

采用雪花模型的电子图书馆数据仓库的分析和设计

Made Dwi Mulyawan, Ida Bagus Alit Swamardika, and Komang Oka Saputra
乌达雅纳大学电气和计算机工程系 印度尼西亚 巴厘岛 80361

摘要: 信息系统已经成为帮助人类工作的一个重要需求, 电子图书馆就是其中之一。被图书馆员使用这个系统来执行图书馆的日常业务流程, 如借书和还书, 管理藏书, 采购最新的藏书, 以更新图书馆书籍。随着时间的推移, 存储在数据库中的数据会大量增加, 所以分析需要一定周期数据的信息时, 会花费相当长的时间。在这种情况下, 需要使用数据仓库来帮助图书馆员提供所需的信息。本研究将按照四步维度建模中的设计阶段, 利用雪花模型提出一个数据仓库设计, 作为在电子图书馆中提供所需的数据仓库设计的解决方案之一。

关键词: 数据仓库; 雪花模型; 四步维度建模; OLTP; 信息系统

Analysis and Design of Data Warehouse for E-Library Using Snowflake Schema

Made Dwi Mulyawan, Ida Bagus Alit Swamardika, and Komang Oka Saputra
Department of Electrical and Computer Engineering, Udayana University, Bali 80361, Indonesia

Abstract: Information systems have become an important requirement in helping human work, one of which is e-library. This system is used by librarians in carrying out daily business processes in the library, such as borrowing and returning books, managing book collections, and procuring the latest book collections so that the library has updated books. Over time, the data stored in the database will increase a lot, so when analyzing information that requires data with a certain period will take quite a long time. So, in this case, the use of a data warehouse is needed to help the librarian in providing the information needed. This research will present a data warehouse design using snowflake schema by following the design stages in Fourth Steps Dimensional Modeling, as one of the solutions in providing the required data warehouse design in the e-library.

Keywords: Data Warehouse; Snowflake Schema; Fourth Steps Dimensional Modelling; OLTP; Information System

1. 引言

当前, 信息技术发展迅速, 开发一个能够简化人类工作的系统变得可能。信息系统可以对一组数据进行存储, 随后可以产生信息。这些信息以后可以作为公司业务流程中的依据。此外, 信息系统也是改变整个系统或工作流程的创新之一, 这些工作流程以传统方式完成或运行, 并与信息系统相结合, 变得更加容易。

在促进管理过程中使用的一个信息系统, 是电子图书馆信息系统。电子图书馆是一个用于执行绩效管理或交易过程的系统, 涉及到借书和还书、分发书籍、监测正在借阅的书籍, 以及对图书馆内的书籍进行盘点。这样, 信息系统就成为图书馆处理所需信息的重要部分。

随着时间的推移, 存储在电子图书馆的数据会越来越多, 所以分析需要一定周期数据的信息时, 将需要相当长的时间。因此, 需要使用数据仓库来帮助图书馆员分析所需的信息。数据仓库被设计成能够快速进行查询^[1]。通过使用数据仓库, 你可以看到模式或趋势, 积累起来进行总结, 向下挖掘详细信息, 生成报告^[2], 如图书馆材料收集报告, 图书借阅报告, 图书馆访问数据收集, 以方便图书馆员快速准确地做出决策。

基于这些问题, 有必要在电子图书馆中设计一个数据仓库, 可用来对电子图书馆中存在的数据库交易进行结构化和详细的处理过程分析。电子图书馆的数据仓库设计是用雪花模型完成的。雪花模型是星型模型的发展, 它应用了维度表的规范化方法, 以简化和加快执行的查

询过程^[3]。

2. 文献回顾

2.1 相关研究

在设计这个数据仓库的研究中，一些文献被用作参考或依据，如下：

1) Narendra 等人^[4]研究进行了数据仓库设计，以应对巴厘岛省的支出报告和评估赠款人。该设计使用 Kimball 的四步骤方法进行，设计一个紧凑、动态的数据仓库，并支持创建赠款资金报告和赠款资金复盘。

2) 在 Budiarta 等人^[5]的研究中，进行了数据仓库的设计，以方便获得信息来支持认证标准，并作为获得结果的参考。数据仓库的设计是使用星型模型，把业务的维度改变成逻辑数据模型。

3) 由 Gede 等人^[6]进行的研究是一个数据仓库，可以作为蜡染公司管理的信息来源。该设计使用雪花模型和 MOLAP（多维在线分析处理）进行。通过使用雪花模型，规范化的表格可以减少数据的冗余，并根据需要更新数据分析。

4) Putra Wijaya 等人^[7]在登巴萨市政府使用的在线社区投诉系统中实现了一个数据仓库。数据仓库的设计采用了雪花模型，旨在最大限度地减少多余的数据。该研究中的数据仓库被用来简化按类别、平台和后续行动分析投诉数量的过程。

5) Prabawa 等人^[8]使用雪花模型进行了数据仓库设计，以优化电子旅游业务。使用雪花模型是企业预测业务趋势、保持质量、提高竞争力和长期生存的一个解决方案。

基于在实施和设计数据仓库方面所做的研究，本研究将通过四步维度建模，使用雪花模型在电子图书馆中设计数据仓库。

2.2 数据仓库

数据仓库是一个数据的集合，它具有面向主题、综合、时变性、固定的数据收集性质，以支持管理决策过程^[9]。该数据库更多的是为查询和分析交易过程而设计，通常包含历史交易数据和可能来自其他来源的数据^[10, 18]。数据仓库将分析工作负载与交易工作负载分开，允许组织结合或合并来自不同来源的数据^[14, 16]。

数据仓库通常存储历史数据，其功能是作为一个全面的数据库，支持一个组织所需的所有决策分析，提供详细信息和总结^[11]。数据仓库有4个特点：

1) 面向主题

数据仓库的设计是为了满足基于特定主题的数据分

析师的需求。

2) 综合

数据仓库必须能够整合不同的数据源作为不同的操作系统，所以在不同来源的数据被存储在数据仓库之前，必须消除不一致的数据。

3) 非易失性

已经输入数据仓库的交易数据将永远不会被更新或删除。存储在数据仓库中的数据不是为了日常活动数据，而是在一定的时间间隔内将数据输入到数据仓库中，这些数据是根据业务需要而制定的。

4) 时变性

历史数据存储（历史数据）是数据仓库中非常重要的一种类型或一块数据。在数据仓库中，数据往往会被多次存储，如交易发生时，交易时间的变化，或交易被取消时。如果数据仓库是有价值的，则被认为是好的数据仓库^[20]。

2.3 雪花模型

雪花模型是一个多维数据库中的逻辑表结构。它的设计使实体-联系图（ER图）类似于雪花的形状^[1]。雪花模型增加了额外的维度，是星型模型的发展。维度表是规范化的，它将数据划分为额外的表。雪花模型的结构如图1所示。

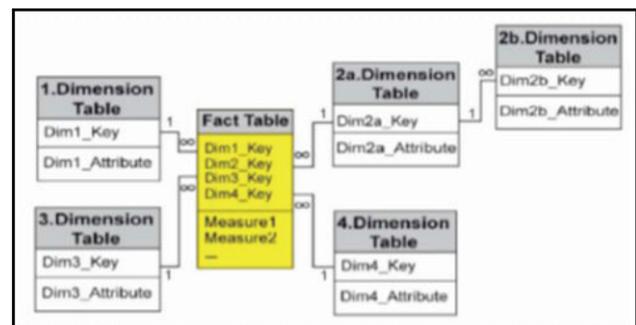


图1 雪花模型结构

图1显示了雪花模型的结构，该结构包含一个事实表，周围有维度和子维度表。维度表进行规范化处理，目的是能够简化和加快在电子图书馆数据仓库中设计的报告的查询过程。

2.4 四步维度建模

在数据仓库中，Kimball 推荐了一种非常独特的开发方法，包括自下而上在每个业务流程中开发数据集市^[12, 13]。该开发方法包括四个步骤：

1) 选择业务过程

这个阶段确定了在数据仓库建模结束时确定维度和事实所需的业务流程。

2) 声明粒度

事实表的内容可以通过粒度陈述来确定。粒度将映射和匹配数据仓库中需要的信息^[15]。

3) 确认维度

这个阶段进行的是开发维度表的过程，以满足事实表中提供的需求^[19]。

4) 确认事实

事实表是数据仓库的体现，是业务需求的主要目标。事实表将显示能够满足业务需求的业务流程的结果。

3. 研究方法

这项研究是通过实现预期结果所需的几个过程来进行的。研究方法如图2所示。

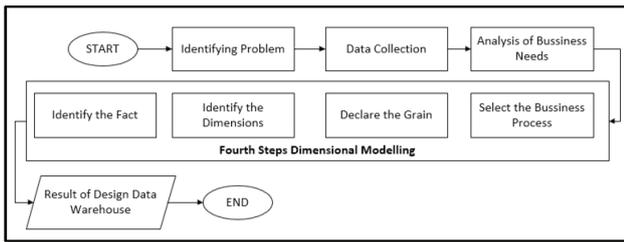


图2 研究方法

图2显示了所进行的研究方法的结构。发现了图书馆的藏书管理问题，它是设计数据仓库的基础。数据的收集是通过观察图书流通过程中发生的流转历史来进行的，如进馆的图书、还在借阅的图书、损坏的图书等等。以及数据仓库开发中的一些相关文献。此外，已经收集到的业务需求将成为设计数据仓库的基础，通过四步维度建模来开发数据仓库，即：

- 1) 选择业务过程；
- 2) 选择粒度；
- 3) 确认维度；
- 4) 确认事实。

通过这些阶段，数据仓库设计的结果有望成为电子图书馆业务需求的解决方案，协助分析藏书管理中需要的信息，作为采购最新藏书的考虑。设计的结果将产生结论，作为一个评估课题，以便数据仓库设计能够有效地运行并应用于电子图书馆。

4. 结果和分析

按照上一节所述的设计方法，本研究的结果将在几个解释中进行回顾。设计图书馆数据仓库的过程分以下几个阶段进行。

4.1 分析业务需求

这个阶段主要是确定要开发的数据仓库设计中需要什么。需求分析是在图3所示的电子图书馆数据库的基

础上进行的。

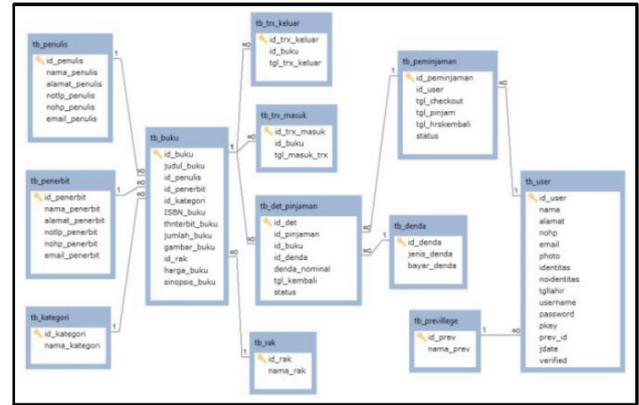


图3 OLTP电子图书馆

图3是作为数据库的OLTP设计，它在电子图书馆的交易过程中作为日常需求发挥作用。基于OLTP设计，图书馆的业务流程是多样化的，即：

- 1) 会员数据是为会员信息储存的各种数据，使其能够进入图书馆信息系统。
- 2) 图书数据是为图书馆所拥有的图书信息而储存的各种数据。
- 3) 出版商数据是指与成为图书馆藏品的图书的出版商有关的数据集合。
- 4) 作者数据是指与成为图书馆藏品的图书作者有关的数据集合。
- 5) 类别数据是用于对图书馆中收藏的图书类型进行分类的各种数据。
- 6) 书架数据是用于存储放置在图书馆书架上的图书的存储位置信息的数据。
- 7) 罚款数据是用于存储会员信息的数据，即逾期还书、损坏图书、淘汰图书馆藏书等。

各种数据是电子图书馆的业务需求，图书馆管理委员会使用一些需要的数据以获得信息。电子图书馆数据仓库设计的业务需求是在图书馆所拥有的藏书的管理方面确定的，如进出书的复述、损坏和丢失书的报告、最受喜爱书籍的报告，以及按类别划分的藏书数量报告。在采购最新的藏书时，需要将这些信息作为一种考虑。藏书管理被分析为一种业务需求，需要一些与目标相关的数据，其中包括：

- 1) 会员数据是交易过程中需要的会员数据信息。
- 2) 书籍数据需要了解电子图书馆中的藏书和会员对书籍的借阅历史。
- 3) 出版商数据是需要了解进入图书馆的图书交易历史，简化现有图书的分类过程。

4) 作者数据用以根据作者对藏书进行分类。

5) 类别数据用以根据电子图书馆的藏书类型对图书进行分组。

6) 数据架用于确定图书馆中的图书存储位置, 以便图书馆员管理现有的图书。

设计者可使用这五个数据来设计电子图书馆数据仓库, 作为图书馆藏书管理过程中的分析需要, 以及作为进行最新藏书采购的重要考虑因素。

4.2 选择业务过程

四步维度建模的初始阶段是对业务过程的选择。业务过程的确定是基于前几个阶段进行的信息需求分析。在数据仓库设计中选择的业务流程是:

- 1) 图书借阅业务
- 2) 最受喜爱的图书
- 3) 丢失和损坏的图书
- 4) 藏书总数

所选的业务流程在图书馆的图书管理中是需要的, 作为采购最新藏书的一种考量。

4.3 声明粒度

记录在事实表中的数据将通过粒度来确定。数据仓库中的粒度用于记录会员在电子图书馆中的图书借阅交易的数据处理。这个阶段粒度的确定可以在下面的表1中描述。

表1 声明粒度

粒度 维度	图书业务	最喜爱的 图书	丢失和损 坏的图书	藏书总数
图书	√	√	√	√
会员	√	√	√	
出版商		√	√	√
作者		√	√	√
类别		√	√	√
书架			√	√
时间	√	√	√	√

表1是根据选定的业务过程, 用实现事实表所需的各种维度来确定粒度的过程。

4.4 确认维度

维度的确定是为了满足事实表的需要。所选择的每个维度表都是有重要数据记录的数据表, 用于从数据仓库输出。表2到表11是确定为实现事实表的支持数据表的维度集合。在设计一个用于分析图书馆员所需信息需求的数据仓库时, 作为采购最新藏书的重要考虑, 需要一个维度表, 比如:

表2 会员维度

会员维度
会员号
身份
姓名
地址
性别
电话号码
邮箱

表3 图书维度

图书维度
图书编号
书名
作者编号
出版商编号
类别编号
出版年份
ISBN
来源编号
状态编号

表4 源库维度

源库维度
来源编号
来源

表5 图书状态维度

图书状态维度
状态编号
状态

表6 出版商维度

出版商维度
出版商编号
出版商名称
出版商地址
电话号码
邮箱

表7 作者维度

作者维度
作者编号
邮箱

表8 类别维度

类别维度
类别编号
类别

表9 书架维度

书架维度
书架编号
书架名称

表 10 时间维度

时间维度
时间编号
日期
月份
年

表 11 业务状态维度

业务状态维度
业务状态编号
状态

4.5 确认事实

在这个阶段，事实表是根据确定的维度表来选择的。数据仓库中的事实表解释了电子图书馆的业务过程，使图书馆员很容易分析信息需求，这些需求被用作采购最新的图书收藏的考虑因素。为电子图书馆数据仓库设计确定的事实表如表 12 所示。

表 12 业务事实

业务事实
业务编号
会员编号
图书编号
时间编号
业务状态编号

表 12 是事实表中使用的一些数据，它是由维度表之间的关系形成的。在案例研究中，作者设计了一个名为“业务事实”的事实表。这个事实表是各种维度的中心，以满足电子图书馆管理图书收藏的信息需求。

4.6 雪花模型设计

为满足电子图书馆藏书管理的信息需求，采用雪花模型的数据仓库设计如图 4 所示。

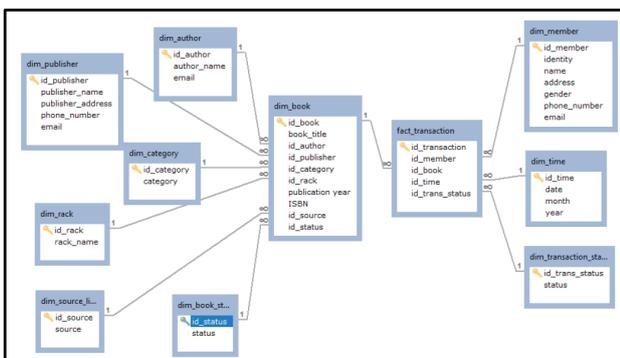


图 4 雪花模型设计

图 4 是一个用于设计电子图书馆数据仓库的雪花模型。有了这个数据仓库的设计方案和模型，电子图书馆的藏书管理将通过对各种可用数据的信息分析能力得到

加强。前面过程中确定的业务过程可以帮助作者设计一个理想和有效的电子图书馆。

4.7 雪花模型分析

使用雪花模型实施数据仓库有其缺点和优点。Gede 等人^[6]通过使用雪花模型，发现规范化的表可以减少数据冗余，加快处理所需的数据分析。使用这种方案的一些基本考虑是它可以减轻所用内存的负担，更容易维护以及更新过程更方便。缺点是会对查询的复杂程度、较高的数据结构、以及寻找数据的难度产生影响。然而，这种方案的设计在某些涉及到大量维与维之间的连接以支持事实表的需要的案例研究中还是可行的。正如本研究中已经完成的电子图书馆中的数据仓库的设计。

5. 结论

这项研究的结论是，在电子图书馆中设计一个数据仓库将非常有可能帮助图书馆员提供管理其藏书所需的信息。通过分析这个数据仓库设计中的预期业务需求，雪花模型的使用在本案例研究中是非常合适的。在电子图书馆中，有许多维度需要作为事实表的支持数据。四步维度建模方法在电子图书馆数据仓库的设计过程中也很有帮助。该方法的每个阶段都非常容易实现。

参考文献：

[1]Connolly, T. M., & Begg, C. E. (2010). Database Systems: A Practical Approach to Design, Implementation, And Management. Boston: Pearson Education.

[2]Foster, E. (2016). Database systems: a pragmatic approach. New York: Apress.

[3]Poniah, P. (2001). Data Warehouse Fundamentals: A Comprehensive Guide for IT Professional. New York: John Wiley & Sons

[4]Narendra, A. N., Murpratiwi, S. I. and Sudarma, M. (2017) ‘Design of E-Grant Application Data Warehouse’ , IJEET (International Journal of Engineering and Emerging Technology), 2(1), p. 11.

[5]Budiarta, K., Ananta Wijaya, P. A. and Partha, C. G. I. (2017) ‘ Analysis and Design of Data Warehouse on Academic STMIK STIKOM Bali’ , IJEET (International Journal of Engineering and Emerging Technology), 2(1), p. 35.

[6]Gede, L., Suardani, P. and Sudarma, M. (2018) ‘Data Warehouse Implementation to Support Batik Sales Information Using MOLAP’ , IJEET (International Journal of Engineering and Emerging Technology), 3(1), pp. 45 - 51.

[7]Putra Wijaya, I. P. A., Wahyudin, W. and Mataram,

- M. (2017) ‘Data warehouse Implementation on Denpasar City Online Community Complaints System’ , IJEET (International Journal of Engineering and Emerging Technology), 2(1), p. 67.
- [8]Prabawa, I. N. A. et al. (2019) ‘Analysis and Design Data Warehouse For E-Travel Business Optimization’ , IJEET (International Journal of Engineering and Emerging Technology), 4(1).
- [9]Inmon, W. H., & Hackathorn, R. D. (1994). Using the Data Warehouse. New York: John Wiley & Sons.
- [10]Lane, Paul. (2002). Oracle9i Data Warehousing Guide. Oracle Corporation.
- [11]Kimball, R., & Margy, R. (2002). The Data Warehouse toolkit: the complete guide to dimensional modelling. New York: John Wiley & Sons.
- [12]Kimball, R., & Margy, R. (2010). The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence. First Edition. Indianapolis: John Wiley & Sons
- [13]K. A. B. Permana, G. B. Subiksa, and M. Sudarma (2017), “Design Data Warehouse for Centralized Medical Record,” IJEET (International Journal of Engineering and Emerging Technology), vol. 2, no. 2, pp. 47 - 51
- [14]Ambara, M. P., Sudarma, M. and Kumara, I. N. S. (2016) ‘Desain Sistem Semantic Data Warehouse dengan Metode Ontology dan Rule Based untuk Mengolah Data Akademik Universitas XYZ di Bali’ , MITE (Majalah Ilmiah Teknologi Elektro), 15(1), p. 8. doi: 10.24843/mite.2016.v15i01p02.
- [15]Iswardani, P. R., Pramana, I. W. S. and Saputra, K. O. (no date) ‘Design of Data Warehouse for Monitoring Hotel’ s Food and Beverage Cost’ , IJEET (International Journal of Engineering and Emerging Technology), vol. 4, no. 1, pp. 1 - 4.
- [16]Shobirin, K. A., Iskandar, A. P. S. and Swamardika, I. B. A. (2017) ‘Data Warehouse Schemas using Multidimensional Data Model for Retail’ , IJEET (International Journal of Engineering and Emerging Technology), 2(1), p. 84. doi: 10.24843/ijeet.2017.v02.i01.p17.
- [17]Sudarmojo, Y. P. (2018) ‘Design of Library Data Warehouse Using OLTP Result of Services Analysis’ , IJEET (International Journal of Engineering and Emerging Technology) 3(1), pp. 62 - 65.
- [18]Pradhana, I. G. N. A., Giriantari, I. A. and Sudarma, M. (2018) ‘Analisis dan Perancangan Sistem Pengelola Data Menuju Implementasi Data Warehouse Untuk Mendukung Administrasi E-Procurement’ , MITE (Majalah Ilmiah Teknologi Elektro), 17(2), p. 245. doi: 10.24843/mite.2018.v17i02.p12.
- [19]Sulastra, I. M. D. J., Sudarma, M. and Kumara, I. N. S. (2015) ‘Pemodelan Integrasi Nearly Real Time Data Warehouse Dengan Service Oriented Architecture Untuk Menunjang Sistem Informasi Retail’ , MITE (Majalah Ilmiah Teknologi Elektro), 14(2), p. 11. doi: 10.24843/mite.2015.v14i02p03.
- [20]Anshari, M., Suryadharma, I. P. and Rinas, I. W. (2017) ‘Implementation of Data Warehouse in Human Resource Information System Using SEM-GeSCA’ , IJEET (International Journal of Engineering and Emerging Technology), 2(1), p. 16.