

上周更新了两篇 Excel 学习，有读者反映写得不够实操。这里解释下，文字真的很难将每一步细化，不然真的要出一本书。**文章更多是一个引路**，告诉大家为什么要学这个，主要学什么，怎么学？深入的操作，还需大家业余加餐啊，已经尽量给大家留了材料了。

也有人提议，出个视频，不错！很赞！明年计划下。即使没有时间，我也会推荐一些我认为 OK 的视频让大家学习。

年底了，近几个晚上都在苦熬年终报告，各位写好了么？

好了，回到正文，本篇讲数据库知识：

经常有刚从事数据分析的职场萌新，问我做数据分析工作要学些什么，应该怎样规划学习路径。我会告诉他：如果你 Excel 还用的不溜的话，就先学学 Excel，当你用 Excel 处理和分析一些小数据集没有问题的時候（具体表现就是：常用函数公式信手拈来，数据透视表，筛选，排序，图表绘制操作熟练），你就去学习 SQL 语言，然后用 BI 去分析去熟悉业务。然后到了一定阶段，你可以上手 R 或者 Python。后面如果你能更进一步，可以去了解一些 Spark 等大数据框架。

为什么要学习 Excel？

首先 Excel 是我们最常用的数据分析和处理工具，Excel 的功能非常丰富，基本可以涵盖我们在之后在其它软件(SQL、BI、Python、R)中要学到的那些功能。

有的人可能会问，既然 Excel 这么强大，为什么还要学其它的工具？这是因为 Excel 是通过菜单的形式来进行操作的，很难实现自动化和功能复用，当然你也可以通过 VBA 来实现，用 VBA 也就是编程了，不过因为 VBA 这种语言学会了基本只能在 office 软件中使用，学习的投入成本和产出收益不成比例，不推荐学习，这是客观原因之一；另外就是 Excel 在处理比较大的数据集的时候，性能很差，

并且经常崩溃。(虽然 Excel2013 及以上版本宣称可以容纳 100+万条记录，但几万条数据就开始卡顿了)。

为什么 Excel 学完要学 SQL ?

客观原因是绝大部分数据分析岗都有 SQL 技能的要求。企业里面为了保证数据的安全性和管理的方便，数据都是统一存放在数据库中，从数据库中提取和查询数据需要使用 SQL 语言，甚至有的公司就是用 SQL 语言来做数据分析。

另外一个原因就是即使你先学了其它的工具，比如 R，Python，甚至 Spark 等大数据框架，你会发现最后你还是得学习 SQL。如果你先学习 SQL，那么很多概念你都能在学习 R，Python，Spark 等更加复杂的工具之前弄清楚。对于后面的学习会有帮助。这就好比建房子，都是先打地基，然后一层一层的盖。

SQL 语言的学习排在 Excel 之后，其它工具之前，还有一个很重要的原因就是，SQL 可以在一定程度上帮 Excel 解决大数据集的问题，同时架起一个通往其它工具的桥梁。

关于数据库和 SQL 的学习，也是分为两篇，**第一篇讲数据库以及表的概念。第二篇是 SQL 语句的掌握和数据库的操作。**

一、数据库基础知识

先谈一下我对数据库的理解。数据库顾名思义就是数据的集合，是由一张张数据表组成的。

放在物理实体上，是一堆写在磁盘上的文件，文件中有数据。这些最基础的数据组成了表(table)，我们把它想象成一张 Excel 的 sheet，如下图：

学生ID	学生姓名	年龄
1231	张三	18
1232	李四	19
1233	王五	18

每一张表都有一个唯一标识，即主键，也就是 ID。ID 是数据库中重要的概念，叫做唯一标识符 / 主键，用来表示数据的唯一性。就相当于我们的身份证，是唯一的，有了身份证，就知道数据在哪了。

ID 通常没有业务含义，就是一种唯一标识，每张表只能有一个主键，且主键通常是整数，主键一旦设立，值通常不允许修改。

数据库是表的集合。一个数据库中可以放多张表，我们给每张表命名，表与表之间能互相联系。联系就是数据能够对应匹配，正式名称叫联接，对应的操作叫做 Join，我们想象成 Excel 中的 vlookup。

学生ID	学生姓名	年龄		老师ID	学生ID	老师姓名
1	张三	18		1	2	王梅
2	李四	19		2	4	李海
3	王五	18		3	3	吴凤娜
4	张柳	19		4	5	邱敏
5	费帆	18		5	1	刘军

比如上面两张图，左图是学生信息表，右图是老师信息表。左图的主键是学生 ID，右图的主键是老师 ID。细心的读者可能发现右图还有一个学生 ID，这里的学生 ID 是专门用来联接用户表的，它并不是主键。只不过两张表通过学生 ID 这个唯一信息来关联。

但两张表关联也并不是信息能一一对应的，也会存在空缺的时候，比如：

学生ID	学生姓名	年龄		老师ID	学生ID	老师姓名
1	张三	18		1	2	王梅
2	李四	19		2	4	李海
3	王五	18		3	3	吴凤娜
4	张柳	19				
5	费帆	18				

那两表建立连接就会变成：

学生ID	学生姓名	年龄	老师ID	老师姓名
1	张三	18	1	王梅
2	李四	19	null	null
3	王五	18	3	吴凤娜
4	张柳	19	2	李海
5	费帆	18	null	null

了解上面的概念，你就知道什么叫关系型数据库。简单说，它是由多张能互相联接的二维行列表格组成的数据库。在数据准备时，我们通常要建立表关联来分析。

关系型数据库是基于关系代数模型发展而来，常用的关系型数据库有 SQL Server、MySQL、Oracle、DB2 等，这个视企业使用为准，我们后续学习都以 MySQL 为主。

各关系型数据库（不感兴趣可跳过）：

DB2: 关系型数据库，适用于大型的分布式应用系统，确实是非常非常好的数据库，无论稳定性，安全性，恢复性等等都无可挑剔，而且从小规模到大规模的应用都非常适合。但是使用起来觉得非常繁琐，安装的时候要求颇多，很多软件都可能和 DB2 产生冲突，因为一般 DB2 都是安装在小型机或者服务器上的，所以在 PC 上安装很费事儿。新建一个库需要设置很多东西，分配各种各样的存储空间。

Oracle: 是目前市场占有率最大的数据库，我在学习 SSH 的时候用的就是 Oracle，安装起来很繁琐，而且居然程序文件有 3G 之多... 用起来非常方便，对于我这样的初学者，有很简单的配置，对于要求

很高的企业级应用, 也有很复杂的配置和管理方法, 有很强大的数据字典, 可以说是最实用的数据库了, 但是查了一下, 价格不菲...

MS SQL: 当初用的是 2000 和 2005 版, 这两个版本差了很多。2000 的数据库做的不错, 程序很小, 操作简单, 功能较全, 算是各方面都很中庸的数据库吧, 是中型数据库, 我的毕业设计就是用 MS SQL 2000 做的。2005 中加入了很多功能, 复杂多了, 有大型数据库的风范了, 而且价格也变高了, 个人觉得除非用的是 Window Server 系统或者针对 Microsoft 产品, 否则不如用 Oracle 好。

MySQL: MySQL 是一个很好的关系型数据库, 免费, 而且功能很全, 程序又小, 安装简单, 现在很多网站都用 MYSQL, 在字段约束上做的差了点儿, 其他的都不错, 和 MS SQL 用着差不多。

Access: 典型的桌面数据库, 觉得做个单机系统, 比如记账, 记事儿什么的还成, 在局域网里跑个小系统都够呛, 数据源连接很简单, 因为是 Office 的数据库, 所以 Windows 自带数据源。

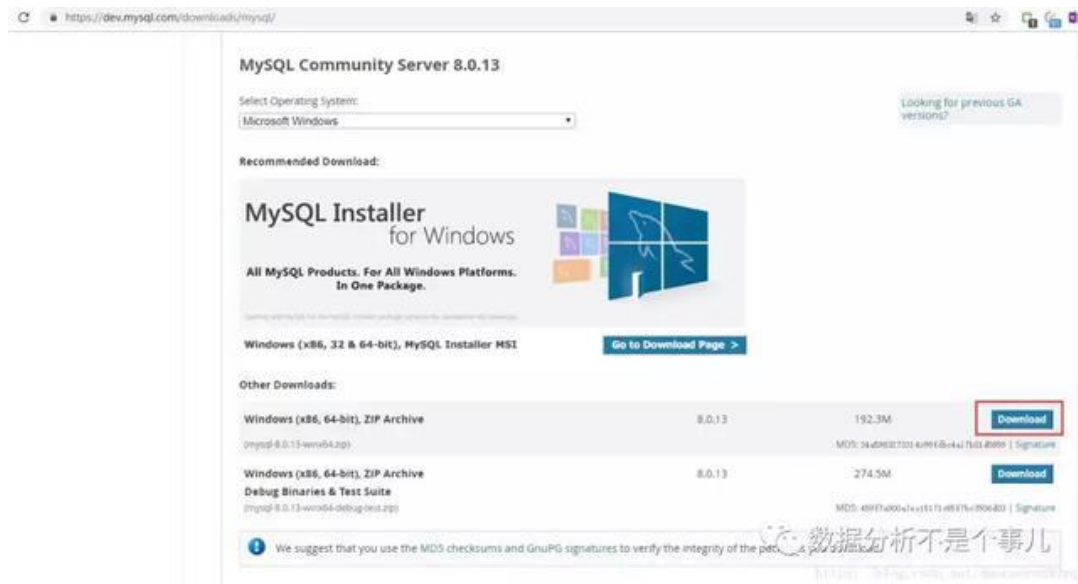
更多关于数据库的知识, 看一本《数据库系统概论》就够了。

二、尝试使用 MySQL 数据库

如果还没有接触过数据库或者说 SQL, 推荐大家下载并安装 MySQL 数据库进行尝试, MySQL 数据库的下载安装比较简单, 安装完成就可以使用。

可以访问 MySQL 官网进行下载, 网址如下 (这里给出的是 Windows 版下载地址) :

[\[链接\]](#)



下载 MySQL 时，需要进行免费注册，注册页面是英文的。下载该工具后，点击运行，这个程序就能够联网自动为你安装 MySQL，并自动进行配置。在安装过程中，需要你设置密码，自己设置一个登陆密码，并记住，下次登陆 MySQL 时，需要这个密码。

解压之后没有 my.ini 文件（我的端口设置的 3308），截图中是我增加的。

```
1 [client]
2 port = 3308
3 default-character-set = utf8
4
5 [mysqld]
6 port = 3308
7 character_set_server = utf8
8
9 basedir=D:\MySQL\mysql-8.0.13-winx64
10 datadir=D:\MySQL\mysql-8.0.13-winx64\data
11
12 group_concat_max_len=20000
13
14
15 [WinMySQLAdmin]
16 D:\MySQL\mysql-8.0.13-winx64\bin\mysqld.exe
```


管理员命令行：

在 MySQL 安装目录的 bin 目录下执行命令：

mysqld --initialize --console

得到如下结果：

```
1 D:\MySQL\mysql-8.0.13-winx64\bin>mysqld --initialize --console
2 2018-10-24T05:32:45.732454Z 0 [System] [MY-013169] [Server] D:\MySQL\mysql-8.0.13-win
3 2018-10-24T05:32:45.870611Z 0 [Warning] [MY-013242] [Server] --character-set-server:
4 2018-10-24T05:33:14.879343Z 5 [Note] [MY-010454] [Server] A temporary password is gen
5 2018-10-24T05:33:37.812581Z 0 [System] [MY-013170] [Server] D:\MySQL\mysql-8.0.13-win
6
7 D:\MySQL\mysql-8.0.13-winx64\bin>
```



数据分析不是个事儿

看到有个警告，查了一下 mysql 建议使用 utf8mb4

修改 my.ini 文件

```
1 [client]
2 port = 3308
3 default-character-set = UTF8MB4
4
5 [mysqld]
6 port = 3308
7 character-set-server = UTF8MB4
8
9 basedir=D:\MySQL\mysql-8.0.13-winx64
10 datadir=D:\MySQL\mysql-8.0.13-winx64\data
11
12 group_concat_max_len=20000
13
14
15 [WinMySQLAdmin]
16 D:\MySQL\mysql-8.0.13-winx64\bin\mysqld.exe
```



数据分析不是个事儿

警告没有了，上面打印了默认密码：nj>uUJkpH4/I

然后，安装服务：

mysqld --install MySQL8.0

因为这是电脑上第二个 mysql 所以服务名改成了 MySQL8.0

```
管理员: 命令提示符
initializing of server in progress as process 43080
2018-10-24T05:32:45.870611Z 0 [Warning] [MY-013242] [Server] --character-set-server: 'utf8' is currently an alias for the
character set UTF8MB3, but will be an alias for UTF8MB4 in a future release. Please consider using UTF8MB4 in order to
be unambiguous.
2018-10-24T05:33:14.879343Z 5 [Note] [MY-010454] [Server] A temporary password is generated for root@localhost: MMQP.YfC
y5XY
2018-10-24T05:33:37.812581Z 0 [System] [MY-013170] [Server] D:\MySQL\mysql-8.0.13-winx64\bin\mysqld.exe (mysqld 8.0.13)
initializing of server has completed

D:\MySQL\mysql-8.0.13-winx64\bin>mysqld --initialize --console
2018-10-24T06:06:51.100372Z 0 [System] [MY-013169] [Server] D:\MySQL\mysql-8.0.13-winx64\bin\mysqld.exe (mysqld 8.0.13)
initializing of server in progress as process 47804
2018-10-24T06:07:18.372699Z 5 [Note] [MY-010454] [Server] A temporary password is generated for root@localhost: nj>uUJkp
H4/I
2018-10-24T06:07:36.024435Z 0 [System] [MY-013170] [Server] D:\MySQL\mysql-8.0.13-winx64\bin\mysqld.exe (mysqld 8.0.13)
initializing of server has completed

D:\MySQL\mysql-8.0.13-winx64\bin>mysqld --install MySQL8.0
Service successfully installed.

D:\MySQL\mysql-8.0.13-winx64\bin>
```

数据分析不是个事儿

<https://blog.csdn.net/daocaoren92>

启动服务：

net start MySQL8.0

```
管理员: 命令提示符
initializing of server in progress as process 43080
2018-10-24T05:32:45.870611Z 0 [Warning] [MY-013242] [Server] --character-set-server: 'utf8' is currently an alias for the
character set UTF8MB3, but will be an alias for UTF8MB4 in a future release. Please consider using UTF8MB4 in order to
be unambiguous.
2018-10-24T05:33:14.879343Z 5 [Note] [MY-010454] [Server] A temporary password is generated for root@localhost: MMQP.YfC
y5XY
2018-10-24T05:33:37.812581Z 0 [System] [MY-013170] [Server] D:\MySQL\mysql-8.0.13-winx64\bin\mysqld.exe (mysqld 8.0.13)
initializing of server has completed

D:\MySQL\mysql-8.0.13-winx64\bin>mysqld --initialize --console
2018-10-24T06:06:51.100372Z 0 [System] [MY-013169] [Server] D:\MySQL\mysql-8.0.13-winx64\bin\mysqld.exe (mysqld 8.0.13)
initializing of server in progress as process 47804
2018-10-24T06:07:18.372699Z 5 [Note] [MY-010454] [Server] A temporary password is generated for root@localhost: nj>uUJkp
H4/I
2018-10-24T06:07:36.024435Z 0 [System] [MY-013170] [Server] D:\MySQL\mysql-8.0.13-winx64\bin\mysqld.exe (mysqld 8.0.13)
initializing of server has completed

D:\MySQL\mysql-8.0.13-winx64\bin>mysqld --install MySQL8.0
Service successfully installed.

D:\MySQL\mysql-8.0.13-winx64\bin>net start MySQL8.0
MySQL8.0 服务正在启动 ...
MySQL8.0 服务已经启动成功。

D:\MySQL\mysql-8.0.13-winx64\bin>
```

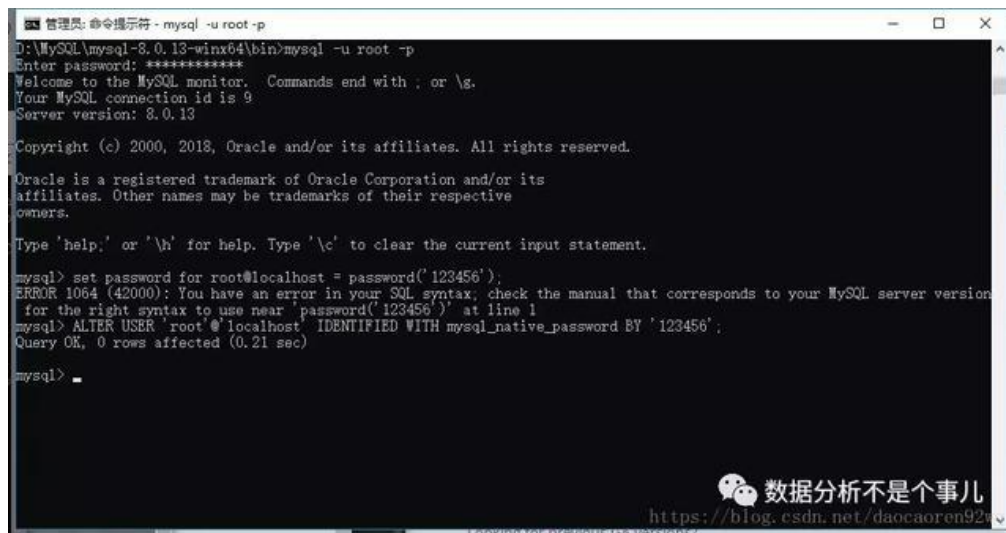
数据分析不是个事儿

<https://blog.csdn.net/daocaoren92>

修改初始密码

登录 mysql 之后执行：

ALTER USER 'root'@'localhost' IDENTIFIED WITH mysql_native_password BY '123456';



```
管理员: 命令提示符 - mysql -u root -p
D:\MySQL\mysql-8.0.13-win64\bin>mysql -u root -p
Enter password: *****
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 9
Server version: 8.0.13

Copyright (c) 2000, 2018, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> set password for root@localhost = password('123456');
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MySQL server version
for the right syntax to use near 'password('123456')' at line 1
mysql> ALTER USER 'root'@'localhost' IDENTIFIED WITH mysql_native_password BY '123456';
Query OK, 0 rows affected (0.21 sec)

mysql>
```

将 MySQL 和 Excel 或其它数据分析工具连接起来

默认安装的情况下，有一个插件允许 MySQL 数据库和 Excel 相连接，试想一下，我们在 MySQL 数据库中通过 SQL 语言进行大量数据的处理和计算，将计算结果存放在特定的数据表中，再通过 Excel 连接上 MySQL 数据库，将数据读取到 Excel 中，用 Excel 进行分析并绘制图表，免去了将数据从数据库导出再导入 Excel 的麻烦，是不是就可以提高效率了呢？

后面会分享一篇通过 ODBC 驱动程序连接一些报表\BI 工具来做分析的操作。

关于 MySQL 的学习推荐书籍《MySQL 必知必会》。