对一个数据流处理系统进行性能评估，主要是吞吐量：其系统在单位时间内可以处理多少元素，和延迟：一个元素从第一次与系统接触到输出需要多长时间，以及可扩展性。使用代码进行评估。

该系统是用idea软件写的一个maven项目，后台语言Java主要用了Flink架构，前台语言JavaScript，使用WebSocket推送信息。

延迟需要说明：【可以使用Flink特殊的LatencyMarkers来测量延迟，在网上可以找到相关信息，此外，可以在flink中定义metrics和markers，然后在flink仪表板上查看。该网站是官方flink monitoring的介绍：https://flink.apache.org/news/2019/02/25/monitoring-best-practices.html】

一个元素从时间点t1（在flink源中摄取元素）到时间点t2（元素从sink发送至前端）到时间点t3（元素在前端界面显示），这个过程中t1至t2，t2至t3的时长。

数据来源使用weibo-data-generator启用WeiboDataGeneratorApplication类，weibo-data-analysis启用WeiboAnalysis类。

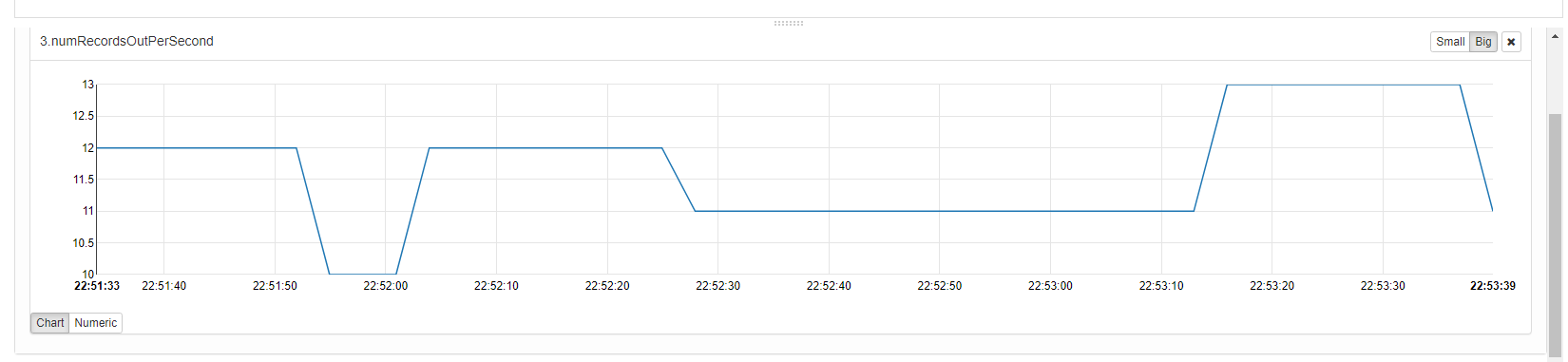
频率：5个元素/秒（weibo-data-generator中的DataGenerator.java 中的 private static int sendNumberPerSecond就是每秒产生元素的个数，改成10即可）

【因为我不知道这个系统一秒钟读多少个元素会出现数据压力的现象，所以这个频率是我自己瞎设置的，你可以根据实际情况选择4到5种情况，直到出现后台无法处理传入数据的现象出现】

完成下表内的不同功能的性能测试：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 性能/功能 | 默认（全都不改变） | Filter功能：Vertex选择User；Edge选择Fans | Filter功能：Vertex选择Tag和Weibo；Edge选择 BelongTo和ReplyOf | Gouping by label功能勾选 |
| 吞吐量 | 12/s | 14/s | 3/s | 12/s |
| 延迟 | 512.75+160 ms | 311.55+308.65ms | 280+172ms | 308.2+200ms |

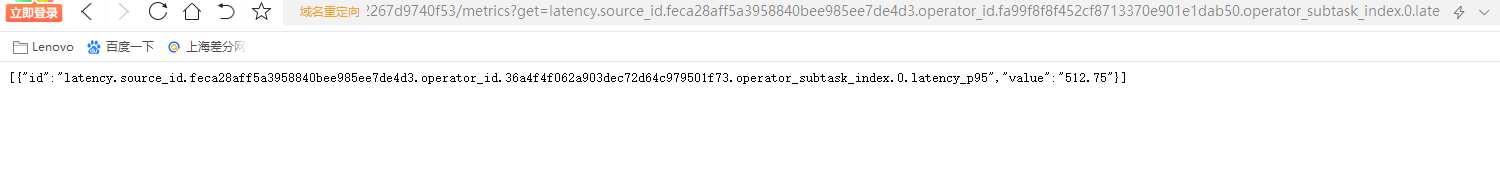
默认（全都不改变）



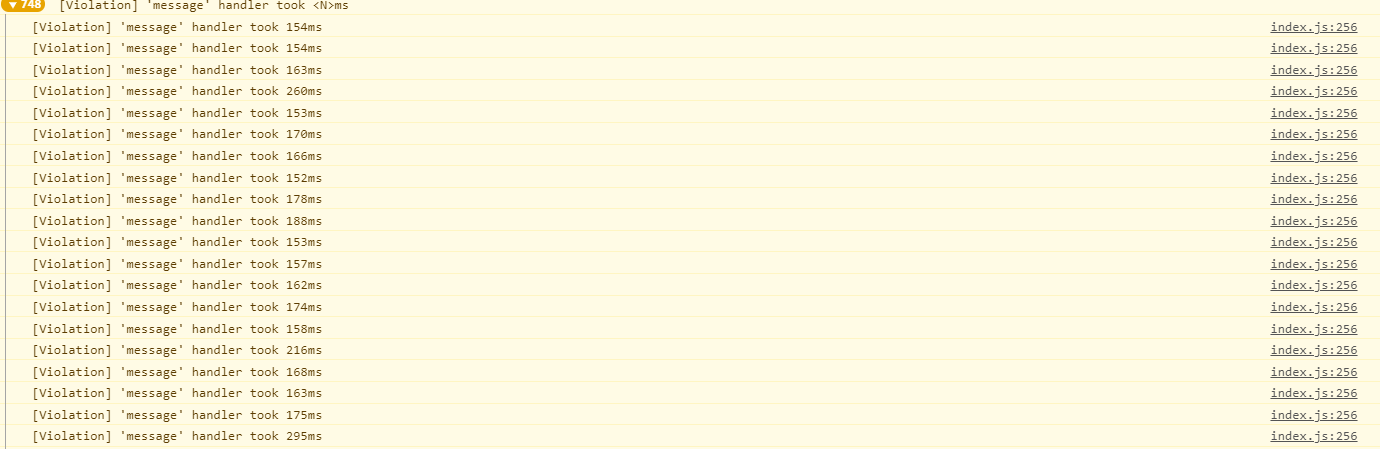
T1-T2 distribution的95分位数

Latency restful api

<http://localhost:6598/jobs/7bc044b8ab547d14c9902267d9740f53/metrics?get=latency.source_id.feca28aff5a3958840bee985ee7de4d3.operator_id.36a4f4f062a903dec72d64c979501f73.operator_subtask_index.0.latency_p95>

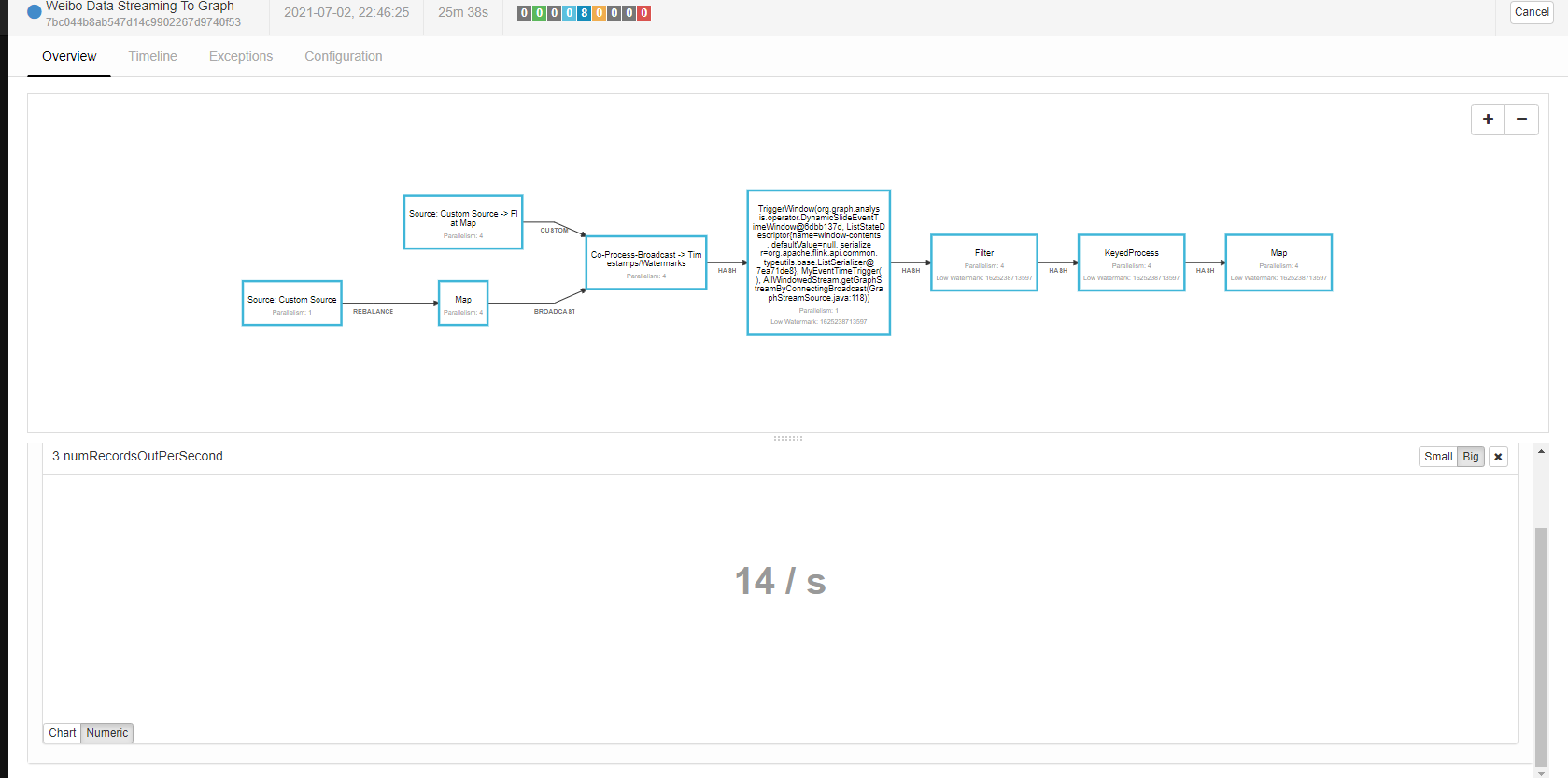


T2-t3

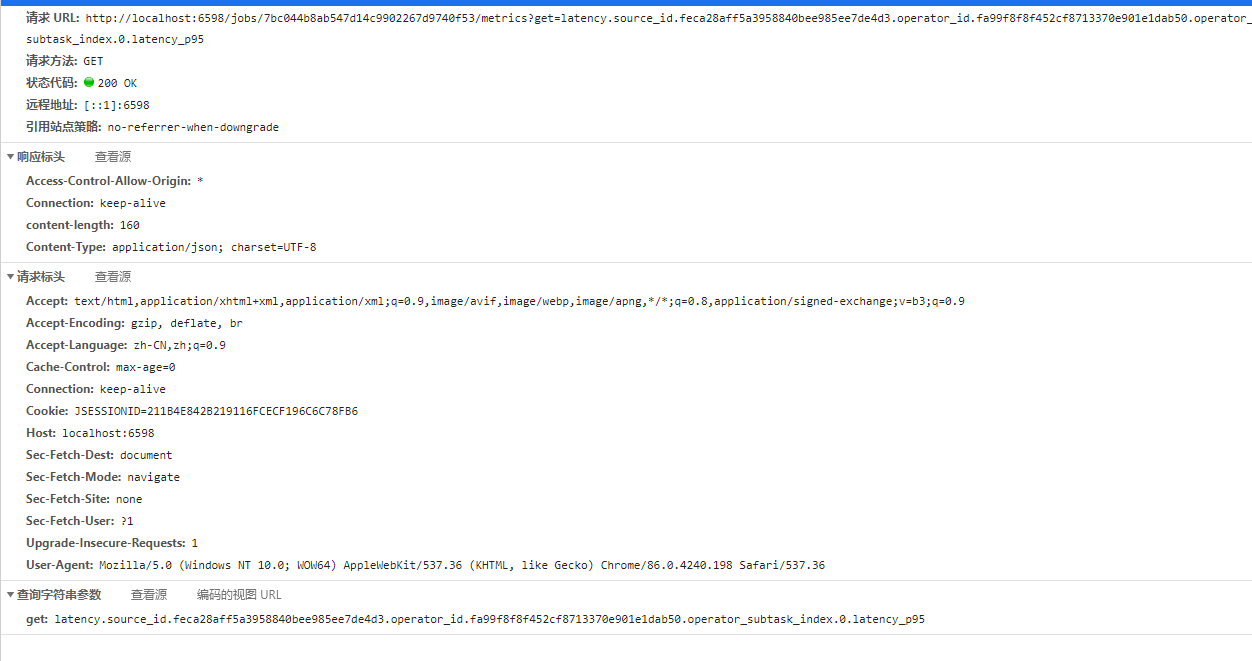


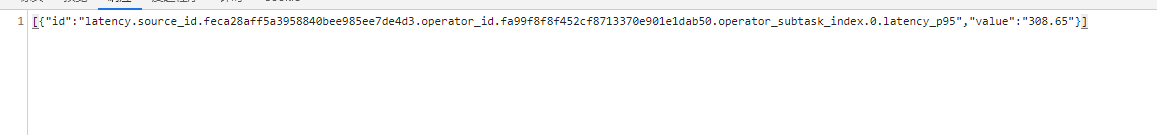
Filter功能：Vertex选择User；Edge选择Fans

吞吐量



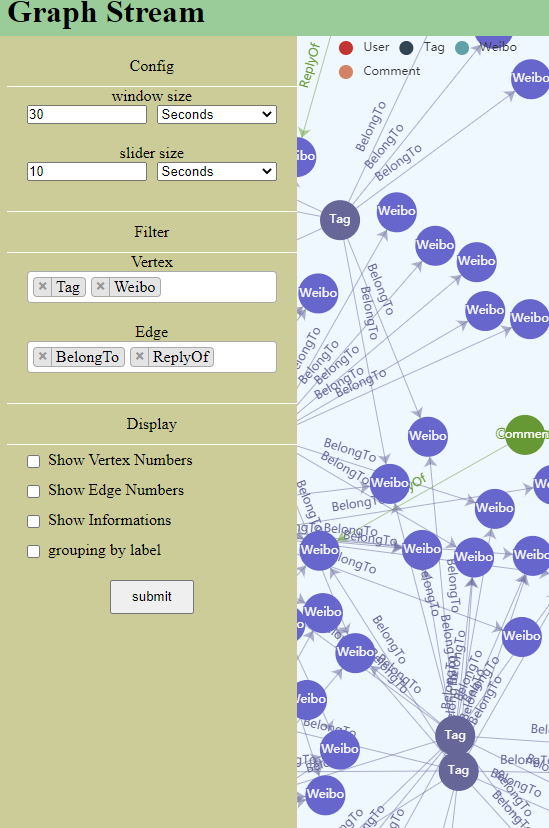
时延

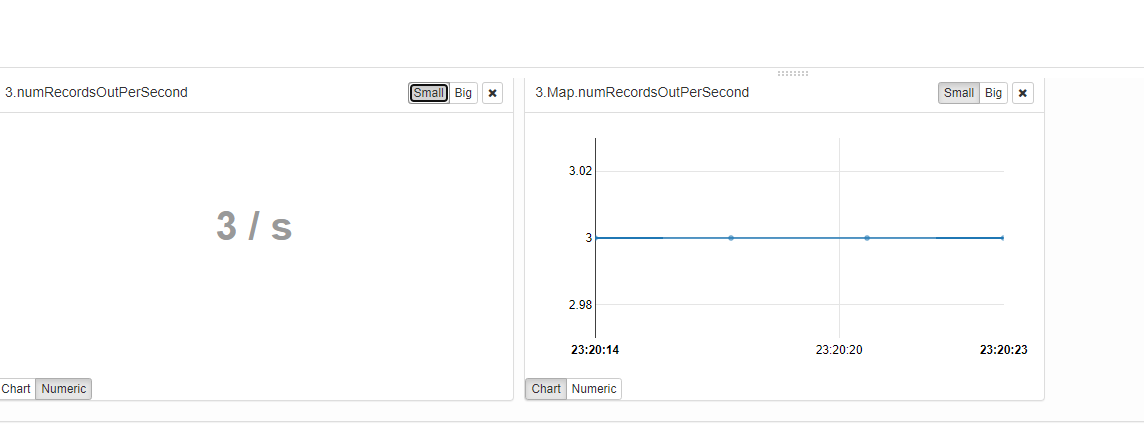


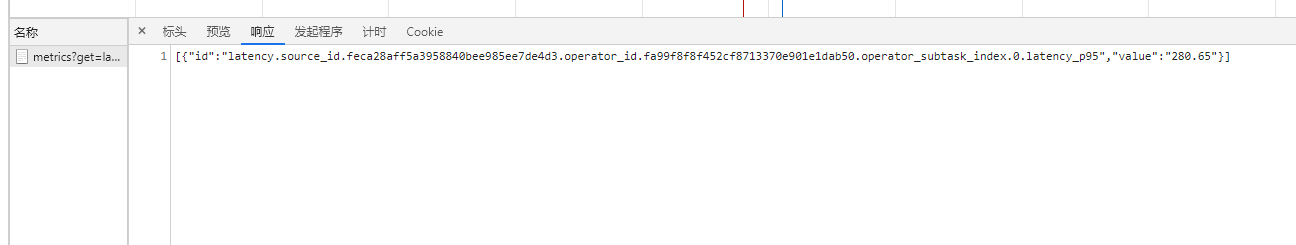


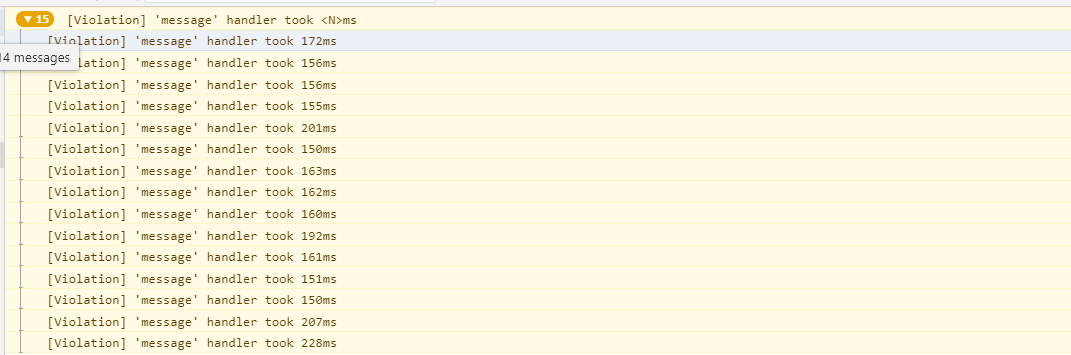


Filter功能：Vertex选择Tag和Weibo；Edge选择 BelongTo和ReplyOf



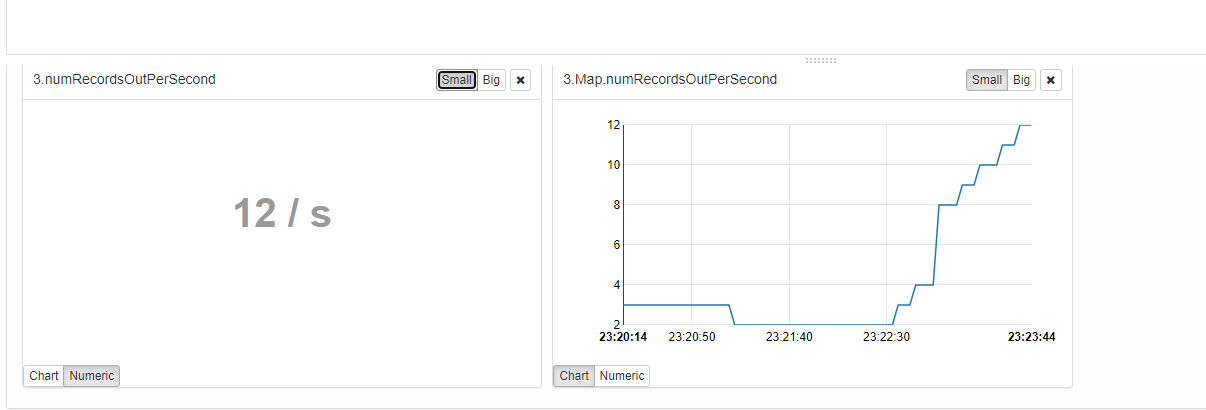




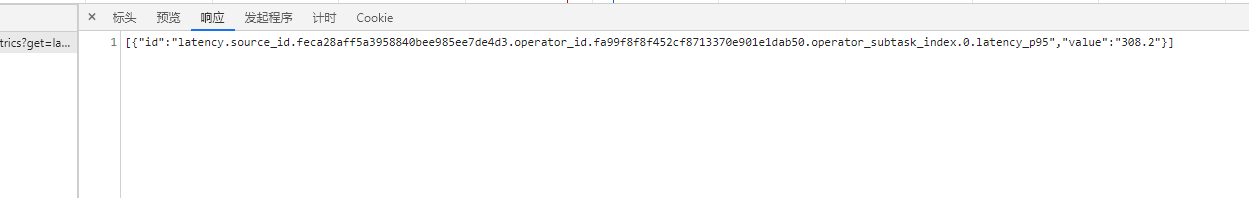


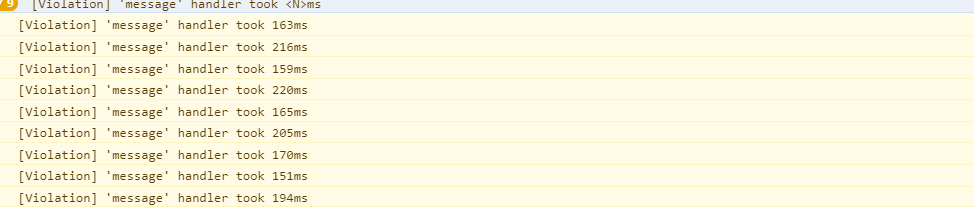
Gouping by label功能勾选

吞吐量



时延



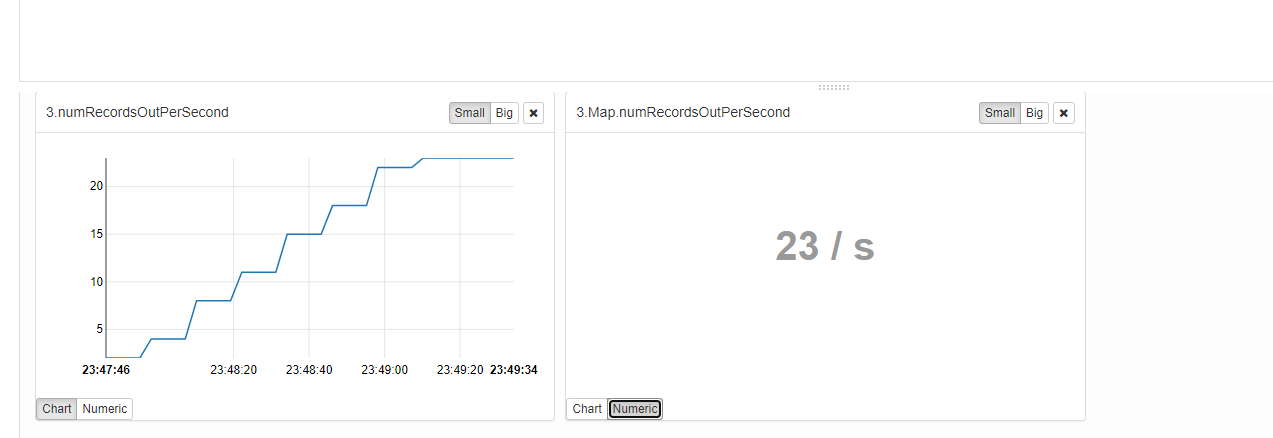


频率改为：10个元素/秒

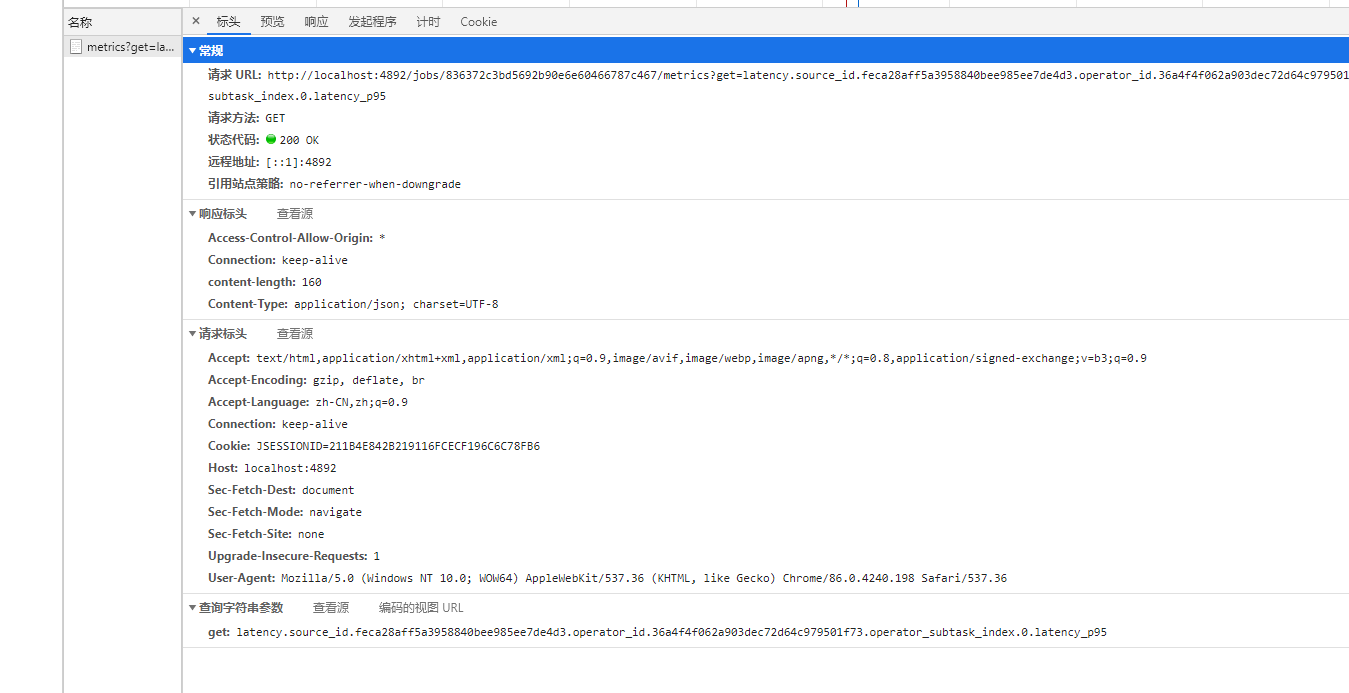
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 性能/功能 | 默认（全都不改变） | Filter功能：Vertex选择User；Edge选择Fans | Filter功能：Vertex选择Tag和Weibo；Edge选择 BelongTo和ReplyOf | Gouping by label功能勾选 |
| 吞吐量 | 23/s | 8/s | 8/s | 21/s |
| 延迟 | 515+200ms | 528+180ms | 528ms+184ms | 526.55+155ms |

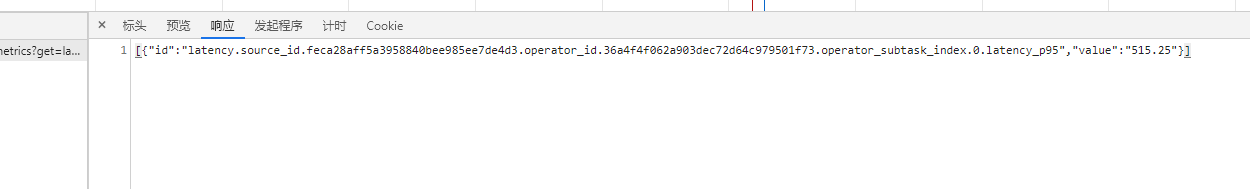
默认（全都不改变）

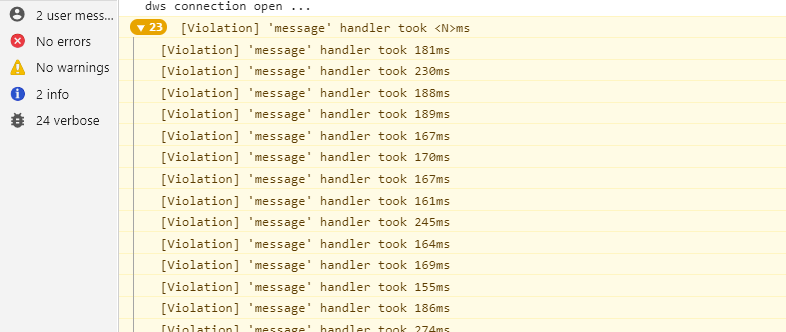
吞吐量



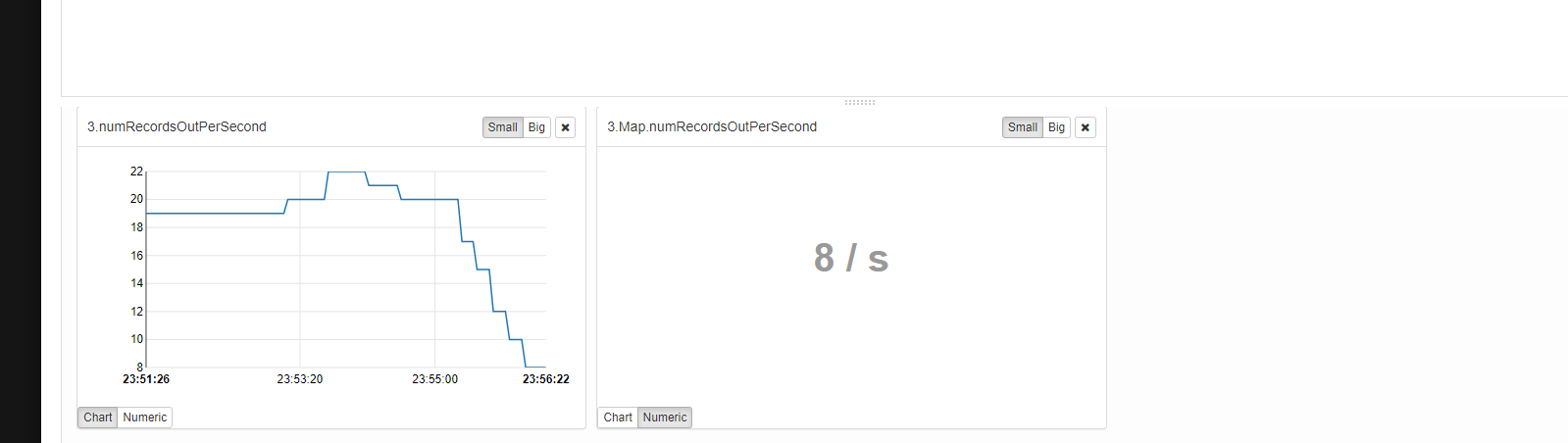
延时

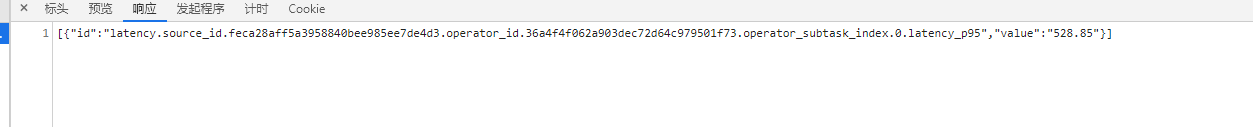


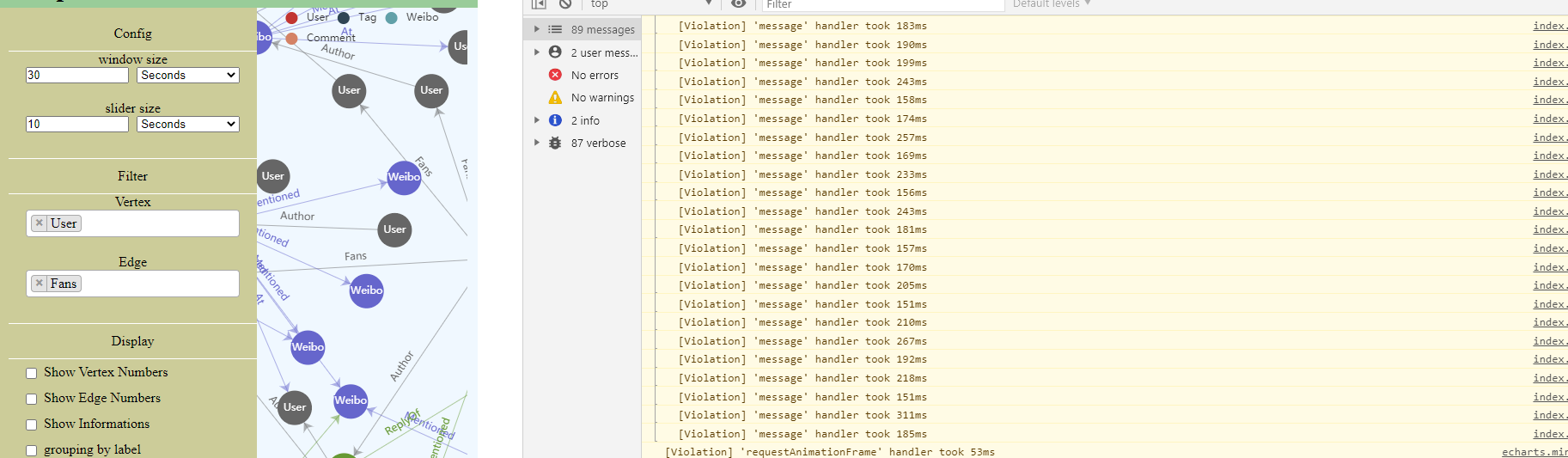




Filter功能：Vertex选择User；Edge选择Fans

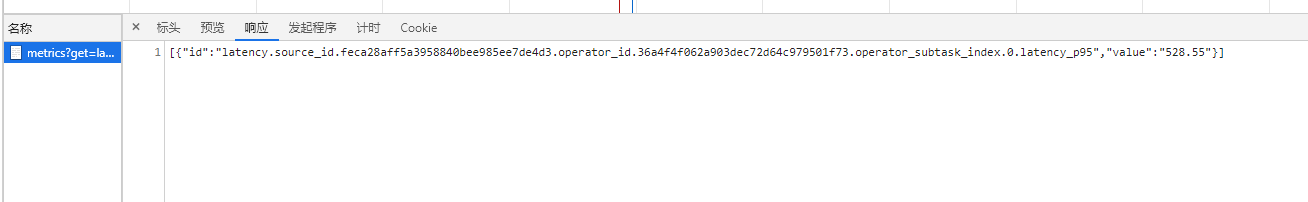


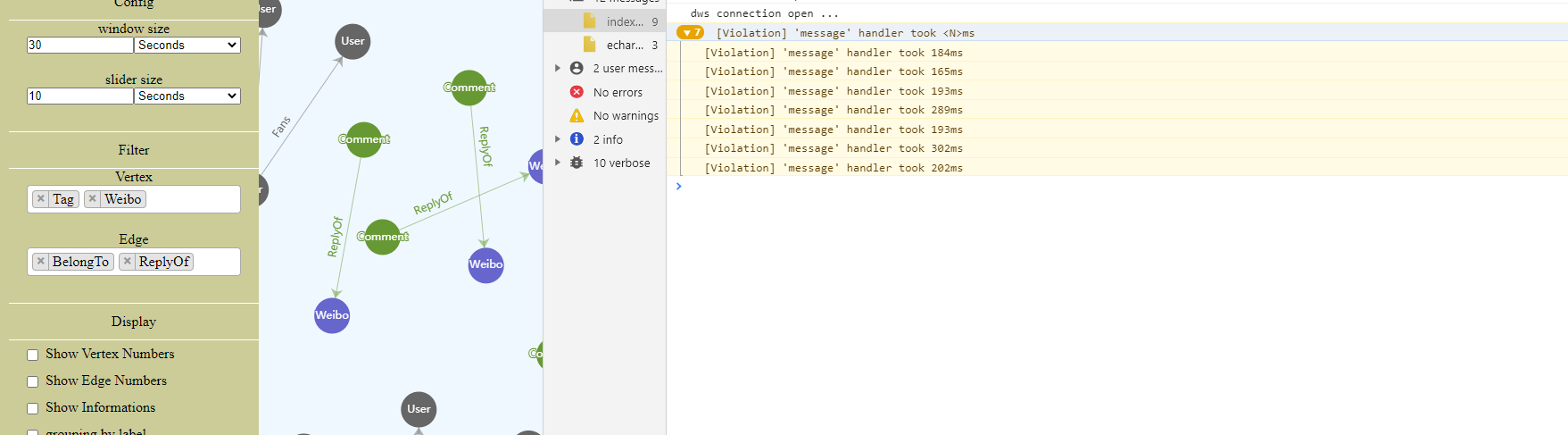




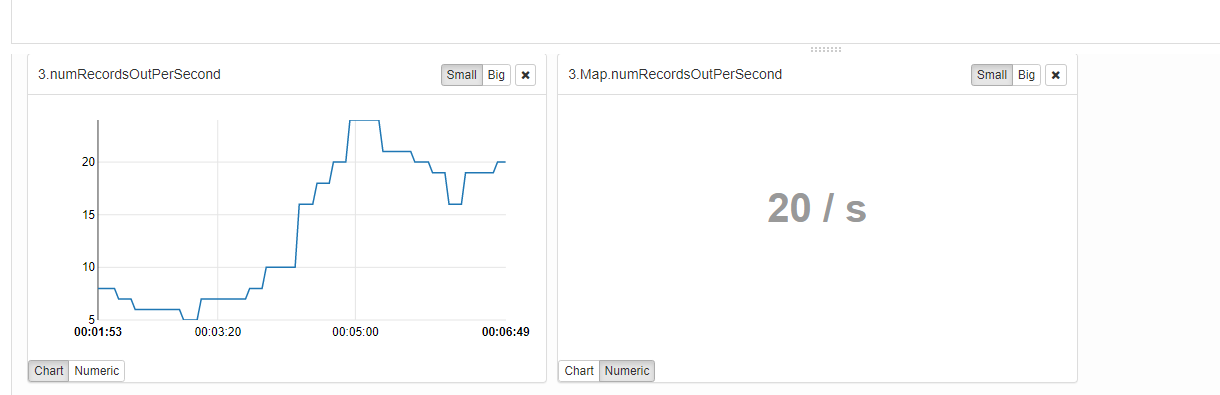
Filter功能：Vertex选择Tag和Weibo；Edge选择 BelongTo和ReplyOf

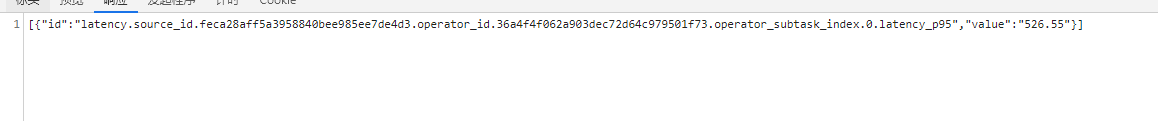


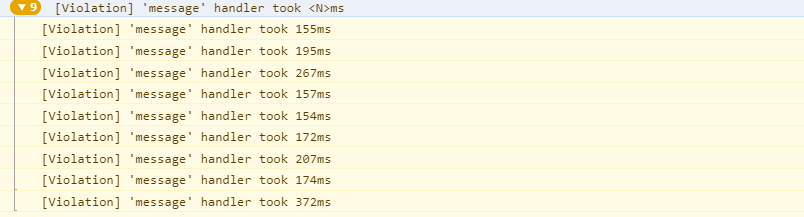




Gouping by label功能勾选

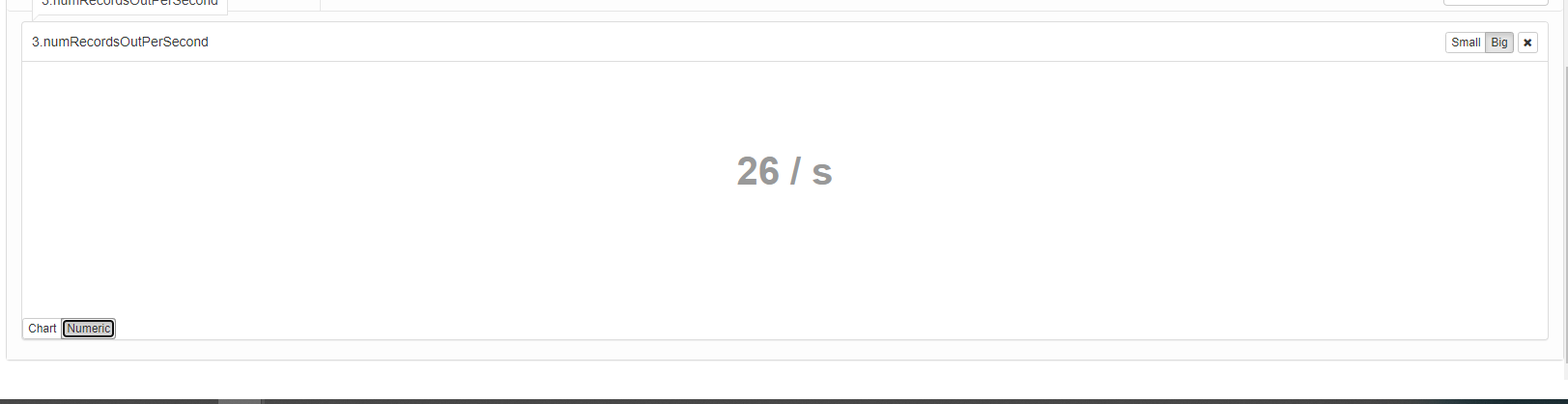


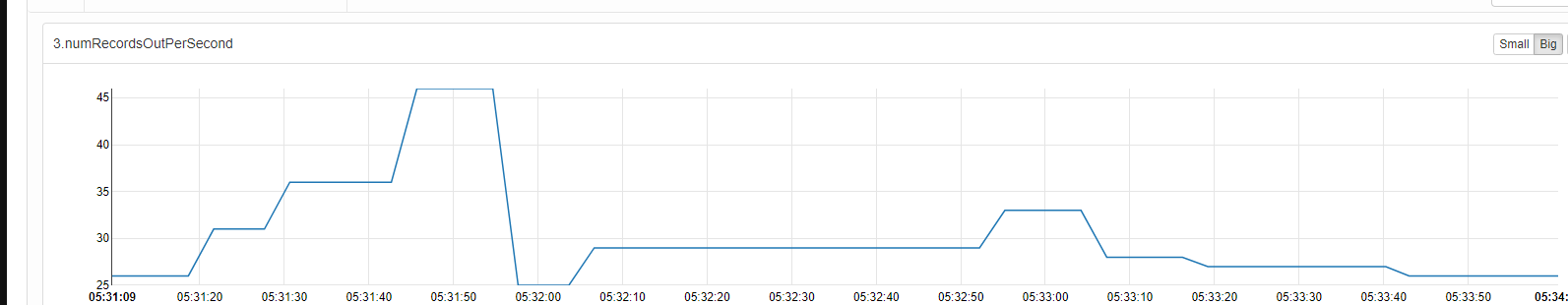


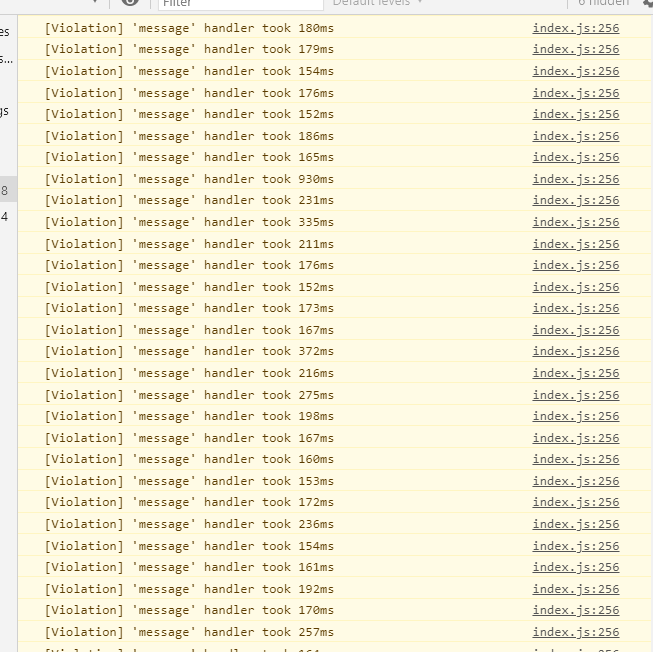


频率改为：15个元素/秒

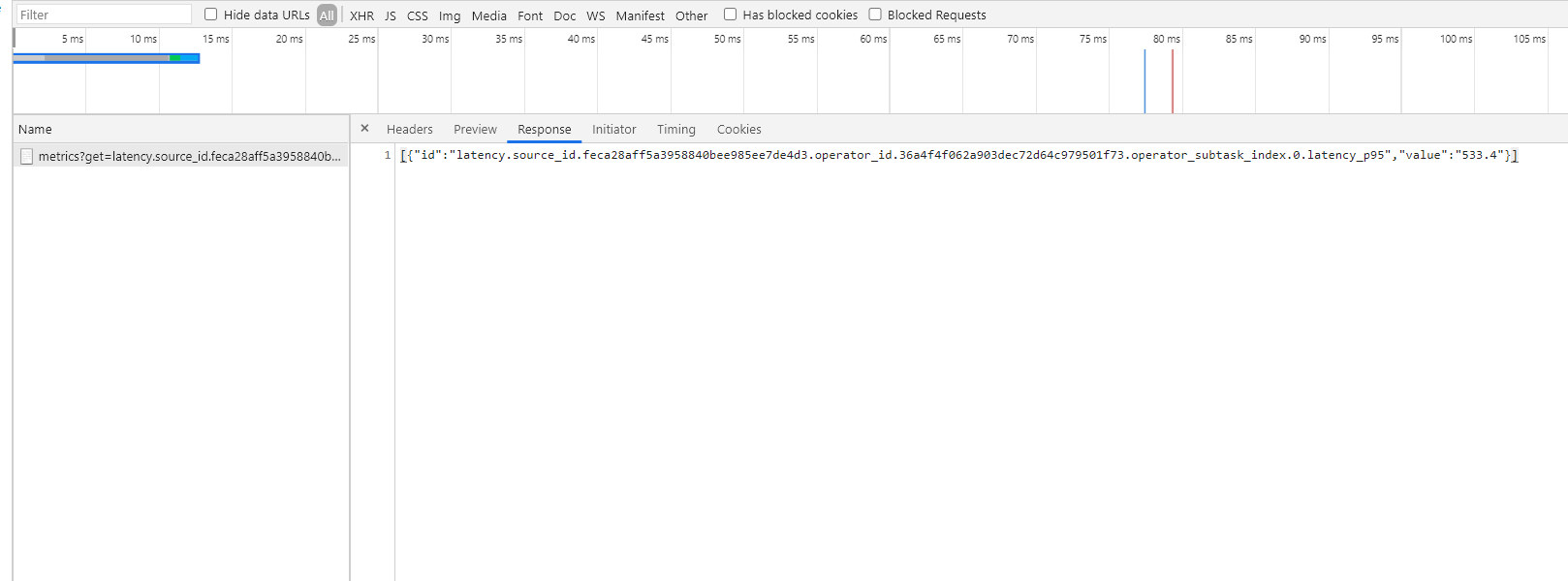
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 性能/功能 | 默认（全都不改变） | Filter功能：Vertex选择User；Edge选择Fans | Filter功能：Vertex选择Tag和Weibo；Edge选择 BelongTo和ReplyOf | Gouping by label功能勾选 |
| 吞吐量 | 26/s | 11/s | 11/s | 25/s |
| 延迟 | 533.04+200ms | 535ms + 200ms | 545+208ms | 533.2+184ms |



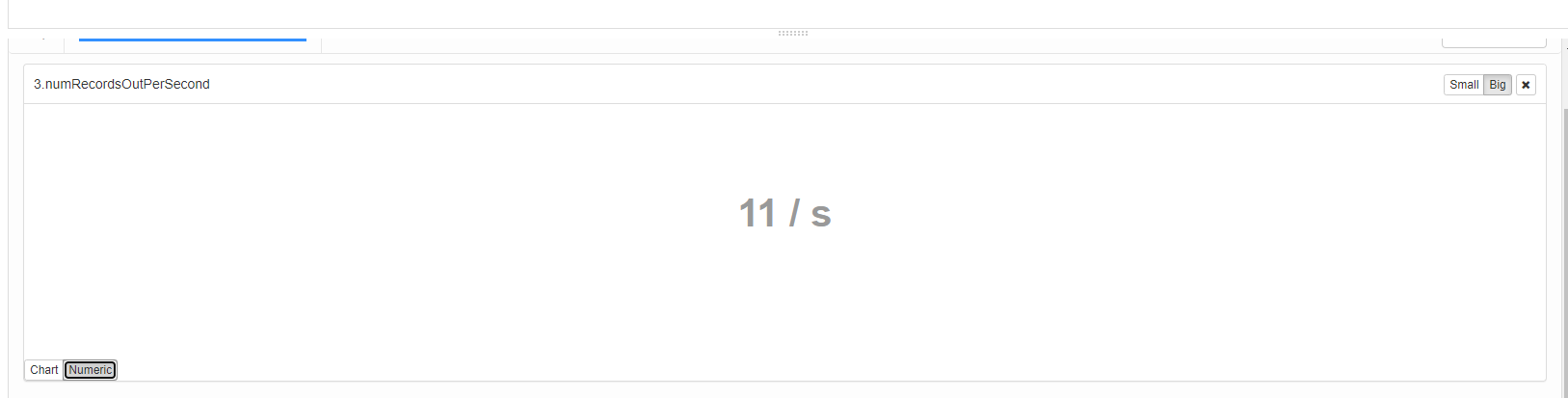


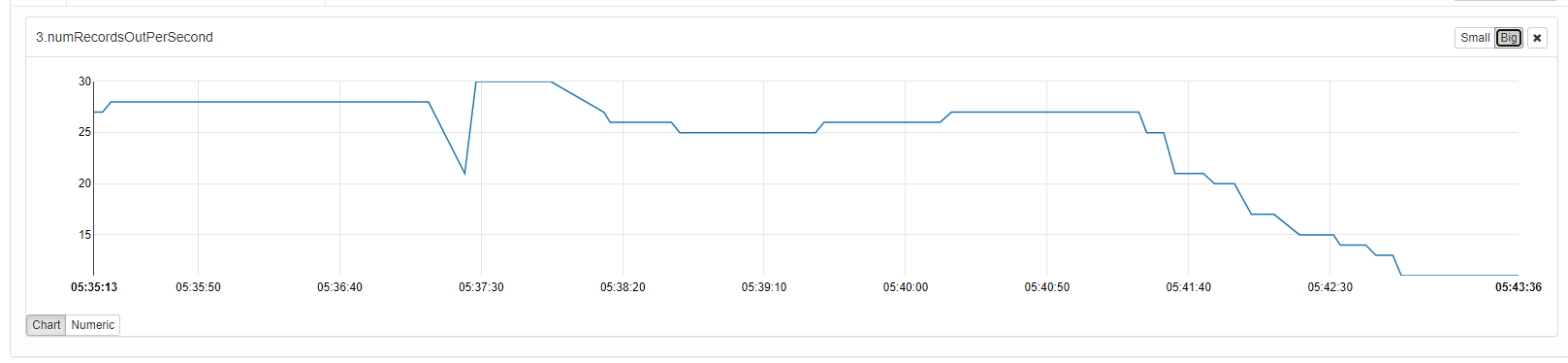


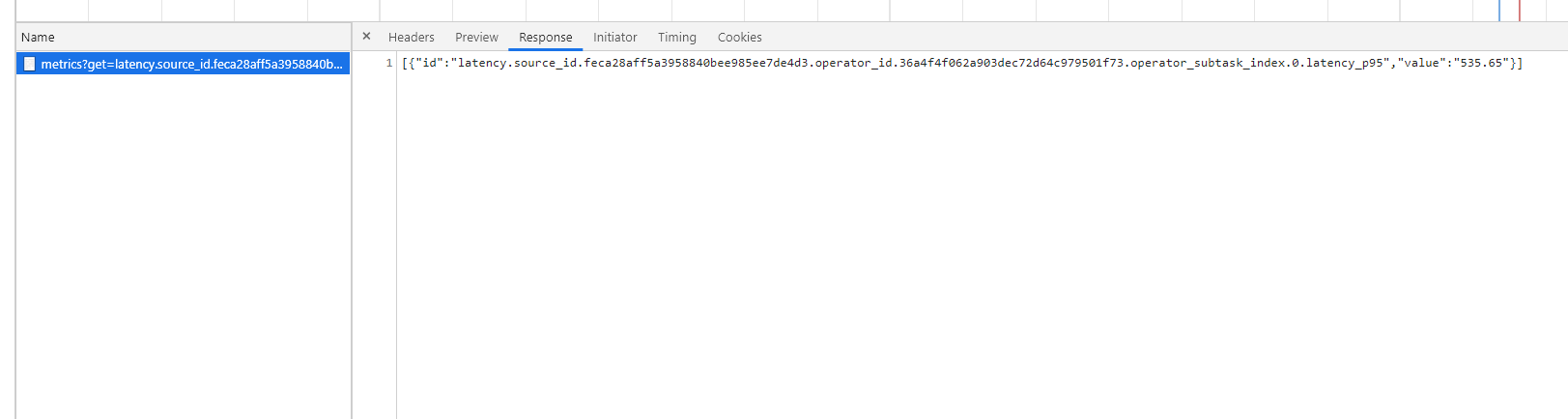
<http://localhost:6598/jobs/7bc044b8ab547d14c9902267d9740f53/metrics?get=latency.source_id.feca28aff5a3958840bee985ee7de4d3.operator_id.36a4f4f062a903dec72d64c979501f73.operator_subtask_index.0.latency_p95>

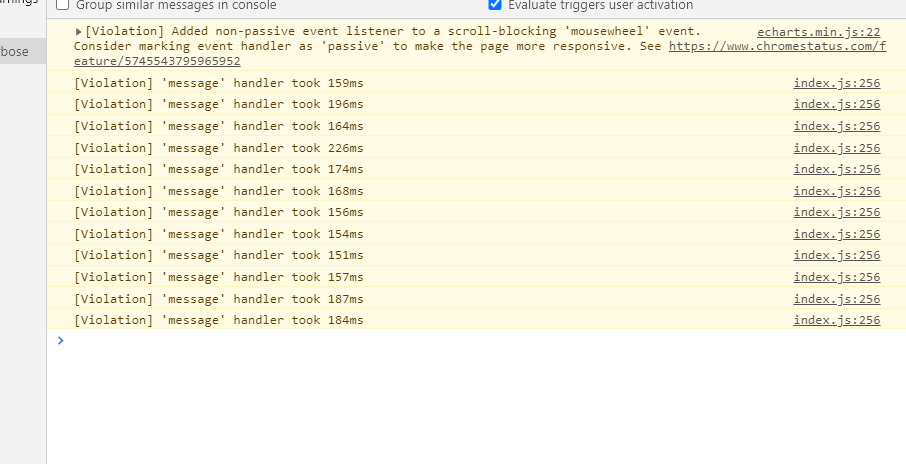


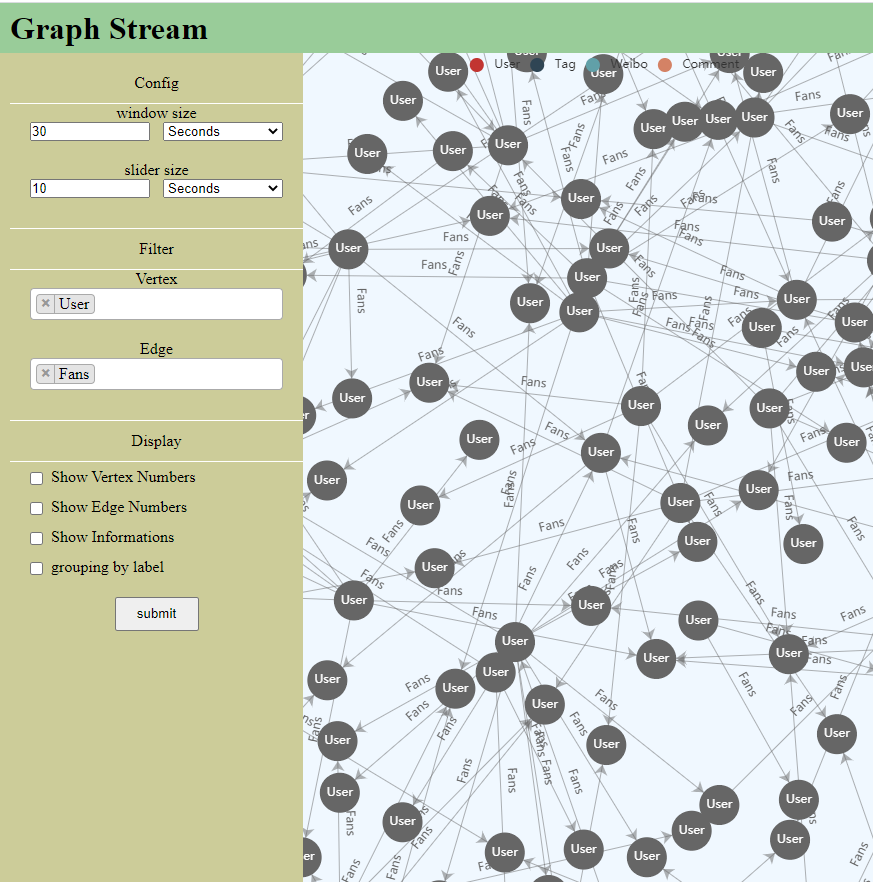
Filter功能：Vertex选择User；Edge选择Fans



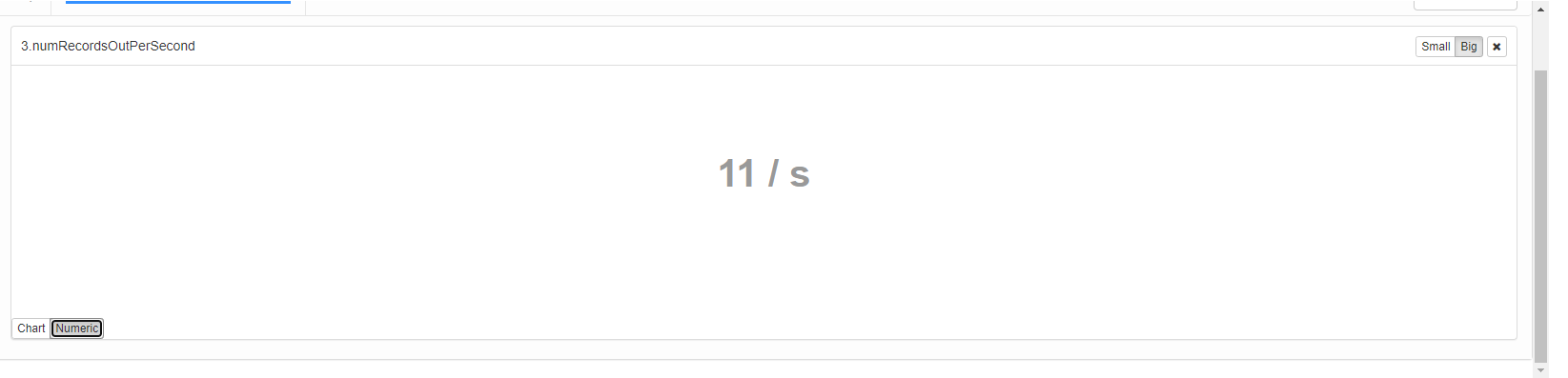


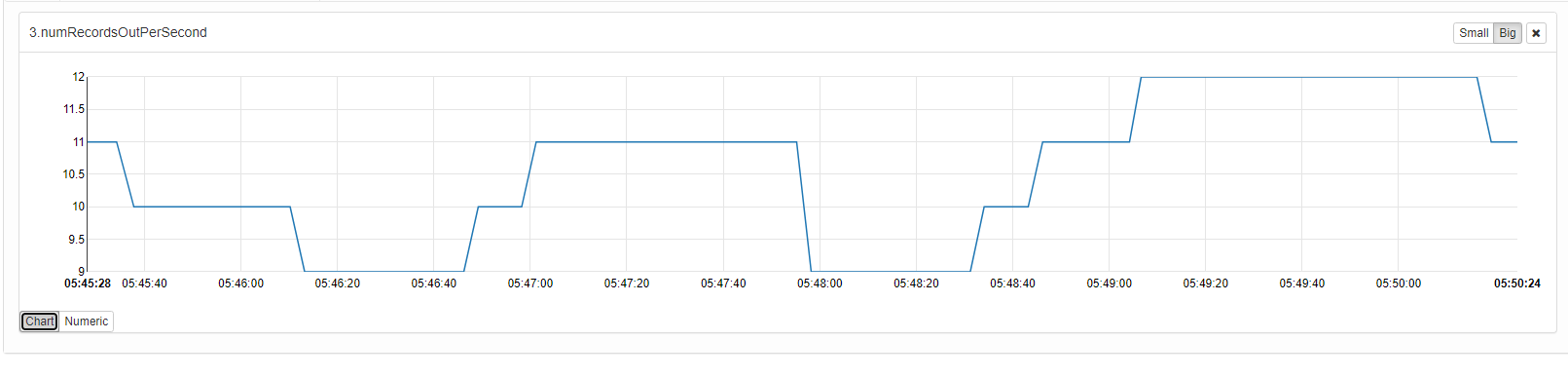


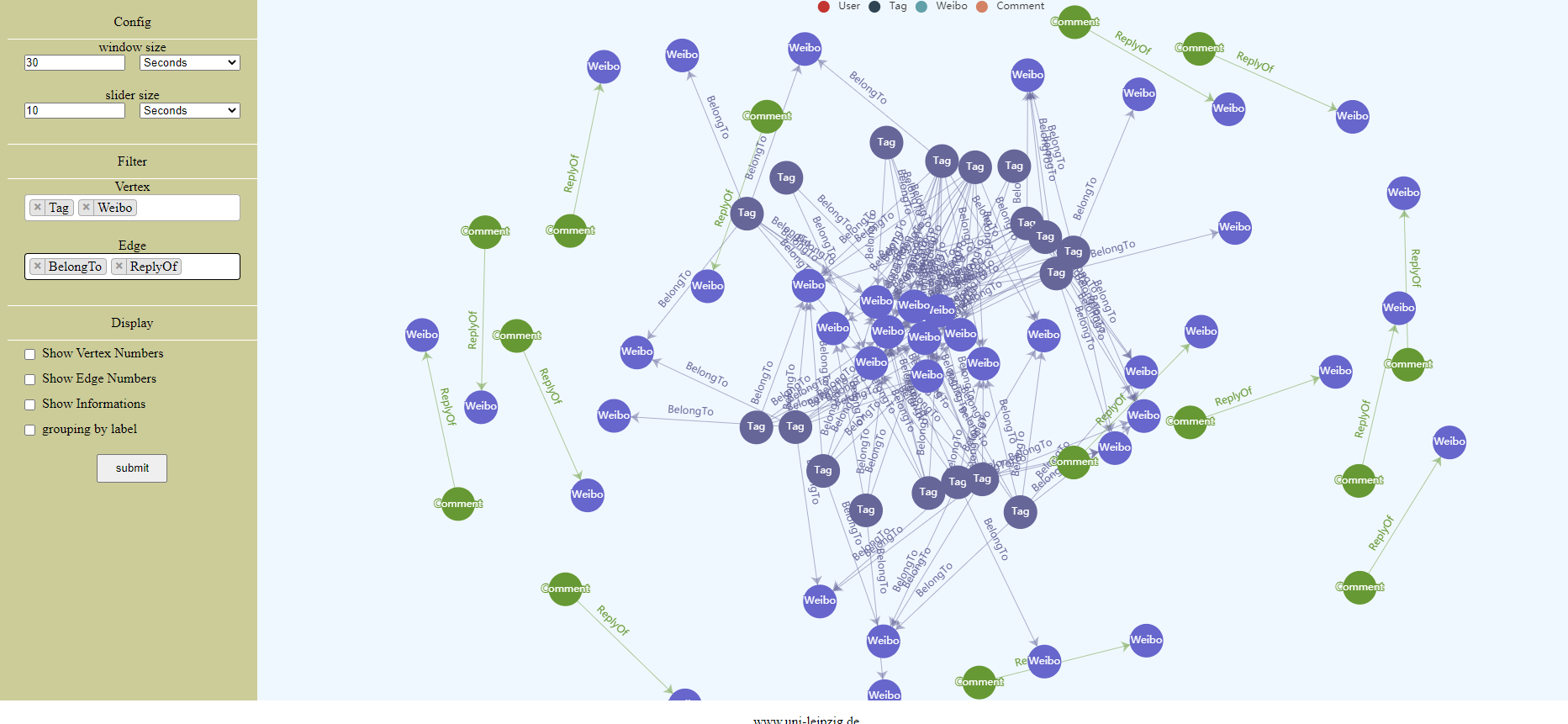


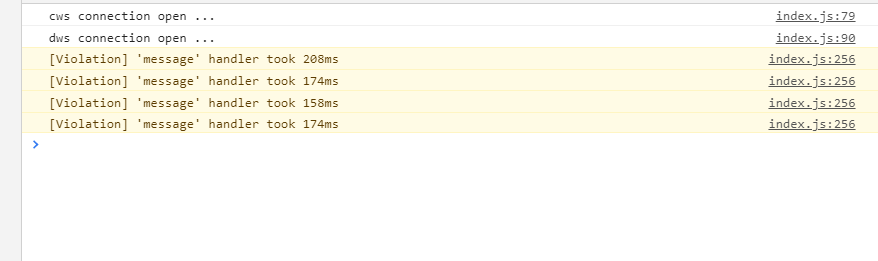


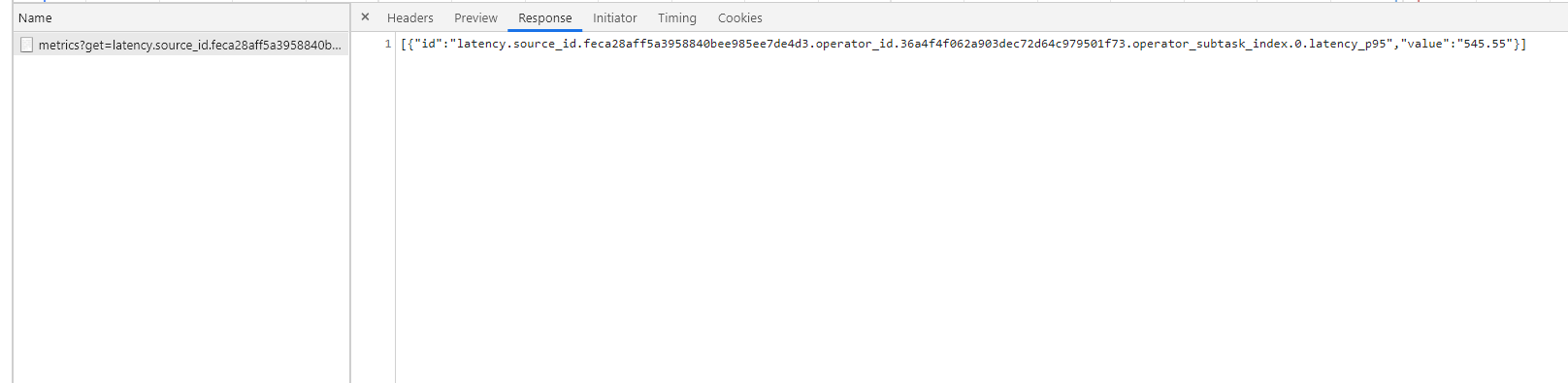
Filter功能：Vertex选择Tag和Weibo；Edge选择 BelongTo和ReplyOf



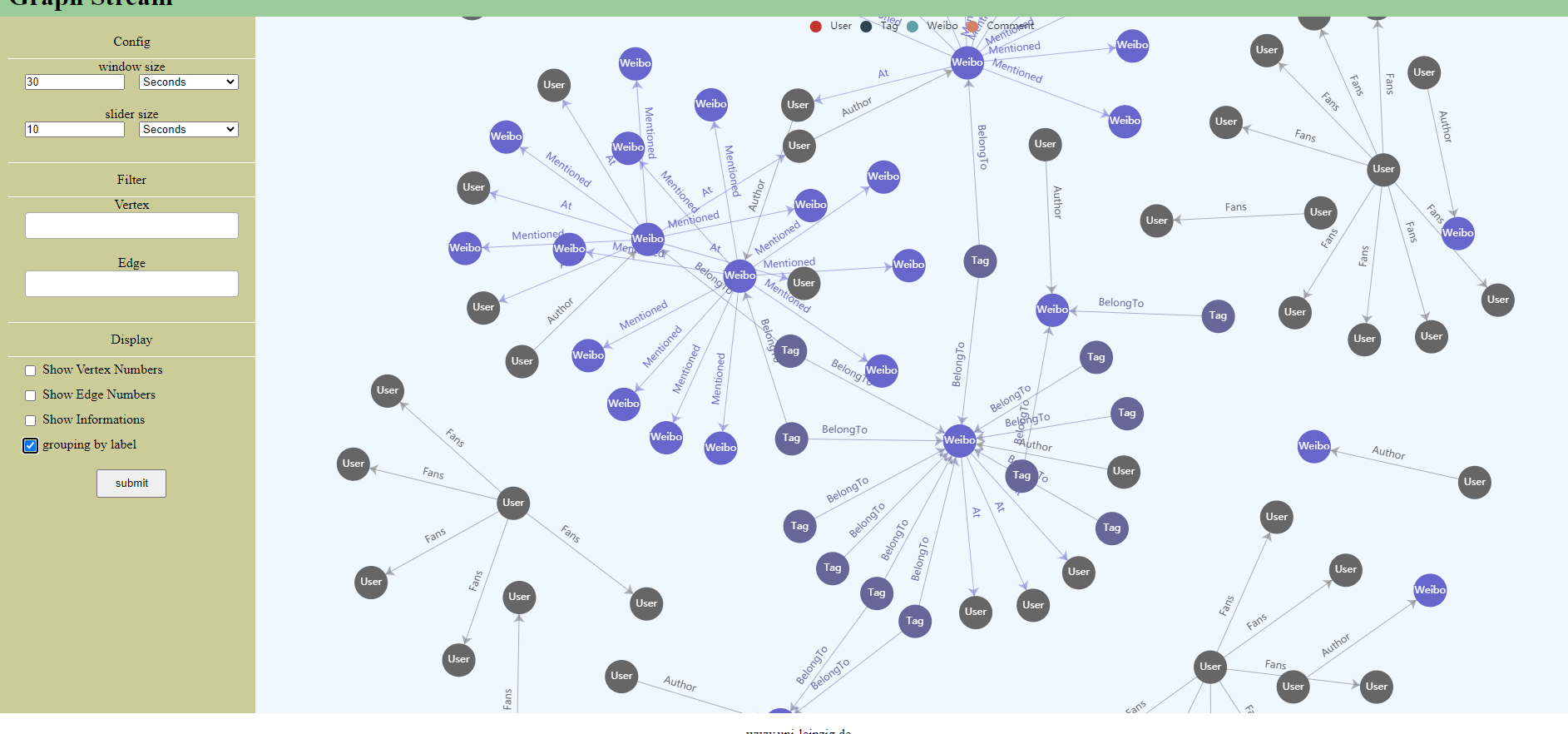


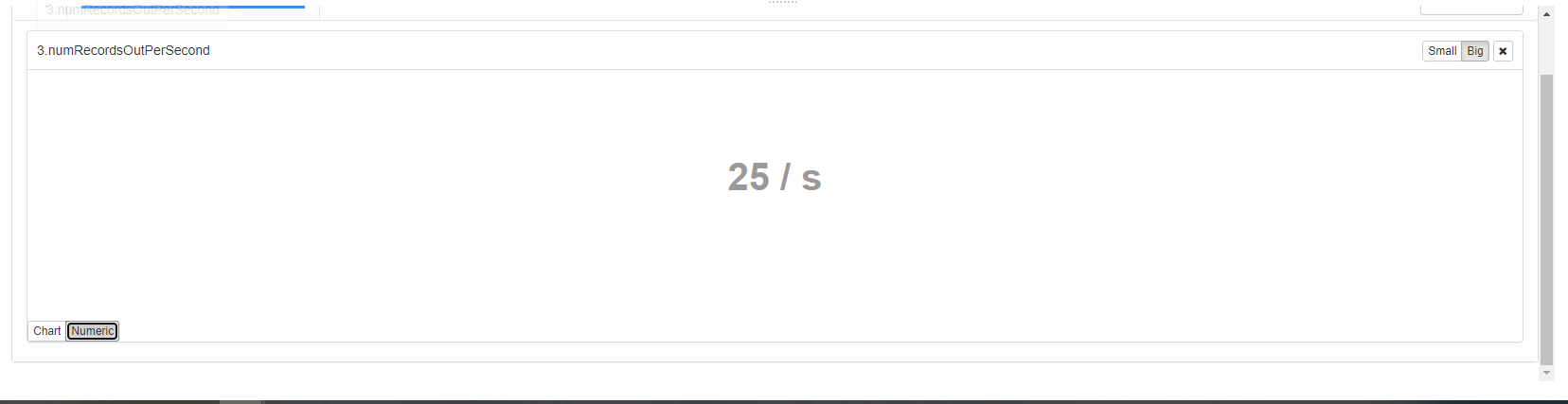


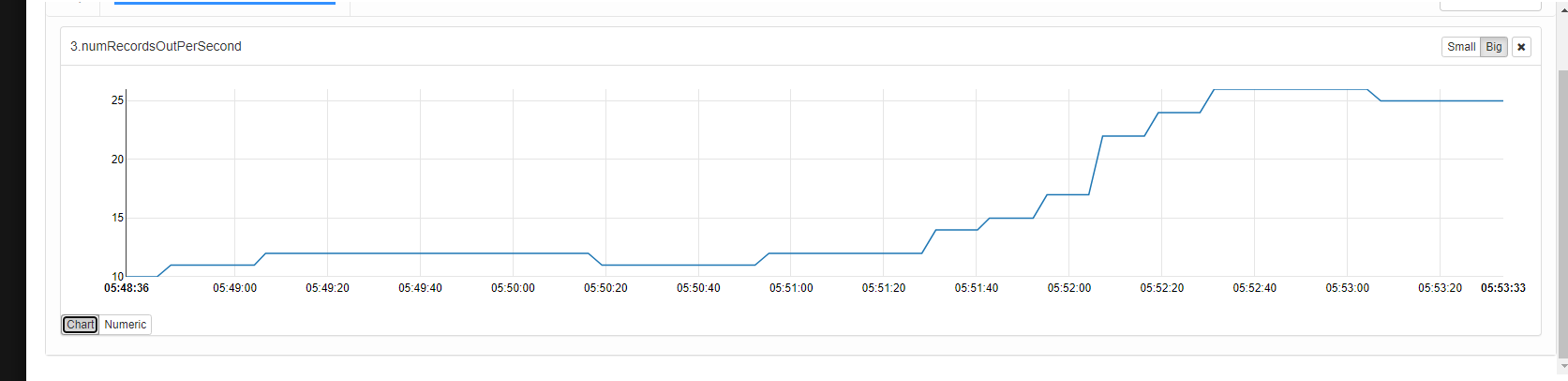


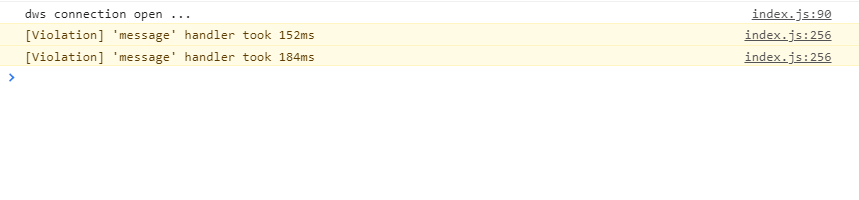


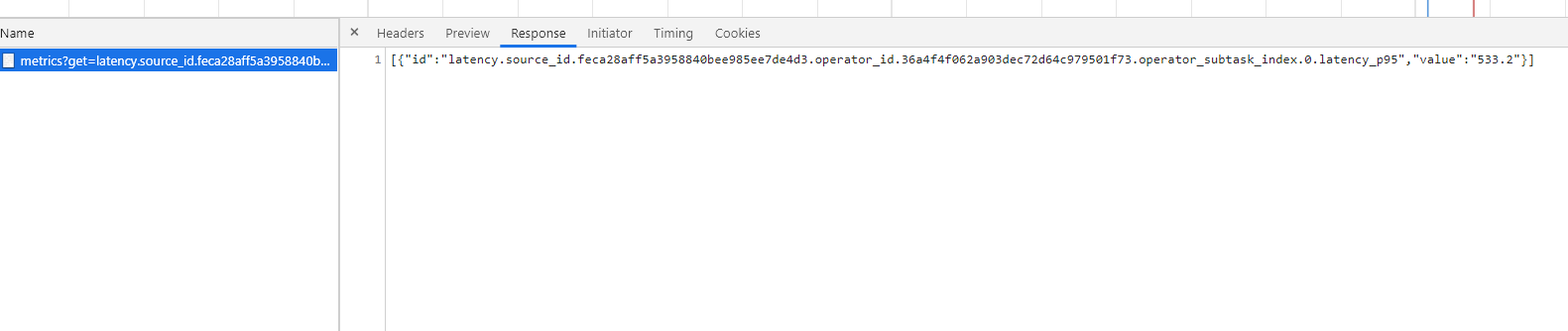
Gouping by label功能勾选





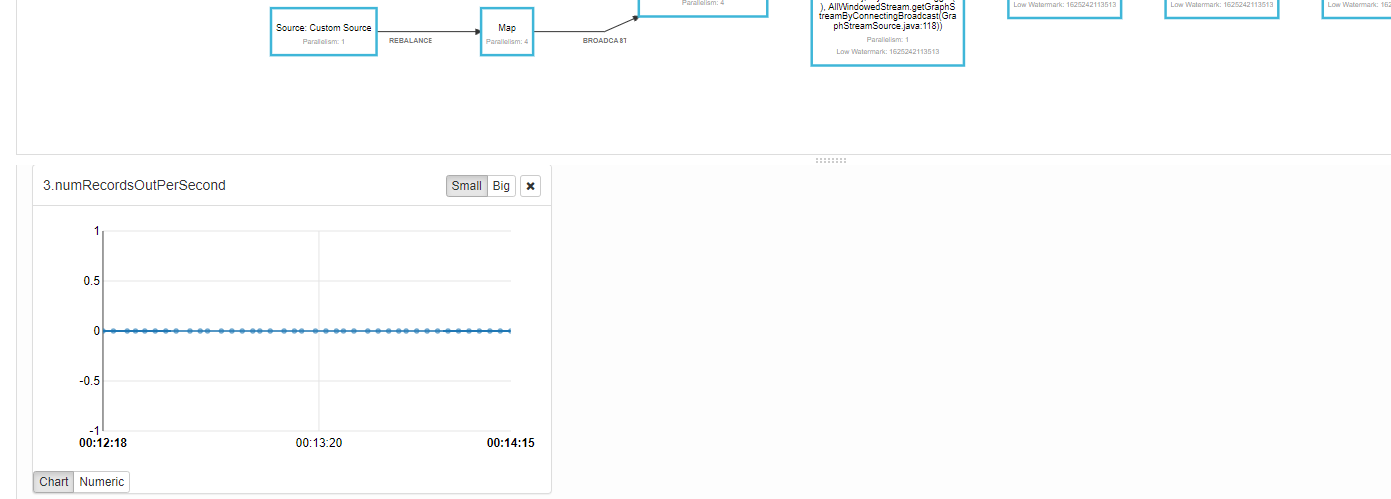






频率改为：20个元素/秒

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 性能/功能 | 默认（全都不改变） | Filter功能：Vertex选择User；Edge选择Fans | Filter功能：Vertex选择Tag和Weibo；Edge选择 BelongTo和ReplyOf | Gouping by label功能勾选 |
| 吞吐量 | 0/s | 0/s | 0/s | 0/s |
| 延迟 |  |  |  |  |



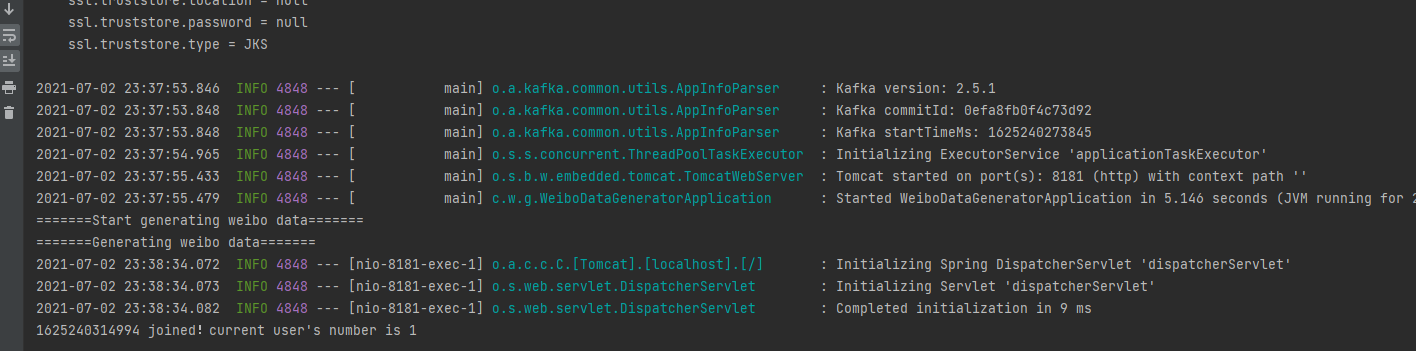
不工作了已经

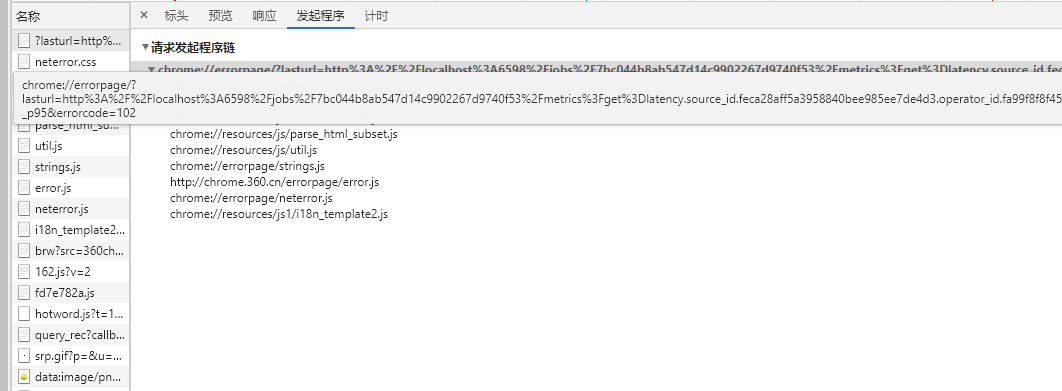
50元素也是爆掉

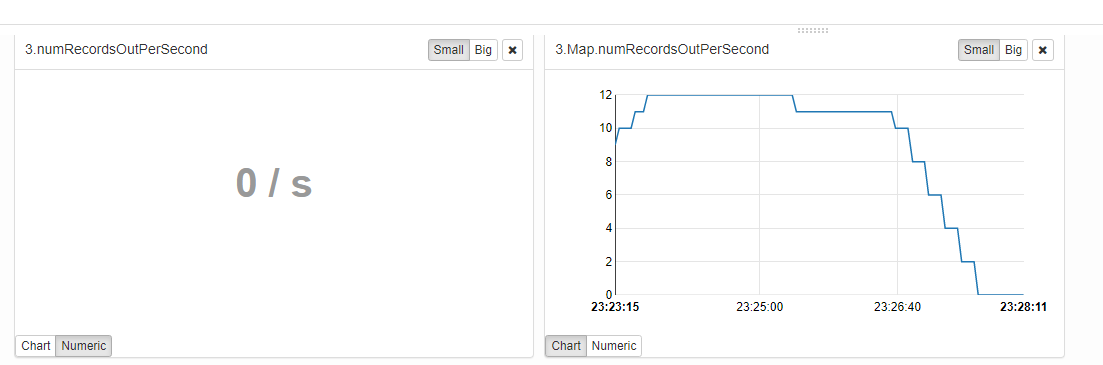
频率改为：100个元素/秒

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 性能/功能 | 默认（全都不改变） | Filter功能：Vertex选择User；Edge选择Fans | Filter功能：Vertex选择Tag和Weibo；Edge选择 BelongTo和ReplyOf | Gouping by label功能勾选 |
| 吞吐量 |  |  |  |  |
| 延迟 |  |  |  |  |

直接卡死







一下均为

默认（全都不改变）

Filter功能：Vertex选择User；Edge选择Fans

Filter功能：Vertex选择Tag和Weibo；Edge选择 BelongTo和ReplyOf

Gouping by label功能勾选

频率改为：1000个元素/秒

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 性能/功能 | 默认（全都不改变） | Filter功能：Vertex选择User；Edge选择Fans | Filter功能：Vertex选择Tag和Weibo；Edge选择 BelongTo和ReplyOf | Gouping by label功能勾选 |
| 吞吐量 |  |  |  |  |
| 延迟 |  |  |  |  |

可扩展性：有了延迟和吞吐量的数值，考虑例如从1台机器到2、4、8、16台机器，数值是如何变化的。在最好的情况下，处理速度翻倍，在最坏的情况下，它保持不变或变得更糟。

频率为：100个元素/秒，2台机器下的：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 性能/功能 | 默认（全都不改变） | Filter功能：Vertex选择User；Edge选择Fans | Filter功能：Vertex选择Tag和Weibo；Edge选择 BelongTo和ReplyOf | Gouping by label功能勾选 |
| 吞吐量 |  |  |  |  |
| 延迟 |  |  |  |  |

4台机器下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 性能/功能 | 默认（全都不改变） | Filter功能：Vertex选择User；Edge选择Fans | Filter功能：Vertex选择Tag和Weibo；Edge选择 BelongTo和ReplyOf | Gouping by label功能勾选 |
| 吞吐量 |  |  |  |  |
| 延迟 |  |  |  |  |

8台机器下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 性能/功能 | 默认（全都不改变） | Filter功能：Vertex选择User；Edge选择Fans | Filter功能：Vertex选择Tag和Weibo；Edge选择 BelongTo和ReplyOf | Gouping by label功能勾选 |
| 吞吐量 |  |  |  |  |
| 延迟 |  |  |  |  |

16台机器下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 性能/功能 | 默认（全都不改变） | Filter功能：Vertex选择User；Edge选择Fans | Filter功能：Vertex选择Tag和Weibo；Edge选择 BelongTo和ReplyOf | Gouping by label功能勾选 |
| 吞吐量 |  |  |  |  |
| 延迟 |  |  |  |  |