Chapter 6

数据库模式设计之层次结构

处理层次结构（Hierarchical Data）

- 树状结构（Tree Structures）
  - 历史…
    - 层次数据库
    - 网状数据库
    - 关系型数据库
  - 直到关系理论出现，数据库设计是“科学（science）”而非“工艺（craft）”
    - 层次性数据广泛存在（XML，LDAP，BOM…）
  - 层次结构复杂度在于
    - 访问树的方式

树状结构VS.主从结构

- 父子结构（parent/child link）--tree structure
- 主从结构（master/detail relationship）
- 差异
  - 树状结构保存只需要一张表
  - 深度
  - 所有权
  - 多重父节点

- 参考书籍：Fabian Pascal：Practical Issues in Database Management（Addion Wesley）

层次结构的实际案例

- Risk exposure
- 档案位置
- 原料使用
- ……
- 不同的案例具有不同的基本特征
- 通常，树中的节点数量偏小。实际上，这也是树的优点，便于高效检索

层次结构的实际案例

```
select building.name building,
       floor.name floor,
       room.name room,
       alley.name alley,
       cabinet.name cabinet,
       shelf.name shelf,
       box.name box,
       folder.name folder
from inventory,
     location folder,
     location box,
     location shelf,
     location cabinet,
     location alley,
     location room,
     location floor,
     location building
where inventory.id = 'AZE087564609'
  and inventory.folder = folder.id
  and folder.located_in = box.id
  and box.located_in = shelf.id
  and shelf.located_in = cabinet.id
  and cabinet.located_in = alley.id
  and alley.located_in = room.id
  and room.located_in = floor.id
  and floor.located_in = building.id
```

用SQL数据库描述树结构

- 只要对象的类型相同，而对象的层树可变，其关系就应该被建模为树结构
- 在数据库设计中，树通常三种模型
  - Adjacency model-邻接模型
  - Materialized path model-物化路径模型
  - Nested set model-嵌套集合模型
    - Joe Celko发明
    - Vadim Tropashko 提出过nested interval model

数据来源http://www.kessler-web.co.uk

树的实际实现：邻接模型

ADJACENCY_MODEL

Name	Null?	Type
ID	NOT NULL	NUMBER
PARENT_ID		NUMBER
DESCRIPTION	NOT NULL	VARCHAR2(120)
COMMANDER		VARCHAR2(120)

表的每一行描述一个部队，parent_id指向树中的上级部队

树的实际实现：物化路径模型

MATERIALIZED_PATH_MODEL

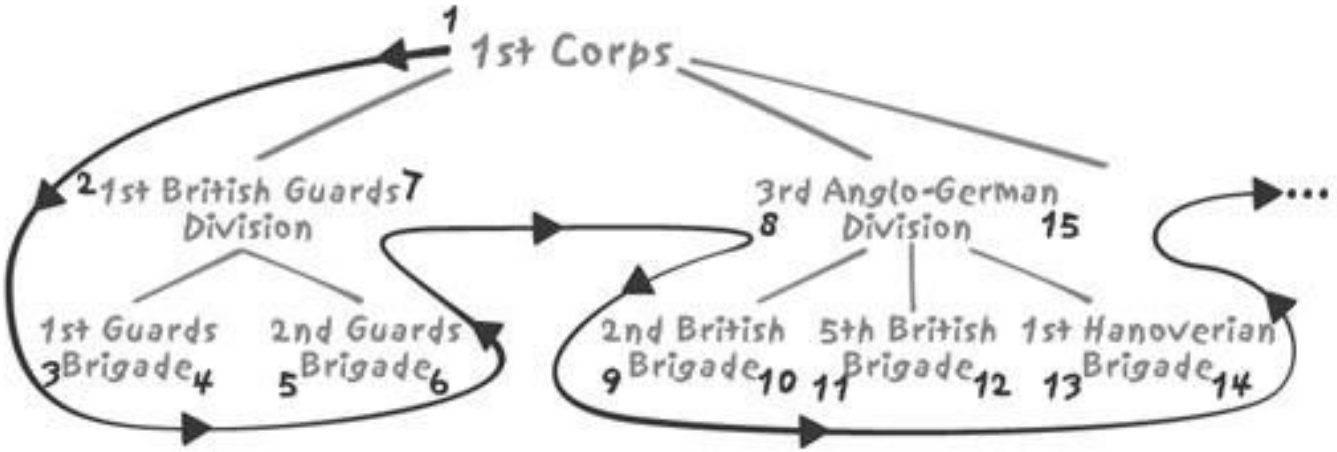
Name	Null?	Type
MATERIALIZED_PATH	NOT NULL	VARCHR2(25)
DESCRIPTION	NOT NULL	VARCHAR2(120)
COMMANDER		VARCHAR2(120)

表中有两个索引，在materialized_path上的唯一性索引以及在commander上的索引，正确的设计应该增加id字段。

树的实际实现：嵌套集合模型

NESTED_SETS_MODEL

Name	Null?	Type
DESCRIPTION		VARCHAR2(120)
COMMANDER		VARCHAR2(120)
LEFT_NUM	NOT NULL	NUMBER
RIGHT_NUM	NOT NULL	NUMBER



用SQL访问树结构

- 为了检查效率和性能，分别用不同模型解决如下两个问题：
- 法国将军Dominique Vandamme指挥哪些部队，以缩排方式或简单列表的方式显示他们。注意，所有的commander字段都构建了索引（简称Vandamme查询）
- Scottish Highlanders的每个团各属于哪个部队（自底向上的查询）。在部队的名称（description字段）上没有索引，唯一的方法是在description字段中查找“Highland”字符串，在没有任何全文索引的情况下，这个问题简称highland问题
  - 注：层次结构Corp-division-brigade-regiment
  - Oracle

自顶向下查询：Vandamme查询

- 邻接模式
  - connect by <*a column of the current row*> = prior <*a column of the previous row*>,
  - connect by <*a column of the previous row*> = prior <*a column of the current row*>

```
select lpad(description, length(description) + level) description,
       commander
  from adjacency_model
  connect by parent_id = prior id
  start with commander = 'Général de Division Dominique Vandamme'
```

邻接模式

```
DESCRIPTION                    COMMANDER
------------------------------ ------------------------------------------
 III Corps                     Général de Division Dominique Vandamme
  8th Infantry Division        Général de Division Baron Etienne-Nicolas Lefol
   2nd Brigade                 Général de Brigade Baron Corsin
    37th Rgmt de Ligne         Colonel Cornebise
   1st Brigade                 Général de Brigade Billard (d.15th)
    23rd Rgmt de Ligne         Colonel Baron Vernier
    15th Rgmt Léger            Colonel Brice
       ...
  10th Infantry Division       Général de Division Baron Pierre-Joseph Habert
   2nd Brigade                 Général de Brigade Baron Dupeyroux
    70th Rgmt de Ligne         Colonel Baron Maury
    22nd Rgmt de Ligne         Colonel Fantin des Odoards
    2nd (Swiss) Infantry Rgmt  Colonel Stoffel
   1st Brigade                 Général de Brigade Baron Gengoult
    88th Rgmt de Ligne         Colonel Baillon
    34th Rgmt de Ligne         Colonel Mouton
   Division Artillery
    18/2nd Foot Artillery      Captain Guérin

40 rows selected.
```

邻接模式:递归实现

- STEP 1： define starting point

```
select 1 level,
      id,
      description,
      commander
from adjacency_model
where commander = 'Général de Division Dominique Vandamme'
```

- STEP 2： define how each child row relates to its parent row

```
select parent.level + 1,
      child.id,
      child.description,
      child.comander
from recursive_query parent, adjacency_model child
where parent.id = child.parent_id
```

邻接模式:递归实现

```
with recursive_query(level, id, description, commander)
as (select 1 level,
          id,
          description,
          commander
    from adjacency_model
    where commander = 'Général de Division Dominique Vandamme'
    union all
    select parent.level + 1,
          child.id,
          child.description,
          child.commander
    from recursive_query parent,
        adjacency_model child
    where parent.id = child.parent_id)
select char(concat(repeat(' ', level), description), 60) description,
      commander
from recursive_query
```

邻接模式:递归实现

```
with recursive_query(level, id, rank, description, commander)
as (select 1,
        id,
        cast(1 as double),
        description,
        commander
   from adjacency_model
   where commander = 'Général de Division Dominique Vandamme'
   union all
   select parent.level + 1,
        child.id,
        parent.rank + ranking.sn / power(100.0, parent.level),
        child.description,
        child.commander
   from recursive_query parent,
      (select id,
            row_number( ) over (partition by parent_id
                        order by description) sn
        from adjacency_model) ranking,
      adjacency_model child
   where parent.id =child.parent_id
     and child.id = ranking.id)
select char(concat(repeat(' ', level), description), 60) description,
     commander
from recursive_query
order by rank
```

邻接模式:递归实现

DESCRIPTION	COMMANDER
III Corps	Général de Division Dominique Vandamme
10th Infantry Division	Général de Division Baron Pierre-Joseph Habert
1st Brigade	Général de Brigade Baron Gengoult
34th Rgmt de Ligne	Colonel Mouton
88th Rgmt de Ligne	Colonel Baillon
2nd Brigade	Général de Brigade Baron Dupeyroux
22nd Rgmt de Ligne	Colonel Fantin des Odoards
2nd (Swiss) Infantry Rgmt	Colonel Stoffel
70th Rgmt de Ligne	Colonel Baron Maury
Division Artillery	
18/2nd Foot Artillery	Captain Guérin
11th Infantry Division	Général de Division Baron Pierre Berthézène
...	
23rd Rgmt de Ligne	Colonel Baron Vernier
2nd Brigade	Général de Brigade Baron Corsin
37th Rgmt de Ligne	Colonel Cornebise
Division Artillery	
7/6th Foot Artillery	Captain Chauveau
Reserve Artillery	Général de Division Baron Jérôme Doguereau
1/2nd Foot Artillery	Captain Vollée
2/2nd Rgmt du Génie	

物化路径模型

- 查询编写不困难
- 计算由路径导出的层次不方便
- 假设mp_depth()函数返回当前节点深度

```
select lpad(a.description, length(a.description)
       + mp_depth(...)) description,
      a.commander
from materialized_path_model a,
    materialized_path_model b
where a.materialized_path like b.materialized_path || '%'
 and b.commander = 'Général de Division Dominique Vandamme')
order by a.materialized_path
```

嵌套集合模型

- 很简单，某节点的后代的left_num和right_num都会在该节点的left_num和right_num范围内

```
select a.description,
       a.commander
from nested_sets_model a,
     nested_sets_model b
where a.left_num between b.left_num and b.right_num
 and b.commander = 'Général de Division Dominique
Vandamme'
```

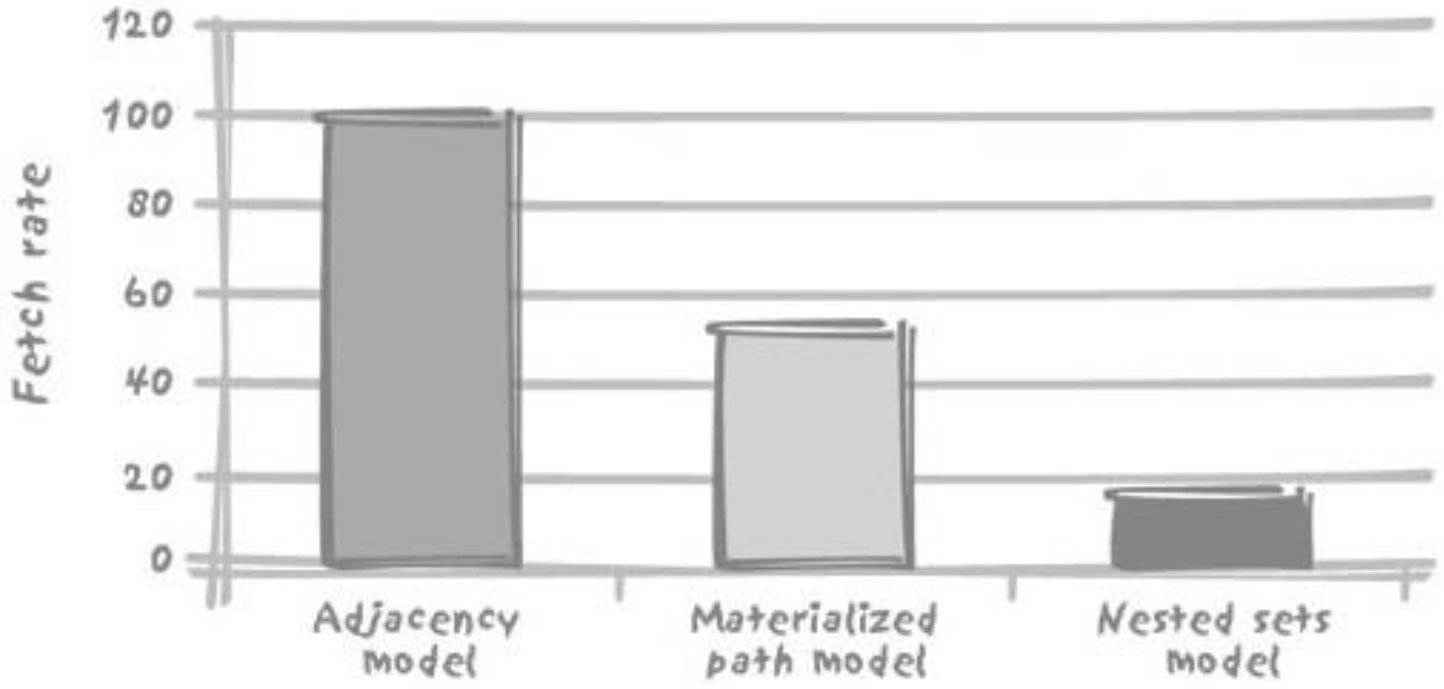
嵌套集合模型

- 缩排怎么办……

```
select lpad(description, length(description) + depth) description,
       commander
from (select count(c.left_num) depth,
             a.description,
             a.commander,
             a.left_num
        from nested_sets_model a,
             nested_sets_model b,
             nested_sets_model c
       where a.left_num between c.left_num and c.right_num
         and c.left_num between b.left_num and b.right_num
         and b.commander = 'Général de Division Dominique Vandamme'
       group by a.description,
                a.commander,
                a.left_num)
order by left_num
```

比较各模型下的Vandamme模型

- 返回40条记录，循环执行每个查询5000次，比较每秒返回的记录数



自底向上访问：Highland查询

- 在description字段中查找“Highland”字符串
- 必然导致完整的表扫描
- 不同模型下Highland查询的差异

邻接模式

- Connect by相当容易实现

```
select lpad(description, length(description) + level) description,
      commander
from adjacency_model
connect by id = prior parent_id
start with description like '%Highland%'
```

```
DESCRIPTION                      COMMANDER
  -------------------------------- ---------------------------------------
   2/73rd (Highland) Rgmt of Foot    Lt-Colonel William George Harris
    5th British Brigade              Major-General Sir Colin Halkett
     3rd Anglo-German Division       Lt-General Count Charles von Alten
      I Corps                        Prince William of Orange
       The Anglo-Allied Army of 1815 Field Marshal Arthur Wellesley, Duke of
                                     Wellington
   1/71st (Highland) Rgmt of Foot    Lt-Colonel Thomas Reynell
    British Light Brigade            Major-General Frederick Adam
     2nd Anglo-German Division       Lt-General Sir Henry Clinton
      II Corps                       Lieutenant-General Lord Rowland Hill
       The Anglo-Allied Army of 1815 Field Marshal Arthur Wellesley, Duke of
                                     Wellington
   1/79th (Highland) Rgmt of Foot    Lt-Colonel Neil Douglas
    8th British Brigade              Lt-General Sir James Kempt
     5th Anglo-German Division       Lt-General Sir Thomas Picton (d.18th)
      General Reserve                Duke of Wellington
       The Anglo-Allied Army of 1815 Field Marshal Arthur Wellesley, Duke of
                                     Wellington
   1/42nd (Highland) Rgmt of Foot    Colonel Sir Robert Macara (d.16th)
    9th British Brigade              Major-General Sir Denis Pack
     5th Anglo-German Division       Lt-General Sir Thomas Picton (d.18th)
      General Reserve                Duke of Wellington
       The Anglo-Allied Army of 1815 Field Marshal Arthur Wellesley, Duke of
                                     Wellington
   1/92nd (Highland) Rgmt of Foot    Lt-Colonel John Cameron
    9th British Brigade              Major-General Sir Denis Pack
     5th Anglo-German Division       Lt-General Sir Thomas Picton (d.18th)
      General Reserve                Duke of Wellington
       The Anglo-Allied Army of 1815 Field Marshal Arthur Wellesley, Duke of
                                     Wellington

  25 rows selected.
```

物化路径模型

- 仅找出适当的记录并缩排显示算容易

```
select lpad(a.description, length(a.description)
                    + mp_depth(b.materialized_path)
                     - mp_depth(a.materialized_path)) description,
       a.commander
  from materialized_path_model a,
       materialized_path_model b
 where b.materialized_path like a.materialized_path || '%'
   and b.description like '%Highland%')
```

  - 重复记录的问题
  - 顺序的问题

物化路径模型

```
select description, commander
from (select distinct lpad(a.description, length(a.description)
                    + mp_depth(b.materialized_path)
                      - mp_depth(a.materialized_path)) description,
                    a.commander,
                    a.materialized_path
      from materialized_path_model a,
           materialized_path_model b
      where b.materialized_path like a.materialized_path || '%'
        and b.description like '%Highland%')
order by materialized_path desc
```

much nicer and more compact result

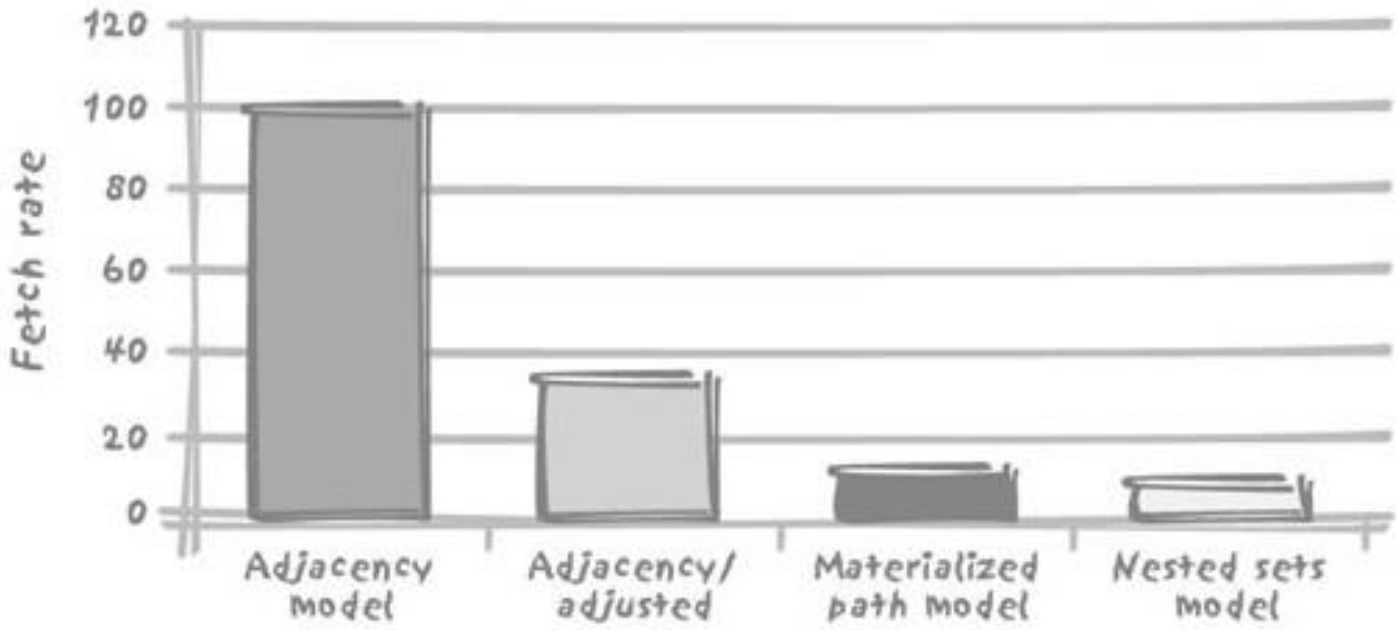
```
DESCRIPTION                       COMMANDER
 --------------------------------- ---------------------------------------
 1/92nd (Highland) Rgmt of Foot    Lt-Colonel John Cameron
 1/42nd (Highland) Rgmt of Foot    Colonel Sir Robert Macara (d.16th)
  9th British Brigade              Major-General Sir Denis Pack
 1/79th (Highland) Rgmt of Foot    Lt-Colonel Neil Douglas
  8th British Brigade              Lt-General Sir James Kempt
   5th Anglo-German Division       Lt-General Sir Thomas Picton (d.18th)
    General Reserve                Duke of Wellington
 1/71st (Highland) Rgmt of Foot    Lt-Colonel Thomas Reynell
  British Light Brigade            Major-General Frederick Adam
   2nd Anglo-German Division       Lt-General Sir Henry Clinton
    II Corps                       Lieutenant-General Lord Rowland Hill
 2/73rd (Highland) Rgmt of Foot    Lt-Colonel William George Harris
  5th British Brigade              Major-General Sir Colin Halkett
   3rd Anglo-German Division       Lt-General Count Charles von Alten
    I Corps                        Prince William of Orange
     The Anglo-Allied Army of 1815 Field Marshal Arthur Wellesley, Duke of
                                   Wellington

 16 rows selected.
```

嵌套集合模型

- 动态计算深度依旧是个问题
- 不要显示人造根节点
- 硬编码最大深度（为了缩排显示）

```
select lpad(description, length(description) + 6 - depth) description,
       commander
from (select distinct b.description,
                      b.commander,
                      b.left_num,
                      (select count(c.left_num)
                       from nested_sets_model c
                       where b.left_num between c.left_num
                                            and c.right_num) depth
      from nested_sets_model a,
           nested_sets_model b
      where a.description like '%Highland%'
       and a.left_num between b.left_num and b.right_num
       and b.left_num > 1)
order by left_num desc
```

比较各种模型下的Highland查询



一些问题

- 物化路径不该是KEY，即使他们有唯一性
- 物化路径不该暗示任何兄弟节点的排序
- 所选择的编码方式不需要完全中立

对保存于叶节点中的值做聚合

- 为人数建模
  - 叶节点包含更多的信息
  - 采用先前的例子，并限定法国第三军，构建UNITS表，记录一个团、一个师或一个旅，不包含关联

ID	NAME	COMMANDER
1	III Corps	Général de Division Dominique Vandamme
2	8th Infantry Division	Général de Division Baron Etienne-Nicolas Lefol
3	1st Brigade	Général de Brigade Billard
4	2nd Brigade	Général de Brigade Baron Corsin
5	10th Infantry Division	Général de Division Baron Pierre-Joseph Habert
6	1st Brigade	Général de Brigade Baron Gengoult
7	2nd Brigade	Général de Brigade Baron Dupeyroux
8	11th Infantry Division	Général de Division Baron Pierre Berthézène
9	1st Brigade	Général de Brigade Baron Dufour
10	2nd Brigade	Général de Brigade Baron Logarde
11	3rd Light Cavalry Division	Général de Division Baron Jean-Simon Domont
12	1st Brigade	Général de Brigade Baron Dommanget
13	2nd Brigade	Général de Brigade Baron Vinot
14	Reserve Artillery	Général de Division Baron Jérôme Doguereau

关联关系的存储

UNIT_LINK_ADJACENCY

ID	PARENT_ID
2	1
3	2
4	2
5	1
6	5
7	5
8	1
9	8
10	8
11	1
12	11
13	11
14	1

UNIT_LINKS_PATH

ID	PATH
1	1
2	1.1
3	1.1.1
4	1.1.2
5	1.2
6	1.2.1
7	1.2.2
8	1.3
9	1.3.1
10	1.3.2
11	1.4
12	1.4.1
13	1.4.2
14	1.5

unit_strength

ID	MEN
3	2952
4	2107
6	2761
7	2823
9	2488
10	2050
12	699
13	318
14	152

计算每一层的人数

- 对于邻接模型
  - 第三军的总人数

```
select sum(men)
from unit_strength
where id in (select id
        from unit_links_adjacency
        connect by prior id = parent_id
        start with parent_id = 1)
```

  - 每级战斗单位的人数

```
select u.name,
      u.commander,
      (select sum(men)
       from unit_strength
       where id in (select id
              from unit_links_adjacency
              connect by parent_id = prior id
              start with parent_id = u.id)
        or id = u.id) men
from units u
```

计算每一层的人数

- 对于物化路径模型
  - 第一步，构建视图EXPLODED_LINKS_PATH

```
SQL> select * from exploded_links_path;

        ID   ANCESTOR      DEPTH
---------- ---------- ----------
        14          1          1
        13          1          2
        12          1          2
        11          1          1
        10          1          2
         9          1          2
         8          1          1
         7          1          2
         6          1          2
         5          1          1
         4          1          2
         3          1          2
         2          1          1
         4          2          1
         3          2          1
         7          5          1
         6          5          1
        10          8          1
         9          8          1
        13         11          1
        12         11          1
```

计算每一层的人数

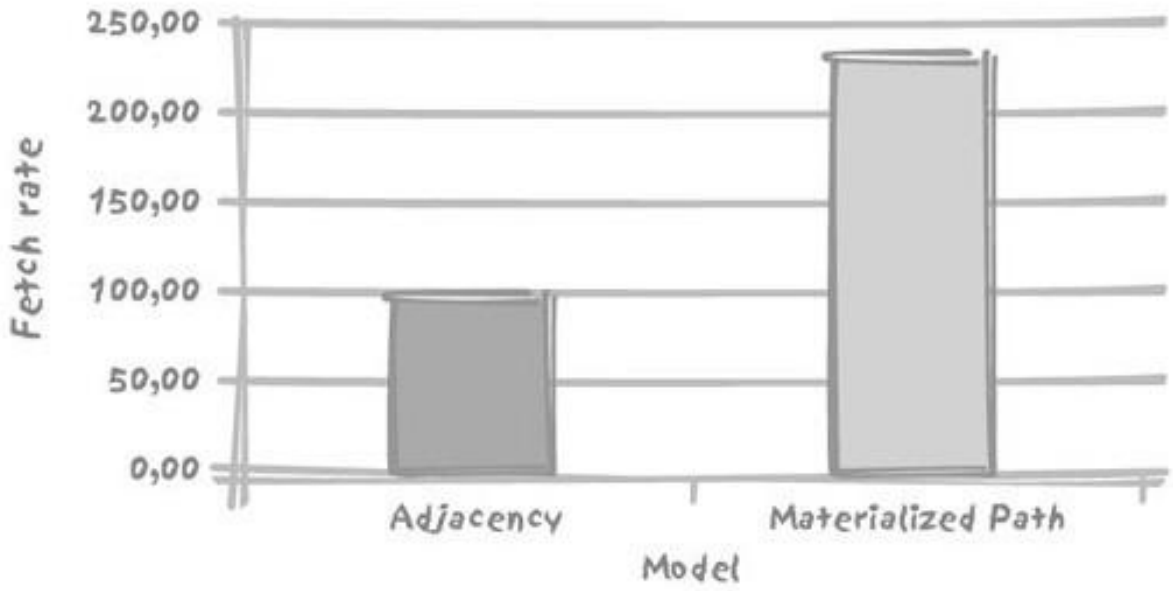
- 然后做sum操作，不考虑叶节点

```
select u.name, u.commander, sum(s.men) men
 from units u,
     exploded_links_path el,
     unit_strength s
 where u.id = el.ancestor
   and el.id = s.id
 group by u.name, u.commander
```

```
NAME                      COMMANDER                                MEN
------------------------- ------------------------------------- -----
III Corps                 Général de Division Dominique Vandamme 16350
8th Infantry Division     Général de Division Baron Etienne-      5059
                          Nicolas Lefol
10th Infantry Division    Général de Division Baron Pierre        5584
                          Joseph Habert
11th Infantry Division    Général de Division Baron Pierre        4538
                          Berthézène
3rd Light Cavalry Division Général de Division Baron Jean-Simon   1017
                          Domont
```

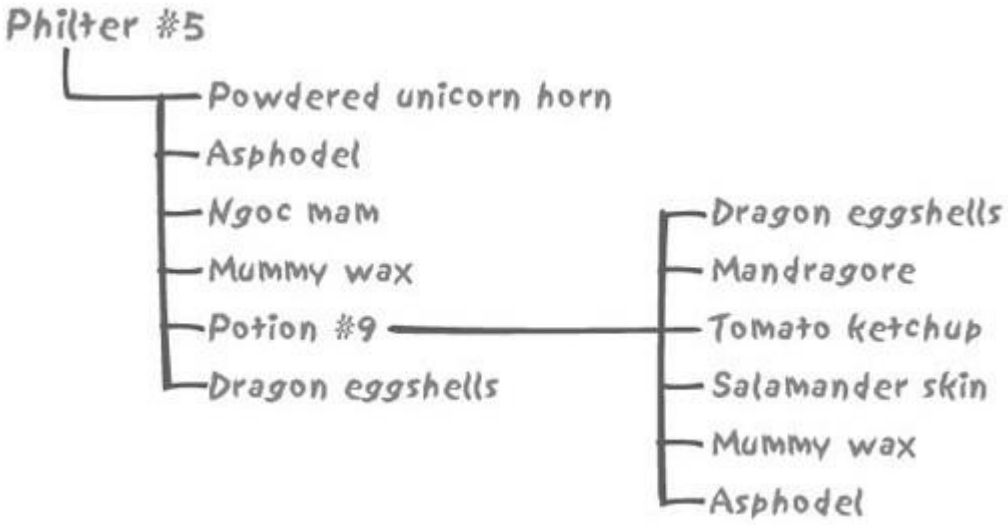
计算每一层的人数

- 执行查询5000次，比较单位时间返回的记录数



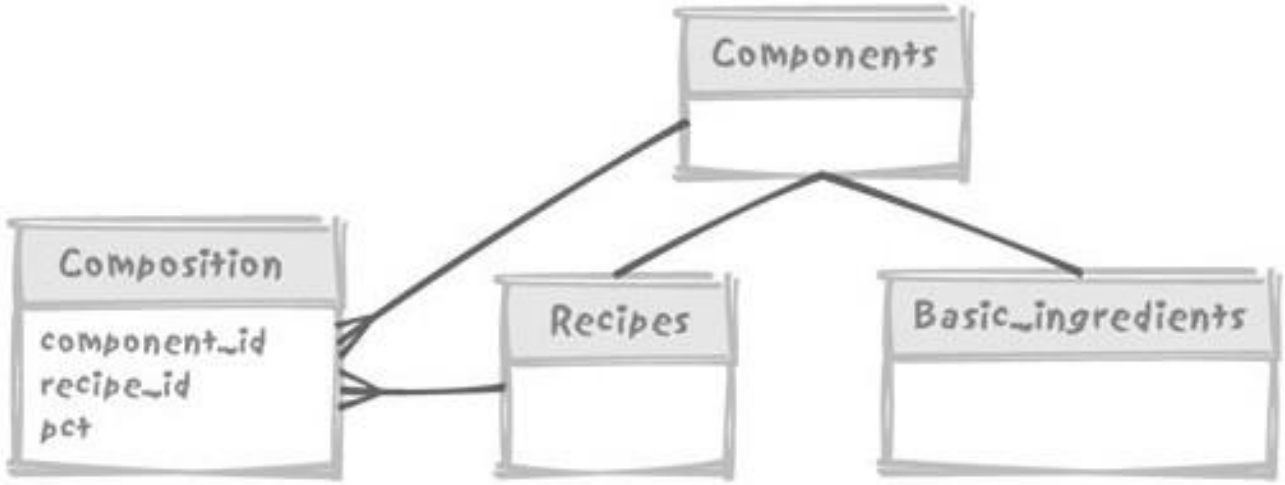
散布在各层的百分比

- 用SQL处理树结构仍有一些局限性
- 假设我们经营魔药。每种魔药由多种成分（ingredient）组成，处方（recipe）列出成分及百分比。处方可以共享某种“基础魔药”，以复合成分（compound ingredient）的形式表示。



散布在各层的百分比

- 某一种可以选择的建模方法
- Components表为通用类型
- 它有recipes和basic_ingredients两种子类型
- Composition表保存处方成分（可以是处方或基本成分及其数量）



散布在各层的百分比

- Connect by 等方法很难使用，由于connect by操作符的过程性本质，我们只能包含两个层次，这虽然对于前面的例子来说已经足够，但不适用于一般情况。

```
SQL> select connect_by_root recipe_id root_recipe,
  2         recipe_id,
  3         prior pct,
  4         pct
  5         component_id
  6  from composition
  7  connect by recipe_id = prior component_id
  8  /

ROOT_RECIPE  RECIPE_ID   PRIORPCT          PCT COMPONENT_ID
----------- ---------- ---------- ------------ ------------
         14         14                      5            3
         14         14                     20            7
         14         14                     15            8
         14         14                     30            9
         14         14                     20           10
         14         14                     10            2
         15         15                     30           14
         15         14         30           5            3
         15         14         30          20            7
         15         14         30          15            8
         15         14         30          30            9
...
```

散布在各层的百分比

- “处方中包含的各种成分，百分比是多少？”是个复杂的问题，用递归with反而是容易的事情。

```
with recursive_composition(actual_pct, component_id)
as (select a.pct,
           a.component_id
    from composition a,
         components b
    where b.component_id = a.recipe_id
      and b.component_name = 'Philter #5'
    union all
    select parent.pct * child.pct,
           child.component_id
    from recursive_composition parent,
         composition child
    where child.recipe_id = parent.component_id)
```

散布在各层的百分比

- 假设components表中有个component_type字段，包含代表基本成分的“I”和代表处方的“R”。最终，查询把处方过滤掉，而且，由于同样的基本成分可以出现在不同层次，所以要以成分做聚合：

```
select x.component_name, sum(y.actual_pct)
from recursive_composition y,
       components x
where x.component_id = y.component_id
    and x.component_type = 'I'
group by x.component_name
```

树状结构的问题

- 本章的方法，在数据量很少的情况下效果令人满意
- 对大数据量的处理“像老爷车一样慢”

- 同样可以采用非规范化模型、或基于触发器的扁平化数据模型。

- 不建议对关系模型“屡遭诟病的缓慢本性”反规范化，这很容易遮掩程序设计中的问题。
- 不过，SQL确实缺乏处理树结构的强大的、可伸缩的手段。