## 实体关系模型——企业数据视图的基础

**摘要**

企业数据视图的概念在数据库设计过程和构建概念图式。本文讨论了使用描述和维护的实体关系方法企业数据视图。基本操作提出了改变企业模式。最后，一个举例说明两者之间的差异实体关系方法和数据结构方法在企业数据视图建模中的应用。

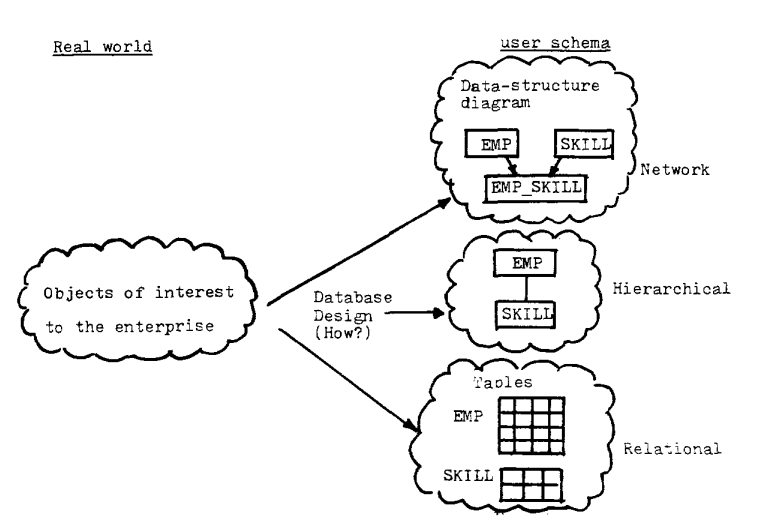
**介绍**

数据逻辑视图的主题吸引了近十年来备受关注。然而，大多数研究人员专注于用户对数据的看法。这不需要研究企业的数据观直到最近才被认可。数据库的不同用户可能对数据库有不同的看法，但企业应具有数据库的唯一且一致的视图。这在设计一个逻辑上有意义且一致的数据库时尤为重要。数据企业视图的概念在数据库设计中非常有用过程和概念图式的设计。

**企业视图和数据库设计**

数据库设计是将数据组织成表格的过程匹配数据库的底层数据模型管理系统。数据库管理系统主要分为三种类型：网络型、分层型和关系型。在网络数据库管理系统中，其中包括霍尼韦尔的 IDS 和 UNIVAC DMS-1100，数据将被组织成不同类型的记录和可以用数据结构图表示1（见图1）。在分层数据库管理系统中，其中包括 IBM 的 IMS，数据将被组织成一个形式类似于但比数据结构更受限制图表。在关系数据库管理系统中，数据将被组织成一组表（或“关系”）。一般来说，设计一个数据库就是决定如何将数据组织成特定的形式（记录类型、表格）和如何访问它们。到目前为止，工具很少可用于帮助数据库设计过程。通常情况下，数据库设计师依靠自己的直觉和经验。因此，生成的数据库可能无法满足公司的目标，并可能导致公司的问题操作。

数据库设计中的另一个相关问题是数据库设计过程的输出——用户模式（a用户对数据视图的描述）——不是“纯粹的”现实世界的代表。原因之一是数据库设计者受限于数据库管理系统的有限能力。例如，实体之间的多对多关系很难在某些数据库系统中直接表示。另一个原因是用户模式可能包含一些特征与数据库的存储表示有关。为了例如，它可以描述哪些记录类型可以直接访问以及如何访问其他记录类型。此外，用户模式通常被设计为对某些类型的数据处理操作。例如，有关员工的数据可以分为两个记录类型，员工主和员工详细信息，以改进检索性能。因此，用户模式是通常不是现实世界的直接表示。这个使用户方案难懂难懂改变。



图一

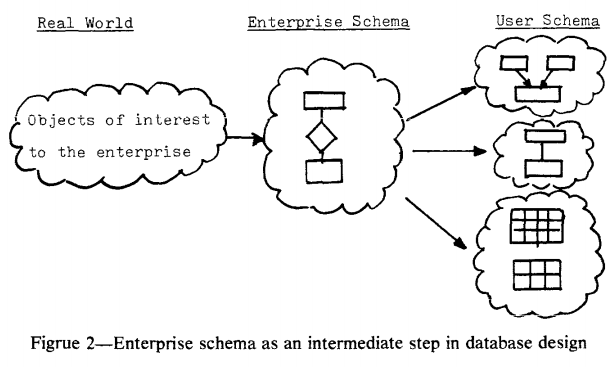
解决上述问题的一种可能方法是引入数据库设计过程的中间阶段：定义企业模式，这是一种“纯”表示真实世界，并且独立于存储和效率考虑。然后，企业模式将是翻译成不同类型的图式数据库管理系统（见图 2）。它也可以翻译成同一个数据库的几个模式管理系统优化不同类型的数据处理操作。这个方法有几个优点：

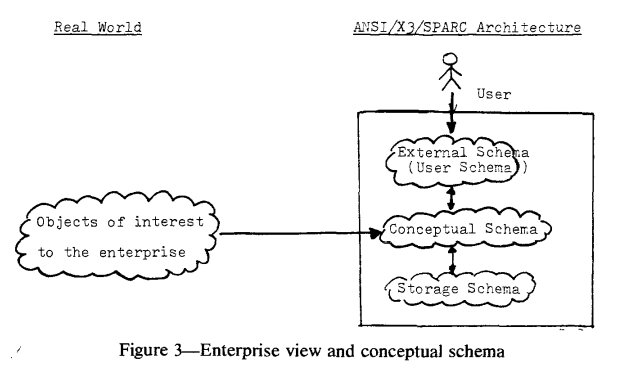
（1） 企业模式比用户模式更容易理解，因为前者没有底层数据库管理系统的限制;

（2）企业架构比用户更稳定模式，因为用户中的某些类型的更改架构可能不需要对企业进行任何更改架构。如果需要更改企业架构反映企业环境的变化，更改可以很容易地执行，因为效率不考虑存储问题。

**企业视图和概念架构**

企业架构和由ANSI/X3/SPARC团体提出的概念模式有什么区别？基本上，它们非常相似，因为两者都是企业数据视图的描述。在SPARC的方法，概念模式作为接口在外部模式（数据的用户视图）和存储模式（数据的物理视图）（参见图 3）。这作为两个其他之间的接口的要求图式可能会在概念图式。如果这个限制在概念上架构被忽略，几乎没有区别概念模式和企业模式。因此，本文讨论的技术也适用用于描述和维护SPARC的架构。





**论文中使用的方法**

为了描述企业对数据的看法，一种心理需要对现实世界进行建模的框架。不同的人们可能习惯于不同的心理框架。这本文使用的心理框架是实体-关系（E-R）模型。E-R 模型和类似方法已发现在模拟真实世界情况中很有用。本文将使用一种称为实体-关系（E-R）图的图表技术来表示企业数据视图。本文分为三个部分。第一部分讨论如何使用 E-R 模型和图表描述企业数据视图的技术。这是个参考文献 5 中报告的工作的扩展。第二个部分描述了改变的基本操作企业数据视图。这是一个很少见的领域工作已经完成。本文提出的操作将有助于维护企业模式。这第三部分使用E-R方法分析一个例子巴赫曼给出的关于概念上的变化架构。

**使用实体关系模型和图解技术对真实世界建模**

在本节中，我们将使用示例来展示如何使用实体-关系 (E-R) 模型和图表描述企业数据视图的技术。一个更模型的正式定义见参考文献 5。假定定义和定义的责任维护企业模式属于一个人打电话给企业管理员。以下是企业管理员的建议程序定义企业模式：

(1) 识别企业感兴趣的实体集

实体是一种“事物”，可以明显地确定。根据企业需要，实体可以分为不同的实体类型，例如员工、股东。实体系列是一组相同类型的实体。在 E-R图中，实体集由一个矩形框表示（见图 4）。术语“设置”和“类型”可以在 E-R 图中互换。这读者可以使用其中任何一个来解释 E-R图表。现实世界中有许多“事物”。在此外，不同的企业可能会看到相同的事情不同。选择实体类型是企业管理员的责任最适合他的公司。

(2) 识别企业感兴趣的关系集

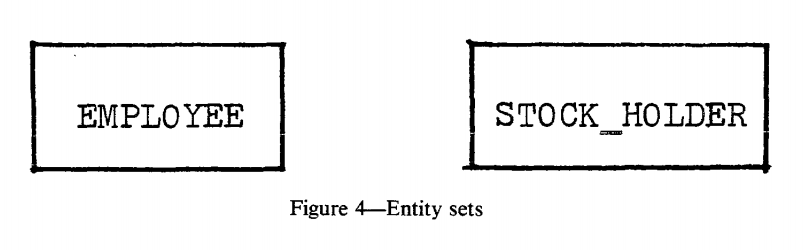
实体是相互关联的。不同类型的不同类型之间可能存在关系实体。关系集是一组关系同一类型。例如，PROJLEMP，它描述了员工对项目的分配，是在两个实体集 EMP 上定义的关系集和项目。关系集也可以定义在两个以上的实体集。例如，PROJ\_SUPP\_PART 是定义在三个实体集 PROJ、SUPP 和 PART。在里面在实体关系图中，关系集由一个菱形框表示，其中的线连接到相关实体集（参见图 5）。这E-R图中与PROJLEMP关系关联的“m”和“n”表示该关系是m:n映射。也就是说，每个员工都可以与多个项目相关联，并且每个项目可能有几个员工。在某些公司，每个员工最多属于一个项目，并且PROJLEMP 关系是 1:n 映射。之间的关系有很多种实体。企业管理员的职责是选择关系集（或类型）企业感兴趣的。他还得指定映射类型（1:1、1:n、m:1 或 m:n）的关系。

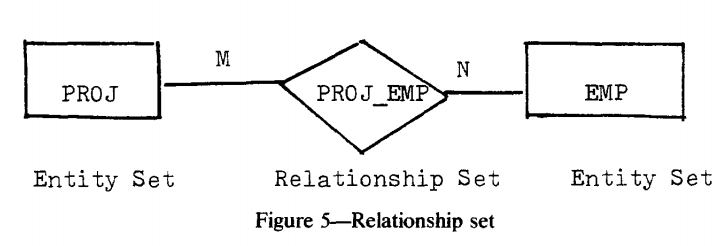
(3) 识别实体和关系的相关属性（即定义值集和属性）

实体和关系具有属性，这些属性可以用属性值对来表示。“Blue”和“4”是值的示例。值可以分为不同的类型，例如颜色或数量。值集是一组值同类型。属性是来自实体的映射集（或关系集）为一个值集（或一组值集）。例如，“地址”是一个属性它将实体集 EMP 中的实体映射到值集 NAMES\_OF\_LOC。请注意，我们放松参考文献 5 中施加的约束，即从实体集到值集的映射必须是函数（即 m: 1 映射）。换句话说，我们现在允许一个属性（例如地址）可以具有同一实体的多个值（例如位置、员工）。属性定义的这种放宽将导致企业观的变化

更简单。这一点在接下来会变得清晰部分。在 E-R 图中，值集由圆圈，属性用箭头表示从实体集（或关系集）指向所需的值集（参见图 6）。选择后实体集和关系集，企业管理员识别属性和值集与公司经营相关的。

上述三个步骤涵盖了主要部分企业架构。为简单起见，我们将不讨论本文与企业架构相关的其他问题比如完整性约束。设计数据库，企业管理员先绘制 E-R 图，如图 7 所示。然后他为每个实体绘制了属性和值集集和关系集。然后翻译 E-R 图放入数据结构图或一组表（“关系”）（见图 2）。使用的规则和程序参考文献 5 中讨论了翻译过程。这里，我们将研究如何改变企业模式（E-R 图）本身。

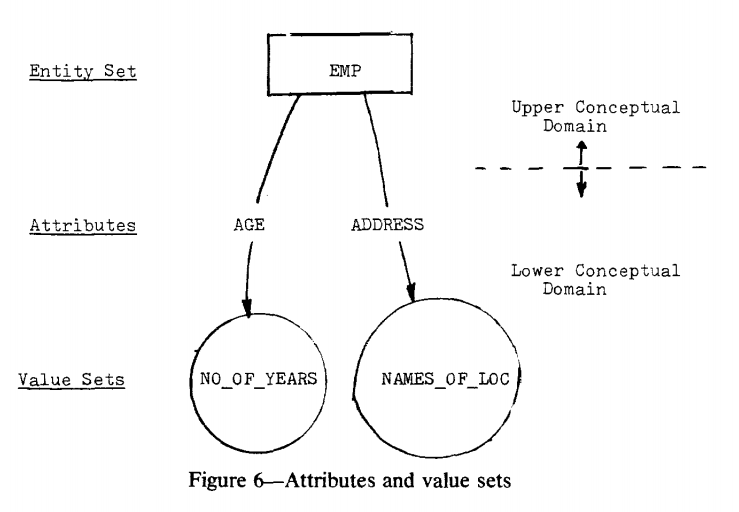


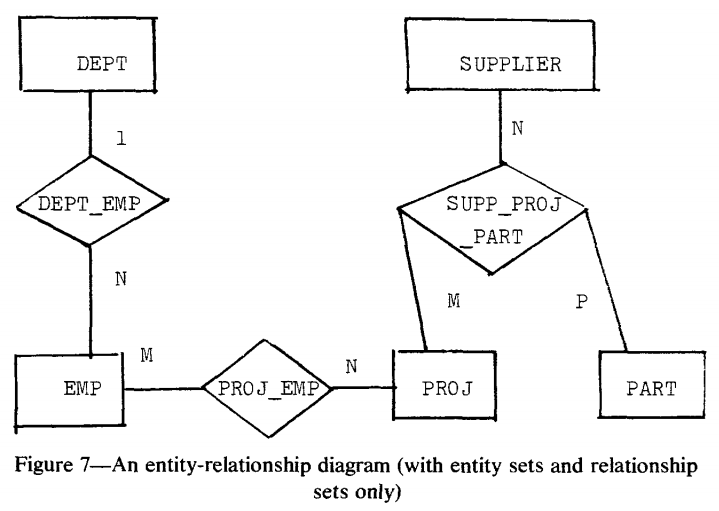


**企业视图的修改**

虽然企业架构比更稳定用户模式，它仍然需要不时更改以反映企业环境的变化。除了巴赫曼的一篇论文，工作很少在这个领域完成。在本文中，我们使用 E-R 模型作为分析企业数据视图中不同类型变化的基础。我们不仅提出了一套操作还要分析这些操作的后果。

有五种基本类型的操作：添加、删除、拆分、合并和转移。前四个操作适用于实体集、关系集、属性和值套。班次操作用于企业时管理员想查看旧版本中设置的值企业模式作为新模式中的实体集或反之亦然。认为 E-R 图很有用由两个概念域组成：（1）由实体集和关系组成的上层概念域套; (2) 下层概念域包括属性和值集。我们将讨论前四个上层和下层概念域的操作。最后，我们将讨论实体集从上概念域到下概念域以及在相反方向移动设置的值。





**上层概念域中的操作**

以下是适用于实体集和关系集的基本操作：

(1) 将一个实体拆分成几个子集

例如，图 8a 中的实体集 EMP 可以是分成两个实体集：图 8b 中的 MALE\_EMP 和 FEMALE\_EMP。这样做的后果操作是关联的关系集实体集也可能需要拆分。例如，PROLEMP 分为 PROJ\_M\_EMP 和PROJLFJEMP（见图 8b）。

(2) 将多个实体集合并为一个实体集

这是（1）的相反操作。结果是相关的关系集可能有被合并。

1. 将关系集拆分成若干个子集

这个操作的一个例子是：关系集图 8a 中的 PROJLEMP 可以拆分为两个关系集，PROJLMANAGER 和 PROJ\_WORKER，在图 8c 中。请注意，类型新关系中的映射可能不同从原来的关系。例如，PROJLMANAGER 中的映射是 1:n 而PROJ\_EMP 中的映射是 m:n。

(4) 将几个关系集合并为一个集合

这是（3）的相反操作。请注意，这些关系集必须定义在相同的一组实体集。

(5) 添加新的实体集

例如，一个名为 SUPPLIER 的新实体集可能被添加到图 8a 中的 E-R 图中。结果如图 9a 所示。请注意，可以有企业模式中的独立实体集，尽管在许多情况下新实体集和现有实体集立即建立（见下一个操作）。

(6) 添加新的关系集

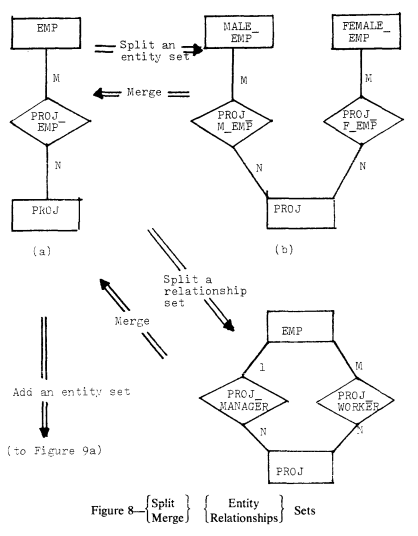
我们可以为新实体添加一个新的关系集集合如图9b中的关系集合PROJLSUPP。我们还可以为现有实体集，例如关系集图 9b 中的 PROJLMANAGER。

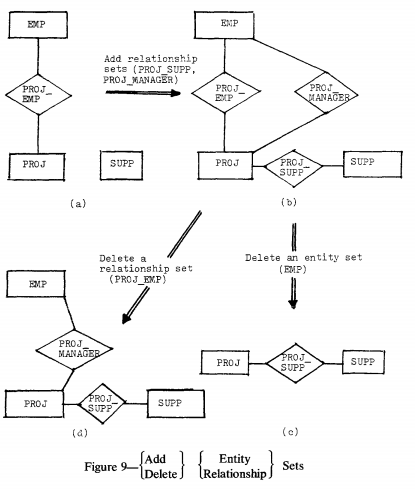
(7) 删除一个实体集

例如，删除实体集 EMP 后图 9b，我们有图 9c。后果是：(i)与实体集相关的关系集是也被删除； (ii) 与删除相关的属性实体集和相关的关系集也被删除。

(8) 删除一个关系集

一个例子是：删除关系集PROJLEMP如图 9b 所示，结果如图 9d 所示。此操作的结果是删除了关系的属性（未显示在图 9d）





**下层概念域中的操作**

假设实体集 EMP 中的实体有两个属性，LEGAL\_NAME 和 PHONE，它们映射实体的值集 NAME 和 PHONE\_#（见图10a)。我们将使用这些属性和值集作为讨论以下操作的基础：

(1) 添加一个值集

例如，一个名为 DOLLARS 的新值集可能被添加到图 10a。结果如图10b。通常，此操作后跟一个“添加属性”操作。

(2) 删除一个值集

删除图 10a 中的值集 PHONE\_# 后，我们得到图 10c。结果就是所有与此值集关联的属性将被删除。

(3) 将一个值集拆分为几个子集

图 10a 中的值集 NAMES 可能是分成两个值集 FIRST\_NAMES 和图 10d 中的 LAST\_NAMES。结果是与值集相关的属性可能必须是调整。虽然属性 LEGAL\_NAME 是图lOd中没有拆分，可以拆分成两个属性：LEGAL\_FIRST\_NAME 和 LEGAL\_LAST\_NAME。这是该部门的责任企业管理员做出此决定。

(4) 将几个值集合并为一个值集

这是（3）的相反操作。

(5) 添加属性

例如，Figure lib 是通过添加将 OTHER\_NAME 归因于图 1 la。

(6) 删除一个属性

从图中删除属性 LEGAL\_NAME11a，我们有图谎言。关联的值集具有该属性的将被另一个操作删除（“删除值集”）如果需要。在某些情况下，值集可能仍与其他属性相关联（见图谎言）。

(7) 将一个属性拆分为多个属性

例如，Figure lid 是通过拆分图 11a 中的 PHONE 属性分为两个属性，OFFICE\_PHONE 和 HOME\_PHONE。

(8) 将多个属性合并为一个属性

这是（7）的相反操作。属性必须在同一实体集（或关系集）上定义。

**两个概念域之间的操作**

假设有两个实体集（EMP 和 PROJ）， 一个关系集（PROJLEMP），四个值集 （NAMES\_OF\_PLACES、SOG\_SEC\_#、PHONE\_# 和PROJLNAMES）和四个属性（ADDRESS，SOC\_SEC\_NO、PHONE 和 NAME）如图12a。我们将把它们作为讨论的基础以下操作：

(1) 将值集从较低的概念域转移到上概念域

当企业环境发生变化时，可能将 PLACE 视为一个实体集变得很自然而不是一个值集。因此，在图 12b 中，“ADDRESS”变成了一个关系集，而“PLACE”有一个指向值的属性“NAME”

设置 NAMES\_OF\_PLACES。由于 PLACE 是一个实体集，我们可以与它建立新的关系其他实体集，例如 PROJ 或添加更多属性和值集来描述“地点”的属性。

(2) 从上层概念域转移一个实体集到较低的概念域

当企业环境再次发生变化时，将 PROJ 视为一个值集可能会变得很自然而不是实体集。在图 12c 中，PROJ 是从上层概念域中删除，并且关系集 PROJLEMP 成为属性INVOLVED\_PROJ。图中的实体集PROJ12b 可能已经与几个值相关联集，但只有值集PROJ\_NAMES 是用于识别实体 PROJ 保留在较低的概念域。

实例分析

在最近的一篇论文中，Bachman10使用数据结构图来说明概念模式的变化。在本节，我们先陈述他的例子，然后用 E-R 图来解释他的例子。

使用数据结构图的示例描述

以下是 Bachman 示例的简化版本：

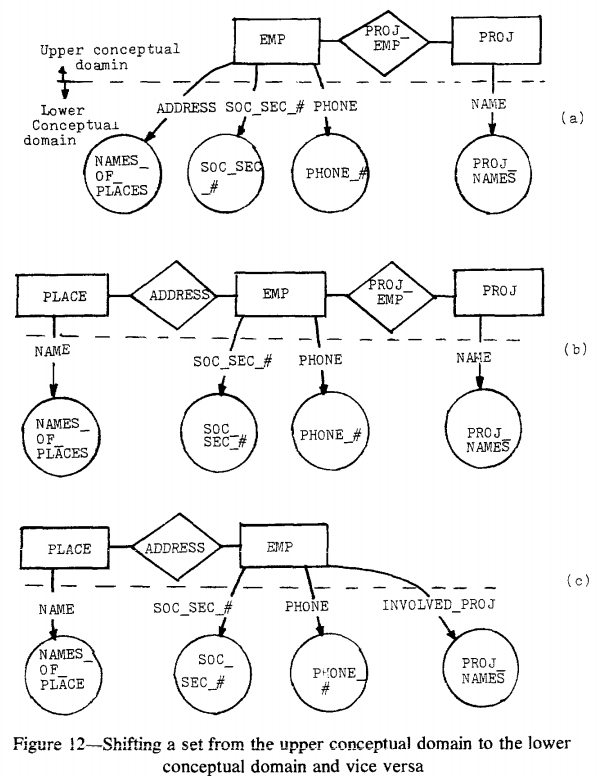
(a) 一开始，企业管理员声明了一个概念模式，如图 13a 所示。假设读者有一定的知识数据结构图1.简单来说，一个长方形的盒子代表一个记录类型，一个箭头表示数据结构集（即记录类型之间的 l:n 关系）。在图 13a 中，有是两种类型的概念记录，COMPANY 和PERSON，以及一个数据结构集“a”，表示每个人都与正好是一家公司，而且每家公司都有一个人员集。

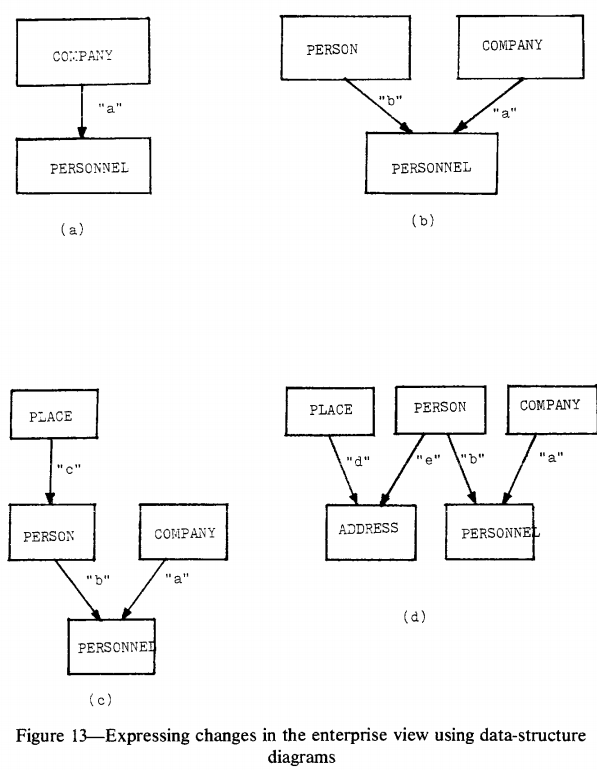
(b) 后来，企业管理员认识到该公司的人员是他们的人自己的权利。这一事实可能在合并时被发现一些人持有的几家公司两个工作，是两个合并的人员公司。图 13b 说明了数据结构新概念模式的图表。基本上，旧的人事类型记录已分成两部分记录类型，PERSONNEL和PERSON。这“PERSON”具有 NAME 和 ADDRESS 属性（图中未显示）。

(c) 过了一会儿，企业管理员决定将该人的居住地址考虑在内记录。图 13c 说明了添加

“PLACE”概念记录类型和数据结构集类型 c。人有一个唯一的地址（地点）。

(d) 现在人们认识到人们从一个地方搬到另一个地方地点，最好知道当前地址以及过去的地址。另一个原因可能是：它发现一个人可能有不止一个地址。在任何一种情况下，一个新的概念记录类型ADDRESS 被添加到概念模式中（参见图 13d)





**使用实体关系图进行分析**

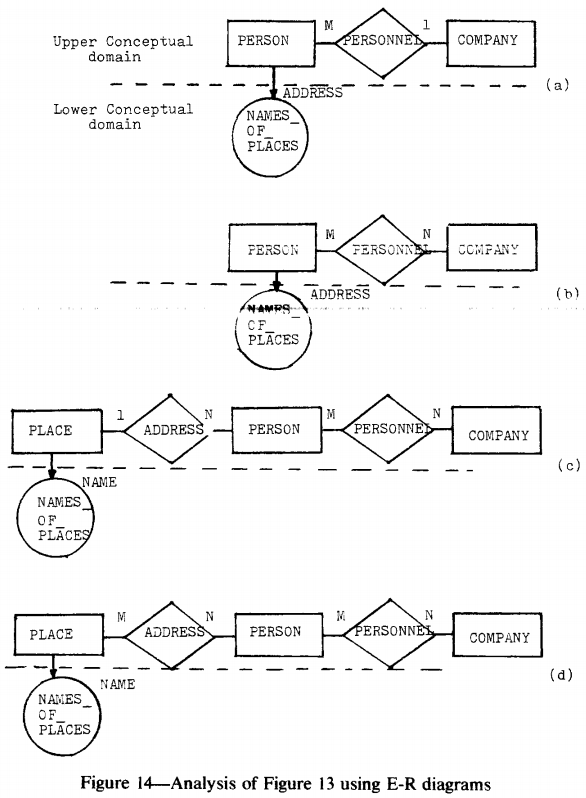
下面，我们将使用 E-R 图来解释上面的例子：上概念领域

(a) 图 14a 中的 E-R 图对应于图 13a 中的数据结构图。有两种类型的实体，PERSON和 COMPANY，在企业观。由于之间的映射COMPANY 和 PERSON 是 1:n，关系集人员由数据结构集表示图 13a 中的“a”。

(b) 图 14b 是对应的 E-R 图图 13b。既然关系集 PERSONNEL是一个 m:n 映射，它由一个关系表示图 13b 中的记录类型 PERSONNEL 和两个数据结构集“a”和“b”。请注意，数字图 14a 和 14b 在上层概念域中具有相同的实体集和关系集，并且区别在于实体之间的映射类型套。

(c) 现在企业管理员更喜欢查看“PLACE”作为实体集而不是值集。因此，我们有图 14c属性地址图 14b 中的关系集变成了图14c。由于 PLACE 和 PERSON 之间的映射是 1:n，关系集 ADDRESS 由图 13c 中的数据结构集“c”表示。

(d) 企业管理员发现PLACE和PERSON之间的映射是m:n映射而不是l:n映射。新企业视图由图 14d 表示。由于映射为m:n，表示关系集ADDRESS由记录类型 ADDRESS 和两个数据结构集“d”和“e”。注意图 14c 和 14d除了映射的类型几乎相同, PLACE 和 PERSON 之间是不同的。



一般来说，E-R图更容易用于分析企业视图的变化比数据结构的变化图表。 Bachman 还提出了图 13d：如果想修改一个人的地址，他是否必须创建一个新的“地址”记录或更改此人居住地的名称？使用 E-R 方法可以很容易地回答这个问题。考虑图 14d。由于 PLACE 是一个实体设置，改变一个人的地址就是改变这个人与“他的地方”的关系。我们不应该更改此人居住地的名称因为“NAME”和“NAMES\_OF\_PLACES”用于描述 PLACE 实体的属性（见图

14d)。

**总结**

企业模式作为中间步骤很有用数据库设计。在本文中，我们展示了如何使用实体关系模型和图解技术来描述企业模式。自企业环境不时变化，企业架构必须更改以反映这些更改。五基本类型的操作（添加、删除、拆分、合并和转移) 这对于修改企业模式很有用已经提出，并讨论了这些操作的后果。最后，我们使用了一个示例分析实体关系方法和网络方法在企业数据视图建模方面的差异。