摘要：通过对视频网站观看量数据的分析，为视频网站cdn策略提供一种思路。

我从 #todo 拿到了 #todo 数据，并对其进行整理分析。

xFig0

1. 如果符合Zipf分布（#todo简单介绍一下zipf分布），那么图象应该是一条直线。图 #fig0 显示了视频的观看量和排名的关系，其中横坐标是视频按照观看量从高到低的排名，Rank = 1为观看量最高的视频；纵坐标是视频的观看量。图象在x<=5时表现出Zipf分布的规律，（#todo这里说明一下有什么意义）但是x>5之后迅速衰减。可能的原因是Youtube启用了某种过滤系统，导致观看量较少的视频更难出现在用户的视野中，排名靠后的视频观看量更难上升，因此曲线的尾部有一个迅速下降的趋势。

xFig1

fig1\_1

1. #fig1\_1展示了视频观看量的分布图，横坐标是视频观看量，纵坐标是视频数量，视频的数量主要集中在观看量小于100，这一部分的视频数量有着极大的占比，y = -klogx + b，因此，观看量较小的视频充斥着视频网站的内容，而在视频分发中，这些重要性不那么突出的视频却与其他观看量大的视频占据了同样的甚至更高的带宽，这是不应当发生的，所以视频分发策略应该保证这些观看量小但是基数巨大的视频不占满 #高速缓存。

fig1\_2

1. #fig1\_2展示了视频观看量的累积曲线，横坐标是视频观看量，纵坐标是观看次数低于x的视频的观看量总和，从图#fig1\_2可以看出，中等观看量（介于3000次到100000次之间）的视频观看量有着最大的占比，这一部分的视频是 #Vod 网站的主力，贡献了最多的观看量。（#有空算一下概率密度函数）

xFig2

fig2\_1，2\_2

1. 图#fig2\_1展示了视频长度的概率密度分布，视频的长度集中在10秒左右，3秒到100秒的视频居多，长度大于700秒的视频几乎不存在。从 #fig2\_2 的累积分布也可以看出，长度在3到100秒的短视频占据了 #Vod 网站的大部分。对于一般的视频提供网站，比如提供电视剧、电影、综艺节目等内容的视频网站的视频长度都较长，少则十几二十分钟，多则两三个小时，相对来说， #Vod 视频网站提供的视频大多数为2分钟以内的短视频，这与视频生产者和盈利模式有关，这里不做赘述。

xFig3

fig3\_1