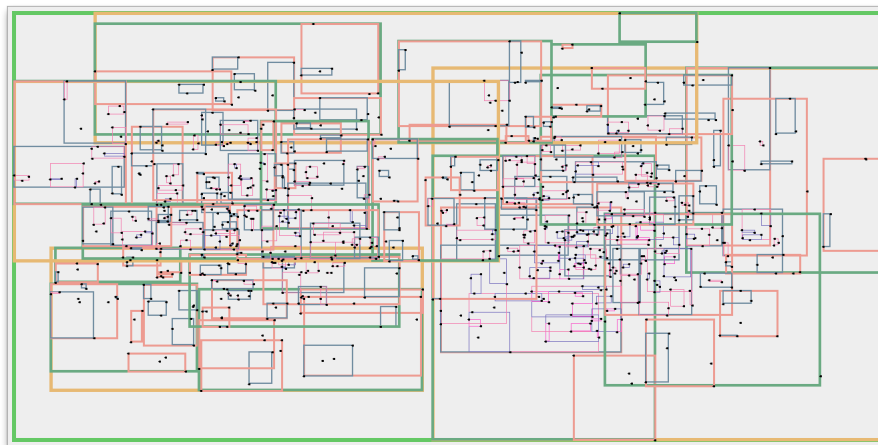


Отчет №3

Амосов Федор

20 ноября 2013 г.



После реализованного построения R-Tree, можно с уверенностью добавить базу в псевдокоде алгоритма,

Алгоритм

Дано множество d мерных точек $P = \{p_i\}_{i=1}^n$. Требуется реализовать построение VoRTree на этих точках с помощью MapReduce.

u_{min}, u_{max} = min and max number of sons
 f = some 2-dimension function
 m = number of java virtual machines in cluster

```
M-Tree(Points  $P$ ) {  
     $root$  = new Node  
    radius of  $root$  = min radius of ball, which contains  $P$   
     $k = f(u_{min}, u_{max})$   
    if ( $|P| < k$ ) {  
        add  $P$  to  $root$   
         $T$  = new Tree only with  $root$   
        return  $T$   
    }  
     $S$  = get  $k$  different random points from  $P$   
     $\{P_i\}$  = split  $P$  into  $m$  parts  
     $\{Pair<P_i, S_i>\}$  = MapReduce( $\{Pair<P_i, S_i>\}$ )  
    for  $P_{s_i}$  in  $\{P_{s_i}\}$  {  
         $T_{s_i}$  = M-Tree( $P_{s_i}$ )  
    }  
     $T$  = hang  $\{T_{s_i}\}$  by  $root$   
    return  $T$   
}
```

```

map(Pair<Points P, Points S>) {
    for p in P {
        s = closest to p from S
        to output: map.entry(s, p)
    }
}

reduce(Point s, Points P_s) {
    return map.entry(s, P_s)
}

```

Вопросы

1. На самом деле, интересно следующее. После того, как в строчке `{Pair< s_i , P_{s_i} >} = MapReduce({Pair< P_i , S >})` отработает MapReduce, все JVM кластера бездействуют. На самом деле, далее они будут загружены внутри рекурсивного построения, но все же, было бы интересно каждое T_{s_i} строить на отдельной JVM. Ясно, что тогда будет рекурсия в MapReduce, что не совсем понятно. Поэтому, хотелось бы узнать что-нибудь о схеме кластера машин и об общении JVM внутри кластера.
2. Когда лучше делать запрос к базе данных? Если в Mapper, то получится слишком много запросов, если перед MapReduce, то надо будет хранить слишком много точек в оперативной памяти.
3. Честно говоря, я не нашел готовой симплификации Делоне в d -мерном пространстве. Есть только для плоскости. Неужели ее придется писать самому?
4. Как искать объемлющую сферу в евклидовой метрике? Это сложная задача — именно поэтому мы используем объемлющие прямоугольники?
5. Откуда точки будут появляться в базе данных? Они там изначально будут, или мы их туда засунем из файла?
6. Есть еще несколько технических вопросов. Но я их пока что не могу сформулировать, ввиду того, что я еще не закомитил соответствующий код.