作业9

519021910913 黄喆敏

1. 请阐述日志结构数据库适合什么样的应用场景？(1分)

答：

（1）适用于**写多读少**，对写性能要求较高的场景。

（2）适用于需要**变长键值对存储**的场景。

（3）适用于**经常访问热数据**的场景，此时数据被存储在内存中，或者上层的SSTable中，查询速度较快。

1. 请阐述日志结构数据库中的读放大和写放大分别是什么意思？(1分)

答：读放大指每次读请求所需的I/O操作数较高。在LSM Tree中，首先从MemTable中查询键值对。若查找不到，则从上到下，遍历SSTable，找出键值对。最坏的情况下，LSM Tree中**没有该键值对**。此时，需要遍历所有的SSTable，才能判断不存在，造成非常严重的读放大。

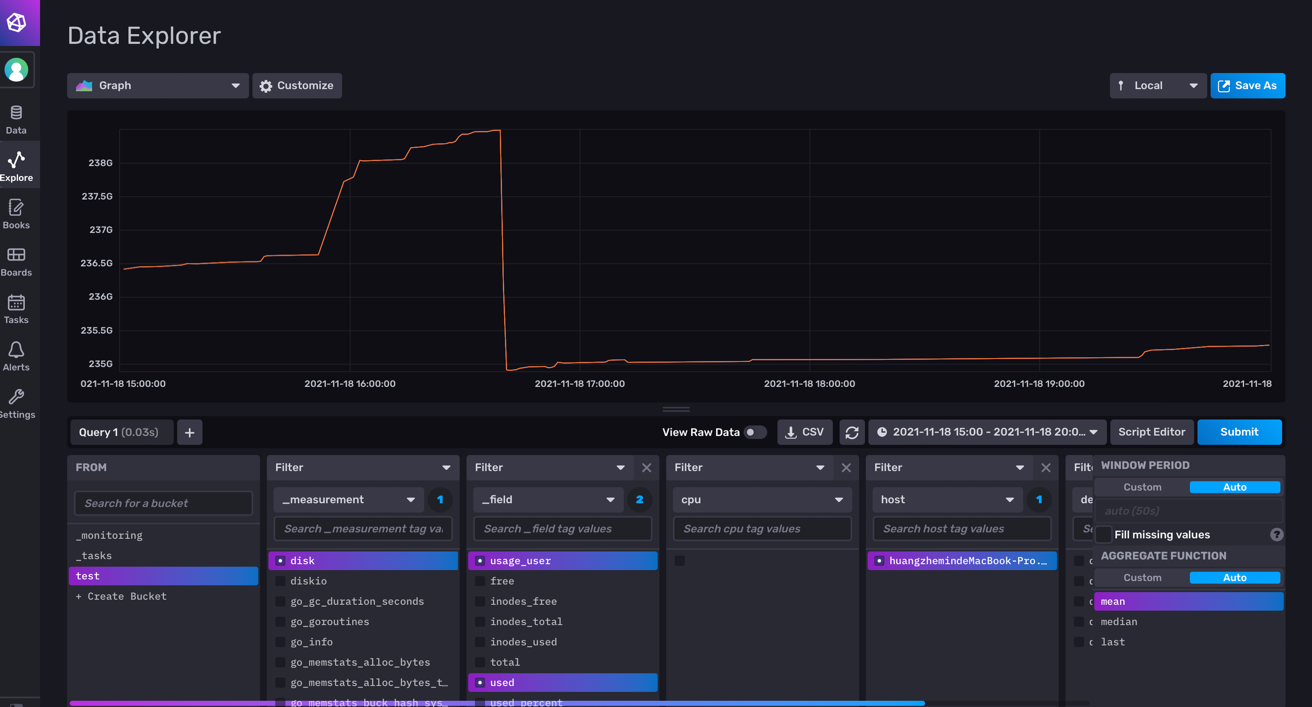
写放大指实际写入的物理数据量是写入数据量的多倍。在LSM Tree中，若MemTable存储大小超过阈值，则会写入SSTable内。而当上层的SSTable存满，则会进行compaction操作，将部分数据存到下一层中。而在部分情况下，**插入一个键值对，会引发多次compaction**，此时写放大非常严重。

1. 日志结构合并树中，WAL的作用是什么？ (1分)

答：根据LSM Tree的实现原理，数据先写到MemTable，即内存中。如果服务器崩溃，则数据会丢失。因此，需要采用预写式日志（Write-ahead log），保证服务器奔溃重启时，可以从log文件中恢复数据。

1. 请你在自己的机器上安装 InfluxDB，并像课程上所演示的一样监控你的笔记本电脑的状态，在Web界面的Explore中截图贴在Word文档中，并根据截图简要说明一下你的笔记本电脑的运行状态。 (2分)

答：我们在15：50时开始下载电影，并在16：40时将其删掉。可以发现，磁盘的占用量会减小3G左右，如下图所示。



我们对CPU的空闲百分比进行监测。可以看到，17：20-18：00左右，系统处于待机状态，因此空闲百分比较高。

