert into table spark\_ya100 partition (dt='1200million') select phonenum,usernick,ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtdouble,amtlong,content,2,1 from spark\_txt where dt='1200million'; |

# 检索过滤性能测试

## count(\*)

|  |  |
| --- | --- |
| **type** | **memo** |
| **txt** | select count(\*) from spark\_txt where dt='100million' limit 10;    select count(\*) from spark\_txt where dt='1200million' limit 10; |
| **parquet** | select count(\*) from spark\_parquet where dt='100million' limit 10    select count(\*) from spark\_parquet where dt='1200million' limit 10 |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu=\*;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('\*',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('\*',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 等值匹配

|  |  |
| --- | --- |
| **type** | **memo** |
| **txt** | select count(\*) from spark\_txt where dt='100million' and phonenum=' 13870282671' and usernick='江峻熙' limit 10;    select count(\*) from spark\_txt where dt='1200million' and phonenum=' 13870282671' and usernick='江峻熙' limit 10; |
| **parquet** | select count(\*) from spark\_parquet where dt='100million' and phonenum=' 13870282671' and usernick='江峻熙' limit 10;    select count(\*) from spark\_parquet where dt='1200million' and phonenum=' 13870282671' and usernick='江峻熙' limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=phonenum='13870282671' and usernick='江峻熙';  set ya100.spark.combine.ydb\_example\_shu=\*;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('\*',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('\*',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 范围查找

|  |  |
| --- | --- |
| **type** | **memo** |
| **txt** | select count(\*) from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10;    select count(\*) from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10; |
| **parquet** | select count(\*) from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10;    select count(\*) from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu=\*;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('\*',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('\*',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 模糊匹配

|  |  |
| --- | --- |
| **type** | **memo** |
| **txt** | select count(\*) from spark\_txt where dt='100million' and content like '%王老吉%' limit 10;    select count(\*) from spark\_txt where dt='1200million' and content like '%王老吉%' limit 10; |
| **parquet** | select count(\*) from spark\_parquet where dt='100million' and content like '%王老吉%' limit 10;    select count(\*) from spark\_parquet where dt='1200million' and content like '%王老吉%' limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= content='王老吉';  set ya100.spark.combine.ydb\_example\_shu=\*;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('\*',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('\*',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 多条件组合count(\*)

|  |  |
| --- | --- |
| **type** | **memo** |
| **txt** | select count(\*) from spark\_txt where dt='100million' and ydb\_sex='女' and ydb\_grade='本科' and (ydb\_age='20到30岁' or ydb\_blood='O') and (amtlong>=3000 and amtlong<=4000) and ydb\_day like '2015111%' limit 10;    select count(\*) from spark\_txt where dt='1200million' and ydb\_sex='女' and ydb\_grade='本科' and (ydb\_age='20到30岁' or ydb\_blood='O') and (amtlong>=3000 and amtlong<=4000) and ydb\_day like '2015111%' limit 10; |
| **parquet** | select count(\*) from spark\_parquet where dt='100million' and ydb\_sex='女' and ydb\_grade='本科' and (ydb\_age='20到30岁' or ydb\_blood='O') and (amtlong>=3000 and amtlong<=4000) and ydb\_day like '2015111%' limit 10;    select count(\*) from spark\_parquet where dt='1200million' and ydb\_sex='女' and ydb\_grade='本科' and (ydb\_age='20到30岁' or ydb\_blood='O') and (amtlong>=3000 and amtlong<=4000) and ydb\_day like '2015111%' limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= ydb\_sex='女' and ydb\_grade='本科' and (ydb\_age='20到30岁' or ydb\_blood='O') and (amtlong like '([3000 TO 4000] )') and ydb\_day like '2015111%';  set ya100.spark.combine.ydb\_example\_shu=\*;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('\*',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('\*',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 等值匹配+返回15列的 order by phonenum desc

|  |  |
| --- | --- |
| **type** | **memo** |
| **txt** | select \* from spark\_txt where dt='100million' and phonenum='13870282671' and usernick='江峻熙' order by phonenum desc limit 10;    select \* from spark\_txt where dt='1200million' and phonenum='13870282671' and usernick='江峻熙' order by phonenum desc limit 10; |
| **parquet** | select \* from spark\_parquet where dt='100million' and phonenum='13870282671' and usernick='江峻熙' order by phonenum desc limit 10;    select \* from spark\_parquet where dt='1200million' and phonenum='13870282671' and usernick='江峻熙' order by phonenum desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=phonenum='13870282671' and usernick='江峻熙';  set ya100.spark.combine.ydb\_example\_shu=;  set ya100.spark.top10000.ydb\_example\_shu=phonenum desc limit 10;  select \* from spark\_ya100 where dt='100million' order by phonenum desc limit 10;    select \* from spark\_ya100 where dt='1200million' order by phonenum desc limit 10; |

## 范围查找+返回15列的order by phonenum desc

|  |  |
| --- | --- |
| **type** | **memo** |
| **txt** | select \* from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 order by phonenum desc limit 10;    select \* from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 order by phonenum desc limit 10; |
| **parquet** | select \* from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 order by phonenum desc limit 10;    select \* from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 order by phonenum desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu=;  set ya100.spark.top10000.ydb\_example\_shu=phonenum desc limit 10;  select \* from spark\_ya100 where dt='100million' order by phonenum desc limit 10;    select \* from spark\_ya100 where dt='1200million' order by phonenum desc limit 10; |

## 模糊匹配+返回15列的order by phonenum desc

|  |  |
| --- | --- |
| **type** | **memo** |
| **txt** | select \* from spark\_txt where dt='100million' and content like '%王老吉%' and ydb\_day like '2015111%' and ydb\_province like '辽%' order by phonenum desc limit 10;    select \* from spark\_txt where dt='1200million' and content like '%王老吉%' and ydb\_day like '2015111%' and ydb\_province like '辽%' order by phonenum desc limit 10; |
| **parquet** | select \* from spark\_parquet where dt='100million' and content like '%王老吉%' and ydb\_day like '2015111%' and ydb\_province like '辽%' order by phonenum desc limit 10;    select \* from spark\_parquet where dt='1200million' and content like '%王老吉%' and ydb\_day like '2015111%' and ydb\_province like '辽%' order by phonenum desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= content='王老吉' and ydb\_day like '2015111%' and ydb\_province like '辽%';  set ya100.spark.combine.ydb\_example\_shu=;  set ya100.spark.top10000.ydb\_example\_shu=phonenum desc limit 10;  select \* from spark\_ya100 where dt='100million' order by phonenum desc limit 10;    select \* from spark\_ya100 where dt='1200million' order by phonenum desc limit 10; |

## 多条件组合+返回15列的order by phonenum desc

|  |  |
| --- | --- |
| **type** | **memo** |
| **txt** | select \* from spark\_txt where dt='100million' and ydb\_sex='女' and ydb\_grade='本科' and (ydb\_age='20到30岁' or ydb\_blood='O') and (amtlong>=3000 and amtlong<=4000) and ydb\_day like '2015111%' and ydb\_province like '辽%' order by phonenum desc limit 10;    select \* from spark\_txt where dt='1200million' and ydb\_sex='女' and ydb\_grade='本科' and (ydb\_age='20到30岁' or ydb\_blood='O') and (amtlong>=3000 and amtlong<=4000) and ydb\_day like '2015111%' and ydb\_province like '辽%' order by phonenum desc limit 10; |
| **parquet** | select \* from spark\_parquet where dt='100million' and ydb\_sex='女' and ydb\_grade='本科' and (ydb\_age='20到30岁' or ydb\_blood='O') and (amtlong>=3000 and amtlong<=4000) and ydb\_day like '2015111%' and ydb\_province like '辽%' order by phonenum desc limit 10;    select \* from spark\_parquet where dt='1200million' and ydb\_sex='女' and ydb\_grade='本科' and (ydb\_age='20到30岁' or ydb\_blood='O') and (amtlong>=3000 and amtlong<=4000) and ydb\_day like '2015111%' and ydb\_province like '辽%' order by phonenum desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= ydb\_sex='女' and ydb\_grade='本科' and (ydb\_age='20到30岁' or ydb\_blood='O') and (amtlong like '([3000 TO 4000] )') and ydb\_day like '2015111%' and ydb\_province like '辽%';  set ya100.spark.combine.ydb\_example\_shu=;  set ya100.spark.top10000.ydb\_example\_shu=phonenum desc limit 10;  select \* from spark\_ya100 where dt='100million' order by phonenum desc limit 10;    select \* from spark\_ya100 where dt='1200million' order by phonenum desc limit 10; |

# 统计性能测试-小范围统计

## 中等纬值列amtlong的sum,avg,max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select count(amtlong),sum(amtlong) ,avg(amtlong),max(amtlong),min(amtlong) from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10;    select count(amtlong),sum(amtlong) ,avg(amtlong),max(amtlong),min(amtlong) from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10; |
| **parquet** | select count(amtlong),sum(amtlong) ,avg(amtlong),max(amtlong),min(amtlong) from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10;    select count(amtlong),sum(amtlong) ,avg(amtlong),max(amtlong),min(amtlong) from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu=amtlong;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('amtlong',ya100\_pipe) ,Ysum('amtlong',ya100\_pipe) ,Yavg('amtlong',ya100\_pipe), Ymax('amtlong',ya100\_pipe), Ymin('amtlong',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('amtlong',ya100\_pipe) ,Ysum('amtlong',ya100\_pipe) ,Yavg('amtlong',ya100\_pipe), Ymax('amtlong',ya100\_pipe), Ymin('amtlong',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 高纬值列amtdouble的sum,avg,max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select count(amtdouble),sum(amtdouble) ,avg(amtdouble),max(amtdouble),min(amtdouble) from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10;    select count(amtdouble),sum(amtdouble) ,avg(amtdouble),max(amtdouble),min(amtdouble) from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10; |
| **parquet** | select count(amtdouble),sum(amtdouble) ,avg(amtdouble),max(amtdouble),min(amtdouble) from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10;    select count(amtdouble),sum(amtdouble) ,avg(amtdouble),max(amtdouble),min(amtdouble) from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu=amtdouble;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('amtdouble',ya100\_pipe) ,Ysum('amtdouble',ya100\_pipe) ,Yavg('amtdouble',ya100\_pipe), Ymax('amtdouble',ya100\_pipe), Ymin('amtdouble',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('amtdouble',ya100\_pipe) ,Ysum('amtdouble',ya100\_pipe) ,Yavg('amtdouble',ya100\_pipe), Ymax('amtdouble',ya100\_pipe), Ymin('amtdouble',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 低纬值列ydb\_sex的max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select count(ydb\_sex),max(ydb\_sex),min(ydb\_sex) from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10;    select count(ydb\_sex),max(ydb\_sex),min(ydb\_sex) from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10; |
| **parquet** | select count(ydb\_sex),max(ydb\_sex),min(ydb\_sex) from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10;    select count(ydb\_sex),max(ydb\_sex),min(ydb\_sex) from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu=ydb\_sex;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('ydb\_sex',ya100\_pipe), Ymaxstring('ydb\_sex',ya100\_pipe), Yminstring('ydb\_sex',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('ydb\_sex',ya100\_pipe), Ymaxstring('ydb\_sex',ya100\_pipe), Yminstring('ydb\_sex',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 低纬值列ydb\_province的max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select count(ydb\_province),max(ydb\_province),min(ydb\_province) from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10;    select count(ydb\_province),max(ydb\_province),min(ydb\_province) from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10; |
| **parquet** | select count(ydb\_province),max(ydb\_province),min(ydb\_province) from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10;    select count(ydb\_province),max(ydb\_province),min(ydb\_province) from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu=ydb\_province;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('ydb\_province',ya100\_pipe), Ymaxstring('ydb\_province',ya100\_pipe), Yminstring('ydb\_province',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('ydb\_province',ya100\_pipe), Ymaxstring('ydb\_province',ya100\_pipe), Yminstring('ydb\_province',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 高纬值列usernick的max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select count(usernick), max(usernick),min(usernick) from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10;    select count(usernick), max(usernick),min(usernick) from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10; |
| **parquet** | select count(usernick), max(usernick),min(usernick) from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10;    select count(usernick), max(usernick),min(usernick) from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu= usernick;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('usernick',ya100\_pipe), Ymaxstring('usernick',ya100\_pipe), Yminstring('usernick',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('usernick',ya100\_pipe), Ymaxstring('usernick',ya100\_pipe), Yminstring('usernick',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 高纬值列phonenum的max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select max(phonenum),min(phonenum) from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10;    select max(phonenum),min(phonenum) from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10; |
| **parquet** | select max(phonenum),min(phonenum) from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10;    select max(phonenum),min(phonenum) from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu=phonenum;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ymax('phonenum',ya100\_pipe), Ymin('phonenum',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ymax('phonenum',ya100\_pipe), Ymin('phonenum',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 11个中低等维值列sum,avg,max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select count(ydb\_sex),count(ydb\_province),count(ydb\_grade),count(ydb\_age),count(ydb\_blood),count(ydb\_zhiye),count(ydb\_earn),count(ydb\_prefer),count(ydb\_consume),count(ydb\_day),count(amtlong), max(ydb\_sex),max(ydb\_province),max(ydb\_grade),max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),max(amtlong) from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10;    **select count(ydb\_sex),count(ydb\_province),count(ydb\_grade),count(ydb\_age),count(ydb\_blood),count(ydb\_zhiye),count(ydb\_earn),count(ydb\_prefer),count(ydb\_consume),count(ydb\_day),count(amtlong), max(ydb\_sex),max(ydb\_province),max(ydb\_grade),max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),max(amtlong) from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10;** |
| **parquet** | select count(ydb\_sex),count(ydb\_province),count(ydb\_grade),count(ydb\_age),count(ydb\_blood),count(ydb\_zhiye),count(ydb\_earn),count(ydb\_prefer),count(ydb\_consume),count(ydb\_day),count(amtlong), max(ydb\_sex),max(ydb\_province),max(ydb\_grade),max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),max(amtlong) from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10;    select count(ydb\_sex),count(ydb\_province),count(ydb\_grade),count(ydb\_age),count(ydb\_blood),count(ydb\_zhiye),count(ydb\_earn),count(ydb\_prefer),count(ydb\_consume),count(ydb\_day),count(amtlong), max(ydb\_sex),max(ydb\_province),max(ydb\_grade),max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),max(amtlong) from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu=ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('ydb\_sex',ya100\_pipe),Ycount('ydb\_province',ya100\_pipe),Ycount('ydb\_grade',ya100\_pipe),Ycount('ydb\_age',ya100\_pipe),Ycount('ydb\_blood',ya100\_pipe),Ycount('ydb\_zhiye',ya100\_pipe),Ycount('ydb\_earn',ya100\_pipe),Ycount('ydb\_prefer',ya100\_pipe),Ycount('ydb\_consume',ya100\_pipe),Ycount('ydb\_day',ya100\_pipe),Ycount('amtlong',ya100\_pipe), Ymaxstring('ydb\_sex',ya100\_pipe),Ymaxstring('ydb\_province',ya100\_pipe),Ymaxstring('ydb\_grade',ya100\_pipe),Ymaxstring('ydb\_age',ya100\_pipe),Ymaxstring('ydb\_blood',ya100\_pipe),Ymaxstring('ydb\_zhiye',ya100\_pipe),Ymaxstring('ydb\_earn',ya100\_pipe),Ymaxstring('ydb\_prefer',ya100\_pipe),Ymaxstring('ydb\_consume',ya100\_pipe),Ymaxstring('ydb\_day',ya100\_pipe),Ymax('amtlong',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('ydb\_sex',ya100\_pipe),Ycount('ydb\_province',ya100\_pipe),Ycount('ydb\_grade',ya100\_pipe),Ycount('ydb\_age',ya100\_pipe),Ycount('ydb\_blood',ya100\_pipe),Ycount('ydb\_zhiye',ya100\_pipe),Ycount('ydb\_earn',ya100\_pipe),Ycount('ydb\_prefer',ya100\_pipe),Ycount('ydb\_consume',ya100\_pipe),Ycount('ydb\_day',ya100\_pipe),Ycount('amtlong',ya100\_pipe), Ymaxstring('ydb\_sex',ya100\_pipe),Ymaxstring('ydb\_province',ya100\_pipe),Ymaxstring('ydb\_grade',ya100\_pipe),Ymaxstring('ydb\_age',ya100\_pipe),Ymaxstring('ydb\_blood',ya100\_pipe),Ymaxstring('ydb\_zhiye',ya100\_pipe),Ymaxstring('ydb\_earn',ya100\_pipe),Ymaxstring('ydb\_prefer',ya100\_pipe),Ymaxstring('ydb\_consume',ya100\_pipe),Ymaxstring('ydb\_day',ya100\_pipe),Ymax('amtlong',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 低维值列ydb\_age Top n排序

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select ydb\_sex, ydb\_age from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 order by ydb\_age desc limit 10;    select ydb\_sex, ydb\_age from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 order by ydb\_age desc limit 10; |
| **parquet** | select ydb\_sex, ydb\_age from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 order by ydb\_age desc limit 10;    select ydb\_sex, ydb\_age from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 order by ydb\_age desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu=;  set ya100.spark.top10000.ydb\_example\_shu=ydb\_age desc;  select ydb\_sex, ydb\_age from spark\_ya100 where dt='100million' order by ydb\_age desc limit 10;    select ydb\_sex, ydb\_age from spark\_ya100 where dt='1200million' order by ydb\_age desc limit 10; |

## 低维值列ydb\_age,ydb\_sex Top n排序

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select ydb\_sex, ydb\_age from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 order by ydb\_age desc, ydb\_sex limit 10;    select ydb\_sex, ydb\_age from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 order by ydb\_age desc, ydb\_sex limit 10; |
| **parquet** | select ydb\_sex, ydb\_age from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 order by ydb\_age desc, ydb\_sex limit 10;    select ydb\_sex, ydb\_age from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 order by ydb\_age desc, ydb\_sex limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu=;  set ya100.spark.top10000.ydb\_example\_shu= ydb\_age desc, ydb\_sex;  select ydb\_sex, ydb\_age from spark\_ya100 where dt='100million' order by ydb\_age desc, ydb\_sex limit 10;    select ydb\_sex, ydb\_age from spark\_ya100 where dt='1200million' order by ydb\_age desc, ydb\_sex limit 10; |

## 高纬值列phonenum, usernick Top n排序

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select phonenum,usernick from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 order by phonenum desc limit 10;    select phonenum,usernick from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 order by phonenum desc limit 10; |
| **parquet** | select phonenum,usernick from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 order by phonenum desc limit 10;    select phonenum,usernick from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 order by phonenum desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu=;  set ya100.spark.top10000.ydb\_example\_shu=phonenum desc;  select phonenum,usernick from spark\_ya100 where dt='100million' order by phonenum desc limit 10;    select phonenum,usernick from spark\_ya100 where dt='1200million' order by phonenum desc limit 10; |

## 返回较多列值的 Top n排序

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select phonenum,usernick, ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 order by ydb\_age desc, ydb\_sex limit 10;    select phonenum,usernick, ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 order by ydb\_age desc, ydb\_sex limit 10; |
| **parquet** | select phonenum,usernick, ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 order by ydb\_age desc, ydb\_sex limit 10;    select phonenum,usernick, ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 order by ydb\_age desc, ydb\_sex limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu=;  set ya100.spark.top10000.ydb\_example\_shu= ydb\_age desc,ydb\_sex limit 10;  select phonenum,usernick, ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong from spark\_ya100 where dt='100million' order by ydb\_age desc,ydb\_sex limit 10;    select phonenum,usernick, ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong from spark\_ya100 where dt='1200million' order by ydb\_age desc,ydb\_sex limit 10; |

## 低维值列ydb\_sex的单列group by count(\*)

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select ydb\_sex,count(\*) from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_sex limit 10;    select ydb\_sex,count(\*) from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_sex limit 10; |
| **parquet** | select ydb\_sex,count(\*) from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_sex limit 10;    select ydb\_sex,count(\*) from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_sex limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu=\*;  set ya100.spark.top10000.ydb\_example\_shu=;  select ydb\_sex,Ycount('\*',ya100\_pipe) from spark\_ya100 where dt='100million' group by ydb\_sex limit 10;    select ydb\_sex,Ycount('\*',ya100\_pipe) from spark\_ya100 where dt='1200million' group by ydb\_sex limit 10; |

## 低维值列ydb\_province的单列group by max(ydb\_age)

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select ydb\_province,count(\*),count(ydb\_age) ,max(ydb\_age) ,min(ydb\_age) from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_province order by ydb\_province limit 10;    select ydb\_province,count(\*),count(ydb\_age) ,max(ydb\_age) ,min(ydb\_age) from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_province order by ydb\_province limit 10; |
| **parquet** | select ydb\_province,count(\*),count(ydb\_age) ,max(ydb\_age) ,min(ydb\_age) from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_province order by ydb\_province limit 10;    select ydb\_province,count(\*),count(ydb\_age) ,max(ydb\_age) ,min(ydb\_age) from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_province order by ydb\_province limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu=\*,ydb\_age;  set ya100.spark.top10000.ydb\_example\_shu=;  select ydb\_province,Ycount('\*',ya100\_pipe), Ycount('ydb\_age',ya100\_pipe), Ymaxstring('ydb\_age',ya100\_pipe), Yminstring('ydb\_age',ya100\_pipe)  from spark\_ya100 where dt='100million' group by ydb\_province order by ydb\_province limit 10;    select ydb\_province,Ycount('\*',ya100\_pipe), Ycount('ydb\_age',ya100\_pipe), Ymaxstring('ydb\_age',ya100\_pipe), Yminstring('ydb\_age',ya100\_pipe)  from spark\_ya100 where dt='1200million' group by ydb\_province order by ydb\_province limit 10; |

## 高维值列usernick的单列group by count(\*)

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select usernick,count(\*) as cnt from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by usernick order by cnt desc limit 10;    select usernick,count(\*) as cnt from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by usernick order by cnt desc limit 10; |
| **spark on parque** | select usernick,count(\*) as cnt from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by usernick order by cnt desc limit 10;    select usernick,count(\*) as cnt from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by usernick order by cnt desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu=\*;  set ya100.spark.top10000.ydb\_example\_shu=;  select usernick,Ycount('\*',ya100\_pipe) as cnt from spark\_ya100 where dt='100million' group by usernick order by cnt desc limit 10;    select usernick,Ycount('\*',ya100\_pipe) as cnt from spark\_ya100 where dt='1200million' group by usernick order by cnt desc limit 10; |

## 高维值列phonenum的单列group by count(\*)

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select phonenum,count(\*) as cnt from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by phonenum limit 10;    select phonenum,count(\*) as cnt from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by phonenum limit 10; |
| **spark on parqueut** | select phonenum,count(\*) as cnt from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by phonenum limit 10;    select phonenum,count(\*) as cnt from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by phonenum limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu=\*;  set ya100.spark.top10000.ydb\_example\_shu=;  select phonenum,Ycount('\*',ya100\_pipe) as cnt from spark\_ya100 where dt='100million' group by phonenum limit 10;    select phonenum,Ycount('\*',ya100\_pipe) as cnt from spark\_ya100 where dt='1200million' group by phonenum limit 10; |

## 低维值列ydb\_sex,ydb\_province的多列group by count(\*)与排序

|  |  |
| --- | --- |
| **compare type** | **memo** |
| **txt** | select ydb\_sex, ydb\_province,count(\*) as cnt from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,count(\*) as cnt from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10; |
| **spark on parqueue** | select ydb\_sex, ydb\_province,count(\*) as cnt from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,count(\*) as cnt from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu=\*;  set ya100.spark.top10000.ydb\_example\_shu=;  select ydb\_sex, ydb\_province,Ycount('\*',ya100\_pipe) as cnt from spark\_ya100 where dt='100million' group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,Ycount('\*',ya100\_pipe) as cnt from spark\_ya100 where dt='1200million' group by ydb\_sex, ydb\_province order by cnt desc limit 10; |

## 低维值列ydb\_sex,ydb\_province与低维值列amtlong的多列group by sum

|  |  |
| --- | --- |
| **compare type** | **memo** |
| **txt** | select ydb\_sex, ydb\_province,sum(amtlong) as cnt from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,sum(amtlong) as cnt from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10; |
| **spark on parqueue** | select ydb\_sex, ydb\_province,sum(amtlong) as cnt from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,sum(amtlong) as cnt from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu=amtlong;  set ya100.spark.top10000.ydb\_example\_shu=;  select ydb\_sex, ydb\_province,Ysum('amtlong',ya100\_pipe) as cnt from spark\_ya100 where dt='100million' group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,Ysum('amtlong',ya100\_pipe) as cnt from spark\_ya100 where dt='1200million' group by ydb\_sex, ydb\_province order by cnt desc limit 10; |

## 低维值列ydb\_sex,ydb\_province与高纬值列amtdouble的多列group by sum

|  |  |
| --- | --- |
| **compare type** | **memo** |
| **txt** | select ydb\_sex, ydb\_province,sum(amtdouble) as cnt from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,sum(amtdouble) as cnt from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10; |
| **spark on parqueue** | select ydb\_sex, ydb\_province,sum(amtdouble) as cnt from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,sum(amtdouble) as cnt from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu=amtdouble;  set ya100.spark.top10000.ydb\_example\_shu=;  select ydb\_sex, ydb\_province,Ysum('amtdouble',ya100\_pipe) as cnt from spark\_ya100 where dt='100million' group by ydb\_sex, ydb\_province order by cnt desc limit 10;  select ydb\_sex, ydb\_province,Ysum('amtdouble',ya100\_pipe) as cnt from spark\_ya100 where dt='1200million' group by ydb\_sex, ydb\_province order by cnt desc limit 10; |

## 3个列group by ,8个列max与sum

|  |  |
| --- | --- |
| **compare type** | **memo** |
| **txt** | select ydb\_sex,ydb\_province,ydb\_grade, max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),sum(amtlong) as cnt from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_sex,ydb\_province,ydb\_grade  order by cnt desc limit 10;  select ydb\_sex,ydb\_province,ydb\_grade, max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),sum(amtlong) as cnt from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_sex,ydb\_province,ydb\_grade  order by cnt desc limit 10; |
| **spark on parqueue** | select ydb\_sex,ydb\_province,ydb\_grade, max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),sum(amtlong) as cnt from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_sex,ydb\_province,ydb\_grade  order by cnt desc limit 10;  select ydb\_sex,ydb\_province,ydb\_grade, max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),sum(amtlong) as cnt from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 and amtdouble >=1090 and amtdouble <=1100 group by ydb\_sex,ydb\_province,ydb\_grade  order by cnt desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' and amtdouble like '([1090 TO 1100] )';  set ya100.spark.combine.ydb\_example\_shu=ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong;  set ya100.spark.top10000.ydb\_example\_shu=;  select ydb\_sex,ydb\_province,ydb\_grade, Ymaxstring('ydb\_age',ya100\_pipe),Ymaxstring('ydb\_blood',ya100\_pipe),Ymaxstring('ydb\_zhiye',ya100\_pipe),Ymaxstring('ydb\_earn',ya100\_pipe),Ymaxstring('ydb\_prefer',ya100\_pipe),Ymaxstring('ydb\_consume',ya100\_pipe),Ymaxstring('ydb\_day',ya100\_pipe),Ysum('amtlong',ya100\_pipe) as cnt from spark\_ya100 where dt='100million' group by ydb\_sex,ydb\_province,ydb\_grade order by cnt desc limit 10;    select ydb\_sex,ydb\_province,ydb\_grade, Ymaxstring('ydb\_age',ya100\_pipe),Ymaxstring('ydb\_blood',ya100\_pipe),Ymaxstring('ydb\_zhiye',ya100\_pipe),Ymaxstring('ydb\_earn',ya100\_pipe),Ymaxstring('ydb\_prefer',ya100\_pipe),Ymaxstring('ydb\_consume',ya100\_pipe),Ymaxstring('ydb\_day',ya100\_pipe),Ysum('amtlong',ya100\_pipe) as cnt from spark\_ya100 where dt='1200million' group by ydb\_sex,ydb\_province,ydb\_grade order by cnt desc limit 10; |

# 统计性能测试-中等范围统计

## 中等纬值列amtlong的sum,avg,max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select count(amtlong),sum(amtlong) ,avg(amtlong),max(amtlong),min(amtlong) from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 limit 10;    select count(amtlong),sum(amtlong) ,avg(amtlong),max(amtlong),min(amtlong) from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 limit 10; |
| **parquet** | select count(amtlong),sum(amtlong) ,avg(amtlong),max(amtlong),min(amtlong) from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 limit 10;    select count(amtlong),sum(amtlong) ,avg(amtlong),max(amtlong),min(amtlong) from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' ;  set ya100.spark.combine.ydb\_example\_shu=amtlong;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('amtlong',ya100\_pipe) ,Ysum('amtlong',ya100\_pipe) ,Yavg('amtlong',ya100\_pipe), Ymax('amtlong',ya100\_pipe), Ymin('amtlong',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('amtlong',ya100\_pipe) ,Ysum('amtlong',ya100\_pipe) ,Yavg('amtlong',ya100\_pipe), Ymax('amtlong',ya100\_pipe), Ymin('amtlong',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 高纬值列amtdouble的sum,avg,max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select count(amtdouble),sum(amtdouble) ,avg(amtdouble),max(amtdouble),min(amtdouble) from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 limit 10;    select count(amtdouble),sum(amtdouble) ,avg(amtdouble),max(amtdouble),min(amtdouble) from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 limit 10; |
| **parquet** | select count(amtdouble),sum(amtdouble) ,avg(amtdouble),max(amtdouble),min(amtdouble) from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 limit 10;    select count(amtdouble),sum(amtdouble) ,avg(amtdouble),max(amtdouble),min(amtdouble) from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' ;  set ya100.spark.combine.ydb\_example\_shu=amtdouble;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('amtdouble',ya100\_pipe) ,Ysum('amtdouble',ya100\_pipe) ,Yavg('amtdouble',ya100\_pipe), Ymax('amtdouble',ya100\_pipe), Ymin('amtdouble',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('amtdouble',ya100\_pipe) ,Ysum('amtdouble',ya100\_pipe) ,Yavg('amtdouble',ya100\_pipe), Ymax('amtdouble',ya100\_pipe), Ymin('amtdouble',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 低纬值列ydb\_sex的max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select count(ydb\_sex),max(ydb\_sex),min(ydb\_sex) from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 limit 10;    select count(ydb\_sex),max(ydb\_sex),min(ydb\_sex) from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 limit 10; |
| **parquet** | select count(ydb\_sex),max(ydb\_sex),min(ydb\_sex) from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 limit 10;    select count(ydb\_sex),max(ydb\_sex),min(ydb\_sex) from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' ;  set ya100.spark.combine.ydb\_example\_shu=ydb\_sex;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('ydb\_sex',ya100\_pipe), Ymaxstring('ydb\_sex',ya100\_pipe), Yminstring('ydb\_sex',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('ydb\_sex',ya100\_pipe), Ymaxstring('ydb\_sex',ya100\_pipe), Yminstring('ydb\_sex',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 低纬值列ydb\_province的max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select count(ydb\_province),max(ydb\_province),min(ydb\_province) from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 limit 10;    select count(ydb\_province),max(ydb\_province),min(ydb\_province) from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 limit 10; |
| **parquet** | select count(ydb\_province),max(ydb\_province),min(ydb\_province) from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 limit 10;    select count(ydb\_province),max(ydb\_province),min(ydb\_province) from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' ;  set ya100.spark.combine.ydb\_example\_shu=ydb\_province;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('ydb\_province',ya100\_pipe), Ymaxstring('ydb\_province',ya100\_pipe), Yminstring('ydb\_province',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('ydb\_province',ya100\_pipe), Ymaxstring('ydb\_province',ya100\_pipe), Yminstring('ydb\_province',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 高纬值列usernick的max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select count(usernick), max(usernick),min(usernick) from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 limit 10;    select count(usernick), max(usernick),min(usernick) from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 limit 10; |
| **parquet** | select count(usernick), max(usernick),min(usernick) from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 limit 10;    select count(usernick), max(usernick),min(usernick) from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' ;  set ya100.spark.combine.ydb\_example\_shu= usernick;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('usernick',ya100\_pipe), Ymaxstring('usernick',ya100\_pipe), Yminstring('usernick',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('usernick',ya100\_pipe), Ymaxstring('usernick',ya100\_pipe), Yminstring('usernick',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 高纬值列phonenum的max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select max(phonenum),min(phonenum) from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 limit 10;    select max(phonenum),min(phonenum) from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 limit 10; |
| **parquet** | select max(phonenum),min(phonenum) from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 limit 10;    select max(phonenum),min(phonenum) from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' ;  set ya100.spark.combine.ydb\_example\_shu=phonenum;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ymax('phonenum',ya100\_pipe), Ymin('phonenum',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ymax('phonenum',ya100\_pipe), Ymin('phonenum',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 11个中低等维值列sum,avg,max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select count(ydb\_sex),count(ydb\_province),count(ydb\_grade),count(ydb\_age),count(ydb\_blood),count(ydb\_zhiye),count(ydb\_earn),count(ydb\_prefer),count(ydb\_consume),count(ydb\_day),count(amtlong), max(ydb\_sex),max(ydb\_province),max(ydb\_grade),max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),max(amtlong) from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 limit 10;    select count(ydb\_sex),count(ydb\_province),count(ydb\_grade),count(ydb\_age),count(ydb\_blood),count(ydb\_zhiye),count(ydb\_earn),count(ydb\_prefer),count(ydb\_consume),count(ydb\_day),count(amtlong), max(ydb\_sex),max(ydb\_province),max(ydb\_grade),max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),max(amtlong) from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 limit 10; |
| **parquet** | select count(ydb\_sex),count(ydb\_province),count(ydb\_grade),count(ydb\_age),count(ydb\_blood),count(ydb\_zhiye),count(ydb\_earn),count(ydb\_prefer),count(ydb\_consume),count(ydb\_day),count(amtlong), max(ydb\_sex),max(ydb\_province),max(ydb\_grade),max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),max(amtlong) from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 limit 10;    select count(ydb\_sex),count(ydb\_province),count(ydb\_grade),count(ydb\_age),count(ydb\_blood),count(ydb\_zhiye),count(ydb\_earn),count(ydb\_prefer),count(ydb\_consume),count(ydb\_day),count(amtlong), max(ydb\_sex),max(ydb\_province),max(ydb\_grade),max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),max(amtlong) from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' ;  set ya100.spark.combine.ydb\_example\_shu=ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('ydb\_sex',ya100\_pipe),Ycount('ydb\_province',ya100\_pipe),Ycount('ydb\_grade',ya100\_pipe),Ycount('ydb\_age',ya100\_pipe),Ycount('ydb\_blood',ya100\_pipe),Ycount('ydb\_zhiye',ya100\_pipe),Ycount('ydb\_earn',ya100\_pipe),Ycount('ydb\_prefer',ya100\_pipe),Ycount('ydb\_consume',ya100\_pipe),Ycount('ydb\_day',ya100\_pipe),Ycount('amtlong',ya100\_pipe), Ymaxstring('ydb\_sex',ya100\_pipe),Ymaxstring('ydb\_province',ya100\_pipe),Ymaxstring('ydb\_grade',ya100\_pipe),Ymaxstring('ydb\_age',ya100\_pipe),Ymaxstring('ydb\_blood',ya100\_pipe),Ymaxstring('ydb\_zhiye',ya100\_pipe),Ymaxstring('ydb\_earn',ya100\_pipe),Ymaxstring('ydb\_prefer',ya100\_pipe),Ymaxstring('ydb\_consume',ya100\_pipe),Ymaxstring('ydb\_day',ya100\_pipe),Ymax('amtlong',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('ydb\_sex',ya100\_pipe),Ycount('ydb\_province',ya100\_pipe),Ycount('ydb\_grade',ya100\_pipe),Ycount('ydb\_age',ya100\_pipe),Ycount('ydb\_blood',ya100\_pipe),Ycount('ydb\_zhiye',ya100\_pipe),Ycount('ydb\_earn',ya100\_pipe),Ycount('ydb\_prefer',ya100\_pipe),Ycount('ydb\_consume',ya100\_pipe),Ycount('ydb\_day',ya100\_pipe),Ycount('amtlong',ya100\_pipe), Ymaxstring('ydb\_sex',ya100\_pipe),Ymaxstring('ydb\_province',ya100\_pipe),Ymaxstring('ydb\_grade',ya100\_pipe),Ymaxstring('ydb\_age',ya100\_pipe),Ymaxstring('ydb\_blood',ya100\_pipe),Ymaxstring('ydb\_zhiye',ya100\_pipe),Ymaxstring('ydb\_earn',ya100\_pipe),Ymaxstring('ydb\_prefer',ya100\_pipe),Ymaxstring('ydb\_consume',ya100\_pipe),Ymaxstring('ydb\_day',ya100\_pipe),Ymax('amtlong',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 低维值列ydb\_age Top n排序

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select ydb\_sex, ydb\_age from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 order by ydb\_age desc limit 10;    select ydb\_sex, ydb\_age from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 order by ydb\_age desc limit 10; |
| **parquet** | select ydb\_sex, ydb\_age from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 order by ydb\_age desc limit 10;    select ydb\_sex, ydb\_age from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 order by ydb\_age desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' ;  set ya100.spark.combine.ydb\_example\_shu=;  set ya100.spark.top10000.ydb\_example\_shu=ydb\_age desc;  select ydb\_sex, ydb\_age from spark\_ya100 where dt='100million' order by ydb\_age desc limit 10;    select ydb\_sex, ydb\_age from spark\_ya100 where dt='1200million' order by ydb\_age desc limit 10; |

## 低维值列ydb\_age,ydb\_sex Top n排序

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select ydb\_sex, ydb\_age from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 order by ydb\_age desc, ydb\_sex limit 10;    select ydb\_sex, ydb\_age from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 order by ydb\_age desc, ydb\_sex limit 10; |
| **parquet** | select ydb\_sex, ydb\_age from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 order by ydb\_age desc, ydb\_sex limit 10;    select ydb\_sex, ydb\_age from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 order by ydb\_age desc, ydb\_sex limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' ;  set ya100.spark.combine.ydb\_example\_shu=;  set ya100.spark.top10000.ydb\_example\_shu= ydb\_age desc, ydb\_sex;  select ydb\_sex, ydb\_age from spark\_ya100 where dt='100million' order by ydb\_age desc, ydb\_sex limit 10;    select ydb\_sex, ydb\_age from spark\_ya100 where dt='1200million' order by ydb\_age desc, ydb\_sex limit 10; |

## 高纬值列phonenum, usernick Top n排序

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select phonenum,usernick from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 order by phonenum desc limit 10;    select phonenum,usernick from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 order by phonenum desc limit 10; |
| **parquet** | select phonenum,usernick from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 order by phonenum desc limit 10;    select phonenum,usernick from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 order by phonenum desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' ;  set ya100.spark.combine.ydb\_example\_shu=;  set ya100.spark.top10000.ydb\_example\_shu=phonenum desc;  select phonenum,usernick from spark\_ya100 where dt='100million' order by phonenum desc limit 10;    select phonenum,usernick from spark\_ya100 where dt='1200million' order by phonenum desc limit 10; |

## 返回较多列值的 Top n排序

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select phonenum,usernick, ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 order by ydb\_age desc, ydb\_sex limit 10;    select phonenum,usernick, ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 order by ydb\_age desc, ydb\_sex limit 10; |
| **parquet** | select phonenum,usernick, ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 order by ydb\_age desc, ydb\_sex limit 10;    select phonenum,usernick, ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 order by ydb\_age desc, ydb\_sex limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' ;  set ya100.spark.combine.ydb\_example\_shu=;  set ya100.spark.top10000.ydb\_example\_shu= ydb\_age desc,ydb\_sex limit 10;  select phonenum,usernick, ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong from spark\_ya100 where dt='100million' order by ydb\_age desc,ydb\_sex limit 10;    select phonenum,usernick, ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong from spark\_ya100 where dt='1200million' order by ydb\_age desc,ydb\_sex limit 10; |

## 低维值列ydb\_sex的单列group by count(\*)

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select ydb\_sex,count(\*) from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 group by ydb\_sex limit 10;    select ydb\_sex,count(\*) from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 group by ydb\_sex limit 10; |
| **parquet** | select ydb\_sex,count(\*) from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 group by ydb\_sex limit 10;    select ydb\_sex,count(\*) from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 group by ydb\_sex limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' ;  set ya100.spark.combine.ydb\_example\_shu=\*;  set ya100.spark.top10000.ydb\_example\_shu=;  select ydb\_sex,Ycount('\*',ya100\_pipe) from spark\_ya100 where dt='100million' group by ydb\_sex limit 10;    select ydb\_sex,Ycount('\*',ya100\_pipe) from spark\_ya100 where dt='1200million' group by ydb\_sex limit 10; |

## 低维值列ydb\_province的单列group by max(ydb\_age)

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select ydb\_province,count(\*),count(ydb\_age) ,max(ydb\_age) ,min(ydb\_age) from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 group by ydb\_province order by ydb\_province limit 10;    select ydb\_province,count(\*),count(ydb\_age) ,max(ydb\_age) ,min(ydb\_age) from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 group by ydb\_province order by ydb\_province limit 10; |
| **parquet** | select ydb\_province,count(\*),count(ydb\_age) ,max(ydb\_age) ,min(ydb\_age) from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 group by ydb\_province order by ydb\_province limit 10;    select ydb\_province,count(\*),count(ydb\_age) ,max(ydb\_age) ,min(ydb\_age) from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 group by ydb\_province order by ydb\_province limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' ;  set ya100.spark.combine.ydb\_example\_shu=\*,ydb\_age;  set ya100.spark.top10000.ydb\_example\_shu=;  select ydb\_province,Ycount('\*',ya100\_pipe), Ycount('ydb\_age',ya100\_pipe), Ymaxstring('ydb\_age',ya100\_pipe), Yminstring('ydb\_age',ya100\_pipe)  from spark\_ya100 where dt='100million' group by ydb\_province order by ydb\_province limit 10;    select ydb\_province,Ycount('\*',ya100\_pipe), Ycount('ydb\_age',ya100\_pipe), Ymaxstring('ydb\_age',ya100\_pipe), Yminstring('ydb\_age',ya100\_pipe)  from spark\_ya100 where dt='1200million' group by ydb\_province order by ydb\_province limit 10; |

## 高维值列usernick的单列group by count(\*)

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select usernick,count(\*) as cnt from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 group by usernick order by cnt desc limit 10;    select usernick,count(\*) as cnt from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 group by usernick order by cnt desc limit 10; |
| **spark on parque** | select usernick,count(\*) as cnt from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 group by usernick order by cnt desc limit 10;    select usernick,count(\*) as cnt from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 group by usernick order by cnt desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' ;  set ya100.spark.combine.ydb\_example\_shu=\*;  set ya100.spark.top10000.ydb\_example\_shu=;  select usernick,Ycount('\*',ya100\_pipe) as cnt from spark\_ya100 where dt='100million' group by usernick order by cnt desc limit 10;    select usernick,Ycount('\*',ya100\_pipe) as cnt from spark\_ya100 where dt='1200million' group by usernick order by cnt desc limit 10; |

## 高维值列phonenum的单列group by count(\*)

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select phonenum,count(\*) as cnt from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 group by phonenum limit 10;    select phonenum,count(\*) as cnt from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 group by phonenum limit 10; |
| **spark on parqueut** | select phonenum,count(\*) as cnt from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 group by phonenum limit 10;    select phonenum,count(\*) as cnt from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 group by phonenum limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' ;  set ya100.spark.combine.ydb\_example\_shu=\*;  set ya100.spark.top10000.ydb\_example\_shu=;  select phonenum,Ycount('\*',ya100\_pipe) as cnt from spark\_ya100 where dt='100million' group by phonenum limit 10;    select phonenum,Ycount('\*',ya100\_pipe) as cnt from spark\_ya100 where dt='1200million' group by phonenum limit 10; |

## 低维值列ydb\_sex,ydb\_province的多列group by count(\*)与排序

|  |  |
| --- | --- |
| **compare type** | **memo** |
| **txt** | select ydb\_sex, ydb\_province,count(\*) as cnt from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,count(\*) as cnt from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10; |
| **spark on parqueue** | select ydb\_sex, ydb\_province,count(\*) as cnt from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,count(\*) as cnt from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' ;  set ya100.spark.combine.ydb\_example\_shu=\*;  set ya100.spark.top10000.ydb\_example\_shu=;  select ydb\_sex, ydb\_province,Ycount('\*',ya100\_pipe) as cnt from spark\_ya100 where dt='100million' group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,Ycount('\*',ya100\_pipe) as cnt from spark\_ya100 where dt='1200million' group by ydb\_sex, ydb\_province order by cnt desc limit 10; |

## 低维值列ydb\_sex,ydb\_province与低维值列amtlong的多列group by sum

|  |  |
| --- | --- |
| **compare type** | **memo** |
| **txt** | select ydb\_sex, ydb\_province,sum(amtlong) as cnt from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,sum(amtlong) as cnt from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10; |
| **spark on parqueue** | select ydb\_sex, ydb\_province,sum(amtlong) as cnt from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,sum(amtlong) as cnt from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' ;  set ya100.spark.combine.ydb\_example\_shu=amtlong;  set ya100.spark.top10000.ydb\_example\_shu=;  select ydb\_sex, ydb\_province,Ysum('amtlong',ya100\_pipe) as cnt from spark\_ya100 where dt='100million' group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,Ysum('amtlong',ya100\_pipe) as cnt from spark\_ya100 where dt='1200million' group by ydb\_sex, ydb\_province order by cnt desc limit 10; |

## 低维值列ydb\_sex,ydb\_province与高纬值列amtdouble的多列group by sum

|  |  |
| --- | --- |
| **compare type** | **memo** |
| **txt** | select ydb\_sex, ydb\_province,sum(amtdouble) as cnt from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,sum(amtdouble) as cnt from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10; |
| **spark on parqueue** | select ydb\_sex, ydb\_province,sum(amtdouble) as cnt from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,sum(amtdouble) as cnt from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 group by ydb\_sex, ydb\_province order by cnt desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' ;  set ya100.spark.combine.ydb\_example\_shu=amtdouble;  set ya100.spark.top10000.ydb\_example\_shu=;  select ydb\_sex, ydb\_province,Ysum('amtdouble',ya100\_pipe) as cnt from spark\_ya100 where dt='100million' group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,Ysum('amtdouble',ya100\_pipe) as cnt from spark\_ya100 where dt='1200million' group by ydb\_sex, ydb\_province order by cnt desc limit 10; |

## 3个列group by ,8个列max与sum

|  |  |
| --- | --- |
| **compare type** | **memo** |
| **txt** | select ydb\_sex,ydb\_province,ydb\_grade, max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),sum(amtlong) as cnt from spark\_txt where dt='100million' and amtlong>=1090 and amtlong<=1100 group by ydb\_sex,ydb\_province,ydb\_grade  order by cnt desc limit 10;    select ydb\_sex,ydb\_province,ydb\_grade, max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),sum(amtlong) as cnt from spark\_txt where dt='1200million' and amtlong>=1090 and amtlong<=1100 group by ydb\_sex,ydb\_province,ydb\_grade  order by cnt desc limit 10; |
| **spark on parqueue** | select ydb\_sex,ydb\_province,ydb\_grade, max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),sum(amtlong) as cnt from spark\_parquet where dt='100million' and amtlong>=1090 and amtlong<=1100 group by ydb\_sex,ydb\_province,ydb\_grade  order by cnt desc limit 10;    select ydb\_sex,ydb\_province,ydb\_grade, max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),sum(amtlong) as cnt from spark\_parquet where dt='1200million' and amtlong>=1090 and amtlong<=1100 group by ydb\_sex,ydb\_province,ydb\_grade  order by cnt desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu= amtlong like '([1090 TO 1100] )' ;  set ya100.spark.combine.ydb\_example\_shu=ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong;  set ya100.spark.top10000.ydb\_example\_shu=;  select ydb\_sex,ydb\_province,ydb\_grade, Ymaxstring('ydb\_age',ya100\_pipe),Ymaxstring('ydb\_blood',ya100\_pipe),Ymaxstring('ydb\_zhiye',ya100\_pipe),Ymaxstring('ydb\_earn',ya100\_pipe),Ymaxstring('ydb\_prefer',ya100\_pipe),Ymaxstring('ydb\_consume',ya100\_pipe),Ymaxstring('ydb\_day',ya100\_pipe),Ysum('amtlong',ya100\_pipe) as cnt from spark\_ya100 where dt='100million' group by ydb\_sex,ydb\_province,ydb\_grade order by cnt desc limit 10;    select ydb\_sex,ydb\_province,ydb\_grade, Ymaxstring('ydb\_age',ya100\_pipe),Ymaxstring('ydb\_blood',ya100\_pipe),Ymaxstring('ydb\_zhiye',ya100\_pipe),Ymaxstring('ydb\_earn',ya100\_pipe),Ymaxstring('ydb\_prefer',ya100\_pipe),Ymaxstring('ydb\_consume',ya100\_pipe),Ymaxstring('ydb\_day',ya100\_pipe),Ysum('amtlong',ya100\_pipe) as cnt from spark\_ya100 where dt='1200million' group by ydb\_sex,ydb\_province,ydb\_grade order by cnt desc limit 10; |

# 统计性能测试-全表统计

## 中等纬值列amtlong的sum,avg,max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select count(amtlong),sum(amtlong) ,avg(amtlong),max(amtlong),min(amtlong) from spark\_txt where dt='100million' limit 10;    select count(amtlong),sum(amtlong) ,avg(amtlong),max(amtlong),min(amtlong) from spark\_txt where dt='1200million' limit 10; |
| **parquet** | select count(amtlong),sum(amtlong) ,avg(amtlong),max(amtlong),min(amtlong) from spark\_parquet where dt='100million' limit 10;    select count(amtlong),sum(amtlong) ,avg(amtlong),max(amtlong),min(amtlong) from spark\_parquet where dt='1200million' limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu=amtlong;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('amtlong',ya100\_pipe) ,Ysum('amtlong',ya100\_pipe) ,Yavg('amtlong',ya100\_pipe), Ymax('amtlong',ya100\_pipe), Ymin('amtlong',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('amtlong',ya100\_pipe) ,Ysum('amtlong',ya100\_pipe) ,Yavg('amtlong',ya100\_pipe), Ymax('amtlong',ya100\_pipe), Ymin('amtlong',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 高纬值列amtdouble的sum,avg,max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select count(amtdouble),sum(amtdouble) ,avg(amtdouble),max(amtdouble),min(amtdouble) from spark\_txt where dt='100million' limit 10;    select count(amtdouble),sum(amtdouble) ,avg(amtdouble),max(amtdouble),min(amtdouble) from spark\_txt where dt='1200million' limit 10; |
| **parquet** | select count(amtdouble),sum(amtdouble) ,avg(amtdouble),max(amtdouble),min(amtdouble) from spark\_parquet where dt='100million' limit 10;    select count(amtdouble),sum(amtdouble) ,avg(amtdouble),max(amtdouble),min(amtdouble) from spark\_parquet where dt='1200million' limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu=amtdouble;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('amtdouble',ya100\_pipe) ,Ysum('amtdouble',ya100\_pipe) ,Yavg('amtdouble',ya100\_pipe), Ymax('amtdouble',ya100\_pipe), Ymin('amtdouble',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('amtdouble',ya100\_pipe) ,Ysum('amtdouble',ya100\_pipe) ,Yavg('amtdouble',ya100\_pipe), Ymax('amtdouble',ya100\_pipe), Ymin('amtdouble',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 低纬值列ydb\_sex的max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select count(ydb\_sex),max(ydb\_sex),min(ydb\_sex) from spark\_txt where dt='100million' limit 10;    select count(ydb\_sex),max(ydb\_sex),min(ydb\_sex) from spark\_txt where dt='1200million' limit 10; |
| **parquet** | select count(ydb\_sex),max(ydb\_sex),min(ydb\_sex) from spark\_parquet where dt='100million' limit 10;    select count(ydb\_sex),max(ydb\_sex),min(ydb\_sex) from spark\_parquet where dt='1200million' limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu=ydb\_sex;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('ydb\_sex',ya100\_pipe), Ymaxstring('ydb\_sex',ya100\_pipe), Yminstring('ydb\_sex',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('ydb\_sex',ya100\_pipe), Ymaxstring('ydb\_sex',ya100\_pipe), Yminstring('ydb\_sex',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 低纬值列ydb\_province的max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select count(ydb\_province),max(ydb\_province),min(ydb\_province) from spark\_txt where dt='100million' limit 10;    select count(ydb\_province),max(ydb\_province),min(ydb\_province) from spark\_txt where dt='1200million' limit 10; |
| **parquet** | select count(ydb\_province),max(ydb\_province),min(ydb\_province) from spark\_parquet where dt='100million' limit 10;    select count(ydb\_province),max(ydb\_province),min(ydb\_province) from spark\_parquet where dt='1200million' limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu=ydb\_province;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('ydb\_province',ya100\_pipe), Ymaxstring('ydb\_province',ya100\_pipe), Yminstring('ydb\_province',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10    select Ycount('ydb\_province',ya100\_pipe), Ymaxstring('ydb\_province',ya100\_pipe), Yminstring('ydb\_province',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10 |

## 高纬值列usernick的max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select count(usernick), max(usernick),min(usernick) from spark\_txt where dt='100million' limit 10;    select count(usernick), max(usernick),min(usernick) from spark\_txt where dt='1200million' limit 10; |
| **parquet** | select count(usernick), max(usernick),min(usernick) from spark\_parquet where dt='100million' limit 10;    select count(usernick), max(usernick),min(usernick) from spark\_parquet where dt='1200million' limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu= usernick;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('usernick',ya100\_pipe), Ymaxstring('usernick',ya100\_pipe), Yminstring('usernick',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10    select Ycount('usernick',ya100\_pipe), Ymaxstring('usernick',ya100\_pipe), Yminstring('usernick',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10 |

## 高纬值列phonenum的max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select max(phonenum),min(phonenum) from spark\_txt where dt='100million' limit 10;    select max(phonenum),min(phonenum) from spark\_txt where dt='1200million' limit 10; |
| **parquet** | select max(phonenum),min(phonenum) from spark\_parquet where dt='100million' limit 10;    select max(phonenum),min(phonenum) from spark\_parquet where dt='1200million' limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu=phonenum;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ymax('phonenum',ya100\_pipe), Ymin('phonenum',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10    select Ymax('phonenum',ya100\_pipe), Ymin('phonenum',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10 |

## 11个中低等维值列sum,avg,max,min统计

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select count(ydb\_sex),count(ydb\_province),count(ydb\_grade),count(ydb\_age),count(ydb\_blood),count(ydb\_zhiye),count(ydb\_earn),count(ydb\_prefer),count(ydb\_consume),count(ydb\_day),count(amtlong), max(ydb\_sex),max(ydb\_province),max(ydb\_grade),max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),max(amtlong) from spark\_txt where dt='100million' limit 10;    select count(ydb\_sex),count(ydb\_province),count(ydb\_grade),count(ydb\_age),count(ydb\_blood),count(ydb\_zhiye),count(ydb\_earn),count(ydb\_prefer),count(ydb\_consume),count(ydb\_day),count(amtlong), max(ydb\_sex),max(ydb\_province),max(ydb\_grade),max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),max(amtlong) from spark\_txt where dt='1200million' limit 10; |
| **parquet** | select count(ydb\_sex),count(ydb\_province),count(ydb\_grade),count(ydb\_age),count(ydb\_blood),count(ydb\_zhiye),count(ydb\_earn),count(ydb\_prefer),count(ydb\_consume),count(ydb\_day),count(amtlong), max(ydb\_sex),max(ydb\_province),max(ydb\_grade),max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),max(amtlong) from spark\_parquet where dt='100million' limit 10;    select count(ydb\_sex),count(ydb\_province),count(ydb\_grade),count(ydb\_age),count(ydb\_blood),count(ydb\_zhiye),count(ydb\_earn),count(ydb\_prefer),count(ydb\_consume),count(ydb\_day),count(amtlong), max(ydb\_sex),max(ydb\_province),max(ydb\_grade),max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),max(amtlong) from spark\_parquet where dt='1200million' limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu=ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong;  set ya100.spark.top10000.ydb\_example\_shu=;  select Ycount('ydb\_sex',ya100\_pipe),Ycount('ydb\_province',ya100\_pipe),Ycount('ydb\_grade',ya100\_pipe),Ycount('ydb\_age',ya100\_pipe),Ycount('ydb\_blood',ya100\_pipe),Ycount('ydb\_zhiye',ya100\_pipe),Ycount('ydb\_earn',ya100\_pipe),Ycount('ydb\_prefer',ya100\_pipe),Ycount('ydb\_consume',ya100\_pipe),Ycount('ydb\_day',ya100\_pipe),Ycount('amtlong',ya100\_pipe), Ymaxstring('ydb\_sex',ya100\_pipe),Ymaxstring('ydb\_province',ya100\_pipe),Ymaxstring('ydb\_grade',ya100\_pipe),Ymaxstring('ydb\_age',ya100\_pipe),Ymaxstring('ydb\_blood',ya100\_pipe),Ymaxstring('ydb\_zhiye',ya100\_pipe),Ymaxstring('ydb\_earn',ya100\_pipe),Ymaxstring('ydb\_prefer',ya100\_pipe),Ymaxstring('ydb\_consume',ya100\_pipe),Ymaxstring('ydb\_day',ya100\_pipe),Ymax('amtlong',ya100\_pipe) from spark\_ya100 where dt='100million' limit 10;    select Ycount('ydb\_sex',ya100\_pipe),Ycount('ydb\_province',ya100\_pipe),Ycount('ydb\_grade',ya100\_pipe),Ycount('ydb\_age',ya100\_pipe),Ycount('ydb\_blood',ya100\_pipe),Ycount('ydb\_zhiye',ya100\_pipe),Ycount('ydb\_earn',ya100\_pipe),Ycount('ydb\_prefer',ya100\_pipe),Ycount('ydb\_consume',ya100\_pipe),Ycount('ydb\_day',ya100\_pipe),Ycount('amtlong',ya100\_pipe), Ymaxstring('ydb\_sex',ya100\_pipe),Ymaxstring('ydb\_province',ya100\_pipe),Ymaxstring('ydb\_grade',ya100\_pipe),Ymaxstring('ydb\_age',ya100\_pipe),Ymaxstring('ydb\_blood',ya100\_pipe),Ymaxstring('ydb\_zhiye',ya100\_pipe),Ymaxstring('ydb\_earn',ya100\_pipe),Ymaxstring('ydb\_prefer',ya100\_pipe),Ymaxstring('ydb\_consume',ya100\_pipe),Ymaxstring('ydb\_day',ya100\_pipe),Ymax('amtlong',ya100\_pipe) from spark\_ya100 where dt='1200million' limit 10; |

## 低维值列ydb\_age Top n排序

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select ydb\_sex, ydb\_age from spark\_txt where dt='100million' order by ydb\_age desc limit 10;    select ydb\_sex, ydb\_age from spark\_txt where dt='1200million' order by ydb\_age desc limit 10; |
| **parquet** | select ydb\_sex, ydb\_age from spark\_parquet where dt='100million' order by ydb\_age desc limit 10;    select ydb\_sex, ydb\_age from spark\_parquet where dt='1200million' order by ydb\_age desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu=;  set ya100.spark.top10000.ydb\_example\_shu=ydb\_age desc;  select ydb\_sex, ydb\_age from spark\_ya100 where dt='100million' order by ydb\_age desc limit 10;    select ydb\_sex, ydb\_age from spark\_ya100 where dt='1200million' order by ydb\_age desc limit 10; |

## 低维值列ydb\_age,ydb\_sex Top n排序

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select ydb\_sex, ydb\_age from spark\_txt where dt='100million' order by ydb\_age desc, ydb\_sex limit 10;    select ydb\_sex, ydb\_age from spark\_txt where dt='1200million' order by ydb\_age desc, ydb\_sex limit 10; |
| **parquet** | select ydb\_sex, ydb\_age from spark\_parquet where dt='100million' order by ydb\_age desc, ydb\_sex limit 10;    select ydb\_sex, ydb\_age from spark\_parquet where dt='1200million' order by ydb\_age desc, ydb\_sex limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu=;  set ya100.spark.top10000.ydb\_example\_shu= ydb\_age desc, ydb\_sex;  select ydb\_sex, ydb\_age from spark\_ya100 where dt='100million' order by ydb\_age desc, ydb\_sex limit 10;    select ydb\_sex, ydb\_age from spark\_ya100 where dt='1200million' order by ydb\_age desc, ydb\_sex limit 10; |

## 高纬值列phonenum, usernick Top n排序

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select phonenum,usernick from spark\_txt where dt='100million' order by phonenum desc limit 10;    select phonenum,usernick from spark\_txt where dt='1200million' order by phonenum desc limit 10; |
| **parquet** | select phonenum,usernick from spark\_parquet where dt='100million' order by phonenum desc limit 10;    select phonenum,usernick from spark\_parquet where dt='1200million' order by phonenum desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu=;  set ya100.spark.top10000.ydb\_example\_shu=phonenum desc;  select phonenum,usernick from spark\_ya100 where dt='100million' order by phonenum desc limit 10;    select phonenum,usernick from spark\_ya100 where dt='1200million' order by phonenum desc limit 10; |

## 返回较多列值的 Top n排序

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select phonenum,usernick, ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong from spark\_txt where dt='100million' order by ydb\_age desc, ydb\_sex limit 10;    select phonenum,usernick, ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong from spark\_txt where dt='1200million' order by ydb\_age desc, ydb\_sex limit 10; |
| **parquet** | select phonenum,usernick, ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong from spark\_parquet where dt='100million' order by ydb\_age desc, ydb\_sex limit 10;    select phonenum,usernick, ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong from spark\_parquet where dt='1200million' order by ydb\_age desc, ydb\_sex limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu=;  set ya100.spark.top10000.ydb\_example\_shu= ydb\_age desc,ydb\_sex limit 10;  select phonenum,usernick, ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong from spark\_ya100 where dt='100million' order by ydb\_age desc,ydb\_sex limit 10;    select phonenum,usernick, ydb\_sex,ydb\_province,ydb\_grade,ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong from spark\_ya100 where dt='1200million' order by ydb\_age desc,ydb\_sex limit 10; |

## 低维值列ydb\_sex的单列group by count(\*)

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select ydb\_sex,count(\*) from spark\_txt where dt='100million' group by ydb\_sex limit 10;    select ydb\_sex,count(\*) from spark\_txt where dt='1200million' group by ydb\_sex limit 10; |
| **parquet** | select ydb\_sex,count(\*) from spark\_parquet where dt='100million' group by ydb\_sex limit 10;    select ydb\_sex,count(\*) from spark\_parquet where dt='1200million' group by ydb\_sex limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu=\*;  set ya100.spark.top10000.ydb\_example\_shu=;  select ydb\_sex,Ycount('\*',ya100\_pipe) from spark\_ya100 where dt='100million' group by ydb\_sex limit 10    select ydb\_sex,Ycount('\*',ya100\_pipe) from spark\_ya100 where dt='1200million' group by ydb\_sex limit 10 |

## 低维值列ydb\_province的单列group by max(ydb\_age)

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select ydb\_province,count(\*),count(ydb\_age) ,max(ydb\_age) ,min(ydb\_age) from spark\_txt where dt='100million' group by ydb\_province order by ydb\_province limit 10;    select ydb\_province,count(\*),count(ydb\_age) ,max(ydb\_age) ,min(ydb\_age) from spark\_txt where dt='1200million' group by ydb\_province order by ydb\_province limit 10; |
| **parquet** | select ydb\_province,count(\*),count(ydb\_age) ,max(ydb\_age) ,min(ydb\_age) from spark\_parquet where dt='100million' group by ydb\_province order by ydb\_province limit 10;    select ydb\_province,count(\*),count(ydb\_age) ,max(ydb\_age) ,min(ydb\_age) from spark\_parquet where dt='1200million' group by ydb\_province order by ydb\_province limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu=\*,ydb\_age;  set ya100.spark.top10000.ydb\_example\_shu=;  select ydb\_province,Ycount('\*',ya100\_pipe), Ycount('ydb\_age',ya100\_pipe), Ymaxstring('ydb\_age',ya100\_pipe), Yminstring('ydb\_age',ya100\_pipe)  from spark\_ya100 where dt='100million' group by ydb\_province order by ydb\_province limit 10    select ydb\_province,Ycount('\*',ya100\_pipe), Ycount('ydb\_age',ya100\_pipe), Ymaxstring('ydb\_age',ya100\_pipe), Yminstring('ydb\_age',ya100\_pipe)  from spark\_ya100 where dt='1200million' group by ydb\_province order by ydb\_province limit 10 |

## 高维值列usernick的单列group by count(\*)

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select usernick,count(\*) as cnt from spark\_txt where dt='100million' group by usernick order by cnt desc limit 10;    select usernick,count(\*) as cnt from spark\_txt where dt='1200million' group by usernick order by cnt desc limit 10; |
| **spark on parque** | select usernick,count(\*) as cnt from spark\_parquet where dt='100million' group by usernick order by cnt desc limit 10;    select usernick,count(\*) as cnt from spark\_parquet where dt='1200million' group by usernick order by cnt desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu=\*;  set ya100.spark.top10000.ydb\_example\_shu=;  select usernick,Ycount('\*',ya100\_pipe) as cnt from spark\_ya100 where dt='100million' group by usernick order by cnt desc limit 10;    select usernick,Ycount('\*',ya100\_pipe) as cnt from spark\_ya100 where dt='1200million' group by usernick order by cnt desc limit 10;    set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu=;  set ya100.spark.top10000.ydb\_example\_shu=;  select usernick,count(\*) as cnt from spark\_ya100 where dt='1200million' group by usernick order by cnt desc limit 10;    select count(\*) as cnt from spark\_ya100 where dt='1200million' limit 10; |

## 高维值列phonenum的单列group by count(\*)

|  |  |
| --- | --- |
| **类型** | **备注** |
| **txt** | select phonenum,count(\*) as cnt from spark\_txt where dt='100million' group by phonenum limit 10;    select phonenum,count(\*) as cnt from spark\_txt where dt='1200million' group by phonenum limit 10; |
| **spark on parqueut** | select phonenum,count(\*) as cnt from spark\_parquet where dt='100million' group by phonenum limit 10;    select phonenum,count(\*) as cnt from spark\_parquet where dt='1200million' group by phonenum limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu=\*;  set ya100.spark.top10000.ydb\_example\_shu=;  select phonenum,Ycount('\*',ya100\_pipe) as cnt from spark\_ya100 where dt='100million' group by phonenum limit 10;    select phonenum,Ycount('\*',ya100\_pipe) as cnt from spark\_ya100 where dt='1200million' group by phonenum limit 10; |

## 低维值列ydb\_sex,ydb\_province的多列group by count(\*)与排序

|  |  |
| --- | --- |
| **compare type** | **memo** |
| **txt** | select ydb\_sex, ydb\_province,count(\*) as cnt from spark\_txt where dt='100million' group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,count(\*) as cnt from spark\_txt where dt='1200million' group by ydb\_sex, ydb\_province order by cnt desc limit 10; |
| **spark on parqueue** | select ydb\_sex, ydb\_province,count(\*) as cnt from spark\_parquet where dt='100million' group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,count(\*) as cnt from spark\_parquet where dt='1200million' group by ydb\_sex, ydb\_province order by cnt desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu=\*;  set ya100.spark.top10000.ydb\_example\_shu=;  select ydb\_sex, ydb\_province,Ycount('\*',ya100\_pipe) as cnt from spark\_ya100 where dt='100million' group by ydb\_sex, ydb\_province order by cnt desc limit 10    select ydb\_sex, ydb\_province,Ycount('\*',ya100\_pipe) as cnt from spark\_ya100 where dt='1200million' group by ydb\_sex, ydb\_province order by cnt desc limit 10 |

## 低维值列ydb\_sex,ydb\_province与低维值列amtlong的多列group by sum

|  |  |
| --- | --- |
| **compare type** | **memo** |
| **txt** | select ydb\_sex, ydb\_province,sum(amtlong) as cnt from spark\_txt where dt='100million' group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,sum(amtlong) as cnt from spark\_txt where dt='1200million' group by ydb\_sex, ydb\_province order by cnt desc limit 10; |
| **spark on parqueue** | select ydb\_sex, ydb\_province,sum(amtlong) as cnt from spark\_parquet where dt='100million' group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,sum(amtlong) as cnt from spark\_parquet where dt='1200million' group by ydb\_sex, ydb\_province order by cnt desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu=amtlong;  set ya100.spark.top10000.ydb\_example\_shu=;  select ydb\_sex, ydb\_province,Ysum('amtlong',ya100\_pipe) as cnt from spark\_ya100 where dt='100million' group by ydb\_sex, ydb\_province order by cnt desc limit 10    select ydb\_sex, ydb\_province,Ysum('amtlong',ya100\_pipe) as cnt from spark\_ya100 where dt='1200million' group by ydb\_sex, ydb\_province order by cnt desc limit 10 |

## 低维值列ydb\_sex,ydb\_province与高纬值列amtdouble的多列group by sum

|  |  |
| --- | --- |
| **compare type** | **memo** |
| **txt** | select ydb\_sex, ydb\_province,sum(amtdouble) as cnt from spark\_txt where dt='100million' group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,sum(amtdouble) as cnt from spark\_txt where dt='1200million' group by ydb\_sex, ydb\_province order by cnt desc limit 10; |
| **spark on parqueue** | select ydb\_sex, ydb\_province,sum(amtdouble) as cnt from spark\_parquet where dt='100million' group by ydb\_sex, ydb\_province order by cnt desc limit 10;    select ydb\_sex, ydb\_province,sum(amtdouble) as cnt from spark\_parquet where dt='1200million' group by ydb\_sex, ydb\_province order by cnt desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu=amtdouble;  set ya100.spark.top10000.ydb\_example\_shu=;  select ydb\_sex, ydb\_province,Ysum('amtdouble',ya100\_pipe) as cnt from spark\_ya100 where dt='100million' group by ydb\_sex, ydb\_province order by cnt desc limit 10    select ydb\_sex, ydb\_province,Ysum('amtdouble',ya100\_pipe) as cnt from spark\_ya100 where dt='1200million' group by ydb\_sex, ydb\_province order by cnt desc limit 10 |

## 3个列group by ,8个列max与sum

|  |  |
| --- | --- |
| **compare type** | **memo** |
| **txt** | select ydb\_sex,ydb\_province,ydb\_grade, max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),sum(amtlong) as cnt from spark\_txt where dt='100million' group by ydb\_sex,ydb\_province,ydb\_grade  order by cnt desc limit 10;    select ydb\_sex,ydb\_province,ydb\_grade, max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),sum(amtlong) as cnt from spark\_txt where dt='1200million' group by ydb\_sex,ydb\_province,ydb\_grade  order by cnt desc limit 10; |
| **spark on parqueue** | select ydb\_sex,ydb\_province,ydb\_grade, max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),sum(amtlong) as cnt from spark\_parquet where dt='100million' group by ydb\_sex,ydb\_province,ydb\_grade  order by cnt desc limit 10;    select ydb\_sex,ydb\_province,ydb\_grade, max(ydb\_age),max(ydb\_blood),max(ydb\_zhiye),max(ydb\_earn),max(ydb\_prefer),max(ydb\_consume),max(ydb\_day),sum(amtlong) as cnt from spark\_parquet where dt='1200million' group by ydb\_sex,ydb\_province,ydb\_grade  order by cnt desc limit 10; |
| **Ya100** | set ya100.spark.filter.ydb\_example\_shu=;  set ya100.spark.combine.ydb\_example\_shu=ydb\_age,ydb\_blood,ydb\_zhiye,ydb\_earn,ydb\_prefer,ydb\_consume,ydb\_day,amtlong;  set ya100.spark.top10000.ydb\_example\_shu=;  select ydb\_sex,ydb\_province,ydb\_grade, Ymaxstring('ydb\_age',ya100\_pipe),Ymaxstring('ydb\_blood',ya100\_pipe),Ymaxstring('ydb\_zhiye',ya100\_pipe),Ymaxstring('ydb\_earn',ya100\_pipe),Ymaxstring('ydb\_prefer',ya100\_pipe),Ymaxstring('ydb\_consume',ya100\_pipe),Ymaxstring('ydb\_day',ya100\_pipe),Ysum('amtlong',ya100\_pipe) as cnt from spark\_ya100 where dt='100million' group by ydb\_sex,ydb\_province,ydb\_grade order by cnt desc limit 10    select ydb\_sex,ydb\_province,ydb\_grade, Ymaxstring('ydb\_age',ya100\_pipe),Ymaxstring('ydb\_blood',ya100\_pipe),Ymaxstring('ydb\_zhiye',ya100\_pipe),Ymaxstring('ydb\_earn',ya100\_pipe),Ymaxstring('ydb\_prefer',ya100\_pipe),Ymaxstring('ydb\_consume',ya100\_pipe),Ymaxstring('ydb\_day',ya100\_pipe),Ysum('amtlong',ya100\_pipe) as cnt from spark\_ya100 where dt='1200million' group by ydb\_sex,ydb\_province,ydb\_grade order by cnt desc limit 10 |