练习 1：高级操作系统：

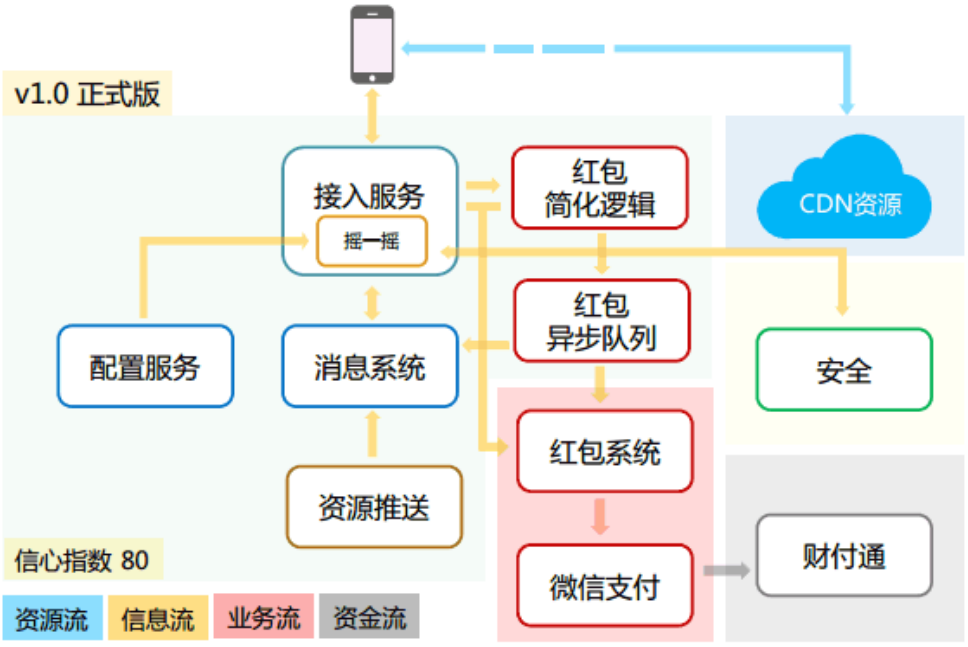
1. **什么是分布式系统？ 什么是分布式系统的设计目标？ 什么是分布式系统的基本特征？查阅资料： 现实生活中的分布式系统举例， 给出架构图并阐述系统组成部分与功能特点。举例说明需要处理的关键技术与特征？ （注意标引参考资料来源）**

分布式系统是若干独立计算机组成的集合，集合中的成员利用网络彼此通信，协同完成任务，而该集合对于用户而言如同一个单独的系统，即用户并不知道处理任务的系统是由多台主机组成的。

分布式系统的设计目标主要有四点，即使资源可访问、透明性、开放性以及可扩展性。分布式系统的最主要的目标就是让用户能够方便的访问远程资源，并实现资源的共享。该目标的实现能够显著的降低经济成本，提供工作效率。透明性的目标实现则在于对用户隐藏分布式系统中的资源和进程实际上是分布于多台计算机之上的这一事实，隐藏该事实在大多数情况下起到了便于用户使用分布式系统的作用。至于开放性，一个开放的分布式系统即能够根据一系列描述了语法和语义的的准则来提供服务的系统。开放性良好的系统将有较好的互操作性与可移植性。而可扩展性的目标，顾名思义，即要让分布式系统能够变“大”。变“大”则体现为在规模、地域以及管理上的扩展。

分布式系统的基本特征主要包括了其并发性（多机并发）、缺乏全局时钟以及故障独立性。

现实生活中的分布式系统样例非常多，而能够跟大部分人都相关甚至是每天都相关的系统却并不是那么常见，而微信的红包系统就是很好的一个例子。接下来介绍的系统是在 2015 年羊年春晚采用的实现了“春晚摇一摇”活动的微信红包系统。该系统的最终架构图如下所示。



微信红包原理系统架构图终版（2015 年春晚所采用）

系统组成部分与功能特点

从用户开始摇手机这个动作开始介绍整个系统。用户摇动手机后，客户端将会产生摇一摇请求，请求将被发送到接入服务中，在接入服务中，系统会根据春晚现场的节目流程，经过一系列的逻辑判断，给客户端返回一个结果：明星拜年或者其他一些特殊用途的界面（配置服务、消息系统与资源推送）或者是红包。如果摇到了红包，由于红包皆为企业赞助，红包界面需要对企业形象进行展示，客户端会从 CDN 中拉回相关企业的 LOGO 等资源，最终展示出一个完整的红包。摇出红包后，用户拆红包，请求就会进入红包系统（包括红包简化逻辑以及红包异步队列），再到支付系统，然后通过财付通系统完成一系列复杂的账务处理，最终取得红包。在上述过程中，为了保证该系统的安全性，即防止红包被多取或恶意领取等情况的发生，该分布式系统中单独设置了一个安全模块。

需要处理的关键技术与特征

（1） **资源预下载**：春晚摇一摇会用到大量的多媒体资源，而这些资源都是从 CDN 下载的，带宽的峰值需求是 3TB/s，这将带来巨大的带宽压力，会造成较长时间的延时，用户体验将受到较大的影响。为了避免上述情况的发生，考虑到上书多媒体资源大多都是静态资源，是可以提前下载到客户端的。故利用资源推送模块将资源上传到 CDN，同时推送资源列表至客户端。推送过程基于微信消息系统实现，可以在极短的时间内把资源列表推送给几亿的用户。客户端随后提前从 CDN 下载资源至本地，春晚时即可无需下载资源，用户可获得较好的体验。

（2） **外网接入梳理**：所有的摇一摇请求都将会到达接入服务，预计春晚当晚会有 3.5 亿的在线人数，面对如此庞大的在线人群，保障外网接入质量的工作刻不容缓。要保证外网的接入质量，除了要保证接入服务本身的稳定性外，需要完成两个功能，一是在某些外网接入出现波动的时候，自动切换到正常接入服务，二是保障网络与服务具备足够的冗余容量。通过在上海 IDC 和深圳 IDC 各设置了 9 个接入集群。每个 IDC 都在 3 个不同的园区分别部署了电信、移动和联通的外网接入线路，实现了上述的基本功能，进而保证了系统的稳定性。

（3） **红包发放**：所有红包都源于红包系统，出于时效性的考虑，红包系统生成的种子红包文件被提前部署到接入服务。为确保红包文件不被重发，有个红包切分程序完成不同机器的文件切分，每台机器只保留自己需要处理的那部分红包；为确保红包不漏发，有一个校验程序完成所有机器红包文件的合并校验。未被用户分享的红包，将会被系统回收，一部分作为种子红包经由摇一摇 agent 回流到接入服务等待被发放。红包是每秒匀速下发的，故系统可精确控制全局红包的下发速度，保持在系统能够处理范围之内。

（4） **安全性保障**：为保证每个用户最多只可以领取 3 个红包，且每个赞助商最多 1 个。系统借鉴 HTTP 协议的 cookie 机制，在后台写入红包领取情况到 cookie，交由客户端保存，客户端在下次请求时发给服务器，服务器进而判断每个用户是否满足红包领取规则。然而恶意用户为获取额外收益等原因，会想办法绕开 cookie 检测机制。对于此类用户，通过红包发放汇总服务，在所有接入服务上同步已经达到红包领取上限的用户，依然可阻止其通过变换不同服务器来获取额外红包的目的。

（5） **春晚实时互动**：春晚摇一摇系统需要配合春晚现场进行互动，实时根据春晚节目信息或者主持人的播报，进行切换活动配置。这一任务的完成既要快又需稳。春晚现场有一个配置前台，现场操作人员可将实时配置通过前台发送至后台，后台有上海和深圳两个接收点。从前台发起变更，至接入服务加载

配置，全程可在 10 秒内完成。

（6） **红包即时分享**：在用户摇到红包、拆红包后，之后的分享红包操作和抢红包操作间存在着延迟问题。为尽可能的降低这种延迟性，获得较好的用户体验，微信红包系统增加了红包简化逻辑和红包异步队列这两个模块。

以上即为羊年春晚的摇一摇微信红包系统的一个较为详细的介绍与分析。为确保该系统在春晚当日能发挥预期的作用，腾讯进行了多轮的全程真实压测，并联合多部门进行了详细的 code review，以及在春晚前进行了两次预热，对系统进行了实战验证。从 2015 年春晚“摇一摇”全程摇动 110 亿次，峰值 8.1 亿/分钟的数据以及该系统的表现来看，该系统运作良好，成功实现了上述关键技术的目标。

参考资料：

微信红包系统设计分享|如何抗住 100 亿次请求？
http://www.woshipm.com/pd/232838.html
内容分发网络（CDN）
http://baike.baidu.com/link?url=Y8CtftGBVX5vGLvKIqbVp6q75NIBsyiQ2w__hYxAvBsrSVcNFL7SvVpc4_Xk-D1_n67GUDnwVFT0vp5Kr9lKFRsJX1MQOSZqvRYMk5DL3K3
互联网数据中心（IDC）
http://baike.baidu.com/link?url=mV7HWaX-ULu4XY49vinYG4SGNAvKSvxzUc1kiQLR2On3g-G0MpeOFddm9ysGO2-4F-FnnzkLc6XqpHbhLpuBIK

2. 现今分布式操作系统的挑战有： Heterogeneity（异构性）， Openness（开放性），Security（安全性）， Scalability（可伸缩性）， Failure handling（故障处理），Concurrency（并发性）， Transparency（透明性）， 服务质量等。 分别选择其中两个挑战， 给出定义， 举例与详细分析， 阐述你的观点和认识，论点论据相结合。（注意标引参考资料来源）

Heterogeneity（异构性）

定义：在一个分布式系统中，组成该系统的软硬件资源、系统内各成员交互所使用的网络、编写应用或系统所使用的编程语言以及开发者的对一项任务的实施方式等都可能是不同的，这就是分布式系统异构型的体现。

而要解决这各方面的异构性从而使得整个分布式系统能够运作良好，并且达到透明性的目标，并不简单。在一个分布式系统中，你所取得的资源有可能是远在地球另一边的服务器所提供的，在这样的情况下，信息的传输势必会经过不同类型的网络，而不同的网络遵循着各自的网络协议，不同协议间如何沟通，以使得信息能够正确的传输，是一个问题。同时，组成分布式系统的各计算机很有可能是不同的，举例而言，就整型数据的存储来说，采用大端方式存储的计算机与采用小端方式存储的计算机，其对于同一个数（回文数字除外）的存储就截然不同，如忽视这种差异，就可能会得到完全相反的结果。此外，不同的计算机上运行的操作系统也可能是不一样的，而不一样的操作系统其对于接口的实现又非常可能是不同的，如 UNIX 系统中对于消息交换的系统调用就与 Windows 系统不同。而不同的编程语言又有着对相同数据结构的不同表示方法，正确识别各编程语言下的数据结构，对于各语言彼此交流非常重要。一个大型的分布式系统的开发通常需要多位开发者，不同的开发者不可避免的存在着自己的编码习惯或者说编码风格，若没有一个统一的开发规范，最终形成的系统很可能不能够成为一个整体运作。

综上，对于任何一个分布式系统，异构性的问题都是亟需解决的一大难点。

Concurrency（并发性）

定义：在一个分布式系统中，存在着若干资源，这些资源可被用户或者说客户端同时访问，如何处理这些访问，以防系统中出现不一致的情况，此即并发性所带来的问题。

举例而言，若 12306 网站无法很好的处理并发，那么很有可能出现多个用户抢到同一张票的情况，尤其是在春运期间。同理，若微信红包系统无法很好的处理并发，那么很有可能出现多个用户抢到同一个红包的情况。上述的问题都是无法容忍的。而简单的利用加锁操作来限制多个用户同时进行操作，在访问需求非常大的情况下，又会降低吞吐量，使得系统性能下降，用户体验下降，故并非一个很好的解决方案。并发性的解决对于系统的稳定高效运行至关重要。

参考资料：

《Distributed Systems: Concepts and Design》 Fifth Edition, Page16-25, George Coulouris, etc.

网络类型

http://baike.baidu.com/link?url=VM4J-dVJBt9nyWnxe7Mo9-4BTyUe8tB7785HXiq1VWMU-kOjN5wFMOmUp_Z3gAjgR8V3Oiue1ooZ5KeUHmCizDhuljgNEb-IBvtfWixUiyyTv94_JPtOYcdYFIFvk9OA

大端与小端存储模式详解

http://blog.csdn.net/favory/article/details/4441361

微信红包系统设计&优化

http://djt.qq.com/article/view/1349