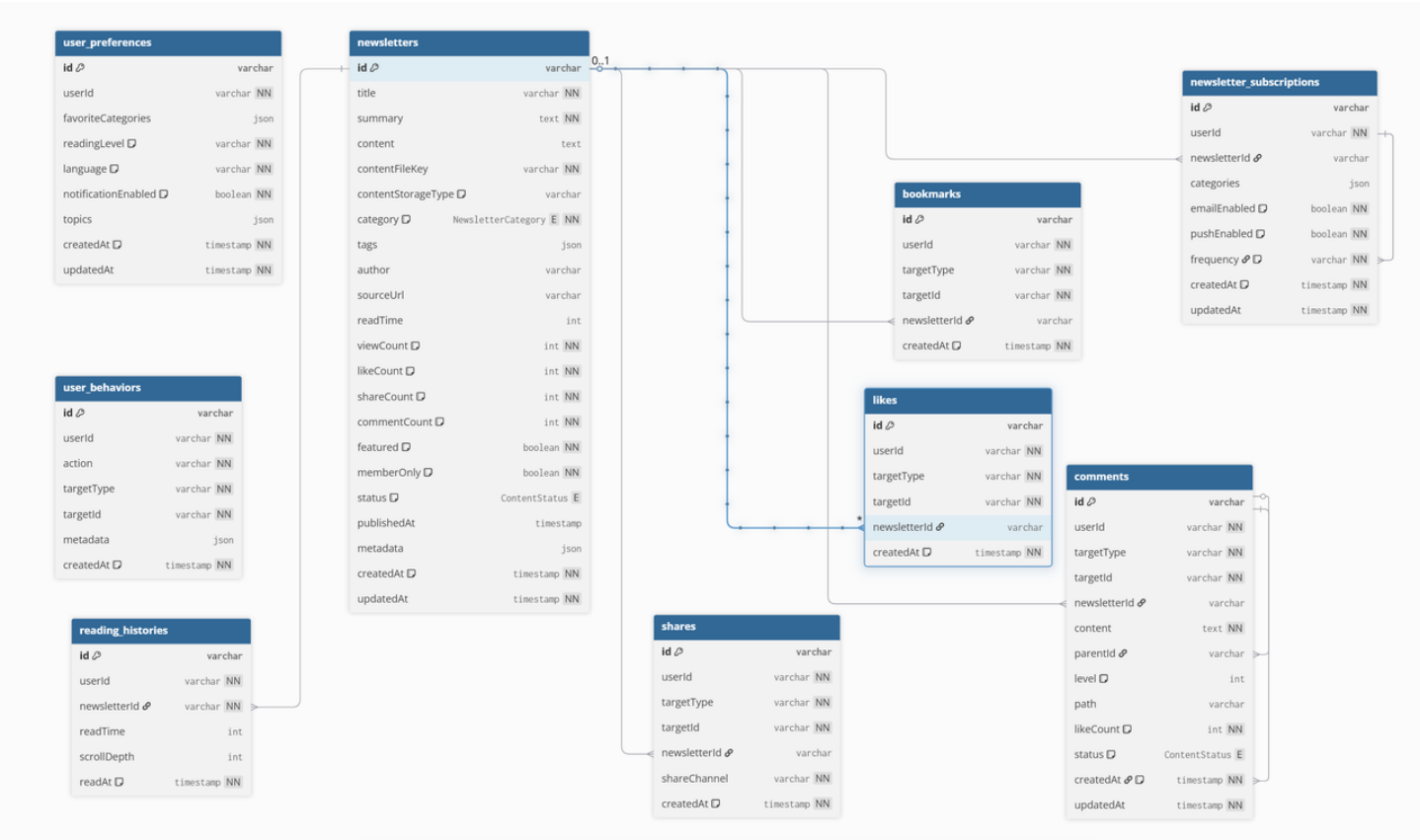
AI Newsletters数据库设计文档

AI Newsletters数据库设计文档

概述

本文档描述了ThinkInOS AI专业资讯平台的数据库模式设计，主要包含用户管理、AI专业资讯内容管理和用户行为跟踪等核心功能模块。

数据模型



平台定位

ThinkInOS是一个专注于AI领域的专业资讯平台，提供5个垂直分类的深度内容：

- **AI Agent**: 智能体技术、多智能体协作、Agent框架、应用案例
- **AI新闻周报**: 行业动态、公司新闻、投融资、政策法规
- **AI Papers**: 学术论文、研究进展、会议报告、技术突破
- **AI Coding**: 开发工具、编程助手、代码生成、技术教程
- **AI Tools**: 新工具发布、工具评测、使用指南、工具推荐

重要设计原则

- **内容存储**: 所有AI专业资讯的文章内容必须存储在MinIO对象存储中，通过Markdown文件管理
- **数据库角色**: 数据库主要负责存储元数据、用户数据和关系数据，不存储大文本内容
- **性能优化**: 通过对象存储和CDN提供高性能的内容访问
- **专业内容**: 专注于AI领域的深度、专业、高质量内容
- **全文检索**: 使用Elasticsearch提供高性能的全文搜索功能

枚举类型

NewsletterCategory (新闻分类)

- `AI_AGENT` - AI智能体技术、多智能体协作、Agent框架、应用案例
- `AI_NEWS` - AI新闻周报、行业动态、公司新闻、投融资、政策法规
- `AI_PAPERS` - AI学术论文、研究进展、会议报告、技术突破
- `AI_CODING` - AI开发工具、编程助手、代码生成、技术教程
- `AI_TOOLS` - AI新工具发布、工具评测、使用指南、工具推荐

ContentStatus (内容状态)

- `DRAFT` - 草稿
- `PUBLISHED` - 已发布
- `ARCHIVED` - 已归档
- `DELETED` - 已删除

核心表结构

1. newsletters (新闻通讯表)

用途: 存储新闻通讯内容

字段	类型	说明	约束
id	varchar	主键	PK
title	varchar	标题	必填
summary	text	摘要	必填
content	text	内容预览/缓存	可选

contentFileKey	varchar	MinIO中Markdown文件路径	必填
contentStorageType	varchar	内容存储类型(markdown/html)	默认: 'markdown'
category	NewsletterCategory	分类	必填
tags	json	标签	可选
author	varchar	作者	可选
sourceUrl	varchar	来源链接	可选
readTime	int	阅读时间(分钟)	可选
viewCount	int	浏览次数	默认: 0
likeCount	int	点赞数	默认: 0
shareCount	int	分享数	默认: 0
commentCount	int	评论数	默认: 0
featured	boolean	是否精选	默认: false
memberOnly	boolean	仅会员可见	默认: false
status	ContentStatus	状态	默认: DRAFT
publishedAt	timestamp	发布时间	可选
metadata	json	元数据	可选
createdAt	timestamp	创建时间	默认: now()
updatedAt	timestamp	更新时间	必填

业务逻辑:

- 支持5个AI垂直分类的专业资讯内容
- **文章内容必须通过MinIO存储的Markdown文件展示**
- contentFileKey字段为必填，指向MinIO中的完整Markdown文件
- content字段为可选，用于存储内容预览或缓存
- 包含阅读统计和互动数据
- 支持会员专属内容

- 支持草稿到发布的完整工作流
- 专注于AI领域的深度、专业、高质量内容

2. newsletter_subscriptions (新闻订阅表)

用途: 管理用户的新闻订阅偏好

字段	类型	说明	约束
id	varchar	主键	PK
userId	varchar	用户ID	FK → users.id
newsletterId	varchar	新闻ID	FK → newsletters.id
categories	json	订阅分类	可选
emailEnabled	boolean	邮件通知	默认: true
pushEnabled	boolean	推送通知	默认: true
frequency	varchar	推送频率	默认: realtime
createdAt	timestamp	创建时间	默认: now()
updatedAt	timestamp	更新时间	必填

业务逻辑:

- 用户可以订阅特定分类的新闻
- 支持邮件和推送两种通知方式
- 可配置推送频率

3. user_behaviors (用户行为表)

用途: 跟踪用户的行为数据

字段	类型	说明	约束
id	varchar	主键	PK
userId	varchar	用户ID	FK → users.id
action	varchar	行为类型	必填
targetType	varchar	目标类型	必填

targetId	varchar	目标ID	必填
metadata	json	行为元数据	可选
createdAt	timestamp	创建时间	默认: now()

业务逻辑:

- 记录用户的各种行为（阅读、点赞、分享等）
- 支持灵活的目标类型和ID
- 用于用户行为分析和个性化推荐

4. user_preferences (用户偏好表)

用途: 存储用户的个性化偏好设置

字段	类型	说明	约束
id	varchar	主键	PK
userId	varchar	用户ID	FK → users.id
favoriteCategories	json	收藏分类	可选
readingLevel	varchar	阅读水平	默认: intermediate
language	varchar	语言偏好	默认: zh
notificationEnabled	boolean	通知开关	默认: true
topics	json	感兴趣话题	可选
createdAt	timestamp	创建时间	默认: now()
updatedAt	timestamp	更新时间	必填

业务逻辑:

- 支持多语言和阅读水平设置
- 记录用户感兴趣的分类和话题
- 用于内容个性化推荐

5. reading_histories (阅读历史表)

用途: 记录用户的阅读历史

字段	类型	说明	约束
id	varchar	主键	PK
userId	varchar	用户ID	FK → users.id
newsletterId	varchar	新闻ID	FK → newsletters.id
readTime	int	实际阅读时间(秒)	可选
scrollDepth	int	滚动深度(%)	可选
readAt	timestamp	阅读时间	默认: now()

业务逻辑:

- 记录用户阅读每篇文章的详细信息
- 用于阅读行为分析和内容优化

6. bookmarks (书签表)

用途: 用户收藏功能

字段	类型	说明	约束
id	varchar	主键	PK
userId	varchar	用户ID	FK → users.id
targetType	varchar	目标类型	必填
targetId	varchar	目标ID	必填
newsletterId	varchar	新闻ID	FK → newsletters.id
createdAt	timestamp	收藏时间	默认: now()

业务逻辑:

- 支持收藏多种类型的内容
- 主要用于收藏新闻文章

7. likes (点赞表)

用途: 用户点赞功能

字段	类型	说明	约束
id	varchar	主键	PK
userId	varchar	用户ID	FK → users.id
targetType	varchar	目标类型	必填
targetId	varchar	目标ID	必填
newsletterId	varchar	新闻ID	FK → newsletters.id
createdAt	timestamp	点赞时间	默认: now()

业务逻辑:

- 支持对多种内容类型点赞
- 主要用于新闻文章的点赞

8. shares (分享表)

用途: 用户分享功能

字段	类型	说明	约束
id	varchar	主键	PK
userId	varchar	用户ID	FK → users.id
targetType	varchar	目标类型	必填
targetId	varchar	目标ID	必填
newsletterId	varchar	新闻ID	FK → newsletters.id
shareChannel	varchar	分享渠道	必填
createdAt	timestamp	分享时间	默认: now()

业务逻辑:

- 记录用户通过不同渠道分享内容
- 支持微信、微博、邮件等多种分享方式

9. comments (评论表)

用途: 用户评论功能

字段	类型	说明	约束
id	varchar	主键	PK
userId	varchar	用户ID	FK → users.id
targetType	varchar	目标类型	必填
targetId	varchar	目标ID	必填
newsletterId	varchar	新闻ID	FK → newsletters.id
content	text	评论内容	必填
parentId	varchar	父评论ID	FK → comments.id
level	int	评论层级	默认: 0
path	varchar	评论路径	可选
likeCount	int	点赞数	默认: 0
status	ContentStatus	状态	默认: PUBLISHED
createdAt	timestamp	创建时间	默认: now()
updatedAt	timestamp	更新时间	必填

业务逻辑:

- 支持两级评论（回复功能），最多2级
- 使用预计算层级和路径优化查询性能
- 评论可以被点赞
- 支持评论审核和状态管理

表关系图

```
users (用户表)
 ├── newsletter_subscriptions (新闻订阅)
 ├── user_behaviors (用户行为)
 ├── user_preferences (用户偏好)
 ├── reading_histories (阅读历史)
 ├── bookmarks (书签)
 ├── likes (点赞)
 ├── shares (分享)
 └── comments (评论)
```

```
10
11  newsletters (新闻通讯)
12  ├── newsletter_subscriptions (新闻订阅)
13  ├── reading_histories (阅读历史)
14  ├── bookmarks (书签)
15  ├── likes (点赞)
16  ├── shares (分享)
17  └── comments (评论)
18
19  comments (评论)
20  └── comments (子评论) [自引用]
```

索引建议

主要索引

1. `newsletters(category, status, publishedAt)` - 新闻分类和状态查询
2. `reading_histories(userId, readAt)` - 用户阅读历史查询
3. `comments(targetType, targetId, status)` - 评论查询
4. `user_behaviors(userId, action, createdAt)` - 用户行为分析

复合索引

1. `likes(userId, targetType, targetId)` - 防止重复点赞
2. `bookmarks(userId, targetType, targetId)` - 用户收藏查询
3. `shares(userId, shareChannel, createdAt)` - 分享统计查询

数据一致性约束

外键约束

- 所有用户相关表必须引用有效的用户ID
- 新闻相关表必须引用有效的新闻ID
- 评论的父评论必须存在

业务约束

1. **内容状态流转**: DRAFT → PUBLISHED → ARCHIVED
2. **评论层级**: 最多支持2级评论嵌套，超过2级自动转为私信
3. **用户行为**: 同一用户对同一目标只能点赞一次
4. **内容存储**: 所有新闻通讯内容必须存储在MinIO中，contentFileKey为必填字段

全文搜索架构

Elasticsearch索引设计

1 Elasticsearch文章索引（thinkinai_newsletters）

```
{
  "mappings": {
    "properties": {
      "id": { "type": "keyword" },
      "title": {
        "type": "text",
        "analyzer": "ik_max_word",
        "search_analyzer": "ik_smart",
        "fields": {
          "keyword": { "type": "keyword" }
        }
      },
      "summary": {
        "type": "text",
        "analyzer": "ik_max_word"
      },
      "content": {
        "type": "text",
        "analyzer": "ik_max_word"
      },
      "category": { "type": "keyword" },
      "tags": { "type": "keyword" },
      "author": { "type": "keyword" },
      "publish_date": { "type": "date" },
      "read_time": { "type": "integer" },
      "view_count": { "type": "integer" },
      "like_count": { "type": "integer" },
      "featured": { "type": "boolean" },
      "member_only": { "type": "boolean" }
    }
  }
}
```

搜索流程设计

2 搜索查询流程

3 数据同步策略

搜索性能优化

4 缓存和优化策略

搜索架构优势

1. **高性能搜索**: Elasticsearch提供毫秒级搜索响应
2. **中文分词**: 使用IK分词器支持中文全文搜索
3. **多字段搜索**: 支持标题、摘要、内容的权重搜索
4. **复杂过滤**: 支持分类、标签、时间等多维度过滤
5. **数据分离**: 搜索数据与业务数据分离，互不影响
6. **实时同步**: 文章发布时自动同步到ES索引

MinIO存储集成

内容存储策略

newsletters表与MinIO的集成:

- `contentFileKey`: **必填**，存储MinIO中Markdown文件的完整路径
  - 格式: `newsletters/{year}/{month}/{newsletterId}.md`
  - 示例: `newsletters/2024/12/newsletter_001.md`
- `contentStorageType`: 标识内容存储类型
  - `markdown`: 原始Markdown文件
  - `html`: 转换后的HTML文件
- `content`: **可选**，数据库中的预览内容或缓存内容

文件管理流程

1. **内容创建**:
   - 编辑Markdown内容
   - 上传到MinIO指定路径
   - 更新数据库中的contentFileKey
2. **内容读取**:

    - **必须从MinIO读取完整内容**
    - 数据库content字段仅作为预览或缓存
3. **内容更新**:
    - 更新MinIO中的文件
    - 同步更新数据库元数据
4. **内容删除**:
    - 删除MinIO中的文件
    - 清理数据库记录

优势

- **存储效率**: 大文件存储在对象存储中
- **版本控制**: 支持文件版本管理
- **性能优化**: 支持CDN加速
- **成本控制**: 按需付费的存储模式

扩展性考虑

水平扩展

1. 按用户ID分片用户相关表
2. 按时间分片历史数据表
3. 使用读写分离

功能扩展

1. 预留JSON字段存储扩展属性
2. 使用targetType支持新的内容类型
3. 通过metadata字段支持自定义行为数据
4. 支持多种文件格式（PDF、Word等）

安全考虑

数据保护

1. 敏感字段加密存储
2. 用户行为数据脱敏
3. 定期数据备份

访问控制

1. 基于角色的权限控制
2. 数据访问审计日志
3. API访问频率限制

AI内容管理策略

内容分类管理

5个垂直分类的内容策略:

1. **AI Agent分类**:
   - 智能体技术发展动态
   - 多智能体协作研究
   - 主流Agent框架对比
   - 实际应用案例分析
   - 技术趋势预测
2. **AI新闻周报分类**:
   - 行业重大新闻事件
   - 公司动态和产品发布
   - 投融资信息汇总
   - 政策法规解读
   - 市场趋势分析
3. **AI Papers分类**:
   - 顶级会议论文解读
   - 重要研究进展
   - 技术突破分析
   - 学术前沿动态
   - 论文实用价值评估
4. **AI Coding分类**:
   - 开发工具评测
   - 编程助手使用指南
   - 代码生成技术
   - 技术教程和最佳实践

    - 开发效率提升方案
5. **AI Tools分类**:
    - 新工具发布信息
    - 工具功能评测
    - 使用指南和教程
    - 工具对比分析
    - 工具推荐和使用建议

内容质量控制

- **专业性**: 确保内容的专业性和准确性
- **时效性**: 及时发布最新AI动态
- **深度性**: 提供深度分析和见解
- **实用性**: 注重内容的实用价值
- **原创性**: 鼓励原创内容和独特观点

用户个性化推荐

基于用户偏好和阅读历史，为不同用户推荐最适合的AI内容：

- 技术开发者 → AI Coding + AI Tools
- 研究人员 → AI Papers + AI Agent
- 行业从业者 → AI News + AI Tools
- 学生群体 → AI Coding + AI Papers

监控指标

业务指标

1. 用户活跃度（阅读、点赞、评论）
2. 内容受欢迎程度
3. 用户参与度
4. AI分类内容质量评分
5. 用户专业度提升指标

技术指标

1. 数据库查询性能

2. 存储空间使用率
3. 并发访问量
4. MinIO文件访问性能
5. Elasticsearch搜索性能
6. 搜索响应时间

文件路径规范

```
MinIO Bucket: thinkinos-content
├── newsletters/
│   ├── ai-agent/
│   │   ├── 2024/
│   │   │   ├── 12/
│   │   │   │   ├── agent-framework-comparison.md
│   │   │   │   └── multi-agent-collaboration.md
│   │   │   └── 11/
│   │   └── 2023/
│   ├── ai-news/
│   │   ├── 2024/
│   │   │   ├── 12/
│   │   │   │   ├── weekly-news-2024-12-01.md
│   │   │   │   └── industry-trends.md
│   │   │   └── 11/
│   │   └── 2023/
│   ├── ai-papers/
│   │   ├── 2024/
│   │   │   ├── 12/
│   │   │   │   ├── neurips-2024-highlights.md
│   │   │   │   └── research-breakthroughs.md
│   │   │   └── 11/
│   │   └── 2023/
│   ├── ai-coding/
│   │   ├── 2024/
│   │   │   ├── 12/
│   │   │   │   ├── coding-tools-review.md
│   │   │   │   └── programming-tutorial.md
│   │   │   └── 11/
│   │   └── 2023/
│   └── ai-tools/
│       ├── 2024/
│       │   ├── 12/
│       │   │   ├── new-tools-release.md
│       │   │   └── tool-comparison.md
│       │   └── 11/
```

```
37  │       └── 2023/
38  └── avatars/
39      └── ...
```

缓存策略

1. **内容缓存**: 热门内容缓存到Redis
2. **元数据缓存**: 新闻列表和统计信息缓存
3. **CDN加速**: 静态文件通过CDN分发
4. **预加载**: 用户浏览时预加载相关内容

参考文档: ThinkAI AI NewsLetters开发环境说明文档