

数据仓库设计

- 1. 对某高等学校部门的分析应用业务进行描述
 - 1.1 数据仓库整体业务背景
 - 1.2 各部门分析型业务描述和主题域分析
 - 1.2.1 教学部门
 - 1.2.2 人事部门
 - 1.2.3 资产管理部门
 - 1.2.4 体育部门
- 2. 对数据仓库进行设计
 - 2.1 数据仓库结构
 - 2.2 公共维度
 - 2.3 数据仓库总线矩阵
 - 2.4 主题分析
 - 2.4.1 课程安排
 - 2.4.2 课程选择
 - 2.4.3 人员招聘
 - 2.4.4 人员考核
 - 2.4.5 资源管理
 - 2.4.6 学生体质健康管理

数据仓库设计

191250133 陶泽华

1. 对某高等学校部门的分析应用业务进行描述

1.1 数据仓库整体业务背景

本数据仓库建立的目的是以统一的方式集中处理高校各个部门分散的数据，方便用户对信息的访问，提供分析和决策信息，提高对历史信息分析的性能。本数据仓库针对学校的的教学、资产管理、人事和体育这四个部门进行分析，其他部门不在本数据仓库的分析范围内。

1.2 各部门分析型业务描述和主题域分析

1.2.1 教学部门

- 1. 职责简述：教学部门负责进行课程安排、课程选择、课程考试安排等与课程相关的任务，记录每个课程的任课老师信息，学生的选课信息以及课程考试的相关信息。
- 2. 涉及主题：课程安排、课程选择
- 3. 主题：课程安排
 - 目的：方便教务部门的人员合理地安排课程，给同学们合理的生活和学习时间，防止课程冲突
 - 分析操作：对学校各个院系的各个专业分析得到各个种类的课程以及课程考试的详细安排。
 - 维度：时间、院系、课程、教师
- 4. 主题：课程选择
 - 目的：方便学生能够方便的选到自己想要学习的课程以及一些必修的课程，方便教务人员更好地安排课

程。

- 分析操作：分析各课程的受欢迎程度。
- 维度：时间、院系、课程、教师、学生

1.2.2 人事部门

1. 职责简述：人事部门负责学校的人员招聘以及考核等工作。
2. 涉及主题：人员招聘、人员考核
3. 主题：人员招聘
 - 目的：方便人事部门人员能够更好地完成人员招聘工作。
 - 分析操作：分析各个岗位的人员情况，得到需要招聘的职工类型以及数量。
 - 维度：院系、职工、教师、岗位
4. 主题：人员考核
 - 目的：帮助人事部门更好地完成人员的考核工作
 - 分析操作：分析得到绩效各个院系各个岗位上绩效最好的职工。
 - 维度：时间、院系、职工、教师、岗位、被考核人

1.2.3 资产管理部门

1. 职责简述：资产管理部门负责管理学校拥有的房产以及各种实验、教学设施等的使用和租借。
2. 涉及主题：资源管理
3. 主题：资源管理
 - 目的：方便资产管理部门的人员了解学校的资源使用情况并进行管理
 - 分析操作：分析得到资源使用情况
 - 维度：时间、资源、学生、职工、教师、院系

1.2.4 体育部门

1. 职责简述：体育部门负责完成对学生的体质健康测试等相关任务。
2. 涉及主题：学生体质健康信息管理
3. 主题：学生体质健康信息管理
 - 目的：方便体育部门的人员掌握学生的体质信息，帮助他们做一些体育活动安排方面的分析和决策。
 - 分析操作：分析得到学生的体质情况
 - 维度：时间、学生、院系、体质测试

2. 对数据仓库进行设计

2.1 数据仓库结构

本数据仓库采用总线结构设计数据仓库。

2.2 公共维度

时间维度	课程维度	学生维度	教师维度	职工维度	院系维度	岗位维度
------	------	------	------	------	------	------

时间维度：时间ID、日期

课程维度：课程ID、课程名称、任课教师、上课时间、课程地点、考试时间、考试地点

学生维度：学生ID、入学年份、学生姓名、性别、出生日期、身份证号、年级、院系、专业、联系电话

教师维度：教师ID、姓名、性别、联系电话、入职年份、院系、身份证号、职称

职工维度：职工ID、姓名、性别、联系电话、身份证号、入职年份

院系维度：院系ID、院系名称、专业、院系老师、院系学生

岗位维度：岗位ID、岗位名称、岗位工作人员ID

2.3 数据仓库总线矩阵

主题\公共维度	时间维度	课程维度	学生维度	教师维度	职工维度	院系维度	岗位维度
课程安排	✓	✓		✓		✓	
课程选择	✓	✓	✓	✓		✓	
人员招聘				✓	✓	✓	✓
人员考核	✓			✓	✓	✓	✓
资源管理	✓		✓	✓	✓	✓	
学生体质健康信息管理	✓		✓			✓	

2.4 主题分析

2.4.1 课程安排

1. 选取业务处理：

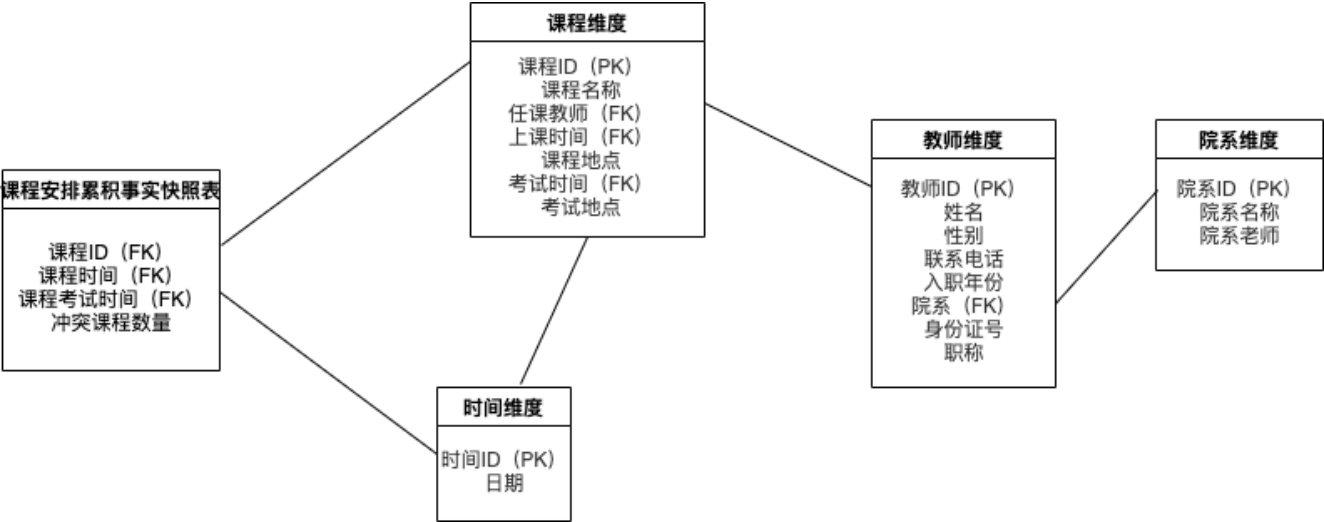
对学校各个院系的各个专业分析得到各个种类的课程以及课程考试的详细安排。
2. 定义粒度：

每一条信息对应一门课程。使用累积快照事务表，记录该课程的上课时间、考试时间、结课时间等等。
3. 选择维度：

时间、教师、院系、课程
4. 选择度量值：

冲突课程数量

5. 多维建模设计



6. 设计原因：

- 课程安排主题使用累积快照事务表，因为课程有明确的开始时间、结束时间以及考试时间，生命周期是确定的。另外课程冲突数量这一数据是一直更新的，且没有明确的周期。
- 课程安排主题选择使用雪花模型，因为学校课程的种类是非常多的，课程对应的院系以及课程涉及到的教师也非常多，数据量很大。使用雪花模型可以节省存储空间，对于性能的损失也是可以接受的。

2.4.2 课程选择

1. 选取业务处理：

根据学生的选课信息分析各课程的受欢迎程度。

2. 定义粒度：

每一行对应一条选课记录。使用累积快照事务表，记录相应课程的选课时间、选课学生ID等。

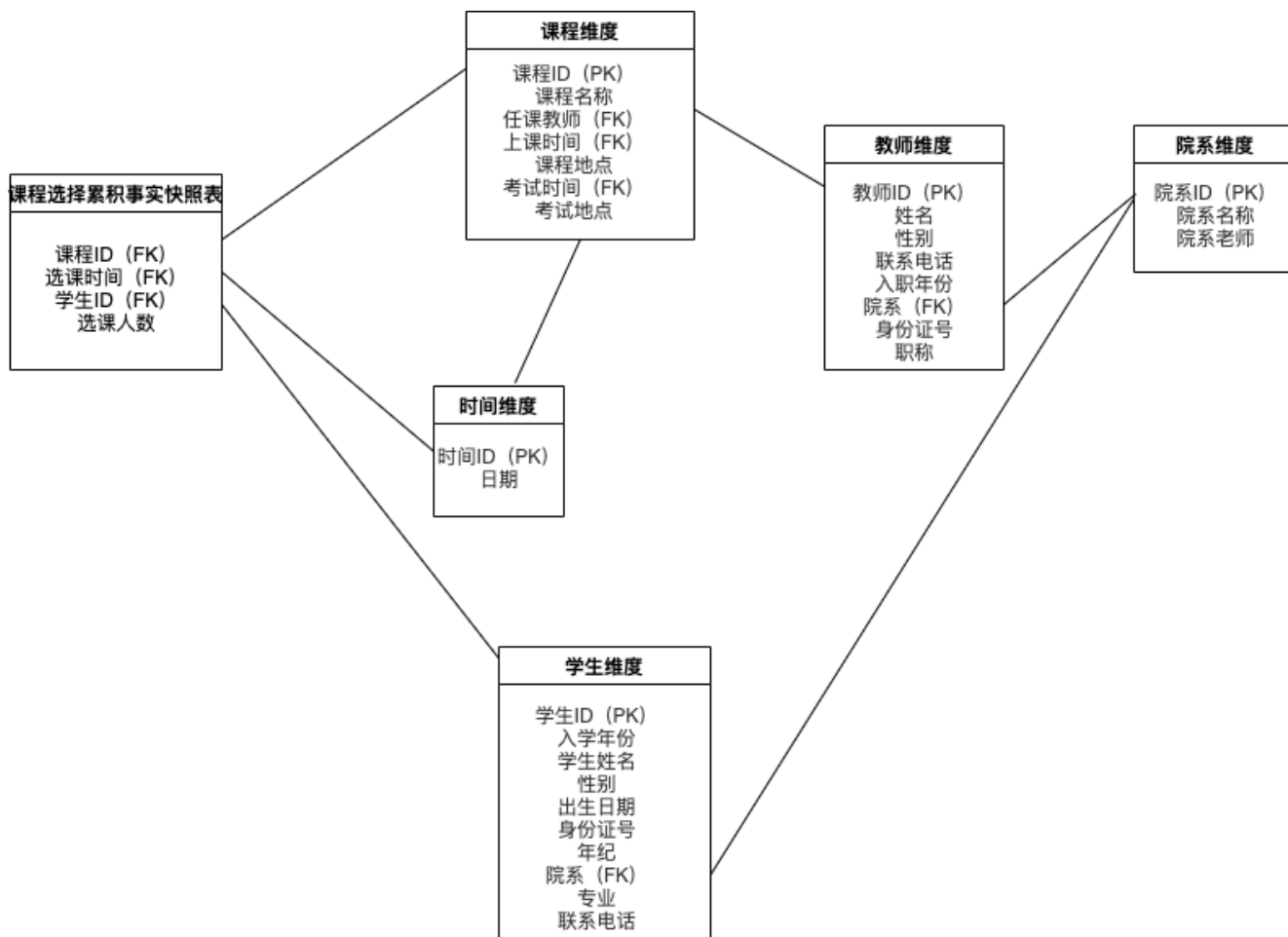
3. 选择维度：

时间、教师、院系、课程、学生

4. 选择度量值：

选课人数

5. 多维建模设计



6. 设计原因：

- 课程选择主题使用累积快照事务表，因为课程的选课人数会随着选课记录的增加而累加，数据会一直更新。课程是有明确的生命周期的。
- 课程选择主题选择使用雪花模型，因为学校课程的种类是非常多的，课程对应的院系以及教师也非常多，另外学生的数量也很多，这就会导致学生选课的数据量很大。使用雪花模型可以节省存储空间，对于性能的损失也是可以接受的。

2.4.3 人员招聘

1. 选取业务处理：

分析各个岗位的人员情况，得到需要招聘的职工类型以及数量。

2. 定义粒度：

每一行对应一个岗位的人员变动记录。使用累积快照事务表，记录相应岗位的人员总数。

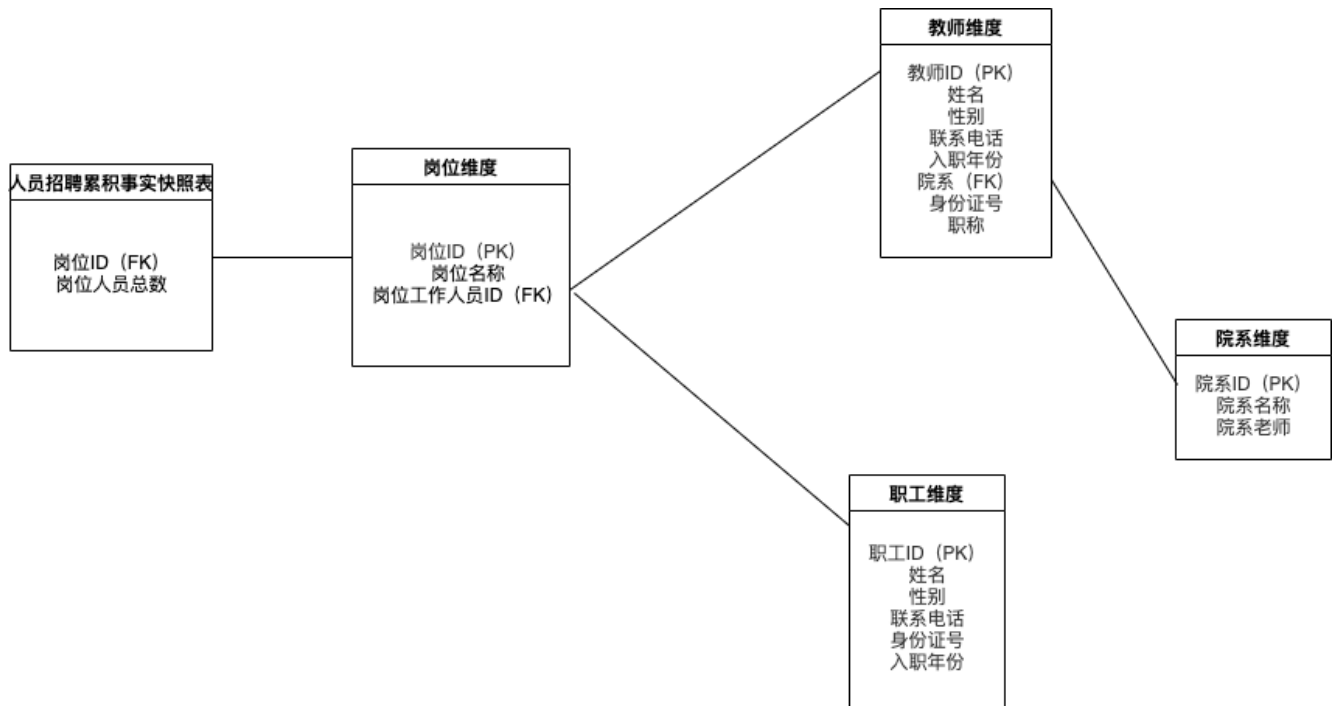
3. 选择维度：

院系、职工、教师、岗位

4. 选择度量值：

岗位人员总数

5. 多维建模设计



6. 设计原因：

- 人员招聘主题使用累积快照事务表，因为学校的教职工人数会随着人员招聘、学校的发展而逐渐增加，数据是一直更新的。
- 人员招聘主题选择使用雪花模型，因为学校教职工人数是非常多的，带来了大量的数据。且教师会属于不同的院系，关联关系复杂。使用雪花模型可以节省存储空间，避免数据冗余，对于性能的损失也是可以接受的。

2.4.4 人员考核

1. 选取业务处理：

分析得到绩效各个岗位上绩效最好的职工。

2. 定义粒度：

每一行对应被考评者的一次考评信息。使用事务事实表，保留原子数据，更加精确。

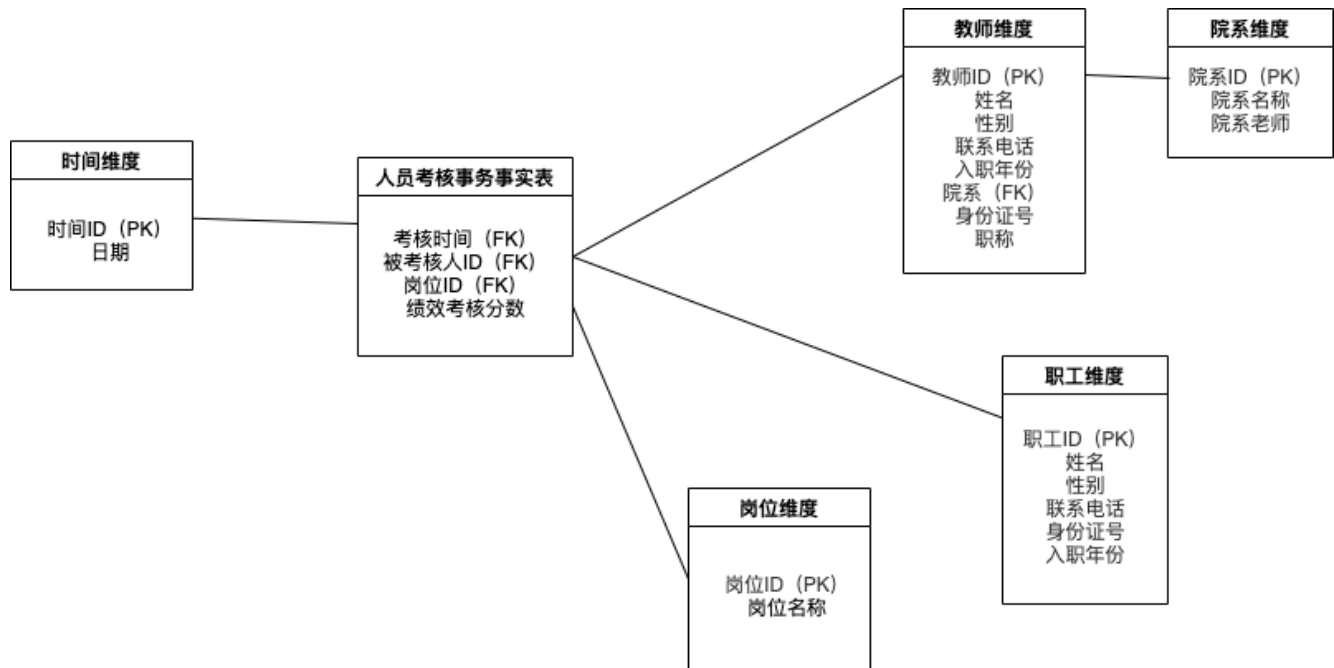
3. 选择维度：

时间、院系、职工、教师、岗位、被考核人

4. 选择度量值：

绩效考核评分

5. 多维建模设计



6. 设计原因：
- 人员考核主题使用事实事务表，保留每一次考核记录以及考核分数，更加详细也更加精确。
 - 人员考核主题选择使用雪花模型，因为绩效考核的数据量较大，且关联关系比较复杂。使用雪花模型可以节省存储空间，减少数据冗余，对于性能的损失也是可以接受的。

2.4.5 资源管理

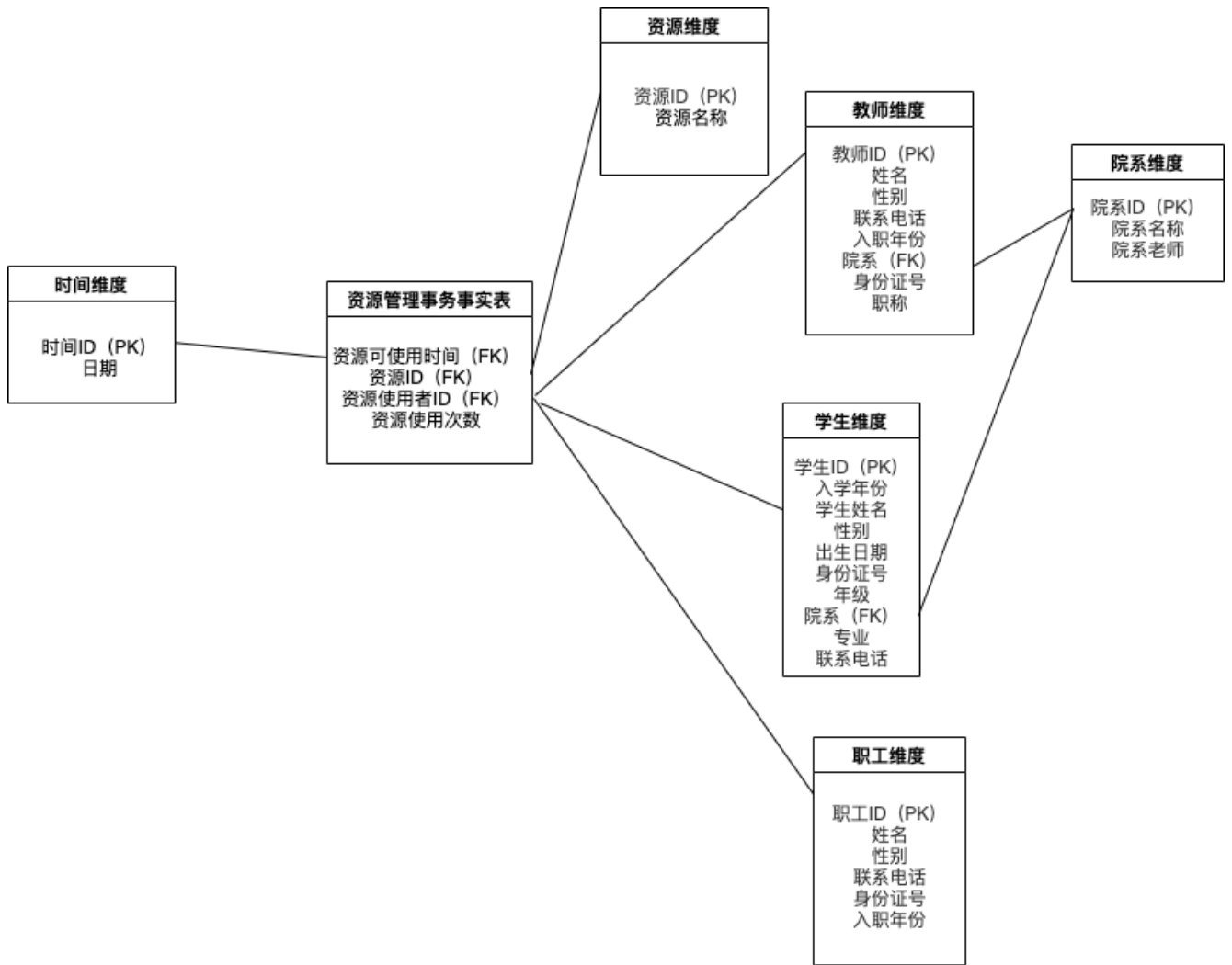
1. 选取业务处理：

分析得到学校资源的使用情况
2. 定义粒度：

每一行对应一次资源租借使用记录。选择使用事务事实表，保留原子数据，详细的记录每一次资源的租借使用记录方便后续查看，且更加精确。
3. 选择维度：

时间、资源、学生、职工、教师、院系
4. 选择度量值：

资源使用次数
5. 多维建模设计



6. 设计原因：

- 资源管理主题使用事实事务表，保留每一次资源的租借使用记录，更加详细也更加精确。
- 资源管理主题选择使用雪花模型，因为资源租借的数据量较大，且关联关系比较复杂。使用雪花模型可以节省存储空间，减少数据冗余，对于性能的损失也是可以接受的。

2.4.6 学生体质健康管理信息

1. 选取业务处理：

分析得到学生的体质情况

2. 定义粒度：

每一行对应一个学生的一次体质测试记录。选择使用事实事实表，保留原子数据，详细的记录每个学生的每次体质测试记录，更加精确。

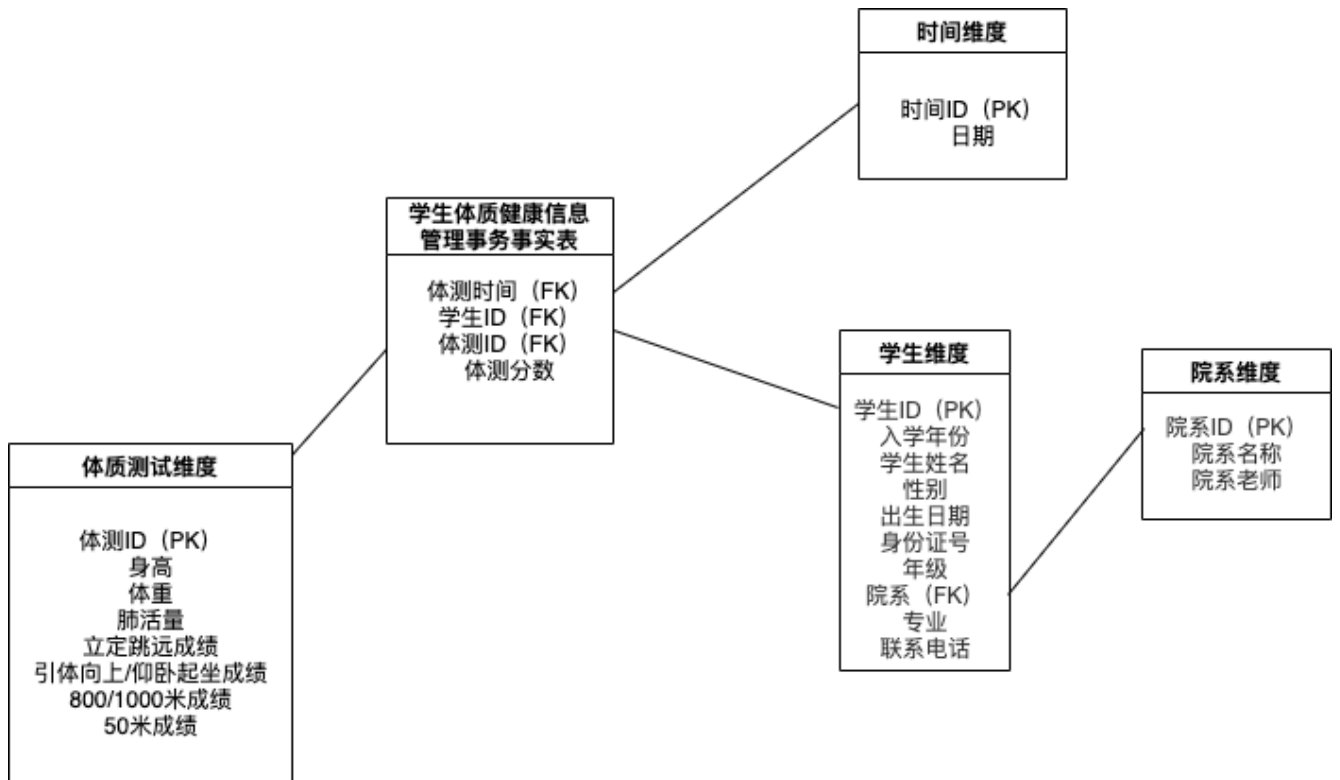
3. 选择维度：

时间、学生、院系、体质测试

4. 选择度量值：

体测分数

5. 多维建模设计



6. 设计原因：

- 学生体质健康信息管理主题使用事实事务表，保留每个学生的每次体质测试记录，更加详细也更加精确。
- 学生体质健康信息管理主题选择使用雪花模型，因为体质测试的数据量较大，且学生和院系之间的关联关系比较复杂。使用雪花模型可以节省存储空间，减少数据冗余。另外由于人员管理的数据存储相对集中，对于性能的损失也是可以接受的。