第8篇

西·莫汉访谈录

Interview with C. Mohan

西·莫汉简介

西·莫汉（C. Mohan）出生在印度泰米尔纳德邦的马尤勒姆镇。1972 年，考入印度理工学院马德拉斯分校，进修化学工程学士学位。第二年因在印度科技协会使用 IBM 系统软件开始对计算机科学感兴趣，转而学习计算机科学，1977 年获得了计算机科学学士学位。然后在美国得克萨斯大学奥斯汀分校攻读博士学位。1981 年 12 月完成数据库方面的毕业论文，然后加入到 IBM 圣何塞研究实验室——关系数据模型、SQL 和 System R、磁盘和原子操作的发源地，直到 2006 年。1998 ~ 1999 年之间在法国巴黎的国家信息与自动化研究所（INRIA）进行学术访问。从 2006 年 6 月到 2009 年 1 月期间，在印度班加罗尔担任 IBM 印度公司首席科学家，主要负责 IBM 在印度的技术领导和外围工作。之后又重新回到 IBM Almaden 研究中心做有关数据库方面的研究。1996 年，由于在数据库系统开发和使用方面的开创性贡献，他获得 ACM SIGMOD 创新奖。1999 年，由于 ARIES 工作广泛的商业和研究影响，他获得 VLDB 十年最佳论文奖。西·莫汉集 ACM 会士、IEEE 会士及 IBM 院士三大头衔于一身，是当今数据库界的权威人士。2009 年，他被选为美国国家工程院院士。目前，莫汉博士是 IBM 软件部门和数据管理架构委员会的成员，他的研究领域包括数据库缓存以及 DB2 和 WebSphere 环境中的新一代消息处理应用。

本专访主要介绍了关于 R*系统和消息队列的技术，印度计算机科学发展的状况，ARIES的由来，以及 IBM 院士西·莫汉的研究生涯，等等。

玛丽安·温丝特

玛丽安·温丝特：感谢 ACM SIGMOD Record 设置了备受尊敬的数据库领域杰出人物系列访谈专栏。我是玛丽安·温丝特，今天我们在 2003 年 SIGMOD 和 PODS 会议的举办地圣地亚哥。今天采访的是西·莫汉（C. Mohan），他是 IBM Almaden 研究中心的 DBCache 项目的技术团队负责人。莫汉因在事务提交、日志管理和恢复方面的工作而闻名于世，他对现有数据库产品有着巨大的影响。对于 IBM 和其他运营商来说，如何把这些技术应用在现有数据库产品中是一个关键问题。莫汉是 1992 年 TODS 上关于 ARIES 的文章的作者，ARIES 成为了每个数据库研究生资格考试的必考科目，这篇文章是唯一一篇超过 50 页限制被接收的 TODS 文章。莫汉是 IEEE 和 ACM 会士，IBM 院士，并且获得了 SIGMOD 创新奖和 VLDB 十年最有影响力论文奖。他在得克萨斯大学奥斯汀分校获得博士学位。接下来，欢迎莫汉。

西·莫汉：谢谢玛丽安·温丝特做了如此详细的介绍。

玛丽安·温丝特：欢迎莫汉，您的毕业论文是讲述理论的——在数据库加锁协议中管理死锁。提出这个理论方向之后，究竟是什么导致您毕业后走入企业？什么导致您从数据库理论转向从事 ARIES 的实际工作？

西·莫汉：实际上，虽然我的论文是理论方面的，但是不是因为我对理论感兴趣。而是因为在我读大学时还没有数据库领域的大规模系统项目，所以我做了一个能够尽快毕业的论文并且获得了我的博士学位。但是，此时我对实际系统比较感兴趣。我写了一篇关于 SSD 1 的评论报道，SSD 1 是一个分布式数据库系统，我详细地分析了 SSD 1 的设计。我把这篇评论发送给不同的人，而且我渴望真正地脱离学校并且更多地参与到开发实际东西中。这就是为什么我没有申请任何高校工作的一个原因。我热衷于加入研究实验室，来获得更多操作实际系统的能力，进而改变它们，让产品体现出一些新技术。

玛丽安·温丝特：当您到了 IBM，您参加到 R* 团队中，自那时起您一直在 IBM 任职。之后有没有想过，在 R* 所做的工作是否是您所希望的？

西·莫汉：我是在项目中途加入的。在那段时间，即使在 IBM 研究实验室，我也没有真正与产品开发人员一起工作过。R* 起初关注的是同构的分布式数据库。我们假设分布式数据库网络上的不同节点都是 System R 节点，然后我们研究两阶段提交协议、复制问题、分布式查询编译和分布式优化。但是，问题是既然我们只关注分布式数据库中的同构方面，当产品商业化的时候，证明了即使是 IBM 自己，还是会推出具有某些不同功能的不同关系数据库产品。解决 R* 研究项目中的异构问题会是一个更加富有成效的实践训练，也可以成为其他项目的基础。

玛丽安·温丝特：这样看来，异构情况比同构情况，是一个更难解决的问题。

西·莫汉：的确是。

玛丽安·温丝特：当时把精力放在解决异构问题上是不是为时过早？

西·莫汉：我不这么认为，当时许多公司（比如美国计算机公司（CCA））

已经有人开始研究这个课题。我们在同构情况上的关注主要是因为开始 R* 项目的人们很熟悉 System R，并且主机上的 DB2 仍然在开发过程中，直到 1984 年才开始发售。

玛丽安·温丝特：当我比较 ARIES 和数据库领域内的其他重要工作时，我发现 ARIES 具有本质上的不同，它是一大堆需要被正确处理的细节聚合体，而不是一个单一的较大创新。这点从不同侧面解释了为什么有关 ARIES 的 TODS 论文的篇幅比其他有影响力的论文长。您能说一说这些本质的不同，更一般来讲，在数据库世界是否能够达到一种状态——一些创新的想法不再起到重要作用？

西·莫汉：我认为 ARIES 论文比较长的原因是它试着综合了所有相关工作，并且对那些可能成为 ARIES 算法一部分的特征作出了解释。我翻阅了过去的论文，发现大部分问题都没有写到发表的论文中（即使它们可能提到了）。作为一个细心的人，我感觉研究团队没能很好地把握整体——以一种高效可靠的方式包含并发控制、恢复和存储管理。我带着这些问题了解一些事情——不但阅读过去的论文还看了较早的 System R 代码，并且了解了一些没有被记录下来的特征。我也和 System R 的人员聊过，他们中的一些人仍然在 IBM 研究实验室工作，我还了解了层次数据库并阅读了代码，发现了 IBM 产品的某些特征。做完这些工作，我感觉学术界和 IBM 产品开发团队会受益于所有这些被收集的信息。所以，我选择写一篇关于恢复、并发和存储管理三方面的整体性文章。

玛丽安·温丝特：您在 ARIES 的技术转化的经历中学到了什么？在今天的 IBM，您和产品组是什么样的关系？

西·莫汉：事实上，论文中的 ARIES 算法来源于与 Don Haderle 的交流，那个时候他是主机上的 DB2 的主要架构师。正是通过与他的交流，使我意识到 DB2 做某些事情的方式不同于 System R 人员做事的方式，后来 System R

作为 SQL/DS 被商业化。System R 使用基于影子页的恢复并且遗留下一个开放问题——怎么样做记录加锁和写前日志。当我试着理解这些事情为什么不同时，我找到了一点感觉——真实问题是什么以及解决方法的特点是什么，这是基于 Don Haderle 关于 DB2 产品的真实经历，所以 ARIES 的产生与产品开发人员和研究人员的紧密合作是分不开的。

ARIES 工作促成了数据库技术研究所的建立，这个数据库技术研究所成为一个合作框架模式，在这种模式下 IBM 的研究人员和产品开发人员在一起工作。技术转化变得容易了，因为我们早就和产品开发人员一起工作，尤其是在我们研究项目的系统设计和实现阶段。从那时起，我开始密切关注产品开发。站在一个研究者角度，我努力尝试做那些具有挑战性的技术工作，论文顺便也就写出来了，并且对于真实问题来说它是适用的，因此把它并入到产品中。在 IBM 工作时，我也试着平衡这两方面。我继续致力于面向产品的东西，不但在数据库领域，并且也有其他领域，比如说 WebSphere 和 Lotus Domino/Notes。给定 ARIES 工作的基本特征，它主要用于任何管理永久保存数据的系统中，ARIES 算法及变体不只被用到到 IBM 关系数据库产品中，也被其他公司的产品（比如微软的 SQL Server）所采纳。它们也被集成到消息系统（比如 MQSeries 和 Lotus Domino）中，用于基于日志的恢复。所以，ARIES 已经深入到很多地方中，我希望在 IBM 继续做这方面的产品和研究。

玛丽安·温丝特：分布式提交——是什么？为什么用户不喜欢它？

西·莫汉：这个工作是我加入 R* 计划之后的第一个工作。给我的任务就是要在 R* 系统中设计并且实现两阶段协调算法。我查找原始的经典的两阶段协议并且致力于开发它的变体，即所谓的假定放弃和假定提交。两阶段提交协议保证了分布式环境下的事务原子性。当一个事务对多个可回收的存储区域（数据库节点或者可回收文件）进行更新的时候，如果事务提交，那么事务所做的更新会被永久记录下来。如果事务回滚，可能是因为用户的命令回滚，或者是因为系统崩溃，那么事务所做的所有更新就会被撤销。

当把这个协议实现到现实系统中，就会遇到问题，在某些环境下，网络上不同节点上的数据拥有者不愿意放弃他们的自治权——在他们的系统内数据是否该被提交还是被回滚，而让其他系统决定，这样可能需要一段时间延迟。在这段不固定周期的时间内，对已被修改而不被提交的数据的访问会被其他事务拒绝。

玛丽安·温丝特：那么解决方法是什么？

西·莫汉：我们已经在寻找解决方案上花了太长的时间，但是在商业化方面，它还不能很好地运行。该解决方案采用了高级事务模型的概念：让节点独立地提交数据的改变。如果你需要回滚数据的改变，那么需要使用日志指出什么被改变了并且应用合适的撤销操作。对于撤销操作，定义了补偿事务的概念并且用它们来逻辑地撤销之前由单节点事务提交的数据改变。有很多文献是关于这个论题的，但是在实际实现过程中，即使在不完整的原型中，也没有一个完全的实现。

玛丽安·温丝特：那么在产品中呢？

西·莫汉：在产品中，当然也缺少。但是在工作流管理系统的背景下，最终一些想法开始拨云见日。标准化的努力，比如 Web 服务事务（WS-TX）和 Web 服务协调（WS-C），正试图给人们提供一些特征——用来建造用户自己的高层事务概念。随着这些规范变得标准化，IBM、BEA 和微软实现了这个标准化，你可以看到越来越多的公司采用这种事务的方式。

玛丽安·温丝特：问题最终得到解决。

西·莫汉：是的，会主导未来很长时间。

玛丽安·温丝特：队列系统：在数据库领域中，我们把它放在分布式应用的什么位置？

西·莫汉：在这个领域，已经有很多商业化的支持，这种支持以一种IMS 的队列事务处理的方式存在很久了，比如说数字产品，并且很多其他公司都在发行自己的事务消息队列系统。在 20 世纪 90 年代初期，IBM 引进了MQSeries 产品以实现异步的事务处理，异步事务处理可以替代用于实现应用内协调和分布式行为的同步 RPC 方法。因为人们不愿意采用两阶段提交协议作为分布式计算的实现方法，所以基于消息的分布式事务工作方式在现实世界中很流行。而许多研究团队都忽视了这个论题，只有少数论文是关于这方面的研究。

最近我们发现业务处理集成能力已经变得越来越重要，随着不同公司大规模地使用 IT 技术，并且组织部门间的工作越来越多地使用自动化，尤其是随着 Web 产生且公司事务都通过 Web 完成，基于通信的方式完成分布式计算变得越来越有意义。一些 DBMS 厂商已经引入一些高级技术到他们的产品中，使得基于数据库的消息队列系统的实现成为可能。

玛丽安·温丝特：那么在消息队列系统中有哪些研究问题？

西·莫汉：事实证明原有的事务并发被引入到关系系统中的那些特点不够好，尤其是遇到并发的增加，这个时候就需要由 API 的消息种类来支持。例如，当一个用户试着从队列获得一个消息，即使有一些比较旧的消息处于未提交状态，这些信息必须被跳过，从而为用户提供一个最近的已提交的可用信息。即使给定了数据库系统中存在的当前事务隔离特性，能够跳过未提交数据而为用户提供快速的请求应答的想法仍是很难被支持的。所以这是一个研究问题。

另一个研究问题是消息可以具有广泛变化的形式。如果你看到过 Java 的报文通信服务，它让用户自己定义新的报文头部。不同的信息运营商可以添加自己的头文件。所以这些信息的格式可以非常不同，各种各样。当你把这些信息映射到关系上时，你就会发现同样的问题，类似于当你试着在关系系统中对 XML 文档进行建模。此外，消息可以很大，所以记录日志对性能的

某些影响也需要我们解决。

第三个研究问题来自于实际，消息进入系统和离开系统很快。一些消息可能永远存在，而其他消息可能不会永存，并且使用这种方式的关系系统的效率目前并不是很理想。

玛丽安·温丝特：从代码上看，这些消息是否等价于 RPC 调用？

西·莫汉：是的。

玛丽安·温丝特：为什么有些消息需要永久保存？

西·莫汉：RPC 开始的时候没有事务的概念，后来有了事务 RPC 的概念，这些 RPC 同步地执行。如果你想使用一个异步的基于消息的方式获得等价于事务 RPC 的功能，那么你不得不保证消息分发时的“一次且仅一次”的语义。一旦我给你发送一个请求需要做一些操作，那么无论发生什么情况，消息都会递交给你。一旦你产生一个应答或者针对我发给你的消息做一些操作，那么你需要反馈一些响应信息给我。不管失败还是其他什么情况，我们都需要确认所有这些信息没有丢失。这就是为什么要永久保存消息的缘故。

有些情况，如果能够提高性能，消息的信息内容可以不必永久保存，就像在研究团队中比较流行的基础能力，即按照应用分组。根据经由这些消息传播什么样的信息，无论你是做一些事务工作还是传播一些像传感器数据以及类似的信息，你可以关心、也可以不关心被处理信息是否被永久保存。

玛丽安·温丝特：莫汉，作为 IBM 56 个院士中的一员，您认为什么使得一个人有资格成为一个院士？成为院士之后怎样改变您所做的事情？

西·莫汉：既然 IBM 有 30 万名雇员，所以成为 56 个被选中的一员，那感觉很棒，因为一个 IBM 院士是公司内可获得的最高技术职位。但是同时我们也会感觉有很大责任，因为公司期望我们成为公司的高级顾问。公司期望我们不要只关注一个项目或一个很窄的问题，而是希望我们监督公司的很多

个部分，并且希望我们在进行组织内部协调时起到领导者的作用。对于那些资历比较浅的年轻人来说，公司也期望我们对他们起到指导者的作用，尤其是那些参与到较深技术活动中的年轻人。所以，我们就不能花太多的时间专注在某一个问题上。因为 IBM 的业务遍布全球，我们需要愿意涉略很多。即使你限制自己在某个研究领域，但是仍然有很多研究实验室遍布全球。

成为一个 IBM 院士有点像多面手。我不但要关注数据库领域，还要致力于与 WebSphere 和 IBM 整个软件组相关的活动，IBM 整个软件组主要负责 Lotus 产品和 Tivoli 产品等等。通过关注 IBM 整个软件组的架构，我可以和 IBM 内任何地方的任何人进行沟通交流。我的工作还涉及到了 IBM 全球服务（IBM Global Services, IGS），它能做当前很多流行的事情，像系统集成项目、顾客意见征求和面向全球多种用户的 IT 业务——外包业务。

玛丽安·温丝特：现在数据库理论学家应该致力于什么问题？

西·莫汉：这个很难说。昨天在小组会议 [SIGMOD2003] 上我们也讨论了这个问题。如你所见，在产品方面当前什么比较盛行，那么就是 XML，目前在这个领域已经取得了很多进展。在形式化的基础之上，使用过去应用的方法仍然有许多工作可以做，因为会遇到一定的新问题，他们需要被形式化。整个 XML 领域，在 DBMS 背景下尤其是查询要怎么处理，以及当被存储为原始形式的时候怎样实现并发控制（而不是转化为关系记录），这些都是很大的研究领域，需要借助理论家的帮助，提供一个更可靠的基础。

玛丽安·温丝特：为什么您认为其他致力于数据库领域的研究型实验室，比如说 AT&T、贝尔实验室和惠普实验室，现在做得不够好呢？

西·莫汉：我认为这个问题实际上是一个事实：至少就惠普实验室来说，他们的确拥有一个 DBMS 产品，但是他们不会以正式的方式出售他们的产品。但是对于惠普实验室来说，仍然有机会通过惠普产品把惠普实验室在数据库领域的研究成果商业化，但是就 AT&T 和朗讯来说，他们根本就不是真

正的计算机公司。更重要的是，他们不是软件产品生产公司。我一直想知道，当他们有大量的数据库研究者的时候，怎样使得他们所做的工作拨云见日，能够被普通用户使用，仅仅是因为这些公司没有……

玛丽安·温丝特：的确是，很好的一点。好，那么微软呢？您怎么看？

西·莫汉：当然，微软不同于其他公司。微软一开始实际上并不真正在做数据库方面的任何严肃工作，即使他们出售一个数据库产品，因为他们出售 Sybase 的重新标识版本。一旦他们决定采纳那个产品，则改变它的内核等，使其成为微软的产品，然后他们就会看到有必要拥有一个研究组。所以，微软研究人员在做技术转化过程中能够获得更大的成功。但是如果比较 IBM 研究院和微软研究院，很明显我们做的研究时间更长。

就 IBM 研究院而言，我们这里传承下来了关系模型的发明者的遗产，因此我们和数据库领域的产品开发人员具有一个长期的合作关系。而微软研究院的例子中，产品形成是在研究组成立之前，所以他们需要一段很艰难的时间来建立他们与其的产品开发公司之间的诚信。

玛丽安·温丝特：莫汉，我们知道您是从印度理工学院马德拉斯分校毕业获得学位的。在过去十年间，印度发生了很大变化。假设您今天在印度理工学院毕业，那么您会做和现在所做的不一样的东西吗？

西·莫汉：1972 年到 1977 年之间，当我在印度理工学院马德拉斯分校学习的时候，没有关于计算机科学的毕业生项目。虽然我是一个化学工程专业的学生，但是我还是对计算机科学具有很浓厚的兴趣。所以如果我现在仍然在那儿，我会重新选择计算机科学学位，并且从某种意义上来说，在获得博士学位之前，我能够学到很多关于计算机科学的知识。为了获得尽可能多的知识，我在空闲时间里，我不得不靠自己获得参考资料，发邮件给美国高校和研究实验室。

在印度，目前学习计算机科学的学生有很多选择，谈到方式，他们可以

采用暑期实习的方式。以前没有太多的印度软件公司能够给本科生提供帮助使其在工作中获得经验。但现在很多具有研究实验室和产品的软件公司提供了相关的实习机会。在印度理工学院德里分校的校园内，IBM 有自己的研究实验室。所以，对于计算机科学的学生来说，将有更多的机会获得实际操作的经验。

玛丽安·温丝特：那么对于数据库研究呢？

西·莫汉：事实上，印度理工学院孟买分校，相比其他地方的计算机科学系，在教员人数方面已经是最大的数据库研究组。

玛丽安·温丝特：哇。

西·莫汉：相当于威斯康星过去所拥有的地位。印度科学院已经有一个组在 VLDB 和 SIGMOD 等类似会议上发表过很多高水平文章。如果印度学生想读一个数据库方面的硕士或博士，在印度就有做世界级研究的场所。

在工业研究实验室方面，在数据库领域，IBM 是唯一一家进入印度的公司。纵然如此，数据库小组仍然很小。因为，如果一个人热衷于工业界做数据库研究，可能还需要假以时日有更大的研究团队形式。但对于做纯学术研究，空间是足够了。

玛丽安·温丝特：在所有过去的研究工作中，有没有一个您喜欢、但是没有做到众所周知的工作？

西·莫汉：之前我向你描述的我在 IBM 做的第一个工作是假定放弃的两阶段提交协议。当它以不同标准（X-Open XA 协议、OSI DTP 协议和其他带有 OTS 和 JTS 的协议）的形式被工业界广泛采用时，并且被实现在不同公司的产品中，实际上人们并不认为它是我们的 R* 团队所做的工作。研究界也没有意识到那么多，但是它是一个很基础的工作。有时在分布式系统课程中会讲到。

我的第二个比较喜欢的工作，是在 ARIES 环境中关于索引并发控制和恢复的工作。虽然 ARIES 的基本恢复模式已经在教科书和课堂中占用很大篇幅和时间，但是做索引并发控制和恢复的更复杂的方面，却没有在课程中得到足够的讲述。那个工作以短文的方式发表在 SIGMOD92 上，描述了极其基础的特征。所以需要更多的教与学。

我的第三个比较喜欢的工作是提交 LSN（日志序列号）概念，一种简单方法，用以识别一页上的所有数据是否提交。这个是 VLDB90 上的一篇文章。

依我看，这些工作都很吸引人，很好也很重要，在将来需要被研究界进一步发掘利用。

玛丽安·温丝特：假设您现在有充足的时间去做一件您应该做但现在还没有做的事情，您想做什么？

西·莫汉：后来，我就没有像发表 ARIES 算法工作那个巅峰时候那样发表更多论文。我更喜欢能够坐下来写更多的我想到的和理解的东西。我没有时间，或者说有时我不能确定自己是否愿意经历写作这个痛苦过程，写出来的东西要让人容易理解，然后再发表出去。

玛丽安·温丝特：您需要一个学生。您的确需要一个学生。

西·莫汉：我非常愿意花很多时间和更多的人一起工作，但是随后我辗转各地忙于学术和管理工作。作为 IBM 院士经常参与各种活动，也使我在一定程度上失去了做这些事情的可能性。

玛丽安·温丝特：作为一个计算机科学研究者，如果您可以改变一件发生在您身上的事情，您想改变什么？

西·莫汉：这个很难回答。我猜可能是在某些特殊的软件的一些细节做得更深入些，并且更加深入地了解它们，就像我过去一直做的。同时我更希望自己能够以一种更容易理解的方式解释那些复杂的概念。以一种抽象的形

式描述算法，就像一些其他研究者能做的一样，我还不是真的在行，这就是为什么许多人感觉我的论文不好读的原因。

玛丽安·温丝特：好的，非常感谢。

西·莫汉：谢谢！

（范玉雷 译，孟小峰 审校）