

2023.6.25

Java二级学习

——学习笔记

猎奇人只爱无机盐

2023.7.12~2023. .

目录

[考试题型和分值 1](#_Toc31502)

[1.单选题（40分） 1](#_Toc32623)

[公共基础（10分） 1](#_Toc30013)

[专业知识（30分） 1](#_Toc7157)

[2.操作题（60分） 2](#_Toc7319)

[基本操作（如：填空等，18分） 2](#_Toc21696)

[简单应用题（如：改错或窗体应用等，24分） 2](#_Toc29898)

[综合应用题（如：Applet程序设计等，18分） 2](#_Toc5899)

[公共基础 2](#_Toc21856)

[第一章、数据结构与算法 2](#_Toc26247)

[1、概述 2](#_Toc23628)

[2、数据结构 2](#_Toc5698)

[3、算法 7](#_Toc4545)

[4、常见的数据结构 12](#_Toc29072)

[第二章 、程序设计基础 28](#_Toc12267)

[1.程序设计的方法与风格 28](#_Toc7539)

[2.结构化程序设计 28](#_Toc21824)

[3.面向对象方法 29](#_Toc32281)

[第三章 、软件工程基础 30](#_Toc24193)

[1.软件工程基本概念 30](#_Toc2401)

[2.软件生命周期 31](#_Toc19880)

[3.软件设计 31](#_Toc30353)

[4.结构化分析方法 32](#_Toc3406)

[第四章、数据库设计基础 32](#_Toc31898)

[1、数据库系统的基本概念 32](#_Toc7259)

[2、数据库的发展和基本特点 37](#_Toc7071)

[3、数据库的系统内部体系结构 39](#_Toc19438)

[4、数据模型的基本概念 43](#_Toc5022)

[5、数据库设计与原理 51](#_Toc18920)

[Java语言程序设计 54](#_Toc31078)

**java计算机二级**

# 考试题型和分值

## 1.单选题（40分）

### 公共基础（10分）

### 专业知识（30分）

## 2.操作题（60分）

### 基本操作（如：填空等，18分）

### 简单应用题（如：改错或窗体应用等，24分）

### 综合应用题（如：Applet程序设计等，18分）

# 公共基础

## 第一章、数据结构与算法

### 1、概述

* 用计算机解决现实生活与生产问题是计算机编程的核心意义
* 数据结构与算法正是为了高效的解决问题引入的两个概念
* 程序=算法+数据结构

• 数据结构是算法实现的基础

• 算法总是要依赖某种数据结构来实现的，算法的操作对象时数据结构

• 算法是编程思想，数据结构是这些思想的基础

### 2、数据结构

* 数据结构是计算机存储、组织数据的方式。
* 数据结构研究的三个方面：

• 1）数据集合中各数据之间固有的关系，即数据的逻辑结构；

• 2）对数据进行处理时，各元素在计算机中的存储关系，即数据的存储结构；

• 3）对各种数据结构的运算

* 数据结构的主要任务

• 是通过分析数据对象的结构特征，包括逻辑结构及数据对象之间的关系，然后把逻辑结构表示成计算机可实现的物理结构，从而便于计算机处理。

• 通常情况下，精心选择的数据结构可以带来更高的运行或者存储效率。

* 2.1、数据结构基本术语

• 1）数据(Data)：

• 是描述客观事物的符号，

• 是计算机中可以操作的对象，

• 是能被计算机识别，并输出给计算机处理的符号集合。

• 数据不仅仅包括整型、实型等数值类型，还包括字符及声音、图像、视频等非数值类型。

• 　2）数据元素（Data Element）：

• 数据的基本单位。

• 一个数据元素由若干数据项构成，在计算机中通常作为整体处理，也被称为记录。

• 　3）数据项（Data Item）：

• 数据项是数据不可分割的最小单位。

• 比如：数据(是一张表)->数据元素(行)->数据项(字段)。

• 　4）数据对象（Data Object）：

• 是性质相同的数据元素的集合，

• 是数据的子集。

• 5）数据类型（Data Type）：

• 是指一组性质相同的值的集合及定义在此集合上的一些操作的总称。

• 　6）数据结构（Data Structure）：

• 是数据的组织形式

• 是数据元素相互之间存在的一种或多种特定关系的数据元素的集合

* 2.2、数据结构的分类

• 2.2.1、逻辑结构

• 数据的逻辑结构，面向的是用户，是从具体问题抽象出来的数学模型，

• 是描述数据元素及其关系的数学特性的。

• 在任何问题中，数据元素都不会是孤立的，在它们之间存在着这样或那样的关系，这种数据元素之间存在的关系称为数据的逻辑结构。

• 分类

• 2.2.1.1、线性结构

• 在该结构中，数据元素除了同属于一个集合外，数据元素之间还存在着一对一的顺序关系。

• 每一个节点都有一个前驱和一个后继,表示为R={d1,d2...dn}，各节点按逻辑关系排列,形成一条'链'。

• 如果一个数据满足下面两个条件：

• 1）有且只有一个根节点；

• 2）每一个节点最多有一个前件，也最多有一个后件

• 则称该数据为线性结构。线性结构又叫线性表。

• 常见的线性结构包括：

• 串（字符串）、队列、链表和栈等。

• 2.2.1.2、非线性结构

• 如果一个数据结构不是线性结构，则称之为非线性结构。

• 树状结构：

• 树形结构中的数据元素之间存在一种一对多的层次关系，每一个元素最多有一个前驱,能够有多个后继,表示为(D,{R}),就像一个树干长了多个树枝具有分支。

• 图状结构：

• 该结构的数据元素之间存在着多对多的任意关系

• 图状结构也称为网状结构。

• 集合结构：

• 在集合结构中，数据元素之间的关系是“属于同一个集合”，表示为R{}。

• 数据元素之间除了同属一个集合外，不存在其他关系。

• 常见的非线性结构包括：

• 广义表，树结构，图结构，多维数组等。

• 2.2.2、存储结构

• 数据结构在计算机中的表示称为存储结构(或物理结构)

• 存储结构主要分为

• 顺序存储、链式存储、索引存储和散列(哈希)存储。

• 1）顺序存储（线性结构）：

• 数据被存储在一组连续的存储单元中，其数据间的逻辑关系和物理关系是一致的。

• 特点：易查难插（删）。

• 2）链式存储（线性结构）：

• 借助元素地址指针表示数据的逻辑结构,每一个元素都会包含指向下一个元素的指针

• 这种结构须要在节点上附加一个指针项,指出后继节点的位置

• 即每一个节点存储单元包含两个部分:[数据项,指针项]。

• 特点：易插（删）难查。

• 3）索引存储：

• 指存储在数据元素信息的同时还建立附加的索引表。

• 索引表中的所有索引项是有序排列的。

• 特点：查找效率高，但额外增加了空间的开销。

• 4）散列存储(哈希表)：

• 根据元素的关键字通过哈希（或散列）函数直接计算出一个值，并将这个值作为该元素的存储地址。

• 特点：查找速度快。

• 一个完整的存储结构必须能够存储数据元素,以及元素之间的逻辑关系。

• 2.2.3、数据的运算

• 数据的运算是指：施加在数据上的运算包括运算的定义和实现。

• 运算的定义是针对逻辑结构的，指出运算的功能；

• 运算的实现是针对存储结构的，指出运算的具体操作步骤。

• 常见的数据运算有：

• 插入（新增）、删除、修改（更新）、查找（检索）、排序等。

### 3、算法

* 概念：

• 算法，字面意思即计算方法，一个算法就是对特定问题求解步骤的一种描述。

• 算法规定了求解给定类型问题所需的全部处理步骤以及执行顺序,使得问题能在优先时间内机械的求解

• 再具体一点,算法是一段有穷的指令序列，算法必须能使用某种语言描述。

• 注：算法并不等于公式，公式是一种高度精简的算法，算法的形式可以比公式更复杂，解决的问题更加广泛。

* 3.1、算法特性

• 1）有穷性：

• 指算法在执行有限的步骤之后，自动结束而不会出现无限循环，并且每一个步骤在可接受的时间内完成；

• 2）确定性：

• 算法的每一步骤都具有确定的含义，不会出现二义性；

• 3）可行性：

• 算法的每一步都必须是可行的，也就是说，每一步都能够通过执行有限次数完成；

• 4）输入：

• 是指算法具有零个或多个输入；

• 5）输出：

• 是指算法至少有一个或多个输出；

* 3.2、算法的分类

• 1）根据应用分类：

• 按照算法的应用领域，可以分为

• 基本算法

• 数据结构相关算法

• 几何算法

• 图论算法

• 规划算法

• 数值分析算法

• 加密解密算法

• 排序算法

• 查找算法

• 并行算法

• 数值算法

• 等等

• 2）根据确定性分类：

• 确定性算法：

• 有限时间内完成，得到结果唯一；

• 非确定性算法：

• 有限时间内完成，得到结果不唯一，存在多值性。

• 3）根据算法的思路分类：

• 递推算法

• 递归算法

• 穷举算法

• 贪婪算法

• 分治算法

• 动态规划算法

• 迭代算法

• 等

* 3.3、算法的表示

• 一般有三种流程结构：

• 顺序结构

• 分支结构

• 循环结构

* 3.4、算法的性能

• 算法的效率作为判断算法优劣的标准。一个算法的优劣往往通过算法复杂度来衡量

• 分类：

• 时间复杂度

• 时间复杂度是指执行算法所需要的计算工作量

• 即通常所说的算法执行所需要耗费的时间，时间越短，算法越好。

• 算法的时间度量，记作：T（n）=O（f（n））。

• 它表示随问题规模n的增大，算法执行时间的增长率和f（n）的增长率相同。

• 常见的时间复杂度量级有：

•



• 空间复杂度

• 空间复杂度是指执行这个算法所需要的内存空间。

• 是对一个算法在运行过程中临时占用存储空间大小的量度，记做S(n)=O(f(n))。

• 其中，n为问题的规模，f(n)为语句关于n所占存储空间的函数。

• 一般而言，常常以算法执行时间作为算法优劣的主要衡量指标。

### 4、常见的数据结构

* 4.1、数组

• 4.1.1、概述

• 数组是一种聚合数据类型，它是将具有相同类型的若干变量有序地组织在一起的集合。

• 数组可以说是最基本的数据结构，在各种编程语言中都有对应。

• 一个数组可以分解为多个数组元素

• 按照数据元素的类型，数组可以分为

• 整型数组

• 字符型数组

• 浮点型数组

• 指针数组

• 结构数组

• 等

• 数组还可以有一维、二维(比较重要，大题必考)以及多维等表现形式。

• 4.1.2、数组运算

• 遍历：

• 遍历所有元素；

• 插入：

• 将一个或多个元素插入数组；

• 删除：

• 从数组中删除元素；

• 搜索：

• 在数组中搜索元素。您可以按元素的值或索引搜索元素；

• 更新：

• 在给定索引处更新现有元素的值。

• 4.1.3、数组的应用

• 用作构建其他数据结构的基础

• 例如数组列表，堆，哈希表，向量和矩阵；

• 用于不同的排序算法

• 例如插入排序，快速排序，冒泡排序和合并排序。

* 4.2、链表

• 4.2.1、概述

• 链表是一种顺序结构，由相互链接的线性顺序项目序列组成。

• 这种存储结构具有在物理上存在非连续的特点。

• 链表由一系列数据结点构成，每个数据结点包括数据域和指针域两部分。

• 其中，指针域保存了数据结构中下一个元素存放的地址。

• 链表结构中数据元素的逻辑顺序是通过链表中的指针链接次序来实现的。

• 链式存储方式既可用于表示线性结构，也可用于表示非线性结构。

• 4.2.2、链表分类

• 单链表：

• 只能沿正向遍历项目；

• 双链表：

• 可以在前进和后退方向上遍历项目。节点由一个称为上一个的附加指针组成，指向上一个节点；

• 循环链接列表：

• 链接列表，其中头的上一个指针指向尾部，尾号的下一个指针指向头；

• 静态链表：

• 线性存储结构的一种，兼顾顺序表和链表的优点，是顺序表和链表的升级；

• 静态链表的数据全部存储在数组中(顺序表)，但存储的位置是随机的。（因为物理位置连续的，逻辑位置不一定连续）

• 4.2.3、链表操作

• 搜索：

• 通过简单的线性搜索在给定的链表中找到键为k的第一个元素，并返回指向该元素的指针

• 插入：

• 在链表中插入一个元素x。

• 插入可以通过3种不同的方式完成

• 在列表的开头插入

• 在列表的末尾插入

• 在列表的中间插入

• 头结点用来记录链表的基地址，是链表遍历的起点

• 尾结点的后继指针不指向任何结点，而是指向一个空地址NULL。

• 删除：

• 从给定的链表中删除元素x。

• 删除可以通过3种不同方式完成

• 从列表的开头删除

• 从列表的末尾删除

• 从列表的中间删除

• 4.2.4、链表的应用

• 用于编译器设计中的符号表管理。

• 用于在使用Alt Tab（使用循环链表实现）的程序之间进行切换。

• 注：

• 数组无法动态扩容

• 如果申请空间不够，需要申请更大的空间，需要数据复制拷贝进入新的数组，非常耗时；

• 而链表天然支持动态扩容，因为他不要内存空间的连续性

• LinkedHashMap的底层实现就是用的双向链表结构。

• 多线程下Hashmap扩容时出现循环链表，jdk1.8把头插法换成了尾插法。

* 4.3、栈

• 4.3.1、概述

• 栈是一种特殊的线性表，

• 它只能在一个表的一个固定端进行数据结点的插入和删除操作。

• 栈按照"后进先出"或“先进后出”的原则来存储数据

• 先插入的数据将被压入栈底，

• 最后插入的数据在栈顶，

• 读出数据时，从栈顶开始逐个读出。

• 栈中没有数据时，称为空栈。

• 4.3.2、栈的顺序存储及其运算

• 栈的基本运算有3种：

• 入栈

• 入栈运算：在栈顶位置插入一个新元素；

• 退栈

• 退栈运算：取出栈顶元素并赋值给一个指定的变量；

• 读栈顶元素

• 读栈顶元素：将栈顶元素赋值给一个指定的变量；

* 4.4、队列

• 4.4.1、概述

• 队列和栈类似，也是一种特殊的线性表。

• 和栈不同的是，队列只允许在表的一端进行插入操作，而在另一端进行删除操作。

• 一般来说，进行插入操作的一端称为队尾，进行删除操作的一端称为队头。

• 队列中没有元素时，称为空队列。

• 4.4.2、队列操作

• 进队：

• 将元素插入队列的末尾。

• 出队：

• 从队列的开头删除元素。

• 4.4.3、队列的应用

• 用于管理多线程中的线程。

• 用于实施排队系统（例如：优先级队列）。

• 扩展理解：例题

• 循环队列的存储空间为 Q(1:40) ，初始状态为 front=rear=40 。经过一系列正常的入队与退队操作后， front=rear=15 ，此后又退出一个元素，则循环队列中的元素个数为（ ）。

• A、39，或0且产生下溢错误

• B、14

• C、40

• D、15

• 注：当队列front=rear=15时，此时队列可能满了或空了，再退出元素，可能是39，或0且产生下溢错误。

• 循环队列

• 基本概念：

• 队列是一种先进先出（FIFO）的线性表

• 循环队列即为队列的顺序表示和实现，

• 其优点主要是入队和出队的时间复杂度均为O（1）

• 这归功于“循环队列将存储结构的理解由序列型转换为环形”的思维方式

• 基本属性：

• 两个指针（其实就是两个整数型变量，因为在这里有指示作用，所以这里理解为指针）

• front

• 一个指示队头

• rear

• 一个指示队尾

• rear和front互相追赶着，这个追赶过程就是队列添加和删除的过程

• 如果rear追到front说明队列满了

• 如果front追到rear说明队列为空

• 即当rear=front时，队列要么为空要么为满

• 我们把他掰弯，用的是求余，这样两个值就不会跑出最大范围，并且可以实现弯曲的效果

• 所以说对于循环队列我们必须给定最大值MAXQSIZE

• 实现过程：

• 添加：

• （rear+1）%MAXQSIZE

• 在jdk实现中，弯曲的效果是tail=(tail+1)&(elements.length-1)，这里elements.length初始值是16，每次容量扩大会乘以2，这里用到公式：当x=2^n（n为自然数）时，a%x=a&(x-1)

• 删除：

• （front+1）%MAXQSIZE

• 当rear=front时，队列

• 可能是满的

• 当队列添加元素到rear的下一个元素是head的时候，也就是转圈子要碰头了，我们就认为队列满了。

• (Q.rear+1)%MAXQSIZE=Q.front

• 也可能是空的

• 当队列删除元素到front=rear的时候，我们认为队列空了。

• Q.rear==Q.front，不一定为0

* 4.5、树

• 4.5.1、树的概述

• 树是典型的非线性结构，它是包括，2个结点的有穷集合K。

• 在树结构中，有且仅有一个根结点，该结点没有前驱结点。

• 在树结构中的其他结点都有且仅有一个前驱结点，而且可以有两个后继结点，m≥0。

• 4.5.2、树的相关概念

• 父节点：

• 每一个节点都只有一个前件，称为父节点，没有前件的节点只有一个，称为根节点；

• 子节点：

• 每一个节点可以有多个后件，称为该节点的子节点，没有后件的节点，称为叶子节点；

• 度：

• 一个节点所拥有的后件个数称为该节点的度，所有节点中最大的度称为树的度；

• 深度：

• 树中结点的最大层次数称为树的深度或高度；

• 4.5.3、树的种类

• 无序树：

• 树中任意节点的子结点之间没有顺序关系，这种树称为无序树,也称为自由树;

• 有序树：

• 树中任意节点的子结点之间有顺序关系，这种树称为有序树；

• 二叉树：

• 每个节点最多含有两个子树的树称为二叉树；

• 完全二叉树

• 满二叉树

• 斜树

• 平衡二叉树

• 霍夫曼树：

• 带权路径最短的二叉树称为哈夫曼树或最优二叉树；

• 4.5.4、二叉树

• 特点：

• 每个结点最多只有两个子结点，可以没有或者只有一个。

• 左子树和右子树是有顺序的，次序不能任意颠倒。

• 性质：

• 在二叉树的第k层上，最多有2k-1个结点；

• 深度为m的二叉树最多有2m-1个结点；

• 具有n个结点的二叉树，其深度至少为log2n+1

• 4.5.5、满二叉树

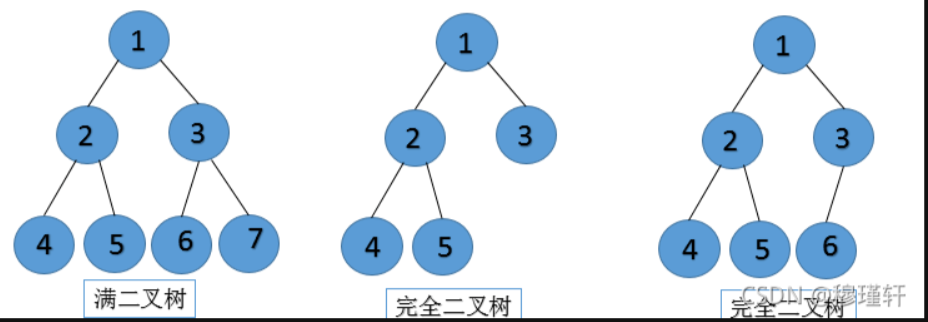
• 除最后一层（叶子节点）外，每一层上的所有结点都有两个子结点。

• 4.5.6、完全二叉树

• 除最后一层外，每一层上的结点数都达到最大值，在最后一层上只缺少右边的若干结点。

• 满二叉树也是完全二叉树，反过来不一定成立。

•



• 4.5.7、平衡二叉树（AVL）

• 它是一 棵空树或它的左右两个子树的高度差的绝对值不超过1，并且左右两个子树都是一棵平衡二叉树，同时，平衡二叉树必定是二叉查找树。

• 4.5.8、二叉查找树

• 若它的左子树不为空，则左子树上所有节点的值均小于它的根节点的值；

• 若它的右子树不为空，则右子树上所有节点的值均大于它的根节点的值；

• 它的左、右子树也分别是二叉排序树。

• 说明它是一颗有顺序的树，左子树节点的值小于根节点的值，右子树节点的值大于根节点的值。

• 红黑树：

• 红黑树的节点是有颜色概念的，即非红即黑。

• 通过颜色的约束，红黑树维持着二叉搜索树的平衡性。

• 一颗红黑树必须满足以下几点条件：

• 1）根节点必须是黑色。

• 2）任意从根到叶子的路径不包含连续的红色节点。

• 3）任意从根到叶子的路径的黑色节点总数相同。

• 4.5.9、霍夫曼树

• 带权路径最短的二叉树称为哈夫曼树或最优二叉树。

• 4.5.10、B-树

• B-树是一种平衡多路查找树，它在文件系统中很有用。

• 4.5.11、B+树

• B+数是B-树的一种变形。

• 4.5.12、字典树（trie树）

• 字典树是一种以树形结构保存大量字符串。

• 以便于字符串的统计和查找，经常被搜索引擎系统用于文本词频统计。

• 优点：

• 利用字符串的公共前缀来节约存储空间，最大限度地减少无谓的字符串比较，查询效率比哈希表高。

• 4.5.13、后缀树

• 包含一则字符串所有后缀的压缩了的字典树。

* 4.6、堆

• 4.6.1、概述

• 堆(Heap)是一种重要的数据结构，是实现优先队列(Priority Queues)首选的数据结构。

• 由于堆有很多种变体，包括二项式堆、斐波那契堆等，但是这里只考虑最常见的就是二叉堆。

• 特点：

• 根结点的值是所有结点中最小的或者最大的

• 并且根结点的两个子树也是一个堆结构。

• 4.6.2、堆的应用

• 用于实现优先级队列，因为可以根据堆属性对优先级值进行排序。

• 可以在O（log n）时间内使用堆来实现队列功能。

• 用于查找给定数组中k个最小（或最大）的值。

• 用于堆排序算法。

* 4.7、图

• 4.7.1、概述

• 图是另一种非线性数据结构。

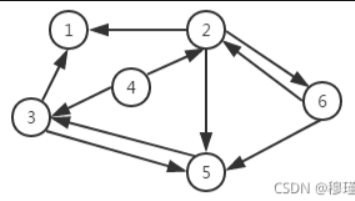
• 在图结构中，数据结点一般称为顶点，而边是顶点的有序偶对。

• 如果两个顶点之间存在一条边，那么就表示这两个顶点具有相邻关系。

• 4.7.2、有向图

• 一个图结构中，边是有方向性的，那么这种图就称为有向图。

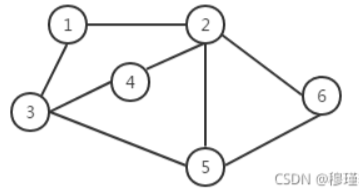
•



• 4.7.3、无向图

• 如果一个图结构中，所有的边都没有方向性，那么这种图便称为无向图。

•



• 4.7.4、图的应用

• 用于表示社交媒体网络。

• 每个用户都是一个顶点，并且在用户连接时会创建一条边。

• 用于表示搜索引擎的网页和链接。

• 互联网上的网页通过超链接相互链接。

• 每页是一个顶点，两页之间的超链接是一条边。

• 用于Google中的页面排名。

• 用于表示GPS中的位置和路线。

• 位置是顶点，连接位置的路线是边。

• 用于计算两个位置之间的最短路径。

* 4.8、哈希表

• 4.8.1、概述

• 散列表源自于散列函数(Hash function)

• 思想：

• 如果在结构中存在关键字和T相等的记录，那么 必定在F(T)的存储位置可以找到该记录，这样就可以不用进行比较操作而直接取得所查记录。

• 4.8.2、哈希表的应用

• 用于实现数据库索引。

• 用于实现关联数组。

• 用于实现"设置"数据结构。

## 第二章 、程序设计基础

### 1.程序设计的方法与风格

* 1.1.源程序文档化

• 符号名的命名

• 程序注释

• 视觉组织

* 1.2.数据说明的方法
* 1.3.语句的结构程序
* 1.4.输入和输出

### 2.结构化程序设计

* 2.1.结构化程序的设计原则

• 自顶向下

• 逐步求精

• 模块化

• 限制使用goto语句

* 2.2.结构化程序的基本结构

• 顺序结构

• 选择结构

• 循环结构

### 3.面向对象方法

* 3.1.对象

• 对象的特征：

• 标识唯一性

• 分类性

• 多态性

• 封装性

• 模块独立性

* 3.2.类和实例

• 类是具有共同属性、共同方法的集合

• 对象则是对其对应类的一个实例

* 3.3.消息

• 消息是实例之间传递的信息

• 它请求对象执行某一处理或回答某一要求的信息

• 它统一了数据流和控制流

* 3.4.继承

• 广义地说，继承是指能够直接获得已有的性质和特征，而不必重复定义它们

* 3.5.多态性

• 对象根据所接收的消息而做出动作，同样的消息被不同的对象接收时可能导致完全不同的行动。

## 第三章 、软件工程基础

### 1.软件工程基本概念

* 1.1.定义与特点

• 研究的主要内容包括：

• 软件开发技术

• 软件工程管理

* 1.2.软件工程三要素：

• 方法

• 工具

• 过程

### 2.软件生命周期

* 2.1.软件定义期

• 问题定义

• 可行性研究

• 需求分析

* 2.2.软件开发期

• 概要设计

• 详细设计

• 实现

• 测试

* 2.3.运行维护期

• 使用和维护

### 3.软件设计

* 3.1.概述
* 3.2.基本原理

• 抽象

• 模块化

• 信息隐藏

• 模块独立性

* 3.3.相关概念

• 内聚性

• 耦合性

### 4.结构化分析方法

* 4.1.概述
* 4.2.常用工具

• 数据流图（DFD）

• 数据字典（DD）

• 判定表

• 判定树

## 第四章、数据库设计基础

### 1、数据库系统的基本概念

* 1.1、数据

• 数据(Data)实际上就是描述事物的符号记录。

• 计算机中的数据一般分为两部分

• 其中一部分与程序仅有短时间的交互关系，随着程序的结束而消亡，它们称为临时性（Transient ）数据，这类数据一般存放于计算机内存中

• 另一部分数据则对系统起着长期持久的作用，它们称为持久性（Persistent ）数据。数据库系统中处理的就是这种持久性数据。

• 软件中的数据是有一定结构的

• 数据有型（Type ）与值（Value）之分

• 数据的型给出了数据表示的类型，如整型、实型、字符型等

• 数据的值给出了符合给定型的值，如整型值15

• .随着应用需求的扩大它包括了将多种相关数据以一定结构方式组合构成特定的数据框架，这样的数据框架称为数据结构（Data structure ）,数据库中在特定条件下称之为数据模式（Data schema ） .

• 在过去的软件系统中是以程序为主体，而数据则以私有形式从属于程序，此时数据在系统中是分散、凌乱的，这也造成了数据管理的混乱，如数据冗余度高，数据一致性差以及数据的安全性差 等多种弊病。

• 近10多年来，数据在软件系统中的地位产生了变化，在数据库系统以及数据库应用系统中数据已占有主体地位，而程序已退居附属地位。

• 在数据库系统中需要对数据进行集中、统一 的管理，以达到数据被多个应用程序共享的目标。

* 1.2、数据库

• 数据库（Database ,简称DB ）是数据的集合，它具有统一的结构形式并存放于统一的存储介质内，是多种应用数据的集成，并可被各个应用程序所共享。

• 数据库存放数据是按数据所提供的数据模式存放的，它能构造复杂的数据结构以建立数据间内在联系与复杂的关系，从而构成数据的全局结构模式。

• 数据库中的数据具有“集成”、“共享”之特点，亦即是数据库集中了各种应用的数据，进行统一的构造与存储，而是他们可以被不同的应用程序所使用。

* 1.3、数据库管理系统

• 数据库管理系统（ Database Management System, 简称 DBMS ）是数据库的机构，它是一种系统软件，负责数据库中的数据组织、数据操纵、数据维护、控制及保护和数据服务等。

• 数据库中的数据是具有海量级的数据，并且其结构复杂，因此需要提供管理工具。

• 数据库管理系统是数据库系统的核心

• 具体功能：

• (1)数据模式定义。

• 数据库管理系统负责为数据库构建模式，也就是为数据库构建其数据框架。

• (2) 数据存取的物理构建。

• 数据库管理系统负责为数据模式的物理存储及构建提供有效的存取方法与手段。

• (3) 数据操纵。

• 数据库管理系统为用户使用数据库中的数据提供方便，它一般提供查询、插入、修改、删除数据的功能。此外，它自身还具有做简单算术运算及统计的能力，而且还可以和某些过程性语言结合，使其具有强大的过程性操作能力。

• (4)数据的完整性、安全性定义与检查。

• 数据库中的数据具有内在语义上的关联性与一致性，他们构成了数据的完整性，数据的完整性是保证数据库中的数据正确的必要条件，因此必须经常检查以维护数据的正确。

• 数据库中的数据具有共享性，而数据共享可能会引发数据的非法使用，因此必须要对数据正确使用作出必要的规定，并在使用时做检查，这就是数据的安全性。

• 数据完整性与安全性的维护是数据库管理系统的基本功能。

• (5)数据库的并发控制与故障恢复。

• 数据库是一个集成、共享的数据集合体，它能为多个应用程序服务，所以就存在着多个应用程序对数据库的并发操作。

• 在并发操作中如果不加以控制和管理，多个应用程序间就会相互干扰，从而对数据库中的数据造成破坏。

• 因此，数据库管理系统必须对多个应用程序的并发执行做必要的控制一保证数据不受破坏，这就是数据库的并发控制。

• 数据库中的数据一旦遭受破坏，数据库管理系统必须有能力及时进行恢复，这就是数据库的故障恢复。

• (6) 数据的服务。

• 数据库管理系统提供对数据的多种服务功能，如数据拷贝、转存、重组、性能检测、分析等。

•

• 1.3.1、数据语言

• 数据定义语言（ Data Definition Language 简称 DDL ）：

• 该语言负责数据的模式定义与数据的物理存取构建。

• 数据操纵语言（ Data Manipulation Language 简称 DML ）：

• 该语言负责数据的操纵，包括查询及增、删、改等操作。

• 数据控制语言（ Data Control Language 简称 DCL ）：

• 该语言负责数据完整性、安全性定义与检查以及并发控制、故障恢复等功能，包括系统初起程序、文件读写与维护程序、存取路径管理程序、缓冲区管理程序、安全性控制程序、完整性检查程序、并发控制程序、事务管理程序、运行日志管理程序、数据库恢复程序等。

• 1.3.2、使用方式

• 交互式命令(又称自含型或自主型语言)

• 它的语言简单，能在终端上即时操作，它又称为自含型或自主型语言。

• 宿主型语言（一般可嵌入某些宿主语言中）

• 它一般可嵌入某些宿主语言（ Host Language ）中，如 C ， C++ 和 COBOL 等高级过程性语言中。

• 此外，数据库管理系统还为用户提供服务的服务性（ Utility ）程序，包括数据初始装入程序、数据转存程序、性能检测程序、数据库再组织程序、数据转换程序、通信程序等。

• 目前流行的DBMS 均为关系数据库系统，比如 ORACLE 、 Sybase 的 PowerBuilder 及 IBM 的 DB2 、微软的 SQL Server 等，他们均为严格意义上的 DBMS 系统。

• 另外一些小型的数据库，如微软的 Visual Foxpro 和 Access 等，它们只具备数据库管理系统的一些简单功能

* 1.4、其他相关概念

• 数据库管理员：

• 对数据库进行规划、设计、维护、监视等的专业管理人员。

• 数据库系统：

• 由数据库（数据）、数据库管理系统（软件）、数据库管理员（人员）、硬件平台（硬件）、软件平台（软件）五个部分构成的运行实体。

• 数据库应用系统：

• 由数据库系统、应用软件及应用界面三者组成。

### 2、数据库的发展和基本特点

* 2.1、数据库的发展

• 人工管理阶段

• 在 20 世纪 50 年代中期以前，主要用于科学计算，硬件无磁盘，直接存取，软件没有操作系统。

• 20 世纪 50 年代后期到 20 世纪 60 年代中期，进入文件系统阶段。

• 文件系统是数据库发展的初级阶段，它提供了简单的数据共享与数据管理能力。

• 但是它无法提供完整的、统一的、管理和数据共享的能力。

• 由于它的功能简单，因此它附属于操作系统而不成为独立的软件，目前一般将其看成是数据库系统的雏形，而不是真正的数据库系统。

• 20 世纪 60 年代之后，数据管理进入了数据库系统阶段。

• 层次数据库与网状数据库阶段 从 20 世纪 60 年代末期起，真正的数据库系统 ------ 层次数据库与网状数据库开始发展，它们为统一管理与数据共享提供了有力支持，这个时期数据库系统蓬勃发展形成了有名的“数据库时代”。

• 关系数据库系统出现于 20 世纪 70 年代，在 80 年代得到蓬勃发展，并逐渐取代前两种系统，关系数据库系统结构简单，使用方便，逻辑性强物理性少，因此在 80 年代以后一直占据着数据库系统的主导地位。

* 2.2、数据库系统的基本特点

• 数据的集成性：

• (1)在数据库系统中采用统一的数据结构方式，如在关系数据库中采用二维表作为统一结构方式。

• (2) 在数据库系统中按照多个应用的需要组织全局的统一的数据结构（即数据模式），数据模式不仅可以建立全局的数据结构，还可以建立数据间的语义关系从而构成一个内在紧密联系的数据整体。

• (3) 数据库系统中的数据模式是多个应用共同的、全局的数据结构，而没个应用的数据则是全局数据结构中的一部分，称为局部结构（即视图），这种全局与局部的结构模式构成了数据库系统数据集成性的主要特征。

• 数据的高共享性与低冗余性：

• 数据的共享自身又可极大的减少数据冗余性，不仅减少了不必要的存储空间，更为重要的是可以避免数据的不一致性。

• 数据独立性（物理独立性与逻辑独立性）：

• 数据独立性是数据与程序间的互不依赖性，即数据库中的数据独立于应用程序而不依赖于应用程序。

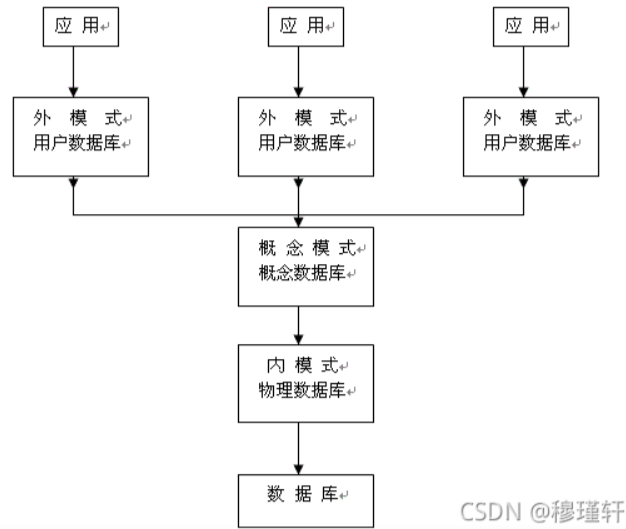
• 也就是说，数据的逻辑结构、存储结构和存取方式的改变不会影响应用程序。

• 数据统一管理与控制：

• 数据库系统不仅为数据提供高度集成环境，同时它还为数据提供统一管理的手段：数据的完整性检查、数据的安全性保护、并发控制。

### 3、数据库的系统内部体系结构

* 数据库系统在其内部具有三级模式及二级映射，三级模式分别是概念模式，内部级模式与外部级模式，二级映射分别是概念级到内部级的映射以及外部级到概念级的映射。



• 3.1、三级模式

• 概念模式（ Conceptual Schema ）：

• 是数据库系统中全局数据逻辑结构的描述，是全体用户（应用）公共数据视图。

• 此种描述是一种抽象的描述，它不设计具体的硬件环境与平台，也与具体的软件环境无关。

• 概念模式主要描述数据的概念记录类型以及它们间的关系，它还包括一些数据间的语义约束，对它的描述可用 DBMS 中的 DDL 语言定义。

• 外模式（ External Schema ）：

• 又称用户模式（ User`s schema ），它是用户的数据视图，也就是用户所见到的数据模式，它由概念模式推导而出。

• 概念模式给出了系统全局的数据描述而外模式则给出每个用户的局部数据描述。

• 一个概念模式可以有若干个外模式，每个用户只关心与它有关的模式，这样不仅可以屏蔽大量无关信息而且有利于数据保护。

• 在一般的 DBMS 中都提供有相关的外模式描述语言（外模式 DDL ）。

• 内模式（ Internal schema ）：

• 又称物理模式（ Physical Schema ），它给出了数据库物理存储结构与物理存取方法，如数据存储文件结构、索引、集簇及 hash 等存取方式与存取路径，内模式的物理性主要体现在操作系统及文件级上，它还未深入到设备级上（如磁盘及磁盘操作）。

• 内模式对一般用户是透明的，但它的设计直接影响数据库的性能。

• DBMS 一般提供相关的内模式描述语言（内模式 DDL ）。

• 这三种数据库中只有物理数据库是真实存在于计算机外存中的，其它两种数据库并不存在于计算机中，而是通过两种映射由物理数据库映射而成。模式的三个级别层次反映了模式的三个不同环境以及它们的不同要求，其中内模式处于最底层，它反映了数据在计算机物理结构中的实际存储形式，概念模式处于中层，它反映了设计者的数据全局逻辑要求，而外模式处于最外层，它反映了拥护对数据的要求。

• 3.2、两级映射

• 两级映射也保证了数据库系统中数据的独立性，亦即数据的物理组织改变与逻辑概念级改变相互独立，使得只要调整映射方式而不必改变用户模式。

• 概念模式到内模式的映射：

• 该映射给出了概念模式中数据的全局逻辑结构到数据的物理存储机构间的对应关系，此种映射一般由 DBMS 实现。

• 外模式到概念模式的映射：

• 概念模式是一个全局模式而外模式是用户的局部模式。

• 一个概念模式中可以定义多个外模式，而每个外模式是概念模式的一个基本视图。

• 外模式到概念模式的映射给出了外模式与概念模式的对应关系，这种映射一般也是由 DBMS 来实现的。

• 扩展理解：例题

• 在数据库的三级模式结构中，外模式有

• A．1个

• B．2个

• C．3个

• D．任意多个

• 注：外模式也称子模式或用户模式，是数据库用户（包括应用程序员和最终用户）能够看见和使用的局部数据的逻辑结构和特征的描述，是数据库用户的数据视图，是与某一应用有关的数据的逻辑表示。一个数据库可以有多个外模式，内模式只能有一个。

### 4、数据模型的基本概念

* 4.1、数据模型的基本概念

• 是数据特征的抽象，从抽象层次上描述了系统的静态特征、动态行为和约束条件，为数据库系统的信息表与操作提供一个抽象的框架。

• 数据模型所描述的内容有三个部分

• 4.1.1、数据结构

• 数据模型中的数据结构主要描述数据的类型、内容、性质以及数据间的联系等。

• 数据结构是数据模型的基础，数据操作与约束均建立在数据结构上。

• 不同数据结构有不同的操作与约束，因此，一般数据模型的分类均以数据结构的不同而分。

• 4.1.2、数据操作

• 数据模型中的数据操作主要描述在相应数据结构上的操作类型与操作方式。

• 4.1.3、数据约束

• 数据模型中的数据约束主要描述数据结构内数据同的语法、语义联系，它们之间的制约与依存关系，以及数据动态变化的规则，以保证数据的正确、有效与相容。

* 4.2、数据模型

• 数据模型按不同的应用层次分成三种类型

• 概念数据模型（ conceptual data model ）

• 概念数据模型简称概念模型，它是一种面向客观世界、面向用户的模型；

• 它与具体的数据库管理系统无关，与具体的计算机平台无关。

• 概念模型着重于对客观世界复杂事物的结构描述及它们之间的内在联系的刻画。

• 概念模型是整个数据模型的基础。

• 目前，较为有名的概念模型有 E-R 模型、扩充的 E-R 模型、面向对象模型及谓词模型等。

• 4.2.1、 E-R 模型

• 实体：

• 现实世界中的事物；

• 属性：

• 事物的特性；

• 联系：

• 现实世界中事物间的关系。

• 实体集的关系有一对一、一对多、多对多的联系。

• E-R模型三个基本概念之间的联接关系：

• 实体是概念世界中的基本单位，属性有属性域，每个实体可取属性域内的值。

• 一个实体的所有属性值叫元组。

• E-R模型的图示法：

• （1）实体集表示法；

• 实体集表示法（矩形表示实体）：

• 如，实体集学生（student）和课程（course）

desc

• （2）属性表法；

• 属性表法（椭圆形表示属性）：

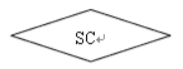
• 如，学生有属性：学号（S#）、姓名（Sn）及年龄（Sa）

desc

• （3）联系表示法

• 联系表示法（菱形表示联系）：

• 如，学生与课程的类型SC



• 逻辑数据模型（ logic data model ）

• 逻辑数据模型又称数据模型，它是一种面向数据库系统的模型，该模型着重于在数据库系统一级的实现。

• 概念模型只有在转换成数据模型后才能在数据库中得以表示。

• 目前，逻辑数据模型也有很多种，较为成熟并先后被人们大量使用过的有；层次模型、网状模型、关系模型、面向对象模型等。

• 4.2.2、层次模型

• 层次模型的基本结构是树形结构

• 具有以下特点：

• 每棵树有且仅有一个无双亲结点，称为根；

• 树中除根外所有结点有且仅有一个双亲。

• 4.2.3、关系模型

• 关系的数据结构关系模型采用二维表来表示，简称表。

• 二维表由表框架 (Frame) 及表的元组—（ Tuple ）组成。

• 表框架由 n 个命名的属性（ Attribute ）组成， n 称为属性元数 (Arity) 。

• 每个属性有一个取值范围称为值域（ Domain ）。

• 表框架对应了关系的模式，即类型的概念。

• 在表框架中按行可以存放数据，每行数据称为元组，实际上，一个元组是由 n 个元组分量所组成，每个元组分量是表框架中每个属性的投影值。

• 一个表框架可以存放 m 个元组， m 称为表的基数（ Cardinality ）。

• 一个 n 元表框架及框架内 m 个元组构成了一个完整的二维表。

• 在二维表中凡能唯一标识元组的最小属性称为键或码。

• 从所有侯选健中选取一个作为用户使用的键称主键。

• 表A中的某属性是某表B的键，则称该属性集为A的外键或外码。

• 关系中的数据约束：

• 实体完整性约束：

• 约束关系的主键中属性值不能为空值；

• 参照完全性约束：

• 是关系之间的基本约束；

• 用户定义的完整性约束：

• 它反映了具体应用中数据的语义要求。

• 物理数据模型（ physical data model ）

• 物理数据模型又称物理模型，它是一种面向计算机物理表示的模型，此模型给出了数据模型在计算机上物理结构的表示。

* 4.3、关系代数

• 关系数据库系统的特点之一是它建立在数据理论的基础之上，有很多数据理论可以表示关系模型的数据操作，其中最为著名的是关系代数与关系演算。

• 关系模型的基本运算：

• 1）插入

• 2）删除

• 3)修改

• 4）查询（包括投影、选择、笛卡尔积运算）

• 五个基本操作：

• 并(∪)

• 差(-)

• 笛卡尔积(×)

• 投影(π)

• 选择(σ)

• 四个组合操作：

• 交(∩)

• 联接(等值联接)

• 自然联接(RS)

• 除法(÷)

• 4.3.1、关系代数表达式

• 建立一个学生选课的关系数据库，它由下面三个关系模式组成：

• 学生关系S(SNO,SNAME,AGE,SEX)

• 学生关系S（学生学号，姓名，年龄，性别）

• 学习关系SC(SNO,CNO,GRADE)

• 学习关系SC（学生学号，课程号，成绩）

• 课程关系C(CNO,CNAME,TEACHER)

• 程关系C（课程号，课程名称，老师）

• 选择(SELECT)运算：

• 查询表中年龄小于20的学生

• SQL方式：

• select \* from S where AGE< 20

• 关系代数方式：

• δAGE<20(S)

• 投影(PROJECTION)运算：

• 投影操作是从列的角度进行运算，查询学生的姓名和年龄，即求S关系上学生姓名和年龄两个属性上的投影

• SQL方式：

• select S.SNAME,S.AGE form S

• 关系代数方式：

• π SNAME,AGE(S)

• 连接(JOIN)运算：

• 连接运算也称为θ连接，从两个关系（R、S）中的广义笛卡尔积中选取属性间满足一定条件的元组形成一个新的连接，θ不为“=”的连接称为非等值连接， θ为“=”的连接称为等值连接。检索学习课程号为C2的学生学号与姓名

• SQL方式：

• select SC.SNO,S.SNAME from SC,S where SC.SNO=S.SNO and SC.CNO='C2'

• 关系代数方式：

• π SNO,SNAME (σCNO='C2'(SSC))

• 其他运算：

• 检索选修课程号为C2或C4的学生学号

• SQL方式：

• select SNO FROM SC where CNO='C2' or CNO='C4'

• 关系代数方式：

• π SNO(σ CNO='C2'∨CNO='C4'(SC))

• 检索至少选修课程号为C2或C4的学生学号

• SQL方式：

• select SA.SNO from SC AS SA,SC AS SB where SA.SNO=SB.SNO and SA.CNO='C2' and SB.CNO='C4'

• 关系代数方式：

• π 1(σ1=4∧2='C2'∧5='C4'（SC×SC）)

•



### 5、数据库设计与原理

* 5.1、数据库设计概述

• 在数据库应用系统中一个核心问题就是设计一个能满足用户要求，性能良好的数据库，这就是数据库设计（Database design）

• 数据库设计的基本任务是根据用户对象的信息需求、处理需求和数据库的支持环境（包括硬件、操作系统与DBMS）设计出数据模式。

• 在数据库设计中有两种方法，

• 一种是以信息需求为主，兼顾处理需求，称为面向数据的方法（data-orented approach）;

• 另一种方法是以处理需求为主，兼顾信息需求，称为面向过程的方法（process-oriented）。

• 数据库设计目前一般采用生命周期（life cycle）法，即将整个数据库应用系统的开发分解成目标独立的若干阶段。

• 它们是：需求分析阶段、概念设计阶段、逻辑设计阶段、物理设计阶段、编码阶段、测试阶段、运行阶段、进一步修改阶段。

* 5.1.1、数据库概念设计

• 数据库概念设计的方法有以下两种：

• 1） 集中式模式设计法这是一种统一的模式设计方法，

• 她根据需求由一个统一机构或人员设计一个综合的全局模式。

• 这种方法设计简单方便，它强调统一与一致，适用于小型或并不复杂的部门或单位，而对大型的或语义关联复杂的单位则并不适合。

• 2） 视图集成设计法这种方法是将一个单位分解成若干个部分，先对每个部分作局部模式设计，建立各个部分的视图，然后以各视图为基础进行集成

• 在集成过程中可能会出现一些冲突，这是由于视图设计的分散性形成的不一致所造成的，因此需对视图作修正，最终形成全局模式。

• 视图集成设计法是一种由分散到集中的方法，它的设计过程复杂但它能较好地反映需求，适合于大型与复杂的单位，避免设计的粗糙与不周到，目前此种方法使用比较多。

• 视图设计一般有三种设计次序：

• 自顶向下

• 由底向上

• 由内向外

• 视图集成的几种冲突：

• 命名冲突

• 概念冲突

• 域冲突

• 约束冲突

* 5.1.2、数据库的逻辑设计

• 数据库的逻辑设计主要工作是将E-R图转换成指定RDBMS中的关系模式。

• 首先，从E-R图到关系模式的转换是比较直接的，实体与联系都可以表示成关系，E-R图中属性也可以转换成关系的属性。实体集也可以转换成关系。

• 逻辑设计的另一个重要内容是关系视图的设计，它又称为外模式设计。

• 关系视图是在关系模式基础上所设计的直接面向操作用户的视图，它可以根据用户需求随时创建，一般RDBMS 均提供关系视图的功能。

• 关系视图的主要作用：

• 1）提供数据逻辑独立性；

• 2）能适应用户对数据的不同需求；

• 3）有一定数据保密功能。

* 5.1.3、数据库的物理设计

• 数据库的物理设计的主要目标是对数据库内部物理结构作调整并选择合理的存取路径，以提高数据库访问速度及有效利用存储空间，

• 在现代关系数据库中已大量屏蔽了内部物理机构，因此留给用户参与物理设计的余地并不多，一般的RDBMS中留给用户参与物理设计的内容大致有如下几种：

• 索引设计

• 集簇设计

• 分区设计

# Java语言程序设计