

作业十：数据库备份

第一题：叙述物理备份和逻辑备份各自的优缺点

- 逻辑备份：先要备份表的结构，然后记录所有的SQL语句的操作，比如所有的增删改查的操作。当一个表崩溃的时候，在另外一个表里面按照顺序执行一遍这些操作，就可以实现备份的恢复。
- 物理备份：拷贝整个表格，索引文件、数据文件、元数据文件，当出现故障的时候，把所有的文件拷贝回去就好了。

先说逻辑备份优点：

- 因为备份的是SQL语句，所以换一个机器也能执行，换一个数据库系统也能执行。比如从Mysql导出的SQL脚本备份，换到MongoDB也一般可以执行，兼容性好。

缺点：

- 空间占用比较大，SQL的脚本文件因为包含了SQL的关键字，所以比只存储数据的表格文件是要大的，牺牲了一部分的空间，在数据量比较大的时候，导出的SQL备份脚本文件会比较浪费空间。
- SQL文件如果泄露了的话，数据库就泄露了，安全性差

先说物理备份优点：

- 适合大的数据库
- 恢复速度快，出现问题的时候直接把文件拷贝回去就好了，方便快捷恢复便利。
- 相对来说安全一些

再说物理备份的缺点：

- 比如从Mysql备份的内容就只能用于Mysql，不能像逻辑备份一样导入到别的类型的数据库。

补充一个：在线备份和离线备份：

- 在线备份是系统在运行的时候备份，缺点管理很复杂，涉及到加锁
- 离线备份是数据库暂停服务，然后专门备份。缺点是用户不能使用了。

第二题：参照上课的举例，详细描述如何通过全量备份和增量备份来实现系统状态的恢复。(2分)

假设我们的备份策略是每个周日下午一点进行一次全量备份，也就是把数据库里面的所有数据全备份一次。

```
1 shell> mysqldump --single-transaction --flush-logs --master-data=2 \  
2      --all-databases --delete-master-logs > backup_sunday_1_PM.sql
```

那么，在两次全量备份之间的时间节点上，数据库每次执行写入操作的时候，都会对bin-log文件进行操作，写入增量备份。也就是记录用户对数据库的所有操作变更。

现在加入在周三的中午，数据库主机突然崩了，那么我们首先要恢复到最近一次做的全量备份：

```
1 shell> mysql < backup_sunday_1_PM.sql
```

好了，现在就已经恢复到最近一次全量备份的状态里，然后要从上次全量备份到崩溃期间的binlog文件全部读取出来，然后恢复。假设是：`gbichot2-bin.000007` `gbichot2-bin.000008`

```
1 | shell> mysqlbinlog gbichot2-bin.000007 gbichot2-bin.000008 | mysql
```

如果每次磁盘坏死或者磁盘错误，基本上就能恢复到原来的状态。但是如果有一个binlog文件崩溃了或者丢了，那就可能恢复不到周三崩溃的时候的状态了，可能只能恢复到周二的下午或者某个稍微更早的时间点。

第三题：分区的好处。(1 分)

- 当一个表里面的数据太多了的时候，占用的空间太大，可能就接近磁盘的存储空间，这时候就需要分区，把表的数据拆开（分区），然后让不同分区的数据存储到不同的磁盘上面
- 提高性能：由于满足给定WHERE子句的数据只能存储在一个或多个分区上，这将自动从搜索中排除任何剩余分区，因此某些查询可以得到极大的优化。
- 失去有用性的数据通常可以通过删除仅包含该数据的分区（或多个分区）轻松地从分区表中删除。相反，在某些情况下，通过添加一个或多个新分区来专门存储数据，可以大大简化添加新数据的过程。

第四题：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？为什么？

这个需要考虑业务逻辑，分区可以解决存储空间不够的问题，也可以对于性能有一定的优化。并不是说空间足够就不需要分区，但是如果出现空间不够的话，一方面可以考虑升级机器的存储空间，另一方面可以考虑数据库表的分区。

例如，如果我的电子书店业务地域分布广泛，在世界各地都有业务。那么，为了访问速度更快，我们可能需要分布式的数据库存储。比如我把订单的表格分区，然后在亚洲设置一个服务器存储亚洲客户的订单数据，然后在欧洲设一个服务器存储欧洲用户的订单数据，等等以此类推，那么假如管理员想要查看全球的业务，由于sql基本的增删改查都是支持分区的，所以也可以查询到全球客户的数据。这样一来，用户访问我的电子数据下订单的速度也变快了，同时不需要修改任何的电子书店后端源代码，管理员也可以查看全球的订单。同时我也可以设置分区的区域管理员，这个区域的管理员也可以非常高效的查看当地的客户数据。

类比我们上课讲的，分区相当于在B+树的顶点再加了一层，所以就是说一般会选择具有特定特性的一列，或者说经常要用到作为筛选条件的一列用来分区，比如不同的区域（亚洲、欧洲）等等。