

Spark-SQL

TalkingData—田永军

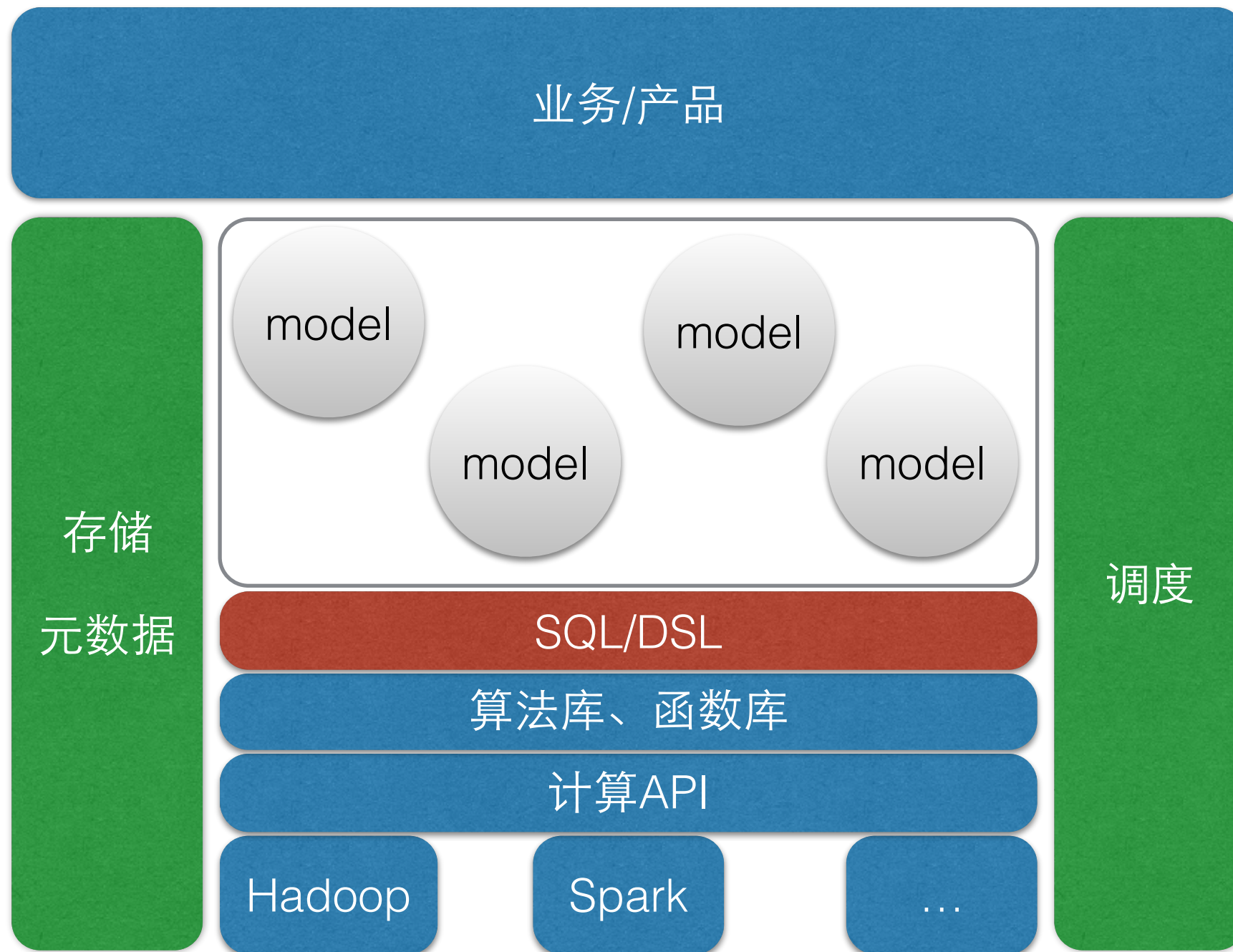
演变

- Hadoop前
- Hadoop、Hive、Impala、Presto ...
- Spark、Spark-SQL

SQL计算理解

- 高层计算
- 高度抽象

位置



为什么Spark-SQL

- 几乎所有的计算是基于Spark
- 支持Parquet
- 不用自己再解析一套SQL

遇到问题

- Parquet Schema REPEATED
 - 通过自定义SqlContext扩展
- 嵌套问题（支持不太好）
 - UDF支持
- 整合现在的元数据管理

后来

```
sqlContext.sql(
  """
    lselect iat(bmap(mergeApps)) as iat,seq.city as city from parquet("/data/parquet/ta/ta")
    lwhere map_value(bmap(mergeApps) , -5208104353007938971 ) > 0
    lgroup by explode(seq.city,map(bmap(mergeApps),"$key"))
    lorder by item(iat,2)
  """.stripMargin).foreach{
    dm =>
      println( dm )
  }
```

```
sqlContext.sql(
  """
    lselect id from text("/data/test/ids" ,"struct( field(id , string , required), field(name ,
    string , required) )" , ",")
  """.stripMargin
).foreach{
  println
}
```

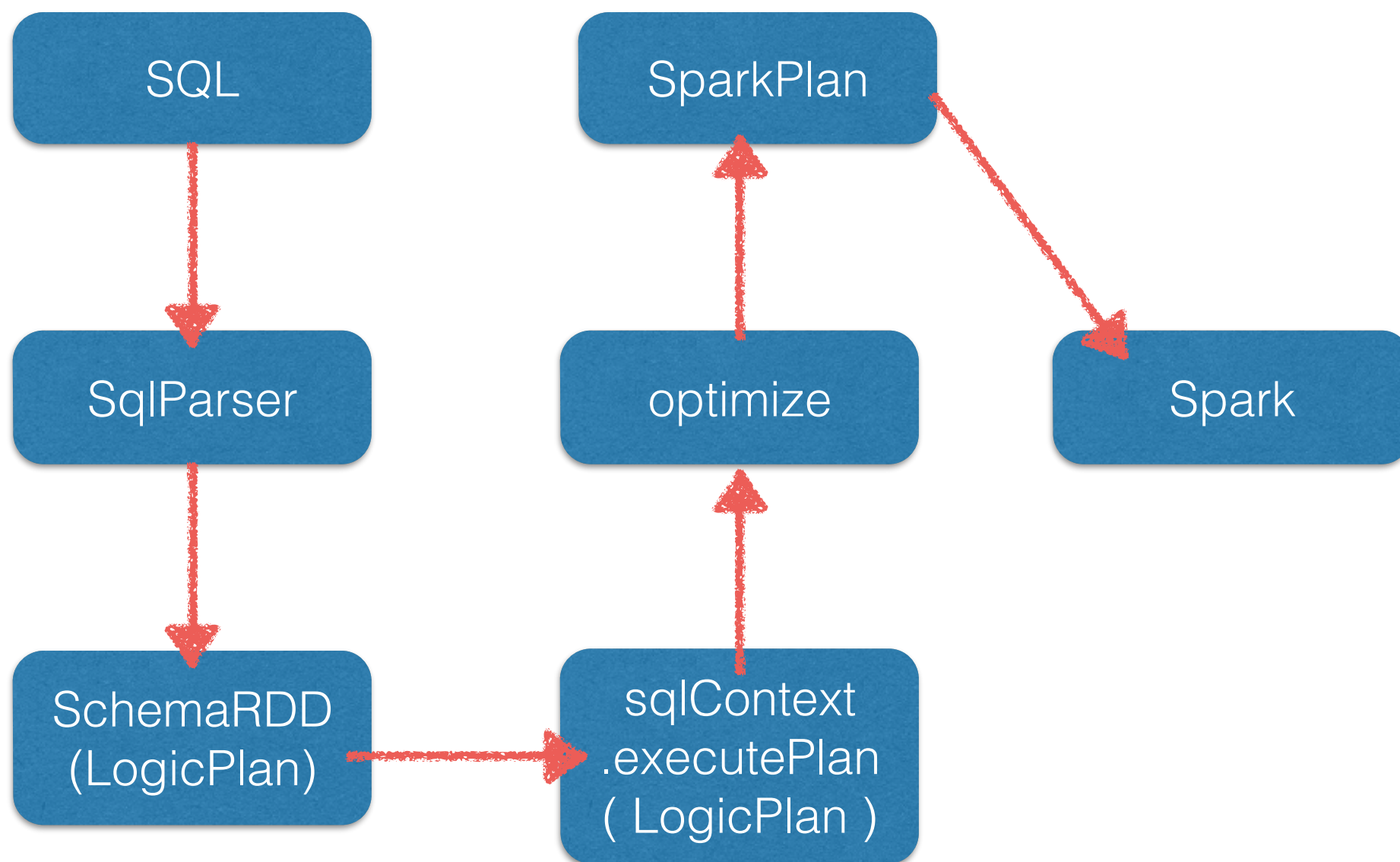
以后

- `select xxx from parquet(xxx) where xxx group by xxx order by xxx into parquet(xxx)`

Spark-SQL使用方式

- 内嵌到spark代码中
 - SqlContext/HiveContext
- JDBC/ODBC 连接

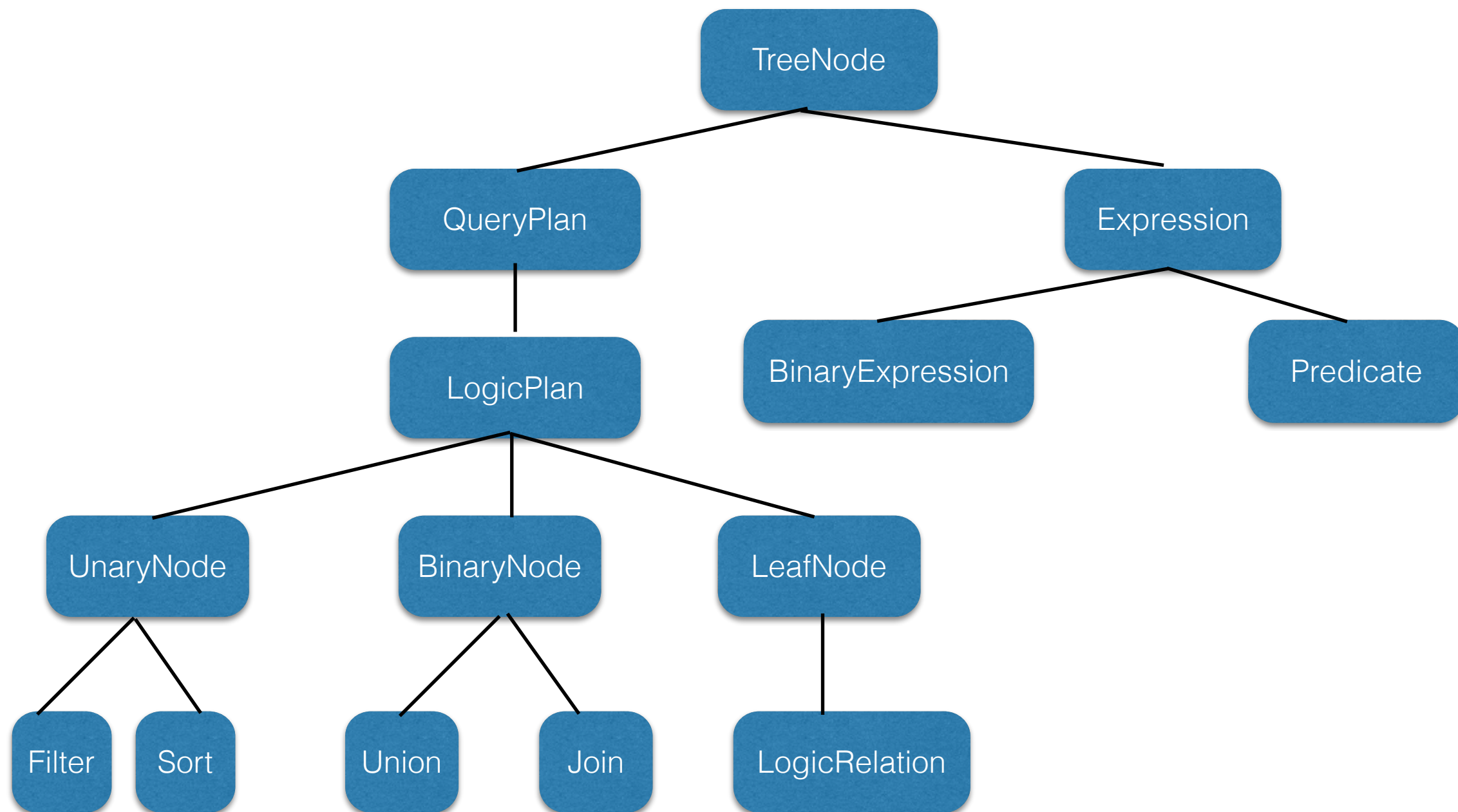
Spark-SQL流程



Catalyst

- sqlParser/ddlParser
- expressions
- analysis
- plans
- types
- optimizer

Catalyst结构



Join

- leftOuter/rightOuter/fullOuter
- inner
- leftSemi

Spark 问题和解决

OOM/Hang

- GC overhead limit exceeded
- Java heap space
- Java perm space
 - `-XX:PermSize=256M -XX:MaxPermSize=256M`
- 线程同步问题

内存问题

- Collect/Broadcast 的数据量大
 - 使用RDD的操作来改善
 - 分批处理
- 计算逻辑导致数据量大
- 数据倾斜/脏数据
 - 过滤掉或者截掉

内存问题

- Full gc
 - 调大shuffle并行度
 - 调低spark.storage.memoryFraction

线程问题

- Base64解码后乱码
- 日期计算出 2015-06-31
- 程序Hang
 - parquet bug(HashMap)

spark-service

[Jobs](#)[Stages](#)[Storage](#)[Environment](#)[Executors](#)[Jobs2](#)[AddJob](#)[Jars](#)

Spark-Service application UI

Spark Jobs (?)

Total Duration: 377.7 h

Scheduling Mode: FIFO

Active Jobs: 0

Completed Jobs: 979

Failed Jobs: 25

Active Jobs (0)

Job Id	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
--------	-------------	-----------	----------	-------------------------	---

Completed Jobs (979)

Job Id	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
1778	runJob at ParquetTableOperations.scala:325	2015/06/26 09:52:16	3.1 h	2/2 (1 skipped)	64/64 (1968 skipped)
1776	runJob at ParquetTableOperations.scala:325	2015/06/26 09:28:20	3.0 h	2/2 (1 skipped)	44/44 (1968 skipped)
1775	runJob at ParquetTableOperations.scala:325	2015/06/26 09:27:59	2.9 h	2/2 (1 skipped)	64/64 (1968 skipped)
1777	runJob at ParquetTableOperations.scala:325	2015/06/26 09:29:07	2.5 h	3/3	104/104
1772	runJob at ParquetTableOperations.scala:325	2015/06/26 09:05:36	2.6 h	2/2 (1 skipped)	64/64 (1968 skipped)
1774	runJob at ParquetTableOperations.scala:325	2015/06/26 09:25:08	1.8 h	3/3	168/168
1770	runJob at ParquetTableOperations.scala:325	2015/06/26 08:36:48	2.6 h	2/2 (1 skipped)	44/44 (1968 skipped)
1771	runJob at ParquetTableOperations.scala:325	2015/06/26 08:43:44	2.0 h	3/3	168/168
1769	runJob at ParquetTableOperations.scala:325	2015/06/26 08:32:19	2.2 h	2/2 (1 skipped)	44/44 (1968 skipped)
1773	runJob at ParquetTableOperations.scala:325	2015/06/26 09:07:47	1.5 h	1/1 (2 skipped)	8/8
1767	runJob at ParquetTableOperations.scala:325	2015/06/26 08:10:19	2.2 h	2/2 (1 skipped)	44/44 (1968 skipped)

spark-service

- 共享RDD
- 减少作业提交时间
- 充分利用资源

QA & 谢谢