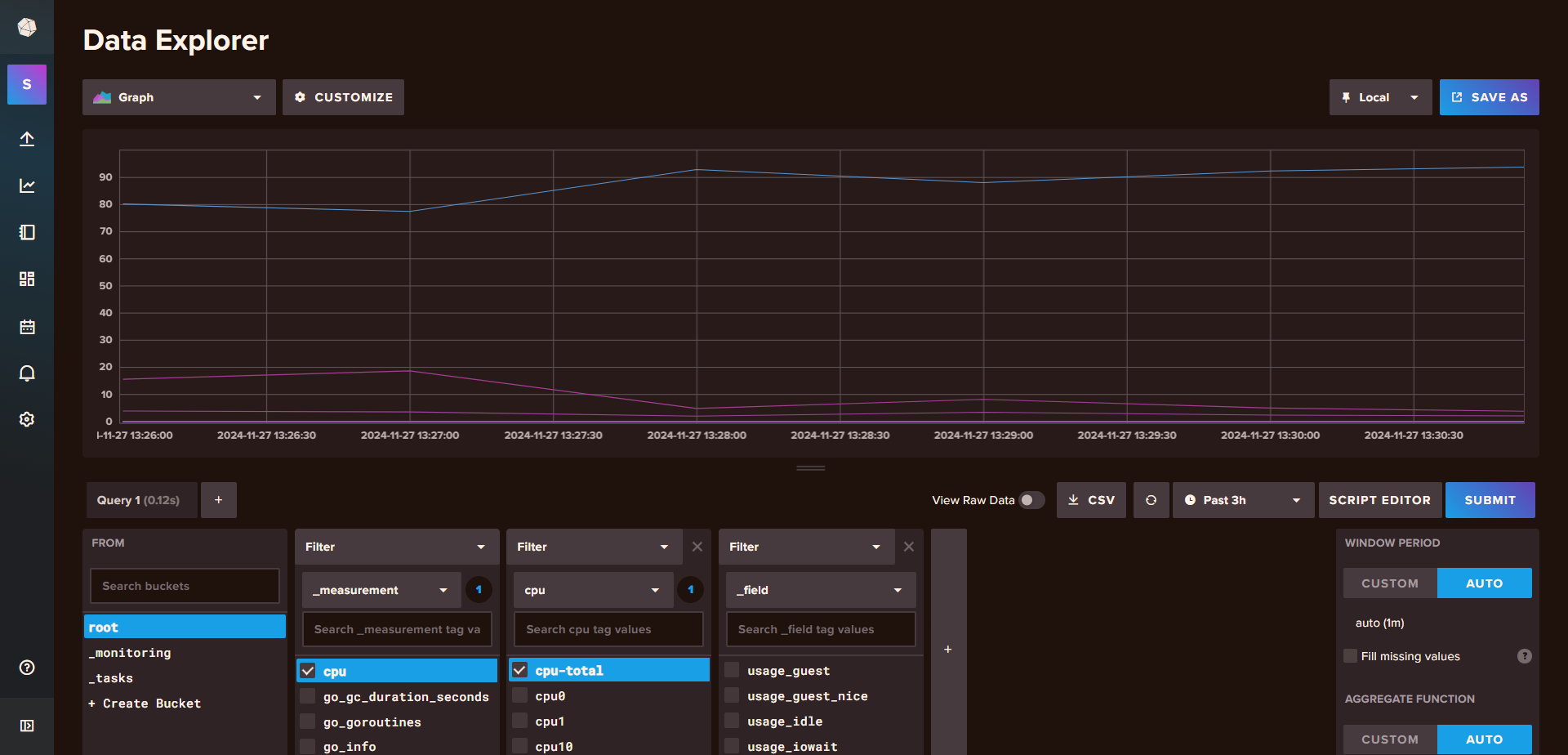
应用系统体系架构 — 作业9

# 学号： 522031910213 姓名： 朱涵 得分：

A.正确运行InfluxDB并正确解释运行结果

答：

如下图是某一时间段内CPU占用的曲线图。其中变化较明显的两条线是usage\_idle(空闲CPU百分比)与usage\_user（用户占用CPU百分比）



可以看到13:26-13:27的时间段内用户的CPU占用短暂提高，这是因为我在此时打开了游戏进程，占用了一部分CPU资源。后续关闭游戏，CPU资源占用基本在10%以下波动，因为电脑基本没有进程需要CPU资源。

## B.正确给出问题的答案

## 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？

## 答：

## 读放大是指实际读取磁盘上的数据量是用户读取到的数据大小的多倍，浪费了大量的IO资源。

## 在日志结构数据库中，读操作需要从上到下一层一层的对SSTable进行查找，直到找到想要的数据。假如读取的数据比较冷，已经下沉到了最底层；甚至可能不存在于数据库中，读操作一直进行到最底层，导致非常多的 IO，也就带来了读放大。解决方法通常有布隆过滤器、行列混合存储等。

## 写放大是指实际写入磁盘的数据量是用户写入的数据大小的多倍，浪费了大量IO资源，对于SSD这种存储介质的寿命非常不利。

## 在日志结构数据库中，写操作遇到SSTable满的情况时，会触发上层SST到下层的Compaction。遇到更坏的情况时，compaction可能会继续触发更下层的compaction，导致一次写操作就需要进行很多层SSTable的合并与重写，也就导致IO的数据量会远大于用户写入的数据大小，带来了写放大。解决方法通常有键值分离、lazy compaction等。

## 请阐述向量数据库中两种以上不同的相似度计算方法中所采用的具体计算方式？

## 答：

## 余弦相似度（Cosine Similarity），公式如下

## 欧几里得距离（Euclidean Distance），公式如下

## 曼哈顿距离（Manhattan Distance），公式如下