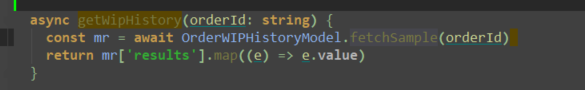
# 浅谈MongoDB中的MapReduce

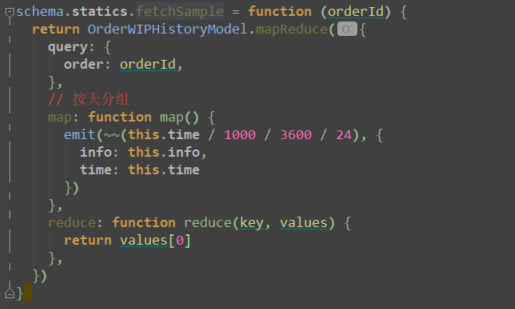
## 问题引导

在checkout老大的代码下，偶然发现有对聚合函数的代码，因为小白，接触的不多，所以就抓紧努力学习了一把。

看到下图中返回值中使用了Map函数遍历出返回值。



一看后面，果然藏着一只大鱼，好吧！



## 分析MapReduce

分析一下上面的内容，是Schema数据的Model层定义了一个fetchSample的查询订单历史的方法，查询参数是orderId。

在mapReduce中可以看到map函数和reduce函数。还有一个方法是emit我暂时不知道它是怎么用的。

## 使用场景

1. 用于计数以及实现聚合函数统计数据。
2. 对于数据进行分组简化或者构造自己想要的格式。
3. 根据条件进行数据筛选。

## 学习与讨论

Map-Reduce是一种计算模型，简单的说就是将大批量的工作（数据）分解（MAP）执行，然后再将结果合并成最终结果（REDUCE）。

使用 MapReduce 要实现两个函数 Map 函数和 Reduce 函数,Map 函数调用 emit(key, value), 遍历 collection 中所有的记录, 将 key 与 value 传递给 Reduce 函数进行处理。

解释一下刚刚提出的emit的问题：emit函数是非常重要的，他的作用是将一条数据放入数据分组集合，这个分组是以emit的第一个参数为key的。你可以这样理解，当你在所有需要计算的行执行完了map函数，你就得到了一组key-values对。基本key是emit中的key，values是每次emit函数的第二个参数组成的集合。

我们的任务就是将这一个key-values变在key-value，也就是把这一个集合变成一个单一的值。这个操作就是Reduce

参数说明:

map：映射函数 (生成键值对序列,作为 reduce 函数参数)。

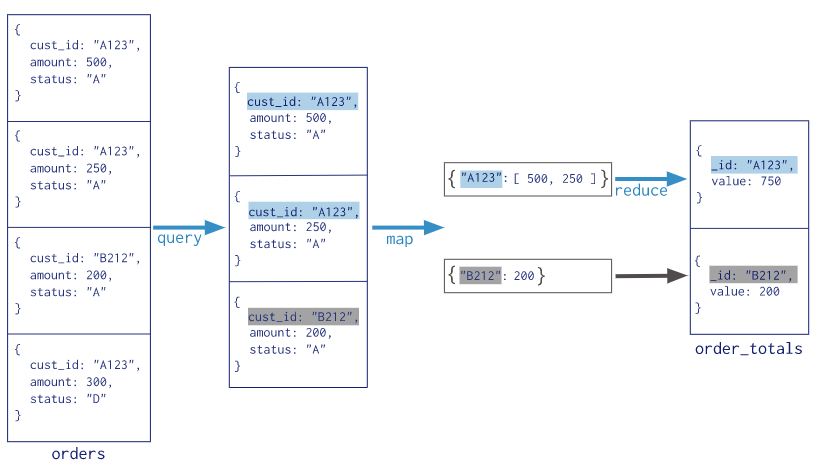
Reduce：统计函数，reduce函数的任务就是将key-values变成key-value，也就是把values数组变成一个单一的值value。。

Out：统计结果存放集合 (不指定则使用临时集合,在客户端断开后自动删除)。

Query：一个筛选条件，只有满足条件的文档才会调用map函数。（query。limit，sort可以随意组合）

Sort：和limit结合的sort排序参数（也是在发往map函数前给文档排序），可以优化分组机制

Limit：发往map函数的文档数量的上限（要是没有limit，单独使用sort的用处不大）

举个被常用的例子：

以上实例在集合 orders 中查找 status:"A" 的数据，并根据 cust\_id 来分组，并计算 amount 的总和。

db.orders.mapReduce({

map：function(){emit(this.cust\_id,this.amount）}

reduce:：function(key,value){

return Array.sum(values)

},

{

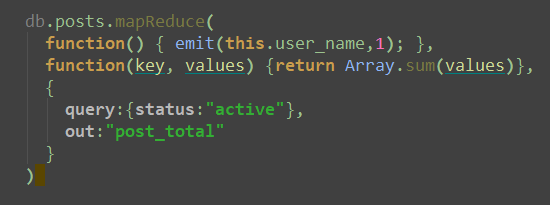
query:{status:”A”}

out:”order\_totals”

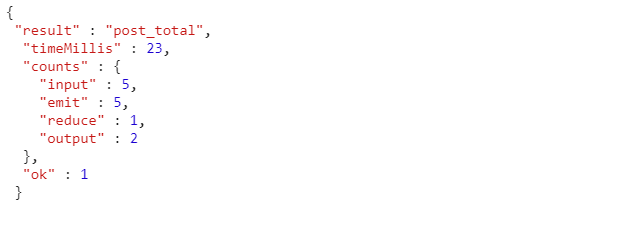
}

})

现在，我们将在 posts 集合中使用 mapReduce 函数来选取已发布的文章(status:"active")，并通过user\_name分组，计算每个用户的文章数：



以上 mapReduce 输出结果为：



结果表明，共有 5 个符合查询条件（status:"active"）的文档， 在map函数中生成了 5 个键值对文档，最后使用reduce函数将相同的键值分为 2 组。

具体参数说明：

result：储存结果的collection的名字,这是个临时集合，MapReduce的连接关闭后自动就被删除了。

timeMillis：执行花费的时间，毫秒为单位

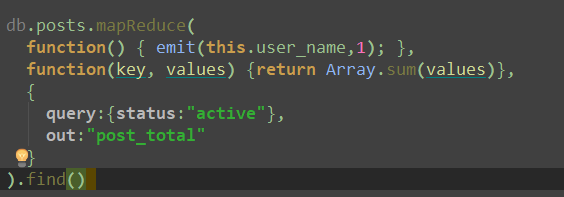
input：满足条件被发送到map函数的文档个数

emit：在map函数中emit被调用的次数，也就是所有集合中的数据总量

ouput：结果集合中的文档个数（count对调试非常有帮助）

ok：是否成功，成功为1

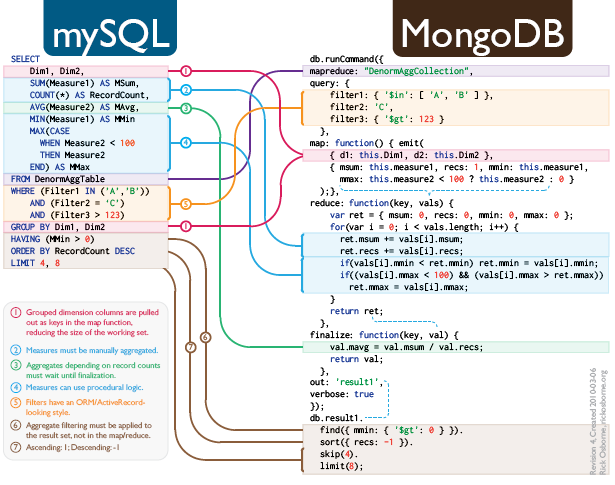
err：如果失败，这里可以有失败原因，不过从经验上来看，原因比较模糊，作用不大

使用 find 操作符来查看 mapReduce 的查询结果：

以上查询显示如下结果，两个用户 tom 和 mark 有两个发布的文章:

用类似的方式，MapReduce可以被用来构建大型复杂的聚合查询。

## 对比MySql

对于mongo的学习，我会经常把它跟mySql的语法做比较。对比与mongo的聚合函数。通过对下面这张图，我们就会对刚才的分析有深刻的理解了

根据上图我再次进行对比的分析，emit函数存放的是分组集合；Reduce通过分组集合进行聚合函数的统计计算；Query确定筛选范围。