

保证睡眠 身体第一 比赛第二

2019中间件复赛会议记录

会议时间

2019年07月31日晚

参赛人员

杨、周、王

主要讨论事项

- 1. t的压缩方式
- 2. msg的压缩方式
- 3. a的数据统计

详细讨论

- 1. t的压缩方式采用边长编码方式进行压缩，变长编码格式如表（以三位变长编码为例）

比特位	1	2	3
数据含义	表示本字节是否是最后一个字节	表示数据	表示数据

- 2. 为了方便t的查找，所以每隔固定的t的个数建立一个t的索引，方便对t进行查找操作，如果每隔16个t建立一个索引，大约需要1g的内存存放（20亿/16 *8)
- 3. 统计得出t的分布情况，请参考日志。统计同线程的t的分布情况，得出前后t的差值分布，主要结果为：0占比约为50% 1占比约为40% 其余占比10%，所以采用2比特的变长编码。统计得出的单线程所使用的存储空间少于50mb
- 4. 统计得出a分布主要集中在-32773，a的压缩方法待定，等待t实现之后对a进行讨论
- 5. msg的数据格式如表,其中所有的msg的盐均相同
- 6. 对于msg压缩最好使用通用的压缩算法进行压缩，可以尝试使用java中的zip进行压缩。测试数据压缩对于cpu的负载以及压缩率等情况

比特位	0-7	8-15	16-32
含义	t	a	盐

下一步的安排

1. 统计a-preva的最大最小以及分布情况
2. 考虑msg的压缩算法，收集多种压缩 算法的实现，根据数据挑选最好的。注：实现准备好非通用压缩以及通用压缩算法，根据比赛情况进行调整使用。
- 3.