

基于多视图的信息系统体系结构描述方法研究

罗爱民 罗雪山 黄力

(国防科技大学信息系统与管理学院 长沙 410073)

摘要 基于多视图的体系结构描述方法是目前流行的体系结构描述方法。本文针对信息系统的特点,定义了体系结构、体系结构描述等概念以及它们之间的关系,分析基于多视图体系结构描述的概念模型,并提出视图选择方法和多视图体系结构描述框架建立的方法。该方法为建立多视图的体系结构描述规范提供指导。

关键词 体系结构,体系结构描述,视图

Research on Information System Architecture Description Method Based on Multi-Views

LUO Ai-Min LUO Xue-Shan HUANG Li

(Information System and Management Academy National University of Defense Technology, Changsha 410073)

Abstract Information system architecture description method based on multi-views is popular method. Aiming at characteristic of information system, architecture, architecture description, and relationship of them are defined. The concept model of architecture description based multi-views is analyzed, the methods are provided, which are used to build multi-views and architecture description framework. The methods offer guidance to build multi-views description criterion of information system architecture.

Keywords Architecture, Architecture description, View

信息系统,特别是军事电子信息系统为满足应用的需求,系统的功能越来越多,规模越来越大,结构越来越复杂,这给体系结构设计和描述带来越来越大的困难。采用简单的模型很难将系统的组成、结构以及相互关系等体系结构内容描述清楚,对于这样复杂的体系结构怎样描述呢?信息系统建设与开发是一项复杂的过程,建设过程中涉及到各种风险承担者。在建设过程中,各种风险承担者必须详细设计框架模型。他们根据自己的需要和所承担的工作和职责,对同一个系统进行不同的描述。由于没有哪一个描述能够清楚说明系统的所有内容,因此,必须由不同的人员根据从不同的角度来描述系统,即采用多视图的描述方法。目前出现了一些基于多视图的信息系统体系结构描述框架,如 Zachman 框架^[3,4]、“4+1”框架^[6]、RM-ODP 模型^[7]、C4ISR 体系结构框架^[1]、DOD 体系结构框架^[5]、联邦企业体系结构框架^[8]等,这些体系结构框架都是采用基于多视图的思想,建立体系结构框架,这些体系结构框架或模型都是根据不同的背景和需求建立的,在各自的领域中发挥重要的作用。本文在研究这些体系结构框架的基础上,分析基于多视图体系结构描述的概念模型,并提出多视图体系结构描述框架的建立方法。

1 基本概念

定义 1 体系结构^[1]定义为“组成部分的结构、它们之间的关系,以及指导它们设计和随时间演化的原则和指南。”

定义 2 体系结构描述是用来描述整个系统结构和行为的模型,它表现系统怎样完成在背景环境中的需求;它确定系统主要组件,它们相互连接和依赖关系以及它们运行中的限制。

体系结构和体系结构描述是两个不同的概念,体系结构是一个抽象概念,体系结构描述则是一个具体的事务,是对体系结构的表现。

体系结构的概念来源于建筑学,相当于建筑学的蓝图设计。在建筑设计上,建筑学的蓝图可以从主体结构、供水管路和供电管路等三个方面(即三个视图)进行设计,形成主体结构图、供水管路图、供电管路图。这三种设计视图的结合可以完整地描述出一个建筑的概貌,若只用其中的任何一个或两个设计视图都不能完整地描述一个建筑的概貌。借鉴建筑学蓝图设计的思想,信息系统体系结构描述也可以采用多视图的方法。

定义 3 基于多视图的体系结构描述是指根据系统风险承担者的关注内容,从各种不同的视角描述信息系统体系结构,构成多视图的体系结构的描述,以形成对体系结构全面、整体的描述。

2 多视图体系结构描述概念模型

IEEE 在综合已有的体系结构描述实践工作基础上,制定了软件密集系统的体系结构描述标准,即 IEEE 标准 P1471-2000^[5]。IEEE 标准 P1471-2000 给出了体系结构描述常用概念、术语的定义,通过建立体系结构描述的概念模型说明了这些概念和术语的相互关系,并提出对体系结构描述的基本要求。多视图体系结构描述的概念模型如图 1 所示^[5]。

按照体系结构描述概念模型,系统存在于一定环境中,系统存在的环境对这个系统产生影响。环境确定开发、运行、政治和其他影响系统因素的背景,环境中包括与该系统发生直

罗爱民 副教授,研究方向:C4ISR 基础理论,C4ISR 体系结构等。罗雪山 教授,博导,研究方向:C4ISR 基础理论,C4ISR 体系结构,建模与仿真等。

接或非直接交互的其他系统,环境确定系统的范围和边界。系统在一定环境中完成一个或多个使命。

一个系统只有一个体系结构,但系统的体系结构可以有一个或多个体系结构描述。体系结构描述必须在有关的理论和方法指导下进行。体系结构描述包括一个或多个视图。每个视图表示系统风险承担者的一个或多个关注点。一种体系结构描述可以选择一个或多个视角,视角的选择应以体系结构描述支持的风险承担者及其关注点为依据。不同的风险承

担者对系统的关注点有相同的地方,也有不同的地方。一个风险承担者可以有一个或多个关注点,多个风险承担者可能都对某一个关注点感兴趣。模型是系统一些方面的抽象表示,模型通过符号、定义和图描述。体系结构描述是模型的集合。一个体系结构视图包括一个或多个模型。体系结构模型按照视角定义的方法建立,一个体系结构模型可以用于描述一个或多个视图。该概念模型较好地规范了基于多视图体系结构描述的含义。

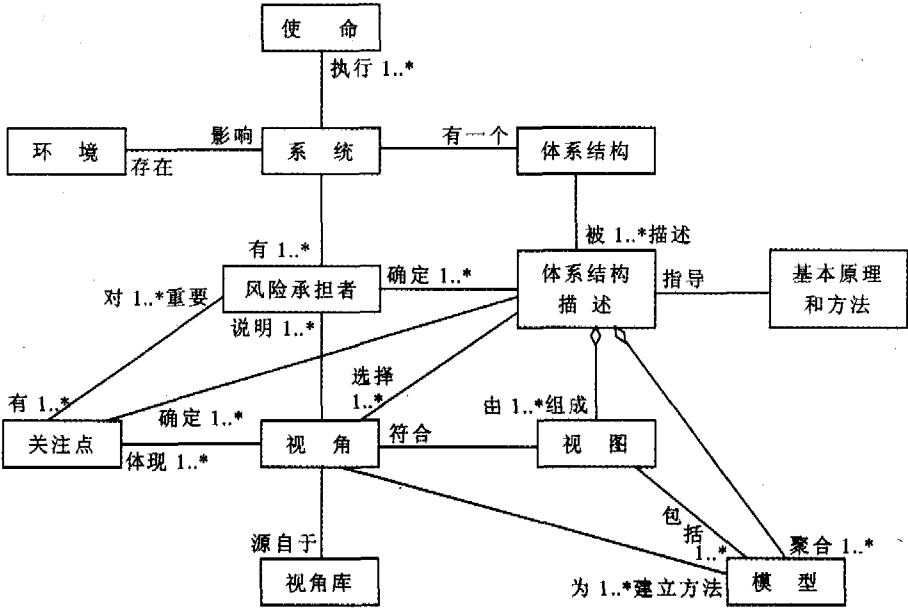


图1 体系结构描述的概念模型

尽管 Zachman 框架在概念模型之前出现,但 Zachman 框架是最典型的基于多视图的信息系统体系结构框架。在 Zachman 框架中,包括了 6 种不同的视角和 6 种不同的方面,Zachman 框架的矩阵表示见表 1,其中每一行(层)表示一个视角,每一列反映关注的一个方面。第 1 层是范围层,代表计划者视角,是对系统的概要描述,为计划者或投资者提供了一个执行摘要,说明系统的范围、开发费用和能够实现的业务目标。第 2 层是企业模型层,代表所有者视角,是对系统的业务设计,主要描述业务实体和业务过程。第 3 层是系统模型层,

代表设计者视角,描述表示业务实体的数据元素和实现业务过程需要具备的系统功能。第 4 层是技术模型层,代表实现者视角,给出系统的技术实现方案,描述系统的组成、结构,规定部件之间的接口。第 5 层是部件模型层,代表子承包商视角,描述系统部件的具体实现细节,而不关心系统的背景和整体结构。第 6 层是代表使用者视角,对应于建筑领域或产品制造领域(系统)而言,使用者层是从最终产品的使用者的角度出发对系统体系结构进行概要描述。

表 1 Zachman 框架模型

	数据/实体 WHAT	功能 HOW	网络 WHERE	人员 WHO	时间 WHEN	目标 WHY
范围(上下文) 计划者视角	业务对象列表	业务过程列表	业务场所列表	重要组织列表	重要事件列表	业务目标列表
企业模型(概念的) 所有者视角	语义模型	业务过程模型	业务分布模型	workflow 模型	主进度表	业务计划
系统模型(逻辑的) 设计者视角	逻辑数据模型	应用体系结构	分布式系统体系结构	人机交互体系结构	处理结构	业务规则模型
技术模型(物理的) 建造者视角	物理数据模型	系统设计	技术体系结构	表现体系结构	控制结构	规则设计
部件模型 (脱离上下文的) 子承包商视角	数据定义	程序	网络体系结构	安全体系结构	时序定义	规则说明
机能企业产品 (使用者视图)	数据	功能	网络	人员	时间	动机

系统体系结构框架的设计内容包括多个方面,如数据、功能、结构等,这些内容组成每个视角完整的结构。Zachman 框

架主要对数据、功能、结构、人员、时间和目标等六部分内容进行描述,这些内容从各个不同的侧面反映信息系统体系结构

的设计内容。在 Zachman 框架模型中,第 1 列数据视图(What 视图),主要描述系统涉及到的实体以及实体之间的关系;第 2 列功能视图(How 视图),主要描述系统要完成怎样的功能;第 3 列是网络视图(Where 视图),主要描述系统的分布和连接,系统各项功能在哪里被执行。第 4 列是人员视图(Who 视图),主要描述系统中由谁来完成各个功能和任务,明确系统中的组织关系、责任关系和功能分配关系;第 5 列是时间视图(When 视图),主要描述系统中各事件的时间关系。第 6 列是目标视图(Why 视图),主要描述为什么要完成这些活动,表明企业的目标。

3 体系结构描述框架构建方法

目前出现多种基于多视图的体系结构描述框架,这些框架虽然都是基于多视图的体系结构描述方法,但是各自采用的视图存在差别,如 C4ISR 体系结构框架采用作战视图、系统视图和技术视图,“4+1”框架采用逻辑视图、过程视图、开发视图和物理视图。这些体系结构框架是针对一定的背景和需求提出的,它们在特定的体系结构开发和领域是成功的。如 C4ISR 体系结构框架主要用于 C4ISR 领域,是以军事作战为目标的方法,适用于军事背景。如果 C4ISR 框架用于网络通信系统和软件设计就会存在问题,那么,在具体使用中,应该选择和采用哪一个框架呢?

由于在实际中,体系结构设计和描述面临不同领域的不同问题,采用一个通用的框架来解决所有问题是不切实际的。不同的开发背景,对体系结构的开发内容是不一样的,如简单独立系统体系结构开发和 SOS(system of system)系统集成体系结构开发是不同的。简单独立系统的体系结构涉及的主要内容是组成、数据和功能等,而在 SOS 系统体系结构开发中,还必须考虑互操作性、公共操作环境(COE)、标准以及演化,技术体制等。因此,必须采用不同的体系结构框架,选用不同的体系结构视图。

在实际应用中,应该以体系结构描述概念模型为基础,按照概念模型定义的概念和关系,根据实际背景和需要选择视图,建立体系结构框架,以满足体系结构设计和描述的需要。前面提到的这些框架都是结合各自的应用背景,选择不同的视图建立各自的体系结构描述规范,并在各自领域内都发挥很好的作用。

选择视图最基本的原则就是考虑不同的开发背景和问题。一个成功的体系结构描述要求对体系结构的应用领域和怎样使用有一个清楚的了解。

选择视图的基本步骤:

(1) 问题和背景分析

首先分析体系结构对应系统的情况,体系结构设计的目的和对应的背景是什么。根据体系结构设计的目的和背景,明确体系结构的功能和作用,建立与项目相关的所有的风险承担者集合 $S = \{s_1, s_2, s_3, \dots, s_m\}$ 。由于风险承担者在体系结构对应系统中的作用和地位的不同,他们对体系结构开发影响的作用也不同。为描述不同风险承担者对体系结构影响程度,建立风险承担者的重要等级 $P = \{p_1, p_2, p_3, \dots, p_m\}$ 。

(2) 风险承担者关注点提取

根据风险承担者集合,建立风险承担者的关注点集合。风险承担者的关注点与体系结构的背景密切相关,同样的风险承担者针对不同的背景和问题,他们关注的问题也有所不同。

第 i 个风险承担者 s_i 的关注点集合为 $C_i = \{c_{i1}, c_{i2}, c_{i3}, \dots, c_{in_i}\}$, $(1 \leq i \leq m)$, 则系统的关注点集合为 $C = \bigcup_{i=1}^m C_i = \{c_{1,1}, c_{1,2}, \dots, c_{m,n_i}\}$ 。

(3) 产生视图集

体系结构视图分为单视图和组合视图。多个关注点或多视角关注内容的组合构成的视图称为组合视图。如果在一定背景下,单纯按照体系结构对应的风险承担者以及他们关注点建立视图,容易产生一些不切实际的视图,而且常常使得体系结构视图较多,提高了设计过程的复杂度。因此,在建立视图集时,必须对视图进行优化选择或合并视图。

建立视图集可以采取以下两种方法。

• 以风险承担者的视角为核心,建立视图集

该方法的基本原则是不同的风险承担者对应不同的视图。首先按照风险承担者的视角建立候选视图;再分析不同风险承担者关注点之间的关系,如一个风险承担者的关注点被另一个风险承担者所包含,则合并这两个候选视图,以重要等级高的风险承担者的关注点的基准。风险承担者集合 $S = \{s_1, s_2, s_3, \dots, s_m\}$, 建立候选视图集 $V_c = \{v_1, v_2, \dots, v_m\}$ 。

若 $C_i \subseteq C_j$, $p_i \leq p_j$, 其中

$$1 \leq i, j \leq m, i \neq j, V = V_c - \{v_i\}$$

这种建立视图集的方法简单。在体系结构具体描述时,设计人员与不同的风险承担者在交流和理解上较容易。但这种视图框架逻辑性不强,视图中包含的内容较多,并且不同视图之间的关系也复杂。C4ISR 体系结构框架和 DOD 体系结构框架就是采用这种方法。

• 以关注点为核心,建立视图集

首先按照关注点,建立候选视图集;其次,分析不同关注点的特点,若两个关注点的内容具有很强的相关性,为降低设计复杂度,可以将两个关注点合并。根据合并后的关注点集,建立视图集。

设关注点集合为 $C = \bigcup_{i=1}^m C_i = \{c_{1,1}, c_{1,2}, \dots, c_{m,n_i}\}$, 若 c_i, c_j 强相关性,即 $c_i \rightarrow c_j$, 则合并 c_i, c_j , 形成新的视图 c'_i 。

若一个视图与另一个视图是等价的,即 $c_i \Leftrightarrow c_j$, 则合并这两个视图,形成新的视图 c'_i 。

这种方法得到的视图集逻辑性强,视图之间的关系简单而明确,但在建立和使用体系结构过程中,由于风险承担者关注的内容分散在不同的视图中,风险承担者交流和使用不方便。

通过上述方式可以减少视图的数量,同时提高视图的科学性和合理性。在具体应用中,设计人员可以根据实际情况选用某种方法建立视图集。

(4) 定义视图中模型体系

按照概念模型,视图是通过一组模型来体现的。这些模型是对各关注点内容的具体描述,因此,在视图集的基础上,根据视图所覆盖的关注点的内容,定义视图中的相关模型。例如, C4ISR 体系结构框架中,在作战视图的基础上,还必须明确关注点的描述模型,即构建相应的体系结构产品。

(5) 分析视图之间的关系,确定该视图集在实施中的可行性和步骤

在建立视图集的基础上,为保证视图集的设计内容符合一致性和逻辑性,必须对视图中的各模型内容以及相关进行分析,确定在体系结构设计的开发步骤和顺序。

(下转第 128 页)

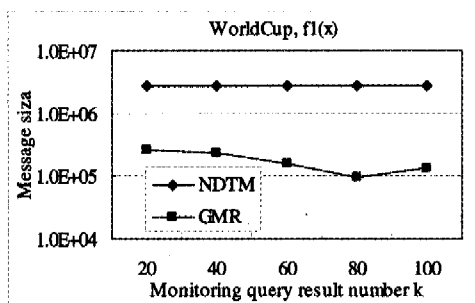


图3 $f_1(x)$ 在不同的结果集大小 k 上的性能比较

6 相关工作

最为著名的通用的 top-k 查询技术是由文[13, 14, 20]分别独立提出的阈值算法(Threshold Algorithm, 简称 TA),它是一种通用的适用于单一数据源的高效的 top-k 查询算法,但不适用于分布式系统。在 TA 的基础上,研究人员提出了许多面向分布式系统的 top-k 查询处理技术^[2, 3, 8, 10, 19],但这些算法都是快照式的,即一次完成,不能支持数据流上的连续查询。

文[4, 5]讨论了如何在分布环境下实现有效的数据流的 top-k 在线监测,通过在远端数据源建立被监测对象的约束关系,保证全局 top-k 集合也是各监测节点上的 top-k 集合,只有在约束被打破时才进行节点间通讯,这样有效地减少了 top-k 在线监测的网络通讯量。但文[4, 5]中的方法在每次约束被打破时都需要传输整个 top-k 集合。

在文[11]中,我们提出一种更为有效的分布式 top-k 算法 MR。MR 在约束被打破时只需要传输打破约束的对象集。因此,和文[4, 5]中的方法相比,MR 显著地减少了监测过程中的网络通讯量,并且 MR 的通讯代价独立于 k 值。但是,这两种方法假定排序函数为加法,都只是针对加法来建立和维护远端数据源的约束关系。而在实际应用中,用户给定的排序函数可能是任意的排序函数。本文提出的 GMR 算法可以支持任意的连续的严格单调聚集函数的分布式 top-k 监测,同时通讯代价和 k 无关。MR 算法可以看作是 GMR 算法的特例。

文[21]讨论了传感器网络中如何监测传感值最高的 k 个传感器的问题,并没有涉及聚集函数下的 top-k 监测。文[17]讨论了单个数据流上的多个聚集函数的 top-k 监测问题,着眼于如何减少单个节点上的 CPU 和内存使用,并没有考虑分布环境下如何减少网络通讯量。

总结 本文研究了如何实现通用并且高效的分布式 top-

k 监测,即在分布的多数据流中根据用户给定的排序函数连续监测最大的 k 个值。在实际应用中,用户给定的排序函数可能是任意的排序函数,然而,目前的分布式 top-k 监测技术只支持加法作为排序函数。在本文中,我们提出了一种通用的支持任意连续的严格单调的聚集函数的分布式 top-k 监测算法 GMR。GMR 通过对远程数据流建立约束关系来维护 top-k 结果集。GMR 只有在约束被打破时才需要节点间通讯,并且通讯时只需要传输少量的对象,通讯代价和 k 无关。通过真实世界数据和模拟数据验证了 GMR 的效率。实验表明,GMR 的网络通讯量比同类算法低一个数量级以上。如何在大规模分布环境下的实现通用的多 top-k 监测是我们下一步研究的目标。

参考文献

- 1 Arlitt M, Jin T. 1998 world cup web site access logs. August 1998. Available at: <http://www.acm.org/sigcomm/ITA/>
- 2 Bruno N, Gravano L, Marian A. Evaluating top-k queries over web-accessible databases. In: ICDE, 2002
- 3 Balke W-T, Nejdl W, Siberski W, et al. Progressive Distributed Top-k Retrieval in Peer-to-Peer Networks. In: ICDE, 2005
- 4 Babcock B, Olston C. Distributed Top-K Monitoring. In: SIGMOD, 2003
- 5 Babcock B, Olston C. Distributed top-k monitoring. [Technical report]. Stanford University Computer Science Department, 2002. <http://dbpubs.stanford.edu/pub/2002-61>
- 6 Carney D, Cetintemel U, Cherniack M, et al. Monitoring streams—a new class of data management applications. In: VLDB, 2002
- 7 Chen J, DeWitt D J, Tian F, et al. NiagaraCQ: A scalable continuous query system for internet databases. In: SIGMOD, 2000
- 8 Chang K C-C, Hwang S-W. Minimal probing: supporting expensive predicates for top-k queries. In: SIGMOD, 2002
- 9 Carey M J, Kossmann D. On saying “Enough already!” in SQL. In: SIGMOD, 1997
- 10 Cao P, Wang Z. Efficient top-k query calculation in distributed networks. In: PODC, 2004
- 11 Deng B, Jia Y, Yang S Q. Supporting Efficient Distributed Top-k Monitoring. To appear in WAIM, 2006
- 12 Fagin R. Combining fuzzy information from multiple systems. J. Comput. System Sci., 1999, 58: 83~99
- 13 Fagin R, Lotem A, Naor M. Optimal aggregation algorithms for middleware. In: PODS, 2001
- 14 Guntzer U, Balke W-T, Kiefling W. Optimizing multi-feature queries for image databases. In: VLDB, 2000
- 15 Gibbons P B, Matias Y. New sampling-based summary statistics for improving approximate query answers. In: SIGMOD, 1998
- 16 何盈捷, 王 珊, 杜小勇. 纯 Peer to Peer 环境下有效的 Top-k 查询. 软件学报, 2005, 16(4): 540~552
- 17 Mouratidis K, Bakiras S, Papadias D. Continuous Monitoring of Top-k Queries over Sliding Windows. To appear in SIGMOD, 2006
- 18 Madden S, Hellerstein J M, Shah M, et al. Continuously adaptive continuous queries over streams. In: SIGMOD, 2002
- 19 Michel S, Triantafyllou P, Weikum G. KLEE: A Framework for Distributed Top-k Query Algorithms. In: VLDB, 2005
- 20 Nepal S, Ramakrishna M V. Query processing issues in image (mul-timedia) databases. In: ICDE, 1999
- 21 Wu M, Xu J, Tang X, et al. Monitoring Top-k Query in Wireless Sensor Networks. In: ICDE, 2006
- 22 Zipf G K. Human Behavior and the Principle of Least Effort. Addison-Wesley Press, 1949

(上接第 121 页)

结束语 基于多视图的信息系统体系结构描述方法是设计信息系统体系结构的一种有效的方法,能够规范体系结构设计,促进体系结构的交流和重用,对提高信息系统设计质量有较强的支持作用。本文深入研究了基于多视图的体系结构描述方法,为设计人员根据不同的背景和需求,构建适合系统开发的体系结构框架提供指导。

参考文献

- 1 C4ISR Architecture Working Group. C4ISR Architecture Framework Version 2.0 [R]. U. S., Department of Defense, 1997
- 2 IEEE Architecture Working Group. IEEE Recommended Practice

- for Architectural Description. IEEE Std 1471, Draft Version 3.0, USA, 1998
- 3 Zachman J A. A Framework for Information Systems Architecture [J]. IBM Systems Journal, 1987, 26(3): 276~292
- 4 Sowa J F, Zachman J A. Extending and Formalizing the Framework for Information Systems Architecture [J]. IBM Systems Journal, 1992, 31(3): 590~616
- 5 DoD Architecture Framework Working Group. DoD Architecture Framework Version 1.0 [R]. U. S. Department of Defense, 2003
- 6 Kruchten P B. Architectural Blueprints—The “4+1” View Model of Software Architecture [J]. IEEE Software, 1995, 12(6): 42~50
- 7 张维明. 信息系统集成技术[M]. 北京: 电子工业出版社, 2002
- 8 Chief Information Officers (CIO) Council. Federal Enterprise Architecture Framework, v 1 [R/OL]. U. S. CIO Council, 1999.
9. <http://www.itpolicy.gsa.gov/mke/archplus/projectmodel-subjects/main-theciocouncilfeaver11.html>