

# 2023秋“大数据软件开发与实践”课程简介

---

## 基本信息

---

本学期授课覆盖课程代码CS32278和CS32283。

### CS32278-大数据软件开发与实践

课程代码：CS32278

课程名称：大数据软件开发与实践

学时：32

上课班级：2103501,21W0351

人数：41

上课时间：3-4周

### CS32283-大数据软件开发与实践

课程代码：CS32283

课程名称：大数据软件开发与实践

学时：64

上课班级：21R0351

人数：5

上课时间：3,4,6,8周

## 授课教师

邹兆年（8学时）、杨东华（8学时）、苗东菁（8学时）、石胜飞（4学时）、刘显敏（8学时）、张炜（8学时）、张开旗（8学时）、王宏志（8学时）。

## 教学内容

---

CS32278（32学时）教学内容包括2项任务：任务1和任务2。

CS32283（64学时）教学内容包括4项任务：任务1-4。

# 任务1：B+树索引实现及可视化软件开发

## 实验目的

1. 加深对B+树基本工作原理的理解。
2. 掌握B+树的算法细节和实现方法。

## 实验过程

1. 定义B+树的数据结构。
2. 实现B+树的基本操作，包括点查找索引项、区间查找、插入索引项、删除索引项。
3. 实现B+树的可视化展示。

## 实验提示

1. 编程语言不限。
2. 访问<https://www.cs.usfca.edu/~galles/visualization/BTree.html>，了解一个类似的B+树可视化工具。
3. B+树的可视化既可以在GUI界面上编程实现，也可以使用Graphviz等图可视化软件实现。访问<https://graphviz.org>获取有关Graphviz的资料。
4. 可以额外实现CS33503数据库系统课程中讲述的B+树优化方法和批量构建。
5. 熟悉Java语言的同学可以结合CS33503数据库系统课程实验的lab2（磁盘管理器），实现基于磁盘的B+树（需要额外实现缓冲区管理器）。
6. 熟悉C++语言的同学可以基于Rucbase来实现B+树（需要额外实现存储管理）。Rucbase提供了实现B+树的代码框架。<https://github.com/ruc-deke/rucbase-lab>
7. 查阅《数据库系统内幕》，了解更多有关B+树的高级知识。

# 任务2：线性哈希表索引实现及可视化软件开发

## 实验目的

1. 加深对线性哈希表基本工作原理的理解。
2. 掌握线性哈希表的实现方法。

## 实验过程

1. 定义线性哈希表的数据结构。
2. 实现线性哈希表的基本操作，包括点查找索引项、插入索引项、删除索引项。
3. 实现线性哈希表的可视化展示。

## 实验提示

1. 编程语言不限。
2. 线性哈希表的可视化既可以在GUI界面上编程实现，也可以使用Graphviz等图可视化软件实现。访问<https://graphviz.org>获取有关Graphviz的资料。

3. 熟悉Java语言的同学可以结合CS33503数据库系统课程实验的lab2（磁盘管理器），实现基于磁盘的线性哈希表（需要额外实现缓冲区管理器）。
4. 熟悉C++语言的同学可以基于Rucbase来实现线性哈希表。Rucbase没有提供线性哈希表的代码框架，需要自己定义线性哈希表的数据结构和操作函数。<https://github.com/ruc-deke/rucbase-lab>

## 任务3：查询计划的可视化软件开发

### 实验目的

1. 加深对查询计划概念的理解。
2. 掌握查询计划解析和可视化实现方法。

### 实验过程

1. 定义表示查询计划的数据结构。
2. 解析DBMS给出的查询计划，并用上面定义的数据结构表示。
3. 实现查询计划的可视化展示。

### 实验提示

1. 编程语言不限。
2. 访问<https://explain.dalibo.com/>，了解一个类似的查询计划可视化工具。
3. 查询计划的可视化既可以在GUI界面上编程实现，也可以使用Graphviz等图可视化软件实现。访问<https://graphviz.org>获取有关Graphviz的资料。

## 任务4：缓冲池的可视化软件开发

### 实验目的

1. 加深对缓冲池基本工作原理的理解。
2. 了解在DBMS上查看缓冲池状态的方法。
3. 通过对缓冲池进行可视化，直观展示缓冲池的状态变化。

### 实验过程

1. 在PostgreSQL上通过pg\_buffercache视图获得缓冲池当前的状态。
2. 实现一个可视化程序，直观展示缓冲池的当前状态。

### 实验提示

1. 编程语言不限。
2. 访问<https://www.postgresql.org/docs/9.1/pgbuffercache.html>，了解如何在PostgreSQL上通过pg\_buffercache视图获得缓冲池当前的状态。

## 教学计划

---

序号	周	日	节	地点	教学内容	上课班级	任课教师
1	3	星期六（9月16日）	1-4	格物208	任务1	2103501,21W0351,21R0351	邹兆年
2	3	星期六（9月16日）	5-8	格物208	任务1	2103501,21W0351,21R0351	邹兆年
3	3	星期日（9月17日）	1-4	格物208	任务1	2103501,21W0351,21R0351	杨东华
4	3	星期日（9月17日）	5-8	格物208	任务1	2103501,21W0351,21R0351	杨东华
5	4	星期六（9月23日）	1-4	格物208	任务2	2103501,21W0351,21R0351	苗东菁
6	4	星期六（9月23日）	5-8	格物208	任务2	2103501,21W0351,21R0351	苗东菁
7	4	星期日（9月24日）	1-4	格物208	任务2	2103501,21W0351,21R0351	石胜飞
8	4	星期日（9月24日）	5-8	格物208	任务2	2103501,21W0351,21R0351	石胜飞
9	6	星期六（10月7日）	1-4	格物208	任务3	21R0351	刘显敏
10	6	星期六（10月7日）	5-8	格物208	任务3	21R0351	刘显敏
11	6	星期日（10月8日）	1-4	格物208	任务3	21R0351	张炜
12	6	星期日（10月8日）	5-8	格物208	任务3	21R0351	张炜
13	8	星期六（10月21日）	1-4	正心401	任务4	21R0351	张开旗
14	8	星期六（10月21日）	5-8	正心401	任务4	21R0351	张开旗
15	8	星期日（10月22日）	5-8	正心401	任务4	21R0351	王宏志
16	8	星期日（10月22日）	9-12	正心401	任务4	21R0351	王宏志