Top 1

亿级别的大量结果集排序、分组（group by）、分页 (Limit)的优化问题

资料：目前假如一个查询SQL跨越30个分片，每个分片上有1000万数据，则总数据规模为3亿，Select × from A orderby field1,field2 Limit 100000,100 即取出排序结果集中从100000到100000+100的这100个记录，所可能采用的一些算法优化策略如下：

* 排序字段放在内存中（限定一个SQL允许占用10M内存？，超过部分放在DirecctBuffer映射的文件中），完整记录存储 在DirecctBuffer映射的文件中或者二进制文件中
* 由于每个分片都是排序的结果集，因此这里的排序有特殊性，即从已经拍好序的N个结果集中进行部分排序，有可能某个分片的结果集被完全排除
* 由于这种排序分页的SQL具有连续性，即第二次查询，只是偏移量发生变化，因此上次排序的一些结果（如某个分片最小值是多少，上次是那个分片的结果）对于下一次排序有优化作用

目前此问题是Mycat 第一大难题，至今无人解决，文档中的《一亿组数据排序》是曾经优化过此问题的同学提供的，可以参考。

Top 2

跨分片JOIN的实现

当需要把SQL结果放到本地进行编程的JOIN时候，遇到的问题比较多，目前设想设计实现一个SQL处理引擎，用编写类似数据库存储过程的方式，“人工”完成JOIN编程，简称人工智能解决JOIN，此问题是新提出来的 S级别任务。

Top3

将阿里的Droid SQL Parser引入Mycat，实现针对MySQL/Oracel语法的特定语法支持

Top 4

实现一个通用的NoSQL转接引擎，将SQL语句翻译成NoSQL所需要的基本信息，比如新增记录（字段名，字段值），虚拟表。。具体实现的Plugin则用此信息完成NoSQL API的调用，实现后端存储。目标是前端统一SQL界面，后端可以灵活替换为各种NoSQL存储。

Top 5

MyCAT智能优化模块，此模块的目标是尽量自动收集Mycat与Mysql运行期的性能参数，提供智能优化指导，包括分片是否合适，是否要建立某些索引，热点数据是哪些，哪些SQL很频繁，Mysql数据库当前的性能指标是否正常。。《Mycat智能优化设计方案》是一些参考的设计。

Top 6 MyCAT Balance

MyCAT Balance有几个重要目标：

* 理解MySQL协议，做到协议层负载均衡，当某个Mycat需要进行维护时候，从集群中脱离，不再接收新请求，当已有的SQL执行完成以后，自动脱离Balance，做到100%不中断业务的更新维护。
* 因为理解MYSQL协议，因此后端的MyCAT集群可以设计为互补模式，以支持更大规模的MYSQL集群，思路如下，假如系统有100个表，则20个表为一组放在一个Mycat上分片，总共5个MyCAT 担负这100个表。对于有Join关系的表，则每个MyCAT都定义一遍。这个特性实现以后，MYCAT集群将会超越很多现有的MySQL集群，适应更强的大数据的计算要求。