

## 36-Redis支撑秒杀场景的关键技术和实践都有哪些？

你好，我是蒋德钧。

秒杀是一个非常典型的活动场景，比如，在双11、618等电商促销活动中，都会有秒杀场景。秒杀场景的业务特点是**限时限量**，业务系统要处理瞬时的大量高并发请求，而Redis就被用来支撑秒杀活动。

不过，秒杀场景包含了多个环节，可以分成秒杀前、秒杀中和秒杀后三个阶段，每个阶段的请求处理需求并不相同，Redis并不能支撑秒杀场景的每一个环节。

那么，Redis具体是在秒杀场景的哪个环节起到支撑作用的呢？又是如何支持的呢？清楚了这个问题，我们才能知道在秒杀场景中，如何使用Redis来支撑高并发压力，并且做好秒杀场景的应对方案。

接下来，我们先来了解下秒杀场景的负载特征。

### 秒杀场景的负载特征对支撑系统的要求

秒杀活动售卖的商品通常价格非常优惠，会吸引大量用户进行抢购。但是，商品库存量却远远小于购买该商品的用户数，而且会限定用户只能在一定的时间段内购买。这就给秒杀系统带来两个明显的负载特征，相应的，也对支撑系统提出了要求，我们来分析下。

**第一个特征是瞬时并发访问量非常高。**

一般数据库每秒只能支撑千级别的并发请求，而Redis的并发处理能力（每秒处理请求数）能达到万级别，甚至更高。所以，**当有大量并发请求涌入秒杀系统时，我们就需要使用Redis先拦截大部分请求，避免大量请求直接发送给数据库，把数据库压垮。**

**第二个特征是读多写少，而且读操作是简单的查询操作。**

在秒杀场景下，用户需要先查验商品是否还有库存（也就是根据商品ID查询该商品的库存还有多少），只有库存有余量时，秒杀系统才能进行库存扣减和下单操作。

库存查验操作是典型的键值对查询，而Redis对键值对查询的高效支持，正好和这个操作的要求相匹配。

不过，秒杀活动中只有少部分用户能成功下单，所以，商品库存查询操作（读操作）要远多于库存扣减和下单操作（写操作）。

当然，实际秒杀场景通常有多个环节，刚才介绍的用户查验库存只是其中的一个环节。那么，Redis具体可以在整个秒杀场景中哪些环节发挥作用呢？这就要说到秒杀活动的整体流程了，我们分析下。

### Redis可以在秒杀场景的哪些环节发挥作用？

我们一般可以把秒杀活动分成三个阶段。在每一个阶段，Redis所发挥的作用也不一样。

第一阶段是秒杀活动前。

在这个阶段，用户会不断刷新商品详情页，这会导致详情页的瞬时请求量剧增。这个阶段的应对方案，一般

是尽量把商品详情页的页面元素静态化，然后使用CDN或是浏览器把这些静态化的元素缓存起来。这样一来，秒杀前的大量请求可以直接由CDN或是浏览器缓存服务，不会到达服务器端了，这就减轻了服务器端的压力。

在这个阶段，有CDN和浏览器缓存服务请求就足够了，我们还不需要使用Redis。

第二阶段是秒杀活动开始。

此时，大量用户点击商品详情页上的秒杀按钮，会产生大量的并发请求查询库存。一旦某个请求查询到有库存，紧接着系统就会进行库存扣减。然后，系统会生成实际订单，并进行后续处理，例如订单支付和物流服务。如果请求查不到库存，就会返回。用户通常会继续点击秒杀按钮，继续查询库存。

简单来说，这个阶段的操作就是三个：库存查验、库存扣减和订单处理。因为每个秒杀请求都会查询库存，而请求只有查到有库存余量后，后续的库存扣减和订单处理才会被执行。所以，这个阶段中最大的并发压力都在库存查验操作上。

为了支撑大量高并发的库存查验请求，我们需要在这个环节使用Redis保存库存量，这样一来，请求可以直接从Redis中读取库存并进行查验。

那么，库存扣减和订单处理是否都可以交给后端的数据库来执行呢？

其实，订单处理可以在数据库中执行，但库存扣减操作，不能交给后端数据库处理。

在数据库中处理订单的原因比较简单，我先说下。

订单处理会涉及支付、商品出库、物流等多个关联操作，这些操作本身涉及数据库中的多张数据表，要保证处理的事务性，需要在数据库中完成。而且，订单处理时的请求压力已经不大，数据库可以支撑这些订单处理请求。

那为啥库存扣减操作不能在数据库执行呢？这是因为，一旦请求查到有库存，就意味着发送该请求的用户获得了商品的购买资格，用户就会下单了。同时，商品的库存余量也需要减少一个。如果我们把库存扣减的操作放到数据库执行，会带来两个问题。

1. **额外的开销。**Redis中保存了库存量，而库存量的最新值又是数据库在维护，所以数据库更新后，还需要和Redis进行同步，这个过程增加了额外的操作逻辑，也带来了额外的开销。
2. **下单量超过实际库存量，出现超售。**由于数据库的处理速度较慢，不能及时更新库存余量，这就会导致大量库存查验的请求读取到旧的库存值，并进行下单。此时，就会出现下单数量大于实际的库存量，导致出现超售，这就不符合业务层的要求了。

所以，我们就需要直接在Redis中进行库存扣减。具体的操作是，当库存查验完成后，一旦库存有余量，我们就立即在Redis中扣减库存。而且，为了避免请求查询到旧的库存值，库存查验和库存扣减这两个操作需要保证原子性。

第三阶段就是秒杀活动结束后。

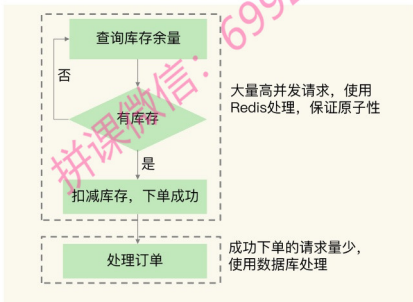
在这个阶段，可能还会有部分用户刷新商品详情页，尝试等待有其他用户退单。而已经成功下单的用户会刷

新订单详情，跟踪订单的进展。不过，这个阶段中的用户请求量已经下降很多了，服务器端一般都能支撑，我们就不重点讨论了。

好了，我们先来总结下秒杀场景对Redis的需求。

秒杀场景分成秒杀前、秒杀中和秒杀后三个阶段。秒杀开始前，高并发压力没有那么大，我们不需要使用Redis，但在秒杀进行中，需要查验和扣减商品库存，库存查验面临大量的高并发请求，而库存扣减又需要和库存查验一起执行，以保证原子性。这就是秒杀对Redis的需求。

下图显示了在秒杀场景中需要Redis参与的两个环节：



了解需求后，我们使用Redis来支撑秒杀场景的方法就比较清晰了。接下来，我向你介绍两种方法。

## Redis的哪些方法可以支撑秒杀场景？

秒杀场景对Redis操作的根本要求有两个。

1. **支持高并发。**这个很简单，Redis本身高速处理请求的特性就可以支持高并发。而且，如果有多个秒杀商品，我们也可以使用切片集群，用不同的实例保存不同商品的库存，这样就避免，使用单个实例导致所有的秒杀请求都集中在一个实例上的问题了。不过，需要注意的是，当使用切片集群时，我们要先用CRC算法计算不同秒杀商品key对应的Slot，然后，我们在分配Slot和实例对应关系时，才能把不同秒杀商品对应的Slot分配到不同实例上保存。
2. **保证库存查验和库存扣减原子性执行。**针对这条要求，我们就可以使用Redis的原子操作或是分布式锁这两个功能特性来支撑了。

## 基于原子操作支撑秒杀场景

在秒杀场景中，一个商品的库存对应了两个信息，分别是总库存量和已秒杀量。这种数据模型正好是一个key（商品ID）对应了两个属性（总库存量和已秒杀量），所以，我们可以使用一个Hash类型的键值对来保存库存的这两个信息，如下所示：

```
key: itemID  
value: {total: N, ordered: M}
```

其中，itemID是商品的编号，total是总库存量，ordered是已秒杀量。

因为库存查验和库存扣减这两个操作要保证一起执行，一个直接的方法就是使用Redis的原子操作。

我们在第29讲中学习过，原子操作可以是Redis自身提供的原子命令，也可以是Lua脚本。因为库存查验和库存扣减是两个操作，无法用一条命令来完成，所以，我们就需要使用Lua脚本原子性地执行这两个操作。

那怎么在Lua脚本中实现这两个操作呢？我给你提供一段Lua脚本写的伪代码，它显示了这两个操作的实现。

```
获取商品库存信息  
local counts = redis.call("HMGET", KEYS[1], "total", "ordered");  
将总库存转换为数值  
local total = tonumber(counts[1])  
将已秒杀的库存转换为数值  
local ordered = tonumber(counts[2])  
如果当前请求的库存量加上已秒杀的库存量仍然小于总库存量，就可以更新库存  
if ordered + k <= total then  
    更新已秒杀的库存量  
    redis.call("INCRBY", KEYS[1], "ordered", k)    return k;  
end  
return 0
```

有了Lua脚本后，我们就可以在Redis客户端，使用EVAL命令来执行这个脚本了。

最后，客户端会根据脚本的返回值，来确定秒杀是成功还是失败了。如果返回值是k，就是成功了；如果是0，就是失败。

到这里，我们学习了如何使用原子性的Lua脚本来实现库存查验和库存扣减。其实，要想保证库存查验和扣减这两个操作的原子性，我们还有另一种方法，就是使用分布式锁来保证多个客户端能互斥执行这两个操作。接下来，我们就来看下如何使用分布式锁来支撑秒杀场景。

## 基于分布式锁来支撑秒杀场景

使用分布式锁来支撑秒杀场景的具体做法是，先让客户端向Redis申请分布式锁，只有拿到锁的客户端才能执行库存查验和库存扣减。这样一来，大量的秒杀请求就会在争夺分布式锁时被过滤掉。而且，库存查验和

的互斥性。

你可以看下下面的伪代码，它显示了使用分布式锁来执行库存查验和扣减的过程。

```
//使用商品ID作为key
key = itemID
//使用客户端唯一标识作为value
val = clientUniqueID
//申请分布式锁，Timeout是超时时间
lock = acquireLock(key, val, Timeout)
//当拿到锁后，才能进行库存查验和扣减
if(lock == True) {
    //库存查验和扣减
    availStock = DECR(key, k)
    //库存已经扣减完了，释放锁，返回秒杀失败
    if (availStock < 0) {
        releaseLock(key, val)
        return error
    }
    //库存扣减成功，释放锁
    else{
        releaseLock(key, val)
        //订单处理
    }
}
//没有拿到锁，直接返回
else
    return
```

需要提醒你的是，在使用分布式锁时，客户端需要先向Redis请求锁，只有请求到了锁，才能进行库存查验等操作，这样一来，客户端在争抢分布式锁时，大部分秒杀请求本身就会因为抢不到锁而被拦截。

所以，我给你一个小建议，**我们可以使用切片集中的不同实例来分别保存分布式锁和商品库存信息**。使用这种保存方式后，秒杀请求会首先访问保存分布式锁的实例。如果客户端没有拿到锁，这些客户端就不会查询商品库存，这就可以减轻保存库存信息的实例的压力了。

## 小结

这节课，我们学习了Redis在秒杀场景中的具体应用。秒杀场景有2个负载特征，分别是瞬时高并发请求和读写多。Redis良好的高并发处理能力，以及高效的键值对读写特性，正好可以满足秒杀场景的需求。

在秒杀场景中，我们可以通过前端CDN和浏览器缓存拦截大量秒杀前的请求。在实际秒杀活动进行时，库存查验和库存扣减是承受巨大并发请求压力的两个操作，同时，这两个操作的执行需要保证原子性。Redis的原子操作、分布式锁这两个功能特性可以有效地来支撑秒杀场景的需求。

当然，对于秒杀场景来说，只用Redis是不够的。秒杀系统是一个系统性工程，Redis实现了对库存查验和扣减这个环节的支撑，除此之外，还有4个环节需要我们处理好。

1. **前端静态页面的设计**。秒杀页面上能静态化处理的页面元素，我们都要尽量静态化，这样可以充分利用CDN或浏览器缓存服务秒杀开始前的请求。

单禁止恶意IP进行访问。如果Redis实例的访问压力过大，为了避免实例崩溃，我们也需要在接入层进行限流，控制进入秒杀系统的请求数量。

3. **库存信息过期时间处理。**Redis中保存的库存信息其实是数据库的缓存，为了避免缓存击穿问题，我们不要给库存信息设置过期时间。
4. **数据库订单异常处理。**如果数据库没能成功处理订单，可以增加订单重试功能，保证订单最终能被成功处理。

最后，我也再给你一个建议：秒杀活动带来的请求流量巨大，我们需要把秒杀商品的库存信息用单独的实例保存，而不要和日常业务系统的数据保存在同一个实例上，这样可以避免干扰业务系统的正常运行。

## 每课一问

按照惯例，我给你提个小问题，假设一个商品的库存量是300，我们使用一个包含了4个实例的切片集群来服务秒杀请求。我们让每个实例各自维护库存量200，然后，客户端的秒杀请求可以分发到不同的实例上进行处理，你觉得这是一个好方法吗？

欢迎在留言区写下你的思考和答案，我们一起交流讨论。如果你觉得今天的内容对你有帮助，也欢迎你分享给你的朋友或同事。我们下节课见。

## 精选留言：

- Kaito 2020-11-13 08:35:44  
使用多个实例的切片集群来分担秒杀请求，是否是一个好方法？

使用切片集群分担秒杀请求，可以降低每个实例的请求压力，前提是秒杀请求可以平均打到每个实例上，否则会出现秒杀请求倾斜的情况，反而会增加某个实例的压力，而且会导致商品没有全部卖出的情况。

但用切片集群分别存储库存信息，缺点是如果需要向用户展示剩余库存，要分别查询多个切片，最后聚合结果后返回给客户端。这种情况下，建议不展示剩余库存信息，直接针对秒杀请求返回是否秒杀成功即可。

秒杀系统最重要的是，把大部分请求拦截在最前面，只让很少请求能够真正进入到后端系统，降低后端服务的压力，常见的方案包括：页面静态化（推送到CDN）、网关恶意请求拦截、请求分段放行、缓存校验和扣减库存、消息队列处理订单。

另外，为了不影响其他业务系统，秒杀系统最好和业务系统隔离，主要包括应用隔离、部署隔离、数据存储隔离。[9赞]

- test 2020-11-13 09:03:49  
用分片集群保存四个分库存，查询库存的时候需要四个实例查询后汇总返回，扣减库存的时候只扣减一个实例，有可能那个实例库存为空。而且要保证上游流量分发均匀（不能用轮询，只能用hash）。问题复杂化了，还不如用一个实例。
- tt 2020-11-13 08:26:11  
每个实例200个库存，当某个请求在某个实例上查询不到库存时，并不是库存真的为0了，还需要去其它实例查询，这个是一个缺点。

好处时可以承受更大的流量。