估算步骤

1. 用户量预估

2. 用户行为建模

3. 性能需求计算

【用户量】

1. 2020.9月月活5.11亿，日活2.24亿（参考《微博2020用户发展报告》）。

【关键行为】

1. 发微博；

2. 看微博；

3. 评论微博（作业的内容，本节课不分析）

用户行为建模和性能估算

【评论微博】

考虑到微博是一个看得多发的少的业务，假设平均每天每人发1条微博（只考虑文字微博），则微博每天的发送量约为2.5亿条。

大部分的人发微博集中在早上8：00~9：00点，中午12：00~13：00，晚上20：00~22：00，假设这几个时间段发微博总量占比为 60%。对用户而言，非热点微博，每100条微博里面，可能有一条有评论；每20条评论中，50%的评论没有回复；30%评论只有一轮回复；10%评论假设有三轮回复。评论微博，回复内容为文字。微博评论，回复时间与微博发布时间重合。

【看微博】

由于绝大部分微博用户看微博的对象是大V和明星，因此我们假设平均一条微博观看人数有100次，则观看微博的次数为：

2.5亿 \* 100 = 250亿。

用户对所看微博，1%的概率进行一轮回复。针对已评论的微博中，40%概率进行第二轮评论。假设用户不会对自己已评论内容进行超过两轮回复。则评论微博次数为0.35亿

2.5亿\*100\*1% +2.5\*100\*1%\*40%=3.5亿

大部分人评论微博的时间段和看微博的时间段基本重合，因此评论微博的平均 TPS 计算如下：

3.5亿 \* 60% / (4\*3600) = 14.5K/s，取15K/s计算

【业务特性分析】

评论微博是一个典型的读多写少的操作，需要对被评论微博进行热度统计。因此需要用缓存(普通评论，以应用内缓存为主；热点微博，以外部缓存为主)，可以用负载均衡。

【架构分析】

用户量过亿，应该要用多级负载均衡架构，覆盖 DNS -> F5 -> Nginx -> 网关的多级负载均衡。

【架构设计】

1. 负载均衡算法选择

评论微博的时候依赖登录状态，登录状态一般都是保存在分布式缓存中的，因此评论微博的时候，评论内容同被评论内容存在关联性，通过Hash算法将请求发给指定服务器。

2. 业务服务器数量估算

评论微博涉及几个关键的处理：内容审核（依赖审核系统）、热度统计、数据写入存储（依赖存储系统）、数据写入缓存（依赖缓存系统），因此按照一个服务每秒处理250来估算，完成15K/s的 TPS，需要60台服务器，加上一定的预留量，80台服务器差不多了。

服务发现

服务发现

DNS

评论微博

评论微博

评论微博

评论微博

评论微博

评论微博

评论微博

评论微博

评论微博

评论微博

服务网关

服务网关

服务网关

服务网关

服务网关

App

F5/LVS

F5/LVS

NGINX

NGINX

NGINX

NGINX

NGINX