

第13篇

杰弗里·诺顿访谈录

Interview with Jeffrey Naughton





杰弗里·诺顿简介

杰弗里·诺顿 (Jeffrey Naughton)，博士，威斯康星大学麦迪逊分校计算机科学教授。1982 年于威斯康星大学麦迪逊分校获得学士学位，1987 年于斯坦福大学获得博士学位。1987 年至 1989 年期间，他在普林斯顿大学任教，之后去了威斯康星大学。杰弗里主要专注于提高数据库管理系统性能和功能方面的研究，先后在 VLDB、SIGMOD 等发表高水平论文百余篇。他于 1991 年获得美国国家科学基金会青年研究者奖，由于其在研究上的杰出成绩，2000 年获得威斯康星“维纳斯合作研究奖”[⊖]，并且于 2002 年提名为美国计算机学会会士 (ACM Fellow)，其参与的 Gamma 项目获得 2008 年 ACM 软件系统奖。他目前担任 Greenplum、Teradata/NCR 等多家知名公司和微软公司吉姆·格雷系统实验室的顾问。

⊖ “国家科学基金会青年研究者奖” (National Science Foundation’s Presidential Young Investigator Award)，这个奖是很早的一个奖 (1984-1991)。是美国 NSF 专门针对大学里年轻助理教授的一种研究资助形式。科研经费在 \$100 000 到 \$125 000 (每年 \$25 000, 4-5 年)。有一些荣誉的性质，美国顶级学校拿这个奖的不少。这个项目 1991 年停了以后，目前对应的奖是 Faculty Early-Career Development (CAREER) award，规模有所扩大。

本专访主要介绍了为什么数据库系统是“控制狂”，如何挑选学生，如何选择问题，如何吸引别人的关注，真实面对自己的重要性等问题。



玛丽安·温丝特

玛丽安·温丝特：欢迎来到本期 ACM SIGMOD Record 数据库领域杰出人物访谈。我是玛丽安·温丝特，今天我们在伊利诺伊大学厄巴纳·尚佩恩分校计算机科学系邀请了威斯康星大学麦迪逊分校的杰弗里·诺顿教授。杰弗里主要专注于提高数据库管理系统性能和功能方面的研究。他是斯坦福大学的博士，曾经获得美国国家自然科学基金会总统杰出青年奖，并且是美国计算机学会会士。欢迎杰弗里。

杰弗里·诺顿：谢谢！

玛丽安·温丝特：杰弗里，据我所知您曾公开说过：数据库人是控制狂，不善与他人相处。

杰弗里·诺顿：我想你指的是我曾在许多地方做过的一个报告标题——“数据库系统是控制狂”。我的意思是数据库系统是控制狂，而不是说系统的开发者是控制狂。尽管您可能会说：“系统会反映出它们的开发者的性格。”

这个标题只是一个吸引人的噱头，还是对一些事情的深入洞察？时间将会说明一切。我曾经和别人讨论过这样一个问题：“当你已经准备好了所有数据并试图管理它们时，为什么不选择使用数据库系统呢？”得到的答案常常

是：“使用数据库系统总是令人痛苦和不快。如果要使用数据库系统，我们必须遵守一个关系模式。我认为并不是所有的数据都可以使用关系模式来表示，但是我们必须事先定义一个模式来描述需要查询的数据。在系统管理数据并运行应用程序之前，我们必须将数据加载到数据库系统中。接下来获取数据的唯一方式是通过 SQL 查询接口，但我并不想写 SQL 查询。”这个报告题目就是源自于此。这就是我所说的“数据库系统不善与他人相处”的含义。如果你想使用一个数据库系统，你必须了解该系统并且接受它的一切要求。这就像一个孩子说：“我想与幼儿园中的其他孩子一起玩，但是我只愿意在我想玩的时候、以我想要玩的方式、玩我想玩的游戏。”

玛丽安·温丝特：接下来，我想问几个关于科学计算领域的问题。该领域的系统运行在很多不同平台上，这些平台多数都是并行的，并且这些不同平台上要存放相同的数据。这意味着数据库管理系统也必须运行在并行计算机上，每隔几年就会有一些新的并行架构出现。数据库管理系统原本应该是廉价的，但并行数据库管理系统无论是购买还是维护都不便宜。因此，应该如何解决这些问题？

杰弗里·诺顿：我希望数据库研究人员可以全身心的研究这个问题，并且设法开发一些新技术解决这些问题。我认为没有简单、容易的方法可以解决该问题。我想一种可行的方法是：数据库研究者可以尝试采用你在之前讨论中所提到的方法。我们研究一下大家所面临的这些问题，并且尝试是否有方法可以既兼顾数据库系统的优点，同时又克服数据库对数据要求过分严格的缺点。

玛丽安·温丝特：这个建议对年轻人来说有些风险，不是对您，而是对年轻人。

杰弗里·诺顿：难道我不年轻吗？

玛丽安·温丝特：我指的是这个研究领域中的新人。目前数据库领域还不能真正接受那些在一个完全不同方向上进行的创新性研究，所以年轻人如果去做一些全新的研究会存在风险。而且对这类论文的一个经典评语是：“这不是数据库领域的研究工作。”

杰弗里·诺顿：对。

玛丽安·温丝特：但是您仍然要给年轻人这样的建议吗？他们怎样才能解决这个问题？或者这些建议只是针对那些“一无所有”的人而言？

杰弗里·诺顿：这是个非常好的问题。我的意思是，我很乐意看到年轻人轻松地追求一些新的东西。当然这也是一个非常难的问题。事实上，我已经试着考虑为什么会有这样的问题。我发现：你要做的事情越接近于工业界研究的问题，就越可能在审稿意见中受到抨击。因为所有这些问题已经被工业界研究，并且他们还知道一些年轻学者并不知道的问题。这意味着，这些论文很难被审稿人录用为数据库基础核心技术级别的论文，审稿人不需要具体说出系统的名字，就会说：“在我们系统中，这个技术永远不可行，因为我们的客户总是提出这样或那样的要求。”因此，你的研究不能太接近于数据库基础核心技术。当然你说的是对的，目前这类研究很难在数据库领域中发表。如果你走得太远，人们会看着你的文章说：“很好，这可能很有趣，但它不是数据库研究。”你必须努力找到一个折中的方法，我想我真的不知道现在数据库的研究范围是什么。这是一个非常好的问题。

玛丽安·温丝特：这个问题与我们今天要讨论的几个问题很相关，其中一个问题就是：如何在 SIGMOD 上发表论文。您在 SIGMOD 上的论文数量很快就要赶上 Mike Carey 的论文纪录。到目前为止他的文章是 28 篇，而你的是 24 篇。如何让论文被 SIGMOD 接收，您可以从这方面给大家一些建议吗？

杰弗里·诺顿：我在 SIGMOD 上发表的论文数已经接近于 Mike Carey 的论文纪录，其中的秘诀是让 Mike Carey 慢下来。我可以告诉大家如何做到

这一点。我觉得我真的很幸运，在威斯康星我有一个非常出色的合作团队和很好的工作环境。我不知道是否做了什么特别的事情。总之，我是在适当的时间、正确的地点和一群出色的人一起工作。

玛丽安·温丝特：在选择学生方面您有什么建议吗？

杰弗里·诺顿：这个问题很好。在选择学生方面我没有任何秘诀。这很难用言语来说清楚。如果你觉得可以和某个学生一起工作，那就去做吧。你不可能列出什么抽象的判断标准，例如：GRE 得分高于多少分，GPA 高于多少分，或者其它类似的标准。如果你可以发现一些擅长讲故事的人，这是一个很好的挑选方法。我并不是指那些擅长讲笑话的人，虽然那很有趣，我指的是那些可以讲清楚其研究工作及研究动机的人。

玛丽安·温丝特：这意味着他们是那些能看到大局的年轻人吗？

杰弗里·诺顿：也许吧，或者是那些可以解释从事当前研究工作原因的年轻人。如果可以认识到自己热衷于写一篇论文的原因，这也是一种能力。你可能会说：“我非常唯利是图，为了获得博士学位和好的工作，我不得不写这篇论文，后来为了保持这一切我必须写更多的论文。”但在某个时刻没有了这些目的时，你就会问：“我为什么要写这些论文？”我认为一个合理的理由是论文中的某些东西是有趣的，它吸引了你的注意力。一个成功的学生必须能够认识并且向他人解释对其研究课题感兴趣的原因，而不是仅仅将研究看成某个课程的作业。

玛丽安·温丝特：数据库研究领域一个尖锐的批评是：有些研究者就像在电线杆下寻找钥匙的醉汉，因为灯光可以照到那里。他们之所以选择一个问题，是因为他们已经有了解决该方法的问题，或者他们认为知道如何解决。所以当选择学生的时候，您是选择那些具有明确研究动机的学生，还是那些只选择可以解决的问题去研究的学生？

杰弗里·诺顿：我认为两方面因素都有一些。但更重要的问题是如何判断什么是有价值的工作、什么工作值得去做。是否有方法可以解决某个问题并不是判断该问题是否有意义的标准。所以我有点纠结于其中……

玛丽安·温丝特：我们别纠结这个问题了。你认为数据库系统研究应该往哪个方向发展？

杰弗里·诺顿：对此我没有明确的答案。我已经听到了一些对数据库界的批评，一些批评来自数据库界本身，对此我有一些不同的看法。另外，所有这些批评都是好的，我们正在进行自我批评，这样很好。

有些批评对我们很不公平，例如有人这样指责：没有人考虑数据库的前景问题。“所有人都在做一些改进性的工作，而不是致力于研究重大的具有突破性的问题”。有一段时间，在所有报告中几乎每个讲者都会使用“巨变”这个词：“我们需要研究一些具有重大突破性的问题，那些将真正改变我们数据管理方式的问题，那些有重大前景的问题。”这是无可辩驳的事实。我的意思是，我们不可能说：“不！我们不应该研究那些具有重大突破性、并将改变数据管理方式的问题”。但有趣的是，为什么我们没有去研究它们呢？有时候，我总觉得大家认为数据库领域的人都在说：“瞧那边有个重大突破性的问题，那真是一个非常好的问题，但是我并不想去研究它。我只是想做一些小的、改进性的研究工作。”事实并非如此，我们都渴望找到一个突破性的问题来研究。我认为数据库领域在许多方面发展得都很快。我们回顾一下你我职业生涯中数据库的发展过程，从逻辑数据库、对象数据库、对象-关系数据库、XML 数据库、流数据库，并且仍在一直在发展下去。我认为这是一个研究领域健康的标志。

但是数据库研究领域现在应该做些什么？要回答这个问题，你必须问：数据库研究者的角色是什么？很多时候，我们纠结于是否应该接受那些对数据库研究者的定义，因为这些定义并不真的符合我们。我认为有一些定义不正确，例如：“数据库研究领域是商业数据库公司开发部门的一个扩展。”如

果你同意这个定义，我们的研究就有了一个非常明确的度量标准：它是否会对产品产生影响？尽管你可能说：“这些商业数据库系统的实现都受到我们的研究和思想的影响。”我并不想诋毁这种度量标准，因为这些影响确实起了很大的作用，但并不是每个研究人员都赞同上面的定义。

我并不认为数据库领域是研发部门的一个扩展。有些时候我们很多人都认可这个定义，但有时候我们又不赞同它。作为一个大学的教授，当我老的时候（你已经说过我老了），应该说我更老的时候……我想我们俩大概同时毕业的吧！

玛丽安·温丝特：嗯，当然。我不采访比我年轻的人。

杰弗里·诺顿：好的。

玛丽安·温丝特：您恰好符合我采访的年龄标准。

杰弗里·诺顿：我认为，到老的时候大多数学者的影响、成绩或贡献是学生。其中一些人可能是你所教课程中的一些学生。尽管你一直希望激发他们的兴趣来思考数据库的问题并从事这个领域的工作，但你并不知道你已经对他们产生了影响。在你的履历上，这将是一个非常重要的成绩。

当你问我数据库研究领域应该做什么的时候，我很难回答。因为我认为做贡献的途径有很多种。你可以研究和设计一个系统来改变数据管理的方式，这是一个很好的途径。你也可以教一门本科生课程来培养人才，让学生对数据管理产生浓厚的兴趣，并将其作为其一生所要研究的事业。数据库界正在做上述所有事情，因此我认为数据库领域是成功的。

玛丽安·温丝特：你已经研究了关系数据库、面向对象的数据库、对象—关系数据库以及现在的 XML 数据库。我们认为你是一个数据库模型的机会主义者，或许更是一个数据模型的不可知论者。你下一步研究什么？XML—关系数据库？

杰弗里·诺顿：我不知道下一步要研究什么，可能我确实有点机会主义。我想作为一个数据库研究者我们所要做的事情之一就是探索那些我们不确定是否会成功的事情。我们不应该探索那些我们确信不能成功的事情——这是我们要避免的。当一个新的数据模型思想出现时，决定是否为其投入研究经费对一个大公司来说是非常困难的事情，因为他们已经在旧的数据模型上投入了很多。所以，当一种新的数据模型值得探索时，我们应该考虑它能走多远。因此，事后说“这个数据模型没有前途，所以我们当初不应该去研究它”是不公平的。我认为失败是成功之母。

玛丽安·温丝特：XML 是一个失败吗？

杰弗里·诺顿：XML 非常有趣。我想，那些 XML 数据模型的批评者认为我是一个愚蠢的年轻人。他们似乎在说：“你是个愚蠢的年轻人，你还没有从你的前辈身上吸取教训。你正在研究的这个树形结构数据模型，我们在 20 世纪 60 年代末 70 年代初已经研究过了，并且结果很糟糕。如果不相信的话，你可以继续试一试，但你一定会憎恨它。”我们并不是说：“对如何存储数据我有一个新的想法，我们应该将数据存储为 XML。我们要将任何东西都以 XML 的形式存储。沃尔玛就将他们的数据仓库以 XML 形式存储。”XML 会被人们接受。大量的 XML 正在被创建、生成和存储，并且人们想要对其管理、查询和处理。我之所以被 XML 吸引并不是因为它是数据处理的一种了不起的新方法，而是因为有一种新的数据类型需要我们处理。所以我并不认为 XML 将是下一代数据模型。如果我必须预言下一代数据模型的话，我会说它是偏关系型的数据模型，或者也可能是支持更多不同对象类型的对象-关系模型，更好地与文本相结合将是一个重大变化。如果必须选择一个将改变未来十年的数据模型，我认为是对象-关系模型。

玛丽安·温丝特：“XML 将会被人们接受，我们在需要时用它即可。”你对 XML 的这一评论让我联想到你之前关于“我们并不只是在做一些改进性

的工作”的评论。您说我们应该走出来看看需要做些什么，因为需要而去做，而不是因为它符合我们之前做事的方式。

杰弗里·诺顿：是的。

玛丽安·温丝特：在你的职业生涯中，你已经从一个纯粹的理论研究者转变为核心系统的研究者。什么导致了这样的转变？

杰弗里·诺顿：我不知道我是否已经成为核心系统的研究者。我认为现在这两个阵营都不会接受我。我想无论是理论者还是核心系统的研究者都不认为我是他们中的一员，我介于两者之间。什么促使我这样做呢？在研究生期间，这两方面的工作我都有所涉猎。（当我告诉大家的时候，很少人能相信。）在斯坦福大学，我从事非常理论的数据库研究。但晚上和周末我将回到 DEC 公司的西部研究实验室，在那里他们有一个新的被称为 Titan 的 ECL 处理器，我们正在研究将 Unix 移植到该处理器上。我总是对核心系统的构建非常好奇，同时也很喜欢做理论研究。可能在斯坦福受到了周围人的影响我的论文起初都很理论化。我的导师杰弗里·乌尔曼（Jeff Ullman）为探索数据库相关的理论问题提供了一个很好的环境。后来当我最终来到威斯康星大学的时候，有一个很好的核心数据库系统研究团队。威斯康星大学的每个助理教授将作为程序员被指派给一位教授。所以，戴维·德维特（David DeWitt）被分配给我做程序员。他构建了一个巨大的并行数据库系统 Gamma。我们在早上喝咖啡的时候讨论一些已有问题的新算法、新方法和一些新问题。我常常会想到一些疯狂的算法，他会否定其中大多数的算法，然后开始实现其中一些可行的算法，我们开始以这种方式合作。这种方式真是棒极了，真的非常好。

玛丽安·温丝特：一些人建议我问：“如今 PODS 上发表的文章是否都值得阅读？”而我更想问的是：“如今数据库理论者应该研究些什么？”

杰弗里·诺顿：要告诉一个理论研究者应该做什么是一个非常棘手的问题。

题。如果实话实说，我认为理论研究者应该研究任何他们感兴趣的东西，因为人类一直热衷于寻找重要而有趣的东西。事实上这看上去也是正确的。我们回想一下，一些非常有趣的工作仅仅是因为研究者认为它们有趣而被研究。一个常用的例子是数值理论。建立数值理论的人曾经说过：“我研究它是因为它不可能被应用。”然而，现在我们发现加密技术使用了大量的数值理论。另一个例子是那些研究数理逻辑的人。他们肯定没有想到数理逻辑可以作为一个基础框架来帮助管理计算机中大量的数据。他们研究的理由仅仅是因为这个问题很有趣。所以我不愿意看到理论研究者去研究那些他们不感兴趣的问题，尽管这些问题可能有一些应用。如果你要成为一个理论研究者，要忠于自己，找到一些由实际应用引发的好问题，研究的动机是你在其中发现了一些有趣的东西。

玛丽安·温丝特：您最有趣的顾问经历是什么？

杰弗里·诺顿：我想这是关于顾问经历的一个故事，它开始于 DEC 公司给我们提供资金的时候。戴维·德维特 (David DeWitt)、迈克·凯里 (Mike Carey) 和我开始做面向对象数据库管理系统的一个综合测试 (OO7 Benchmark)。早前，里克·卡特尔 (Rick Cattell) 已经做了对象操作 1 的测试 (OO1)。迈克·凯里建议我们跳过 2 ~ 6 直接做 OO7 测试。对我们来说这是一段非常奇妙的经历，工业界和一些知名行业媒体都对我们所做的事情产生了兴趣，这还是第一次。我们做这个测试时，有三到四家面向对象数据库公司与我们合作，我们在他们的数据库系统上开始进行测试。不知这件事是怎么被泄露出去的，我们开始接到各行业媒体的电话。编辑们会打电话问：“谁赢了？”我试图解释：“我们并不对它们进行排名，事实上我们这里没有获胜者。我们只是测试这些系统的不同方面，大家应该关注并且解决这些问题。”但是他们仍然不停地逼我。记得有一次，我被邀请到斯坦福大学关于 OO7 数据库研讨会上做一个报告。我刚好在报告前到达房间，我从来没有过在如此满的一个房间里做报告的经历。这个房间不是很大，但当我做报告的

时候房间非常满以至于我不能挪动一步，所以我不得不站在同一个地方做报告。而且后面人们都带着摄影机录下这个报告。那真是一次奇特的经历。我建议那些刚刚起步的年轻学者：如果你们希望吸引很多注意力，那么你们要做的就是找一种测试，并且测试商业系统，然后发布你们的结果。虽然这些公司可能要起诉你，但是你将引起很多人的注意。

玛丽安·温丝特：在普林斯顿大学您谋到了第一个教师职位，但是您很快就离开那里去了威斯康星大学。什么原因促使您做了这样的调动？

杰弗里·诺顿：这完全是因为个人的原因。我在斯坦福大学毕业以后去了普林斯顿大学，我真的很喜欢那里。那是一个很好的地方。那里有和我一起工作的几个朋友：首先是李凯（Kai Li），他不是一个真正从事数据库研究的人；迪克·立顿（Dick Lipton），他是一个非常重要的人；赫克托·加西亚·莫利纳（Hector Garcia-Molina），他是一个非常出色的顾问。那里一切都非常好。之后发生的事情是，在普林斯顿大学有一门编程导论课程，有这样的一个规律：如果你教了这门课程，你的家庭将会添丁或者你将成为肯·施泰格茨（Ken Steiglitz）这样的学者。这是一个定理，因为每个人在教授这门课程期间都有了孩子。我犯了一个错误，我连续两个学期教授这门课程，所以我有了一对双胞胎。我和我的妻子都来自麦迪逊，所以我们回去过了一个寒假。德威特是一个非常精明的人，在这两周的寒假期间他给了我一个办公室。白天我可以潜心工作，晚上回家，孩子们看上去非常高兴，我们也很高兴。对我们整个家庭来说这似乎是一个非常好的地方。无论对于我个人还是我的专业来说都是一个非常出色的地方。两周后德威特说：“为什么你不调到这边来呢？”我回到家并且说：“为什么我们不调到这里来呢？”事情就是这样，所以这真的与专业考虑无关。

玛丽安·温丝特：我相信很多有孩子的年轻教师也面临着类似的压力。

杰弗里·诺顿：是的，太难了。孩子在我们生活艰难的时候到来。

玛丽安·温丝特：你可以给没有经验或处于职业生涯中期的研究者一些其它建议吗？

杰弗里·诺顿：你必须要知道自己最适合做什么类型的研究者，并且照着它去做。尽管这个建议可能太抽象不实用，但是这一点非常重要。不要盯着某个大牛说：“我希望像他一样，有个可以开发很多系统的团队，这就是我要做的。”如果这真是你所想要的，那就照着去做吧。但不要试图模仿那些与你研究特点不相符的人，这会让你迷失自我。

玛丽安·温丝特：如果你有足够时间去做一件从未做过的事情，你会做什么？

杰弗里·诺顿：我知道还有很多东西我不太了解。在一些会议上或者像 UIUC 这样的地方，我经常会遇到一些追求感兴趣东西的年轻人。从他们那里我了解了一些以前不知道的东西，我希望能花更多的时间去了解他们所做的的事情。

玛丽安·温丝特：作为一个计算机科学家，如果您可以改变关于自己的一件事情，那会是什么？

杰弗里·诺顿：到目前为止，对我来说一切都如我想象的发展，所以我特别高兴有机会和这些出色的伙伴一起工作，能够有这样的经历。因此，如果说我想要改变一些什么，这似乎看上去有点忘恩负义。我认为我很满意现有的一切。

玛丽安·温丝特：好的。非常好。谢谢你的到来。

杰弗里·诺顿：谢谢。

（马如霞 译，孟小峰 审校）