八、复制

- **通过设置slaveof配置选项或者执行SLAVEOF命令，可以让一个服务器复制另外一个服务器，复制分为：**

复制分为：

- 初次复制：服务器之前没有复制过其他服务器，或者服务器执行了SLAVEOF NO ONE命令；
- 断线后重复制：主从服务器在命令传播阶段断开网络连接，之后从服务器通过自动重连重新连接至主服务器继续复制；

1.旧版复制功能

实现：

同步阶段：

- 将从服务器的数据库状态更新至与主服务器当前数据库状态一致，通过从服务器发送SYNC命令实现。

具体流程：

- 从服务器发送SYNC命令；
- 主服务器接收SYNC命令后，触发BGSAVE命令的执行，在后台生成一个RDB文件，并将BGSAVE命令执行期间主服务器执行的所有写命令保存到内存缓冲区中。然后将生成的RDB文件发送给从服务器；
- 从服务器接收并载入RDB文件，将从服务器的数据库状态更新至主服务器开始执行BGSAVE命令那一刻所处的数据库状态；
- 主服务器将内存缓冲区中的所有命令发送给从服务器，从服务器接受并执行，保证主从服务器数据库状态一致；

命令传播阶段：

- 同步阶段完成后，主服务器每执行一条写命令，就将其发送给属下的所有从服务器，从服务器接收并执行，保证主从服务器的数据库状态实时一致。

缺陷：

- 对于初次复制来说，旧版复制功能能够很好的完成任务；
- 对于断线后重复制来说，如果断开的时间很短，为了补足主服务器执行的一小部分命令而重新执行SYNC命令是一个非常消耗资源的过程。

SYNC命令特点：

- 需要触发主服务器执行BGSAVE命令从而消耗大量的CPU资源、内存资源、IO资源；
- 主服务器向从服务器发送RDB文件需要消耗大量的网络资源；
- 从服务器接收并载入RDB文件时需要阻塞挂起，无法处理客户端发送的命令请求；

2.新版复制功能

在REDIS2.8引入了PSYNC命令代替SYNC命令，新版复制功能模式：

- 完整重同步。用于处理初次复制情况，和SYNC命令执行情况大体相同；
- 部分重同步。用于处理断线后重复制情况；

部分重同步功能实现：

服务器运行ID。

- 每个redis服务器在启动时都会随机生成一个16进制的40字节长的字符串作为服务器运行ID；

复制偏移量。

- 处于复制模式的主从服务器都会在内存维护一个复制偏移量。
- 主服务器每次发送N字节的数据，就将自己的复制偏移量加N；
- 从服务器每次接收N字节的数据，就将自己的复制偏移量加N。
- 通过对比主从服务器的复制偏移量，很轻松的知道主从服务器的数据库状态是否一致。

复制积压缓冲区。

- 处于复制模式的主服务器在自己的内存维护一块固定大小、先进先出的队列，默认大小1MB；
- 主服务器执行的每一条写命令不仅会被写入到aof_buf缓冲区中，还会保存到复制积压缓冲区中；
- 主服务器还会为复制积压缓冲区中的每个字节维护对应的复制偏移量。

PSYNC命令执行流程：

从服务器发送命令：

- 如果从服务器之前没有复制过其他服务器，或者从服务器执行了SLAVEOF no one命令，则向主服务器发送PSYNC ？ -1命令，主动请求主服务器对其执行完整重同步操作；
- 如果从服务器断线之前复制过其他服务器，则向主服务器发送PSYNC runid offset命令。其中runid代表断线之前复制的主服务器的运行ID，offset代表从服务器自己的复制偏移量。

主服务器接收、处理并返回命令回复：

返回+FULLRESYNC runid offset命令。

- 表示对从服务器执行完整重同步操作。
- runid代表自己的服务器运行id，从服务器接收并保存起来，待下次断线重连后发送给主服务器；
- offset代表自己的复制偏移量，从服务器接收并将该值作为自己的复制偏移量的初始值。

返回+CONTINUE回复。

- 对从服务器执行部分重同步操作，根据从服务器发送命令中携带的offset参数，到复制积压缓冲区中找到对应的命令，将之后的所有命令发送给从服务器。从服务器接收并执行，保证主从服务器数据库状态实时一致。

返回-ERR回复。

- 主服务器版本低于2.8，无法识别PSYNC命令。从服务器向主服务器发送SYNC命令执行完整重同步操作。

slaveof选项或SLAVEOF命令执行流程， 以SLAVEOF命令为例：

- 从服务器接收客户端发送的SLAVEOF命令，将命令中携带的参数信息即IP地址和端口号保存到服务器状态redisServer结构中的masterHost属性和masterPort属性中。

从服务器建立连向主服务器的套接字，套接字建立完成后：

- 主服务器将从服务器看做自己的一个客户端，并为其创建对应的客户端状态redisClient结构，并添加到自己的服务器状态redisServer结构中的clients链表表尾；
- 从服务器也会套接字分配一个专门的文件事件处理器用于处理复制工作；

从服务器发送一条PING命令，作用：

- 作用1、确保套接字的读写状态是正常的；
- 作用2、确保主服务器能够正常处理命令请求；
- 如果主服务器没有返回PONG回复，需要返回至第二步重新建立套接字连接。

进行身份验证。

- 情况1、如果从服务器没有设置masterauth选项，不需要进行身份验证，该步骤跳过；
- 情况2、如果从服务器设置了masterauth选项，向主服务器发送一条AUTH命令，命令中的参数即为masterauth的值。
- 主服务器在接收到从服务器发送的AUTH命令后，如果没有返回PONG回复，需要返回至第二步重新建立套接字连接。

发送端口信息。

- 从服务器向主服务器发送REPLCONFIG listening_port命令，将从服务器的监听端口号发送给主服务器。
- 主服务器接收并将其保存到从服务器对应客户端状态redisClient结构中的listening_port属性中。

执行同步操作。

- 从服务器发送SYNC命令或者PSYNC命令执行同步操作。
- 同步操作执行结束之后，主服务器也将成为从服务器的客户端。从服务器为其创建对应的客户端状态redisClient结构并将其添加到服务器状态redisServer结构中的clients链表表尾。

- 执行命令传播操作。

心跳检测：

处于命令传播阶段的从服务器会以每秒一次的频率向主服务器发送一条REPLCONFIG ACK命令，该命令作用：

- 检测主从服务器网络连接状态；

检测命令丢失；

- 如果主服务器接收到的从服务器的复制偏移量小于自己的复制偏移量，会到复制积压缓冲区中找到从服务器复制偏移量对应的命令，并将复制积压缓冲区之后的所有命令发送给从服务器。

辅助min-slaves选项实现，保护主服务器在不安全的情况下拒绝执行写命令。

- 1、min-slaves-to-write。从服务器最低数量限制；
- 2、min-slaves-max-lag。从服务器允许最大延迟限制；
- 如果从服务器数量小于min-slaves-to-write，或每个从服务器延迟值均大于等于min-slaves-max-lag毫秒，则主服务器拒绝执行写命令。