

MongoDB Sharding cluster backup and recovery solution



document author: Yongjie Lyu Yongjie.L@outlook.com

阿里云备份与恢复方案概览

备份与恢复方案概览

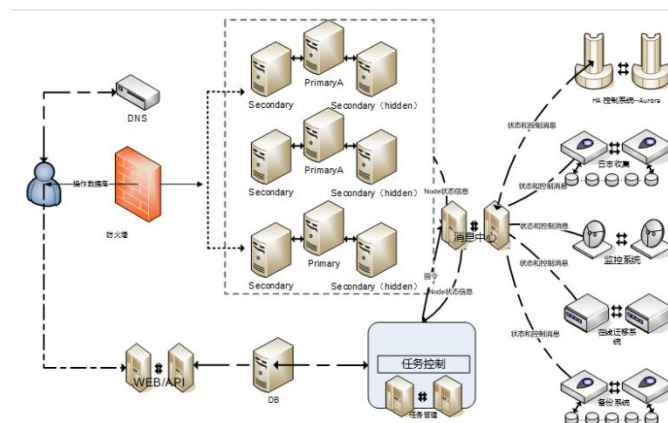
为防止系统故障等因素导致数据丢失或损坏，您可以通过云数据库MongoDB提供备份功能对数据进行备份，并在进行数据恢复时使用。

https://help.aliyun.com/document_detail/309504.html

备份数据库及数据恢复至MongoDB实例

阿里云MongoDB备份服务

Name	Tags
物理备份	自动备份策略
逻辑备份	手动备份策略
全量备份+增量备份	时间点备份



云数据库MongoDB自动搭建好3节点的副本集供用户使用，用户可以直接操作Primary节点和一个Secondary节点。

分片集群实例全量备份采用物理&逻辑备份的方式，不影响主节点(Primary)及从节点(Secondary)的读写性能，所有备份都在MongoDB实例的隐藏节点(Hidden)上进行。

MongoDB数据库备份手段

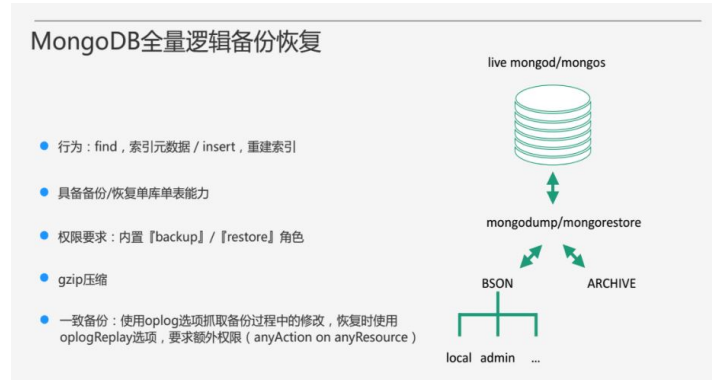
全量备份

1.Mongodump/Mongorestore全量逻辑备份/恢复

逻辑备份恢复就是通过使用官方mongodump和mongorestore两个工具在数据库层将MongoDB的数据进行导出和导入来备份恢复。mongodump可以连上一个正在服务的mongod节点进行逻辑热备份，支持并发dump多个集合。

- 备份

其主要原理是遍历所有集合，然后将文档数据一条条全部查询出来。数据为索引时仅导出元数据（例如索引建在哪个字段上、什么类型的索引、索引有哪些选项这些元数据，并没有把索引的数据本身导出来），实际索引数据在恢复时重新将数据insert进数据库，数据量大的话很耗时。



在mongodump执行过程中由于数据库还有新的修改，直接运行dump出来的结果不是一个一致的快照，需要将过程中的oplog也一块dump下来。

由于MongoDB的oplog是一个固定大小的特殊集合，如果dump过程很长，oplog空间又不够，oplog被滚掉就会dump失败。因此在dump前最好检查一下oplog的配置大小以及目前oplog的增长情况（可结合业务写入量及oplog平均大小进行粗略估计），确保dump不会失败。

阿里云MongoDB服务针对oplog做了弹性扩缩容的优化，能够确保在逻辑备份过程中oplog不被滚掉，一定能够备份成功。

全量逻辑备份恢复可以输出为两种格式：

- BSON格式：按照数据库生成各自BSON格式文件，方便单库单表的备份恢复
- 单文件：归档、压缩、传输，可以输出到一个文件方便备份恢复
- 恢复
 - mongorestore则是连上一个正在服务的mongod节点进行逻辑恢复。其主要原理是将备份出来的数据再一条条写回到数据库中。
- 优劣
 - 备份过程中数据库正常读写
 - 具备恢复单库单表的能力（便于某些场景的应用，例如紧急状态下需要恢复某一数据库的某一个表，此时不需要下载整个全量的数据备份，只需单独把想要恢复的表进行恢复即可）
 - 可跨存储引擎

2.全量物理备份/恢复

- 备份及恢复

物理备份的作用更接近底层一些，通过物理拷贝数据文件实现备份（例如作用在文件系统上，通过cp和tar文件系统工具将MongoDB的物理文件拷贝以进行备份。）

恢复时可以直接使用物理备份拷贝出来的数据文件启动mongod实例。
- 分类：
 1. 快照备份：依赖系统组件（linux逻辑卷管理器）快照功能，保留某一时间点磁盘的数据状态。

在单盘多租户的情况下，无法做到对块盘上每一个MongoDB实例单独进行备份

依赖于MongoDB的WiredTiger存储引擎的Checkpoints检查点+预写日志Journal实现宕机恢复。
 2. 文件拷贝备份：对目录进行拷贝。

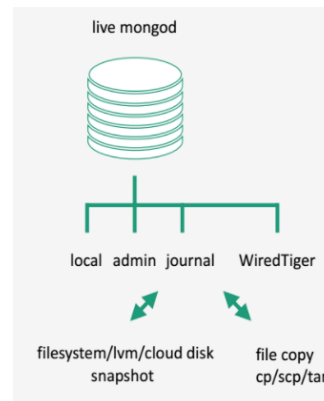
在文件拷贝开始之前需要执行db.fsyncLock的命令，即对MongoDB的全局加一个写锁，物理文件拷贝完毕执行db.fsyncUnlock解锁命令，间接达到一致备份的效果。

支持单盘多租户情况下的MongoDB实例单独备份

MongoDB全量物理备份恢复

- 行为：存储引擎相关，恢复时即插即用，无需重建索引
- 不具备备份/恢复单库单表能力
- snapshot VS file copy

	snapshot	file copy
实施依赖	filesystem/lvm/cloud disk provider	cp/scp/tar
单盘多租户支持	不支持	支持
一致备份	有Journal，万事无忧	db.fsyncLock/ db.fsyncUnlock



- 数据一致性

文件拷贝方式的热备份

- 优劣

速度快恢复时也不需要再建索引。

无法将某一个数据库的单独文件恢复出来，不具备单库单表的备份恢复能力。

MongoDB全量备份 逻辑备份 & 物理备份

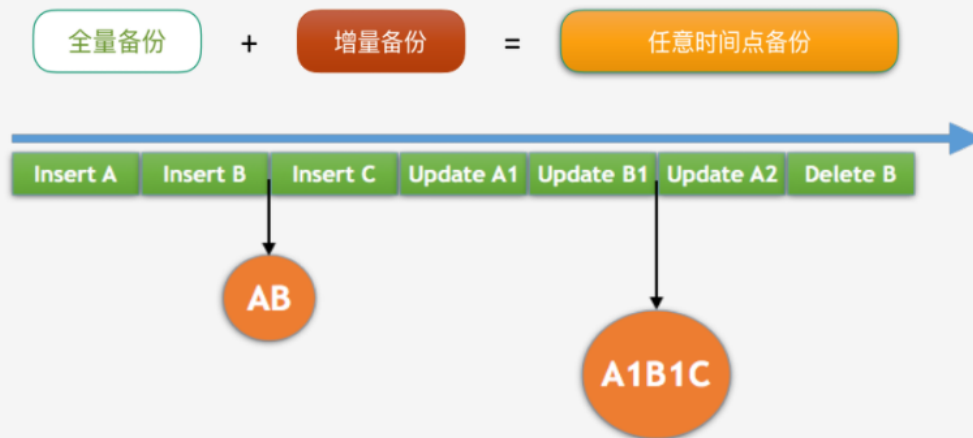
备份指标	逻辑备份	物理备份
备份效率	低	高
恢复速度	低	高
备份影响	与业务争抢数据库资源	数据库暂不可访问
备份文件大小	原文件更小	原文件相同
兼容性	兼容大部分版本	无法跨存储引擎备份，成功率高

增量备份

MongoDB的增量备份可以通过持续抓取oplog来实现。

抓取oplog主要的难题也和使用mongodump进行全量备份一样，需确保要抓取的oplog不被滚掉。

MongoDB增量备份



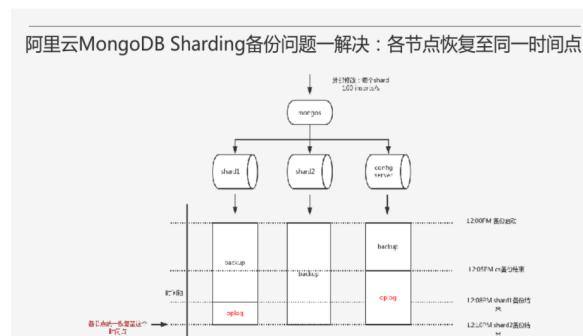
分片集群备份

1. 分片集群各节点恢复至同一时间点

集群多个节点在外部修改情况下如何取得一致备份。

集群当中每个节点容量的不同导致节点备份的耗时不同，当对应用进行写入时，由于每个节点备份结束时间不同，有些节点的备份会多包含一些新写入的数据，其中备份结束的时间点很难进行确定。

采用全量备份加增量备份做到各节点备份恢复至同一时间点，在备份结束比较早的节点可以多抓取一些oplog，备份结束比较晚的节点可以少抓取一些oplog，从而保证各自节点的备份加oplog能够对应到同一个时间点。



2. 集群内部有数据迁移时需要停止备份

集群进行扩容的时候（增加或删除Shard）或数据分布不均时自发地进行数据迁移的时候不能进行备份。因为Chunk迁移可能导致数据的重复与丢失问题。

1. 计划时间窗口内进行数据迁移（阿里云解决方案）

用户在MongoDB Sharding备份时可以配置一个迁移的时间段，即用户可以根据业务访问行为指定迁移在哪段时间进行，从而保证迁移在预期时间段内进行，其它时间段可以进行备份恢复。

2. 备份之前加Balancer锁

备份开始之前使用 `sh.stopBalancer()` 加锁，之后开始备份操作，备份完成之后开锁

引用资源

开发者头条

最近数据库真是多灾多难，前段时间刚刚爆出MongoDB数据库安全问题。这两天又被炉石传说数据库故障给刷屏了。 ...

 <https://toutiao.io/posts/9iy3n8/preview>


我们十分理解广大玩家的焦急的心情，也曾在事故发生后的最初两天尽力做出其他的各种尝试，但效果及进度均不理想。在此期间游戏环境已变得不稳定，而游戏的维护时间也已超过24小时。

在尝试了各种解决方案后，暴雪和网易最后综合考虑，决定将所有游戏数据回档至事故发生前状态（即2017年1月14日15:20）。我们需要向大家说明，游戏回档是我们最后的无奈决定，暴雪和网易对被迫做出这个艰难的决定深表遗憾。

游戏回档意味着，自事故发生以来的所有英雄等级提升、卡牌变动以及天梯排名等均无法复原。我们一贯重视玩家的游戏体验，也珍惜玩家在炉石当由投入的心血和时间。由于此次炉石推出以

技术分享|技术分享|数据库的自我修炼--阿里云MongoDB备份恢复功能说明和原理介绍|虚拟机备份专家云祺科技

本次直播视频精彩回顾，戳这里！直播涉及到的PPT，戳这里！演讲嘉宾简介：郑泮（花名：明俨） 阿里云技术专家，2011年加入阿里，曾参与TFS、Tengine研发，目前主要参与阿里云MongoDB云数据库服务研发，主要关注分布式存储、数据库等领域。 ...

 <https://www.vinchin.com/blog/vinchin-technique-share-details.html?id=9654>



https://s3-us-west-2.amazonaws.com/secure.notion-static.com/b629881d-71aa-493a-9329-89d454eb0d65/技术分享_技术分享_数据库的自我修炼--阿里云MongoDB备份恢复功能说明和原理介绍_虚拟机备份专家云祺科技.html