

2023 年上半年系统分析师考试下午真题（专业解析+ 参考答案）

1、某公司拟开发一个电商平台用户行为分析系统，该系统可支持对电商平台用户行为数据的采集和实时/高效数据处理分析，统计分析用户访问情况。搜集 PV/UV 商品热度等相关数据。系统应具备以下能力

(1) 对海量用户行为数据的存储和处理能力

(2) 为适应用户行为的随机性，保证系统不会被爆发的日志数据压垮，系统应具备对突发流量的处理能力

(3) 对用户数据的营私保护能力

针对项目需求，工时召开会议，会上张工提出买那个对大量数据，传统的文件存储于关系型数据库，无法完全解决当前系统问题，应基于大数据技术设计并实现系统，李工提出可就要数据脱敏的方法解决用户隐私的问题。

问题内容：

【问题 1】

根据项目要求，该系统需考虑电商平台用户行为的随机性，具有应对突发数据，流量数据的处理能力。请基于大数据相关技术的了解，说明可选那个技术支持流系统应对突发流量暴增情况？并用 200 字以内说明该技术优势。

【问题 2】

根据张工建议，项目组拟采用大数据技术开发该系统。图 5 给出本项目架构设计方案，请选择 (a) ~ (m) 选项中选出合适的内容，完善 (1) ~ (7) 处

(a) 网络层、(b) 业务层、(c) 接入层、(d) 数据处理、(e) 数据采集、(f) Flink 实时计算、(g) Flink 批处理、(h) Hbase、(i) Redis、(j) NoSQL、(k) kafka、(l) 边缘计算、(m) 设备层

图暂缺

【问题 3】

根据李工的建议，项目组拟应用数据脱敏方法对该系统的敏感数据进行脱敏。请用 300 字以内文字说明常用 5 种数据脱敏方法。

试题答案：

【问题 1】

可选消息中间件技术支持流系统应对突发流量暴增情况。

消息中间件具有的技术优势：

(1) 异步通信，通过消息中间件，不同的服务之间进行异步通信，可以提高系统的吞吐量和并发性能。(2) 解耦合，将消息发送方和接受方进行解耦，从而能独立开发和部署，提高系统的可扩展性和可维护性。(3) 流量控制或者消息缓冲，可以设置流量控制，当消息量过大时，暂停消息的发送，流量小时，继续发送，实现削峰填谷功能。也可以作为消息缓冲功能，作为缓冲区暂存消息，从

而实现系统的负载和压力。

【问题 2】

- (1) c
- (2) k
- (3) 业务层
- (4) f
- (5) g
- (6) h
- (7) i

【问题 3】

常用数据脱敏方法为：（1）替换。将字符或整数值替换为查找或标准模式中的字符或整数值。例如，可以用列表中的随机值替换名字。（2）混排。在一个记录中交换相同类型的数据元素或者在不同行之间交换同一属性的数据元素。例如，在供应商发票中混排供应商名称，以便将发票上的原始供应商替换为其他有效供应商。（3）时空变异。把日期前后移动若干天（小到足以保留趋势），足以使它无法识别。（4）数值变异。应用一个随机因素（正负一个百分比，小到足以保持趋势），重要到足以使它不可识别。（5）取消或删除。删除不应出现在测试系统中的数据。

2、某国际大型企业拟向企业开发一个在线社交网络服务（SNS）系统，让企业内部的员工能够通过类似社交网络的方式进行工作和自我管理，以实现企业内部员工高效、透明、便捷的沟通与协作。其基本的工序包括：

- 1、企业内部各部门、员工等基本信息的维护和管理；
 - 2、提供常见的 SNS 服务，如 Blog，Wiki，数钱，RSS，即时通信和邮件等；
 - 3、提供社交网络，推荐系统等基本功能提高业务协作除了上述功能性的需求外，还需要满足系统的可拓展、灵活部署以及高性能和高可用性需求
- 企业设计系统小组……

问题内容：

【问题 1】

李工建议采用 SNS 中常见的 NoSQL 数据库技术，原因是传统的关系数据库不能满足实际需求，请用 200 字内说明该系统不能采用传统关系数据库的原因，并给出 NoSQL 数据的基本特点。

【问题 2】

李工对当前主流的 NoSQL 数据库镜像了调研形成了如下表
暂缺

【问题 3】

针对李工的建议，王工认为单纯使用 NoSQL 数据库无法完全满足系统要求，考虑到系统功能所涉及数据类型建议使用混合数据库模型来实现，请用 100 字以内文字说明。混合数据库模式如何组织。

试题答案：

【问题 1】

根据在线社交网络服务（SNS）系统的需求，要求实现即时通信和邮件等功能需求，还要满足系统的可扩展、灵活部署以及高性能和高可用性需求等。所以该系统要求对数据库高并发读写的需求、对海量数据的高效率存储和访问的需求、对数据库的高可扩展性和高可用性的需求。而传统的数据库是无法实现这“三高”需求。所以该系统是不能单纯使用传统关系数据库来实现。

NoSQL 数据库基本特点：

优点：

NoSQL 数据库支持高并发数据访问，性能较高。

NoSQL 数据库的数据存储结构松散，能够灵活支持多种类型的数据格式。

NoSQL 数据库能够支持海量数据的存储，且易于横向扩展。

NoSQL 数据库基于分布式数据存储，不存在单点故障和性能瓶颈，系统可用性高。

缺点：数据模型和查询语言没有经过数学验证。不支持 ACID 特性等。

【问题 2】

(a) 大数据、OLAP 应用

(b) 图由节点和关系组成

(c) 高性能读写

(d) 以文档为存储信息的基本单位

【问题 3】

混合数据库模型是混合了各种类型的数据库，比如传统关系型数据库、NoSQL 数据库等。根据题干的要求，可以用关系型数据库实现部门、员工基本信息的维护和管理；可以用 key-value 数据库或者列式存储实现即时通讯的功能；可以用图数据库实现社交网络和推荐系统等。

3、随着嵌入式计算资源快速提升，容器技术（Docker）发挥重要作用，某公司对原有平台升级，公司将平台升级任务交给了张工，张工经过分析、调研，提出在园嵌入式操作系统平台上采用容器技术的升级方案，但该方案引发了争议。

问题内容：

【问题 1】

争论焦点是采用容器技术还是虚拟机（VM）技术。

李工指出由于容器技术共享主机内核能像虚拟机一样完全隔离，系统存在安全问题；如果采用虚拟机技术除满足需求外，还保证了系统的安全和稳定，会上领导根据系统升级的初衷选择了张工的升级方案，请用 300 字以内的文字说明容器技术和虚拟技术的含义，并简要论述公司领导采纳容器技术的原因。

【问题 2】

表 3-1 给出了虚拟技术和容器技术的性能对比表，请根据下面的 (a) ~ (h) 的 8 个性能指标；判断这些指标属于哪类对比项，补充完善 3-1 的 (1) - (8) 的空白处。

(a) 分钟级、(b) 包含 GuestOS, G 两级以上、(c) 跨操作系统平台迁移、
(d) CPU 与内存按核、按 G 分配 (e) 毫秒级、(f) Cgroups, 进程级别、(g)
VM 伸缩, cpu/内存手动伸缩、(h) 实例自动伸缩、cpu 内存自动在线伸缩。

表 3-1 虚拟技术和容器技术的对比表

对比项	虚拟机技术	容器技术
镜像大小	①	仅包含运行的Bin/Lib,M量级
资源要求	②	CPU与内存按单核、低于G量级分配
启动时间	③	④
可持续性	跨物理机迁移	⑤
弹性伸缩	⑥	⑦
隔离策略	操作系统、系统级别	⑧

试题答案:

【问题 1】

容器技术是一种内核轻量级的操作系统层虚拟化技术,能隔离进程和资源。虚拟机(VM)技术是一种创建于物理硬件系统,充当虚拟计算机系统的虚拟环境,该虚拟机可以独立运行在一个完全隔离的环境中,向本地计算一样运行计算机的程序。

容器技术与虚拟机技术相比有更多的优点,比如容器技术是轻量级的虚拟化技术,启动时间更快,毫秒级别。另外就是容器技术扩展更方便,易于移植,并且能自动实现弹性伸缩等。基于以上优点,公司领导采纳容器技术。

【问题 2】

- ① b
- ② d
- ③ a
- ④ e
- ⑤ c
- ⑥ g
- ⑦ h
- ⑧ f

4、阅读以下关于企业信息系统结构化分析的叙述,回答问题 1 和 2。

说明:某软件公司为企业开发一套员工在线教育系统支持员工利用业余时间开展专业技术培训,提升员工技能。在项目开展初期,采用结构化分析进行开发,并对系统中培训部员工和培训教师的相关功能进行分析,具体需求如下:

(1) 培训部根据企业技术发展需求,负责策划培训课程,并形成课程计划,针对不同的员工设置不同的课程;

(2) 员工首先在系统进行注册,填写自己的编号,学历,专业,岗位等信息,

生员工注册信息，然后将自己的培训需求录入系统，系统自动评估并进行课程推荐，员工确认后形成课程需求；

(3) 培训教师也通过系统进行注册, 填写自己的编号、学历、专业等信息, 形成培训教师注册信息。

(4) 系统根据课程计划、员工注册信息, 课程需求和培训教师注册信息, 为员工和培训教师生成对应的课程表。

工时系统分析师对上述流程进行了审核, 并指出需补充数据字典, 从而更完整地
对系统建模。

问题内容:

【问题1】

数据流图 (DFD) 是结构分析方法的重要工具。请用 300 字以内的文字描述 DFD 的定义。

【问题2】

项目组针对题干描述的业务需求，初步绘制了系统流程图（2-1），请分析途中的三类错误并对每类错误进行简单解释。



图2-1 员工在线教育系统数据流图

试题答案:

【问题1】

数据流图或称数据流程图,是一种便于用户理解、分析系统数据流程的图形工具。它摆脱了系统的物理内容,精确地在逻辑上描述系统的功能、输入、输出和数据存储等,是系统逻辑模型的重要组成部分。数据流图是结构化分析方法中使用的工具,它以图形的方式描绘数据在系统中流动和处理的过程,由于它只反映系统必须完成的逻辑功能,所以它是一种功能模型。在结构化开发方法中,数据流图是需求分析阶段产生的结果。

【问题 2】

第1类错误,加工只有输入没有输出。比如“需求评估与课程推荐”这个加工只有输入没有输出。

第2类错误，数据流没有方向箭头，比如“员工注册信息”这条数据流没有标注方向箭头。

第3类错误，所有的加工没有编号，比如“需求评估与课程推荐”、“培训教师注册”等等。

第4类错误，数据流没有名称，比如“培训课程规划”到“培训课程安排”这条数据流没有名字。

第5类错误，不能从一个存储直接到另一个存储，比如“员工课程表”到“员工”表这条数据流就是错误的。

5、阅读以下关于软件系统分析与建模的叙述，在纸上回答问题1至3。

说明：某软件公司拟开发一套汽车租赁系统，科学、安全和方便的管理租赁公司的各项业务，提高公司效率，提升利率。注册用户在使用系统镜像车辆预约时需执行以下操作：（a）用户登录系统（b）查询车辆信息（c）预订车辆（d）提取车辆（e）归还车辆（f）计算结余（g）查询还车时间。

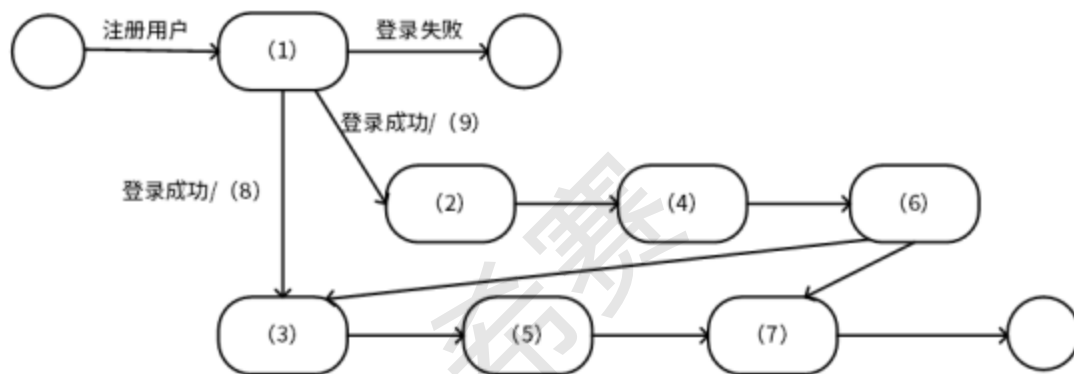


图1-1 系统状态机图

问题内容：

【问题1】（6分）

在面向对象软件分析过程中，状态机图和活动图用于建立软件的动态模型，主要描述系统随时间变化的行为，请分别阐述这两种图的概念和特点。

【问题2】（10分）

图1-1给出了采用状态机图描写的注册用户的状态交换过程，请使用题干给出的（a）~（g）补充完善图1-1中的（1）~（7），并给出（8）和（9）空白处的状态转移执行动作。将正确的答案填写在答题纸上。

【问题3】（9分）

RUP 中有9个核心工作流，分6个核心过程和3个核心支持工作流。9个核心工作流在项目中迭代被使用在每次迭代中以不同的重点和强度重复。结合题干案例，用300字以内的文字简要说明 RUP 中9个核心工作流的主要作用。

试题答案：

【问题 1】

状态图用来描述对象状态和事件之间的关系，强调一个实体基于事件反应的动态行为。状态图适合用于表述在不同用例之间的对象行为。

活动图描述的是对象活动的顺序关系所遵循的规则，它着重表现系统的行为，强调了行为活动的顺序和条件控制。

活动图是由状态图变化而来的，他们各自用于不同的目的。活动图依据对象状态的变化来捕获动作（将要执行的工作或活动）与动作的结果。活动图中一个活动结束后立即进入下一个活动（在状态图中，状态的变迁可能需要事件的触发）。

【问题 2】

- (1) a
- (2) b
- (3) g
- (4) c
- (5) e
- (6) d
- (7) f
- (8) 查询到有预定车辆
- (9) 查询到无预定车辆

【问题 3】

RUP 中有 9 个核心工作流，分别是业务建模、需求、分析和设计、实现、测试、部署、配置和变更管理、项目管理、环境。【注意各个工作流的作用结合题干信息描述下。】