

# 复习资料

## (一) 微服务应用目前成熟的部署方案中，有哪些关键模块，分别的作用是什么？

微服务应用部署目前主要有负载均衡、服务注册发现、智能路由、自动熔断、日志服务、配置中心等关键模块。具体作用如下：

- ✓ 负载均衡：实现服务网关的负载均衡，常用解决方案 F5、Nginx。
- ✓ 服务注册发现：服务端通过 REST 协议暴露服务，提供应用服务的注册和发现的功能，常用解决方案 Eureka、Consul。
- ✓ 智能路由：进行安全验证、动态路由、负载分配等，常用解决方案 Zuul。
- ✓ 自动熔断：在服务出现超时、异常时，自动熔断，避免服务器瘫痪，常用解决方案：Hystrix。
- ✓ 统一日志服务：各应用日志文件采集到统一的服务器上，方便日志查看，常用解决方案：ELK(Elasticsearch、Logstash、Kibana)等。
- ✓ 分布式配置中心：例如有配置文件发生变化，会通过消息总线统一发布到服务，服务实时更新内存，不需要重启，常用解决方案：Spring Cloud Config、携程 Apollo。

## (二) 传统 BI 的关键技术包括哪些，分别详细描述。

传统 BI 的关键技术包括 ETL 数据抽取、转换、加载；联机分析处理；数据仓库模型设计；数据的发布（报表开发）等技术。

- ETL 数据抽取：在构造数据仓库时，要经过数据的清洗、数据的抽取转换、数据集成和数据加载等过程。面向不同的需求，对数据进行清洗以保证数据的正确性，然后对数据进行抽取，转换成数据仓库所需形式，并实现加载到数据仓库。

- 联机分析处理（OLAP）：又称多维分析，对数据仓库中的数据进行多维分析和展现，是使分析人员、管理人员或执行人员能够从多种角度对从原始数据中转化出来的真实反映企业特性的信息进行快速、一致、交互地存取，从而获得对数据更深入了解的一类软件技术。它的技术核心是“维”这个概念，因此 OLAP 也可以说是多维数据分析工具的集合。

- 数据仓库建模：数据仓库的数据模型有星型模式、雪花模式。星型模式最为常见，有一个包含大批数据并且不含冗余的中心表，每维一组小的附属表。雪花模式中某些维表是规范化的，因而把数据进一步分解到附加的表中，模式图形成了类似雪花的形状。

- 数据的发布：为了使分析后的数据直观、简练地呈现在用户面前，通常采用的是一些查询和报表工具和信息可视化技术。所谓信息可视化是指以图形、图像、虚拟现实等易为人们所辨识的方式展现原始数据间的复杂关系、潜在信息以及发展趋势，以便我们能够更好地利用所掌握的信息资源。

## (三) 描述一下保险行业大数据四层架构，在保险行业的应用场景有哪些，选择一个案例详细描述实施步骤。

保险行业大数据四层架构：

L1：基础设施、数据源：关键需要引入保险行业适配的数据，如中保信数据、医疗数据等等。

L2：数据平台层：统一的数据集成、存储和分析，统一的数据治理和应用平台

L3: 数据服务层: 整合第三方组件的能力, 如: DB、DWH, Hadoop、分析工具、可视化软件

L4: 数据应用层: 保险行业驱动的应用精准营销、智能核保、风险预警等

精准营销实施步骤:

#### 1、数据采集

数据采集分实时采集和批量采集。

实时数据采集内容: 用户的行为数据, 包括设备信息、浏览轨迹、页面停留时间、地域等。实时数据来源: 网销平台、APP、官网、微信公众号、EPOS、微信客服、呼叫中心、掌 e 通、互联网公开数据。实时采集方式: 前端或后端埋点、爬虫方案。

批量采集数据内容: 保单数据、理赔数据、保全数据、核保数据、财务分析数据。

批量数据来源: 关系型数据库、Excel、xml、JSON、日志文件、音频、视频文件等。

批量采集方式: 全量或增量同步。

#### 2、数据导入

实时数据采用 Flume+ Kafka+spark streaming, 批量数据采用 sqoop

#### 3、数据清洗及标准化

对多源多格式的数据进行标准化处理, 摒弃一些无效的数据。根据客户号、身份证号、港澳通行证、台湾同乡证、军官证、护照、驾照、车牌号、IP 地址、电话、mac 地址、邮箱、银行卡等数据对客户身份进行归一化处理。

#### 4、建立标签及客户画像

建立多维标签, 进行数据建模, 进行客户 360 度画像。

#### 5、机器建模

将客户的各种标签属性和行为特征交给精准营销机器学习模型进行建模和训练, 针对不同的场景建立客户购买意愿评分、产品推荐、客户忠诚度等模型。由于营销渠道、营销场景及营销目标等方面存在差异, 可以根据情况建立不同的模型。

#### 6、模型应用及验证

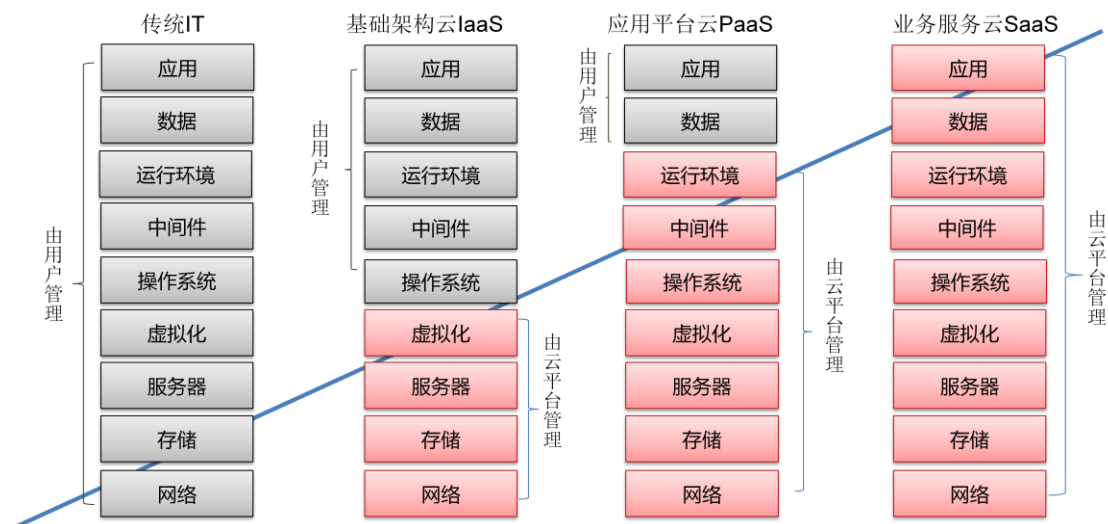
配合模型设计 3 个月的营销活动, 根据机器学习模型的预测目标, 对客户进行探索, 评估是否满足模型要求, 并对模型的有效性进行验证。

#### 7、模型迭代优化

营销活动结束后, 分析模型的实际预测效果, 并对模型进行优化。模型的优化需要不断整合更新的数据, 让机器学习模型能够使用更新的客户数据做进一步的模型训练。模型的优化过程是一个持续不断、周而复始的过程。

(四) 请用图形方式描述云服务架构的三个层次分别是什么？保险企业一般重点建设的是哪个层次的云架构？

基础架构云 IaaS、应用平台云 PaaS 和业务服务云 SaaS 是云服务提供的三种层次，最基础的是 IaaS，中间的为 PaaS，最后直观呈现出来的是 SaaS。如图：



#### 1) IaaS 基础设施即服务

(Infrastructure-as-a-Service)

- 拥有了 IaaS，就可以将IT外包到别的地方去。IaaS 会提供场外服务器，存储和网络硬件，也可以选择租用。节省了维护成本和办公场地，公司可以在任何时候利用这些硬件来运行其应用。
- 目前比较知名的 IaaS 公司有亚马逊、Bluelock、CSC、GoGrid、IBM 等。

#### 2) PaaS 平台即服务

(Platform-as-a-Service)

- 某些时候也被叫作中间件。所有的开发都可以在这一层进行，节省时间与资源。PaaS 可以提供各种开发和分发应用的解决方案，比如虚拟服务器和操作系统等，可以节省硬件上的费用。
- PaaS 公司与 IaaS 公司有许多重叠，除了上面列出的那些之外，还有 Google、Microsoft Azure、Force.com、Heroku、Engine Yard 等。

#### 3) SaaS 软件即服务

(Software-as-a-Service)

- SaaS 是我们目前普通用户接触最多的层面，在网络上任意一个远程服务器上的应用都是属于 SaaS。
- 比如现在阿里的钉钉、JIBUU 以及苹果的 iCloud 都属于这一类。比较知名的 SaaS 公司有 Salesforce、workday、Slack 等。

目前对于保险企业来说，主要建设的是 PaaS 云平台。