

MySQL Backup, Recovery & Partition

请你详细描述如何通过全量备份和增量备份来实现系统状态恢复。

应该将全量备份和增量备份两者结合。每隔一段时间（一般而言相对长一点，而且可以挑在服务器负载比较低的时间段进行，假设每隔6小时一次）就在另一台服务器上进行全量备份，将数据库的所有内容进行备份。增量备份将数据库变化写入bin-log文件（bin-log文件也应该位于另一台服务器上，或者存储在不同的物理设备上）。当数据库服务器挂掉的时候，应该先通过全量备份恢复到某个时间段数据库的状态，然后删除该时间段之前的所有bin-log文件（因为已经用了全量备份，该时间段之前的增量备份就没有必要存储了），接着再用该时间段之后的所有增量备份对数据进行恢复。

请你根据MySQL缓存的工作原理，描述预取机制的优点。

MySQL默认引擎InnoDB有两种预读取方式，分别为线性预读和随机预读。常用线性预读，后者默认为OFF。线性预读每次都会读一个extent（InnoDB以64个page为一个extent），可以通过配置 `innodb_read_ahead_threshold` 参数，指定读取下一个extent的时机（如果一个extent中的被顺序读取的page超过或者等于该参数变量时，InnoDB将会异步的将下一个extent读取到buffer pool中）。通过预读，MySQL可以将可能在将来被用到的数据提前加载到缓存中，这在很多场景下都是有很大的好处的。比如一些数据查询需要获取表中一定范围的数据，而一张表如果已经进行过optimize操作的话，其中的数据将会是连续存放的。这样一来，通过预读的方式，可以一次性获取较多的、将来会使用的数据。这样等将来要从表中获取数据的时候，就可以直接从缓存中获取，效率会得到提升。实际上很多操作都具有空间局部性，使用预读机制则很好地利用了这一性质。

请你按照你的理解，阐述Partition机制有什么好处？如果数据文件在一台机器上有足够的存储空间存储，是否还需要进行Partition？

Partition机制可以将一张大表根据某些指标或者数据特征分成多个分区进行存储。Partition机制有以下好处：

- 由于分区文件可以存储在不同的物理磁盘上，所以使用的分区的表可以比没有使用分区的表存储更多的内容；
- 由于增删改查操作可以在查询中指定分区，所以使用分区可以提升查询的效率。比如只查询某个分区中的数据，这样会比查询整张表效率高很多。
- 一般而言表都会根据某一特征或者指标进行分区，所以这一特性也会加快增删改查的速度：比如有张表存储学生的信息，那么可以将该表分为研究生和本科生两个分区。如果我要查询有关本科生的信息，那么我只需要在本科生的那个分区中进行搜索就行了；再比如如果我需要删除所有本科生的信息，那么我只要删除本科生那个分区对应的所有数据即可。这样会比在一张表上进行相应操作快很多；如果按照某个连续值进行分区，那么根据该连续值的范围搜索的效率也会得到提升。
- 将分区文件存储在不同的物理磁盘上，可以在多个磁盘上传播数据，实现更高的查询吞吐量。

如果数据文件在一台机器上有足够的存储空间存储，是否还需要进行Partition？答案是可以根据业务需求决定是否需要进行分区。如果表是类似于上面提及的学生信息的表，那么使用分区还是很有必要的。因为很多增删改查对于研究生和本科生是有区别的，比如统计GPA等等。将这张表分成研究生和本科生两个分区，那么就可以很好地应对针对某一种学生的操作，比如统计研究生有多少人，统计本科生平均GPA等等。所以，分区的意义不仅仅在于增加表的容量，合理利用存储资源，而且可以通过分区增加某些增删改查操作的效率。

结合我的E-BookStore进行说明

- 备份

对于我的E-BookStore而言，书籍、订单、用户都是非常重要的信息，是需要重点进行保护的，所以必须要定期进行维护和备份。三者均使用全量备份+增量备份相结合的方案，开启bin-log选项。可以每12个小时进行一次全量备份，将数据库内容dump到另一台服务器或者另一个物理存储设备上（具体的时间间隔应该根据各种信息和统计决定；由于全量备份对系统性能有一定影响，为了防止系统性能受到波动，应该选择服务器负载比较小的时候进行全量备份工作，比如半夜）。注意全量备份和增量备份的数据都必须存储在另一台服务器上，或者至少是另一台物理存储设备上（可以使用RAID）

- 预取机制

由于我的E-BookStore有大量涉及范围搜索的服务，所以使用预取机制可以增加效率。预取对于一些需要分页的内容也十分友好，比如首页显示的书籍，每次都是获取一页的内容，用预取就非常适合。

- 分区机制

可以根据书籍分类（比如科幻、言情）将书籍进行分区。由于用户查询很多时候会指定书籍分类，比如有些用户只想看科幻书籍，那么这时候使用分区就可以很好地优化相关查询，只要显示的指定只在科幻书籍的那个分区搜索就行。对于用户，可以根据用户权限分类：（被封禁、普通用户、会员、管理员），这样也能提升一定的效率（比如我可能需要研究分析所有会员的消费习惯，那么只需要研究会员那个分区的就好）

由于书籍表会很大，订单也是（不过这小破店不像是有这么多订单的样子XD），所以分区是很有必要的，可以合理利用存储资源，增加表的容量，容纳更多的内容。

分区可以和分表一起配合使用，两者不冲突。