

数据分析师的SQL自学之路

曲箫扬

(长江职业学院 湖北武汉 430074)

【摘要】SQL是数据分析师的基础能力之一。本文就数据分析师的SQL自学方法,以个人经验为基础,以职业能力为导向,从书籍、视频、刷题、文档四个方面分别进行了分析与探讨,给初学者提供了一个SQL自学路线图。

【关键词】SQL;数据分析师;自学

DOI: 10.18686/jyfzyj.v3i4.40594

互联网本身具有数字化和互动性的特征,这种属性特征给数据搜集、整理、研究带来了革命性的突破。以往,在“原子世界”中,数据分析师要花较高的成本(资金、资源和时间)获取支撑研究、分析的数据,并且数据的丰富性、全面性、连续性和及时性都比互联网时代差很多。

与传统的数据分析师相比,互联网时代的数据分析师面临的问题,不是数据匮乏,而是数据过剩。因此,互联网时代的数据分析师必须学会借助技术手段进行高效的数据处理。而SQL就是作为一名数据分析师必须掌握的基本技术手段之一。作为对一名数据分析师在SQL方面的要求,从完成简单的数据的增删改查,到包括数据库本身的DDL、DCL、查询调优和锁机制等,再进一步外延,数据仓库作为进阶要求。现将一名数据分析师在SQL方面的自学历程从四个方面进行梳理总结。

个人SQL学习之路,推荐以最为广泛使用的数据库之一的MySQL作为对象展开学习,下面从书籍、视频、刷题和文档四个部分展开介绍,这四个部分本身并无先后之分,甚至会有经常性交叉和重复,利用多源资料相互促进、不断深化,方可温故知新。

1 书籍

在目前移动互联网蓬勃发展的背景下,传统意义上的教材在学习一门新知识过程中的所占比重正变得越来越弱,相较于文字的生涩和古板,很多初学者可能更习惯于看一套视频教程。但即便如此,有一些经典书籍仍然值得提倡和阅读的。在这里介绍两本MySQL经典书籍,《MySQL必知必会》与《高性能MySQL》,前者适合建立学习SQL之初的总体框架,后者适合具备一定SQL基础之后的提高。两本书都堪称是经典,其中前者是图灵必知必会系列丛书之一,与这本书内容很相近的还有该系列图书《SQL必知必会》。后者是REILLY动物书系列之一,从其封面图中的那只鹰就可以感受到本书对MySQL性能要求的犀利程度。作为教材中的经典,两本书的缺点也很是显而易见:那就是较为古老,时效性差。例如,MySQL

必知必会是出版于2009年,那时的MySQL版本大概是5.1,刚好是SUN公司收购MySQL之后不久。而高性能MySQL经典版本是2013年的第三版,这也是市面最为流行的版本,本人所使用的也是这一版,对应MySQL5.5,是目前比较新的版本。所以,在学习两本书的过程中也要注意甄别版本之间的区别。在这里说明一下,当前MySQL的最新版本是8.0。

两本书各自的定位分别是,《MySQL必知必会》一书适合通读,侧重帮助初学者迅速建立数据库知识框架。《高性能MySQL》适合专题学习,用于数据分析从业者进阶,不断精进技术。

2 视频

在学完《MySQL必知必会》,建立起基本知识框架之后,就可以找一门视频教程进行专项学习了。实际上,在学习一门新知识之初看视频可能是更为高效和直观的方式,但在入门之后再观看视频则不再合适,而应追求高质量的专题研讨甚至学术论文。

学习视频教程有众多渠道可以获取,以个人经验,本人首先推荐Bilibili。Bilibili曾经是一个以二次元为主的视频站点,如今已经成长成一个集合了众多优秀学习资源的开发学习平台。该站上面有大量的开源视频教程,有的是培训机构的录制视频,也有一些技术大佬录制的教程,种类繁多。本人推荐该站点上的一套完整视频教程《MySQL基础+高级全套视频教程》,视频非常齐全,共241集,累计37个多小时,个人感觉是目前对MySQL讲解较为通俗和全面的。初学者可能对很多知识点,当时不甚理解,但在后续复习中往往会有新的理解和认知,一套教程下来基本上涵盖了MySQL入门和进阶,非常值得反复研读。

3 刷题

纸上得来终觉浅,绝知此事要躬行。当看书和刷视频到一定程度之后,就应该用刷题来检验所学深浅了。刷题

对初学者来说,可能是最能短时间快速学习和掌握新知识的方式,也是应试或参加面试笔试的有效备考方法。个人在 MySQL 方面也找了很多刷题资源,经比较后,值得推荐的是以下 4 个:

第一个是互联网广泛流传的《MySQL 经典 50 题》。实际上,这套题目暂不知其最初起源,但这 50 题确实广为流传,实际操作拿来练手也会确有收获。由于 50 题没有平台收录,所以需要自己在本机搭建好 MySQL 环境,并将样例数据建表写入库中,方可使用。经典 50 题的定位整体是入门级,部分题目会有些拔高,网上有答案,易于寻找,真正完整刷一遍相信可以达到入门级水平。

第二个是 SQLzoo。SQLzoo 是一个支持多语言的 SQL 练习平台,定位也是偏入门级,会系统按照 MySQL 的各大关键语法进行分章练习。例如,比照 MySQL 必知必会中的章节划分,题目难度,由浅入深,循序渐进,有的题目还会给出基本的 SQL 语句样例,回答正确后会提示笑脸,回答错误时会给出简单的错误描述,例如返回记录过多等,必要时还可提前查看下正确答案,整体教学设计十分贴心。

第三个是牛客网。牛客网上相关题目共计 60 余道,主要是查询题目,也有少数题目考查其他语法。牛客网作为 1 个 SQL 刷题平台,其特点具有两面性。该平台上的很多题目具有代表性和深度,很能启发思维。但是另一方面,SQL 是以 SQLite 语法作为蓝本,跟 MySQL 还是有些许区别,而且牛客网上面的少数题目的表达不太准确,理解起来存在歧义。会出现看了解答之后方能准确理解的情况。个别题目还有字符错误问题。

最后推荐的一个是 Leet Code。之所以把 Leet Code 放在最后,是因为作为一个程序员刷题网站,Leet Code

在数据库题目方面比较丰富,按简单、中等、困难 3 个难度进行区分,题目在严谨性和代表性方面也较前 2 个平台有很大提升。但最大的问题是 Leet code 中免费可刷的数据库题目太少,在 100 多道题目中,仅有 20 题左右是免费开放的,所以光靠这 20 题很难带来 SQL 能力的实质提升。只有把题目全部做完之后,SQL 能力才会有很大提升,在面对多重嵌套、临时变量、窗口函数等知识点时不至于束手无策。

4 文档

在互联网这个知识日新月异的领域,必须经常性地学习跟踪技术前沿才不至于落伍。所以,在 MySQL 学习路上,最后值得推荐的其实是 MySQL 官网和官方文档。一部官网记录了 MySQL 的方方面面,一份官方文档则涵盖了 MySQL 的所有知识要点。当然,几千页的文档是不可能短期看完的,所以本人也只是针对其中的重点章节仔细研读,例如,个人在学习 MySQL 中的 4 种隔离级别和不同锁机制时,学习的主要资料就是阅读官方文档。以个人阅读经验,官方文档虽是英文文档,但其撰写规范严谨,理解起来毫无歧义。更重要的是,官方文档实际上才是最为权威的学习资料,而不像某些网络教程鱼龙混杂,出现掺杂了个人理解和三人成虎之后,很多知识点其实是会存在误解的。

5 总结

现在在 SQL 自学之路上的四种学习方法从 4 个方面进行对比,初学者可根据各自特点,依据自身情况,选择合适的方法进行学习,具体内容参见表 1。

表 1 四类自学方法的异同点

	书籍	视频	刷题	文档
获取方式	付费	免费	大部分免费	免费
权威性	次高	中等	略低	最高
友好度	较友好	友好	较友好	较不友好,对英文阅读水平要求较高
适用层次	入门至高级	入门至高级	入门,考试,面试	进阶,高级

以上内容是从个人自学 MySQL 出发总结的经验,仅能称得上是学习 SQL 之万里长征的第一步,学习的过程本身就是只有进行时而没有完成时。

作者简介:曲箫扬(1983.6—),女,湖北武汉人,硕士,副教授,研究方向:移动网络规划与优化,数据分析。

【参考文献】

- [1] Ben Forta.MySQL 必知必会.[M].北京:人民邮电出版社,2009.
- [2] 施瓦茨,扎伊采夫,特卡琴科.高性能 MySQL[M].北京:电子工业出版社,2013.
- [3] 董国钢 Bilibili 基于“互联网+”的 IT 类课程教学模式创新与教学改革——以《MySQL 数据库应用》为例[J].石家庄理工职业学院学术研究,2020,15(Z1):35-39+34.