**数据流图DFD**

　逻辑模型的工具——只反映信息在系统中流动和处理情况的图称为数据流图，它是描述系统逻辑模型的工具之一。数据流图(Data Flow Diagram，简称DFD)是便于用户理解系统数据流程的图形表示。它能精确地在逻辑上描述系统的功能、输入、输出和数据存贮等，而摆脱了其物理内容。数据流图是系统逻辑模型的重要组成部分。

　　系统分析阶段必须进行全面准确的收集、整理、分析收集的数据及其流程。

一、数据收集

　　 数据收集工作量很大， 故要求系统研制人员应具备经营管理的素质，耐心细致地深入实际，配合业务人员收集与系统有关的一切数据。

　　1.数据收集的渠道

　　现行的组织机构；现行系统的业务流程；现行的决策方式；各种报表、报告、图示。

　　2.数据的来源

(1)组织的正式报告(对于手工系统而言)：各种卡片、报表；会议决议。

(2)现行系统的说明性文件(对于已局部计算机化了的系统而言)：各种流程图；计算机文件（数据库）系统的数据组织结构。

(3)组织外的数据来源：上级下达的各种文件和各项任务指标；与本单位密切相关的其它单位的有关信息。

　　3.收集数据的方法

(1)查阅档案：到各个科室按收集数据的类型，查阅档案材料。有时候没有现成的档案，系统分析员就要帮助这些部门建立档案材料。如一个企业的各种报表应该汇编成册，每张报表编上号，注明用途、填报单位、报送单位、月用量、年用量等。如果企业没有做这方面的工作，我们只好自己动手去收集这些报表，编成册，统一标号，调查各种使用情况，作为技术档案资料保存起来，以备日后查阅。

(2)面谈调查：对各级管理人员和工作人员要自上而下地进行访问。调查有关系统总貌、系统目标、环境约束、近年内信息的需求情况，以及他们对现有信息系统的看法(包括有哪些信息是多余的,有哪些或哪方面的信息是急需补充和加强的等等)。

(3)发调查表：对于要作普遍调查的问题，可以发调查表进行调查。

(4)测定：有些数据，如业务的吞吐量、各项工作的时间和费用要实测一段时间。

(5)采样：对于大规模的统计，因不可能收集到数据的全部，可以采用抽样的办法解决。抽样的方式有随机抽样和系统抽样两种，它们的区别在于是不是按一定的规则来抽取样本。样本的大小应根据抽样理论和实际要求来确定。

(6)实际动手：深入实际，亲自动手参加信息的处理工作，这样能加深体会，对我们今后的工作是很有帮助的。

　　4.数据调查内容

输入信息：输入信息名称；使用目的；搜集方式；发生周期；信息量；编码方式；保存期；相关业务；使用文字；其它。

输出信息：输出信息名称；使用单位；使用目的；发行份数；发送方法；使用文字；输出时间；输出方式；其它。

信息处理过程：处理内容；处理周期；处理方法；处理时间；处理场所；其它。

存储方式：文件名称；保管单位；保存时间；总信息量；保密要求；使用频率；删除周期；追加周期；增加、删除比率。

代码信息：代码名称；分类方式；编码方式；使用目的；起始码；终止码；未使用码；贝码率；追加或废弃频率；其它。

信息需求：所需信息名称；需求目的；需求单位；需求者；时间和期限；所需信息的形式；信息表达的要求。

二、数据分析

    收集上来的数据是“ 原材料”，其中有些数据不能用作系统设计的依据，要把这些原材料加工成系统设计可用的资料，就必须做数据的分析工作。数据分析包括以下几个方面：

　　1.围绕系统目标进行分析

    (1)从业务处理角度来看。为了满足正常的信息处理业务，需要哪些信息，哪些信息是冗余的，哪些信息暂缺，有待于进一步收集。

    (2)从管理角度来看。为了满足科学管理的需要，应该分析这些信息的精度如何，能否满足管理的需要；信息的及时性如何，可行的处理区间如何，能否满足对生产过程及时进行处理的需求；对于一些定量化的分析(如预测、控制等)能否提供信息支持等等。

　　2.弄清信息源周围的环境

    对数据进行分析就必须分清，这些信息是从现存组织结构中哪个部门来的，目前用途如何，受周围哪些环境影响较大(如有的信息受具体统计人员的计算方法影响较大；有的信息受检测手段的影响较大；有的受外界条件影响起伏变化较大)，它的上一级(或称层次)信息结构是什么，下一级的信息结构是什么等等。

　　3.围绕现存的业务流程进行分析

　　围绕现存的业务流程进行分析包括：

(1)分析现有报表的数据是否全面，是否满足管理的需要，是否正确反映业务实物流。

(2)分析业务流程，现存的业务流程有哪些弊病，需要做出哪些改进；做出这些改进以后对信息与信息流应该做出什么样的相应改进，对信息的收集、加工、处理有哪些新要求等等。

(3)根据业务流程分析，哪些信息是多余的，哪些是系统内部可以产生的，哪些需要长期保存。

　　4.数据特征分析

　　数据特征分析是下一步设计工作的准备工作。特征分析包括以下几方面的内容：

(1)数据的类型以及长度

　　是数字型还是字符型，是定长的还是变长的， 长度多少(字节数)，以及有何特殊要求(如精度、正负号)等等。

(2)合理的取值范围

　　 这对于将来设计校验和审核功能都是十分必要的。

(3)数据所属业务

　　哪些业务要用到这个数据。

(4)数据业务量

　　每天、每周、每月的业务量 (包括平均数量、最低的可能值、最高的可能值)以及要存储的量有多少，要输入、输出的频率有多大。

(5)数据重要程度和保密程度

　　重要程度即对于检验功能的要求有多高，对后备储存的必要性如何。保密度即是否需要有加密措施，它的读、写、改、看权限如何等等。

三、数据流图（DFD）

1.数据流图的基本符号

　　数据流图由四种基本符号组成，见图5－4－1所示。

|  |
| --- |
| 逻辑模型的工具－数据流图DFD |
|  |

图5－4－1  数据流图的基本符号

　　例：图5－4－2是一个简单的数据流图，它表示数据X从源S流出，经P加工转换成Y，接着经P加工转换为Z，在加工过程中从F中读取数据。

|  |
| --- |
| 逻辑模型的工具－数据流图DFD |
|  |

图5－4－2　数据流图举例

下面来详细讨论各基本符号的使用方法。

2.数据流

　　数据流由一组确定的数据组成。例如“发票”为一个数据流，它由品名、规格、单位、单价、数量等数据组成。数据流用带有名字的具有箭头的线段表示，名字称为数据流名，表示流经的数据，箭头表示流向。数据流可以从加工流向加工，也可以从加工流进、流出文件，还可以从源点流向加工或从加工流向终点。

　　对数据流的表示有以下约定：

对流进或流出文件的数据流不需标注名字，因为文件本身就足以说明数据流。而别的数据流则必须标出名字，名字应能反映数据流的含义。

 数据流不允许同名。

　　两个数据流在结构上相同是允许的，但必须体现人们对数据流的不同理解。例如图5－4－3(a)中的合理领料单与领料单两个数据流，它们的结构相同，但前者增加了合理性这一信息。

　　两个加工之间可以有几股不同的数据流，这是由于它们的用途不同，或它们之间没有联系，或它们的流动时间不同，如图5－4－3(b)所示。

|  |
| --- |
| 逻辑模型的工具－数据流图DFD |
|  |

　　　　　　　　（a）        　　　　　　　　　              （b）                      　　　　　　　　　 （c）

图5－4－3 简单数据流图举例

　　数据流图描述的是数据流而不是控制流。如图5-4-3 (c)中，“月末”只是为了激发加工“计算工资”，是一个控制流而不是数据流，所以应从图中删去。

3.加工处理

　　加工处理是对数据进行的操作，它把流入的数据流转换为流出的数据流。每个加工处理都应取一个名字表示它的含义，并规定一个编号用来标识该加工在层次分解中的位置。名字中必须包含一个动词，例如“计算”、“打印”等。

　　对数据加工转换的方式有两种：

改变数据的结构，例如将数组中各数据重新排序；

产生新的数据，例如对原来的数据总计、求平均等值。

4.文件

　　文件是存贮数据的工具。文件名应与它的内容一致，写在开口长条内。从文件流入或流出数据流时，数据流方向是很重要的。如果是读文件，则数据流的方向应从文件流出，写文件时则相反；如果是又读又写，则数据流是双向的。在修改文件时，虽然必须首先读文件，但其本质是写文件，因此数据流应流向文件，而不是双向。

    例如，在图5-4-3 (a)中，检查合理性加工时，只从库存帐目文件中读出库存信息与领料单核对，所以数据流从文件流出，箭头指向加工。

5.数据源或终点

　　数据源和终点表示数据的外部来源和去处。它通常是系统之外的人员或组织，不受系统控制。

为了避免在数据流图上出现线条交叉，同一个源点、终点或文件均可在不同位置多次出现，这时要在源(终)点符号的右下方画小斜线，或在文件符号左边画竖线，以示重复，如图5－4－4所示。

|  |
| --- |
| 逻辑模型的工具－数据流图DFD |
|  |

图5－4－4 重复的源点、终点或文件

　　由上图可见，数据流图可通过基本符号直观地表示系统的数据流程、加工、存贮等过程。但它不能表达每个数据和加工的具体、详细的含义，这些信息需要在“数据字典”和“加工说明”中表达。

6.DFD的画法

　　一般遵循“由外向里”的原则，即先确定系统的边界或范围，再考虑系统的内部，先画加工的输入和输出，再画加工的内部。即：

　　 (1)识别系统的输入和输出。

　　 (2)从输入端至输出端画数据流和加工，并同时加上文件。

　　 (3)加工的分解“ 由外向里”进行分解。

　　(4)数据流的命名，名字要确切，能反映整体。

　　(5)各种符号布置要合理，分布均匀，尽量避免交叉线。

　　(6)先考虑稳定态，后考虑瞬间态。如系统启动后在正常工作状态，稍后再考虑系统的启动和终止状态。

　　对于不同的问题，数据流图可以有不同的画法。一般情况下，应该遵守“由外向里”的原则。即先确定系统的边界或范围，再考虑系统的内部，先画加工的输入和输出，再画加工内部。具体实行时可按下述步骤进行：

（1）识别系统的输入和输出，画出顶层图

　　即确定系统的边界。在系统分析初期，系统的功能需求等还不很明确，为了防止遗漏，不妨先将范围定得大一些。系统边界确定后，那么越过边界的数据流就是系统的输入或输出，将输入与输出用加工符号连接起来,并加上输入数据来源和输出数据去向就形成了顶层图。

(2)画系统内部的数据流、加工与文件,画出一级细化图

　　从系统输入端到输出端(也可反之)，逐步用数据流和加工连接起来，当数据流的组成或值发生变化时，就在该处画一个“加工”符号。

　　画数据流图时还应同时画上文件，以反映各种数据的存贮处，并表明数据流是流入还是流出文件。

　　最后，再回过头来检查系统的边界，补上遗漏但有用的输入输出数据流，删去那些没被系统使用的数据流。

(3)加工的进一步分解，画出二级细化图

　　同样运用“由外向里”方式对每个加工进行分析，如果在该加工内部还有数据流，则可将该加工分成若干个子加工，并用一些数据流把子加工联接起来，即可画出二级细化图。二级细化图可在一级细化图的基础上画出，也可单独画出该加工的二级细化图，二级细化图也称为该加工的子图。

(4)其它注意事项

　　一般应先给数据流命名，再根据输入/输出数据流名的含义为加工命名。名字含义要确切，要能反映相应的整体。若碰到难以命名的情况，则很可能是分解不恰当造成的。应考虑重新分解。

　　从左至右画数据流图。通常左侧、右侧分别是数据源和终点，中间是一系列加工和文件。正式的数据流图应尽量避免线条交叉，必要时可用重复的数据源、终点和文件符号。此外，数据流图中各种符号布置要合理，分布应均匀。

　　画数据流图是一项艰巨的工作，要做好重画的思想准备，重画是为了消除隐患，有必要不断改进。

　　因为作为顶层加工处理的改变域是确定的，所以改变域的分解是严格的自顶向下分解的。由于目标系统目前还不存在，应此分解时开发人员还需凭经验进行，这是一项创造性的劳动。同时，在建立目标系统数据流图时，还应充分利用本章讲过的各种方法和技术，例如:分解时尽量减少各加工之间的数据流；数据流图中各个成分的命名要恰当；父图与子图间要注意平衡等等。

　　当画出分层数据流图，并为数据流图中各个成分编写词典条目或加工说明后，就获得了目标系统的初步逻辑模型。

7.DFD特性

　　抽象性：在DFD中具体的组织机构、工作场所、物质流等都已经去掉，只剩下信息和数据存储、流动、使用以及加工的情况。故描述的是抽象出来的数据。

　　概括性：它把系统对各种业务的处理过程联系起来考虑，形成一个总体，可反映出数据流之间的概括情况。

 8.DFD用途

自顶而下分析系统的信息流程。

在图上确定需要计算机处理的部分。

向数据库设计过渡。

根据数据流向确定存取方式。

能够对应一个处理过程。

9.DFD优缺点

总体概念强：每层明确“ 干什么”、“ 需要什么”、“给出什么”。

可反映出数据流向的处理过程。

容易及早发现系统各部分逻辑错误。

易与计算机处理对照。

不直观。

人工绘制太麻烦，工作量较大。

四、画分层数据流图时应注意的问题

　　下面从四个方面讨论画分层数据流图时应注意的问题。

1．合理编号

　　分层数据流图的顶层称为0层,称它是第1层的父图，而第1层既是0层图的子图，又是第2层图的父图，依此类推。由于父图中有的加工可能就是功能单元，不能再分解，因此父图拥有的子图数少于或等于父图中的加工个数。

为了便于管理，应按下列规则为数据流图中的加工编号：

l          子图中的编号为父图号和子加工的编号组成。

l          子图的父图号就是父图中相应加工的编号。

　　为简单起见，约定第1层图的父图号为0，编号只写加工编号1、2、3...，下面各层由父图号1、1.1等加上子加工的编号1、2、3...组成。按上述规则，图的编号即能反映出它所属的层次以及它的父图编号的信息，还能反映子加工的处理信息。例如1表示第1层图的1号加工处理，1.1、1.2、1.3...表示父图为1号加工的子加工，1.3.1、1.3.2、1.3.3...表示父图号为1.3加工的子加工。

　　为了方便，对数据流图中的每个加工，可以只标出局部号，但在加工说明中，必须使用完整的编号。例如图5-4-5可表示第1层图的1号加工的子图，编号可以简化成图中的形式。

|  |
| --- |
| 逻辑模型的工具－数据流图DFD |
|  |

图5-4-5  简化子图编号示例

2．注意子图与父图的平衡

　　子图与父图的数据流必须平衡，这是分层数据流的重要性质。这里的平衡指的是子图的输入、输出数据流必须与父图中对应加工的输入、输出数据流相同。但下列两种情况是允许的，一是子图的输入/输出流比父图中相应加工的输入/输出流表达得更细。例如，在图5－4－6中，若父图的“订货单”数据流是由客户、品种、帐号、数量四部分组成，则图中的子图和父图是平衡的。在实际中，检查该类情况的平衡，需借助于数据词典进行。二是考虑平衡时，可以忽略枝节性的数据流。例如图5-4-6，在4号加工的子图中4.3号子加工中增加了一个输出，表示出错的数据流(由虚线所示)，则子图和父图仍可看作是平衡的。

|  |
| --- |
| 逻辑模型的工具－数据流图DFD |
|  |

图5-4-6   子图和父图的平衡图片

3．局部文件

　　图5-4-7中的父图和子图是平衡的，但子图中的文件W并没在父图中出现。这是由于对文件W的读、写完全局限在加工3.3之内，在父图中各个加工之间的界面上不出现，该文件是子图的局部文件或为临时文件。

|  |
| --- |
| 逻辑模型的工具－数据流图DFD |
|  |

                                图5-4-7  数据流图中的局部文件

　　应当指出的是，如果一个临时文件在某层数据流图中的某些加工之间出现，则在该层数据流图中就必须画出这个文件。一旦文件被单独画出，那么也需画出这个文件同其它成分之间的联系。

4．分解的程度

　　对于规模较大的系统的分层数据流图,如果一下子把加工直接分解成基本加工单元,一张图上画出过多的加工将使人难以理解，也增加了分解的复杂度。然而，如果每次分解产生的子加工太少，会使分解层次过多而增加作图的工作量，阅读也不方便。经验表明，一般说来一个加工每次分解量最多不要超过七个为宜。同时，分解时应遵循以下原则：

l          分解应自然,概念上要合理、清晰。

l          上层可分解的快些(即分解成的子加工个数多些)，这是因为上层是综合性描述，对可读性的影响小。而下层应分解得慢些。

l          在不影响可读性的前提下,应适当地多分解成几部分,以减少分解层数。

l          一般说来，当加工可用一页纸明确地表述时，或加工只有单一输入/输出数据流时(出错处理不包括在内)，就应停止对该加工的分解。另外，对数据流图中不再作分解的加工(即功能单元)，必须作出详细的加工说明，并且每个加工说明的编号必须与功能单元的编号一致。

五、数据流图的修改

　　前面介绍了画数据流图的基本方法。对于一个大型系统来说，由于在系统分析初期人们对于问题理解的深度不够，在数据流图上也不可避免地会存在某些缺陷或错误。因此还需要进行修改，才能得到完善的数据流图。这里介绍如何从正确性和可读性方面对数据流图进行改进。

1．正确性

　　数据流图的正确性，可以从以下几个方面来检查：

(1)数据守恒

(2)文件使用

　　在数据流图中，文件与加工之间数据流的方向应按规定认真标注，这样也有利于对文件使用正确性的检查。例如，在图5-4-8中，因为文件1和文件2是子图的局部文件，所以在子图中应画出对文件的全部引用。但子图中文件2好象一个“渗井”，数据只流进不流出，显然是一个错误。

|  |
| --- |
| 逻辑模型的工具－数据流图DFD |
|  |

图5-4-8  局部文件使用错误

（3) 子、父图平衡

　　造成子图与父图不平衡的一个常见原因是在增加或删除一个加工时，忽视了对相应父图或子图的修改。在检查数据流图时应注意这一点。

（4) 加工与数据流的命名

　　加工和数据流的名字必须体现被命名对象的全部内容而不是一部分。对于加工的名字，应检查它的含义与被加工的输入/输出数据流是否匹配。

　　一个加工的输出数据流仅由它的输入数据流确定，这个规则绝不能违背。数据不守恒的错误有两种，一是漏掉某些输入数据流，二是某些输入数据流在加工内部没有被使用。虽然有时后者并不一定是个错误，但也要认真考虑，对于确实无用的数据就应该删去，以简化加工之间的联系。

　　在检查数据流图时，应注意消除控制流。

2．可读性

　　数据流图的可读性，可以从以下几个方面来提高：

(1)简化加工之间的联系

　　各加工之间的数据流越少，各加工的独立性就越高。因此应当尽量减少加工之间数据流的数目，有必要时可采用后面要介绍的步骤对数据流图重新分解。

(2)分解应当均匀

　　在同一张数据流图上，应避免出现某些加工已是功能单元，而另一些加工却还应继续分解好几层的情况出现。否则应考虑重新分解。

(3)命名应当恰当

　　理想的加工名由一个具体的动词和一个具体的宾语(名词)组成。数据流和文件的名字也应具体、明确。

3. 数据流图重新分解的步骤

　　有时需要对作出的部分或全部数据流图作重新分解，可按以下步骤进行：

 (1)把需要重新分解的所有子图连成一张；

 (2)根据各部分之间联系最少的原则，把图划分成几部分；

 (3)重建父图，即把第二步所得的每一部分画成一个圆圈，各部分之间的联系就是加工之间的界面；

 (4)重建各张子图，只需把第二步所得的图，按各自的边界剪开即可；

 (5)为所有加工重新命名、编号。

　　例如，图5-4-9中加工2和其它加工的联系太复杂以致很难独立理解，所以其结构不太合理。因为图5-4-9的子图是由不相连的两部分组成的，所以将它们的父图加工分成两个更为合适。

|  |
| --- |
| 逻辑模型的工具－数据流图DFD |
|  |

  图5-4-9 结构不合理的数据流图

六、数据流图的示例

百货商店业务管理系统顶层数据流程图；

百货商店业务管理系统数据流程图一级分解；

销售处理二级数据流程；

采购处理二级数据流程；

会计处理二级数据流程；

七、数据字典(Data Dictionary,DD)

　　在数据流图的基础上，还需对其中的每个数据流、文件和数据项加以定义， 我们把这些定义所组成的集合称为数据字典(Data Dictionary)。数据流图是系统的大框架，而数据字典以及下面将要介绍的加工说明则是对数据流图中每个成分的精确描述。它们有着密切的联系，必须结合使用。

　　我们把上述每一个定义作为数据字典中的一个条目。因此，在数据字典中有三种类型的条目：数据流条目、文件条目和数据项条目。下面分别讨论。

    1.数据流条目

　　数据流条目对每个数据流进行定义，它通常由四部分组成：数据流名、别名、组成和注释。其中，别名是前面已定义的数据流的同义词；组成栏是定义的主要部分，通常是列出该数据流的各组成数据项；注释栏用于记录其它有关信息，例如该数据流在单位时间中传输的次数等。

　　如果数据流的组成很复杂，则可采用“自顶向下，逐步分解”的方式来表示。例如，“课程”数据流可写成：

     课程＝课程名＋教师＋教材＋课程表

     课程表＝{ 星期几＋第几节＋教室 }

     只要依次查这两个条目，就可确切了解“课程”的含义。

      在数据字典各条目的定义中，常使用下述符号：

     ＝ 表示“等价”；

      ＋ 表示“与”；

     [ | ] 表示“或”，即选括号中某一项，括号中各选择项用“|”隔开。例如，三好学生＝[ 甲|乙|丙|丁 ]；

     ( ) 表示“可选”，即从括号从中任选一项，也可一项都不选；

    {  } 表示“重复”，即重复括号内的项，重复次数的上、下界标在括号右边。例如{X}51 表示把X加工重复1-5次。若在重复括号上没有附加重复次数的上下界时，则表示0次或多次重复。

　　数据流条目的编写格式见表5-4-1、5-4-2“职工基本情况”和“查询条件”数据流条目。