

Отчет №2

Амосов Федор

15 февраля 2014 г.

Алгоритм

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

Далее, привожу набросок алгоритма построения произвольного M-tree с MapReduce. База (построение листьев) не приводится.

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

M-Tree(Points  $P$ ) {
     $k = f(u_{min}, u_{max})$ 
     $S = \text{get } k \text{ different random points from } P$ 
     $\{P_i\} = \text{split } P \text{ into } m \text{ parts}$ 
     $\{\text{Pair}\langle s_i, P_{s_i} \rangle\} = \text{MapReduce}(\{\text{Pair}\langle P_i, S \rangle\})$ 
    for  $P_{s_i}$  in  $\{P_{s_i}\}$  {
         $T_{s_i} = \text{M-Tree}(P_{s_i})$ 
    }
     $T = \text{hang } \{T_{s_i}\} \text{ by new root}$ 
    radius of root = min radius of ball, which contains  $P$  *
    return  $T$ 
}

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

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

Вопросы

1. Правда, что мы никак не можем внутри маппера иметь доступ к каким-то «большим глобальным» коллекциям данных? (к примеру, чтобы все мапперы видели один и тот же объект). Видимо нет, т.к. мапперы работают на разных машинах.
2. Зачем нам вообще диаграмма Вороного на этапе построения VoR-Tree?
3. Я правильно понимаю, что мы не можем себе позволить хранить одновременно все точки в оперативной памяти? Если это так, то над некоторыми, на первый взгляд простыми действиями, придется призадуматься.

4. Надо научиться что-то понимать про функцию f с этапа $k = f(u_{min}, u_{max})$.
5. Еще у меня есть много всяких дурацких вопросов по тому, как конкретно мы будем работать с точками. Они у нас будут храниться где-нибудь и мы будем передавать в методы только их айдишники? Или мы будем передавать сами точки? Тогда, если точки будут иметь координаты в `double`, то после маппера их не удастся «склеить» по ключам. И т.п.