

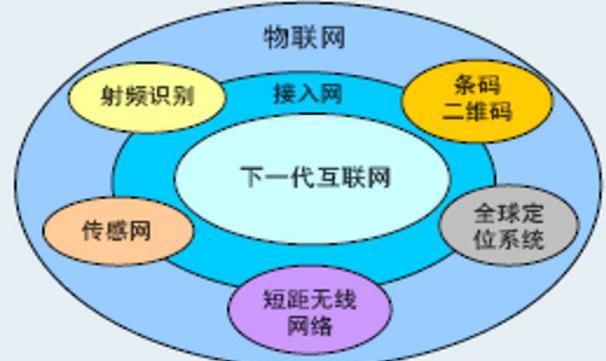
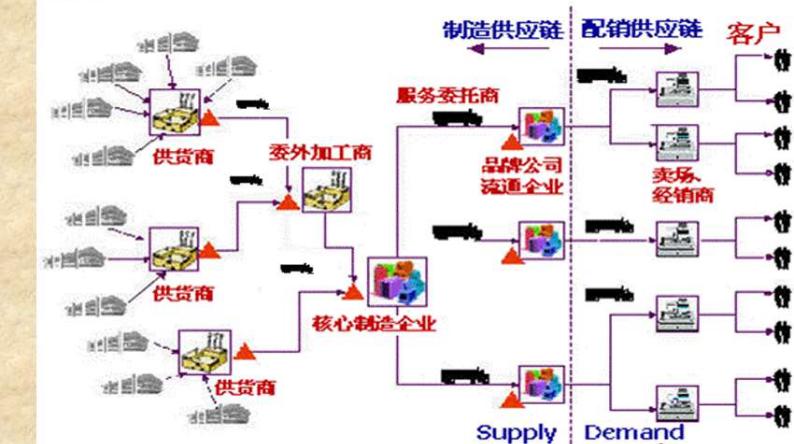
物流的定义

■ CLM定义

— 物流是供应链过程的一部分⁹⁸, 以满足客户需求为目的, 以高效和经济的手段来组织产品、服务⁹¹以及相关信息从供应到消费的运动和储存的计划、执行和控制的过程。



第四方物流——提供完整的供应链解决方案。



物联网概念模型

	智能Agent i不合作D	智能Agent i合作C
智能Agent j不合作D	4	4
智能Agent j合作C	1	1

刻画智能Agent i 对于可能结局的偏好:

$D, D \geq D, C \geq C, D \geq C, C$

智能Agent i 宁可选择不合作D产生的结局;



物联网技术体系模型

网站联盟营销

一个网站同意向给它推荐新业务机会的网站支付一定的佣金。

根据销售额付费

根据访问量或注册人数付费

合作伙伴

汽车用品 | 鲜花

折扣服装 | 月饼

最新秋装 | 女装

合作伙伴资讯

- 中秋送礼超值精选

- 李宁新品折扣卖场

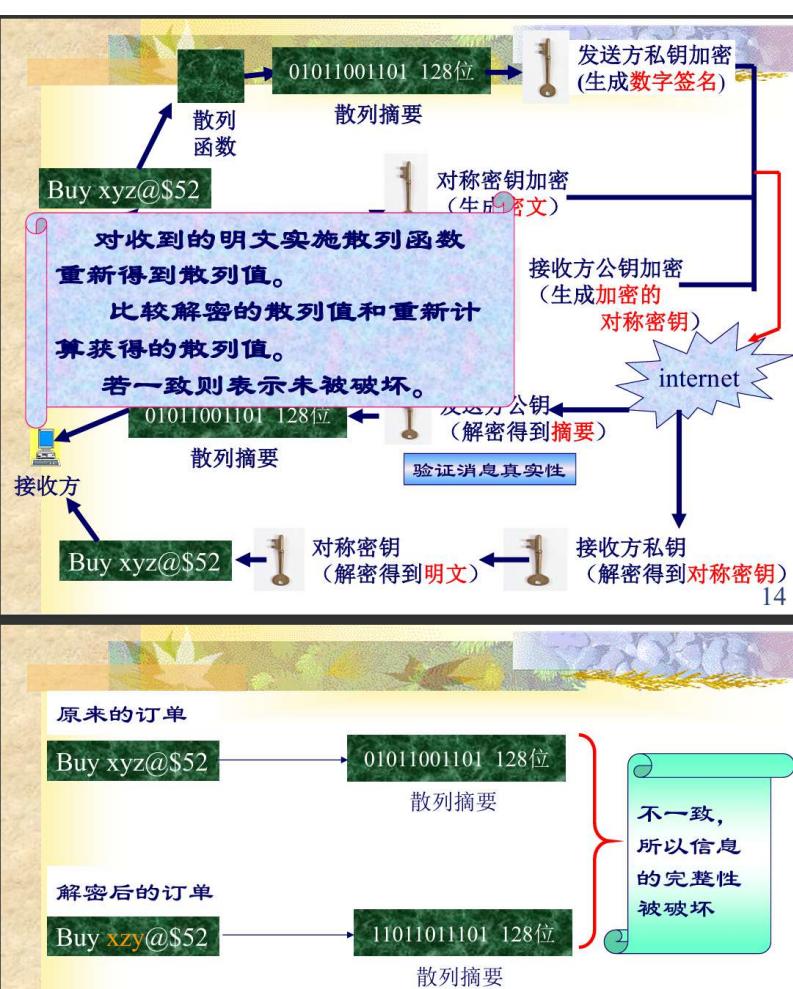
- 中秋好礼在莎啦啦

- 兰蔻品牌护肤5折

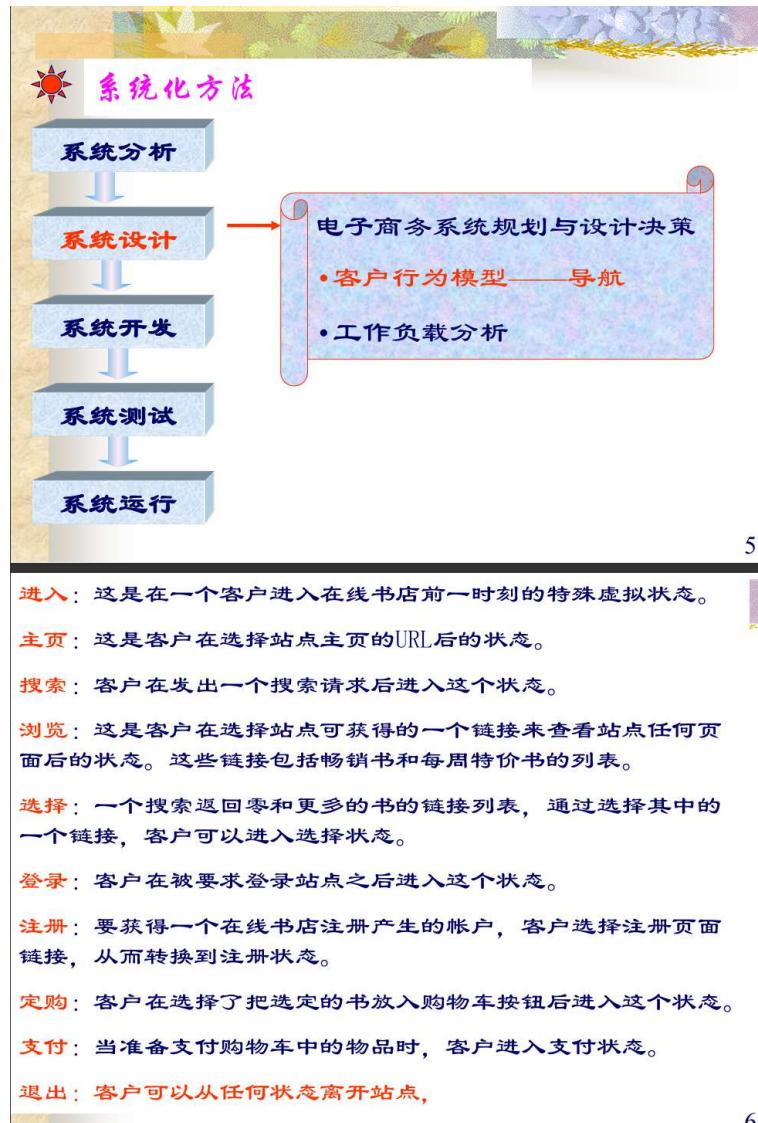
RSA的算法

- 选取两个质数P和Q; 如: P=11, Q=19
- 计算P和Q相乘所产生的乘积n = P×Q; 如: n=209
- $t=(P-1) \times (Q-1)$, 取任何一个数e, 要求满足 $e < t$ 并且e与t互素; $t=180$ 如: 取 $e=97$
- 另找一个数d, 使其满足 $(e \times d) \bmod [(P-1) \times (Q-1)] = 1$ (其中mod为相除取余); 如: 取 $d=13$
- (n, e) 即为公开密钥; (n, d) 即为私用密钥;
- 将明文X分组, $X=X_1X_2\dots X_r$ ($X_i \leq n$);
- 加密: $Y_i = (X_i^e) \bmod n$, 得密文 $Y=Y_1Y_2\dots Y_r$ 如: 明文c=02, 密文m=128
- 密文 $c = (m^e) \bmod n$ 明文 $m = (c^d) \bmod n$ 即无论哪一个质数先与原文加密, 均可由另一个质数解密。用一个质数来求出另一个质数, 则是非常困难的。

32



17



提 纲

① 云计算概念与发展现状

② 典型云计算技术方案

云计算是一种商业计算模型。它将计算任务分布在大量计算机构成的资源池上，使各种应用系统能够根据需要获取计算力、存储空间和信息服务。



亚马逊云计算AWS (Amazon Web Service)

- 弹性计算云EC2
- 简单存储服务S3
- 简单队列服务SQS



简单存储服务S3 (Simple Storage Service)

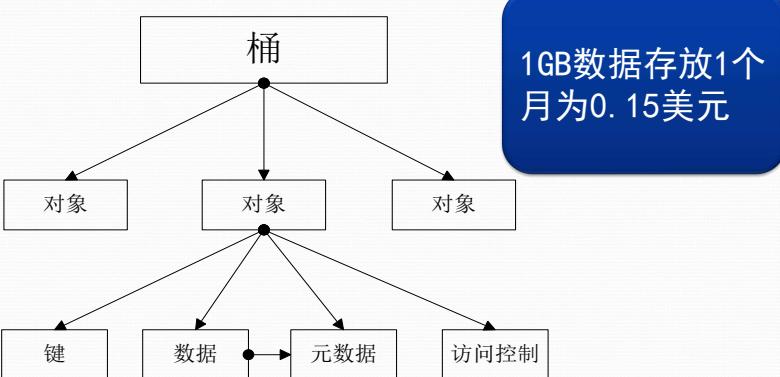
■ 基本概念

- 对象：S3的基本存储单元（数据、元数据），数据类型
- 键：对象的唯一标识符
- 桶：存储对象的容器（不能嵌套、在S3中名称唯一、每个用户最多创建100个桶）



简单存储服务S3 (Simple Storage Service)

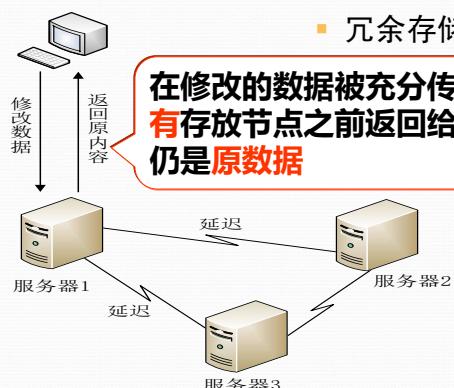
Amazon云计算



简单存储服务S3 (Simple Storage Service)

■ 冗余存储

在修改的数据被充分传播到所有存放节点之前返回给客户的仍是原数据



Amazon云计算



弹性计算云EC2 (Elastic Computing Cloud)

EC2 (Elastic Compute Cloud)

简言之，EC2就是一部具有无限采集能力的虚拟计算机，用户能够用来执行一些处理任务。

- **灵活性：**可自行配置运行的实例类型、数量，还可以选择实例运行的地理位置。

- **低成本：**按小时计费

Amazon云计算



弹性计算云EC2 (Elastic Computing Cloud)

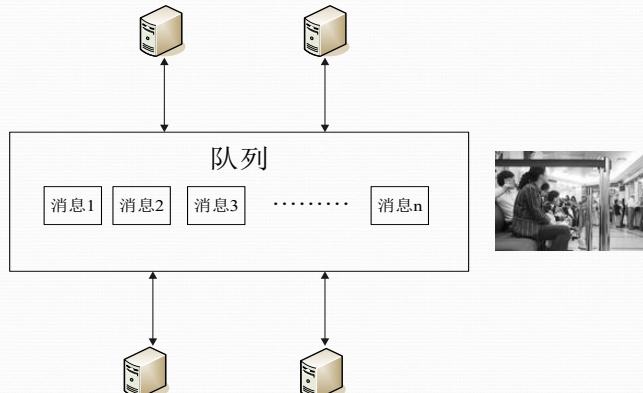
- **Amazon机器映像AMI (Amazon Machine Image)**

——由一个**操作系统**和当虚拟机启动时你想要预先载入的任何的**应用程序**组成。

——用户使用EC2服务的**第一步**就是要创建一个自己的AMI。

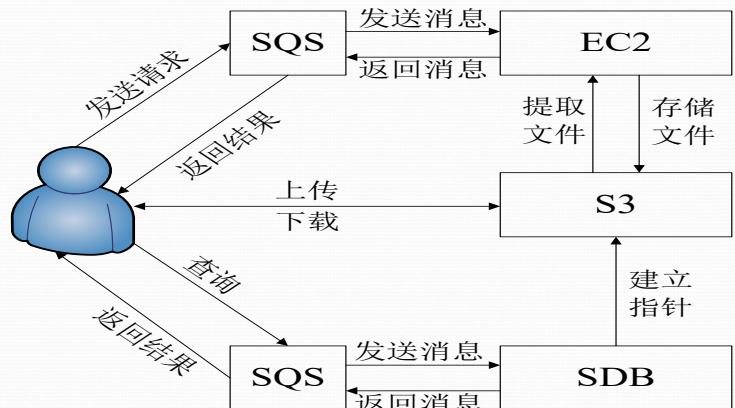


简单队列服务SQS (Simple Queue Service)



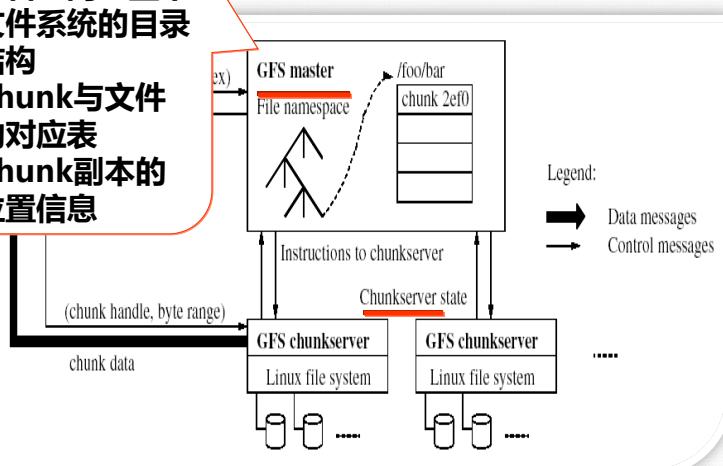
目标：解决低耦合系统间的通信问题

总结—AWS的结合使用



三种元数据
命名空间：整个文件系统的目录结构
Chunk与文件的对应表
Chunk副本的位置信息

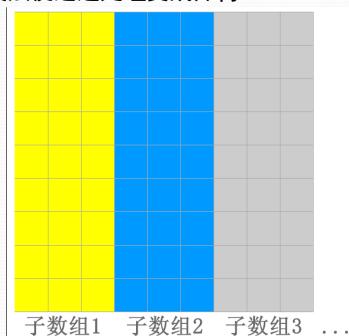
Google文件系统 (GFS)



并行计算基础

- 什么样的问题适合并行计算？

- 如果有大量结构一致的数据要处理，且数据可以分解成相同大小的部分，那我们就可以设法使这道处理变成并行



为什么需要MapReduce?

Jeffery Dean设计一个新的抽象模型，使我们只要执行的简单计算，而将并行化、容错、数据分布、负载均衡的等杂乱细节放在一个库里，使并行编程时不必关心它们。这就是MapReduce



Google MapReduce
架构设计师
Jeffrey Dean

逻辑视图

- 总体上，与关系数据库中的表类似

(row:string, column:string, time:int64) -> string

Row Key	Time Stamp	Column Contents	Column Anchor		Column "mime"
			cnnsi.com	my.look.ca	
"com.cnn.ww w"	T9		CNN		
	T8			CNN.COM	
	T6	"<html>.. "			Text/html
	T5	"<html>.. "			
	t3	"<html>.. "			

表中的数据通过行关键字和列关键字以及一个时间戳进行索引

引言

• 背景：

- 人工智能领域的巨大飞跃，尤其是大型深度学习模型的兴起，改变了处理信息和数据的方式。
- 大模型，如GPT-5，具有令人瞩目的上下文理解和学习能力。

• 关键特点：

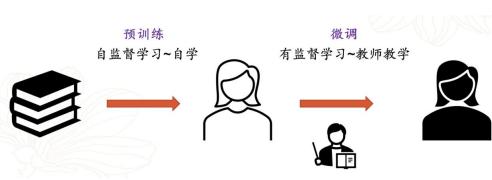
- 超越传统机器学习，大模型在处理复杂任务上表现出色。
- 通过深度学习，大模型能够从大规模数据中提取抽象模式，为各行业带来新的机遇。

• 为何关注金融领域：

- 金融领域对于准确、迅速的信息处理至关重要。
- 大模型在金融领域的应用有望优化决策、降低风险，并提供更智能的客户服务。

大模型

- 大模型是指参数规模庞大、层次深度的深度学习模型。
- 这些模型利用数以亿计的参数，通过训练数据自动学习任务，无需人为规定特定规则。



抑制模型幻觉，增强输出准确性

1. 提示工程(PE)

2. 模型微调(FT)

3. 检索信息增强 (RAG)

提示语的设计

1. 零样本提示 (Zero shot)

角色：希望模型扮演的角色
指令：指定您希望语言模型执行的任务或指令。
上下文：包含相关信息或额外上下文，以帮助语言模型更好地响应。
输入数据：您输入的内容或问题。
输出指示：指定您需要的输出类型或格式。

2. 少样本提示 (Few shot)

通过提供少量的样本就可以让gpt回答的更加精确。

3. 思维链 (Chain of thought)

通过让大模型解释其推理过程，从而实现更加精准的答案。

抑制模型幻觉，增强输出准确性

1. 提示工程(PE)

2. 模型微调(FT)

3. 检索信息增强 (RAG)

模型微调效果

模型微调前后效果对比：

输入：你好，我有一笔5万元的闲钱，近一两年可能用不到，想获得比定期存款高一点的收益，有什么推荐吗？，

原始模型输出：感谢您的咨询，投资涉及风险，选择产品前建议您充分了解自身的风险承受能力，并咨询专业的理财顾问以获得个性化的建议。……

微调后的模型输出：“根据您的情况——5万元闲置资金、投资期限1-2年、追求高于定存的稳健收益——我为您提供以下参考方向：

中短期纯债基金：主要投资于国债、金融债等，波动相对较小，收益通常高于同期定存，流动性也较好。适合您作为主要配置考虑。

银行中低风险理财 (R1/R2级)：您可以关注银行APP或网点销售的1-2年期封闭式理财，这类产品历史业绩较为稳健，是替代定存的常见选择。

• 结论：微调后的模型回复更专业，能够结合需求提供具体的理财推荐。

• 原因：微调使大模型具有领域知识。

模型微调简介

目的：使大模型能够理解并遵循人类的指令，具备对话的能力。

数据格式：指令 (instruction) 、输入 (input, 可选) 和输出 (output) 。

```
{  
  "instruction": "根据给定的描述，生成一个创意的营销文案，旨在吸引年轻人购买最新款智能手机。描述：这款手机拥有超强的摄像头和全天候的电池续航，适合喜欢拍照和玩游戏的年轻人。",  
  "input": "",  
  "output": "探索无限可能！这款全新智能手机，不仅拥有顶级摄像头记录每个瞬间，还有持久电力陪伴你畅玩每一刻。立即拥有，让生活更精彩！"  
}
```

微调数据示例

模型微调流程

数据收集

金融领域微调数据收集：

数据类别	数量 (条)	示例
金融术语解释	1259	INPUT：什么是固定资产 OUTPUT：固定资产是指为生产商品、提供服务、出租或经营管理而持有的、使用……
情感分析任务 ^[21]	2000	INPUT：假设你是一个股票文本情感分析专家，以下是一段文本，…… OUTPUT：假设你是一个专业机构的股票分析师，你需要对给定的股票的风险进行分析。……
财报分析任务	6000	INPUT：根据所提供的财务数据，可以看出该公司存在一些风险因素：1. 净利润波动大；公司净…… OUTPUT：假设你是一个专业机构的行业研究员，你需要对给定的新闻进行详细的分析……
新闻分析任务	1000	INPUT：假设你是一个专业机构的行业研究员，你需要对给定的新闻进行详细的分析…… OUTPUT：根据提供的新闻文本，我们可以分析以下几个方面……

9

模型微调流程

微调基座模型选择

- 假设需要微调一个多语言聊天机器人，则需要考虑基模型的多语言能力，
可以选择语言评分最高的DeepSeek-V2.5。



模型微调流程

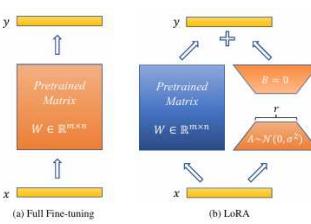
模型微调

重参数化方法-低秩适配微调：

对模型参数进行特定变换或替换，将原始的模型参数以某种形式重新表示或调整，以减少训练成本、提高训练效率

代表性方法：LoRA (目前最受欢迎的PEFT方法)

- 做法：对指定参数增加额外的低秩矩阵，也就是在原始PLM旁边增加一个旁路，做一个降维再升维的操作



LoRA方法示意图

- 假设需要处理一个金融理财问答任务，涉及复杂的知识查询和推理，
可以选择知识和推理评分都高的Qwen2.5-72B-instruction。

模型	评估日期	类别	分数	得分	语言	推理
Qwen2.5-72B-Instruct	2024/5/19	对话	728	57.3	57.3	73.5
Metal-Large-Instruct-2411	2024/1/18	对话	1238	54.1	54.1	73.3
DeepSeek-V2.5	2024/5/19	对话	2368	53.9	53.7	69.1

抑制模型幻觉，增强输出准确性

1. 提示工程(PE)

2. 模型微调(FT)

3. 检索信息增强 (RAG)

检索增强生成 (Retrieval-Augmented Generation, RAG)

RAG 检索增强：根据问题在知识库中检索相关的分析样例（知识），输入给大模型LLM，以提高模型的分析推理能力，即通过语义检索为LLM提供额外知识。

大模型的缺陷：

- 幻觉
- 信息过时
- 缺乏专业领域知识
- 推理能力弱

应用需求：

- 领域精准问答
- 数据频繁更新
- 生成内容可解释可溯源
- 成本可控
- 数据保护

ECUST

RAG流程——以财报分析为例

