浙 江 林 学 院 学 报 2009，**26**(2)：149－154
Journal of Zhejiang Forestry College

森林资源数据库系统查询效率分析

张茂震[1]，唐小明[2]，谢阳生[2]，丁丽霞[1]

（1. 浙江林学院 环境科技学院，浙江 临安 311300；2. 中国林业科学研究院 资源信息研究所，北京 100091）

摘要：对森林资源数据库系统的查询特点及效率进行了分析，提出了森林资源数据库应用系统的4种应用查询模式及其制约查询效率的关键因素。实例分析证明：数据库管理系统(DBMS)参数配置、数据库逻辑和物理设计、SQL查询表达以及应用程序设计都是影响查询效率的关键因素。改进森林资源数据库系统查询优化应充分考虑不同查询应用模式的特点，分别从不同角度采用不同的优化方法和策略。对于以网络传输为主要特征的浏览查询，查询算法和数据组织是优化的关键；对于计算密集的统计分析查询，则应优先考虑数据库调整和使用中间表、聚簇等措施。图3表4参14

关键词：森林经理学；森林资源；数据库系统；查询优化
中图分类号：S757　　**文献标志码**：A　　**文章编号**：1000-5692(2009)02-0149-06

Analysis of query efficiency of forest resources database system

ZHANG Mao-zhen[1]，TANG Xiao-ming[2]，XIE Yang-sheng[2]，DING Li-xia[1]

（1. School of Environmental Science and Technology，Zhejiang Forestry College，Lin'an 311300，Zhejiang，China；
2. Research Institute of Fores Resource Information Techniques，The Chinese Academy of Forestry，Beijing 100091，China）

Abstract：Analyzed the query characteristics and its efficiency on forest resources database system，proposed that all the queries on forest resources database can be divided into 4 types，and showed the key factors affecting the query efficiency for each type. It is evident that DBMS parameters，the logical and physical design of database，the SQL expression，and the application system design are all the key factors affecting the system efficiency. The optimization of forest resources database application system should give full consideration to the characteristics of each query type. The policy should be made according to their query characteristics by choosing different optimizing method from different angle of view. For the type of browsing queries through network，the query algorithm and the data organizing are most important for query optimization. For the type of query of statistics and the analyzing，the priority should be given to database tuning，using of middle table and cluster indexing etc.［Ch，3 fig. 4 tab. 14 ref.］

Key words：forest management；forest resources；database system；query optimization

查询是数据库管理系统(DBMS)提供的基本功能之一。在查询过程中，SQL语句被翻译成关系代数的扩展形式，而查询求解计划则表示为关系操作构成的树。一个查询由若干个操作组成，对查询的求解，需要选择这些操作的不同组合来完成，各种组合的执行效率各不相同，从中选出较好的组合被称为查询优化[1]。查询优化的最终目的是缩短查询请求的响应时间[2]。对于一个数据库应用系统，要使整个系统的效率提高，必须从算法、源代码编写、SQL表达、数据组织等多方面进行优化[3-6]。因此，在应用中，查询优化往往是广义的，它代表了以提高整个数据库应用系统的用户请求响应效率为核心的多方面的优化工作[7-9]。对于森林资源管理等具体的专题数据库系统，除数据库管理系统的特

收稿日期：2008-07-10；修回日期：2008-12-12
基金项目："十一五"国家科学技术支撑项目(2006BAD23B0204-4)；浙江林学院科学研究发展基金资助项目(2006FR058)；浙江省林业厅资助项目(07A16)
作者简介：张茂震，副教授，博士，从事地理信息系统与数据库应用技术研究。E-mail：zhangmaozhen@126.com

性和数据库设计以外，业务规则及其应用特点也是查询优化应考虑的重要因素[10-11]。因此，从森林资源数据库的数据组织方式和应用查询特点入手，分析各相关因素对系统效率的影响，有针对性地在不同层次上提出查询优化策略，对森林资源管理信息系统设计乃至整个林业信息化建设都具有重要意义[12]。

1 森林资源数据库应用系统查询处理的特点

1.1 应用查询类型

在数据库系统中，执行任何一个 SQL(DML)语句，系统内部一般都有相应的查询处理过程，因此，包括 Update 在内的对数据的所有操作都可以归结为查询。在森林资源数据库中，主要查询模式可以归纳为以下 4 种类型。

1.1.1 *报表统计* 报表统计是完成各级政府和林业主管部门要求的报表而进行的查询处理。其特点是查询涉及的表多，查询内容广，查询语句复杂，查询的执行集中，而且查询由专人操作，有季节性。进行这类查询的用户主要是森林资源监测和管理机构、林业行政管理部门。其体系结构多为 C/S。

1.1.2 *浏览查询* 浏览查询是指仅查询一般信息的操作，查询的模式和目标都相对稳定。这类查询主要是对统计分析结果的查询，数据量不大，涉及的数据对象和类型也不多，但查询的用户最多，包括政府办公用户和公众用户两大部分。浏览查询主要在 B/S 体系结构中进行，在目前情况下，查询效率与数据库及其应用系统有关，但在很大程度上取决于广域网的带宽。

1.1.3 *分析处理* 分析处理是指为了特定的目标，运用一定的专业知识和计算工具，对森林资源数据进行各种数学运算，得到满足设计要求的查询。常见于森林资源管理研究、林业生态及其相关工程设计、林业管理与森林经营辅助决策等领域。分析处理查询用户较少，主要是林业决策部门、森林资源监测部门、森林经营部门等。此类查询对数据库系统性能要求较高，查询的响应时间一般都比较长。

1.1.4 *数据维护* 数据维护是指对数据库中数据进行管理和更新维护的操作，是数据库管理的重要内容。数据维护的操作较复杂，涉及数据定义、数据操作及数据完整性约束的各种操作。作为 DBA，不仅需要随时掌握 DBMS 的状态，而且需要经常通过对数据库的查询了解和掌握实时数据状态，以便对数据进行各种更新操作。这类用户数量最少，操作一般都在 C/S 体系结构下实现，查询效率较高。

1.2 查询特点

1.2.1 *基本特征* 根据以上分析可知，森林资源数据库系统中各种查询应用类型都有自己的特点(表 1)。

表 1 森林资源数据库系统查询的主要特点

Table 1 The main features for the query on forest resources database

编号	类型	查询特点
1	报表统计	SQL 语句多,类型相对固定,运行集中,C/S 结构 查询语句以聚集查询为主
2	浏览查询	SQL 语句少,类型相对固定,运行分散,B/S 结构 查询语句以 SPJ 查询为主
3	分析处理	SQL 语句数量有限,类型不固定,运行分散,C/S 结构+B/S 结构 查询语句包括聚集查询和 SPJ 查询的各种类型
4	数据维护	SQL 语句多,类型不固定,运行集中,C/S 结构 查询语句多为包括等值连接的各种更新操作,也包括 SPJ 查询和聚集查询

1.2.2 *森林资源数据库系统中主要 SQL 语句类型* ①简单选择、投影和连接查询(SPJ queries)：$\pi_{c1[,c2...]}[\sigma_{c1=a[c2=b[...]]}]([(R.c1=S.c1[\wedge R.c2=S.c2[...]])](R[\times S...])$。②非分组聚集查询(non-group-by aggregation queries)：$\pi_{SUM\ [AVG\mid COUNT\mid MAX\mid MIN](Area\ [Volume\mid Growth]}(\sigma_{c1=a\ [c2=b\ [...]]}([(R.c1=S.c1\ [\wedge R.c2=S.c2\ [...]]](R[\times S...]))))$。③分组聚集查询(group-by aggregation queries)：$\pi_{c1[,c2...]SUM[AVG\mid COUNT\mid MAX\mid MIN](Area[Volume\mid Growth]}([Having\ \_[Count(*)[...]]]...(Group\ \_By\ \_c1[,c2...])(\sigma_{c1=a[c2=b[...]]}([(R.c1=S.c1[\wedge R.c2=S.c2[...]]](R[\times S...])))$。④嵌套查询(nested queries)：$\pi_{R.c1[,c2...]}[\sigma_{R.c1=}(\sigma_{s.c1=b[...]}(S)))(R))$。

嵌套查询是在 WHERE 子句中含查询的语句，包含上面 3 种查询中带嵌套的查询。在更新操作

(update manipulation)中，以上 4 种类型都可能包含，但以 SPJ 查询和嵌套查询居多。

森林资源数据库系统中，查询语句以聚集查询为主，其中分组聚集查询占极大比例。因此，解决聚集查询效率问题是优化的关键问题之一。

2　影响森林资源数据库查询效率的因素

森林资源数据库是一个共享数据源，各类应用系统建筑在此数据源基础之上。数据库系统的效率与其具体应用有关，数据量、应用逻辑等都是影响效率的重要因素。在各类应用系统中，以聚集查询为主要查询类型的统计查询是查询的主体。因此，对聚集查询效率分析是分析的主要对象。为了更有效地分析查询响应速度，森林资源数据库系统按其应用类型分子系统测试。下面以国家森林资源连续清查和森林资源规划设计调查数据库系统(福建省)为例，对主要类型查询的效率和优化方法与潜力进行分析。分析内容涉及 DBMS、数据库设计和应用程序设计等方面。

2.1　DBMS 参数配置

DBMS 的性能直接影响数据库查询效率，可以通过调整其参数对数据库查询性能进行优化。主要调整是 I/O 模式和内存分配，如数据高速缓存、全局共享区等，不同的 DBMS 参数虽有一定区别，但基本方法相同。下面以 Oracle9i[13]为 DBMS 来进行测试分析，用同一条 SQL 语句运行于具有不同参数的 DBMS 环境，得出响应时间曲线。以下为用于测试的 SQL 语句：

```
select a.plot_no, sum(nvl(a.volume, 0)-nvl(b.volume, 0))from tree2003_350 a, tree1998_350 b
where a.plot_no = b.plot_no and a.tree_no = b.tree_no
and a.tree_type = 1 and a.tally_type = 11  group by a.plot_no
```

测试表明，缓冲区高速缓存的大小对查询相应时间有巨大影响，在此条件下，高速缓存的大小是优化的主要问题。数据及结果如图 1 所示：对于 SGA 区的 Shared pool size 和 PGA 大小的调整，效果均不太明显(图 2)。图 1 和图 2 反映了 DBMS 的几个主要参数对查询效率的影响。其中缓冲区高速缓存对查询影响最大。

以上调整均在其他参数不变的条件下进行：调整 PGA 大小时，SGA 区为 64 MB，Shared pool size 为 12 MB、DB Cache size 为 12 MB。调整 Shared pool 大小时，PGA 大小为 12 MB，其他不变。

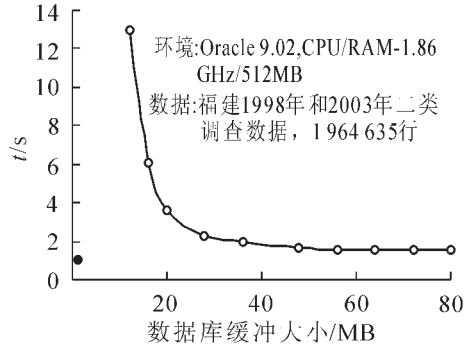

图 1　查询响应时间与数据高速缓存的关系

Figure 1　The relation between response time and the DB cache size

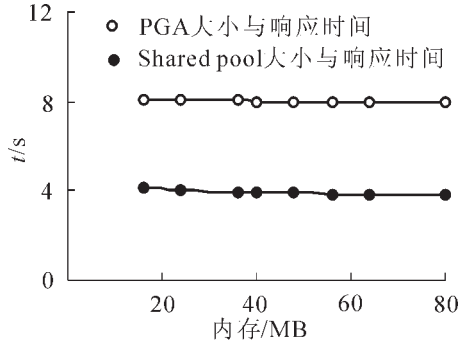

图 2　查询响应时间与数据库其他参数的关系

Figure 2　The relation between response time and other DB parameters

上图侧重反映查询响应时间的趋势，在绝对时间上可能因不同的环境而有所不同，但查询响应时间趋势一致。环境因素在测试中保持前后一致，未进行专门测试和分析。

2.2　数据库设计

2.2.1　*数据规范化和反规范化*　模式是数据库逻辑设计的关键。通过模式分解，将一个低一级范式的关系模式转化为若干个高一级范式的关系模式，这一过程称为规范化。规范化使数据结构在逻辑上合理，在保证数据一致性、完整性和降低数据维护难度等方面有重要作用[14]。但是，当一个查询包含较

多表连接时其查询效率会受较大影响，即不符合范式要求的数据结构可能会有更高的查询效率。如在一类调查数据中，按 3 N 要求，有样地基本情况、生态环境、经济林、灌木林、林分蓄积等多个关系。如果保留数据冗余，这些关系可以概括为一个关系。但这样会浪费较多存储容量，而且还会由于属性间的函数依赖关系带来数据操作异常等问题。在实际应用中，应具体分析，权衡利弊。

2.2.2 *数据物理存储* 数据库日志文件通常执行连续写，数据文件可能执行连续 I/O、随机 I/O 或两者都有。将连续 I/O 分配到与随机 I/O 不同的磁盘上，同时将不同的连续 I/O 文件分离到它们各自的磁盘上，都可以使数据的访问速度得以提升。例如，在 Oracle9i DBMS 中，建立一个表空间，并为这一表空间指定 3 个数据文件，这些数据文件分别存储于 3 个不同的物理硬盘上。用同样 SQL 语句查询(福建省二类调查数据)，其响应时间相差十分明显(表 2)。

表 2 数据文件在不同物理位置的查询响应时间

Table 2 The query efficiency affected by the physical location of data files

语句序号	SQL 查询语句	数据逻辑组织方式	响应时间 t_1/s	响应时间 t_2/ s
1	select LandType,Sum (Area) from SubComp where AgeGroup=2 and ct_code='D'group by LandType	Where 子句中,列上无索引	17.603	8.210
2	select LandType,Sum(Area) from SubComp where AgeGroup=2 group by LandType	Where 子句中,列上有索引	3.825	2.191

说明:响应时间 t_1 为数据存储在一个物理盘上,响应时间 t_2 为数据分别存储在 3 个物理盘上。

2.3 应用程序及 SQL 语句

2.3.1 *应用程序* ①数据预处理。将规定的统计查询和常用的查询要求进行分类，对每一类的数据源进行分析，为查询提供最接近目标的数据源，以降低查询代价。例如，对福建省二类调查统计查询语句分类，按分类语句的最小粒度将查询结果存入中间表，供报表统计查询，大幅度提高了查询效率(表 3)。②查询处理算法。查询处理算法是指为组织查询语句和实现查询过程的一系列逻辑，对这些逻辑的优化有相当大的潜力。查询处理算法涉及分类、多表连接和复杂查询的分解等多个方面。例如，为了提高执行效率，可将多个语句合成一个来执行。在报表统计中，可以用一个 SQL 语句从数据源中一次获得一个列的统计数据，也可以一次仅获取一个分量。究竟采用哪种方式，取决于两类 SQL 查询语句执行效率的比较。

表 3 森林资源数据库系统中数据预处理对查询效率的影响

Table 3 The affects on query efficiency by data pre-processing of forest resources database

数据源名称	查询目标	查询方式	响应时间 t/s
小班基本数据	1. 分权属、林种、优势树种的面积和蓄积	将面积、蓄积按县、地类、权属、林种、优势树种分组,并存储成临时表作为查询数据源	0.10
小班林分数据	2. 分地类、权属的面积		0.06
小班林分数据	1. 分权属、林种、优势树种的面积和蓄积	不进行任何预处理,直接查询数据源	245.02
小班基本数据	2. 分地类、权属的面积		201.00

2.3.2 *SQL 语句* 同一查询目的可以用不同的 SQL 语句实现，不同查询表达方式有不同的执行效率。在森林资源二类调查数据库系统中，查询主要分为 SPJ 查询和聚集查询。统计查询一般都为带聚集函数的聚集查询或分组聚集查询，如：Select forcateg，sum (area)，sum (volume) from SubComp where ownership = 1 group by forcateg 就是最常用和最简单的查询模式之一。在这个语句中，有投影运算、选择运算以及分组和聚集运算。随机查询中较多用到 SPJ 查询，如 select sublotid from SubComp where ownership = 1 and Domspecies > 100 and Domspecies <= 200 and volume > 10；SPJ 查询主要是选择

和投影。无论是聚集查询或 SPJ 查询，每个语句都有多种不同的等价语法表达，效率各不相同[7,13]。

2.3.3　索引　①非聚簇索引。非聚簇索引是一般数据库管理系统中缺省的索引，在不给定任何条件的时候，建立的索引即非聚簇索引。在非聚簇索引条件下，数据在物理上随机存放在数据页上，在范围查询时，必须执行一次表扫描才能找到这一范围内的全部行。非聚簇索引包括 B 树索引、位图索引、组合索引。②聚簇索引。聚簇索引是对磁盘上数据重新组织，以按指定的一个或多个列的值排序。一个聚簇索引是一个 B 树，其底层包含了表中所有的数据页，并且数据的物理存储顺序与索引顺序完全相同，即聚簇索引的数据是按照一定的物理排序方式来保存的。由于聚簇索引的索引页面指针指向数据页面，所以使用聚簇索引检索数据要比非聚簇索引快，而且它适用于检索连续键值[9]。

以下是在小班数据中查询国有用材林中的成过熟林林分的面积和蓄积的 SQL 语句：

Select sublotid，area，volume from SubComp where ownership = 1 and landtype = 111 and forcateg = 11 and age_group ＞ 3

该查询返回记录 11 865 条，占总记录数的 0.6%。

结果表明：建立与查询语句相对应的组合索引效率最高，比无索引条件下的查询速度提高将近 110 倍。在各列上分别建立索引，其查询效率略低于组合索引(表 4)。

表 4　二类数据库中不同索引对查询的影响

Table 4　The query efficiency with different kinds of indexes

序号	索引方式	物理读盘次数	查询响应时间 *t*/s
1	无索引	59 066	23.960
2	权属、地类、林种、龄组上分别建 B 树索引	6	0.300
3	权属、地类、林种、龄组上建组合 B 树索引	0	0.210
4	权属、地类、林种、龄组上建聚簇索引	6	0.430
5	权属上建位图索引,其他列建 B 树索引	0	0.200

2.3.4　计算策略　在统计查询中，对于同一个问题，可以用多种编程方案解决。这些方案之间在效率上可能存在很大差异。一个最简单的例子是各地类、林种的面积统计。对于这个问题，至少有 4 种解决方案：①直接用两维的分组查询，即 Group By 地类，林种的形式；②先找出哪几种地类，将地类代码放在内存中作为参数，循环组成按林种分组的面积查询语句；③先查询地类，同样将地类代码作为参数存储于内存，组织好按林种分组查询的聚集查询游标，地类可作为游标所带的输入参数，用 Open ＜cursorname＞ Using ＜parameter＞的形式调用；④将全部记录的地类、林种和面积 3 个列的值读入内存，用程序分组处理。图 3 是用福建省森林资源数据进行查询的测试结果，纵坐标为总查询时间，序列 1，3，4 包括查询和处理时间。结果显示，就总时间看，采用一次 SQL 分组查询效果最好。

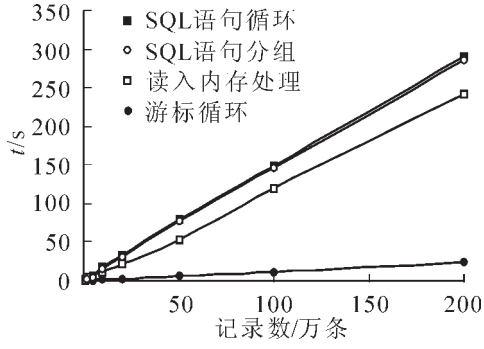

图 3　分地类和林种的面积统计方案与效率

Figure 3　The time consumption on different plans for summing area by land type and forest category

3　结语

森林资源数据库系统的性能与 DBMS 配置、数据库设计、SQL 语句表达及应用程序设计都有密切关系，其中，数据库缓存大小、数据的组织方式、应用系统的业务逻辑和查询特点至关重要。要提高系统的性能，除运用一般查询优化理论、方法和策略以外，还应结合应用特点，采用特定的综合策略

来实现。在应用程序中，计算策略是影响系统效率的最重要因素之一。

从查询处理的角度，森林资源数据库应用系统查询可概括为 4 类：报表统计查询、浏览查询、分析处理查询、数据维护查询。浏览查询一般采用 B/S 体系结构，数据和运算量小，网络依赖性强，特别是远程网络，因而其优化策略应该以减少网络传输量为主。报表统计查询、分析处理和维护查询一般采用 C/S 体系结构在局域网内实施，其特点是运算较集中，占用系统资源较多，优化策略应以优化算法和数据库设计为主，包括中间表的应用，同时考虑 DBMS 配置。

参考文献：

[1] 苗雪兰，刘瑞新，宋会群. 数据库系统原理与应用[M]. 北京：机械工业出版社，2005.

[2] 董玉杰，李小军. 关系数据库系统的查询优化策略[J]. 计算机与现代化，2005 (8)：72 – 74.
DONG Yujie，LI Xiaojun. Query optimization strategy of RDBMS[J]. *Comput & Mod*，2005 (8)：72 – 74.

[3] 朱鸿宇，刘瑰，唐福华，等. 数据库查询优化中的智能预取技术[J]. 计算机应用研究，2007，**24** (5)：35 – 40.
ZHU Hongyu，LIU Gui，TANG Fuhua，*et al*. Intelligent prefetch algorithm on database query optimization[J]. *Appl Res Comput*，2007，**24** (5)：35 – 40.

[4] 徐丽萍，金雄兵，赵小松. 并行数据库查询优化技术研究[J]. 华中科技大学学报：自然科学版，2006，**34** (3)：11 – 20.
XU Liping，JIN Xiongbing，ZHAO Xiaosong. Query optimization technique for parallel databases[J]. *J Huazhong Univ Sci Technol Nat Sci Ed*，2006，**34** (3)：11 – 20.

[5] 魏士伟，黄文明，康业娜，等. 分布式数据库中基于半连接的查询优化算法研究[J]. 计算机应用，2007，**27**(增刊 1)：34 – 39.
WEI Shiwei，HUANG Wenming，KANG Yena，*et al*. A Study on query optimazation based on semi join in distributed database[J]. *Comput Appl*，2007，**27** (supp 1)：34 – 39.

[6] 李春生，罗晓沛. 基于. NET 实现分布式数据库查询[J]. 计算机工程与设计，2007，**28** (12)：2937 – 2939.
LI Chunsheng，LUO Xiaopei. Implementation of distributed networks database query based on.NET[J]. *Comput Eng Des*，2007，**28** (12)：2937 – 2939.

[7] 张茂震. 森林资源数据库查询优化策略与技术研究[D]. 北京： 北京林业大学，2006.
ZHANG Maozhen. *A Study on the Optimazation of Query Strategy and the Technique in Forest Resources Database*[D]. Beijing： Beijing Forestry University，2006.

[8] 傅明，张桂平. 关系型数据库查询优化及实践[J]. 长沙交通学院学报，1995，**11** (3)：6 – 11.
FU Ming，ZHANG Guiping. Query optimization and practice on relational database [J]. *J Changsha Commun Univ*，1995，**11** (3)：6 – 11.

[9] 宋薇，董占球. 聚簇索引在数据库查询中的重要作用[J]. 微机发展，2000，**10** (5)：70 – 73.
SONG Wei，DONG Zhanqiu. Cluster index plays an important role in database query[J]. *Microcomput Develop*，2000，**10** (5)：70 – 73.

[10] 曹世恩. 森林资源信息处理自动化系统的研究[J]. 林业资源管理，1991(增刊)：124 – 129.
CAO Shien. Study on the automation of forest resources data processing[J]. *For Resour Manage*，1991(supp)：124 – 129.

[11] 罗光斗. 高峰林场森林资源核算微机管理系统的研制[J]. 中南林业调查规划，1996 (2)：6 – 9.
LUO Guangdou. The design and development of the forest resource accounting management system in Gaofeng forest farm [J]. *Central South For Inventory Planing*，1996 (2)：6 – 9.

[12] 陈瑞吕. 森林资源管理信息系统的研究现状及发展[J]. 林业资源管理，2001 (6)：73 – 79.
CHEN Ruilu. The recent development of the forest resource management information system and its trend[J]. *For Resour Manage*，2001 (6)：73 – 79.

[13] 陈向辉，王敬乐. 基于 Oracle 的应用软件系统检索性能的优化[J]. 河北科技大学学报，2002，**23** (1)：60 – 63.
CHEN Xianghui，WANG Jingle. Performance optimization of searching based on oracle application software system[J]. *J Hebei Univ Sci Technol*，2002，**23** (1)：60 – 63.

[14] 刘靖. 数据库规范化及系统优化[J]. 企业技术开发，2005，**24** (2)：12 – 14.
LIU Jing. Database standardization and system optimization[J]. *Technol Develop Enterp*，2005，**24** (2)：12 – 14.