

## 吉姆·格雷访谈录

探索对象-关系之间的联系；为何性能不再是问题；糟糕的想法会变成好的想法；重塑数据库领域；帆船；垃圾论文；争取一场全垒打；更多关于他的趣闻轶事。

玛丽安妮 温丝莱特

**玛丽安妮：**欢迎来到本期 ACM SIGMOD RECORD 访谈。该系列访谈主要采访数据库领域中的杰出人物。现在，我正在威斯康星州的麦迪逊城，PODS2002 和 SIGMOD2002 联合会议的主会场，将要对吉姆·格雷进行采访。吉姆·格雷在计算机领域，特别是在事务处理方面做出了杰出贡献，于 1998 年获得图灵奖。他担任过总统信息技术顾问委员会委员，主要是为总统的技术难题提供参考性意见，也曾担任过国会图书馆的咨询委员。他于 1969 年获得伯克利计算机博士，曾先后在 IBM、Tandem 和 DEC 等公司工作过，目前在微软工作。格雷先生，欢迎您。

**吉姆·格雷：**谢谢，很高兴来到这儿。

**玛丽安妮：**很高兴能邀请到您。您在今年二月份的 FTP 在线的采访中说过“我们已经找到了对象-关系之间的联系，并且很快就要实现了”。但是有人却说市场上基于对象-关系的产品，还没有预期的那样成功。是理论研究的不足，还是对象-关系技术需要一个应用契机，还是误解了商业市场，还是其它原因导致那些人的误解？

**吉姆·格雷：**我认为仅仅是因为人们缺乏耐心。科技的产品化和实用化，比想象的需要更长时间。以散列连接为例，大约经过 15 年的时间它才成为数据库的核心技术，现在大多数数据库系统都包含这种技术。

八十年代，有一大批的创业者从事对象-关系系统的研发。为了改善现状，他们认为有必要消除产品与市场的不协调，有必要对产品进行包装。他们提出了许多有意义的方法，解决了人们想往数据库中存储对象，而不仅仅是数字、字符串和超长二进制字符串。那些方法都很有意义，并且也很正确。但坦率地讲，他们开发的产品和人们的实际需求确实存在着差距。实际上，搞面向对象数据库的研究人员，在处理对象方面做了一件不错的工作。但是，在对象如何存入磁盘，事务，或安全方面，他们做得差强人意。相比之下，IBM、Oracle、微软等公司却抢了风头，这些公司说：“哦，我们可以存储对象，我们可以这样做”。这样说确实有点不近人情，但是，像 SQL-J 和 ADO 一样，语言和数据库确实正在统一，这些公司为数据库提供一个相当好的对象模型。根据以上情况，我们有好多的答案来回答这些问题：“什么样的对象是一个数据库？”，“什么样的对象库是一个数据库？”。相反地，好的对象模型告诉你如何把对象嵌入到关系数据库中去。虽然，目前这些工作还没有标准化，但我相信再过十年，SQL 标准，或者它的后继产品会有对象嵌入。

在对象-关系世界里，最适合我的活动是与微软的同事继续从事所喜爱的研究。同事们做的是把通用语言的运行系统与数据库系统进行集成，这样你可以使用自己擅长的语言，编写出像数据库存储过程一样的程序，而且程序使用对象模型，来操作记录和记录的字段。相信 Oracle，IBM 和 DB2 的研究也正做同样的事情。衡量对象-关系技术成功的简单标准是“你能买到什么？”，或者“下一年或两年内人们希望的是什么？”。以上问题的答案是所有系统要具有对象-关系的特征，并且用户可以使用。所以说，我们要有足够的耐心等待。

**玛丽安妮：**据我所知，你的博士论文是关于语法分析理论的，什么导致你转向数据库领域，又是什么让你远离理论方面的研究？

**吉姆·格雷：**是的，由于对任何领域都有几分兴趣，我觉得应该能够写出一篇偏于实践的，或偏于理论的论文。我是那种急于求成的人，仅用一年半的时间就拿到了博士学位。

如果你急于毕业，就写偏向于理论方面的论文，原因是这方面的论文，需要你证明定理、挖掘证据，需要你描述已经完成的和正在做的工作。然而，在一个偏于系统的或者是偏向于实践的论文里，你要搭建一个系统，并且要解释是怎样做的。而这种论文的工作量是理论论文的两倍，原因是你要完成一个的系统，还要围绕该系统写论文。

我做博士论文时候，是计算时代的初期。后来 60 年里，操作系统领域出现了许多动荡。实际上，那个时候我和布特尔、戴维、保罗、查尔斯、霍华德、布鲁斯这些家伙一起研究面向对象操作系统，当时，也研究城市结构良性发展的仿真系统。我做了许多不同类型的研究，但是必须选择某一个与学位论文相关的课题去做，这样对于写文章比较容易。

博士毕业后，继续从事着理论研究与系统开发，先在伯克利做了两年的博士后，而后去了 IBM 工作。你刚才问我是怎样从操作系统转向数据库领域的，答案比较有意思。我上司的上司，一个叫雷纳德的家伙，不仅友好，而且是个不错的管理者。一天他来到我的办公室，坐下来：“你知道，吉姆，现在市场上已经出现了很多操作系统，但是到目前为止，还没一个像样的网络操作系统和数据库系统，如果你真想在 IBM 做点事，研究网络操作系统和数据库系统是很有前途的”。当时我正研究操作系统。我听从了他的建议。事实上已经有很多人为操作系统做出了伟大的贡献，我确实没有好的想法来提高当时操作系统的性能。我曾经从事的，似乎只有最聪明的人才能做的面向对象操作系统的研究，彻底被放弃了。所以，我做了件顺其自然的事情，开始研究数据库操作系统。

**玛丽安妮：**这否意味着你曾经从事过 SystemR 的开发工作？

**吉姆·格雷：**哦，是的，确实在 SystemR 组工作过。当时我负责从构建进程、优化虚拟内存、启动授权方面来提高 SystemR 的性能。后来，从用户的角度，研究了并发控制、故障恢复，以保证 SystemR 的正常运行。实际上，雷蒙德已完成了 SystemR 的所有 I/O 操作，并给出一个低瓶颈的操作接口。我完成了该系统的配置、启动、进程构建、锁机制、系统日志以及进程间的通信。这些就是我在虚拟机 370 上，使用一种简单的语言 PLS 所做的工作。

**玛丽安妮：**听说当时你的主管经常敲门来给你警示，那是怎么回事？

**吉姆·格雷：**导师麦克·哈里森常教我要把一些事情记录下来。所以，无论何时去旅行，都要写一个旅行报告；无论何时与人谈话得到的想法，都要做备忘录，并归档。凭借这种习惯我写了许多文章，参加了许多国际会议，并出名了。这可能对那些做同样事情的人来说是不公平的，但那就是生活。我常说自己的文章思想大都来自佛朗哥和欧文。但是，记笔记和做报告的习惯，使我在圈内得到好评。

实际上，我最终更像一个研究者，而不太像一个研发者，尽管参与研发了 SystemR。佛朗哥为 SystemR 一年写了两万行代码，并调试通过；我一年内才写了一万行代码。SystemR 有各种各样的产品，而管理者想展示原型系统，这对我们来说，发行版本有一定的压力。所以，上司经常敲我的门说：“快点编码”。有时，他会因我花时间写作、旅游或偷懒耍滑而感到沮丧。我确实因他的敲打快速编码，但代码中出现了许多缺陷。在 SystemR 中，我大概写了五万至七万行代码，具体多少，也记不清楚了，主要是涉及并发控制、系统恢复、系统

启动、安全性管理等工作。

**玛丽安妮** 那么,现在这些核心数据库的内核研究是过时了呢,还是要投入更多的研究精力?

**吉姆·格雷**:随着时间推移,摩尔定律一直操纵着和改变着许多事情。许多过去争辩的观点已经消失了。我曾反对指令操作,担心数据库将来变成多大,发展多快。在毫秒级下运行超越函数运算器的事实一直让我震惊。在泰勒系列级别时间内,处理器就耗尽了。然而,寄存器里却出现了双曲线反余弦函数级的时间。后台编程的时代过去了,现在计算机能在自我管理、自我组织、自我救治情况下完成某些任务。机器的性能已不是问题,最大的问题是如何使其简单化。

过去,我们认为在缓冲池中能否放一千个页是个大问题。而现在,能否放下两亿个页是个大问题了,并且相应的算法也不尽相同。缓冲区中处理检查点的方法、查找文件的方式,以及页替换策略也发生了变化。目前,并发问题与以前相比已完全不同了,比着过去我们很多并发事务要处理。今天,数据库的并发性已不是问题。在我们那个时代,大约有 50 至 70 人研究数据库的并发性,如果至今还有 50 至 70 人从事这个工作,就太好了。实际上,这个领域还有很多好的工作可以做,只不过其它领域更容易实现目标罢了。

**玛丽安妮**:一些研究者争论说讯盘是个糟糕的想法,以前打着数据库的幌子来试验它,最后总是失败。而现在你认为讯盘是个好的想法,为什么呢?如果如你说的那样。

**吉姆·格雷**:有趣的是,我昨天与戴维也争论了这个问题。他说讯盘在过去是个好想法,而是现在是个糟糕的想法。这有悖于我的观点。讯盘在过去确实是糟糕的想法,它只是被当做专用的计算机和专用的软件,偶尔提高一下机器的性能罢了。实际上,辩论这个问题是错误的,原因是计算机的性能并没有因此得到提高。过去,人们提出这个想法,只是想从中获利而已。

问题是,人们在那个年代本想设计出自己的专用硬件,等设计出了专用硬件,通用硬件早已跑到他们前头了。再后来,磁盘容量越来越大,价格越来越便宜,并且磁盘机已有了盘片、磁头和一些机械设备、打印电路块、处理器、内存以及网络接口(类似于今天的 IDE、SCSI 接口)。当时的网络要么是私有专用的,要么是比较怪异的,如光纤信道网 SCSI、ATA 网,当然你可以使用 Ethernet 来替换那些怪异的网络。现在这种讯盘驱动器已有几兆赫或几百兆赫的处理器,几百兆赫的随机存取存储器和一个网络接口。

当然,你希望这种磁盘机具有报错功能,配置功能和交流功能。要实现这些功能,必须有超文本传输协议服务和 Web 应用接口,通过 Web 服务和简单对象访问协议,就可以与讯盘交流。讯盘里有一个携带网络堆栈和 Web 服务堆栈的操作系统。如今已证明讯盘中的处理机性能比起八十年代后期的 Sybase、Oracle 和 DB2 要有优势的多。所以,在这种磁盘驱动器中,可以运行任何的数据库系统,当然也可以引导任何操作系统。磁盘驱动器中可以放入通用的软件,来优化磁盘臂,这样做很有意义。磁盘臂是最珍贵的设备,电子设备的价格逐渐趋于零,而真正有实价的是些物理器件。因此,磁盘臂越具有智能,越有意义。

几年前,剑桥的埃里克,伯克利的金·基顿等几位博士做论文时,大约只有 200 兆赫的磁盘空间来测试他们的数据库基准。几年之后,磁盘强大的处理能力和足够的存储空间,可以运行所有当今的数据库系统软件。这就是为什么我认为软件堆栈会逐渐移入磁盘驱动器事务原因。顺便提一句,这样会产生有意思的计算机体系结构:体系结构中不存在处理器。所有的处理器都移到磁盘中,或者移到网络适配器中,或者移到打印机或显示器中,或者移到键盘或麦克中。它们也可移到下一代变频器中。

然而,问题是这种系统如何自我组织?新添加的设备如何与之前的计算机系统集成?以及其它的一些问题驱使我们要在每个设备上添加更多的智能。今天我们开始看到了这种趋势,并且五年或十年之后,这种趋势会变成事实。假如开发软件需要很长的时间,我们如今所从事的软件开发即是五年或十年之后做准备。

**玛丽安妮:**听起来此领域可能有很多研究点,人们应当关注。

**吉姆·格雷:**是的,自我组织系统确实值得研究。IBM 采用自动计算作为描述它的术语,而个人机用户采用插件作为描述它的术语。

独立冗余磁盘阵列如何在这样的系统中工作?两块一起工作的磁盘如何实现容错性?独立冗余磁盘阵列是错误的概念吗?我们仅仅在数据库层做复制操作吗?新的设备加入时,系统如何重新平衡自己?如何做查找?当你有一万个磁盘并且每个都安装了数据库系统时,如何做查询优化?这些点都值得研究。但是,可扩展性与可管理性是两种不同的研究点。

**玛丽安妮:**不远将来,磁盘空间大部分是免费的,事务严格可串行性的情况如何?我们要采用版本化的方法来考虑事务并发控制吗?要重新考虑恢复机制吗?

**吉姆·格雷:**我们当然要考虑事务的版本化。关于事务的版本化,20年前,大卫·里德开发出一个很漂亮的系统:Swallow。该系统是在当时的光盘只是一次写,多次读的情况下完成的。Swallow 是个面向对象的数据库系统,从不过多写,它的对象被放到缓冲池中。有时,该系统把一个要提交的记录扔进缓冲池,当所提交的记录到达缓冲池时,相应的事务也被提交。Swallow 系统中的对象在它们有效期内是时变的。一个更新操作会终止该对象当前值的有效性,并在新的有效起始点创建一个新的事务。所以,该领域里,有很多有趣的点可做,如讯盘、面向对象数据库等。技术的变革使过去那些糟糕的想法变成了非常好的想法。

**玛丽安妮:**斯通布雷克在 SIGMOD2002 主题演讲时,他不是故意告诉我们这种多版本事务处理方法由于速度慢,有可能行不通吧?

**吉姆·格雷:**他所说的 1996 年的事情。当时 POSTGRES 系统的垃圾回收器没有成功实施。(POSTGRES 系统从不过多写数据,它的垃圾回收器负责回收那些已更新数据所占有的空间资源。),而我们所讨论的是将要在 2006 年或 2016 年发生的事情。我曾十分谨慎地跟踪过磁盘价格的变化,由于垃圾回收器的失败,磁盘价格以百为基数下降。磁盘臂的价格也以十为基数下降。价格的下调比率确实左右了一些事情。首先是,还继续发行多版本数据吗?当然要!还按照 POSTGRES 系统方式组织那些多版本数据吗?可能不需要。那么怎样组织它们?这是个好问题,似乎对我来说是个好的研究点。

**玛丽安妮:**你已经从事数据库研究近 30 年了,这个领域还能再活跃 30 年吗?

**吉姆·格雷:**不可能。我们计算机人在字节和算法上有优势,在信息科学上有优势,研究信息科学和设计算法也正是要做的事情。而在数据库领域里,我们已经非常狭隘。眼下所研究 SQL 语言,没有数据显示和可视化功能。SIGMOD 这个词中的 MOD 表示“数据管理”。而对我来说,数据管理包括很多工作,如收集数据、存储数据、组织数据、分析数据、表示数据,特别是数据的表示部分。针对数据查询已经做了相当多的工作,但这些工作仅仅围绕查询画了个艾普西龙球面,而没有真正超越它。所以,如果我们还像以前一样把研究与现实脱

离开来，这得再次回到斯通布雷克的话题，还继续保持狭隘的眼光审视自己所做的研究，数据库领域将要消失，因为那些研究越来越偏离此领域。眼光狭隘会使我们口无遮拦地说：“我们管理信息数据，数据库不会消失”。现在人们已经拥有太多的数据，而我对许多人说我只希望拥有更多的时间。所以，整个数据搜集、数据分析和数据简单化的工作，就是能准确地给予人们所要的数据，而不是把所有的数据都提供给他们。这个问题不会消失，而是变得越来越重要。如果你用一种大而广的眼光看 SIGMOD 关注的研究点，数据库确实是一个蓬勃发展的领域。如果采用审视的眼光看它们，现在做的很多研究对 30 年后的人们不会产生任何影响，特别是关于性能的研究。我不会再关注计算机指令问题，而是关注 I/O，但是也许有一天 I/O 问题也不必关注了。

**玛丽安妮：**数据库领域的研究，越靠近人类生活，成功的可能性就越小。比如，早期的自然语言查询响应与用户接口研究都相应失败。时代发生变化了吗？现在我们能够在过去失败的地方取得成功吗？

**吉姆·格雷：**自然语言数据库的研究证明了自然语言不适合数据库。但是通过例子查询语言 (Query by Example, QBE) 给人们一种视觉暗喻。基于 QBE 也产生了相应软件：VisiCalc，第一个屏幕表格显示软件。QBE 是一种高度非过程化的查询语言，用户通过终端屏幕编辑程序以填写表格的方式构造查询要求，而查询结果也是以表格形式显示。但我不知道它将来将是什么样子？

**玛丽安妮：**这是我下一个问题

**吉姆·格雷：**我还没想好。但是越来越多的人发现 QBE 语言、VisiCalc，和它们的后继产品非常有用。微软的 Access 非常成功，它基本上采取了 QBE 的思想，仅仅对其泛化了而已。所以，提供操作对象的虚拟接口，可使很多人对数据库进行访问。

**玛丽安妮：**这次访问前，有人坚持让我问问你关于沉船的事，听说差点溺亡。另外，还有其它的故事让我们分享吗？

**吉姆·格雷：**那件事有些尴尬。

**玛丽安妮：**你不打算透露细节了？

**吉姆·格雷：**还是比较有意思的，比较有趣。

我一个浪漫的人。人应该有一个完整的生活，并且要劳逸结合。我曾经对帆船非常着迷，有个帆船可以环游，于是，就买了一艘帆船，并且在上面呆了 10 年。我从来不会航海，买船时，专门去图书馆找本书学习如何驾驶。有件事情没预料到，帆船要有停泊位，而买车不用担心停车位，但是帆船不行，必须租借一个滑道。

寻找滑道时，我找到一个相当友好的港口管理者，并告诉他我的意图。这家伙说可以，然后给我张申请表，填好表后问他等多长时间能轮到我。15 年或者 25 年吧，他说。我认为这也太不靠谱了，就在一家制糖厂附近，租借了一个相当肮脏的，不中意的滑道。我在阿拉梅达口岸练习过帆船，那是个学习航行的好地方。

有次我赌气说“忘记这些，搬往旧金山。”于是，我抛了锚泊线，搬到旧金山。我找到港口主管对他说：“听着，伙计，我要租赁这里，要把船停靠这儿，这是我的电话号码。如

果你想让我移开船,打这个号码就行,我三个小时到达这里。”如果你打算把船短暂性停靠,他们不得不租借给你,于是我的船在那儿停靠了三个月。

每天早晨和深夜,我在码头踱来踱去时,注意到码头的一只船正逐渐下沉。我把该船捆绑了几次,给船主打电话要 500 美金买下它。船主接完电话后一个小时,我就拥有那个快沉的船和一个不错的滑道。我现在有两艘船了,但感觉不是特别方便,于是我需要处理掉 SouSea 号。怎样处理掉一艘快要下沉的船?对谁来说,都是个挑战。当然你可以喊海岸警卫队帮忙把它拖走,但那时我不知道可以这样做。那晚极度漆黑。我靠近甲板时,布鲁斯和我那些醉醺醺的朋友们开车跑了,伴随着旧金山的犬叫,披着漆黑的夜晚,我只有继续航行。我就坐在这艘即将沉没的船上等待,所幸船没沉没。我的朋友逛了一圈回来开始找我,因为玩笑结束了。我坐在那里喊叫。这段比较有意思。

**玛丽安妮:**我猜他们回来救你,是因为今天你要和我们在一起,你是否和他们进行一次严厉的谈话?

**吉姆·格雷:**哦,没有,没有。我一直感到那件事可笑,他们也是。

**玛丽安妮:**在你过去的研究中,你最引以自豪的研究是什么?

**吉姆·格雷:**我确实认为并发控制事务模型是项漂亮的研究。其他人也做同样的研究,我是碰巧第一个公布它而已。这个模型已被发现或者重新发现几十次或上百次了,那些发现或重新发现这个模型的人会因他们的所做的工作而感到自豪的。

数据立方体做的也不错。

基本上,我喜爱的一些研究具有相关性和优美性。你知道,当某一研究是个新奇的想法时,你可以品味它,像这样(咬手指);它模糊不清时,确实对自己的科研有相当大的影响。你的生活中没碰到过诸如此类的事情,至少,我还没有。

**玛丽安妮:**总统的信息技术咨询委员会产生了持久性的影响吗?

**吉姆·格雷:**去华盛顿是因为我相信科学研究的一分投入,会带给社会十分的回报。我曾参加过委员会的会议,坐在会议室里却很难相信自己在做的事情能对社会产生正面影响。

总统信息技术顾问委员会(PITAC)是受政府里的一些人的控制,那些人是我们乐于倾听的听众。实际上他们要做的工作是划拨钱,可能每年经他们手里划拨出去的钱数将近 2 亿美元。国家科学基金、信息技术研究基金,这些都是 PITAC 直接资助的项目。资助机构在国会上听了 PITAC 的报告后,可以上下调动资助额度。相当多被 PITAC 资助的科学家做出了非常伟大的研究,PITAC 将有着深远的意义。

顺便提一句,科学基金资助正处于一个关键期。高性能技术和通信(HPCC)的资金受国家科学基金(NSF)基数的控制,而信息技术研究(ITR)资金却高于这个基数。我们希望两年或三年内 ITR 的资助资金也受 NSF 基数的控制。在信息技术领域,资助横向项目还没先例,NSF 应该找一些具有前瞻性的项目给予资助。在国家科学基金委当顾问时,针对如何找到有前瞻性的研究,我们也很困惑。如果提出错误的研究,可能得不到资助,这很糟糕。然而,如果错误的研究得到资助,正确的研究就失去机会,这同样很糟糕。大部分人对方向的把握不太敏感,而数据库领域的资深学者具有这种能力。所以,资历浅的学者若有好的想法,应找资历深的学者进行交流。而资历深的学者应该一起行动起来,提出好的有前瞻性的研究。

**玛丽安妮：**对那些刚起步或者处于事业中期的数据库研究者，你要给出一些建议吗？

**吉姆·格雷：**你和我刚刚谈论了斯通布雷克的报告，他的建议很切题，适合大家。

**玛丽安妮：**他的建议与大卫·帕特森的直接冲突。

**吉姆·格雷：**没错。做研究与本垒打比赛一样，不要追求单人赛和双人赛，甚至不要追求外场赛，而是要追求在露天看台里比赛，要敢于挑战。但是做之前，你要问问自己为什么这样做，在做什么。

我不相信来世，尽可能把时间花费到自己最擅长的事情上，尽可能花时间去做自己引以为自豪的事情，尽量不做无意义的事情，实际上，某些研究可能有些意义。我并不担心在某个领域持续多长时间，不担心如何得到某个职位。当主管敲门给我警示，让快些编码时，我会说，“警官，这比较容易，开除我吧。如果你不喜欢我这样做，让我滚蛋好了。”我不太在意他说什么，如果有必要的话，某一天我会辞职。我想这就是自由。当然，这常使主管做噩梦。

但是，不是每个人都想按照这样的方式生活，如果你想长期被聘用，并且保险的话，你必须使自己更强。这个世界需要几个疯子，需要很多纯粹的研究。我们不希望会议的文章都是垃圾，5%的比例可以接受，而不是 0%。斯通布雷克困惑的事情是这 0%。如果你有眼光，并且知道自己很有眼光，你就一定要追求有远见的研究。相反，如果你是个科学家，你要遵循科学的方法，重复前人的实验，拓展前人的实验，这是条捷径。你可以尽可能像巴斯德一样做创新性的实验，或者做一个不怎么出名的科学家，追随巴斯德来做自己的研究。

**玛丽安妮：**刚才你提到了会议论文，有人说会议的收稿系统被垃圾文章撑爆了，你同意这样的观点吗？

**吉姆·格雷：**对 VLDB 和 SIGMOD 两大会议的收稿，我们一直尝试能有一套防伪收稿分类机制。两大会议的程序委员会一致认为确实该这样做，但是却想不出如何去做。所有的尝试都以失败告终。我们想要的机制就是能够筛选出杰出的人才。

会议的稿件接收系统并没有崩溃。实际上，对于那些在做科学研究和做长线性研究的人来说，这种收稿系统很有效，但对那些做非线性研究的人来说效果不是很明显。几个有名的事情可以说明这些。初始的 B-树文章被拒了，早期我提出的数据立方体文章被拒了，提交的事务处理文章也被拒了。那些非线性研究的文章将来也有可能被拒。

**玛丽安妮：**你原来那些文章都被发表了吗？

**吉姆·格雷：**是的，它们都被发表了。

**玛丽安妮：**怎样使退回的文章被接受呢，您有什么诀窍吗？

**吉姆·格雷：**诀窍就是反复投稿。

**玛丽安妮：**退回的稿件，是做修改还是仅仅重新投稿？

**吉姆·格雷：**数据立方体那篇文章投到声望较低的 ICDE 会议，当时就录用了。满足五分钟陈述规则的论文基本上是被接受的，这样的论文真的不错。我也写了许多文章，这些文章可

能永远不被发表,只是放到我的网站上。与那些看这次访谈,或者读这次访谈记录的人相比,对待研究的态度,我和他们完全不同,因为我不打算自己长期被聘用,也不想追求提升,只想做好的有意义的研究。坦率的讲,那些试图长期被聘用、追求提升的人将有他们大学的同事们来评价。我很怀疑他们是否在做伟大的研究,他们的同事也会设法想出某种途径,追求长期被聘用和提升。实际上,我反感这种过程。工业界这种情况也是如此。

**玛丽安妮**：如果你有足够的额外时间去做以前没做过的事情,你希望是什么？

**吉姆·格雷**：我目前在微软工作,却与那里的研发团队接触的较少,特别是 SQL 小组。我没有足够的时间一边做着研究,一边又做着其它的事情,比如参加斯隆数字天空调查,处理华盛顿信息技术委员会的工作,诸如此类的事情。所以,与微软研发团队只是偶尔的接触。在数据库组度过了令人兴奋的时光,我们开发出的系统已经合成了文件系统、对象、数据库、XML 等。我本应该沉醉于这些成果中,但我没有。坦率地将,我错过了这些。

**玛丽安妮**：作为一个计算机科学家,如果你能够改变自己的一个缺点的话,希望是什么？

**吉姆·格雷**：应该变得更加细心些。虽然凭直觉做了许多研究,但有时直觉是错误的。我的衡量标准有时是正确的,但大部分是错误的。

**玛丽安妮**：你犯的最大错误是什么？我们知道你很多大的成就,这些大的成就使你获得了图灵奖。

**吉姆·格雷**：我曾认为许多超比例扩展的并行运行系统,不可能成为世界一流的系统。大卫·德维特和我曾写过一篇关于并行数据库系统的文章,认为到 2000 年并行机和并行系统才会出现。然而,在此之前这种系统出现了,并且已经有人能够设计出 64 通道,256 通道,对称多处理计算机系统;认为强大中央处理机不会有机会出现,当时设计一个 16 处理器的计算机对我们来说都是非常困难的。坦率地讲,1990 年以前,没有人真正成功地设计一台超过 16 个处理器的计算机。但随着软硬件变化、高级缓冲算法出现、总线结构更为合理,搞硬件的伙计们在单机上就完成了处理器的扩展。所以,我的看法是错误的。我还一直认为超尺寸范围问题会阻碍计算机体系结构的扩展,但我彻底错了。

**玛丽安妮**：非常感谢您和我们一起度过一段美好时光。

**吉姆·格雷**：谢谢,玛丽安妮。你做的这些事情太伟大了。

(张啸剑译,马友忠审校)