

数据库开发课程 练习、作业与考核

南京大学软件学院

自我介绍

数据库开发技术助教

南京大学软件学院 软件工程

南京大学智能软件工程实验室

博士二年级在读~~牛马~~

研究方向：复杂分布式软件系统质量保障

林哲远 zheyuanlin@smail.nju.edu.cn



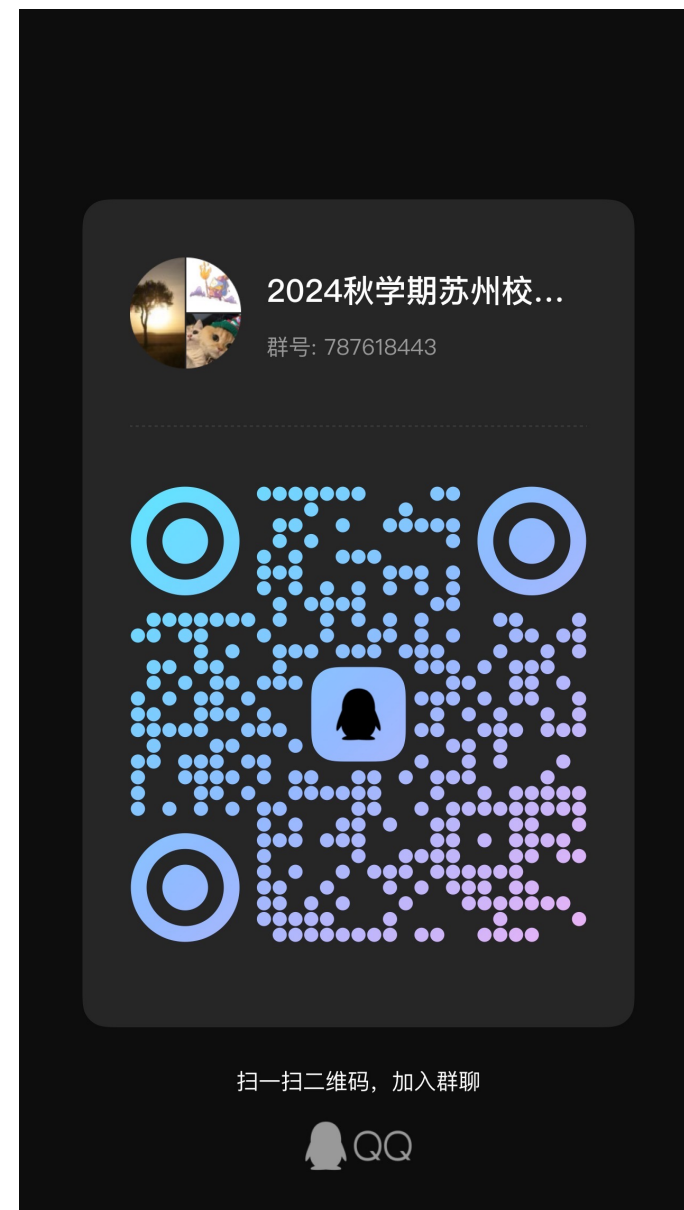
欢迎通过邮件/QQ和我沟通

考核组成

Grading:

- Practice 40%
 - SQL Exam 20%
 - Implementation of miniDataBase 20%
- Final Exam 60%

比例是明确的，分数是稳定的，努力了是能过的。



SQL练习

SQL 在线平台

练习系统：

1. 请在校园网或校园网VPN环境下，访问SQL Platform 网址: <http://172.29.4.28>。
2. 注册并自由练习。
3. 练习题目不提供统一的标准答案，请同学们独立/借助LLM解决。如有需求可以与助教进行探讨。
4. 题目可能有bug，会热修(请随时发邮件拷打助教谢谢)，不算分，请大家放心。

考试系统：

1. 请在校园网或校园网VPN环境下，访问SQL Exam网址: <http://172.29.4.69>。
2. 考前会发布测试考试，考试系统和练习系统有一些不一样。
3. 11月下旬和12月上旬进行2次考试，取最高分，考试内容为SQL编程题。

SQL 练习系统

按知识点筛选

题目名称	已通过	提交未通过	完成进度	操作
公司职员	1 / 4	1 / 4	<div><div></div></div> 25%	✍ 去做题
船只租赁	1 / 4	2 / 4	<div><div></div></div> 25%	✍ 去做题
顾客与商品	1 / 4	0 / 4	<div><div></div></div> 25%	✍ 去做题
选手与比赛	0 / 4	0 / 4	<div><div></div></div> 0%	✍ 去做题
顾客与商品2	0 / 4	0 / 4	<div><div></div></div> 0%	✍ 去做题
好友申请	0 / 4	0 / 4	<div><div></div></div> 0%	✍ 去做题
阅读记录	0 / 4	0 / 4	<div><div></div></div> 0%	✍ 去做题
网吧登录	1 / 4	0 / 4	<div><div></div></div> 25%	✍ 去做题
岗位成绩	0 / 4	0 / 4	<div><div></div></div> 0%	✍ 去做题

SQL 练习系统

公司职员

返回

设有一个公司内部信息管理数据库，表结构如下：

有 employees 表，

Column Name	Type
eno	int
ename	varchar
salary	int
dno	int

eno为该表主键。
该表包含员工的工号，姓名，工资，部门编号等信息

有 projects 表，

Column Name	Type
pno	int
pname	varchar
city	varchar
dno	int

pno为该表主键。
该表包含项目的项目编号，项目名称，所在城市，负责部门编号等信息。

有 works 表，

Column Name	Type
-------------	------

问题1

问题2

问题3

问题4

EASY 编写一个sql语句，查询累计工作时间超过1000的职工，结果返回职工工号eno。

数值

分组

聚合函数

深色模式

1 SELECT * FROM employees

☆ 收藏

📄 保存草稿

▶ 运行

✔ 提交

错误检查，时间可能比较久

2024-10-09 15:08:25 未通过

🔄

SQL 考试系统

- 考试每次包含4~5道题目，时间3小时。
- 2次机会，取最高分。所以，第二次可以不来，但要提前发邮件申请，避免我十点在群里@你，让其他同学考完被凡尔赛到。
- 线上平台MySQL版本为5.7版本，注意语法差异(with等)，近期在加8.x的容器，争取考试前上线，但题目都可以5.7版本完成。
- 考试期间不得使用除电脑外的手机或其他电子设备，禁止访问校外网页，不允许使用大模型辅助答题，如发现出现违规情况，本次考试计0分。

为什么不允许使用LLM

- 去年有很多同学有疑问，为什么不允许使用LLM
 - 使用LLM完成作业——加速任务完成
 - 使用LLM完成考试——虽然不明白，但我完成了
 - 幻觉，和人类一样但又不一样
 - 理解、审查和修改SQL语句的能力，或者说背锅的能力是最重要的


SQL 考试系统

SQL EXAM

当前考试

历史考试

Xie Kailian

 剩余时间: 581:43:41

交卷

展开查看数据库表结构

mysql 5.7

保存代码, 快捷键ctrl+s/command+s

代码格式化, 快捷键shift+enter

大题:

有如下表:

blank_data表:

id	name	amount
1	Tom	?
2	Jerry	?

题目的场景描述

address_data表:

name	address
Tom	?
Jerry	?

小题:

查询Tom和Jerry的存款

一个大题的题目场景对应多个小题, 每个小题会要求你编写sql完成相应要求

20分

1

select * from |

sql答题区域

10分

2

运行并保存

SQL 考试系统

数据集

剩余时间：259:41:58

交卷

查看公告

搜索表名、列名

tables

customers

customer_id int
name varchar

orders

order_id int
order_date date
customer_id int
product_id int

products

大题:

对于顾客购买产品，系统中有以下表结构:

表: customers

Column Name	Type
customer_id	int
name	varchar

customer_id 是该表主键。
该表包含消费者的id和姓名。

表: orders

Column Name	Type
order_id	int
order_date	date
customer_id	int
product_id	int

order_id 是该表主键。
该表包含消费者产生的订单编号，订单日期，顾客id和商品id。
不会有商品被相同的用户在一天内下单超过一次。

表: products

Column Name	Type
product_id	int
product_name	varchar
price	int

product_id 是该表主键。
该表包含所有商品id，商品名称和商品价格。

表的示例如下:

customers

小题:

< 1 2 3 4 >

编写一个sql语句，查询每件商品都购买过的用户，结果返回用户编号customer_id。

openGauss作业实践

数据库简史

数据库是一门非常“长寿”的学科，历史往往重演(What Goes Around and Comes Around……And Around)

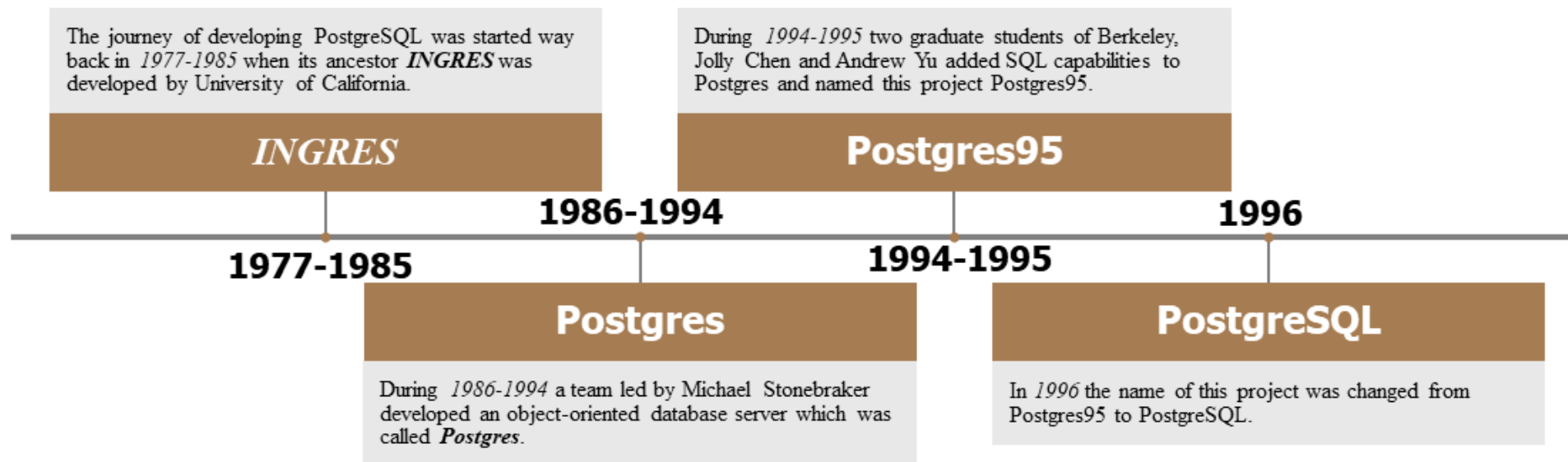
- 数据库：IDS，1962，Charles W.Bachman，10年后(1973)获得图灵奖
- 数据库概念：1963年，CODASYL，网状模型
- 关系模型：1969-1970，Edgar F. Codd，11年后(1981)获得图灵奖
- 事务处理：1969-1980，System R，James Nicholas Gray，离职18年后(1998)获得图灵奖
- Postgre：1986，Michael Stonebraker，28年后(2014)获得图灵奖
- （照此推算，下一个数据库领域的图灵奖应该在2027-2031年之间）

PostgreSQL简史

- PostgreSQL由UCB计算机科学教授Michael Stonebraker于1986年创建。在此之前，Stonebraker教授领导了关系数据库Ingres研究项目，1982年他离开UCB并将Ingres商业化使之成为Relational Technologies (RT) 公司的一个产品。而后该公司被Computer Associations (CA) 公司收购。
- Stonebraker教授在返回UCB后开始了一项post-Ingres计划，该计划致力于解决基于关系模型的数据库管理系统产品的局限性，这即是Postgres的开端。
- 1986-1993年，Postgres用户开始剧增，并且特性需求急剧增加。随后，Stonebraker再次创业，成立Illustra公司提供对Postgres的商业支持，Illustra在1997年被Informix收购。

PostgreSQL简史

- 1994年，两个UCB的研究生，Andrew Yu和Jolly Chen，增加了一个SQL语言解释器来替代早先的基于Ingres的QUEL系统，建立了Postgres95。代码随后被发行到互联网上。1996年，该计划被重新命名为PostgreSQL，以反映数据库的新查询语言SQL。



PostgreSQL

Rank			DBMS	Database Model	Score		
Oct 2024	Sep 2024	Oct 2023			Oct 2024	Sep 2024	Oct 2023
1.	1.	1.	Oracle	Relational, Multi-model	1309.45	+22.85	+48.03
2.	2.	2.	MySQL	Relational, Multi-model	1022.76	-6.73	-110.56
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	802.09	-5.67	-94.79
4.	4.	4.	PostgreSQL	Relational, Multi-model	652.16	+7.80	+13.34
5.	5.	5.	MongoDB	Document, Multi-model	405.21	-5.02	-26.21
6.	6.	6.	Redis	Key-value, Multi-model	149.63	+0.20	-13.33
7.	7.	11.	Snowflake	Relational	140.60	+6.88	+17.36
8.	8.	7.	Elasticsearch	Multi-model	131.85	+3.06	-5.30
9.	9.	8.	IBM Db2	Relational, Multi-model	122.77	-0.28	-12.10
10.	10.	9.	SQLite	Relational	101.91	-1.43	-23.23
11.	11.	12.	Apache Cassandra	Wide column, Multi-model	97.61	-1.34	-11.21
12.	12.	10.	Microsoft Access	Relational	92.15	-1.61	-32.16
13.	13.	14.	Splunk	Search engine	91.27	-1.75	-1.10
14.	14.	17.	Databricks	Multi-model	85.60	+1.35	+9.78
15.	15.	13.	MariaDB	Relational, Multi-model	84.89	+1.45	-14.77
16.	16.	15.	Microsoft Azure SQL Database	Relational, Multi-model	74.53	+1.58	-6.40
17.	17.	16.	Amazon DynamoDB	Multi-model	71.85	+1.78	-9.07
18.	18.	18.	Apache Hive	Relational	52.57	-0.50	-16.61
19.	19.	20.	Google BigQuery	Relational	51.18	-1.48	-5.39
20.	20.	21.	FileMaker	Relational	44.40	-0.80	-8.92

PostgreSQL

- PostgreSQL是一种先进的对象-关系数据库管理系统（ORDBMS），它不仅支持关系数据库的各种功能，而且还具备类、继承等对象数据库的特征。它是目前功能最强大、特性最丰富和结构最复杂的开源数据库管理系统。



PostgreSQL

诞生于伯克利 POSTGRES 软件包发展而来的。经过十几年的发展，PostgreSQL 是世界上可以获得的最先进的开放源码的数据库系统。

PostgreSQL XC

Postgres-XC 是一种高性能，提供写可靠性，多主节点数据同步，数据传输的开源集群方案，是PostgreSQL唯一可以确保事务ACID的分布式数据库存储解决方案。



PostgreSQL XL

从Postgres-XC衍生而来的一款产品，是一个完全满足ACID的、开源的、可方便进行水平扩展的、多租户安全的、基于PostgreSQL的分布式数据库解决方案，对MPP做了比较大的改进。

PostgreSQL X2

PostgreSQL的分布式数据库集群产品，同时适合OLTP 和 OLAP应用。整合了Postgres-XC 和 Postgres-XL源代码及功能，PostgreSQL X2在分布式功能和性能及稳定性上有极大的提升。

openGauss背景

课程背景

目前，我国的软件短板主要在于底层基础软件、重要支撑性软件环境和大型数据库等方面，越底层的软件的开发难度越大，重要性也越高，核心技术受制于人不仅会带来供应链断链风险，同时也会带来安全风险，所以实现高端服务器和数据库替代“IOE”（IBM主机、Oracle数据库、EMC存储设备构成的系统）不仅是大势所趋，也势必将是一件不得不完成的事。目前大部分高校的《数据库》课程中主要使用的数据库为MySQL、SQL Server 或者Oracle，都不是国内自主研发的数据库。

华为多年来致力于数据库领域的自主研发，GaussDB持续战略投入，作为华为自研数据库品牌，自建华为生态+兼容开源生态，发布了多款GaussDB产品。

本课程将华为数据库技术的内容和院校教学体系紧密融合，围绕数据库基础知识结合华为openGauss开源数据库，使学生具有使用openGauss进行数据库设计、开发、管理和维护等基本专业能力。

openGauss是一款开源关系型数据库管理系统，采用木兰宽松许可证v2发行。openGauss内核深度融合华为在数据库领域多年的经验，结合企业级场景需求，持续构建竞争力特性。

openGauss

- openGauss是一款开源关系型数据库管理系统，采用木兰宽松许可证v2发行。openGauss内核源自PostgreSQL，深度融合华为在数据库领域多年的经验，结合企业级场景需求，持续构建竞争力特性。同时openGauss也是一个开源的数据库平台，鼓励社区贡献、合作。



高性能

两路鲲鹏性能150万tpmC；面向多核架构的并发控制技术；NUMA-Aware存储引擎；SQL-Bypass智能选路执行技术；面向实时高性能场景的内存引擎。



高安全

业务无忧，故障切换时间RTO<10s；精细安全管理: 细粒度访问控制、多维度审计；全方位数据保护: 存储&传输&导出加密。



易运维

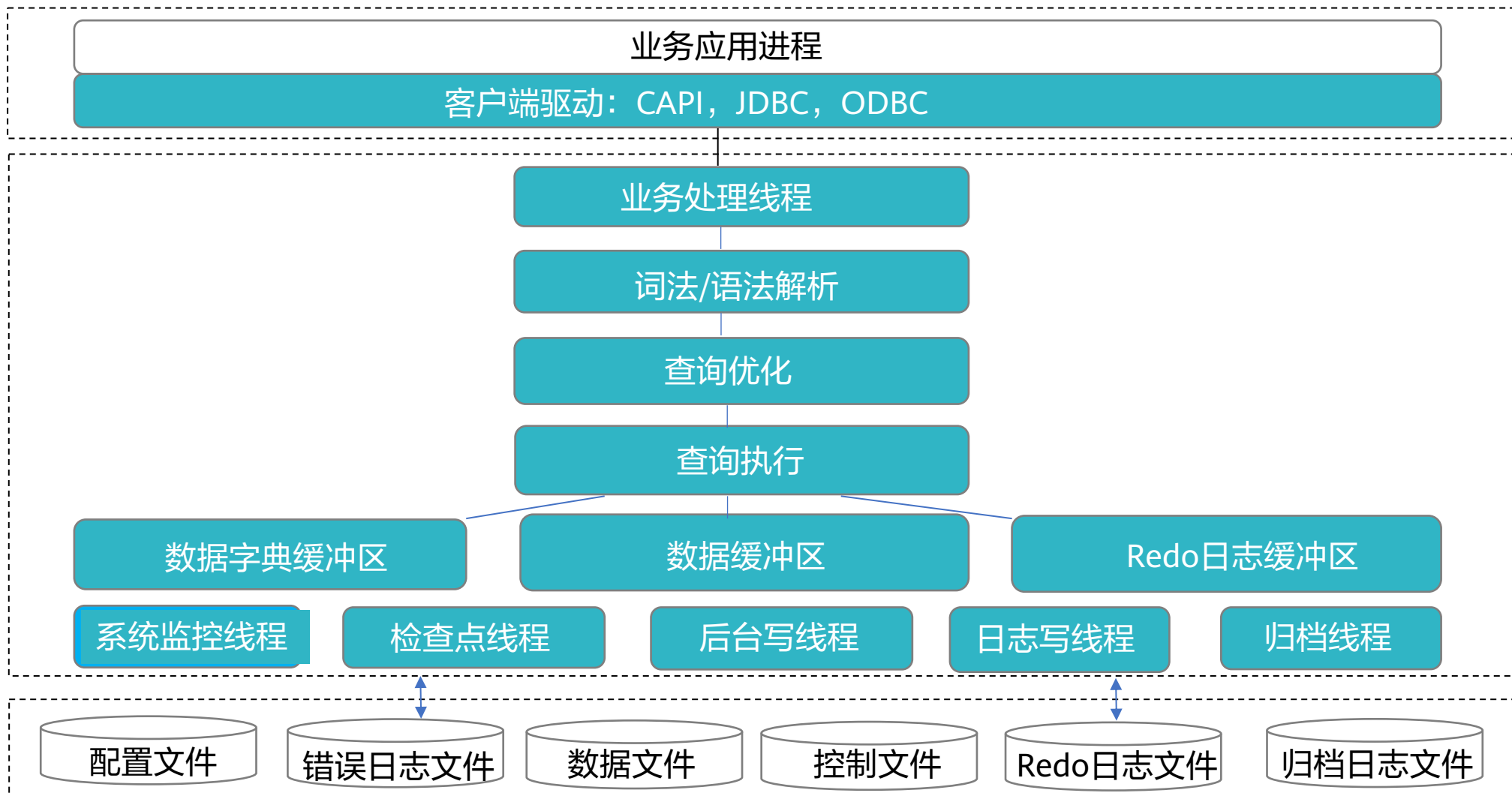
智能参数调优: 结合深度强化学习和启发式算法, 实现参数自动推荐; 慢SQL诊断, 多维性能自监控视图, 实时掌控系统性能表现; 提供在线自学习的SQL时间预测、快速定位、急速调优。



全开放

采用木兰宽松许可证协议，允许对代码自由修改、使用、引用; 完全开放数据库内核能力, 联合开发者和伙伴共同打造工具等数据库周边能力; 开放伙伴认证、培训体系及高校课程。

openGauss



openGauss架构 VS PostgreSQL架构 关键技术对比

- openGauss是衍生自PostgreSQL-XC，单机逻辑架构与PG接近。
- openGauss和PG在架构和关键技术上有根本性差异，尤其是存储引擎和优化器两大核心能力。

关键差异化因素		openGauss	PostgreSQL
运行时模型	执行模型	线程池模型，高并发连接切换代价小、内存损耗小，执行效率高，一万并发连接比最优性能损耗<5%。	进程模型，数据库进通过共享内存实现通讯和数据共享。每个进程对应一个并发连接，存在切换性能损耗，导致多核扩展性问题。
事务处理	并发控制	64位事务ID，使用CSN解决动态快照膨胀问题；NUMA-Aware引擎优化改造解决“五把大锁”。	事务ID回卷，长期运行性能因为ID回收周期大幅波动；存在“五把大锁”的问题，导致事务执行效率和多处理器多核扩展性存在瓶颈。
	日志和检查点	增量Checkpoint机制，实现性能波动<5%。	全量checkpoint，性能短期波动>15%。
	鲲鹏NUMA	NUMA改造、cache-line padding、原生spin-lock。	NUMA多核能力弱，单机两路性能TPMC <60w。
数据组织	多引擎	行存、列存、内存引擎，在研DFV存储和原位更新。	仅支持行存。
SQL引擎	优化器	支持SQL Bypass, CBO吸收工行等企业场景优化能力。	支持CBO，复杂场景优化能力一般。
	SQL解析	ANSI/ISO标准SQL92、SQL99和SQL2003和企业扩展包。	ANSI/ISO标准SQL92、SQL99和SQL2003。

运行环境

- 支持的硬件平台
 - openGauss 支持运行在鲲鹏服务器和通用的x86服务器上：
 - 支持鲲鹏服务器和基于x86_64的通用PC服务器。
 - 支持本地存储（SATA、SAS、SSD）。
 - 支持千兆、万兆Ethernet网络。
- 支持的操作系统
 - openEuler release 20.03 (LTS) on ARM。推荐采用此操作系统。
 - openEuler 20.03 on X86-64。
 - CentOS 7.6 on X86-64。

运行指标

技术指标	最大值	技术指标	最大值
数据库容量	受限于操作系统与硬件	复合索引包含列数	32
单表大小	32 TB	单表约束个数	无限制
单行数据大小	1 GB	并发连接数	10000
每条记录单个字段的大小	1 GB	分区表的分区个数	32768
单表记录数	2^{48}	分区表的单个分区大小	32 TB
单表列数	250~1600（随字段类型不同会有变化）	分区表的单个分区记录数	2^{55}
单表中的索引个数	无限制		

基本功能

- 标准SQL支持
 - 支持标准的SQL92/SQL99/SQL2003/SQL2011规范，支持GBK和UTF-8字符集，支持SQL标准函数与分析函数，支持存储过程。
- 数据库存储管理功能
 - 支持表空间，可以把不同表规划到不同的存储位置。
- 提供主备双机
 - 事务支持ACID特性、单节点故障恢复、双机数据同步，双机故障切换等。

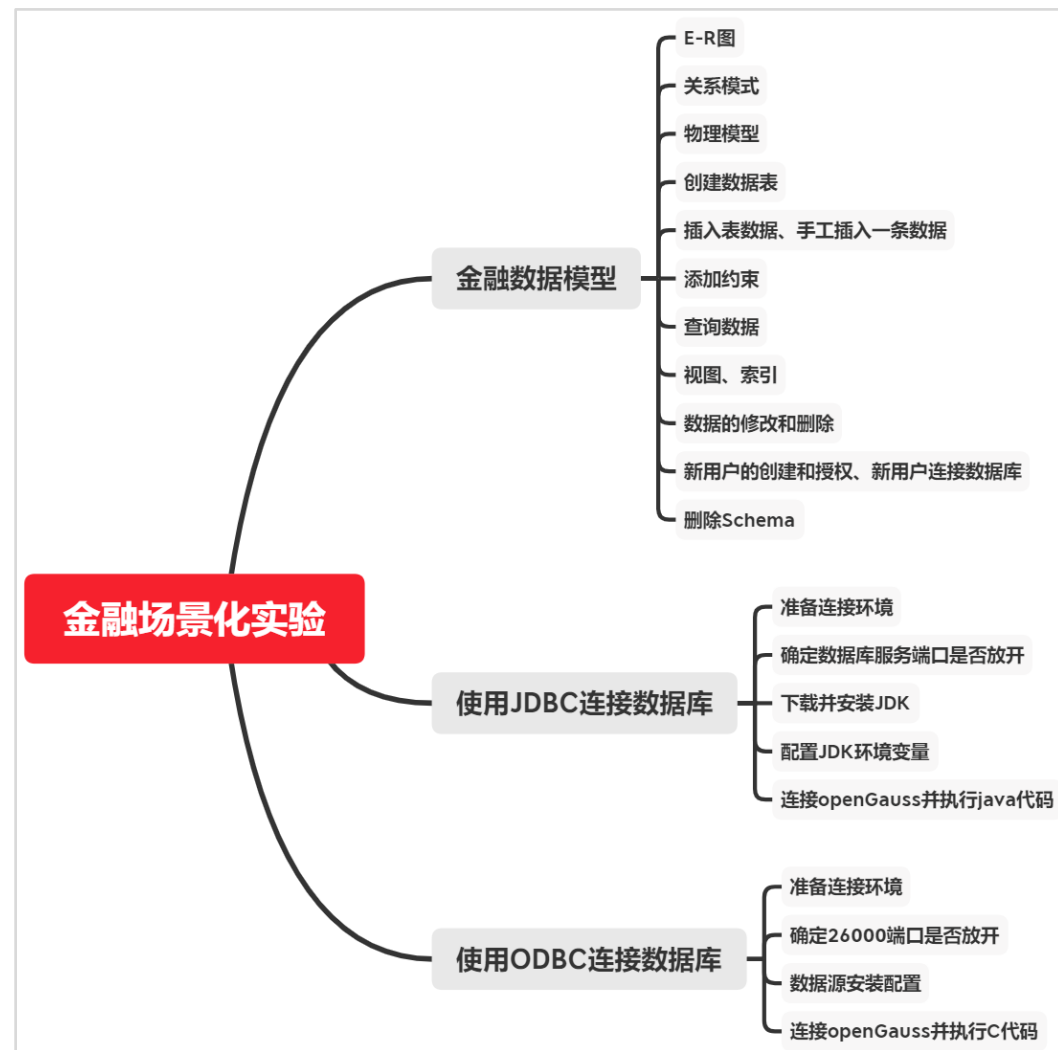
基本功能

- 应用程序接口
 - 支持标准JDBC 4.0的特性、ODBC 3.5特性。
- 管理工具
 - 提供安装部署工具、实例启停工具、备份恢复工具。
- 安全管理
 - 支持SSL安全网络连接、用户权限管理、密码管理、安全审计等功能，保证数据库在管理层、应用层、系统层和网络层的安全性。

openGauss实验

实验要求：

1. 学习openGauss数据库基本操作及OS基本操作。
2. 根据《指导手册》购买ECS弹性云服务器，安装openEuler操作系统构建openGauss数据库部署环境。
3. 根据《openGauss场景化综合应用实验》完成openGauss数据库的编译和安装，并完成金融场景化实验，体验应用过程。
4. 不会很难，学习为主。



openGauss实验

邀请链接：

<https://edu.hicomputing.huawei.com/teaching/voucher-details/invite/202210095K51Hh1l>

DDL：期末考试前一天

提交方式：<https://table.nju.edu.cn/dtable/collection-tables/fb1a226e-ee1f-421e-8dd2-38019bcb2005>

目前发现存在的问题：

1. 部分下载链接（sourceforge等）由于内部限制可能无法下载，可通过更换镜像地址或者FTP上传解决。

如果遇到其他问题可以邮件和我联系。

openGauss实验-升级版

实验要求：

1. 学习openGauss数据库基本操作及OS基本操作。
2. 购买ECS弹性云服务器，安装openEuler操作系统构建openGauss数据库部署环境。
3. 根据《openGauss数据库维护管理指导手册》完成数据库运维相关任务。
4. 根据《openGauss场景化综合应用实验》完成openGauss数据库的编译和安装，并完成金融场景化实验，体验应用过程。
5. 根据《openGauss开源数据库二次开发指导手册》为数据库添加一个新函数。

```
postgres=# select gauss_hello('openGauss User');
          gauss_hello
-----
Hello, openGauss User
(1 row)

postgres=#
```

期末考试

期末考试

不能透露更多信息

但同学们可以提出自己的建议*

例如：

SQL编程题



SQL语句填空、纠错、解释、写运行结果

*：婉拒降低难度的要求，因为现在的题目真的不难。

Q&A

谢谢各位同学的聆听