

# 系统架构设计师案例模拟题

一、希赛公司欲为某教育机构开发一个在线教育平台，支持用户完成网上下单购买课程，在线点播视频，在线测试等功能。在系统开发之初，企业对该平台提出了如下要求和质量场景：

- (a) 10000 用户并发点播视频时，视频不卡顿。
- (b) 用户口令在保存进数据库之前，必须进行 MD5 码处理。
- (c) 视频点播支持多平台，当一个平台出现故障后，30 秒内将用户请求切换至另一平台。
- (d) 平台支持的高清视频码率至少达到 3Mb/s。
- (e) 学员不必查看帮助文档，也能正常使用系统。
- (f) 学员信息数据库授权必须保证 99.9%可用。
- (g) 更改视频加密的方式将对安全性和性能产生影响。
- (h) 目前对学员学习情况分析功能尚未达成共识，这可能导致部分业务功能模块开发后期需要调整。
- (i) 用户名要求必须以字母开头，可以包含下划线和数字，长度不少于 6 个字符。
- (j) 系统要扩容时，应保证在 2 人•月内完成所有的部署与测试工作。
- (k) 对用户隐私的保护要求将影响系统的数据传输协议和处理过程的设计。
- (l) 系统对学员的操作行为有完整的日志记录。
- (m) 系统必须提供远程调试接口，并支持系统的远程调试。

(o) 更新系统的 Web 界面接口必须在 1 人月内完成。

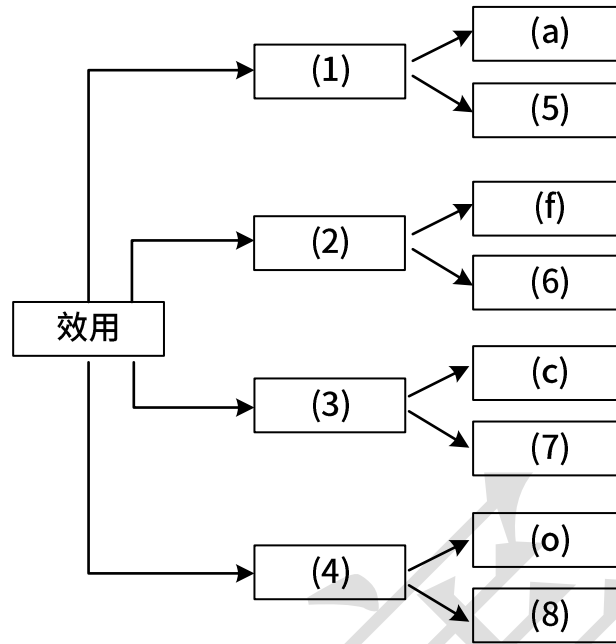
(p) 若每秒学员交卷请求的数量是 30 个，处理单个请求的时间为 20 毫秒，则系统保证在 1 秒内完成用户的请求是可以做到的。

(q) 当在线教育平台发生网络失效后，需要在 60 秒内发现错误并启用备用网络。

在对系统需求，质量属性描述和架构特性进行分析的基础上，系统的架构师给出了三个候选的架构设计方案，公司目前正在组织系统开发的相关人员对系统架构进行评估。

**【问题 1】 (9 分)**

在架构评估过程中，质量属性效用树 (utilitytree) 是对系统质量属性进行识别和优先级排序的重要工具。请给出合适的质量属性，填入图 1-1 中 (1) ~ (4) 空白处；并选择题干描述的 (a) ~ (q)，填入 (5) ~ (8) 空白处 (注：1 个空可能对应多个选项)，完成该系统的效用树。(评分规则：对于多选的空，少选按答对个数得分，错选不得分)



【问题 2】（4 分）

在架构设计过程中,除了需要考虑经典的 4 大质量属性,还要考虑其它属性,如:易用性,可测试性。题目中 (9) 是易用性的示例, (10) 是可测试性的示例。

【问题 3】（12 分）

在架构评估过程中;需要正确识别系统的架构风险点、非风险点、敏感点和权衡点,并进行合理的架构决策。请用 300 字以内的文字给出系统架构风险点、非风险点、敏感点和权衡点的定义,并从题干描述中的 (a) ~ (q) 各选出 1 个属于系统架构风险点、非风险点、敏感点和权衡点的描述。

答案:

【问题 1】（9 分）

(1) 性能 (2) 安全性 (3) 可用性 (4) 可修改性

(5) (d) (6) (b) (l) (7) (q) (8) (j)

**【问题 2】（4 分）**

(9) (e) (10) (m)

**【问题 3】（12 分，定义 1 分，实例 2 分）**

敏感点是一个或多个构件（和／或构件之间的关系）的特性，它能影响系统的某个质量属性。

权衡点是影响多个质量属性的特性，是多个质量属性的敏感点。

风险点是指架构设计中潜在的、存在问题的架构决策所带来的隐患。

非风险点是指不会带来隐患的分析与描述。

敏感点：(k) 权衡点：(g) 风险点：(h) 非风险点：(p)

解析：

(a) 10000 用户并发点播视频时，视频不卡顿。【性能】

(b) 用户口令在保存进数据库之前，必须进行 MD5 码处理。【安全性】

(c) 视频点播支持多平台，当一个平台出现故障后，30 秒内将用户请求切换至另一平台。【可用性】

(d) 平台支持的高清视频码率至少达到 3Mb/s。【性能】

(e) 学员不必查看帮助文档，也能正常使用系统。【易用性】

(f) 学员信息数据库授权必须保证 99.9%可用。【安全性】

(g) 更改视频加密的方式将对安全性和性能产生影响。【权衡点】

(h) 目前对学员学习情况分析功能尚未达成共识，这可能导致部分业务功能模块开发后期需要调整。【风险点】

(i) 用户名要求必须以字母开头，可以包含下划线和数字，长度不少于 6 个字符。【设计约束】

- (j) 系统要扩容时，应保证在 2 人•月内完成所有的部署与测试工作。【可修改性】
- (k) 对用户隐私的保护要求将影响系统的数据传输协议和处理过程的设计。【敏感点】
- (l) 系统对学员的操作行为有完整的日志记录。【安全性】
- (m) 系统必须提供远程调试接口，并支持系统的远程调试。【可测试性】
- (o) 更新系统的 Web 界面接口必须在 1 人月内完成。【可修改性】
- (p) 若每秒学员交卷请求的数量是 30 个，处理单个请求的时间为 20 毫秒，则系统保证在 1 秒内完成用户的请求是可以做到的。【非风险点】
- (q) 当在线教育平台发生网络失效后，需要在 60 秒内发现错误并启用备用网络。【可用性】

二、希赛公司的在线教育平台因业务扩展，访问量不断增大，现有系统访问速度缓慢，有时甚至出现系统故障瘫痪等现象。面对这一情况，公司召开项目组讨论会议，寻求该商务平台的改进方案。讨论会上，陈工提出可以利用镜像站点、CDN 内容分发等方式解决并发访问量带来的问题。而王工认为，仅仅依靠上述外网加速技术不能完全解决系统现有问题，如果访问量持续增加，系统仍存在崩溃的可能。王工提出应同时结合 Web 内网加速技术优化系统改进方案，如综合应用负载均衡、缓存服务器、Web 应用服务器等。经过讨论，公司最终决定采用王工的思路，完成改进系统的设计方案。

【问题 1】（10 分）

针对王工提出的改进方案，从 a~o 中分别选出各技术的相关描述和对应常见支持工具/技术填入表 1 中的（1）~（10）处。

- (a) 响应 Web 请求处理
- (b) 可采用软件级和硬件级任务分配机制实现业务分流
- (c) 将内存的对象映射存储到数据库中
- (d) 对单次请求的处理，不依赖其他请求
- (e) 通过部署本地化结点，尽可能避开互联网上有可能影响数据传输速度和稳定性的瓶颈和环节，使内容传输得更快、更稳定
- (f) 文件存储系统，快速查找文件
- (g) HDFS
- (h) Redis
- (i) Hibernate

- (j) 主从复制
- (k) cookie 存 session
- (l) 响应式图片
- (m) JSON
- (n) Nginx
- (o) WebLogic

表 1 技术描述与常见支持工具/技术

技术	相关描述	支持工具/技术
ORM	(1)	Mybatis、(2)
内存数据库	数据存储在内存中的数据库	Memcached、(3)
负载均衡	(4)	(5)、HAproxy
无状态服务	(6)	(7)，Redis 共享存储
Web 服务器	(8)	(9)，Jboss
CDN	(10)	数据复制

**【问题 2】 (9 分)**

负载均衡的算法可分为静态算法与动态算法，请用 200 字以内的文字解释两类算法的含义，并各列举两种算法。

**【问题 3】 (6 分)**

当单机服务器升级为集群时，会面临一些挑战，有状态服务便是其中一个。请判断以下构件是有状态服务还是无状态服务：

- (1) 登录认证构件
- (2) 试卷搜索构件
- (3) 在线答题构件

答案:

【问题 1】

- (1) (c) (2) (i) (3) (h) (4) (b) (5) (n)
- (6) (d) (7) (k) (8) (a) (9) (o) (10) (e)

【问题 2】

静态算法是不考虑服务器动态负载的算法，包括：

- (1) 轮转算法：轮流将服务请求（任务）调度给不同的节点（即：服务器）。
- (2) 加权轮转算法：考虑不同节点处理能力的差异。
- (3) 源地址哈希散列算法：根据请求的源 IP 地址，作为散列键从静态分配的散列表找出对应的节点。
- (4) 目标地址哈希散列算法：根据请求目标 IP 做散列找出对应节点。
- (5) 随机算法：随机分配，简单，但不可控。

动态算法是考虑服务器动态负载的算法，包括：

- (1) 最小连接数算法：新请求分配给当前活动请求数量最少的节点，每个节点处理能力相同的情况下。
- (2) 加权最小连接数算法：考虑节点处理能力不同，按最小连接数分配。
- (3) 加权百分比算法：考虑了节点的利用率、硬盘速率、进程个数等，使用利用率来表现剩余处理能力。



**【问题 3】**

(1) 有状态服务 (2) 无状态服务 (3) 有状态服务

解析:

静态算法是不考虑服务器动态负载的算法，包括：

- (1) 轮转算法：轮流将服务请求（任务）调度给不同的节点（即：服务器）。
- (2) 加权轮转算法：考虑不同节点处理能力的差异。
- (3) 源地址哈希散列算法：根据请求的源 IP 地址，作为散列键从静态分配的散列表找出对应的节点。
- (4) 目标地址哈希散列算法：根据请求目标 IP 做散列找出对应节点。
- (5) 随机算法：随机分配，简单，但不可控。

动态算法是考虑服务器动态负载的算法，包括：

- (1) 最小连接数算法：新请求分配给当前活动请求数量最少的节点，每个节点处理能力相同的情况下。
- (2) 加权最小连接数算法：考虑节点处理能力不同，按最小连接数分配。
- (3) 加权百分比算法：考虑了节点的利用率、硬盘速率、进程个数等，使用利用率来表现剩余处理能力。

三、近年来，随着互联网行业的迅猛发展，公司或组织业务的不断扩张，需求的快速变化以及用户量的不断增加，传统的单体（Monolithic）软件架构面临着越来越多的挑战，已逐渐无法适应互联网时代对软件的要求。在这一背景下，微服务架构模式(MicroserviceArchitecturePattern)逐渐流行。

希赛公司目前使用的在线教育平台已开发多年，很多方面的设计已不符合业务需求，因此公司决定重新开发新一版的在线教育平台，为了完成开发目标公司架构设计师张工建议采用 SOA 架构，而王工表示反对，建议采用微服务架构，经公司研究决定采纳王工的解决方案。

【问题 1】（5 分）

请简述微服务的特点。

【问题 2】（10 分）

请结合案例背景分析希赛公司为何选用王工的解决方案。

【问题 3】（10 分）

请从以下维度分析微服务与 SOA 的差异，并完成填空。

	微服务	SOA
业务划分方式	(1)	水平多层
粒度	(2)	(3)
部署方式	(4)	整体部署
通信方式	(5)	企业服务总线（ESB）充当了服务之间通信的角色

答案:

**【问题 1】（5 分）**

- (1) 服务粒度小
- (2) 专注于做 1 件事情
- (3) 轻量级的通信机制
- (4) 松耦合
- (5) 独立部署

**【问题 2】（10 分）**

回答从以下几个方面进行展开分析，每个方面 2 分，最高得 10 分。

- (1) 技术异构性
- (2) 弹性
- (3) 扩展
- (4) 简化部署
- (5) 与组织结构相匹配
- (6) 可组合性
- (7) 对可替代性的优化

**【问题 3】（10 分）**

- (1) 纵向业务划分
- (2) 细粒度
- (3) 粗粒度
- (4) 独立部署
- (5) 使用轻量级的通信方式，如 HTTP

解析:

	微服务	SOA
业务划分方式	纵向业务划分	水平多层
粒度	细粒度	粗粒度
部署方式	独立部署	整体部署
通信方式	使用轻量级的通信方式， 如 HTTP	企业服务总线（ESB）充当了服务 之间通信的角色

四、企业信息集成是解决“信息孤岛”问题的需要，由于“信息孤岛”的现象广泛存在，所以企业信息集成也为企业所重视。

企业集成的水平在很大程度上取决于企业内部各种系统、应用或服务的集成化运行水平，良好的软件支持工具可以帮助企业加快实现企业系统集成。作为支持企业集成化运行的使能工具，企业集成平台的主要功能是为企业中各种数据、系统、过程等多种对象的协同运行提供各种公共服务及运行时的支撑环境，从而降低实现企业内部的信息孤岛集成的复杂度，提高应用间集成的有效性，将信息系统实施规划中确定的企业中各种应用系统、服务、人员、信息资源及数字化设备的协同关系物化到集成化运行的可执行系统中去。

【问题 1】（10 分）

1、EAI 一般包括：（1）、数据集成、控制集成、（2）。其中（1）的主要作用是把各应用系统的界面集成起来，统一入口，使用户能够对集成系统产生一个“整体”的感觉。（3）为实现整体的业务目标，要定义、关联和管理不同的业务过程，并通过相应的业务信息系统中实现所需要的信息交换，从而降低成本，更高效地实现客户目标。并可以进行 B2B 集成。

2、数据集成主要有：数据联邦、数据复制和基于接口的数据集成三种模式。其中：（4）是指不同的应用共同访问一个全局虚拟数据库，通过全局虚拟数据库管理系统为不同的应用提供全局信息服务。（5）是指不同的应用系统之间利用适配器来实现相互调用以达到集成的目标。

【问题 2】（7 分）

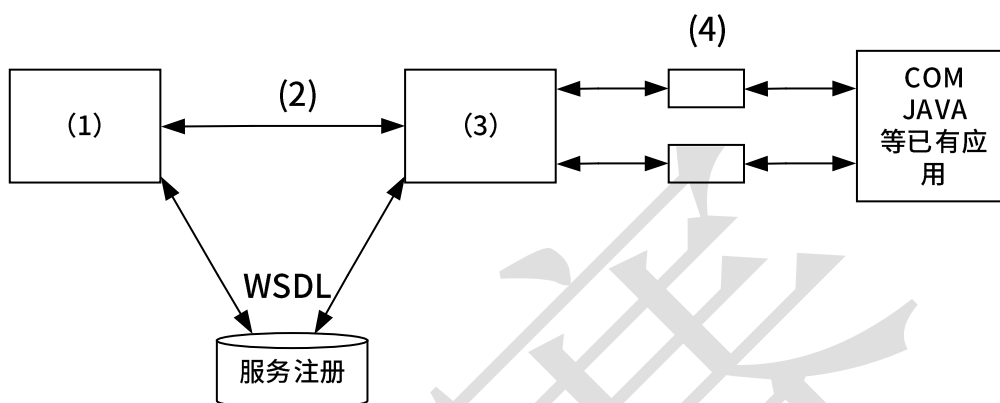
请简述应用之间开发一对一专用接口方式进行集成的优缺点。

【问题 3】（8 分）

面向服务的集成，经常会用到 WebService 技术对遗留系统进行集成。以下为面向服务的集成结构图，请使用以下词，根据自己的理解补充（1）~（4）。

A、ARPB、SOAPC、WSDLD、DCOME、客户端

F、适配器 G、服务提供者 H、扩展开发接口



答案:

【问题 1】（10 分）

- (1) 表示集成（界面集成）
- (2) 业务流程集成（过程集成）
- (3) 业务流程集成（过程集成）
- (4) 数据联邦
- (5) 基于接口的数据集成模式

【问题 2】（7 分，答对 1 点给 2 分，答对 4 条以上得全分）

优点:

直观，当企业应用数量少时容易实现。

缺点:

- 1、工作量大
- 2、集成系统的维护费用高，系统升级与扩展困难

3、不易于标准化，由于接口数量多，给系统管理造成比较大的困难

4、一般只能解决应用系统之间的数据集成问题，难以用来支持过程集成和应用之间的协调

【问题 3】（8 分）

(1) 客户端

(2) SOAP

(3) 服务提供

(4) 适配器

解析：

EAI 一般包括：表示集成（界面集成）、数据集成、控制集成（应用集成）、业务流程集成（过程集成）。

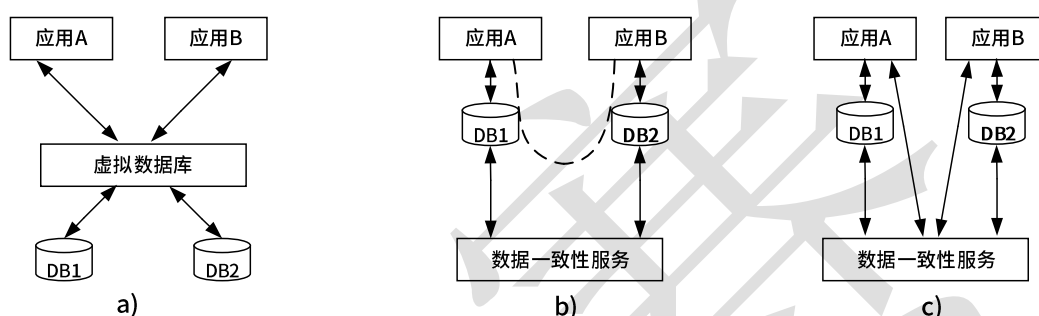
(1) 界面集成：把各应用系统的界面集成起来，统一入口，使用户能够对集成系统产生一个“整体”的感觉。

(2) 数据集成：数据集成是应用集成和业务过程集成的基础，可以提供企业之间的信息共享能力。在集成以前，要对数据进行统一标识、分类，并进行元数据建模。这三个步骤完成后，就可以实现企业范围的数据共享和数据分布了。

(3) 应用集成：这一水平的集成目的是指将多个应用系统进行“绑定”，使之象一个实时运行的系统一样接受信息输入和产生数据输出，实现多个系统功能的“叠加”。应用集成广泛用于 B2B 集成、在后端服务应用基础上建立的客户关系管理系统、集成多个应用的 Web 门户等等。在 ERP 应用实施后，也要经常进行与新的应用系统的集成。

(4) 过程集成（业务流程集成）：为实现整体的业务目标，要定义、关联和管理不同的业务过程，并通过相应的业务信息系统中实现所需要的信息交换，从而降低成本，更高效地实现客户目标。BPI 的要素包括过程管理，过程建模和工作流。

数据集成主要有以下三种模式：数据联邦、数据复制和基于接口的数据集成。如图所示，它们分别描述了对多个异构数据源透明、一致访问的三种实现方法。



图三种典型的数据集成模式

a) 数据联邦 b) 数据复制 c) 基于接口的数据集成

### (1) 数据联邦

数据联邦是指不同的应用共同访问一个全局虚拟数据库，通过全局虚拟数据库管理系统为不同的应用提供全局信息服务，实现不同的应用和数据源之间的信息共享和数据交换，其具体实现由客户端应用、全局信息服务和若干个局部数据源三部分组成。

### (2) 数据复制模式

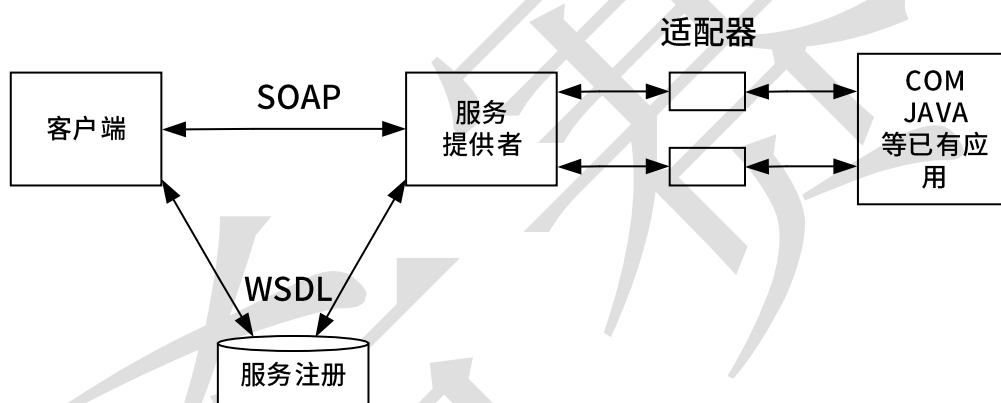
在数据复制模式中，通过底层应用数据源之间的一致性复制来实现（访问不同数据库的）不同应用之间的信息共享和互操作，其实现的关键是必须能够提供在两



个或多个数据库系统之间实现数据转换和传输的基础结构（以屏蔽不同数据库间数据模型的差异）。

### （3）基于接口的数据集成模式

在基于接口的数据集成模式中，不同的应用系统之间利用适配器（或接口代理）提供的编程接口来实现相互调用。应用适配器或接口代理通过其开放或私有接口将业务信息从其所封装的具体应用系统中提取出来，进而实现不同的应用系统之间业务数据的共享与互交换。接口调用的方式可以采用同步调用方法，也可以采用基于消息中间件的异步方法来实现。



五、某证券公司计划更新其网络架构，以提高安全服务能力。该公司技术总监王工通过对公司内部调研，以及对现有系统分析发现，公司网络经常遭受同一种网络攻击。被攻击时的现象为：

- (1) 被攻击主机上有大量等待的 TCP 连接；
- (2) 网络中充斥着大量的无用的数据包，源地址为假；
- (3) 高流量无用数据，造成网络拥塞，使受害主机无法正常和外界通讯；
- (4) 在遭受严重攻击时，主机系统出现崩溃死机的状况。

【问题 1】（6 分）

以下哪些行为属于信息安全中的被动攻击方式：

A、重放攻击 B、主观抵赖 C、网络监听 D、假冒身份 E、信息截取 F、散播病毒  
G、拒绝服务 H、数据篡改 I、流量分析

【问题 2】（8 分）

信息系统安全属性主要包括：保密性、完整性、可用性。

请分析以下攻击方式，是破坏哪一种安全属性。

攻击方式	安全属性
重放攻击	(1)
网络监听	(2)
拒绝服务	(3)
数据篡改	(4)

【问题 3】（11 分）

请根据王工所描述的现象，分析这是一种什么样的网络攻击。可以采取何种方式来抵抗这种攻击。

答案:

【问题 1】 (6 分)

C、E、I

【问题 2】 (8 分)

(1) 完整性 (2) 保密性 (3) 可用性 (4) 完整性

【问题 3】 (11 分)

DDos 攻击或分布式拒绝服务攻击。(3 分)

- (1) 关闭不必要的服务
- (2) 及时更新系统补丁
- (3) 限制同时打开的 Syn 半连接数目
- (4) 缩短 Syn 半连接的 timeout 时间
- (5) 限制特定 IP 地址的访问

注：每答对 1 空得 2 分，最高得 8 分。

解析:

主动攻击：主动攻击(activeattack)可能改变信息或危害系统。威胁信息完整性和有效性的攻击就是主动攻击。主动攻击通常易于探测但却难于防范，因为攻击者可以通过多种方法发起攻击。

被动攻击：在被动攻击(passiveattack)中，攻击者的目的只是获取信息，这就意味着攻击者不会篡改信息或危害系统。系统可以不中断其正常运行。然而，攻击可能会危害信息的发送者或者接收者。威胁信息机密性的攻击--窃听和流量分析均属被动攻击。信息的暴露会危害信息的发送者或接收者，但是系统不会受到影响。因此，在信息发送者或者接收者发现机密信息已经泄露之前，要发现这种

攻击是困难的。然而，被动攻击可以通过对信息进行加密而避免。

从以上概念可以看出：网络监听、信息截取、流量分析都属于“获取信息”而不会影响系统的正常运行，所以属于被动攻击。

问题 2 属于概念题，答案参考下表。

攻击方式	安全属性
重放攻击	完整性
网络监听	保密性
拒绝服务	可用性
数据篡改	完整性

本题中，问题 3 难度最高。

从描述中的“被攻击主机上有大量等待的 TCP 连接”、“网络中充斥着大量的无用的数据包，源地址为假”、“高流量无用数据，造成网络拥塞”可以看出该公司遭受的是 DDos 攻击（即分布式拒绝服务攻击）。该攻击方式目前没有很好的手段完全防御，但可以考虑采取以下手段，一定程度的缓解。

- (1) 关闭不必要的服务
- (2) 及时更新系统补丁
- (3) 限制同时打开的 Syn 半连接数目
- (4) 缩短 Syn 半连接的 timeout 时间
- (5) 限制特定 IP 地址的访问

六、某软件公司计划开发一套教学管理系统，用于为高校提供教学管理服务。该教学管理系统基本的需求包括：

- (1) 系统用户必须成功登录到系统后才能使用系统的各项功能服务；
- (2) 管理员(Registrar)使用该系统管理学校(University)、系(Department)、教师(Lecturer)、学生(Student)和课程(Course)等教学基础信息；
- (3) 学生使用系统选择并注册课程，必须通过所选课程的考试才能获得学分；如果考试不及格，必须参加补考，通过后才能获得课程学分；
- (4) 教师使用该系统选择所要教的课程，并从系统获得选择该课程的学生名单；
- (5) 管理员使用系统生成课程课表，维护系统所需的有关课程、学生和教师的信息；
- (6) 每个月到了月底系统会通过打印机打印学生的考勤信息。

项目组经过分析和讨论，决定采用面向对象开发技术对系统各项需求建模。

**【问题 1】 (7 分)**

用例建模用来描述待开发系统的功能需求，主要元素是用例和参与者。请根据题目所述需求，说明教学服务系统中有哪些参与者。

**【问题 2】 (7 分)**

用例是对系统行为的动态描述，用例获取是需求分析阶段的主要任务之一。请指出在面向对象系统建模中，用例之间的关系有哪几种类型？对题目所述教学服务系统的需求建模时，“登录系统”用例与“注册课程”用例之间、“参加考试”用例与“参加补考”用例之间的关系分别属于哪种类型？

**【问题 3】 (11 分)**

类图主要用来描述系统的静态结构，是组件图和配置图的基础。请指出在面向对象系统建模中，类之间的关系有哪几种类型？对题目所述教学服务系统的需求建模时，类 University 与类 Student 之间、类 University 和类 Department 之间、类 Student 和类 Course 之间的关系分别属于哪种类型？

答案:

【问题 1】

学生、教师、管理员、时间、打印机。

【问题 2】

用例之间的关系包括：包含、扩展、泛化。

“登录系统”用例与“注册课程”用例之间的关系为：包含关系。

“参加考试”用例与“参加补考”用例之间的关系为：扩展关系。

【问题 3】

类之间的关系包括：关联、聚合、组合、依赖、泛化、实现（可写可不写，因为实现是接口与类之间的关系，而接口是一种特殊的类）

类 University 与类 Student 之间的关系是：聚合关系。

类 University 与类 Department 之间的关系是：组合关系。

类 Student 与类 Course 之间的关系是：关联关系。

解析:

【问题 1】

参与者是指系统以外的，需要使用系统或与系统交互的事物，包括：人或组织、设备、外部系统等。在本题中，较为容易识别的参与者包括：学生、教师、管理

员，比较隐晦的参与者包括：时间、打印机。

### 【问题 2】

用例之间的关系主要有泛化 (Generalization)、包含 (Include) 和扩展 (Extend)。

(1) 当可以从两个或多个用例中提取公共行为时，可以使用包含关系来表示。

(2) 如果一个用例混合了两种或两种以上不同场景，即根据情况可能发生多种分支，则可以将这个用例分为一个基本用例和一个或多个扩展用例。

(3) 当多个用例共同拥有一个类似的结构和行为的时候，可以将它们的共性抽象成父用例，其他的用例作为泛化关系中的子用例。

在题目要求中，用例“登录系统”是用例“注册课程”和其他用例执行的公共行为，两者是包含 (Include) 关系。用例“参加补考”是用例“参加考试”的一种分支和特殊场景，两者之间的关系是扩展 (Extend) 关系。

### 【问题 3】

依赖关系：一个事物发生变化影响另一个事物。泛化关系：特殊/一般关系。

关联关系：描述了一组链，链是对象之间的连接。

聚合关系：整体与部分生命周期不同。

组合关系：整体与部分生命周期相同。

实现关系：接口与类之间的关系。

在题目要求中，类 University 与类 Student 之间的关系是整体与部分关系，而且具有不同的生存周期，所以是聚合关系。类 University 和类 Department 之间的关系是整体与部分的关系，两者具有相同的生存周期，所以是组合关系。类 Student 和类 Course 之间为连接关系，所以属于关联关系。

七、某软件公司成立项目组为某高校开发一套教职工信息管理系统。与教职工信息相关的数据需求 and 处理需求如下：

(1) 数据需求：在教职工信息中能够存储学校所有在职的教工和职工信息，包括姓名、所属部门、出生年月、工资编号、工资额和缴税信息；部门信息中包括部门编号、部门名称、部门人数和办公地点信息。

(2) 处理需求：能够根据编制内或外聘教职工的工资编号分别查询其相关信息；每个月的月底统一核发工资，要求系统能够以最快速度查询出教工或者职工所在部门名称、实发工资金额；由于学校人员相对稳定，所以数据变化及维护工作量很少。

项目组王工和李工针对上述应用需求分别给出了所设计的数据模型（如图 2-1 和图 2-2 所示）。王工遵循数据库设计过程，按照第三范式对数据进行优化和调整，所设计的数据模型简单且基本没有数据冗余；而李工设计的数据模型中存在大量数据冗余。

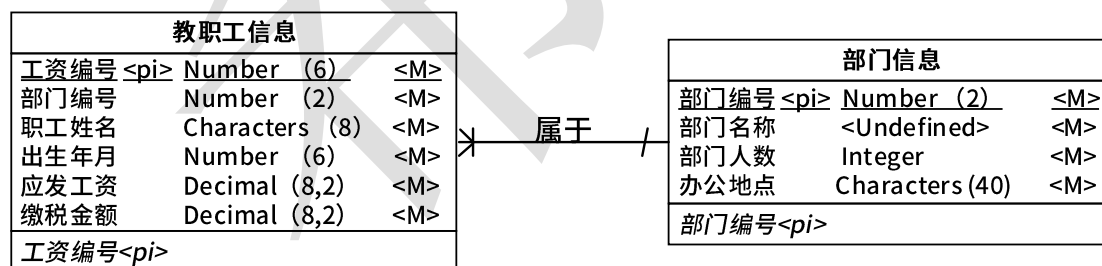


图2-1 王工设计的数据模型



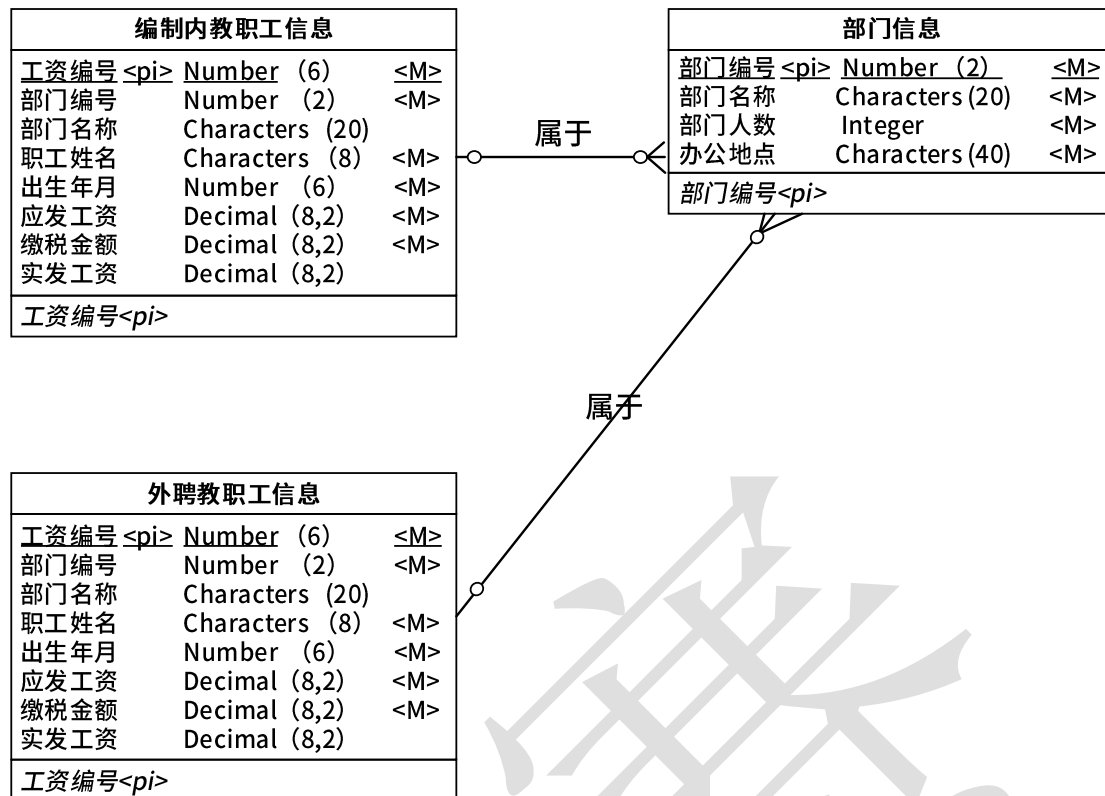


图2-2 李工设计的数据模型

项目组经过分析和讨论，特别是针对数据处理中对数据访问效率的需求，最终选择了李工给出的数据模型设计方案。

【问题 1】（9 分）

请用 300 字以内的文字，说明什么是数据库建模中的反规范化技术，指出采用反规范化技术能获得哪些益处，可能带来哪些问题。

【问题 2】（8 分）

请简要叙述常见的反规范化技术有哪些。

【问题 3】（8 分）

请分析李工是如何应用反规范化技术来满足教职工信息管理需求的。

答案:

### 【问题 1】

规范化设计后，数据库设计者希望牺牲部分规范化来提高性能，这种从规范化设计的回退方法称为反规范化技术。

采用反规范化技术的益处：降低连接操作的需求、降低外码和索引的数目，还可能减少表的数目，能够提高查询效率。

可能带来的问题：数据的重复存储，浪费了磁盘空间；可能出现数据的完整性问题，为了保障数据的一致性，增加了数据维护的复杂性，会降低修改速度。

### 【问题 2】

(1) 增加冗余列：在多个表中保留相同的列，通过增加数据冗余减少或避免查询时的连接操作。

(2) 增加派生列：在表中增加可以由本表或其他表中数据计算生成的列，减少查询时的连接操作并避免计算或使用集合函数。

(3) 重新组表：如果许多用户需要查看两个表连接出来的结果数据，则把这两个表重新组成一个表来减少连接而提高性能。

(4) 水平分割表：根据一行或多行数据的值，把数据放到多个独立的表中，主要用于表数据规模很大、表中数据相对独立或数据需要存放到多个介质上时使用。

(5) 垂直分割表：对表进行分割，将主键与部分列放到一个表中，主键与其他列放到另一个表中，在查询时减少 I/O 次数。

### 【问题 3】

在教职工信息管理系统的需求中，能够根据编制内或外聘教职工的工资编号分别

查询其相关信息，数据查询要求有很高的处理效率。李工所设计的数据模型中采用了三种反规范化技术：

- (1) 增加冗余列：增加“部门名称”列，消除了数据查询中“教职工信息”表和“部门信息”表之间的连接；
- (2) 增加派生列：增加“实发工资”列，消除了实发工资的计算过程；
- (3) 水平分割表：将教职工信息表分割为“编制内教职工信息”表和“外聘教职工信息”表，减少了数据查询的范围。

解析：

本题考查数据库相关的知识。题目涉及反规范化技术。

规范化技术在数据库中的应用十分常见。当数据库对数据模型进行规范化处理后，会发现这些经过规范化处理的模型在进行查询操作时效果并不理想。因为经过规范化处理的数据模型形成了一系列的小表，每个表的数据量较小，进行查询操作时往往需要应用程序对这些表进行动态的连接操作，这就要在不同的表中进行多次 I/O 操作。表的连接操作对于数据较少的小容量表也许不会产生较大的影响，但是对于数据量十分庞大的数据库，这种多表连接操作在时间上是很难让用户接受的。此时，提高效率的最好方法就是使这些小表合并在一起，这就是数据的反规范化处理。

#### 【问题 1】

经过以上对于规范化处理的描述可以得知：反规范化技术主要是为了提升查询性能而进行的一系列处理，包括对数据表进行连接，合理规划数据的存放，以及引入合理的数据冗余。其益处在于：降低连接操作的需求、降低外码和索引的数目，

还可能减少表的数目，能够提高查询效率。带来的问题是：数据的重复存储，浪费了磁盘空间；可能出现数据的完整性问题，为了保障数据的一致性，增加了数据维护的复杂性，会降低修改速度。

### 【问题 2】

问题 2 为纯概念题，请参看本题参考答案。

### 【问题 3】

问题 3 需要利用问题 2 的结论并结合题目要求进行解题。教职工信息管理系统要求“能够根据编制内或外聘教职工的工资编号分别查询其相关信息”，并从题目“特别是针对数据处理中对数据访问效率的需求”可以看出系统对数据访问效率非常重视。而李工正是采用了反规范化技术，使得效率能明显提升。对比王工和李工所设计的数据模型可以发现：在李工设计的数据模型中，增加了“部门名称”列，这属于增加冗余列的技术；增加了“实发工资”列，这属于增加派生列的技术；将教职工信息表分割为“编制内教职工信息”表和“外聘教职工信息”表，这属于水平分割表。

八、某牙科诊所拟开发一套信息系统，用于管理病人的基本信息和就诊信息。诊所工作人员包括：医护人员(DentalStaff)、接待员(Receptionist)和办公人员(OfficeStaff)等。系统主要功能需求描述如下：

1.记录病人基本信息(Maintainpatientinfo)。初次就诊的病人，由接待员将病人基本信息录入系统。病人基本信息包括病人姓名、身份证号、出生日期、性别、首次就诊时间和最后一次就诊时间等。每位病人与其医保信息(MedicalInsurance)关联。

2.记录就诊信息(Recordofficevisitinfo)。病人在诊所的每一次就诊，由接待员将就诊信息(OfficeVisit)录入系统。就诊信息包括就诊时间、就诊费用、支付代码、病人支付费用和医保支付费用等。

3.记录治疗信息(Recorddentalprocedure)。病人在就诊时，可能需要接受多项治疗，每项治疗(Procedure)可能由多位医护人员为其服务。治疗信息包括：治疗项目名称、治疗项目描述、治疗的牙齿和费用等。治疗信息由每位参与治疗的医护人员分别向系统中录入。

4.打印发票(Printinvoices)。发票(Invoice)由办公人员打印。发票分为两种：给医保机构的发票(InsuranceInvoice)和给病人的发票(PatientInvoice)。两种发票内容相同，只是支付的费用不同。当收到治疗费用后，办公人员在系统中更新支付状态(Enterpayment)。

5.记录医护人员信息(Maintaindentalstaffinfo)。办公人员将医护人员信息录入系统。医护人员信息包括姓名、职位、身份证号、家庭住址和联系电话等。

6.医护人员可以查询并打印其参与的治疗项目相关信息(Searchandprintprocedureinfo)。

现采用面向对象方法开发该系统，得到如图 3-1 所示的用例图和 3-2 所示的初始类图。

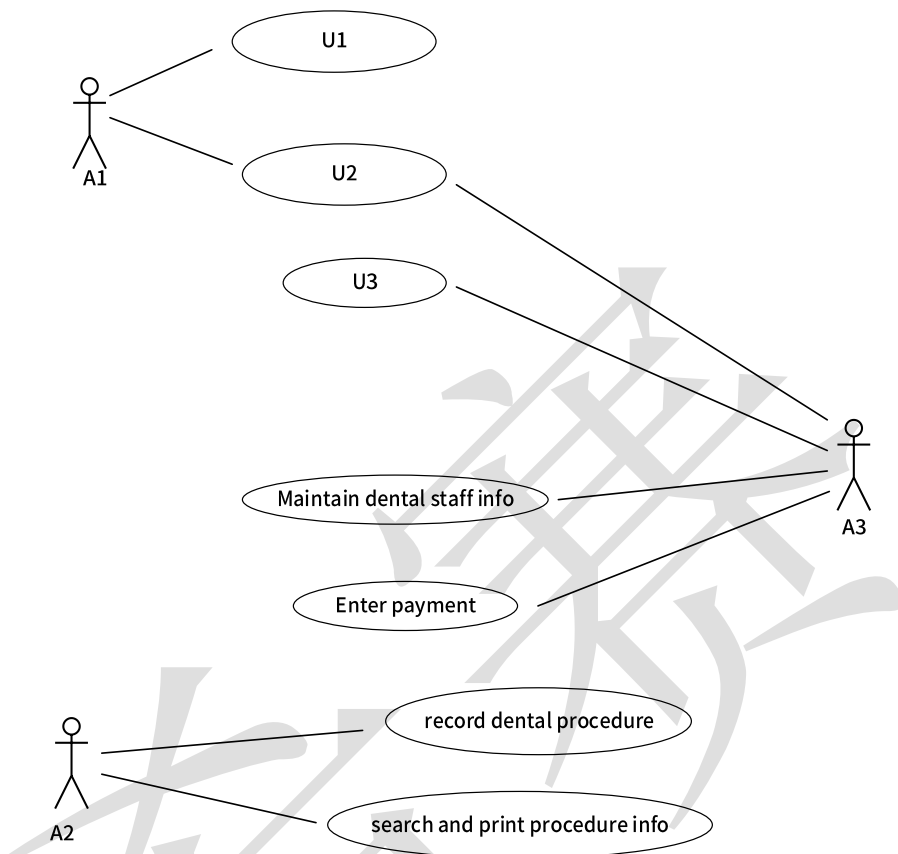


图3-1 用例图

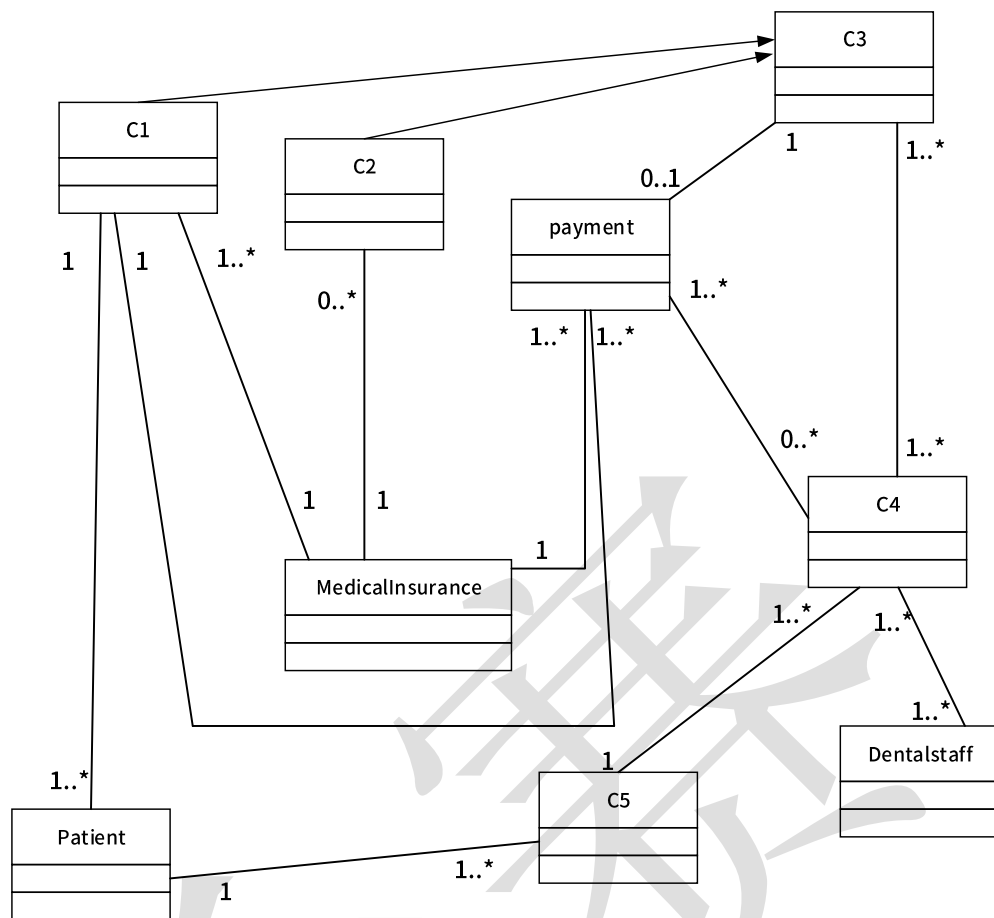


图3-2 初始类图

【问题 1】（6 分）

根据说明中的描述，给出图 1 中 A1~A3 所对应的参与者名称和 U1~U3 所对应的用例名称。

【问题 2】（10 分）

根据说明中的描述，给出图 2 中 C1~C5 所对应的类名。

【问题 3】（9 分）

根据说明中的描述，给出图 2 中类 C4、C5、Patient 和 DentalStaff 的必要属性。

答案:

**【问题 1】**

A1: Receptionist (接待员) A2: DentalStaff (医护人员) A3: OfficeStaff  
(办公人员)

U1: Maintainpatientinfo(记录病人基本信息)U2: Recordofficevisitinfo(记录就诊信息)U3: Printinvoices(打印发票)

**【问题 2】**

C1:PatientInvoiceC2:InsuranceInvoiceC3:InvoiceC4:ProcedureC5:  
OfficeVisit

**【问题 3】**

C4: 治疗项目名称、治疗项目描述、治疗的牙齿和费用

C5: 病人就诊时间和费用、支付代码、病人支付费用、医保支付费用

Patient: 姓名、身份证号、出生日期、性别、首次就诊时间和最后一次就诊时间

DentalStaff: 姓名、职位、身份证号、住址、联系电话。

解析:

**【问题 1】**

本题属于常规考题，考查对参与者和用例名的补充，系统的参与者一般为人员、机构或第三方系统。用例名一般为动词+名词或名词+动词，是对系统功能的概括和描述。

本题根据题干说明，参与者即诊所工作人员包括：医护人员(DentalStaff)、接待员(Receptionist)和办公人员(OfficeStaff)等。



根据用例图已有信息，A2 使用用例 Record dental procedure 和 Search and print procedure info，根据题干说明记录治疗信息

(Record dental procedure) 由每位参与治疗的医护人员分别录入，医护人员可以查询并打印其参与的治疗项目相关信息 (Search and print procedure info)，因此 A2 为医护人员 (Dental Staff)。

根据用例图已有信息 A3 使用用例 Maintain dental staff info 和 Enter payment，根据题干说明记录医护人员信息 (Maintain dental staff info)，由办公人员录入系统，所以 A3 为 Office Staff (办公人员)。并且根据题干办公人员还需要打印发票 (Print invoices)、更新支付状态 (Enter payment)，缺失的 U3 应该是打印发票 (Print invoices)。根据题干描述 A3 与 U2 之间的联系没有意义，不参考。根据题干描述和图示，A1 对应的参与者应该是 A1: Receptionist (接待员)，接待员需要参与的功能有记录病人基本信息 (Maintain patient info) 和记录就诊信息 (Record office visit info)，分别对应 U1、U2，二者与办公人员都没有明确联系，位置可以互换。

### 【问题 2】

本题属于常规考查题型，补充缺失的类名，常见的实体类类名一般为名词形式，也会有特殊的边界类/接口类和协调类。需要参照题干描述和类图中类与类之间的关系，来确定缺失的类名及其位置。

根据初始类图，存在一组泛化关系，C3 是 C1、C2 的泛化，即 C3 是 C1、C2 的父类，根据题干描述存在这样泛化关系的只有发票 (Invoice)、病人发票 (Patient Invoice)、医保机构发票 (Insurance Invoice)，又根据多重度来分析，由于可能存在全自费的情况，即医保发票不存在，所以多重度 0..\* 对应的 C2 类

名应该是医保机构发票(InsuranceInvoice)，C1 对应的是病人发票

(PatientInvoice)，父类 C3 对应的是发票(Invoice)。

又根据图示和题干，与医护人员(DentalStaff)相关的用例有记录治疗信息

(Recorddentalprocedure)、查询并打印其参与的治疗项目相关信息

(Searchandprintprocedureinfo)，其中能够找到相关内容治疗信息

procedure，即 C4 对应的是实体类治疗信息 Procedure，与治疗相关的是就诊

信息 officevisit，即 C5 对应的是实体类就诊信息 OfficeVisit。

### 【问题 3】

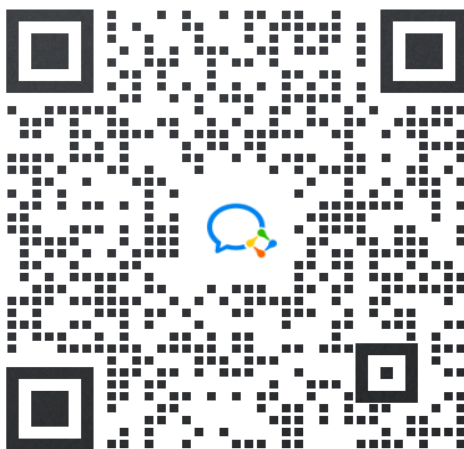
根据题干描述，“病人基本信息包括病人姓名、身份证号、出生日期、性别、首次就诊时间和最后一次就诊时间等”，因此 Patient 的必要属性包括病人姓名、身份证号、出生日期、性别、首次就诊时间和最后一次就诊时间。

根据题干描述，“医护人员信息包括姓名、职位、身份证号、家庭住址和联系电话等”，因此 DentalStaff 的必要属性包括姓名、职位、身份证号、家庭住址和联系电话等。

根据题干描述，C5OfficeVisit 就诊信息包括就诊时间、就诊费用、支付代码、病人支付费用和医保支付费用等。

根据题干描述 C4:Procedure 治疗信息包括：治疗项目名称、治疗项目描述、治疗的牙齿和费用等。

更多备考资料和学习福利，可扫码添加希赛嘉儿老师，申请入群



希赛网