

21-缓冲区：一个可能引发“惨案”的地方

你好，我是蒋德钧。今天，我们一起来学习下Redis中缓冲区的用法。

缓冲区的功能其实很简单，主要就是用一块内存空间来暂时存放命令数据，以免出现因为数据和命令的处理速度慢于发送速度而导致的数据丢失和性能问题。但因为缓冲区的内存空间有限，如果往里面写入数据的速度持续地大于从里面读取数据的速度，就会导致缓冲区需要越来越多的内存来暂存数据。当缓冲区占用的内存超出了设定的上限阈值时，就会出现缓冲区溢出。

如果发生了溢出，就会丢数据了。那是不是不给缓冲区的大小设置上限，就可以了呢？显然不是，随着累积的数据越来越多，缓冲区占用内存空间越来越大，一旦耗尽了Redis实例所在机器的可用内存，就会导致Redis实例崩溃。

所以毫不夸张地说，缓冲区是用来避免请求或数据丢失的惨案的，但也只有用对了，才能真正起到“避免”的作用。

我们知道，Redis是典型的client-server架构，所有的操作命令都需要通过客户端发送给服务器端。所以，缓冲区在Redis中的一个主要应用场景，就是在客户端和服务端之间进行通信时，用来暂存客户端发送的命令数据，或者是服务器端返回给客户端的数据结果。此外，缓冲区的另一个主要应用场景，是在主从节点间进行数据同步时，用来暂存主节点接收的写命令和数据。

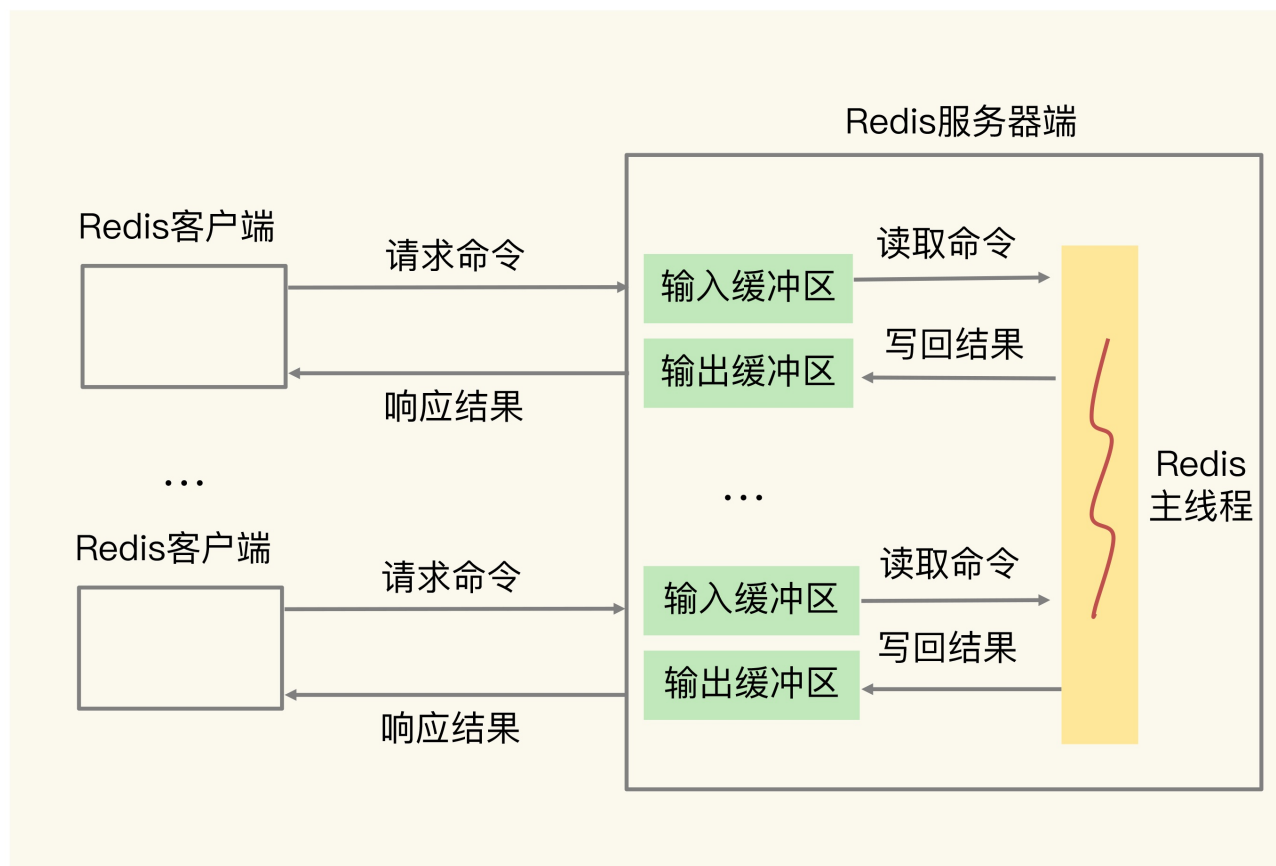
这节课，我们就分别聊聊服务器端和客户端、主从集群间的缓冲区溢出问题，以及应对方案。

客户端输入和输出缓冲区

我们先来看看服务器端和客户端之间的缓冲区。

为了避免客户端和服务端的请求发送和处理速度不匹配，服务器端给每个连接的客户端都设置了一个输入缓冲区和输出缓冲区，我们称之为客户端输入缓冲区和输出缓冲区。

输入缓冲区会先把客户端发送过来的命令暂存起来，Redis主线程再从输入缓冲区中读取命令，进行处理。当Redis主线程处理完数据后，会把结果写入到输出缓冲区，再通过输出缓冲区返回给客户端，如下图所示：



下面，我们就分别学习下输入缓冲区和输出缓冲区发生溢出的情况，以及相应的应对方案。

如何应对输入缓冲区溢出？

我们前面已经分析过了，输入缓冲区就是用来暂存客户端发送的请求命令的，所以可能导致溢出的情况主要是下面两种：

- 写入了bigkey，比如一下子写入了多个百万级别的集合类型数据；
- 服务器端处理请求的速度过慢，例如，Redis主线程出现了间歇性阻塞，无法及时处理正常发送的请求，导致客户端发送的请求在缓冲区越积越多。

接下来，我们就从**如何查看输入缓冲区的内存使用情况，以及如何避免溢出**这两个问题出发，来继续学习吧。

要查看和服务端相连的每个客户端对输入缓冲区的使用情况，我们可以**使用CLIENT LIST命令**：

```
CLIENT LIST
id=5 addr=127.0.0.1:50487 fd=9 name= age=4 idle=0 flags=N db=0 sub=0 psub=0 multi=-1 qbuf=26 qbuf-free=3274
```

CLIENT命令返回的信息虽然很多，但我们只需要重点关注两类信息就可以了。

一类是与服务器端连接的客户端的信息。这个案例展示的是一个客户端的输入缓冲区情况，如果有多个客户端，输出结果中的addr会显示不同客户端的IP和端口号。

另一类是与输入缓冲区相关的三个参数：

- cmd，表示客户端最新执行的命令。这个例子中执行的是CLIENT命令。
- qbuf，表示输入缓冲区已经使用的大小。这个例子中的CLIENT命令已使用了26字节大小的缓冲区。
- qbuf-free，表示输入缓冲区尚未使用的大小。这个例子中的CLIENT命令还可以使用32742字节的缓冲区。qbuf和qbuf-free的总和就是，Redis服务器端当前为已连接的这个客户端分配的缓冲区总大小。这个例子中总共分配了 $26 + 32742 = 32768$ 字节，也就是32KB的缓冲区。

有了CLIENT LIST命令，我们就可以通过输出结果来判断客户端输入缓冲区的内存占用情况了。如果qbuf很大，而同时qbuf-free很小，就要引起注意了，因为这时候输入缓冲区已经占用了很多内存，而且没有什么空闲空间了。此时，客户端再写入大量命令的话，就会引起客户端输入缓冲区溢出，Redis的处理办法就是把客户端连接关闭，结果就是业务程序无法进行数据存取了。

通常情况下，Redis服务器端不止服务一个客户端，当多个客户端连接占用的内存总量，超过了Redis的maxmemory配置项时（例如4GB），就会触发Redis进行数据淘汰。一旦数据被淘汰出Redis，再要访问这部分数据，就需要去后端数据库读取，这就降低了业务应用的访问性能。此外，更糟糕的是，如果使用多个客户端，导致Redis内存占用过大，也会导致内存溢出（out-of-memory）问题，进而会引起Redis崩溃，给业务应用造成严重影响。

所以，我们必须得想办法避免输入缓冲区溢出。**我们可以从两个角度去考虑如何避免，一是把缓冲区调大，二是从数据命令的发送和处理速度入手。**

我们先看看，到底有没有办法通过参数调整输入缓冲区的大小呢？答案是没有。

Redis的客户端输入缓冲区大小的上限阈值，在代码中就设定为了1GB。也就是说，Redis服务器端允许为每个客户端最多暂存1GB的命令和数据。1GB的大小，对于一般的生产环境已经是比较合适的了。一方面，这个大小对于处理绝大部分客户端的请求已经够用了；另一方面，如果再大的话，Redis就有可能因为客户端占用了过多的内存资源而崩溃。

所以，Redis并没有提供参数让我们调节客户端输入缓冲区的大小。如果要避免输入缓冲区溢出，那我们就只能从数据命令的发送和处理速度入手，也就是前面提到的避免客户端写入bigkey，以及避免Redis主线程阻塞。

接下来，我们再来看看输出缓冲区的溢出问题。

如何应对输出缓冲区溢出？

Redis的输出缓冲区暂存的是Redis主线程要返回给客户端的数据。一般来说，主线程返回给客户端的数据，既有简单且大小固定的OK响应（例如，执行SET命令）或报错信息，也有大小不固定的、包含具体数据的执行结果（例如，执行HGET命令）。

因此，Redis为每个客户端设置的输出缓冲区也包括两部分：一部分，是一个大小为16KB的固定缓冲空间，用来暂存OK响应和出错信息；另一部分，是一个可以动态增加的缓冲空间，用来暂存大小可变的响应结果。

那什么情况下会发生输出缓冲区溢出呢？ 我为你总结了三种：

- 服务器端返回bigkey的大量结果；

- 执行了MONITOR命令；
- 缓冲区大小设置得不合理。

其中，bigkey原本就会占用大量的内存空间，所以服务器端返回的结果包含bigkey，必然会影响输出缓冲区。接下来，我们就重点看下，执行MONITOR命令和设置缓冲区大小这两种情况吧。

MONITOR命令是用来监测Redis执行的。执行这个命令之后，就会持续输出监测到的各个命令操作，如下所示：

```
MONITOR
OK
1600617456.437129 [0 127.0.0.1:50487] "COMMAND"
1600617477.289667 [0 127.0.0.1:50487] "info" "memory"
```

到这里，你有没有看出什么问题呢？MONITOR的输出结果会持续占用输出缓冲区，并越占越多，最后的结果就是发生溢出。所以，我要给你一个小建议：**MONITOR命令主要用在调试环境中，不要在线上生产环境中持续使用MONITOR**。当然，如果在线上环境中偶尔使用MONITOR检查Redis的命令执行情况，是没问题的。

接下来，我们看下输出缓冲区大小设置的问题。和输入缓冲区不同，我们可以通过client-output-buffer-limit配置项，来设置缓冲区的大小。具体设置的内容包括两方面：

- 设置缓冲区大小的上限阈值；
- 设置输出缓冲区持续写入数据的数量上限阈值，和持续写入数据的时间的上限阈值。

在具体使用client-output-buffer-limit来设置缓冲区大小的时候，我们需要先区分下客户端的类型。

对于和Redis实例进行交互的应用程序来说，主要使用两类客户端和Redis服务器端交互，分别是常规和Redis服务器端进行读写命令交互的普通客户端，以及订阅了Redis频道的订阅客户端。此外，在Redis主从集群中，主节点上也有一类客户端（从节点客户端）用来和从节点进行数据同步，我会在介绍主从集群中的缓冲区时，向你具体介绍。

当我们给普通客户端设置缓冲区大小时，通常可以在Redis配置文件中进行这样的设置：

```
client-output-buffer-limit normal 0 0 0
```

其中，normal表示当前设置的是普通客户端，第1个0设置的是缓冲区大小限制，第2个0和第3个0分别表示缓冲区持续写入量限制和持续写入时间限制。

对于普通客户端来说，它每发送完一个请求，会等到请求结果返回后，再发送下一个请求，这种发送方式称为阻塞式发送。在这种情况下，如果不是读取体量特别大的bigkey，服务器端的输出缓冲区一般不会被阻塞的。

所以，我们通常把普通客户端的缓冲区大小限制，以及持续写入量限制、持续写入时间限制都设置为0，也就是不做限制。

对于订阅客户端来说，一旦订阅的Redis频道有消息了，服务器端都会通过输出缓冲区把消息发给客户端。所以，订阅客户端和服务端间的消息发送方式，不属于阻塞式发送。不过，如果频道消息较多的话，也会占用较多的输出缓冲区空间。

因此，我们会给订阅客户端设置缓冲区大小限制、缓冲区持续写入量限制，以及持续写入时间限制，可以在Redis配置文件中这样设置：

```
client-output-buffer-limit pubsub 8mb 2mb 60
```

其中，pubsub参数表示当前是对订阅客户端进行设置；8mb表示输出缓冲区的大小上限为8MB，一旦实际占用的缓冲区大小要超过8MB，服务器端就会直接关闭客户端的连接；2mb和60表示，如果连续60秒内对输出缓冲区的写入量超过2MB的话，服务器端也会关闭客户端连接。

好了，我们来总结下如何应对输出缓冲区溢出：

- 避免bigkey操作返回大量数据结果；
- 避免在线上环境中持续使用MONITOR命令。
- 使用client-output-buffer-limit设置合理的缓冲区大小上限，或是缓冲区连续写入时间和写入量上限。

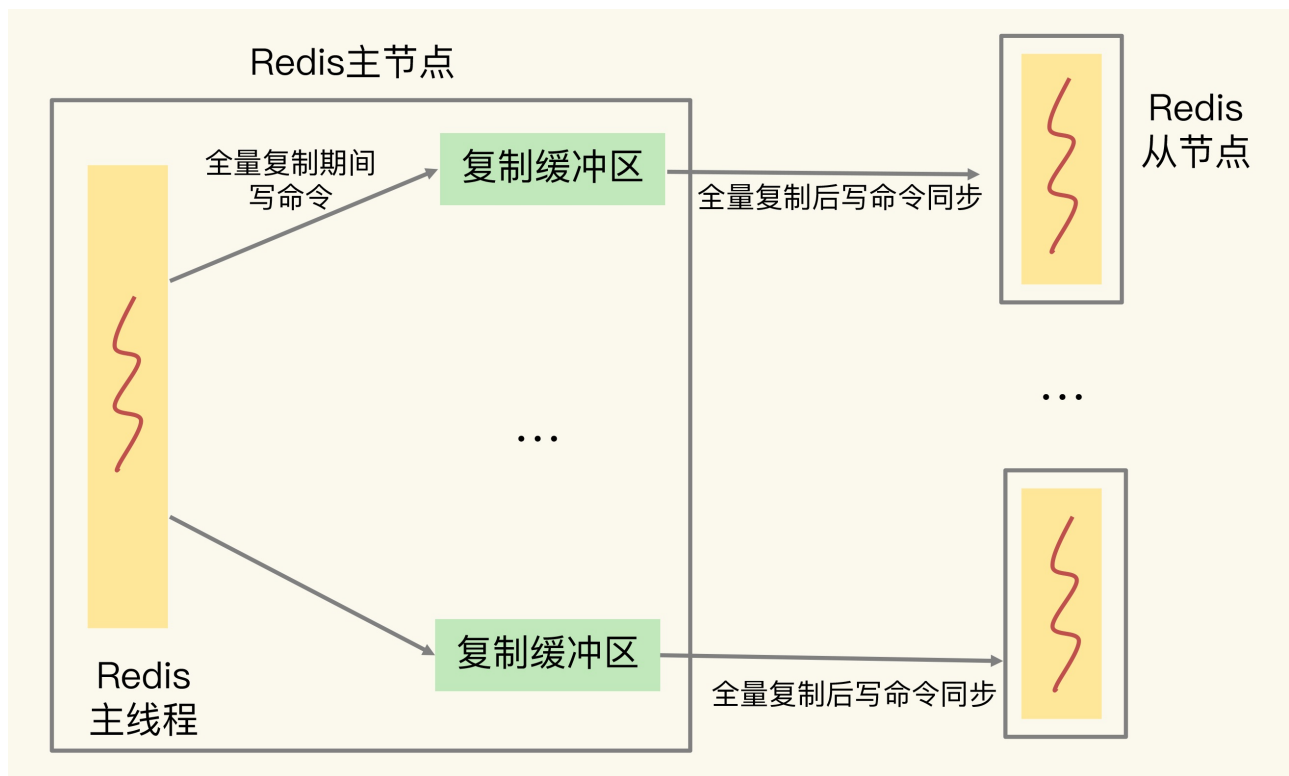
以上就是关于客户端缓冲区，我们要重点掌握的内容了。我们继续看看在主从集群间使用缓冲区，需要注意什么问题。

主从集群中的缓冲区

主从集群间的数据复制包括全量复制和增量复制两种。全量复制是同步所有数据，而增量复制只会把主从库网络断连期间主库收到的命令，同步给从库。无论在哪种形式的复制中，为了保证主从节点的数据一致，都会用到缓冲区。但是，这两种复制场景下的缓冲区，在溢出影响和大小设置方面并不一样。所以，我们分别来学习下吧。

复制缓冲区的溢出问题

在全量复制过程中，主节点在向从节点传输RDB文件的同时，会继续接收客户端发送的写命令请求。这些写命令就会先保存在复制缓冲区中，等RDB文件传输完成后，再发送给从节点去执行。主节点上会为每个从节点都维护一个复制缓冲区，来保证主从节点间的数据同步。



所以，如果在全量复制时，从节点接收和加载RDB较慢，同时主节点接收到了大量的写命令，写命令在复制缓冲区中就会越积越多，最终导致溢出。

其实，主节点上的复制缓冲区，本质上也是一个用于和从节点连接的客户端（我们称之为从节点客户端），使用的输出缓冲区。复制缓冲区一旦发生溢出，主节点也会直接关闭和从节点进行复制操作的连接，导致全量复制失败。那如何避免复制缓冲区发生溢出呢？

一方面，我们可以控制主节点保存的数据量大小。按通常的使用经验，我们会把主节点的数据量控制在2~4GB，这样可以全量同步执行得更快些，避免复制缓冲区累积过多命令。

另一方面，我们可以使用client-output-buffer-limit配置项，来设置合理的复制缓冲区大小。设置的依据，就是主节点的数据量大小、主节点的写负载压力和主节点本身的内存大小。

我们通过一个具体的例子，来学习下具体怎么设置。在主节点执行如下命令：

```
config set client-output-buffer-limit slave 512mb 128mb 60
```

其中，slave参数表明该配置项是针对复制缓冲区的。512mb代表将缓冲区大小的上限设置为512MB；128mb和60代表的设置是，如果连续60秒内的写入量超过128MB的话，也会触发缓冲区溢出。

我们再继续看看这个设置对我们有啥用。假设一条写命令数据是1KB，那么，复制缓冲区可以累积512K条（512MB/1KB = 512K）写命令。同时，主节点在全量复制期间，可以承受的写命令速率上限是2000条/s（128MB/1KB/60 约等于2000）。

这样一来，我们就得到了一种方法：在实际应用中设置复制缓冲区的大小时，可以根据写命令数据的大小和应用的实际负载情况（也就是写命令速率），来粗略估计缓冲区中会累积的写命令数据量；然后，再和所设

置的复制缓冲区大小进行比较，判断设置的缓冲区大小是否足够支撑累积的写命令数据量。

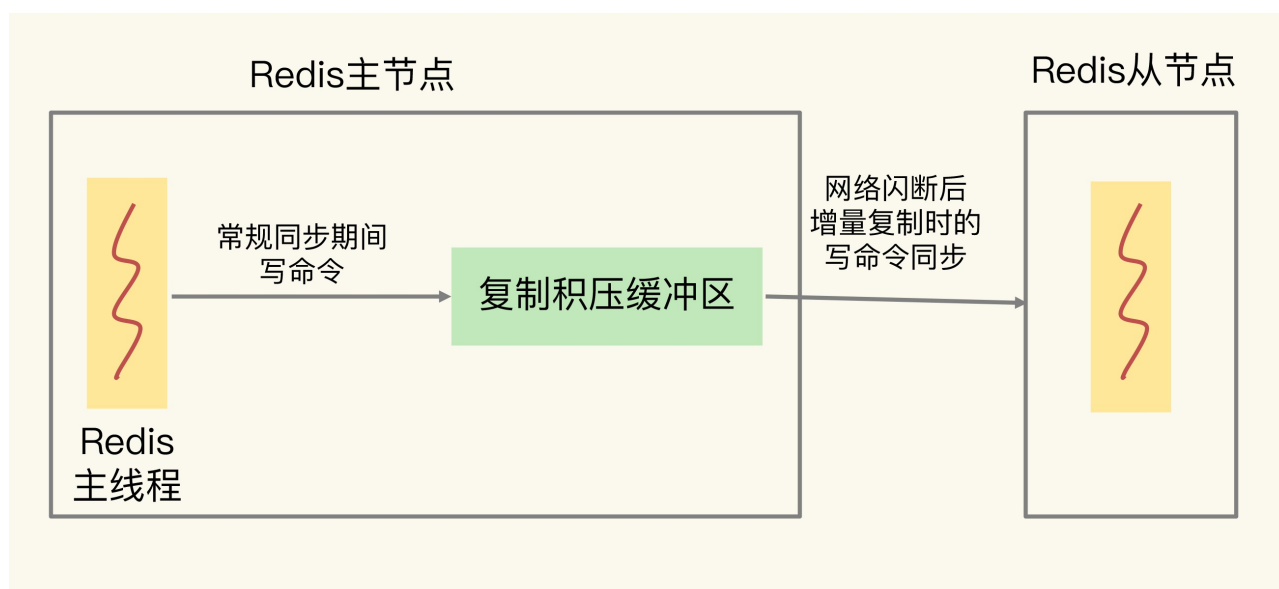
关于复制缓冲区，我们还会遇到一个问题。主节点上复制缓冲区的内存开销，会是每个从节点客户端输出缓冲区占用内存的总和。如果集群中的从节点数非常多的话，主节点的内存开销就会非常大。所以，我们还必须得控制和主节点连接的从节点个数，不要使用大规模的主从集群。

好了，我们先总结一下这部分的内容。为了避免复制缓冲区累积过多命令造成溢出，引发全量复制失败，我们可以控制主节点保存的数据量大小，并设置合理的复制缓冲区大小。同时，我们需要控制从节点的数量，来避免主节点中复制缓冲区占用过多内存的问题。

复制积压缓冲区的溢出问题

接下来，我们再来看下增量复制时使用的缓冲区，这个缓冲区称为复制积压缓冲区。

主节点在把接收到的写命令同步给从节点时，同时会把这些写命令写入复制积压缓冲区。一旦从节点发生网络闪断，再次和主节点恢复连接后，从节点就会从复制积压缓冲区中，读取断连期间主节点接收到的写命令，进而进行增量同步，如下图所示：



看到这里你是不是感觉有些熟悉？没错，我们在[第6讲](#)时已经学过复制积压缓冲区了，只不过我当时告诉你的是它的英文名字`repl_backlog_buffer`。所以这一讲，我们从缓冲区溢出的角度再来回顾下两个重点：复制积压缓冲区溢出的影响，以及如何应对复制积压缓冲区的溢出问题。

首先，复制积压缓冲区是一个大小有限的环形缓冲区。当主节点把复制积压缓冲区写满后，会覆盖缓冲区中的旧命令数据。如果从节点还没有同步这些旧命令数据，就会造成主从节点间重新开始执行全量复制。

其次，为了应对复制积压缓冲区的溢出问题，我们可以调整复制积压缓冲区的大小，也就是设置`repl_backlog_size`这个参数的值。具体的调整依据，你可以再看下[第6讲](#)中提供的`repl_backlog_size`大小的计算依据。

小结

这节课，我们一起学习了Redis中使用的缓冲区。使用缓冲区以后，当命令数据的接收方处理速度跟不上发送方的发送速度时，缓冲区可以避免命令数据的丢失。

按照缓冲区的用途，例如是用于客户端通信还是用于主从节点复制，我把缓冲区分成了客户端的输入和输出缓冲区，以及主从集群中主节点上的复制缓冲区和复制积压缓冲区。这样学习的好处是，你可以很清楚Redis中到底有哪些地方使用了缓冲区，那么在排查问题的时候，就可以快速找到方向——从客户端和服务器的通信过程以及主从节点的复制过程中分析原因。

现在，从缓冲区溢出对Redis的影响的角度，我再把这四个缓冲区分成两类做个总结。

- 缓冲区溢出导致网络连接关闭：普通客户端、订阅客户端，以及从节点客户端，它们使用的缓冲区，本质上都是Redis客户端和服务端之间，或是主从节点之间为了传输命令数据而维护的。这些缓冲区一旦发生溢出，处理机制都是直接把客户端和服务端的连接，或是主从节点间的连接关闭。网络连接关闭造成的直接影响，就是业务程序无法读写Redis，或者是主从节点全量同步失败，需要重新执行。
- 缓冲区溢出导致命令数据丢失：主节点上的复制积压缓冲区属于环形缓冲区，一旦发生溢出，新写入的命令数据就会覆盖旧的命令数据，导致旧命令数据的丢失，进而导致主从节点重新进行全量复制。

从本质上看，缓冲区溢出，无非就是三个原因：命令数据发送过快过大；命令数据处理较慢；缓冲区空间过小。明白了这哥，我们就可以有针对性地拿出应对策略了。

- 针对命令数据发送过快过大的问题，对于普通客户端来说可以避免bigkey，而对于复制缓冲区来说，就是避免过大的RDB文件。
- 针对命令数据处理较慢的问题，解决方案就是减少Redis主线程上的阻塞操作，例如使用异步的删除操作。
- 针对缓冲区空间过小的问题，解决方案就是使用client-output-buffer-limit配置项设置合理的输出缓冲区、复制缓冲区和复制积压缓冲区大小。当然，我们不要忘了，输入缓冲区的大小默认是固定的，我们无法通过配置来修改它，除非直接去修改Redis源码。

有了上面这些应对方法，我相信你在实际应用时，就可以避免缓冲区溢出带来的命令数据丢失、Redis崩溃的这些“惨案”了。

每课一问

最后，我给你提个小问题吧。

在这节课上，我们提到Redis采用了client-server架构，服务器端会为每个客户端维护输入、输出缓冲区。那么，应用程序和Redis实例交互时，应用程序中使用的客户端需要使用缓冲区吗？如果使用的话，对Redis的性能和内存使用会有影响吗？

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所帮助，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 2020-09-28 00:06:09
应用程序和Redis实例交互时，应用程序中使用的客户端需要使用缓冲区吗？如果使用的话，对Redis的性能和内存使用有什么影响？

客户端需要使用缓冲区，好处如下。

1、客户端和服务端交互，一般都会制定一个交互协议，客户端给服务端发数据时，都会按照这个协议把数据拼装好，然后写到客户端buffer中，客户端再一次性把buffer数据写到操作系统的网络缓冲区中，最后由操作系统发送给服务端。这样服务端就能从网络缓冲区中读取到一整块数据，然后按照协议解析数据即可。使用buffer发送数据会比一个个发送数据到服务端效率要高很多。

2、客户端还可以使用Pipeline批量发送命令到服务端，以提高访问性能。不使用Pipeline时，客户端是发送一个命令、读取一次结果。而使用Pipeline时，客户端先把一批命令暂存到buffer中，然后一次性把buffer中的命令发送到服务端，服务端处理多个命令后批量返回结果，这样做的好处是可以减少来回网络IO的次数，降低延迟，提高访问性能。当然，Redis服务端的buffer内存也会相应增长，可以控制好Pipeline命令的数量防止buffer超限。

缓冲区其实无处不在，客户端缓冲区、服务端缓冲区、操作系统网络缓冲区等等，凡是进行数据交互的两端，一般都会利用缓冲区来降低两端速度不匹配的影响。没有缓冲区，就好比一个个工人搬运货物到目的地，每个工人不仅成本高，而且运输效率低。而有了缓冲区后，相当于把这些货物先装到一个集装箱里，然后以集装箱为单位，开车运送到目的地，这样既降低了成本，又提高了运输效率。缓冲区相当于把需要运送的零散数据，进行一块块规整化，然后分批运输。

另外，关于Redis服务端为客户端分配的输出缓冲区，我想补充一点：主库上的从库输出缓冲区（slave client-output-buffer）是不计算在Redis使用的总内存中的，也就是说主从同步延迟，数据积压在主库上的从库输出缓冲区中，这个缓冲区内存占用变大，不会超过maxmemory导致淘汰数据。只有普通客户端和订阅客户端的输出缓冲区内存增长，超过maxmemory时，才会淘汰数据。[29赞]

- 无名小卒 2020-09-29 17:16:06
蒋老师的课程很深入，学到了很多底层redis知识，迫不及待的想看下面的章节，如果能更新快一点就更好了。催更啦😁!
- yeek 2020-09-28 23:50:30
从节点客户端输出缓冲区会用来传输rdb文件吗？如果会的话，该设置项的大小，是参考rdb文件大小吧？这可能是所有缓冲区中实际可能使用的最大缓冲区了
- yeek 2020-09-28 23:42:09
输出缓冲区，动态增加的部分，会产生内存碎片吗？何时回收呢？增加后会有机会缩小吗？
- yeek 2020-09-28 23:35:35
服务器端处理请求的速度过慢，例如，Redis 主线程出现了间歇性阻塞，无法及时处理正常发送的请求，导致客户端发送的请求在缓冲区越积越多。

这句话有点没理解，redis大部分请求是阻塞的，对客户端来说需要等待服务器的相应结果，虽然相应结果不一定有用，但这样的情况，redis服务端如果处理请求较慢，那么客户端输入缓冲区应该是当前请求一直在hold吧？服务器处理慢会导致等待的客户端变多，整体积压的输入缓冲变多，但对单个缓冲区来说，溢出应该不是主要的吧？

不知道上述理解是否正确……

- 脱缰的野马__ 2020-09-28 20:23:26
老师你好，在讲主从全量复制的时候，如果发生复制缓冲区溢出会导致全量复制失败，这个失败是指全部失败还是指rdb文件剩余部分同步复制失败，因为我知道的是每次从节点连接上主节点时会把当前对rdb复制偏移量给到主节点，主节点再判断是否重新全量复制还是从收到的偏移量位置开始复制

- 漫步oo0云端 2020-09-28 16:37:24

我发现redis2.8版本执行CLIENT LIST，没有 qbuf 和 qbuf-free 这两个值。