## 05-内存快照: 宕机后, Redis如何实现快速恢复?

你67. 我早菇德约.

上节课,我们学习了Redis避免数据丢失的AOF方法。这个方法的好处,是每次执行只需要记录操作命令,需要持久化的致密量不大。一般而言,只要你采用的不是always的持久化策略,就不会对性能造成太大影 total

但是,也正因为记录的是操作命令,而不是实际的数据,所以,用AOF方法进行放真恢复的时候,需要逐一把操作日志都执行一遍,如果操作日志非常多,Redis就会恢复得很缓慢。\$\%\$\正常使用。这当然不是理想的结果。那么,还有没有既可以保证可靠性,还能在岩机对实现快速恢复的基础方法呢?

当然有了,这就是我们今天要一起学习的另一种持久化力点: 內容後國。所谓內存快期,就是指內存中的数据在某一个时刻的状态记录。这就类似于照片,当你给那支伯里时,一张照片就能把朋友一瞬间的形象完全

对Redis来说,它实现类似照片记录效果的方式。故是把某一时刻的状态以文件的形式写到磁盘上,也就是 快期。这样一来,即使否机,使刚交换也先丢步失,数据的可靠性也就得到了保证。这个快期文件就称为 RDB安件、其中、RDB就是Redis Database的需求。

和AOF相比,RDB记录的是某一时刻的数据,并不是操作,所以,在做数据恢复时,我们可以直接把RDB文件读入内存。级域特殊的恢复。 听起来好像很不错,但内容惊醒也并不是最优选项。为什么这么说呢?

# 我们还要考虑两个关键问题:

- 对哪些数据做快照?这关系到快照的执行效率问题;
- · 做快照时,数据还能被增删改吗?这关系到Redis是否被阻塞,能否同时正常处理请求。

这么说可能你还不太好理解, 我还是拿拍照片来举例子。我们在拍照时, 通常要关注两个问题:

- 如何取景? 也就是说,我们打算把哪些人、哪些物拍到照片中;
- 在按快门前,要记着提醒朋友不要乱动,否则拍出来的照片就模糊了。

你看,这两个问题是不是非常重要呢?那么,接下来,我们就来具体地聊一聊。先说"取景"问题,也就是 我们对哪些数据做快照。

### 给哪些内存数据做快照?

Redis的数据都在内存中,为了提供所有数据的可塑性保证,它执行的是**全核期**,也就是说,把内存中的 所有数据都记录到继盘中,这就类似于结100个人拍台影,把每一个人都拍进照片里。这样做的好处是,一 次性记录了所有数据。一个都不少。

当你给一个人拍照时,只用协调一个人就够了,但是,拍100人的大合影,却需要协调100个人的位置、状态,等等。这当然会更费时费力。同样,给内存的全量数据使换照,把它们全部写入磁盘也会花费很多时间。而且、全着数据越多。RDS文件就越去、往城盘上节驾旅的时间开除战越大。

对于Redsli而言。它的单线程模型就决定了,我们要尽量避免所有会阻塞主线程的操作。所以,针对任何操作,我们都会提一个灵魂之问: "它会阻塞主线程吗?" RDB文件的主成是否会阻塞主线程,这就关系到是否会降低免纪纺炸能。

Redis提供了两个命令来生成RDB文件、分别是save和hosave。

- save: 在主线程中执行,会导致阻塞;
- bgsave:创建一个子进程,专门用于写入RDB文件,避免了主线程的阻塞,这也是Redis RDB文件生成的 默认配置。

好了,这个时候,我们就可以通过bgsave命令来执行全量快照,这既提供了数据的可靠性保证,也避免了 对Redis的性能影响。

接下来,我们要关注的问题就是,在对内方数据做快期时,这些数据还能"动"吗? 也就是说,这些数据还能被待众吗? 这个问题非常重要,这是因为,如果数据被被修改,那就意味着Redis还能正常处理与操作。 否则,所有写操作都得等到找明识了才像感,性能一下去的摊任了

### 快照时数据能修改吗?

在给别人拍照时,一旦对方动了,那么这张照片就拍糊了,我们就需要重拍,所以我们当然希望对方保持不动。对于内存铁照语言,我们也不希望数据"动"。

举个例子。我们在考别结内存做快照,假设内存数据量是4G。磁盘的写入带宽是0.2G8/s,简单来说,至 少需要20 (4/02 20) 才能整元。如果在时刻+55时,一个还没有被写入磁盘的内存数据A、被特效成了 A、那么就会牵环快期的完整性。因为A、不是时刻时的状态。因此,和拍照类似,我们在做快期时也不 希望数据"站",也就是不能被概念。

個量, 如果快照執行期间数据不能被修改, 组会有潜在问题的。对于例例的例子来说,在微快照的20-9时间 里, 如果这4GB的数据都不能被修改, Redis就不能处理对这些数据的写操作, 那无疑就会给业务服务造成 巨大的影响。

你可能必無別,可以用的Ssava避免阻塞啊。这里我就要说到一个常见的似区了,**避免阻塞和正常处理写接 作并不是一回事**。此时,主线程的确没有阻塞,可以正常接收请求,但是,为了保证快照完整性,它只能处 理读操作。因为不能修改正在执行快期的数据。

为了快照而暂停写操作,肯定是不能接受的。所以这个时候,Redis就会借助操作系统提供的写时复制技术 (Copy-On-Write, COW) ,在执行快照的同时,正常处理写操作。

简单来说,bgsave子进程是由主线程fork生成的,可以共享主线程的所有内存数据。bgsave子进程运行 后,开始读取主线程的内存数据,并把它们写入RDB文件。

此时,如果主线程对这些数据也都是读操作(例如图中的键值对A),那么,主线程和bgsave子进程相互不 影响。但是,如果主线程要的这一块数据(例如图中的键值对C),那么,这块数据检查被复制一份,生成 该数据的影本。然后,bgsave子进程会把这个副本数据写入RDB文件,而在这个过程中,主线程仍然可以 育单格学简单命的转程。



这既保证了快照的完整性,也允许主线程同时对数据进行修改,避免了对正常业务的影响。

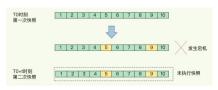
到这里,我们就解决了对"喝些放弃做快照"以及"做快照时数据能否修改"这两大问题:Redis会使用 bgsave对当前内存中的所有就这做快照,这个操作是子进程在后台完成的,这就允许主线程同时可以修改 数据。

现在,我们再来有另一个问题:多久做一次快照?我们在拍照的时候,还有项技术叫"连拍",可以记录人或物连续多个瞬间的状态。那么,快赔也适合"连拍"吗?

#### 可以每秒做一次快照吗?

对于快照来说,所谓"连拍"就是指连续地做快照。这样一来,快照的间隔时间变得很短,即使某一时刻发 生宕机了,因为上一时刻快照别执行,丢失的数据也不会太多。但是,这其中的快照间隔时间就很关键了。

如下围所示,我们先在TD时刻除了一次快期,然后又在TO+b时刻做了一次快期,在这期间,数据块5和9被 传改了。如果在这段时间内,机器右机了,那么,只能按照TD时刻的按照进行恢复。此时,数据块5和9的 传改值因为9倍转期记录。被干法恢复了。



呢,比如说是不是可以每秒做一次快照? 毕竟,每次快照都是由bgsave子进程在后台执行,也不会阻塞主 练程

这种想法其实是错误的。虽然bgsave执行时不阻塞主线程,但是,如果频繁地执行全量快服,也会带来两方面的开端。

一方面,频繁将全量数据写入磁盘,会给磁盘带来很大压力,多个快照竞争有限的磁盘带宽,前一个快照还没有做完,后一个又开始做了,容易造成恶性循环。

另一方面,bgsave于进程需要通过fork操作从主线程创建出来。虽然,于进程在创建后不会两阳寡主线程,但是,fork这个创建过程本身会阻塞主线程,而且主线程的内存超大,但要训制器长。如果频繁fork出 bssave于测度,这数余频繁测度主线程了。那么、有什么其他例子选择。

此时,我们可以做增量快照,所谓增量快照,就是指,做了一次全量快照后,后续的快照只对修改的数据进 行快照记录,这样可以避免每次全量快照的开**诉。** 

在第一次使完全量长限后,T1和T2时到19平560。据,我们只需要将被修改的数据写入快取文件就行。但 是,这么做的前提是,我们需要记住哪些数据被修改了。你可不要小嘴这个"记住"功能,它需要机使用 额外的元数据信息表记录那些数据管辖设了,这会带来题外的空间开销问题。如下图所示:



如果與们对每一个键值对的传改、都做个记录,那么,如果有1万个被传效的键值对,我们就需要有1万条额 外的记录。而且,有的时候,键值对专案小、比如只有32字节,而记录它被修改的元效规信息,可能依需要 8字节,这样的面,为了"记住" 傅改,引入的额外空间开销比较大。这对于内存资源宝贵的Redis来说,有 标稿不由年

到这里,你可以发现。虽然跟AOF相比,快照的恢复速度快,但是,快照的频率不好把握,如果频率太低, 两次快照间一旦右别,级可能有比较多的数据丢失。如果频率太高,又会产生额外开销,那么,还有什么方 法裁据利用ROB的快速按复,又能以较小的升销便到尽量少去数据呢?

Redis 4.0中提出了一个**混合使用AOF日志和内存快服**的方法。简单来说,内存快照以一定的频率执行,在两 次快照之间,使用AOF日志记录这期间的所有命令操作。 快照间的操作,也就是说,不需要记录所有操作了,因此,就不会出现文件过大的情况了,也可以避免重写 开细

如下图所示,T1和T2时刻的修改,用AOF日志记录,等到第二次做全量快照时,就可以清空AOF日志,因为 此时的修改都已经记录到快照中了,恢复时就不再用日志了。



这个方法既能享受到RDB文件快速恢复的好处,又能享受到AOF只记录操作命令的简单优势,颇有点"鱼和熊掌可以兼得"的感觉,建议你在实践中用起来。

#### 小结

这节语,我们学习了Redis用于避免数据丢失的内容快服方法。这个方法的优势在于,可以快速恢复数据 库,也就是只需要把RDB文件直接读入内容,这就避免了ADF需要顺序、逐一重新执行操作命令带来的低效 性能问题。

不过,内存快圈也有它的局限性。它拍的是一张内存的"大台影",不可避免地会耗时耗力。虽然,Redis 设计了Ussawe简写过复新方式,尽可能减少了你存快搁对正常读等的影响。但是,频繁按照仍是是不太能 接受的。而混合使用RDBMAOF,正好可以取两者之长,海两者之短,以较小的性能并研保证数据可靠性和 性能。

最后,关于AOF和RDB的选择问题,我想再给你提三点建议:

- 数据不能丢失时,内存快照和AOF的混合使用是一个很好的选择;
- · 如果允许分钟级别的数据丢失,可以只使用RDB;
- 如果只用AOF, 优先使用everysec的配置选项, 因为它在可靠性和性能之间取了一个平衡。

到这里,关于持久化我们就讲完了,这块儿内容是熟练掌握Redis的基础,建议你一定好好学习下这两节 课。如果你觉得有政莽,希望你能提到分享给事多的人。帮助更多人解决特久化的问题。

#### 精洗留言:

#### . Kaito 2020-08-14 01:23:00

2æ\_CPU36€4GBá+--à-"36€500G£€ç;"TWŒRediså"⊕åkná∳;" 26BWŒā+"\*ê'\*æ'" ākiā,°
8.2WŒace=--{à-\$SRDBæŒ-\$-à--à-Z-WŒ\*5'ç' 'yơ, liệ-ệ-"Qā, -è|-∳āœ''ā'∳ CPUèµ,æ'-∳ā'
Ή+--à\*"9µ,æ'-∳ê';"2æ-'è-∳0'W3

al@a1 — a1 m. a1 — 4 m. a1 — 4 m. a2 — 4 m. a

bide  $O(DU)_{ijk}m^2\phi(e^2\phi^2m)(e^2\phi^2m)(e^2\phi^2m)$  and  $(\Phi e^2D(B(E^2\phi^2m)-4E^2)(EE))(e^2\phi^2m^2\phi(E))$  bide  $(\Phi e^2\phi^2m^2\phi^2m^2\phi^2m)$  and  $(\Phi e^2\phi^2m^2$ 

### Darren 2020-08-14 09:26:10

Kaitoçš"å.♦ç" ä,°å · ¥è€♦è°©æ° '觉得臰å·±é,£ä'°è♦œå '¢í½í,,

 $x^{-1}(x) = x^{-1}(x) + x^{-1}(x) + x^{-1}(x) = x^{-1}(x) + x^{-$ 

圔redis4.0⻥ā‰♠7¼Œredis AOFçā,é‡♠写朰å″¶æ"ïæŒ‡äפæ·′ã♠″7¼°è€♠å, "ā,Šā çēš,ģ. "¼å"²ç+♠°"è;‡7¼‰7¼Œå9††æ"ï圔redis4.0ā×¼6♠7¼Œredis54,AOFé‡♠å\*7%,æ å◆ ~⻥通èz ‡ aof-use-rdb-preamble é···◆ç½°å◆»è\*¾ç½°æ" ¹åŠŸè∰ã€

# When rewriting the AOF file, Redis is able to use an RDB preamble in the
# AOF file for faster rewrites and recoveries. When this option is turned
# on the rewritten AOF file is composed of two different stanzas:

# # [RDB file][AOF tail]

#
# When loading Redis recognizes that the AOF file starts with the "REDIS"

# string and loads the prefixed RDB file, and continues loading the AOF

# tail.

aof-use-rdb-preamble yes

[5赞]

Geek\_cc6b96 2020-08-14 14:43:43
 RDBå' ŒAOF ä ◆å\*±æ⁻⁻æu◆å" ä½ "c¼-c ◆ä cš...lå §ã€◆På § æl.āṣuā⁻>

Spring4J 2020-08-14 11:13:12
 c" ±ä°�āj\*æ" 'æ "�āl/ɔœā�āï

. c䎣⊕æ⊕¥ǧŸ⊕ § 2020-08-14 10:38:06

\$\delta(\text{2}\), \text{ \text{def}} \(\delta(\text{def}) \), \(\delt

 $\delta|_{,m} \bullet cm^{-} \delta|_{,\kappa} \circ cm^{-} \delta|_{$ 

å» ä , "redisæ" "å♦ · ç"¿ç" ·çš "ī¼Œæ‰€ä×¥å♦Œæ , 够ç" ",ä ,€ä , "æ‰ § è jŒå "½äפī¼Œä ,€ ä. °c" ~ze�Vbgsave,ä. �会é~»â;�ä. »c°;c~\冷··å~4g,å\*�é™···å� c" ~2gå-lå�³ï¼ Ό; æ¤å † ···å "æ" "ä .€ä .\*ä . 'c · Œc. 'ī¼Œæ" "ål, fork å�è; ;åŸ�cš.时候å¾ "有å� "èf½ è€--尽所有çš,冷··å⁻,åŠä,Šè⁻»å†™æ⁻"例æ⁻-写多è⁻»å°ï¼Æé,£ä¹^æ⁴æ�° save cš\_é···• c½°å¾°a:cv%å• ~èf½å ± °c• °ét 'c¹•cš\_ bgsave,æ¤æ—¶å+···å°å°+ æ°•å.°

æ^ 'æf³æ��ā, €ā, "å,»é-"é¢~,æ^ '作ä, °å^�ā|耷··æf³é-",å|,æ�œredisæœ�āŠjæ Œ,䰆,⤇ä׽有ä׀买ç" "?èf½æ�¢ä¤�çš"ä‰�æ��ä,�æ" "æœ�äŠjè; "å"æ' » � - 『¼Ÿé\$¼é� "####ÅŠ;#Œ ä° † ä¼\$è ‡ "Š" #;%èu : ###\$Š;Ĩ¼Ÿè ‡ "Š"è; "Â� ŸÂ� - Ĩ¼Ÿ

å»ä, "Redisæ" "å♦ · ç°¿ç" ‹æ" ¡å♦ «ï¼Œæ‰€ä»¥Redisä, »è¿¿ç" ‹ä¹Ÿå°±æ" "ä, €ä, °ç°¿ç" «ï¼Œæ° "ä , ۏ°¬ā¹Ÿç§°ä,°ä,×ç°¿ç⁻⟨ä°†ã€,

�ā, "ĭ¼Œè¿™é‡Œæ" "ä, »ç°¿ç" «è¿ "æ" "ä, »è¿»ç" «ï¼Ÿ

āVscn耷··ā-•⤕2020-08-14 08:43:37

é.£æ—¶å\*×2020-08-14 09:34:20

 m—ā♠aā♠ā %-2020-08-14.08-07-10 ç"€å• · æ•¥è" "bgsave å•迸ç" ·æ" "ç" ±ä, »ç°¿ç" · fork ç" Ÿæ"•çš"å• "以共䰫ä .»c°/c ~c\$,所有å† ···å\*æ · °æ�\*ã€,bgsave å�è/>c ~è/�è/Œã��,ã¼€å § ·è\* »å�-ä ,»c °2c~c5.å † ···å~æ · °æ�\*,å¹ ¶æŠŠå\*f们å † ™å···¥ RDB æ- ‡ ä» ¶ã€.

ä °é€‰é:¹ï¼Ÿè;™ä °è™c'y°æ "有什ä¹ ā···¶ä»-cš.ä¼⁻c.¹ä′ Œä½;c" "地æ™ å♠—? 386 \$ 20 - 3 \$ 20 C \$ 2 æ"iå¼�ï¼₩i¼Œè;™æ ·å�šcš,è"�æ""å�^é€,cš,æŒ�ā'···åŒ-选æ:©å�—?ā¼šå" Redisè >c"-a" 9 c" Yax€a1^a½±a "�i¾Ÿ

" \_cs\_éf/ym\_\_åt�é ± �â; sc- \$ W(Ré, tā ^è; ""ā "m " Redisā�šā; sc- \$ m — ¶ és "è"acs\_mop°ā^ ¶ è¿~æ~~需è|�æ^ 'ä×na \*a\$ ie \*••¢½\*? 2�既c. ¶ savea 16a\*pa 48a\*¤ä >c°;c~;é~sā;�ī¼Œé.£ä °ä»€ä¹^Redisè; ~è!�è°¾c½°è;™ä €

耕à, ae 'æce‰ā‡ ä, 'é-'é¢ æf'é-'\%s .

 è.±c¼°cš.é± ◆éD- 2020-08-14-09:17:36 �ā ^,ā½ è " 'cš..ä »c°¿c "‹forkā�c°¿c "‹ä¹Ÿä¼šé "-ā;♠¼Æë€Œä " ä »c°¿c "‹ā†····å"è¶Šā¤ § é~»ā;�è¶Šā¹···ā€,è€�ā, ^ā� "以ā†�æ·±ā···¥è™è§£ā, €ä, ‹forkā�ç°¿ç¨،çš"过ç¨،å~;?

è · é - °è € ♦ å , î ¼ Œãœ "æ · å ♦ °ã ½ ; ç " "aofã" Œã † · · å °ã ; «ç · · § çš "æ - ¹æ³ · æ - ¶ã € ™ã € ♦ æ ♦ ♦å\*\*åœ\*\*å♦šå···\*é‡♦å;«ç···§cš,æ-¶å€™(¼Œæ,···c©\*AOFæ-¥å;-ã€.æ\*\*åœ\*å···\*é‡♠å;«c ... 5 â% • Øæ .... c©°AOFæ—¥å; —è; Tæ⁻¯å... "é ‡ • å; «c... 5 å • Øæ .... c©°AOFæ—¥å; —å '¢ï¼Ÿ c²-c · ¥cš\_æf²ï¼Œåœ "å�šå··· "é‡�å¿«c··· § æ-¶å€™ä° § c" Yæ-°cš"AOFæ-¥å;-,ål.æ�œ æ ...cl0°AOFæ—¥å;—cš.è \* �T¼(F会抚æ-°cš.æ—¥å;—æ ...æ•%

Laperpap "Tå t "tap "e ä Wopå" t åne å " fl 

#### ä, ♦çŸYé♦ "有没有å^†æ♦♦æ£ç;®ğŸ",

å°♠å°‡å♠2B 2020-08-14 00:28:53

\$\delta\phi\rm \frac{\partial}{\partial}\phi\rm \frac{\partial}{\p

排形性: 69923