**ETL过程设计**

**赵辰 肖强**

目录

[一、挂号数量指标分析 3](#_Toc318379230)

[1 需要数据源表 3](#_Toc318379231)

[2 ETL数据流任务设计 4](#_Toc318379232)

[2.1 门急诊标识维表（原表可复用）（不要了） 4](#_Toc318379233)

[2.2 科室代码维表 5](#_Toc318379234)

[2.3 员工基本信息维表 5](#_Toc318379235)

[2.4 挂号数量指标事实表1（下钻到科室） 6](#_Toc318379236)

[2.5 挂号数量指标事实表2（下钻到员工） 7](#_Toc318379237)

[二、收费指标分析 8](#_Toc318379238)

[1 需要数据源表 8](#_Toc318379239)

[2 ETL数据流任务设计 9](#_Toc318379240)

[2.1科室代码维表 9](#_Toc318379241)

[2.2员工基本信息维表 10](#_Toc318379242)

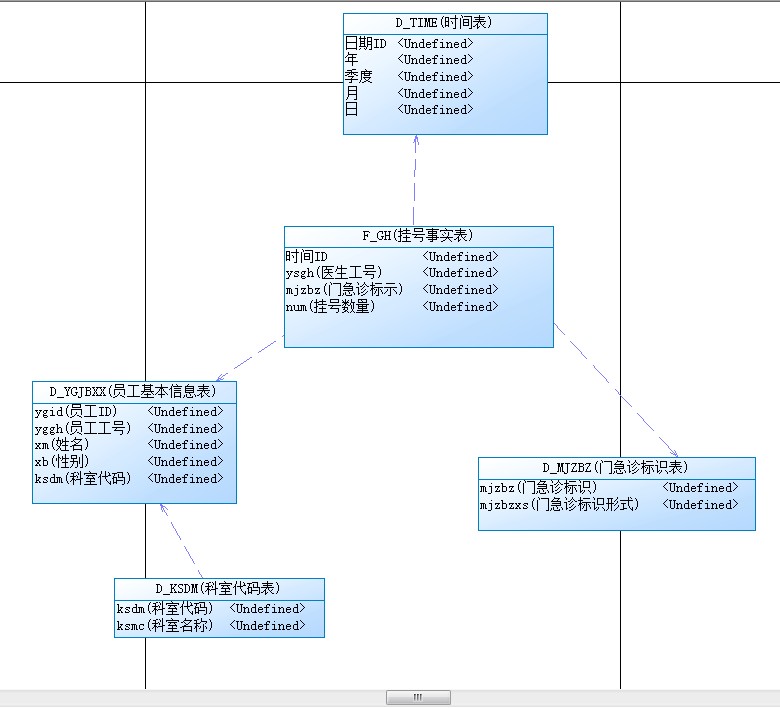
[2.3身份单列代码表 10](#_Toc318379243)

[2.4身份单列代码表 11](#_Toc318379244)

[2.5 收费指标事实表1（下钻到科室） 12](#_Toc318379245)

[2.6 收费指标事实表2（下钻到员工） 13](#_Toc318379246)

# 一、挂号数量指标分析



## 1 需要数据源表

**D\_KSDM（科室代码表）**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| ksdm | Nvachar(7) | 科室代码 | 4430100代表‘急诊内科’，  4060300代表‘急诊神内’，  0110700代表‘专家门诊’ |
| ksmc | Nvachar(20) | 科室名称 |  |

**D\_MJZBZ 门急诊标志表（不要了）**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| mjzbz | char(1) | 门急诊标志 |  |
| mjzbzxs | varchar(10) | 门急诊标志形式 | 1-门诊，2-急诊 |

**D\_XB 性别表**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| Xb | samllint | 性别 |  |
| Xbmc | Varchar（10） | 性别名称 | 1-🡪男，2-🡪女 |

**D\_YGJBXX 员工基本信息**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| Ygid | nvarchar（14） | 员工ID |  |
| Yggh | nvarchar（5） | 员工工号 |  |
| Xm | nvarchar（20） | 姓名 |  |
| Xb | nvarchar（1） | 性别 |  |
| Ksdm | nvarchar（7） | 科室代码 |  |

**M\_GH 挂号表**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| Key\_time | int | 挂号日期 |  |
| Ghsj | datetime | 挂号时间 | 不需要抽取 |
| Ghlsh | nvarchar（14） | 挂号流水号 | 不需要抽取 |
| Xm | nvarchar（30） | 姓名 | 不需要抽取 |
| Xb | nvarchar（1） | 性别 | 不需要抽取 |
| Nl | smallint | 年龄 | 不需要抽取 |
| Sfsfdm | nvarchar（3） | 收费身份代码 | 不需要抽取 |
| Ghlx | nvarchar（1） | 挂号类型 | 不需要抽取 |
| Ksdm | nvarchar（7） | 科室代码 |  |
| Ysgh | nvarchar（5） | 医生工号 |  |
| Mjzbz | nvarchar（1） | 门急诊标志 | 不需要抽取 |

## ETL数据流任务设计

### 门急诊标识维表（原表可复用）（不要了）

数据源：

**D\_MJZBZ 门急诊标志表**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| mjzbz | char(1) | 门急诊标志 |  |
| mjzbzxs | varchar(10) | 门急诊标志形式 | 1-门诊，2-急诊 |

1. 复制和转换

复制mjzbz字段和mjzbzxs字段

1. 导出到数据仓库

使用SQL语句在数据仓库中建立新的数据维表

Create table Dim\_MJZBZ(

Id integer Identity(1,1) as primary\_key,

mjzbz char(1),

mjzbzxs varchar(10)

)

### 科室代码维表

数据源：

**D\_KSDM（科室代码表）**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| ksdm | Nvarchar(7) | 科室代码 | 4430100代表‘急诊内科’，  4060300代表‘急诊神内’，  0110700代表‘专家门诊’ |
| ksmc | Nvarchar(20) | 科室名称 |  |

1. 复制和转换

复制字段ksdm和字段ksmc，并进行数据转换：

Ksdm: nvarchar(7)->char(7)

Ksmc: nvarchar(20)->varchar(20)

1. 导出到数据仓库

使用SQL语句在数据仓库中建立新的数据维表

Create table Dim\_KSDM(

Id integer Identity(1,1) as primary\_key,

ksdm char(7),

ksmc varchar(20)

)

### 员工基本信息维表

数据源：

**D\_YGJBXX 员工基本信息**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| Ygid | nvarchar（14） | 员工ID |  |
| Yggh | nvarchar（5） | 员工工号 |  |
| Xm | nvarchar（20） | 姓名 |  |
| Xb | nvarchar（1） | 性别 |  |
| Ksdm | nvarchar（7） | 科室代码 |  |

1. 复制和转换

复制字段ygid，字段yggh，字段xm，字段xb和字段ksdm，并进行数据转换：

ygid: nvarchar(14)->char(14);

yggh: nvarchar(5)->char(5);

xm: nvarchar(20)->varchar(20);

xb: nvarchar(1)->char(1);

ksdm: nvarchar(7)->char(7)

1. 查找外键

查找和科室关联的外键：

D\_YGJBXX.ksdm->Dim\_KSDM. Ksdm

1. 导出到数据仓库

使用SQL语句在数据仓库中建立新的数据维表

Create table Dim\_YGJBXX (

Id integer Identity(1,1) as primary\_key,

ygid char(14),

yggh char(5,)

xm varchar(20),

xb char(1),

ksdm char(7),

constraint ksdm\_fk foreign key (ksdm) references Dim\_KSDM (ksdm)

)

### 挂号数量指标事实表1（下钻到科室）

数据源：

**M\_GH 挂号表**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| Key\_time | int | 挂号日期 |  |
| Ghsj | datetime | 挂号时间 | 不需要抽取 |
| Ghlsh | nvarchar（14） | 挂号流水号 | 不需要抽取 |
| Xm | nvarchar（30） | 姓名 | 不需要抽取 |
| Xb | nvarchar（1） | 性别 | 不需要抽取 |
| Nl | smallint | 年龄 | 不需要抽取 |
| Sfsfdm | nvarchar（3） | 收费身份代码 | 不需要抽取 |
| Ghlx | nvarchar（1） | 挂号类型 | 不需要抽取 |
| Ksdm | nvarchar（7） | 科室代码 |  |
| Ysgh | nvarchar（5） | 医生工号 | 不需要抽取 |
| Mjzbz | nvarchar（1） | 门急诊标志 | 不需要抽取 |

1. 复制和转换

复制字段Key\_time和字段ksdm，并进行转换：

ksdm: nvarchar(7)->char(7)

1. 派生列

ghsj = (Key\_time / 1000).toString() + “-” + (Key\_time / 100 % 100).toString() + “-” + (Key\_time % 100)

1. 转换

ghsj: varchar->dateTime

1. 查找外键

M\_GH.ksdm->Dim\_KSDM.ksdm

1. 聚合

gh\_num: count

Group by ghsj;

Group by ksdm;

1. 导出到数据仓库

使用SQL语句在数据仓库中建立新的数据维表

Create table Fact\_GH1 (

Id integer Identity(1,1) as primary\_key,

ghsj dateTime,

ksdm char(7),

gh\_num numric(12),

constraint ksdm\_fk foreign key (ksdm) references Dim\_KSDM (ksdm)

)

### 挂号数量指标事实表2（下钻到员工）

数据源：

**M\_GH 挂号表**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| Key\_time | int | 挂号日期 |  |
| Ghsj | datetime | 挂号时间 | 不需要抽取 |
| Ghlsh | nvarchar（14） | 挂号流水号 | 不需要抽取 |
| Xm | nvarchar（30） | 姓名 | 不需要抽取 |
| Xb | nvarchar（1） | 性别 | 不需要抽取 |
| Nl | smallint | 年龄 | 不需要抽取 |
| Sfsfdm | nvarchar（3） | 收费身份代码 | 不需要抽取 |
| Ghlx | nvarchar（1） | 挂号类型 | 不需要抽取 |
| Ksdm | nvarchar（7） | 科室代码 | 不需要抽取 |
| Ysgh | nvarchar（5） | 医生工号 |  |
| Mjzbz | nvarchar（1） | 门急诊标志 | 不需要抽取 |

1. 复制和转换

复制字段Key\_time和字段ysgh，并进行转换：

ysgh: nvarchar(5)->char(5)

1. 派生列

ghsj = (Key\_time / 1000).toString() + “-” + (Key\_time / 100 % 100).toString() + “-” + (Key\_time % 100)

1. 转换

ghsj: varchar->dateTime

1. 查找外键

M\_GH. ysgh ->Dim\_YGJBXX. ysgh

1. 聚合

gh\_num: count

Group by ghsj;

Group by ysgh;

1. 导出到数据仓库

使用SQL语句在数据仓库中建立新的数据维表

Create table Fact\_GH2 (

Id integer Identity(1,1) as primary\_key,

ghsj dateTime,

ysgh char(5),

gh\_num numeric(12),

constraint ksdm\_fk foreign key (ysgh) references Dim\_YGJBXX (ysgh)

)

# 二、收费指标分析



## 1 需要数据源表

**D\_KSDM（科室代码表）**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| ksdm | Nvachar(7) | 科室代码 | 4430100代表‘急诊内科’，  4060300代表‘急诊神内’，  0110700代表‘专家门诊’ |
| ksmc | Nvachar(20) | 科室名称 |  |

**D\_YGJBXX 员工基本信息**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| Ygid | nvarchar（14） | 员工ID |  |
| Yggh | nvarchar（5） | 员工工号 |  |
| Xm | nvarchar（20） | 姓名 |  |
| Xb | nvarchar（1） | 性别 |  |
| Ksdm | nvarchar（7） | 科室代码 |  |

**D\_SFDLDM 身份单列代码表**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| Sfdldm | nvarchar（2） | 收费单列代码 |  |
| Sfdlmc | nvarchar（12） | 收费单列名称 |  |
| Bzw | nvarchar（1） | 标志位 |  |
| Ybmzdldm | nvarchar（2） | 医保名字大类代码 |  |
| Ybzydldm | nvarchar（2） | 医保？？ |  |

**D\_SFSFB 收费身份表**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| Sfsfdm | nvarchar（3） | 收费身份代码 |  |
| Sfmc | nvarchar（30） | 身份名称 |  |
| Yblx | nvarchar（1） | 医保类型 |  |
| Bjqk | nvarchar（1） | 报价情况 |  |
| Ztdm | nchar（1） | 状态代码 |  |

**M\_SF 收费表**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| Rq | nvarchar（8） | 日期 |  |
| mjzks | nvarchar（7） | 门急诊科室 |  |
| Mjzbz | nvarchar（1） | 门急诊标志 | 无需抽取 |
| Ysid | nvarchar（14） | 医生id |  |
| Sfsfdm | nvarchar（3） | 收费身份代码 | 无需抽取 |
| Sfdldm | nvarchar（2） | 收费大类代码 | 无需抽取 |
| Je | numeric（12,2） | 金额 |  |
| flzfje | numeric（12,2） | 福利支付金额 | 无需抽取 |
| zfje | numeric（12,2） | 自费金额 | 无需抽取 |

## ETL数据流任务设计

### 2.1科室代码维表

数据源：

**D\_KSDM（科室代码表）**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| ksdm | Nvarchar(7) | 科室代码 | 4430100代表‘急诊内科’，  4060300代表‘急诊神内’，  0110700代表‘专家门诊’ |
| ksmc | Nvarchar(20) | 科室名称 |  |

(1)复制和转换

复制字段ksdm和字段ksmc，并进行数据转换：

Ksdm: nvarchar(7)->char(7)

Ksmc: nvarchar(20)->varchar(20)

(2)导出到数据仓库

使用SQL语句在数据仓库中建立新的数据维表

Create table Dim\_KSDM(

Id integer Identity(1,1) as primary\_key,

ksdm char(7),

ksmc varchar(20)

)

### 2.2员工基本信息维表

数据源：

**D\_YGJBXX 员工基本信息**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| Ygid | nvarchar（14） | 员工ID |  |
| Yggh | nvarchar（5） | 员工工号 |  |
| Xm | nvarchar（20） | 姓名 |  |
| Xb | nvarchar（1） | 性别 |  |
| Ksdm | nvarchar（7） | 科室代码 |  |

(1)复制和转换

复制字段ygid，字段yggh，字段xm，字段xb和字段ksdm，并进行数据转换：

ygid: nvarchar(14)->char(14);

yggh: nvarchar(5)->char(5);

xm: nvarchar(20)->varchar(20);

xb: nvarchar(1)->char(1);

ksdm: nvarchar(7)->char(7)

(2)查找外键

查找和科室关联的外键：

D\_YGJBXX.ksdm->Dim\_KSDM. Ksdm

(3)导出到数据仓库

使用SQL语句在数据仓库中建立新的数据维表

Create table Dim\_YGJBXX (

Id integer Identity(1,1) as primary\_key,

ygid char(14),

yggh char(5,)

xm varchar(20),

xb char(1),

ksdm char(7),

constraint ksdm\_fk foreign key (ksdm) references Dim\_KSDM (ksdm)

)

### 2.3身份单列代码表

数据源：

**D\_SFDLDM 身份单列代码表**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| sfdldm | nvarchar（2） | 收费单列代码 |  |
| sfdlmc | nvarchar（12） | 收费单列名称 |  |
| bzw | nvarchar（1） | 标志位 |  |
| ybmzdldm | nvarchar（2） | 医保名字大类代码 |  |
| ybzydldm | nvarchar（2） | 医保？？ |  |

(1)复制和转换

复制字段sfdldm，字段sfdlmc，字段bzw，字段ybmzdldm和字段ybzydldm，并进行数据转换：

sfdldm: nvarchar(2)->char(2);

sfdlmc: nvarchar(12)->varchar(12);

bzw: nvarchar（1）->char(1);

ybmzdldm: nvarchar(2)-> char(2);

ybzydldm: nvarchar(2)->char(2);

(2)导出到数据仓库

使用SQL语句在数据仓库中建立新的数据维表

Create table Dim\_SFDLDM (

Id integer Identity(1,1) as primary\_key,

sfdldm char(2),

sfdlmc varchar(12)

bzw char(1),

ybmzdldm char(2),

ybzydldm char(2),

)

### 2.4身份单列代码表

数据源

**D\_SFSFB 收费身份表**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| sfsfdm | nvarchar（3） | 收费身份代码 |  |
| sfmc | nvarchar（30） | 身份名称 |  |
| yblx | nvarchar（1） | 医保类型 |  |
| bjqk | nvarchar（1） | 报价情况 |  |
| ztdm | nchar（1） | 状态代码 |  |

(1)复制和转换

复制字段sfsfdm，字段sfmc，字段yblx，字段bjqk和字段ztdm，并进行数据转换：

sfsfdm: nvarchar(3)->char(3);

sfmc: nvarchar(30)->varchar(30);

yblx: nvarchar（1）->char(1);

bjqk: nvarchar(1)-> char(1);

ztdm: nchar(1)->char(1);

(2)导出到数据仓库

使用SQL语句在数据仓库中建立新的数据维表

Create table Dim\_SFSFB (

Id integer Identity(1,1) as primary\_key,

sfsfdm char(3),

sfmc varchar(30)

yblx char(1),

bjqk char(1),

ztdm char(1)

)

### 2.5 收费指标事实表1（下钻到科室）

数据源

**M\_SF 收费表**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| rq | nvarchar（8） | 日期 |  |
| mjzks | nvarchar（7） | 门急诊科室 |  |
| mjzbz | nvarchar（1） | 门急诊标志 | 不需抽取 |
| ysid | nvarchar（14） | 医生id | 不需抽取 |
| sfsfdm | nvarchar（3） | 收费身份代码 | 不需抽取 |
| sfdldm | nvarchar（2） | 收费大类代码 | 不需抽取 |
| je | numeric（12,2） | 金额 |  |
| flzfje | numeric（12,2） | 福利支付金额 | 不需抽取 |
| zfje | numeric（12,2） | 自费金额 | 不需抽取 |

(1)复制和转换

复制字段rq, 字段mjzks和字段je，并进行转换：

mjzks: nvarchar(7)->char(7)

(2)派生列

sfsj = (rq / 1000).toString() + “-” + (rq / 100 % 100).toString() + “-” + (rq % 100)

(3)转换

sfsj: varchar->dateTime

(4)查找外键

M\_SF. ksdm ->Dim\_KSDM. ksdm

(5)聚合

je: sum

Group by sfsj;

Group by ksdm;

(6)导出到数据仓库

使用SQL语句在数据仓库中建立新的数据维表

Create table Fact\_SF1 (

Id integer Identity(1,1) as primary\_key,

sfsj dateTime,

ksdm char(7),

je numeric(12,2),

constraint ksdm\_fk foreign key (ksdm) references Dim\_KSDM (ksdm)

)

### 2.6 收费指标事实表2（下钻到员工）

数据源

**M\_SF 收费表**

|  |  |  |  |
| --- | --- | --- | --- |
| 列名 | 类型 | 描述 | 备注 |
| rq | nvarchar（8） | 日期 |  |
| mjzks | nvarchar（7） | 门急诊科室 | 不需抽取 |
| mjzbz | nvarchar（1） | 门急诊标志 | 不需抽取 |
| ysid | nvarchar（14） | 医生id |  |
| sfsfdm | nvarchar（3） | 收费身份代码 | 不需抽取 |
| sfdldm | nvarchar（2） | 收费大类代码 | 不需抽取 |
| je | numeric（12,2） | 金额 |  |
| flzfje | numeric（12,2） | 福利支付金额 | 不需抽取 |
| zfje | numeric（12,2） | 自费金额 | 不需抽取 |

(1)复制和转换

复制字段rq, 字段ysid和字段je，并进行转换：

ysid: nvarchar(14)->char(14)

(2)派生列

sfsj = (rq / 1000).toString() + “-” + (rq / 100 % 100).toString() + “-” + (rq % 100)

(3)转换

sfsj: varchar->dateTime

(4)查找外键

M\_SF. Yggh ->Dim\_ YGJBXX. Yggh

(5)聚合

Je:sum

Group by sfsj;

Group by Yggh;

(6)导出到数据仓库

使用SQL语句在数据仓库中建立新的数据维表

Create table Fact\_SF2 (

Id integer Identity(1,1) as primary\_key,

sfsj dateTime,

ysid char(14),

je numeric (12,2),

constraint ksdm\_fk foreign key (Yggh) references Dim\_ YGJBXX (Yggh)

)