

迈克尔·斯通布雷克访谈录

本专访介绍了迈克尔·斯通布雷克开发 INGRES 过程中的痛苦和辛酸,以及他在数据库理论研究中的成功与失败、对创业的价值和中间件等问题的看法,更多内容请看访谈录。

— 玛丽安 温斯特

最近, SIGMOD 学界的两位成员 Phil Bernstein 和 Hector Garcia-Molina 被推选为国家工程院院士,他们是继 David DeWitt 和 Jim Gray 之后再次获此殊荣的两个人。在此我代表 SIGMOD 学界,向 Phil 和 Hector 表示祝贺。感兴趣的读者可以阅读他们相应的访谈录。同时也要向 Avi Sliberschatz 表示祝贺,他获得了今年的 IEEE 泰勒教育奖,相应的访谈最近也将出版。

细心的读者可能已经注意到,上期 Jim Gray 的访谈录里缺少之前栏目中的编辑标记。在编辑的过程中我们将这些访谈的口语转整理成正式的书面文档,编辑过程中的文档引入了编辑标记,在转换为书面文档之后我们再将这些标记去掉。口语和书面语之间存在着一定的差异,口语更加直白易懂,而书面文档可读性更强一些。访谈的 Web 视频版本即将完成,不久后读者就可以下载并观看这些原始的访谈视频。有了这些原始视频,读者可参照两者来对照阅读,这样我们也就不用担心两者在内容上存在的差异了。因此,我们以后的访谈录也将采用这种方式进行编辑。

玛丽安: 欢迎来到 ACM SIGMOD RECORD 数据库界杰出人物系列访谈栏目,我是玛丽安·温斯特,我们现在在威斯康星州麦迪逊——2002 年 PODS 和 SIGMOD 的主会场。我身边的这位是 Michal Stonebraker,它不久将成为麻省理工大学的客座教授,此前他在伯克利大学任教近 30 年。他领导开发的 INGRES 系统证明了关系模型是一个切实可行的理论,并且可以在数据库中被实际应用的优美模型。随后他创办了 INGRES 公司,同时他还是 Illustra 信息技术公司的创始人之一,他将对象-关系数据库方面的研究成果商业化。迈克博士毕业于密歇根大学。接下来,欢迎迈克。

迈克: 谢谢!

玛丽安: 迈克,当年你还是伯克利大学的助教,着手去构建一个关系数据管理系统,做了一个非常大胆的尝试。当时关系模型在数据库领域的大多数人看来是一个不切实际的想法,并且还没有一个成型的关系数据库系统存在。那么当你决定去做 INGRES 时,你考虑了多久?

迈克: 博士期间我研究的是运筹学中的应用马尔科夫理论,来到伯克利后,我很快意识到如果继续我的博士课题将无法获得大学的终身教职。所以我必须要做一些全新的、无中生有的东西,这时我决定做数据库。1971 年 Codd 写了篇开创性的文章,立刻引发了传统数据库学者和关系数据库学者之间的一场激烈争论。传统数据库学者攻击我:“你不可能构建出来一个全新的数据库系统,即使你可以,也没有人能够理解你的查询语言。”关系数据库学者争论到:“只要不和传统数据库一样复杂的东西都可能是一件正确的事情。”一些研究者认为根据关系模型尝试构建出一个关系数据库系统是一个有趣而重要的工作,我就是其中的一个。

我们很快构建了一个最初的原型系统,在这个阶段我们集中了 90% 的精力去努力使这个原型尽可能成为一个真实的系统,并使它差不多可以运行起来。回顾开发 INGRES 的那段经历,我们曾经做过的最明智的事情就是把接下来 90% 的精力集中在努力使 INGRES 真

正的运行起来，我认为这是 INGRES 与其他的传统的学术项目最大的区别所在。INGRES 项目花费了我们团队整整 5 年的时间，这个团队包括一个全职的程序员和四五个聪明的研究生和本科生。构建 INGRES 在当时看起来仅仅是一个很好的想法，刚开始的时候，我对数据库系统是什么以及需要多长时间可以去构建这样一个系统没有一点概念，但是我必须取得终身教职，于是我赌了一把。

玛丽安：你认为 INGRES 项目中最大的成功和失败分别是什么？

迈克：我认为我们最大的成功是让 INGRES 系统运作起来，这在当时是一个非常有创意的工作。回想起来，我们当时做了很多非常幸运而又意外的决定。我们选择了 UNIX 和 C，当时这些都是一些疯狂的想法。我认为我们做出这些选择的仅有的一个原因是 Ken Thompson，他是伯克利的毕业生，来我们这做过一次访问，而且我们相信他很清楚他自己正在做些什么。我们成功的另外一个因素是我们遵循这样一句名言，“如果你觉得它错了，就扔掉它重写”。所以，在 INGRES 真正能运作之前，我们对 INGRES 中的很多的代码都进行了多次重写，并且乐此不疲。总之我认为我们最大的成功就是使 INGRES 正常地运作起来了。

我认为最大的失败是真实用户对系统的认可度问题。由于我们有了可以真正运行的关系系统，到 70 年代后期人们已经不说“你不可能使它运作起来并且人们也不能理解这种查询语言。”这样的话了，但是他们还在不停地攻击我们“它不能很好的扩展！谁会是你最大的用户？关系数据库并不合时宜！”当时，我认为具有说服力的案例是，1978 年亚利桑那州立大学打算将大约 40,000 名学生的信息放入 INGRES 数据库中。但是亚利桑那州立大学需要面对两个严峻的现实问题，首先他们必须从 AT&T 获得还未得到任何软件支持的 UNIX 操作系统，其次是要从伯克利的菜鸟那儿获取同样没有任何其它应用软件支持的数据库系统；当他们意识到在 UNIX 系统中不能使用 COBOL 时，这个项目自然就流产了。而且正是如此，INGRES 系统渗透到现实应用中的希望又一次破灭，因为他们的技术依赖于 COBOL，而当时还没有任何应用软件提供对 UNIX 和 INGRES 的支持。INGRES 作为一个学术原型系统，不可能很好的克服这些障碍，唯一可行的途径是商业化。

玛丽安：你觉得在 INGRES 项目中获得了哪些经验教训？

迈克：如果能预知这一切的话，可能我根本就不会尝试去构建 INGRES，它对于当时的我们来说是如此之难，大大超出了我的想象。在构建系统的过程中有大量的工作需要做，这些在项目开始之初我们根本无法考虑到。如果我比较明智，我根本就不会开始它，因为就如我所说的，它真的太难了。但是当时我已经做出了那个不明智的决定，所以我对年轻的自己的唯一建议就是：停止所有怀疑，放手去做吧！不积跬步，无以至千里；不积小流，无以成江海。你需要做的是一步一步坚定地迈出你的脚，不管前面的路有多艰难。

另外一些建议是招募一些好的本科生为你工作，尽量给他们提供最好的设备，即使代价非常高昂。给出这个建议的原因是 INGRES 很多的代码都是由伯克利的本科生写的。项目中我们招募一些比较聪明的大一和大二的学生，给他们提供最好的设备，让他们可以全心全意为我们写代码，他们做出了卓越的工作。所以我非常热衷于招一些天才的本科生，并为他们创造出一个良好的氛围。

玛丽安：你在伯克利还有做其他一些后续项目吗？跟在 MIT 有什么不同？

迈克：MIT 与 Berkely 是非常不同的，在 MIT 我有点儿像个传教士，这里没有做数据库的人，所以我正试着找出如何在 MIT 开展工作。

玛丽安 :作为一个有两次创业经验的前辈,你对现在打算创办公司的新人有什么好的建议?

迈克 :这个问题很好。如果你觉得自己有了一个好的想法,那么尝试着从无到有把它创造出来,我认为这是检验你的想法是否真的具有意义的一个好方法。另外,如果你真的能从无到有通过创办一个公司去实现你的想法,那是一个满足用户的真正需求,以及和一些拥有难题的聪明应用者讨论你的产品的好方法。这也是产生新问题和有趣想法的一个重要来源。在学术界或研究领域,我们相对来说比较孤立,创业是让研究建立在实际问题基础上的一个好方法。

玛丽安 :听起来似乎你也支持在工业界或者是刚刚起步的团队休假?

迈克 :我认为学术界最大的问题就是它促使我们每个人都在为得到终身教职而挤破头皮,大家的目标是尽可能把自己的简历填的更完美,而这就意味着你需要花费你的休假时间去写很多无用的论文,这是不能培养出全面的研究型人才的。如果研究人员解决的是一些实际的研究问题,我非常赞同他们进行学术休假,暂时离开学校去工业界或工业研究实验室去实现和检验自己的想法,尤其是自己创业。我认为在休假的时候去其他的大学进行交流访问是一件最没有意义的事情,因为这意味着你把时间花在跟你做同样事情的人身上。

玛丽安 :迄今为止,数据库理论产生过一些真正有价值的东西吗?现阶段数据库理论研究人员应该关注什么样的研究问题呢?

迈克 :我猜接下来你是想要讨论关于 Laguna Beach 的未来数据库研究报告,它在 1989 或者 1988 年出现的,距离现在有一段时间了。当时,递归查询是理论研究界关注的一个重要的研究问题。所以当时几乎所有的人都在研究递归查询,在 Laguna Beach 就有 10 到 12 个实习者从事这个方面的研究,但是他们中没有任何人注意到其实用户对这个问题没有任何兴趣。

玛丽安 :难道传递闭包查询算不上有价值的研究吗?

迈克 :一些用户需要的仅仅是线性递归查询,没有真正的用户在查询中需要用到它,而数据库理论人员却都在研究着这方面的内容。因此,我们向理论学家提出挑战:在你解决这个理论难题之前,你应该问一下自己是否有人真的需要它或对它感兴趣!

我觉得理论界最大的问题是他们把自己孤立起来,他们按照自己地方式一个课题接着一个课题做下去,并不关心他们所做的事情是否有用。我认为对递归的研究就是其中的一个最大的失败,因为现在我仍然找不到对非线性递归查询感兴趣的人。这些理论家中存在的最大的问题是他们感兴趣的是那些可解决的问题,而不是那些真正重要的问题。我对理论学家们的建议和我刚谈论的是相同的:花一些时间去了解现实世界中的真正需求,并尝试去解决这些问题。

玛丽安 :你认为理论学家花一些时间处理实际问题的时候,他们是继续从事理论研究还是从中找一些非理论的实际问题进行研究?在处理实际问题的时候,他们是转向为系统类型还是继续理论研究?他们能找到理论问题研究吗?

迈克 :我认为理论界最大的成功是并发控制和可串行化理论。他们在这个领域做出了卓越的

贡献。我认为 Avi Silberschatz 去贝尔实验室并转换为做系统类型的研究是一件很好的事情。事实上这样的案例并不乏其人，据说康奈尔大学的 John Hopcroft 也转向系统方面的研究，虽然他的研究领域不是数据库。所以说如果在这条路上能做出了重要贡献，那是非常值得去做的。我认为数据库理论研究总是脱离实际问题太远，这些研究并不符合我们的实际需要，而且他们也不能很好地进行角色转换，把研究的成果转化为系统。

玛丽安：面向对象数据库是否也曾脱离了实际需求呢？

迈克：我个人的观点是这个主题前一段时间已经被激烈地讨论过了，而且我也和一个人面向对象数据库公司的创始人讨论过这个问题，我的第一反应是：“这个产品的市场在哪里？”，这是一个高雅的技术，可惜的是没有市场。面向对象数据库在某种意义上来看好像是一种有趣但没有实用价值的技术，但是我认为现在没有人需要它的原因是它太超前了。用 C++ 构建一个持久性机制不是一个很大的市场，传统的大型数据库公司都没有在意它，它是一个利基市场。可能会有人花 5 年的时间去打开这个市场，也许他们不会这样做，事实上我也不能确定。但是我认为已经有了一些很有趣的对象数据库产品，而且这个技术最终会得到广泛的应用。

玛丽安：现在仍然有一些公司在做面向对象数据库。所以，就像你所说的，这里是存在一个利益市场的。

迈克：我觉得一个更加有趣的问题是，“为什么关系对象技术没能够迅速流行起来呢？”

玛丽安：那正是我的下一个问题！人们一直在说对象关系技术并没有达到供应商所预期的那种影响。那是什么意思？是我们缺失对一些关键技术的研究，或是至今还没有找到一个杀手级的应用，还是有更深层次的内在原因？

迈克：我认为对象-关系技术很明显是一个好想法，而且已经流行过了，并且获得了一定的成功。Jeffery Moore 的著作《跨越鸿沟》能很好支持我的观点，任何一项新技术的流行都需要一个过程，最早接受它的是那些技术发烧友，如果这些人喜欢，他们就会把这个技术推荐给周围的朋友，这样就会有更多的人去尝试它。当越来越多人使用它的时候，这个技术会迎来它的主流用户，并进入一个广阔的市场，随后它才被晚期用户所接受。这个过程需要一些时间，最困难的是如何跨过早期用户和主流用户之间的这个鸿沟。

我认为对象关系技术在其早期用户接受者中获得了广泛的成功。在 GIS 系统和媒体资源管理的应用中，它是非常成功的。像 CNN、BBC、Televisa 这样的拥有海量数据的公司都非常需要对象关系技术，他们都是对象-关系技术的早期使用者，但是它没有能够跨过这个技术鸿沟。我认为这主要有两个方面的原因：第一个原因是标准的缺失，不同的供应商提供的产品扩展能力是不相同的。在这个时代，所有人都想让他们的应用能跨平台运行，这是他们所不愿面对的。就像 J2EE 一样，除了微软之外的每个人似乎都把它作为扩展机制的标准。所以，除了缺失一个标准之外，我们还是取得了一些成果。第二个原因是微软和 Oracle 这两个主流的数据库生产商没有参与推动对象-关系技术的发展，仅仅依赖 Informix 和 IBM 两个公司的力量是很难做到的，我认为这是阻碍对象-关系技术得到广泛应用的最大的原因。

我认为当前人们对 XML 狂热又一次的证实了为什么可扩展的系统是一个好的想法。Informix 可以很好的支持 XML，通过添加各种用户自定义的操作符就可以支持各种非标准的 XQuery 和 XPath。例如，在 Informix 中，我们把 XML 中的“//”（用在层次搜索中）看

成一个由用户定义的操作符。所以说提供可扩展性的技术是非常有用的，因为不管接下来出现什么事情都能够适应。在 XML 之后保证还会有另外一个新的事物出现，所以我认为对象关系技术最终会取得实质性的胜利，这只是时间的问题，因为 XML 不是数据库系统将要处理的最后一个新数据类型。

玛丽安：所以你觉得我们仅仅需要耐心的等待，同时做好市场营销工作，对象-关系数据库总有一天会被广泛使用的？

迈克：我也是这么认为的，或者更简单的办法是直接给微软的头头们换换脑筋。

玛丽安：退休是你所期望的吗？对我们这些接近退休的人有什么建议吗？是否我们都应该搬到新英格兰吗？

迈克：哇哦，你还很年轻。

玛丽安：实际上我已经退休了，我退的较早。

迈克：首先，“退休”是对我当前状态的一种误称。我猜 7 年前伯克利开始有预谋地贿赂所有可能符合退休计划的人们。他们提供的条件非常有诱惑力，可以满足你想要的一切。我答应了他们的条件，退了下来，成为了一个退休人员——一个领取退休金的人。我从来没有感觉到真的退休了，现在我跟以前一样忙碌，一半时间在纽约忙于一个数据库公司的创办，同时我还成为了 MIT 的老师。当时我妻子比较沮丧，她以为我离开了伯克利以后，就会像那些退休下来的人们那样，每天坐在花园里晒太阳或者去打打高尔夫。这种退休状态是我不敢想象的，这会让我觉得厌倦。

所以接下来的问题就来了，为什么 1999 年我会从加利福尼亚搬到了新罕布什尔州。主要有两个原因：第一个是我的妻子来自新英格兰的一个大家族，近几年她一直在游说我搬到一个离家族近一点的地方，现在我退休了，也是时候陪她过她想要的生活了，所以这次搬迁并不是基于职业的考虑。另外一个主要因素是加利福尼亚的生活质量在不断下降，到处都特别拥挤，在交通高峰期高速路上的等待让我觉得非常疲倦。除了职业的原因，新罕布什尔州的生活质量非常好。我非常喜欢这种生活状态，而且在这里也没有太多的赋税压力。如果你离开城市去乡村生活，我认为新英格兰地区是一个非常合适的选择，这里生活非常棒。当然，如果你想留在城市生活，加利福尼亚是一个非常棒的地方。

玛丽安：很少有人像你一样曾经在两个世界顶级的大学任职，你可以评价一下伯克利和 MIT 之间的区别在哪里吗？

迈克：首先，我想说明的是，在 MIT 我是一个传教士的角色。自从 1981 年 Mike Hammer 离开之后，MIT 就没有做数据库的人了，所以我在那里是唯一做数据库的人。而在伯克利有很多人从事数据库研究，在这两个学校我做的工作是非常不同的。

我认为伯克利是一个学术氛围非常浓厚的大学，在伯克利大家喜欢在一起工作，这些人花大量的时间在他们的办公室相互交流。这里有很多合作项目，大家都努力为一些共同的利益而工作。MIT 的大学氛围与伯克利完全不一样，那里的老师更多的时候是在单兵作战。拿办公室这件事来举个例子，在伯克利，所有的办公室都以系为单位来进行管理，而在 MIT 系里只负责处理学术上的事务，如果你想要间办公室，你必须从每个独立的实验室比如计算

机科学实验室、人工智能实验室、多媒体实验室等实验室那里申请。以至于所有人一下子就能划清界线，因为从哪个实验室申请办公室已经确定了你将忠于那个阵营。

我认为这两所学校的学科侧重点非常不同。MIT 在人工智能和语音处理方面很强，而伯克利在传统系统方面比较强。这两所学校都有非常出色的学生和教员，所以我认为最大的不同点在于学校的特质方面的不同。比如，在伯克利，在学校的年度活动中，每个人都会为聘请新的教师出谋划策，而在 MIT 教师的委任完全由一个小的委员会来决定。

玛丽安：如果有充足额外的时间去做一件你想做但是还没做的一件事情，那会是什么呢？

迈克：我最讨厌的一件事情就是我今天早上在 SIGMOD 2002 的邀请发言中谈到的：有太多的中间件了。大多数公司都买了门户网站系统、企业应用集成系统、ETL（抽取、转换和加载）系统和一个应用服务器，可能还有一个联邦数据库系统。所有的这些系统都运行在中间层上，它们在功能上高度重叠、复杂并且需要系统管理员。大多数企业都至少有一个这样的系统，这样的中间件环境错综复杂，维护和使用这些系统的代价非常昂贵。

玛丽安：所有人都意识到了这个问题的严重性，传统的解决方案（商业智慧）是通过增强应用层服务器的功能，使得它能完成这些中间件的综合功能。IBM 的 Web Sphere 就是这样的一个例子，一个功能丰富的包，可以实现许多中间件的功能。

迈克：我认为联邦数据库系统比应用服务器更适用于用来构建中间件平台。因为应用服务器仅仅管理代码，而数据移交给底层来处理。如果一个应用程序需要一些数据，则它运行在中间层并且需要从底层请求数据，最终将数据移动到代码。如果你有一个联邦数据库系统，使得数据管理器运行在中间层和底层，而对象关系引擎可以很好的存储和激活功能，这样代码和数据就可以共存于一个平台上了。然后，可以通过这种方式进行物理数据库设计，把数据放在需要它们的代码附近，因为你不可能一直把数据移动到代码中。我认为这是一种更加长期的、更健壮构建一个成熟的中间件的方式。我一直都在尝试证明这是一个好的想法，如果给我更多一点的时间的话，我一定去做好这件事情。

玛丽安：过去你在学术和工业方面的工作中，你认为你最大的成功和失败分别是什么？

迈克：最大的成功毫无疑问是领导 INGRES 的开发，它在学术界和工业界都确实产生了巨大的影响。至于 POSTGRES 是否会产生同样的影响，正如我们前面所说的，还需要“跨越鸿沟”，谁也不知道它是否会发生。我认为最大的失败是 Cohera，它是一个非常成熟的联邦数据库系统，只是因为当时市场对这个技术的需求非常小，我们花费了大量精力去推动它，但是仍然很难找到相应的应用需求，最终彻底失败了。

玛丽安：所以，以前你本以为联邦数据库将在今天的市场上占据重要位置，但实质上它并没有如你预期的那样。你在今天的讲话中举了 Grainger 拥有 60,000 供应商这个例子，是否想说联邦数据库并没有很大的市场？

迈克：你可以这么认为。我认为联邦数据库技术将来成为商业的重点这是毋庸置疑的，这同样只是时间的问题。但是在商业市场中时机就是一切，我认为 Cohera 失败的原因是在时间上有点超前。对于一个成功的创业来说，一个好的创意和对的时机是必不可少的两个因素，Cohera 只是生不逢时而已。

玛丽安：作为一个计算机科学专家，如果你可以改变一件你自己的事情，那是什么呢？

迈克：我认为有两件我要做的事情。第一，我要把编码学好。我从来都没有时间成为一个真正的程序员，我们在前面谈到为什么人们需要利用学术休假去工业界做一些事情，对于我来说，如果我有休假，我会去学习如何成为一个产品程序员。

玛丽安：但是你并没有在学术休假的时候去做这两件事情？是因为创业太忙了吗？

迈克：是的，学习编程是需要花费大量的时间的，它总是不停地被往后推，因为还有很多更重要的事情需要去做。所以我一直都没有时间去学好编程。

我认为如果期望成为一名研究者，我们应该做的一件事情就是首先成一名合格的程序员，但是我不是。我认为问题在于我并没有接受大量的编程训练就完成了本科和研究生的教育。所以在学校的时候我从来没有获得足够的编程经验，毕业以后再也没有时间去学习。

另外一件想要改变的事情是在毕业之前缺乏对计算机科学全面的了解，当时我对编译技术了解很少。我发现现在的很多学生知识面非常广泛，他们对许多技术都有很深的了解，而这些东西我却知之甚少。所以，如果再重给我一次机会，我想选择接受现在学生们所接受的教育，对计算机科学有一个更全面的了解，而这在当时是不可能实现的。

玛丽安：非常感谢你今天接受我的采访。

（张金增译，霍峥审校）