链滴

记一次清空数据仓库的过程

作者：someone9891
原文链接：https://ld246.com/article/1553941852324
来源网站：链滴

始于删库

2019年3月30日

生产环境数据仓库被我清空了。

目前，通过DBA大神的帮助，数据已经还原，DBA还原完数据，还不忘安慰我，让我不要有什么心里担；其他同事也一起安慰我……

同事打趣的说：毕竟删过库，也算没白当一回程序员。

领导轻描淡写的说：那赶紧想办法恢复回来吧。

在他们看来，也许数据删了，还能恢复，就不算什么天塌下来的大事。

但对于我来说，是对灵魂对一次清洗。天天调侃删库跑路，何曾想，自己也会有删库的这一天。

那一刻，突然开始怀疑人生，怀疑自己这么多年的从业经验，一直以为删库这种事情，只会发生在经尚浅、入行不久的程序员身上。可怕的不是删库了，而是，自信的认为自己的每一步操作，都是经过思熟虑的，任何敏感的操作，都不会轻易去做，而正是这种思想，让一些常规简单操作，变成了不必证即可执行的操作。

思于后

其实，很多人经常调侃删库跑路，甚至连圈外的人也在说，但没有真正经历过的人，都不会明白，这味着什么。在目前这个信息化的时代，任何有自身业务的企业，数据都是企业之本，数据的丢失，是个企业不可挽回的灾难。

我是经历过两次数据丢失的，加上身边发生的，应该说是有三次。

我觉得非常有必要把这个过程记录下来，一来，虽说数据都还原了，但是也给我们敲响了一记警钟：据的安全不仅仅是各种安全策略防攻击、防脱库，最重要的，还得防自己人的脑残操作。

事情虽然是过去了，但是想想还是挺恐怖的，好在使用的数据库是oracle，虽然目前还没来得及做备，但好在数据原本是从另外一个库迁移过来不久的，假如使用的是mysql呢？假如数据原本就是这个里面但，也没有良好但备份呢？很多事情都是需要去认真思考的。

记其所以删库

如前面所说，数据丢失这种事情，耳闻目睹亲手所为，一共三次，天灾人为均有之。

耳闻

第一次是在广州的时候，公司旗下的一家子公司，有员工误操作，删除了ERP系统的数据库，具体细不得而知，只知道，请了我旁边的同事去外援，最后听说是恢复了……

目睹

第二次还是在广州的时候，非人为，属于服务器硬件故障。清晰的记得，那是一个星期天，早上8点右，我还躺在床上睡觉，突然接到了好几个电话，说系统故障，无法使用，我从床上爬起来，尝试访系统，失败…… 联系同事查看服务器信息，同事说 SSH 已无法连接。

因为同组有同事住的近，于是他前往公司机房查看，我在家通过企业qq、电话等方式，向公司数千等使用系统的同事 发公告、说明情况。正值业务高峰期，公司群内炸开了锅，询问信息、电话如洪水般来，我只能复制了信息，一一粘贴 发送……直到下午，才赶去公司帮忙。

那一次的事故，应该说是灾难性的，数据库磁盘阵列两块硬盘损坏，而要命的是，数据库备份策略都备份在本机的，所以短时间内，是没法通过备份文件等途径来恢复数据库的。

当时做了两件事情：

1. 立即换新硬盘，重新搭建数据库环境
2. 坏硬盘交给专业数据恢复的公司进行恢复

不幸中的万幸是：在事故发生三天前，我同事把数据库导出来一份在自己电脑上，搭建好环境，可以将这份数据导入进去，但是毕竟三天啊，中间丢失的数据也是不少的。

恢复工作持续到半夜1点左右，技术总监开车送的我回家。

事后，听说恢复原旧硬盘需要耗费一周多的时间，于是我们决定，在同事三天前的数据基础上，人工录丢失的数据，号召各分公司、各部门同事回忆三天来的业务数据并进行补录，期间一些数据的时间题，我们后台协助。

事故造成 的影响持续来一周左右，光系统停运行三天左右，这三天中，公司的业务一下子回到了原始会，所有的业务数据通过员工个人手动记录。

亲为

接下来就说发生在昨天的这件事了

差不多下午四点多的时侯，远在杭州的DBA打电话给我，说数据仓库有两个表数据被清空了，会不会我的程序哪里出了问题。因为我最近一直在负责一套ETL 的流程，会将十余个业务系统的数据，进行洗、转换、检核后，集中到数据仓库中，所有的程序都是大面积的会 频繁操作数据库，所以第一时间问我也是合适的。

但是我很自信的告诉他，不可能是我这里的问题，对于数据库我还是比较谨慎的，在没有十足把握的况下，我都没有直连生产库，并且就算我连了生产库，也是没有任何 drop、delete、truncate 相关令的，最多就是某个表数据清洗不对，造成数据异常。

但是没多久，我就听到消息说，并不是两个表，虽然第一时间两个表出现了问题，但是事实证明，是有的表都被清空了。大家都在忙着排查问题，看是否黑客入侵，或者程序BUG。通过查看日志，证明据库 被循环所有表执行了delete命令，证明是人为的，所以必须要找出是通过什么途径执行的命令。

与此同时，我努力的在想，我最近做过什么操作。

我想起来，在3月29日晚上，我在生产库做了导出所有表结构的操作，因为我自己建了一个临时的schma ,用于测试，但是我需要跟生产库保持同样的表结构，最简单的方法，就是在生产库导出表结构，导入到 我自己建的schema 里面。

但是这个操作我也是不是第一次做，怎么也不会直接清空了数据吧，况且，我是29 号进行的导出，据是在30 号下午被清空的，那一定不是我导出造成的。

我又回忆了一下，30 号我还做过什么操作。

没错，29号导出的表结构，我在30号的时候，进行了导入操作。时间正好就是在30号下午4点左右。为我做的ETL 流程需要做演示，为了模拟一个完全一样的数据仓库，我是在数据仓库的 oracle 里面建的 schema ，平时我都是执行哪个表，就新建哪个表，但是如果要演示，程序跑一半报表不存在建表的话，会造成演示不太顺利，所以，我就打算批量进行表结构的导入。

但是，在另外有一个 schema ，通过另外的用户名密码登录，执行的 导入操作，怎么会清空原schem 里面的数据呢？ 我突然想到，为了方便，我是通过导出 sql 语句，直接 去新的 schema 执行的方式，难道是sql 语句的问题？

我打开 sql 语句进行查看，果然就是这里出问题了，sql 语句部分如下：

```
DROP TABLE "LZUDATALWL1"."BKS_XWXLZL";
CREATE TABLE "LZUDATALWL1"."BKS_XWXLZL" (
  "ID" VARCHAR2(32 BYTE) VISIBLE DEFAULT SYS_GUID()   NOT NULL ,
  "BYYXXHDW" VARCHAR2(60 BYTE) VISIBLE NOT NULL ,
  "BYZYZH" VARCHAR2(20 BYTE) VISIBLE ,
  "HDXWM" VARCHAR2(3 BYTE) VISIBLE ,
  "HXWMLM" VARCHAR2(2 BYTE) VISIBLE NOT NULL ,
  "HXWRQ" VARCHAR2(8 BYTE) VISIBLE ,
  "HXWZYM" VARCHAR2(10 BYTE) VISIBLE ,
  "JSXYNY" VARCHAR2(6 BYTE) VISIBLE NOT NULL ,
  "RXNY" VARCHAR2(6 BYTE) VISIBLE NOT NULL ,
  "SXWDWMC" VARCHAR2(60 BYTE) VISIBLE ,
  "SXWGJDQM" VARCHAR2(3 BYTE) VISIBLE ,
  "SXZYM" VARCHAR2(10 BYTE) VISIBLE ,
  "XH" VARCHAR2(20 BYTE) VISIBLE NOT NULL ,
  "XLM" VARCHAR2(2 BYTE) VISIBLE NOT NULL ,
  "XLZSH" VARCHAR2(20 BYTE) VISIBLE ,
  "XWWYHZXXM" VARCHAR2(36 BYTE) VISIBLE NOT NULL ,
  "XWZSH" VARCHAR2(20 BYTE) VISIBLE ,
  "XXFSM" VARCHAR2(1 BYTE) VISIBLE NOT NULL ,
  "XXXSM" VARCHAR2(2 BYTE) VISIBLE ,
  "XZ" NUMBER(3) VISIBLE NOT NULL ,
  "XZXM" VARCHAR2(36 BYTE) VISIBLE NOT NULL
)
TABLESPACE "LZU_DL"
LOGGING
NOCOMPRESS
PCTFREE 10
INITRANS 1
STORAGE (
  INITIAL 65536
  NEXT 1048576
  MINEXTENTS 1
  MAXEXTENTS 2147483645
  BUFFER_POOL DEFAULT
)
```

你没看错，我就是脑残的执行了这个sql 语句，清空了原 schema 里面的数据，sql 的逻辑是先drop ，后create，而 drop 和 create 的时候，表名前，都是包含 用户名 作为前缀的，所以，不管我拿这 sql 去哪里执行，只要是在对原 schema 有操作权限的地方，其实都是在操作 原来的库，而我这个有表结构，所以，执行一遍以后，表结构还在，数据都没有了。

三省

事情已经发生，数据也已经恢复，但是有些教训还是必须要记住的：

- 数据库备份，一定不能备份在本机
- 常用的数据库用户，权限一定不能太高
- 每一步操作，尤其是执行sql,记得先查看sql 内容。

另外：如果真的发生了删除这类事情，不要慌，第一时间回忆自己做过的操作，如果真是自己行为导的，要第一时间站出来说出来，不要让同事在排查黑客攻击与程序bug 上无谓的浪费时间，因为假如的是程序bug 或者黑客攻击，可能第一时间恢复数据是起不到实质性作用的，必须要 先解决 这类问，才能真正的恢复数据，既知 是人为，就站出来给大家省点宝贵的时间吧。