

戴维·迈尔访谈录





戴维·麦尔简介

戴维·麦尔 (David Maier), 博士, 波特兰州立大学计算机科学系教授。1978 年于普林斯顿大学电子工程与计算机科学系取得博士学位。戴维·麦尔博士主要专注于数据流、叠加信息管理、科学信息管理与声明式编程等方面的研究, 先后在 JCAM、TODS、VLDB、SIGMOD 等发表高水平论文 200 余篇并三获美国国家自然科学基金会颁发的青年研究者奖, 多次获得俄勒冈研究所杰出教育奖和杰出教师奖。他于 1988 ~ 1989 年获得总统奖, 于 1997 年获得 SIGMOD 创新奖。同时, 他还曾担任 TODS 和 IJDL 的副主编, 他于 1998 年成为 ACM 会士。

本专访介绍了对象数据库产生的影响、为什么对象关系数据库的特征至今较少被人们使用、数据库理论研究者应该做什么、如何在一个小研究机构获得成功等。



玛丽安·温丝特

玛丽安·温丝特：欢迎来到 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 2002 年 PODS 和 SIGMOD 会议的举办地威斯康星州的麦迪逊城。今天和我在一起的是戴维·迈尔（David Maier），目前是俄勒冈健康和科学大学科学与工程学院计算机科学与工程系教授，ACM 会士以及 DARPA 基金机构的顾问。他由于在数据库理论和面向对象数据库方面的杰出工作而闻名，这包括他作为 Gemstone、ObjectStore 和 O2 对象数据库产品的顾问时做出的工作。现在，他依然活跃在科学数据库和科学数据管理领域。戴维于 70 年代在普林斯顿大学拿到博士学位，在加入俄勒冈研究所前曾于 80 年代在纽约州立大学石溪分校任教授。下面欢迎戴维。

戴维·迈尔：谢谢

玛丽安·温丝特：戴维，10 或 15 年以前，面向对象数据库被认为是下一个浪潮，现在它们还存在吗，为什么它们失败了呢？

戴维·迈尔：这是一个有趣的问题，而且对我来说很重要，因为我的很多工作都和对象数据库有关。我认为一部分对象数据库公司针对小众市场，

比如计算机辅助设计这个最多只有 5 千万到 1 亿美元的市场。与此相比，商用数据应用的市场在 10 亿美元以上。另一方面，依然有很多重要的应用在使用面向对象数据库，比如每月交易超过数十亿美元的 Gemstone 金融应用。CERN 使用它们处理粒子物理数据。至于对象数据库衍生出的技术，一部分被用于对象关系数据库，大部分被用于 XML 数据库。

玛丽安·温丝特：因此，对象数据库也许会消失，但产品依然存在且仍在销售。

戴维·迈尔：是的，一些公司已经倒闭或者被收购。对象数据库公司 O2 被 Unidata 收购，而 Unidata 被 Informix 收购，最终 Informix 被 IBM 收购。因此很难看出 O2 的技术在哪里被使用。一些公司重新定义自己为 XML 数据的管理者，比如 ObjectStore 以它的 XML 管理产品 Execlon 闻名。还有一些对象数据库公司开始尝试中间件引擎和集成引擎。

玛丽安·温丝特：我听说对象关系数据库并不能满足销售商们的预期，为什么？从长远看，对象关系数据库会怎样呢？

戴维·迈尔：我认为一个原因是，对象关系数据库中的新特征扩展了模式设计的空间，但目前尚不具备类似关系数据库的范式理论那样好的设计理论来规范关系设计。当我还在威斯康星休假时，我曾与拉古·罗摩克里希（Raghu Ramakrishnan）探讨过简单的对象特性有哪些可能的设计，结论是可能的设计有很多很多。并且，已有的对对象进行建模的技术（比如 UML）并不适合。因为对象关系数据库的对象模型与 UML 的对象模型并不吻合。我看不到哪些未来的终端应用会使用这些新特性。也许我是错的，但当我对一屋子的人说时，没有人站起来指出我错了。这些新特性似乎只在人们构建数据库系统的扩展时被使用，比如 Informix 的 DataBlade。对象关系特性在构建这些扩展时被使用，之后人们购买这些扩展后的产品。因此我想他们在用一种封装的方式使用这些特性。

玛丽安·温丝特：研究界需要对对象关系数据库做更多的工作吗？还是他们应该把精力放在产品上？

戴维·迈尔：我认为为对象关系提供设计工具的研究是有趣的。我认为在人们知道如何把问题映射到对象关系模式之前他们不会使用对象关系的新特性。我不认为目前的对象关系数据库具有性能问题。当人们知道如何使用这些特性以后性能才可能成为问题。我确定在一些新特性上会涉及一些有趣的优化问题。

玛丽安·温丝特：那么，模式设计的工具是应该通过理论研究者们的研究还是通过用户的实践才能解决的问题？

戴维·迈尔：数据库界对这个问题最感兴趣的是那些做建模的人。他们研究实体-关系模型等类似的东西。如果搞建模的人能够提出一些像实体-关系模型那样的工具，我认为将是很有用的一件事情。

玛丽安·温丝特：您认为哪些对象关系数据库的研究成果会产生长远的影响？

戴维·迈尔：我认为对象数据库体系结构的工作具有较大的影响。比如如何把数据库系统划分成服务器端和客户端？通常，数据库运行在一台机器上，而应用运行在另一台机器上。现在，一些面向对象的系统，实际上把数据库运行在客户端。这是一件好事，因为客户端可能是计算最密集的地方。探索这些体系结构对关系系统也是很重要的，这样他们可以把更多的处理放在客户端。

另一件我认为很重要的是如何基于代价对象查询语言做优化，我认为这对 XML 查询语言很有用。

玛丽安·温丝特：我听戴维·德维特（David Dewitt）说对象数据库失败是因为没人知道如何对查询做优化。这话现在还正确吗？

戴维·迈尔：嗯，一件事做2年和做10年可能会有很大区别。我们确实需要很多专门的人来做查询优化，但很多对象数据库公司无力承担。也许学术界做这件事更合适。

玛丽安·温丝特：您想过一旦我们知道如何优化，它将很快进入什么产品吗？

戴维·迈尔：我不确定。一个有趣的产品也许是 JavaBeans，这是一个中间件体系结构。我刚刚读过 JavaBeans 2.0 的说明书，它具有关系模式和查询语言。因此，JavaBeans 越来越像一个对象数据库了。一些查询处理的技术可以用来优化 JavaBeans 的查询。另外，我认为最适合的应用是 XML 查询语言。

玛丽安·温丝特：你认为 XML 将比对象数据库更适合市场吗？

戴维·迈尔：我认为事实已经是这样了。人们还没有意识到 XML 已经渗透市场了。一些人对我说，“我爬取的网页是 XML 文档，但我没有看到它们”。人们需要意识到网页不是互联网，互联网上发生的事情也不仅仅是上传网页。实际上，你看的是 XML，只是它们在进入浏览器前被翻译成了 HTML。在我最新的对象数据库课程中，我要求学生构建数据库从 bioDAS 服务器中获取的生物序列做标注。我的学生们构建的数据库接受用户的查询并返回 XML。这里 XML 可以直接进入某种特殊的浏览器。因此 XML 就在那里只要你知道如何找到它们。

另外消息通信中也使用 XML。

玛丽安·温丝特：今天数据库理论研究的角色是什么？

戴维·迈尔：我认为数据库理论家们应该寻找真正的问题类并抽象出这些问题中真正难的那些部分或寻找它们的相同点。然后他们应该研究我们如何解决这些问题，或者至少如何思考这些问题。我不认为数据库理论家们应

该继续昨天的理论。

玛丽安·温丝特：那么他们应该去哪里寻找这些新的难的问题？

戴维·迈尔：我认为流数据是一个好方向，会有很多的理论和语义的工作需要做。

我认为科学家们使用的时序数据、多维数组、有限元素网格是我们长期以来忽略的地方。很多搞理论的态度是只研究记录的集合以及具有较好模式的数据。科学家们使用的数据中已经有一些有趣的问题，比如优化数组查询。我认为现在很需要这种观察科学家们实际使用的数据结构然后看看可以做什么的研究。

玛丽安·温丝特：您最近的很多研究都是和科学数据库有关的。这是一个很小的问题，因为没有公司靠卖给科学家数据库赚大钱。您认为更多的人应该考虑这个领域中的问题还是应该坚持具有更大的市场和更大的潜在影响力的主流应用？

戴维·迈尔：我想我要纠正您的一些假设。并不是卖给科学家数据库管理系统而是卖给科学家数据库，这是能赚钱的。比如，具有大量基因表达数据的数据库可以卖给科学家。

我认为科学家对数据库的需要和市场很大的领域对数据库的需要有很多共同点。比如，金融应用、健康和医疗记录应用中的时序数据。科学家们对基因序列数据、基因表达数据库很感兴趣，企业同样也很感兴趣。因此，我认为研究如何支持这些数据是值得的。

从经济上为科学家设计新的数据库管理系统是不划算的。但很多人只是为科学家的数据结构进行简单的扩展，混合传统数据库和特制的文件管理。这看上去是一个很可行的方案，尤其很多科学数据都是只写一次的。

玛丽安·温丝特：在我对数据库领域杰出人物进行访谈的名单中，您是唯一一个不在著名大学或研究所的。这对您的生活有什么影响吗？

戴维·迈尔：如果我去 MIT 或类似的地方，我会从那里获得很高的声望。现在有点相反，有点像我的系或学院因为我的成功得到了声望。因此，你会对自己能够在自己所处的环境中成功而有信心，而不会因为身处在名校而假设人们认为自己很厉害。

玛丽安·温丝特：您对年轻的数据库研究者，那些希望产生重大影响但却没有进入排名前 30 的研究机构的人想说些什么吗？

戴维·迈尔：根据一些度量标准，俄勒冈在美国大学中排名前 30，但也许不在前 20。我认为你应该少想一些你的影响力有多大。比如回想我自己，我总是在对我现在所在的系进行建设。当我刚到这里时它还很小，而现在它正越来越被人们所熟知。

另一件事情是你应该集中在少一些领域。不要试图覆盖做所有的事情，试图集中在一个或两个你得到国家性承认的领域。我们在美国新闻和世界报道中在数据库领域排名前 5。因此不要认为你被你所在的系的整体情况限制住，选择一个小的领域不断进行突破。

而且，你应该利用你独特的条件。有一些事情如果你在别的地方可能是不能做的。以我自己为例，Gemstone 在波特兰的乡下创业，我和他们有长期的咨询管理方面的联系。我不认为如果我呆在石溪可以获得这种与工业界接触的机会。

另外你可以做联合项目。我和斯坦福大学、威斯康星大学和布朗大学的人做了很多工作。和牛人共事也有帮助。

玛丽安·温丝特：现在俄勒冈研究所与健康科学研究所合并了，这是否意味着接下来你需要研究生物信息？

戴维·迈尔：是的，我现在基本上是生物信息科学家的数据库管理员。其实并非如此 [幽默]。

这不意味着我必须研究生物信息学。很多人都将继续原来的研究，但是

合并带来了一些有趣的机会。基因、神经和心脏病发展方面有很好的研究组，因此进行合作研究也是有趣的。我并没有被强迫研究生物信息学，但是我认为我具有其他地方的数据库研究者所没有的机会。我想要抓住这些机会。实际上，在合并开始前我就和医药信息部的人一起研究了。

玛丽安·温丝特：我明白了，听起来生物信息学家们已经准备好和您合作研究了。

戴维·迈尔：是的。很多生物学家突然发现他们进入了计算领域。尤其是基因测序、基因表示使得他们有大量的数据，这些数据他们之前都没有处理过。因此，至少在数据管理和信息层面他们正在寻求帮助。他们很乐意和我们沟通。

玛丽安·温丝特：您对刚开始或者已经在职业生涯中的数据库研究者们有什么话说吗？

戴维·迈尔：我认为我的主要忠告是成为无敌的人，而且知道有很多方法都可以让你成为无敌的人。一个方法是选择一个没人比你更可能成为专家的小领域。但你需要承担没人注意的风险。另一个方法是对很多领域都有知识，使得没人比你懂得更多。

以我自己为例，我认为我的无敌来自于同时做系统和理论。最美好的一件事是：戴维·德维特说，你是一个了解很多系统的做理论的人。同时一个做理论的人会说，你是一个了解很多理论的做系统的人。我认为我的影响力来自于在理论方面发表了很多文章而且还了解很多系统。

了解数据库的很多领域和相关的应用也是成为无敌的好方法。

玛丽安·温丝特：我也同时做两件事。既然您提到了，我想不到很多同时做理论和系统的人。这是一件坏事吗？

戴维·迈尔：我认为很多人是先做理论后做系统。大多数数学家在他们

25 到 30 岁之间发表了他们重要的成果。因此他们可能试图转变，就像我从理论研究转到系统研究。但你是对的，很少有人长期同时做这两件事。我认为这样的人多些会更好。

玛丽安·温丝特：很多人认为美国的高等教育系统将无法继续，正如我们知道的那样，由于网络的影响。实际上，终身教职制度并不能为社会带来最好利益的传言也有很多年了。您在研究所工作了很多年并没使用终身教职制度，我猜这让您对这个问题有独特的看法。您可以说一些吗？

戴维·迈尔：是的，我认为终身教职系统已经偏离了其保护学术自由的初衷并沦为无效的系统了。

至少在计算机科学领域，我并不担心人们有自己的政治信仰。

如果决定是否与我续约的管理者不喜欢我的工作，我注意到这些管理者一般 5 年后会离职，因此如果他们经过一段时间发现不喜欢我，我们签署续约的时间会保证他们比我离开的早。所以我不担心我的合约不被续签。

在研究所，我们确实需要晋升，并且我晋升到副教授的时候也是和我同时毕业的人得到终身教职的时候。然而，他们为了得到终身教职把日子过得太疯狂了。对我来说，这里的晋升并不困难。总之，我避免了为了获得终身教职而经历的艰难。

过去的 20 年中有一个有趣的趋势，大学中拥有终身教职的员工的比例从 80% 降低到 50%。这是因为很多员工退休了，他们的位置被合同工、临时工代替。因此，终身教职并不像以前那样很重要。

另一个优势是我们没有太多没用的东西。通过与其他机构的人们交谈，我知道获得终身教职的人会比零产出还差。他们组织的委员会以及申请的项目都可能成为别人的阻碍。

玛丽安·温丝特：如果您有时间做一件您现在没有做的事情，您将做什么？

戴维·迈尔：我会选择选修我们系里的课程。在这里我们有很聪明的同

事讲授很好的课程。我通过让学生去听课间接地从课程材料中吸收了一些。但我应该亲自去那里更新我的知识。

最有用的事情之一是我利用公休上了两个学期的 Marv Solomon 的操作系统课程。这是因为当我读研究生时还没有操作系统的课程。那时只有共享内存、并行处理以及网络。

玛丽安·温丝特：您在您的职业生涯中遇到过难以处理的事情吗？

戴维·迈尔：是的，有一次我读到了一篇论文，他们解决了对象数据库中的重要问题，我很开心，并认为我不能比他们说的更好。突然，我意识到我自己说过这些东西，意识到这篇论文中的一部分是在剽窃我的一篇论文。我向作者说明剽窃是不允许的。在我即将忘了这件事时我发现这篇论文将被授予最佳论文奖。因此我联系了程序委员会主席。但奖项不能撤销，尽管我知道有 15% 的论文是我的，也许我应该获得 15% 的奖。[笑]

玛丽安·温丝特：您可以把这列在您的简历上。

戴维·迈尔：是的，列出一个得奖论文的简介和写作动机。我再也没有看到有关这个剽窃的任何公开信息，但我应该和那些人所在的系说明他们应该更加注意这些事情。

玛丽安·温丝特：作为计算机科学家，如果您可以改变，您想做什么？

戴维·迈尔：让我的脾气更怪一些，对更多的事情说不。我不想让别人不高兴，因此我很难说不。结果我的很多时间花在评阅工作或程序委员会中以及采访中。

玛丽安·温丝特：好吧，这是我今天的所有问题。

戴维·迈尔：你还没有问有关巴黎的问题，那是我最喜欢的问题。

玛丽安·温丝特：您是否曾经去过巴黎的每个博物馆？您有最喜欢的吗？

戴维·迈尔：我没有去过每个博物馆，我还没有去过面包博物馆。但我去过的博物馆比巴黎人平均去过的博物馆要多。我想我最喜欢的是我最近将去的假货博物馆。假货博物馆展示更多的假商品而不是假钱。因此，这个博物馆中有假的范思哲手提包、假香水、假表。去看看人们觉得什么值得造假很有意思。

玛丽安·温丝特：您可以分别假的东西和真的东西吗？

戴维·迈尔：一些很显然，你搞不清为什么人们会买假的还以为是真的。一些很难分辨。人们很聪明，抄袭商标和包装进行造假是很容易的。汽车零件也可以造假，如果没有材料检测机很难分辨。我不认为可以凭借肉眼分辨。

还有一些其他的有趣博物馆，比如魔术博物馆，我喜欢巴黎工艺博物馆，那里有拉瓦锡（Lavoisier）的化学实验室的装备。任何对法国博物馆感兴趣的人都应该看看我的网页，那里有我在 INRIA（法国国家信息与自动化研究所）休假时写的笔记。

玛丽安·温丝特：谢谢大家的收听，谢谢您接受采访，戴维。

戴维·迈尔：不客气。

（王春凯 译，孟小峰 审校）