

数据库设计说明书

文档名称：福录娃学生获奖智能统计系统数据库设计说明书

版本号：V1.0

编写日期：2025 年 10 月 20 日

团队名称：月下调试人

文档状态：正式发布

目录

数据库设计说明书 1

1. 引言 5

 1.1 编写目的 5

 1.2 项目背景 5

 1.3 适用范围 5

 1.4 术语定义 5

2. 数据库环境说明 6

 2.1 数据库选型配置 6

 2.1.1 主数据库 6

 2.1.2 辅助数据库 6

 2.2 数据库连接配置 6

3. 数据库结构设计 7

 3.1 总体 E-R 图设计 7

 3.2 核心业务表设计 7

 3.2.1 用户表 (users)..... 7

 3.2.2 奖项规则表 (awards)..... 9

 3.2.3 任务表 (tasks)..... 10

 3.2.4 审核记录表 (audit_logs) 12

 3.3 统计报表相关表 13

 3.3.1 统计摘要表 (statistics_summary) 13

4. 外部存储设计 14

 4.1 Elasticsearch 索引设计 14

 4.1.1 OCR 文本索引 (ocr_text_index)..... 14

 4.2 Redis 数据结构设计 15

 4.2.1 会话管理 15

4.2.2 任务队列管理	16
5. 安全性设计	17
5.1 数据加密策略	17
5.1.1 敏感字段加密	17
5.1.2 传输加密	17
5.2 权限控制设计	17
5.2.1 数据库用户权限	17
6. 性能优化策略	18
6.1 索引优化方案	18
6.1.1 复合索引设计	18
6.1.2 覆盖索引优化	18
6.2 查询优化建议	18
6.2.1 避免全表扫描	18
6.2.2 分页查询优化	19
6.3 分区策略设计	19
6.3.1 按时间范围分区	19
6.3.2 按业务维度分区	19
7. 备份与恢复策略	20
7.1 备份策略设计	20
7.1.1 全量备份计划	20
7.2 恢复测试方案	20
7.2.1 恢复演练计划	20
8. 监控与维护	21
8.1 监控指标体系	21
8.1.1 数据库性能监控	21
8.1.2 业务指标监控	21

8.2 定期维护任务	21
8.2.1 日常维护	21
8.2.2 月度维护	22
9. 风险评估与应对	22
9.1 技术风险应对	22
9.2 业务风险应对	22
10. 版本更新记录	22

1. 引言

1.1 编写目的

本文档旨在详细描述"福录娃学生获奖智能统计系统"的数据库设计方案，为开发团队提供数据库结构、表关系、索引策略等方面的技术指导，确保数据存储的合理性、完整性和高效性。

1.2 项目背景

本系统采用 AI 驱动+人工复核的混合模式，处理学生奖状图片的识别、分类和计分。数据库设计需支持高并发上传、AI 处理流水线、多维度查询统计等复杂业务场景。

1.3 适用范围

- 数据库开发人员
- 后端开发工程师
- 系统架构师
- 测试工程师
- 运维工程师

1.4 术语定义

术语	说明
OCR	光学字符识别技术
CLIP	图像-文本预判模型
ES	Elasticsearch 搜索引擎
JWT	JSON Web Token 身份认证

2. 数据库环境说明

2.1 数据库选型配置

2.1.1 主数据库

- 数据库类型：MySQL 8.0
- 字符集：utf8mb4
- 排序规则：utf8mb4_unicode_ci
- 存储引擎：InnoDB
- 连接池：HikariCP

2.1.2 辅助数据库

数据库类型	版本	用途	配置说明
Elasticsearch	8.x	全文检索和向量搜索	3 节点集群
Redis	7.x	缓存和队列管理	哨兵模式
MinIO	最新版	对象存储	分布式存储

2.2 数据库连接配置

数据库连接配置

spring.datasource.url=jdbc:mysql://localhost:3306/fuluwa

spring.datasource.username=fuluwa_user

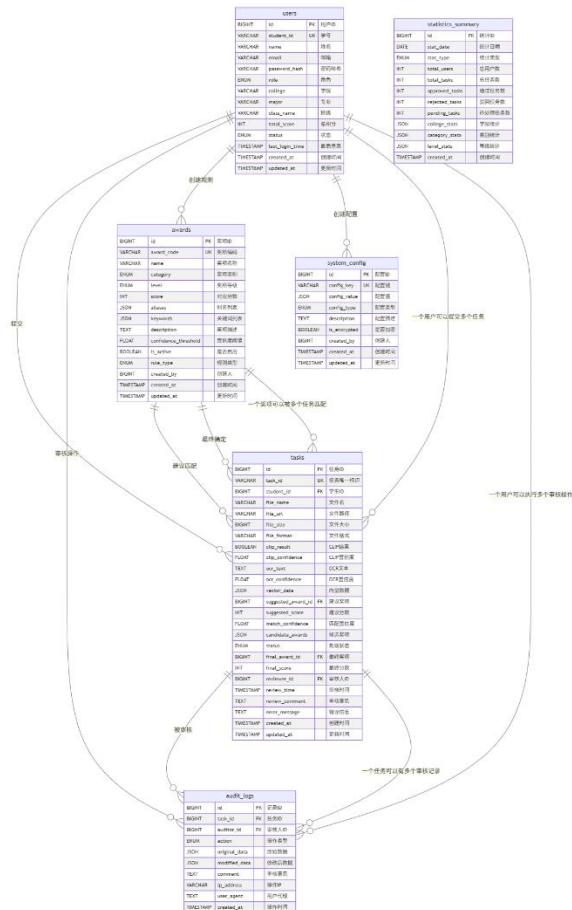
spring.datasource.password=encrypted_password

spring.datasource.hikari.maximum-pool-size=20

spring.datasource.hikari.minimum-idle=5

3. 数据库结构设计

3.1 总体 E-R 图设计



3.2 核心业务表设计

3.2.1 用户表 (users)

CREATE TABLE users (

id BIGINT PRIMARY KEY AUTO_INCREMENT COMMENT '用户 ID',

student_id VARCHAR(20) UNIQUE NOT NULL COMMENT '学号',

```

name VARCHAR(50) NOT NULL COMMENT '姓名',

email VARCHAR(100) UNIQUE COMMENT '邮箱',

password_hash VARCHAR(255) COMMENT '密码哈希',

role ENUM('STUDENT', 'ADMIN', 'AUDITOR') DEFAULT 'STUDENT' COMMENT '用户角色',

college VARCHAR(100) COMMENT '学院',

major VARCHAR(100) COMMENT '专业',

class_name VARCHAR(50) COMMENT '班级',

total_score INT DEFAULT 0 COMMENT '总积分',

status ENUM('ACTIVE', 'INACTIVE') DEFAULT 'ACTIVE' COMMENT '状态',

last_login_time TIMESTAMP NULL COMMENT '最后登录时间',

created_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP COMMENT '创建时间',

updated_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP ON UPDATE
CURRENT_TIMESTAMP COMMENT '更新时间'
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COMMENT='用户信息表';

```

字段说明：

字段名	类型	约束	说明	示例
id	BIGINT	PRIMARY KEY	自增主键	1
student_id	VARCHAR(20)	UNIQUE NOT NULL	学号	"20210001"
role	ENUM	DEFAULT 'STUDENT'	用户角色	"STUDENT"
total_score	INT	DEFAULT 0	累计积分	85

索引设计：

-- 主要索引

INDEX idx_student_id (student_id),

INDEX idx_email (email),

INDEX idx_role (role),

```
INDEX idx_college (college),  
  
INDEX idx_class (class_name),  
  
INDEX idx_status (status)
```

3.2.2 奖项规则表 (awards)

表功能说明： 存储系统认可的奖项类型及其评分规则。

```
CREATE TABLE awards (  
  
    id BIGINT PRIMARY KEY AUTO_INCREMENT COMMENT '奖项 ID',  
  
    award_code VARCHAR(50) UNIQUE NOT NULL COMMENT '奖项编码',  
  
    name VARCHAR(200) NOT NULL COMMENT '奖项名称',  
  
    category ENUM('HONOR', 'COMPETITION', 'SPORTS', 'INNOVATION') NOT NULL  
COMMENT '奖项类别',  
  
    level ENUM('NATIONAL', 'PROVINCIAL', 'UNIVERSITY', 'COLLEGE') NOT NULL  
COMMENT '奖项等级',  
  
    score INT NOT NULL COMMENT '对应分数',  
  
    aliases JSON COMMENT '别名列表',  
  
    keywords JSON COMMENT '关键词列表',  
  
    description TEXT COMMENT '奖项描述',  
  
    confidence_threshold FLOAT DEFAULT 0.8 COMMENT '置信度阈值',  
  
    is_active BOOLEAN DEFAULT TRUE COMMENT '是否启用',  
  
    rule_type ENUM('DEFAULT', 'TEMPORARY') DEFAULT 'DEFAULT' COMMENT '规则  
类型',  
  
    created_by BIGINT COMMENT '创建人',  
  
    created_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP COMMENT '创建时间',  
  
    updated_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP ON UPDATE  
CURRENT_TIMESTAMP COMMENT '更新时间'  
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COMMENT='奖项规则表';
```

字段说明：

字段名	类型	约束	说明	示例
award_code	VARCHAR(50)	UNIQUE NOT NULL	奖项 编码	"HONOR_SANHAO"
category	ENUM	NOT NULL	奖项 类别	"HONOR"
level	ENUM	NOT NULL	奖项 等级	"UNIVERSITY"
aliases	JSON	NULL	别名 列表	["三好学生", "优秀学 生"]

索引设计：

INDEX idx_category (category),

INDEX idx_level (level),

INDEX idx_is_active (is_active),

INDEX idx_rule_type (rule_type),

FULLTEXT idx_name_desc (name, description)

3.2.3 任务表 (tasks)

表功能说明： 存储奖状处理任务的全流程信息。

CREATE TABLE tasks (

id BIGINT PRIMARY KEY AUTO_INCREMENT COMMENT '任务 ID',

task_id VARCHAR(50) UNIQUE NOT NULL COMMENT '任务唯一标识',

student_id BIGINT NOT NULL COMMENT '学生 ID',

file_name VARCHAR(255) NOT NULL COMMENT '文件名',

file_url VARCHAR(500) NOT NULL COMMENT '文件存储路径',

file_size BIGINT COMMENT '文件大小(字节)',

file_format VARCHAR(10) COMMENT '文件格式',

-- AI 处理相关字段

```

clip_result BOOLEAN COMMENT 'CLIP 预判结果',
clip_confidence FLOAT COMMENT 'CLIP 置信度',
ocr_text TEXT COMMENT 'OCR 识别文本',
ocr_confidence FLOAT COMMENT 'OCR 置信度',
vector_data JSON COMMENT '向量化数据',

-- 匹配结果
suggested_award_id BIGINT COMMENT '建议奖项 ID',
suggested_score INT COMMENT '建议分数',
match_confidence FLOAT COMMENT '匹配置信度',
candidate_awards JSON COMMENT '候选奖项列表',

-- 审核状态
status ENUM(
    'PENDING',          -- 待处理
    'PROCESSING',       -- 处理中
    'CLIP_CHECKED',     -- CLIP 验证完成
    'OCR_COMPLETED',    -- OCR 识别完成
    'AI_REVIEW',        -- AI 审核中
    'MANUAL_REVIEW',    -- 人工审核中
    'APPROVED',         -- 已通过
    'REJECTED',         -- 已驳回
    'DUPLICATE',        -- 重复提交
    'FAILED'            -- 处理失败
) DEFAULT 'PENDING' COMMENT '处理状态',

```

```

    final_award_id BIGINT COMMENT '最终奖项 ID',

    final_score INT COMMENT '最终分数',

    reviewer_id BIGINT COMMENT '审核人 ID',

    review_time TIMESTAMP NULL COMMENT '审核时间',

    review_comment TEXT COMMENT '审核意见',

    error_message TEXT COMMENT '错误信息',


    created_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP COMMENT '创建时间',

    updated_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP ON UPDATE
CURRENT_TIMESTAMP COMMENT '更新时间'
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COMMENT='任务处理表';

```

索引设计：

```

INDEX idx_task_id (task_id),

INDEX idx_student_id (student_id),

INDEX idx_status (status),

INDEX idx_created_at (created_at),

INDEX idx_reviewer_id (reviewer_id),

INDEX idx_final_award_id (final_award_id)

```

3.2.4 审核记录表 (audit_logs)

表功能说明： 记录所有人工审核操作日志。

```

CREATE TABLE audit_logs (

    id BIGINT PRIMARY KEY AUTO_INCREMENT COMMENT '记录 ID',

    task_id BIGINT NOT NULL COMMENT '任务 ID',

    auditor_id BIGINT NOT NULL COMMENT '审核人 ID',

    action ENUM('APPROVE', 'REJECT', 'MODIFY') NOT NULL COMMENT '操作类型',

    original_data JSON COMMENT '原始数据',

```

```
modified_data JSON COMMENT '修改后数据',  
comment TEXT COMMENT '审核意见',  
ip_address VARCHAR(45) COMMENT '操作 IP',  
user_agent TEXT COMMENT '用户代理',  
created_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP COMMENT '操作时间'  
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COMMENT='审核记录表';
```

3.3 统计报表相关表

3.3.1 统计摘要表 (statistics_summary)

```
CREATE TABLE statistics_summary (  
    id BIGINT PRIMARY KEY AUTO_INCREMENT COMMENT '统计 ID',  
    stat_date DATE NOT NULL COMMENT '统计日期',  
    stat_type ENUM('DAILY', 'WEEKLY', 'MONTHLY') NOT NULL COMMENT '统计类型',  
  
    -- 基础统计  
    total_users INT DEFAULT 0 COMMENT '总用户数',  
    total_tasks INT DEFAULT 0 COMMENT '总任务数',  
    approved_tasks INT DEFAULT 0 COMMENT '通过任务数',  
    rejected_tasks INT DEFAULT 0 COMMENT '驳回任务数',  
    pending_tasks INT DEFAULT 0 COMMENT '待处理任务数',  
  
    -- 学院统计  
    college_stats JSON COMMENT '学院维度统计',  
    category_stats JSON COMMENT '奖项类别统计',  
    level_stats JSON COMMENT '奖项等级统计',
```

```
created_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP COMMENT '创建时间'  
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COMMENT='统计摘要表';
```

4. 外部存储设计

4.1 Elasticsearch 索引设计

4.1.1 OCR 文本索引 (ocr_text_index)

```
{  
  "settings": {  
    "number_of_shards": 3,  
    "number_of_replicas": 1,  
    "analysis": {  
      "analyzer": {  
        "ik_smart_custom": {  
          "type": "custom",  
          "tokenizer": "ik_smart",  
          "filter": ["lowercase"]  
        }  
      }  
    }  
  },  
  "mappings": {  
    "properties": {  
      "taskId": {"type": "keyword"},  
      "studentId": {"type": "long"},
```

```

    "studentName": {"type": "text", "analyzer": "ik_smart_custom"},
    "ocrText": {
      "type": "text",
      "analyzer": "ik_smart_custom",
      "fields": {
        "keyword": {"type": "keyword"}
      }
    },
    "ocrVector": {
      "type": "dense_vector",
      "dims": 768,
      "index": true,
      "similarity": "cosine"
    },
    "confidence": {"type": "float"},
    "createTime": {"type": "date"}
  }
}

```

4.2 Redis 数据结构设计

4.2.1 会话管理

用户会话存储

Key: SESSION:{sessionId}

Value: {

"userId": "123",

```
"studentId": "20210001",  
"role": "STUDENT",  
"loginTime": "2024-01-01T10:00:00Z"  
}
```

用户会话索引

Key: USER_SESSIONS:{userId}

Type: Set

Value: [sessionId1, sessionId2]

4.2.2 任务队列管理

待处理任务队列

Key: TASK_QUEUE:PENDING

Type: List

Value: [taskId1, taskId2, ...]

任务状态缓存

Key: TASK_STATUS:{taskId}

Value: {

 "status": "PROCESSING",

 "progress": 50,

 "currentStep": "OCR_EXTRACTION"

}

5. 安全性设计

5.1 数据加密策略

5.1.1 敏感字段加密

-- 敏感数据加密存储示例

```
CREATE TABLE sensitive_data (  
    id BIGINT PRIMARY KEY,  
    encrypted_field VARBINARY(512) COMMENT 'AES-256 加密字段',  
    iv VARBINARY(16) COMMENT '加密向量',  
    created_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP  
);
```

5.1.2 传输加密

- 所有数据库连接使用 SSL/TLS 加密
- API 接口强制 HTTPS 协议
- 敏感数据传输前进行加密处理

5.2 权限控制设计

5.2.1 数据库用户权限

-- 只读用户

```
CREATE USER 'fuluwa_read'@'%' IDENTIFIED BY 'password';  
GRANT SELECT ON fuluwa.* TO 'fuluwa_read'@'%';
```

-- 读写用户

```
CREATE USER 'fuluwa_write'@'%' IDENTIFIED BY 'password';  
GRANT SELECT, INSERT, UPDATE ON fuluwa.* TO 'fuluwa_write'@'%';
```

-- 管理用户

```
CREATE USER 'fuluwa_admin'@'localhost' IDENTIFIED BY 'password';
```

```
GRANT ALL PRIVILEGES ON fuluwa.* TO 'fuluwa_admin'@'localhost';
```

6. 性能优化策略

6.1 索引优化方案

6.1.1 复合索引设计

-- 任务查询优化索引

```
CREATE INDEX idx_task_status_time ON tasks(status, created_at);
```

-- 用户查询优化索引

```
CREATE INDEX idx_user_college_class ON users(college, class_name, status);
```

-- 统计查询优化索引

```
CREATE INDEX idx_stat_date_type ON statistics_summary(stat_date, stat_type);
```

6.1.2 覆盖索引优化

-- 奖项查询覆盖索引

```
CREATE INDEX idx_award_search ON awards(category, level, is_active)
```

```
INCLUDE (name, score, description);
```

6.2 查询优化建议

6.2.1 避免全表扫描

-- 不推荐

```
SELECT * FROM tasks WHERE status = 'APPROVED';
```

-- 推荐

```
SELECT id, task_id, student_id, final_score  
  
FROM tasks  
  
WHERE status = 'APPROVED'  
  
AND created_at >= '2024-01-01';
```

6.2.2 分页查询优化

-- 传统分页（数据量大时性能差）

```
SELECT * FROM tasks ORDER BY id LIMIT 10000, 20;
```

-- 优化分页（使用游标）

```
SELECT * FROM tasks WHERE id > 10000 ORDER BY id LIMIT 20;
```

6.3 分区策略设计

6.3.1 按时间范围分区

-- 任务表按年分区

```
ALTER TABLE tasks PARTITION BY RANGE (YEAR(created_at)) (  
  
    PARTITION p2024 VALUES LESS THAN (2025),  
  
    PARTITION p2025 VALUES LESS THAN (2026),  
  
    PARTITION p_future VALUES LESS THAN MAXVALUE  
);
```

6.3.2 按业务维度分区

-- 统计表按类型分区

```
ALTER TABLE statistics_summary PARTITION BY LIST COLUMNS(stat_type) (
```

```
PARTITION p_daily VALUES IN ('DAILY'),  
  
PARTITION p_weekly VALUES IN ('WEEKLY'),  
  
PARTITION p_monthly VALUES IN ('MONTHLY')  
  
);
```

7. 备份与恢复策略

7.1 备份策略设计

7.1.1 全量备份计划

备份类型	执行频率	保留周期	存储位置
全量备份	每周日 02:00	1 个月	本地磁盘+云存储
增量备份	每天 01:00	7 天	本地磁盘
二进制日志	实时	15 天	本地磁盘

7.2 恢复测试方案

7.2.1 恢复演练计划

- **月度演练：** 每月第一个周末进行恢复测试
- **季度演练：** 每季度进行灾难恢复演练
- **年度审计：** 每年进行备份完整性审计

8. 监控与维护

8.1 监控指标体系

8.1.1 数据库性能监控

监控指标	告警阈值	检查频率
连接数使用率	>80%	1 分钟
慢查询数量	>10 个/分钟	5 分钟
锁等待时间	>5 秒	实时
存储空间使用率	>85%	15 分钟

8.1.2 业务指标监控

监控指标	正常范围	异常处理
任务处理成功率	>95%	检查 AI 服务
平均处理时间	<60 秒	优化队列
并发任务数	<1000	扩容处理

8.2 定期维护任务

8.2.1 日常维护

-- 索引碎片整理

ANALYZE TABLE tasks, users, awards;

-- 表状态检查

CHECK TABLE tasks QUICK;

```
-- 统计信息更新

UPDATE statistics_summary

SET college_stats = calculate_college_stats()

WHERE stat_date = CURDATE();
```

8.2.2 月度维护

- 数据库参数调优
- 历史数据归档
- 备份策略评估
- 性能瓶颈分析

9. 风险评估与应对

9.1 技术风险应对

风险类型	影响程度	应对措施	负责人
数据丢失	高	多副本备份，异地容灾	DBA
性能瓶颈	中	读写分离，缓存优化	架构师
安全漏洞	高	定期安全扫描，漏洞修复	安全工程师

9.2 业务风险应对

风险类型	影响程度	应对措施	监控指标
数据一致性	高	事务控制，数据校验	错误日志
处理积压	中	自动扩容，流量控制	队列长度
识别准确率	中	人工复核，模型优化	准确率统计

10. 版本更新记录

版本号	更新日期	更新内容	更新人
V1.0	2025-10-20	初始版本创建	月下调试人