

# 基于大模型的运维知识管理平台

# 背景：金融科技系统技术架构的变化与挑战

大型机、小型机、DB2、Oracle、集中式存储、集中式应用系统等

## 稳态架构

- 1、架构相对简单
- 2、运行稳定
- 3、节点数量少
- 4、技术组件种类少
- 5、运维工作量和队伍规模可控
- 6、人员的技术能力和知识体系要求固定

VS

云计算（IaaS、PaaS）、容器、分布式数据库、分布式存储、分布式微服务应用系统等

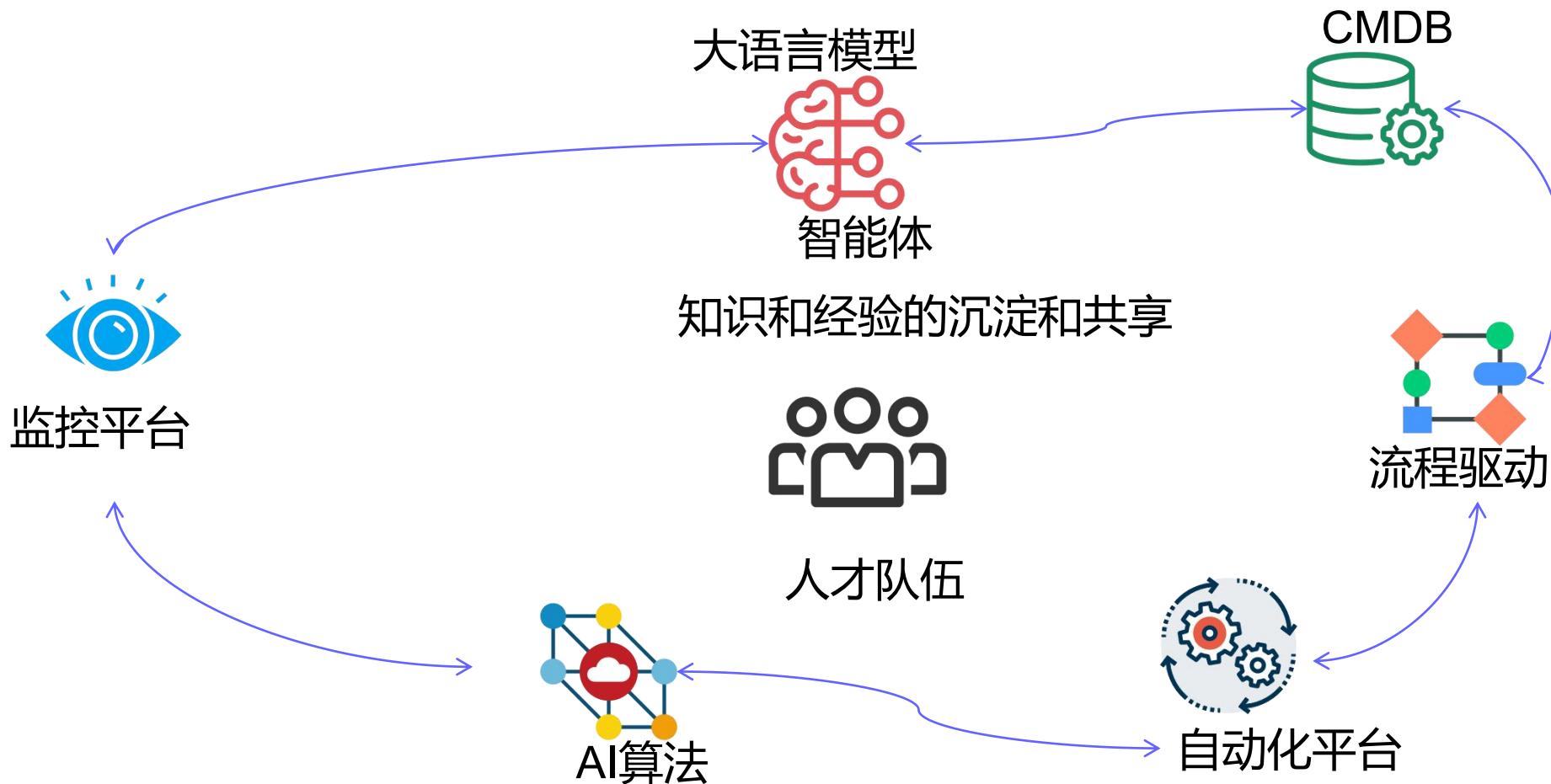
## 敏态架构

- 1、架构更加复杂
- 2、稳定性相对较差
- 3、节点数量庞大
- 4、技术组件种类繁多
- 5、运维工作量暴增
- 6、人员的技术能力和知识体系要求日新月异

# 应对策略：运维数智化转型



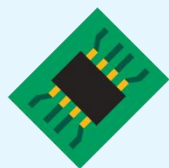
SRE理论：技术队伍的规模不应随系统的规模和复杂度的增加而增长。



# 大模型实践场景

## 实践目标

适配和运用**国产算力**，  
提升运维工作效能。



国产GPU算力池

### • 运维知识管理平台

- 设计文档模版，输入知识素材，引入反馈机制，并且和行内运维系统打通。（**监控平台、CMDB、流程平台**）

**>2000**文档

知识库中已有2000余篇运维文档和100余篇技术管理条例。

### 可持续优化

用户可随时对问题结果进行反馈，为后续模型的持续优化提供依据。

### 打通行内系统

针对用户的问题，系统可自主决定是否调取相应的API或者SQL，抓取对应数据进行回答。

### • 公文扩写

- 基于输入的提纲，自动扩写成公文，并根据用户选择的润色要求润色和生成对应的结尾、标题。

**1600**字/分钟

生成/润色公文

### 风格可选

用户可根据需要选择不同的报告类型和润色标签，从而获得符合自己要求的公文。

### • SQL语句生成

- 根据数据库连接信息和表结构分析，基于用户自然语言需求，定位并分析表格与字段，生成精准的SQL语句。

**5**秒以内

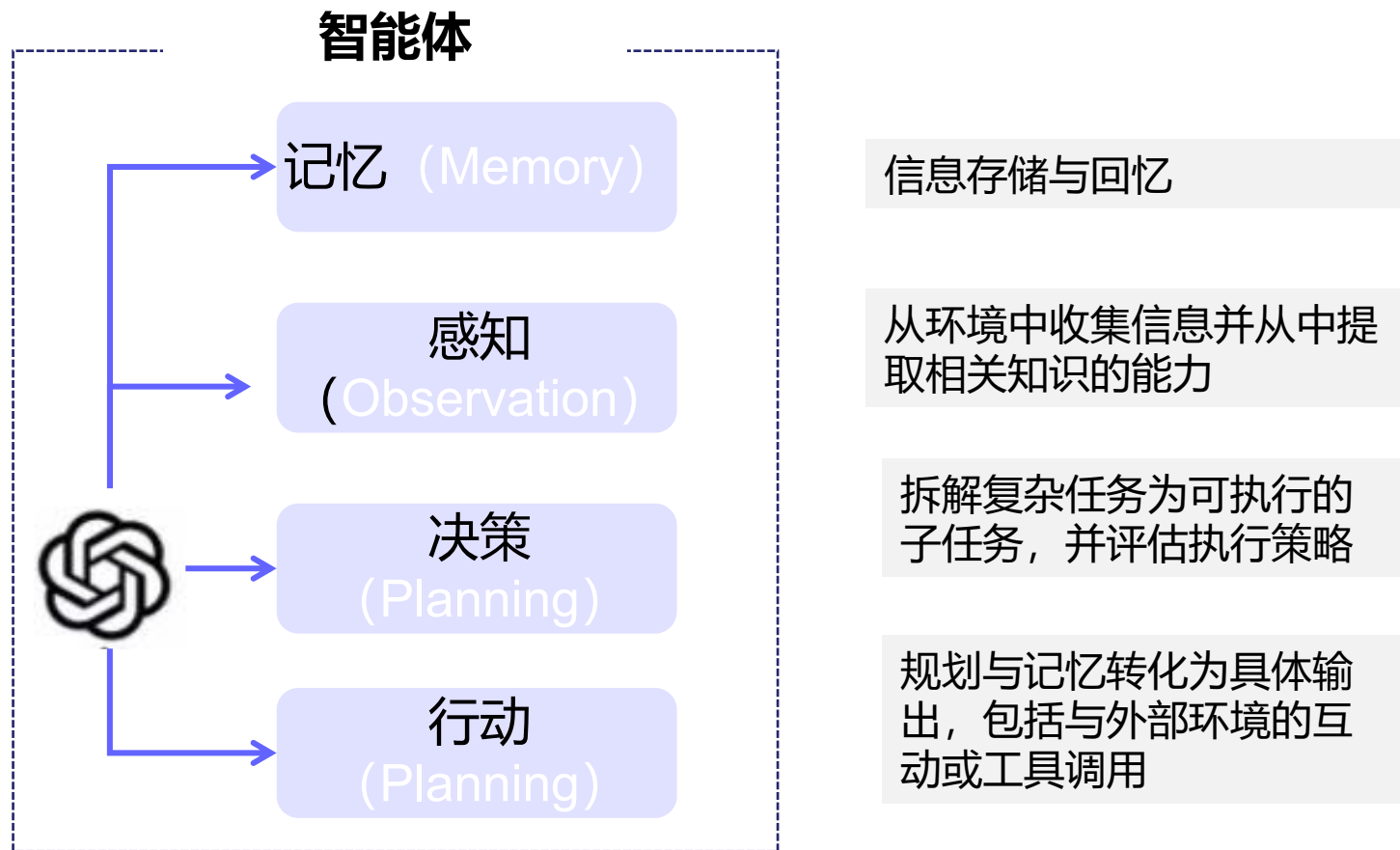
生成对应的SQL代码

**>800**张表

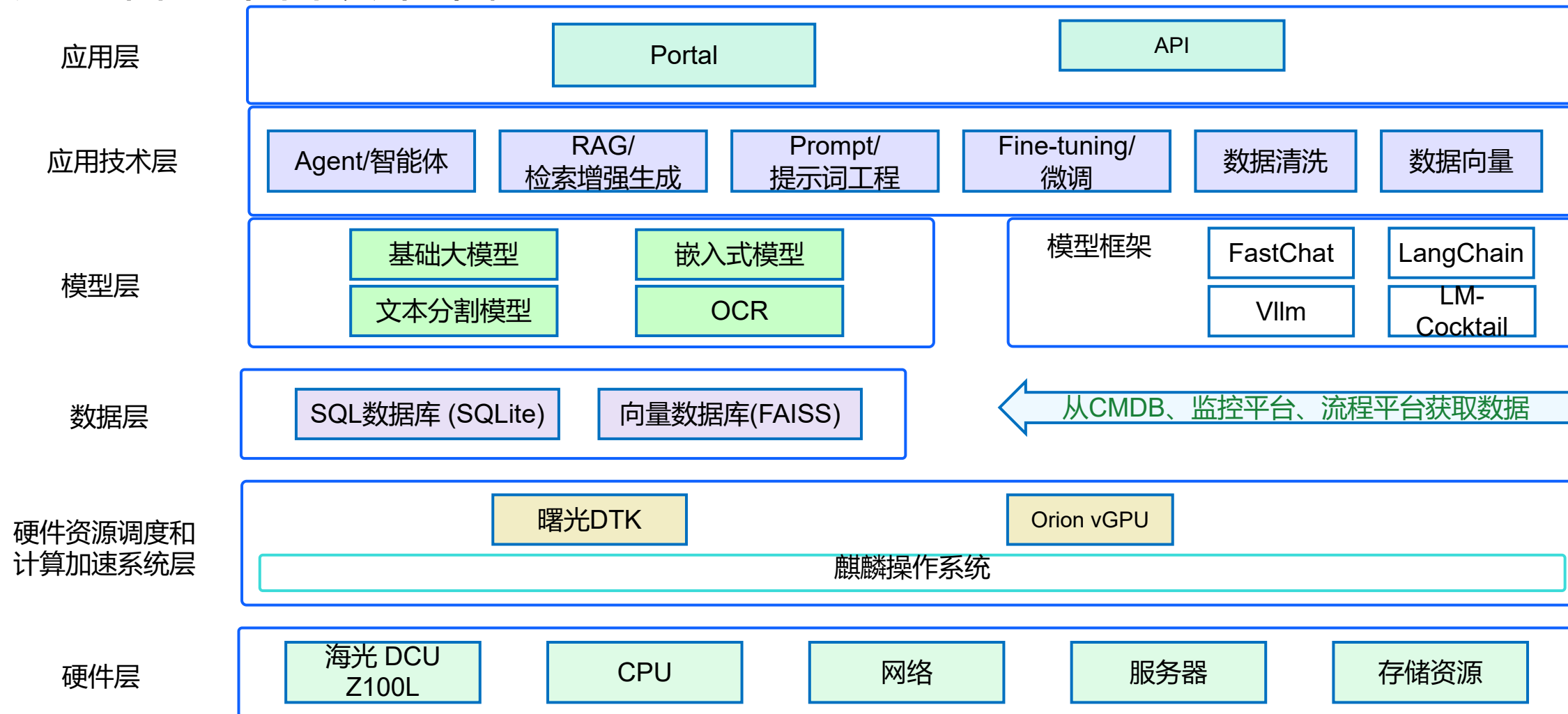
可分析超过八百张表格的数据结构

# 运维Agent智能体

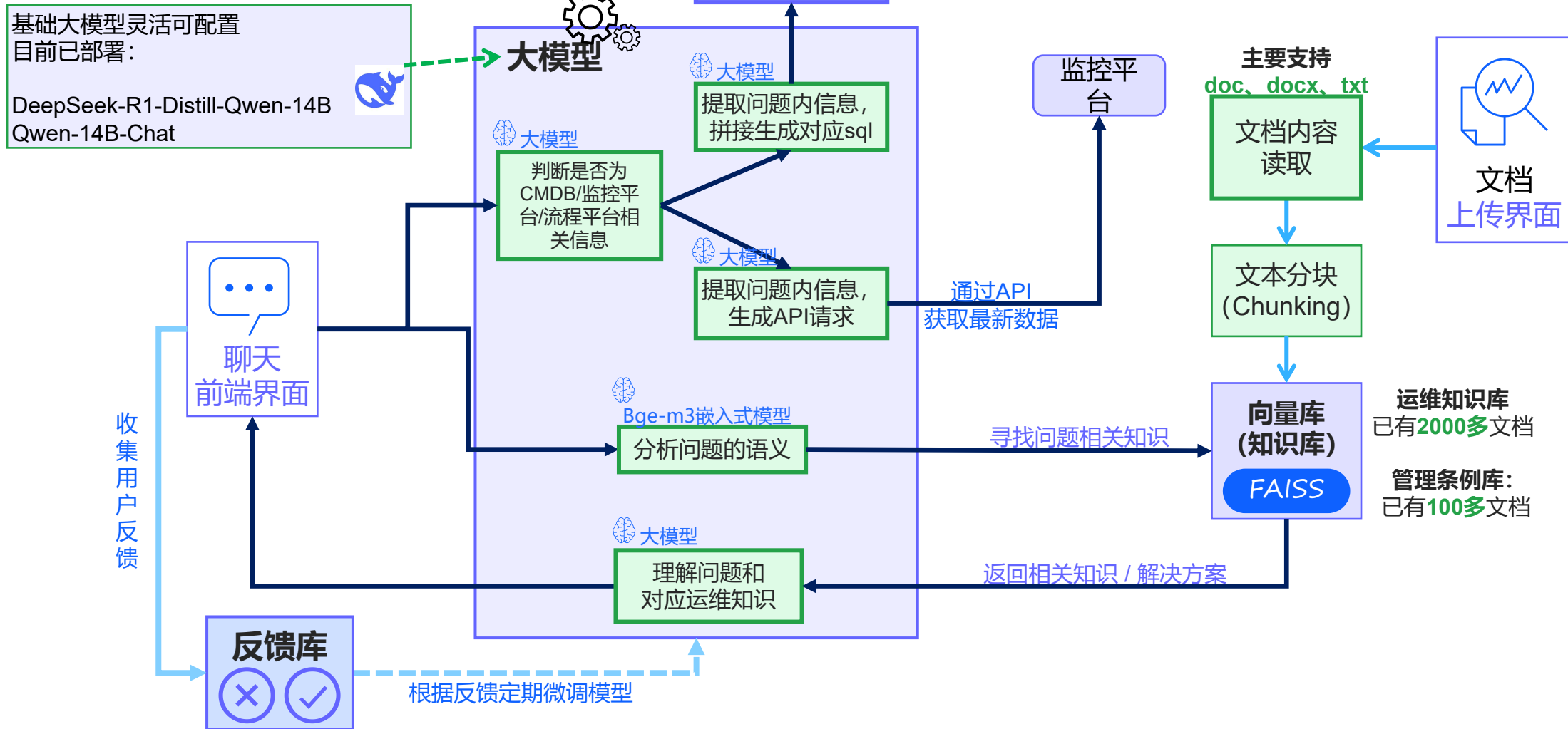
- 构建于大语言模型（LLM）之上
- 具备**环境感知能力、自主理解、决策制定及执行行动**的能力
- 能够模拟独立思考，灵活调用各类工具，逐步达成预设目标的过程



# 运维知识管理平台技术架构



# 运维知识管理平台



# 运维知识管理平台

江苏省农村信用社联合社  
JIANGSU RURAL CREDIT UNION

运维知识管理平台

对话

知识库管理

当前会话: default

请选择对话模式: 知识库问答

选择LLM模型: deepseek-r1-distill

知识库配置  
请选择知识库: JSRCCB

导出记录 清空对话

DB2数据库表空间扩容的步骤是什么?

好, 我现在要回答用户关于DB2数据库表空间扩容步骤的问题。首先, 我需要仔细阅读已知信息, 提取相关的步骤和注意事项。

已知信息中提到, 变更内容是扩表空间, 实施阶段有四个步骤。我需要将这些步骤整理出来, 并确保语言简洁专业。

第一步是确定扩充的空间大小, 并检查系统盘是否足够, 如果不够, 需要提前申请存储资源。这一步很重要, 因为存储不足可能导致后续步骤无法进行。

第二步是扩大相应的逻辑卷 (LV)。在HACMP双机环境中, 使用smitty命令进行操作, 具体路径是cl\_admin下的Shared Logical Volumes, 然后设置共享逻辑卷的特性, 增加大小。这部分需要详细说明命令和路径, 确保用户能正确操作。

第三步是扩大数据库表空间。需要切换到实例用户, 连接到数据库, 然后执行alter tablespace命令, 指定设备和扩展大小。这里要注意命令的格式, 特别是设备路径和扩展大小的单位。

第四步是查看表空间状态, 确认扩容是否成功。使用db2 list tablespaces show detail命令, 检查表空间详细信息, 确保状态正常。

变更影响范围和风险评估部分也需要简要提及, 比如确认HA状态、裸设备属组正确、避免表空间rebalance等, 这些是实施过程中需要注意的事项。

最后, 总结一下整个流程, 确保步骤清晰, 注意事项明确, 这样用户在操作时能有全面的了解, 避免出现问。

DB2数据库表空间扩容的步骤如下:

- 确定扩充的空间大小:** 检查当前系统盘的存储情况, 确保有足够的空间进行扩容。如果系统盘不足, 需提前申请存储资源。
- 扩大逻辑卷 (LV):** 在HACMP双机环境中, 使用smitty命令进行操作:
  - 打开smitty, 选择路径: HACMP Logical Volume Management -> Shared Logical Volumes -> Set Characteristics of a Shared Logical Volume -> Increase the Size of a Shared Logical Volume。
- 扩大数据库表空间:**
  - 切换到DB2实例用户: `su - 实例名`

请输入对话内容, 换行请使用Shift+Enter。输入/help查看自定义命令

- 执行表空间扩容命令:** `db2 "alter tablespace dms_idx extend (device '/dev/rslv_name' 1G)"`, 其中 1G 为扩展大小, 可根据需要调整。
- 查看表空间状态:** 执行命令 `db2 list tablespaces show detail`, 检查表空间详细信息, 确保扩容成功。

# 运维知识管理平台

江苏省农村信用社联合社  
JIANGSU RURAL CREDIT UNION

## 运维知识管理平台

对话

知识库管理

当前会话: default

请选择对话模式: 知识库问答

选择LLM模型: deepseek-r1-distill

知识库配置  
请选择知识库: JSRCCB

导出记录 清空对话

空间rebalance等，这些是实施过程中需要注意的事项。

最后，总结一下整个流程，确保步骤清晰，注意事项明确，这样用户在操作时能有全面的了解，避免出现问题。

DB2数据库表空间扩容的步骤如下：

- 确定扩充的空间大小：**检查当前系统盘的存储情况，确保有足够的空间进行扩容。如果系统盘不足，需提前申请存储资源。
- 扩大逻辑卷（LV）：**在HACMP双机环境中，使用smitty命令进行操作：
  - 打开smitty，选择路径：`HACMP Logical Volume Management -> Shared Logical Volumes -> Set Characteristics of a Shared Logical Volume -> Increase the Size of a Shared Logical Volume`。
- 扩大数据库表空间：**
  - 切换到DB2实例用户：`su - 实例名`
  - 连接到数据库：`db2 connect to 数据库名`
  - 执行表空间扩容命令：`db2 "alter tablespace dms_idx extend (device '/dev/r$LV_NAME' 1G)"`，其中1G为扩展大小，可根据需要调整。
- 查看表空间状态：**执行命令`db2 list tablespaces show detail`，检查表空间详细信息，确认扩容成功且状态正常。

**注意事项：**

- 确保在HA双机状态下正常下进行扩容。
- 确认裸设备的属组和权限已正确修改。
- 避免表空间rebalance，确保表空间与对应裸设备一致。

✓ 知识库匹配结果

出处 [1] DB2变更.txt

标题：DB2变更 变更内容： 扩空间 备份阶段： nan 实施阶段： 1.确定要扩充的空间大小，如果所在系统盘不够，需要提前申请存储资源 2.扩大相应的LV通过HACMP双机环境中扩大LV,smitty &#x2013; admin --&#x2013;HACMP Logical Volume Management--&#x2013;Shared Logical Volumes--&#x2013;Set Characteristics of a Shared Logical Volume--&#x2013;Increase the Size of a Shared Logical Volume 3.扩大数据库表空间 su - 实例名 db2 connect to 数据库名 db2 "alter tablespace dms\_idx extend (device '/dev/r\$LV\_NAME' 1G)" 4.查看表空间状态是否正常 db2 list tablespaces show detail 变更影响范围： 影响所扩充的表空间 变更风险评估： 确认HA状态下扩表空间 确认裸设备的属组正确修改 避免表空间rebalance 表空间与对应裸设备要一致

请输入对话内容，换行请使用Shift+Enter。输入/help查看自定义命令

# 知识库管理

The screenshot displays the '运维知识管理平台' (Operation Knowledge Management Platform) interface. The left sidebar contains the logo of '江苏省农村信用社联合社' (Jiangsu Rural Credit Union), the title '运维知识管理平台', a search bar with '知识库管理' selected, and a progress indicator for '匹配条数' (Match Count) showing 3 out of 100.

The main content area is titled '请选择或新建知识库:' (Please select or create a new knowledge base:). A dropdown menu shows 'JSRCCB (faiss @ bge-m3)'. Below this, there is a section for '上传知识文件:' (Upload knowledge files:), which includes a drag-and-drop area with supported file formats (HTML, MHTML, MD, JSON, CSV, PDF, DOCX, DOC, PPT, PPTX, PNG, JPG, JPEG, BMP, EML, MSG, RST, RTF, TXT, XML, EPUB, ODT, TSV, EML, MSG, EPUB, XLSX, XLS, XLSB, IPYNB, ODT, PY, RST, RTF, SR, TOML, TSV, DOCX, DOC, XML, PPT, PPTX, ENEX, HTM) and a '选择文件' (Select file) button.

Below the upload section is a text input field for '请输入知识库介绍:' (Please enter the knowledge base introduction:), with the text '关于JSRCCB的知识库' (Knowledge base about JSRCCB). A '添加文件到知识库' (Add files to knowledge base) button is located below the input field.

The bottom section, titled '知识库 JSRCCB 中已有文件:' (Files already in the JSRCCB knowledge base:), contains a table with the following data:

序号	文档名称	文档加载器
1	银联跨行业务应急预案.docx	RapidOCRDocLoader
2	大机DB2变更实施规范_d210620.txt	UnstructuredFileLoader
3	DB2数据库配置、日志、状态管理.txt	UnstructuredFileLoader
4	江苏省农村信用社联合社小型机操作系统安全配置规范.txt	UnstructuredFileLoader
5	DB2数据库使用及命名规范.txt	UnstructuredFileLoader
6	江苏省联社DB2数据库应急手册.txt	UnstructuredFileLoader
7	江苏省农村信用社联合社大型机系统应急预案.txt	UnstructuredFileLoader
8	DB2备份管理.txt	UnstructuredFileLoader
9	DB2变更.txt	UnstructuredFileLoader

# 核心大模型升级—DeepSeek赋能

及时迭代核心大模型和部署框架，进行适配和优化，改善系统效能，提供更加优质、高效的知识库服务

升级前

## Qwen-14B-Chat

- 部分专业相关问题回答不够精确，存在对知识库信息遗漏或错误解读的情况
- 大模型的推理过程为黑盒，使用者难以知晓依据，不具备可解释性

## FastChat部署框架

- 处理并发请求时，速度稍慢
- 内存管理采用传统方式，内存消耗增长快，容易发生内存不足现象

VS

升级后



DeepSeek-R1-Distill-Qwen-14B

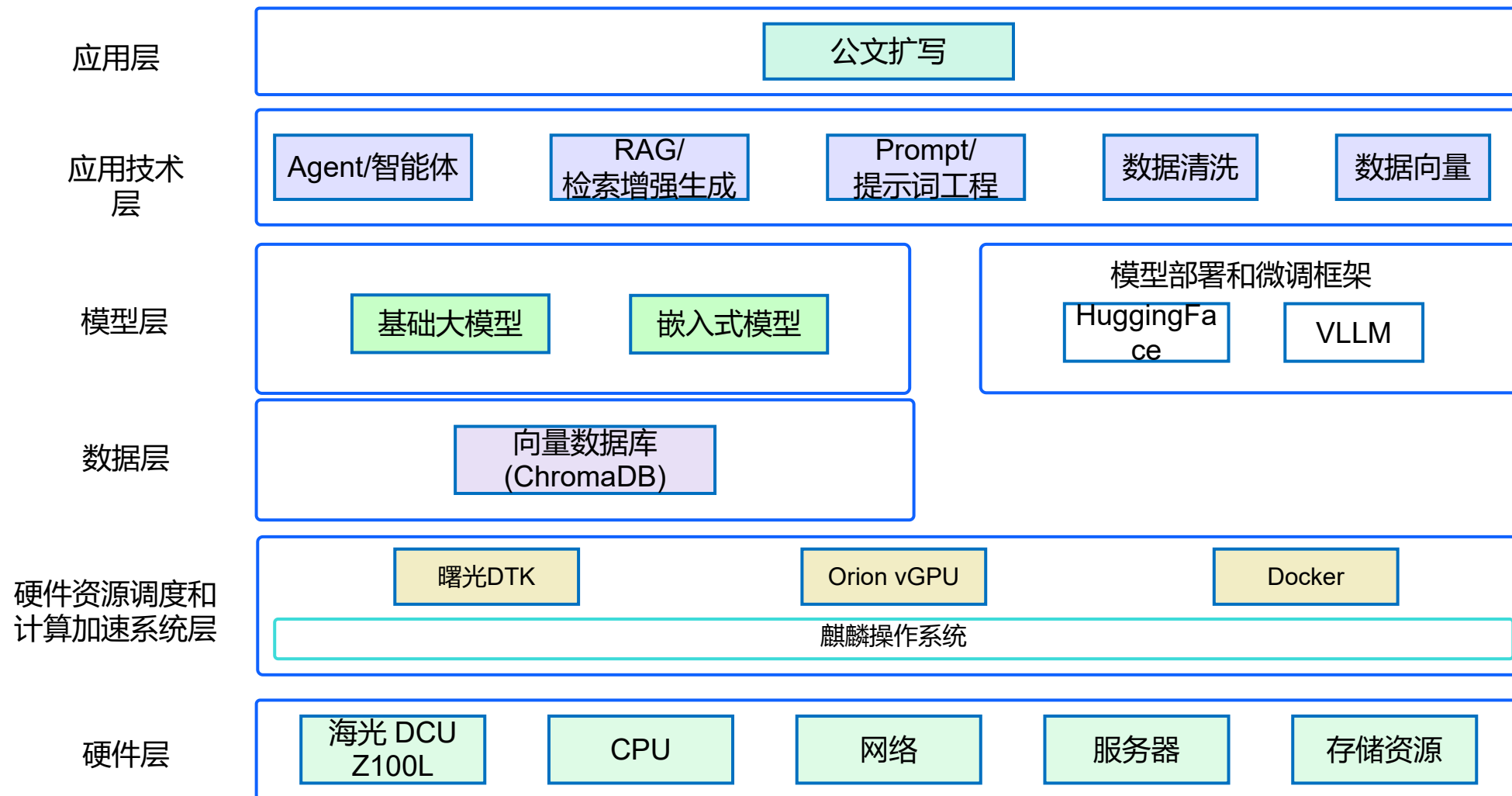
## DeepSeek

- **准确性更高**：凭借其强大的语言理解能力，能够更准确理解问题含义，并提供更精准的答案
- **透明可解释性**：展示推理过程，使用者在获得答案的同时，理解知识运用过程，增强对回答的信任

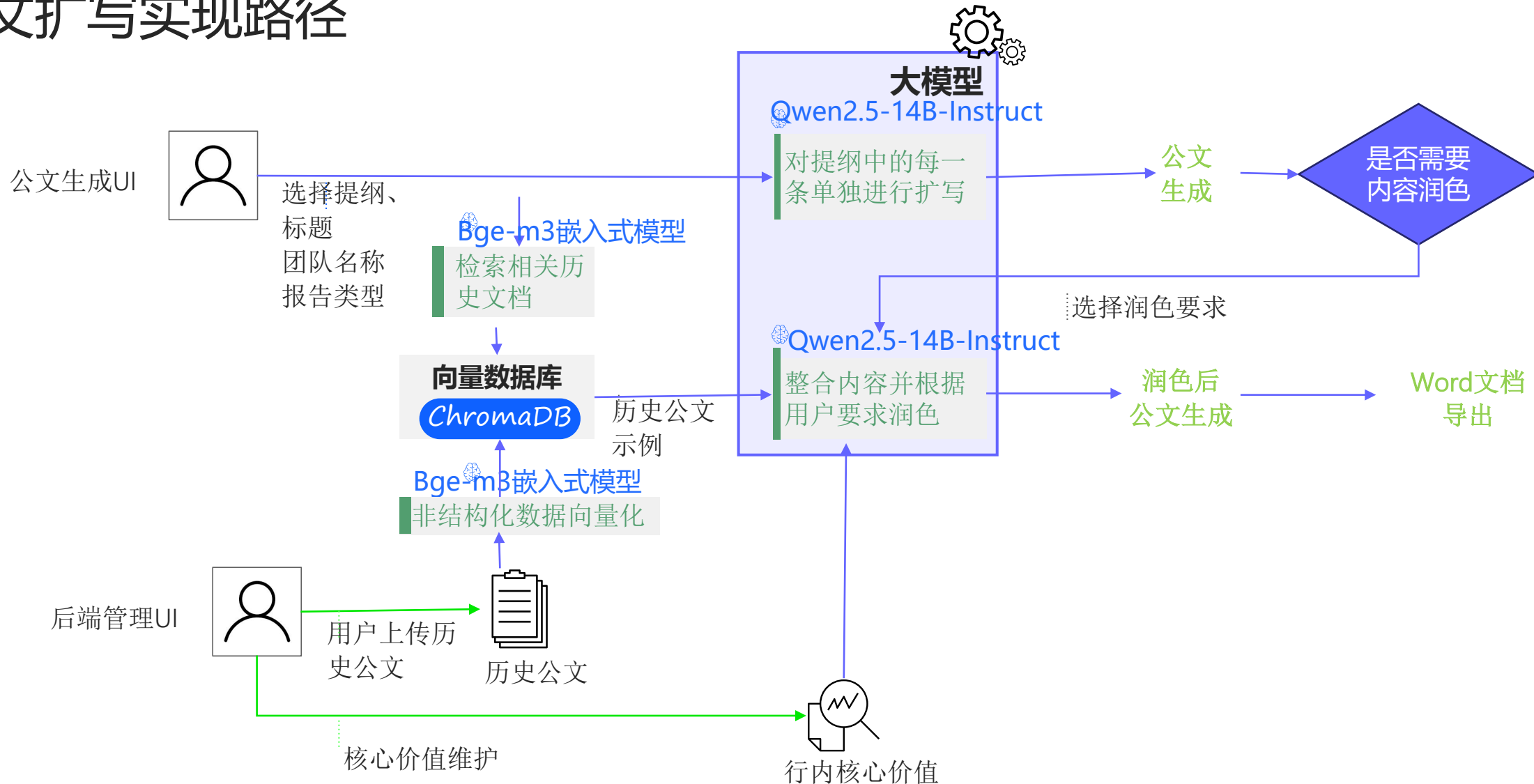
## VLLM部署框架

- **推理速度更快**：推理过程中无需频繁内存重分配，推理速度快，尤其在处理并发和长文本请求时。
- **有效管理内存**：基于PagedAttention技术，内存利用效率提高

# 公文扩写功能技术架构



# 公文扩写实现路径



# 公文扩写前端界面

## 公文写作助手

### 报告类型

年报

### 团队名称

系统维护团队

### 文章主题

2025年系统维护团队工作思路

### 工作要点（不同的工作内容以换行区分）

2025年计划完成5台支付AFA应用、5台卡AFA应用、省级和县级中间业务AFE集群等重要系统的操作系统、小型机硬件、存储升级遵循2025年制定的将基于VPLEX和SVC存储网关的原存储池整合、逐步迁移到基于国产高端存储搭建的3DC存储池  
2025年，系统维护团队计划在数据库监控方面继续深入分析和研究，采用zabbix等软件对相关软件进行监控

根据历史公文进行改写

生成文档

## 生成结果

原存储池整合，并逐步迁移至由国产高端存储设备构建的3DC（三数据中心）存储池。此计划旨在提升存储系统的整体性能、可靠性和安全性，同时推动国产化替代进程，增强自主可控能力。

本项目将按照既定的总体思路和具体目标，分阶段推进。首先，我们将进行全面的现状调研和技术评估，明确现有存储架构的优势与不足，以及新架构的需求点和可行性。随后，根据调研结果，制定详细的实施方案，包括技术选型、方案设计、测试验证等环节，确保迁移过程的安全可控。在实施过程中，将注重协调各部门资源，保障项目顺利开展。预计到2025年底，完成所有关键业务系统的迁移工作，并通过严格的测试和优化，达到预期的技术指标和服务水平。

### 三、深入研究，完善监控体系，确保信息系统安全稳定

为进一步提升我单位信息系统运行的安全性和稳定性，系统维护团队拟在2025年度围绕数据库监控系统的深化研究与应用展开全面工作。根据以往的经验积累和技术发展趋势，我们将充分利用Zabbix等先进的监控工具，实现对数据库运行状态的实时监控，并结合数据分析技术，不断优化和完善监控体系，确保能够及时发现并处理潜在的技术故障，保障业务连续性和服务质量。

本年度的工作将分为三个阶段进行。第一阶段，我们将通过详细的市场调研和技术评估，确定最适合我单位需求的数据库监控解决方案，包括但不限于Zabbix等成熟工具的选择与配置。第二阶段，是实施阶段，重点在于监控系统的部署与调试，确保其能够在实际环境中稳定运行。第三阶段，则是持续改进与优化阶段，通过定期的数据分析与反馈机制，不断调整和完善监控策略，提高系统的智能化水平和响应速度。

### 润色标签

结尾总结要点与收获  正式庄重型  数据支撑  段落标题对仗

润色文章

回退

导出Word

# 历史公文和核心价值管理

## 历史公文管理系统

上传、管理和检索您的公文档案

### 上传文件

选择文件 (最多20个文件)

Choose Files No file chosen

支持的格式: .txt, .doc, .docx

报告类型

月报

团队名称

系统维护团队

自定义标签

输入自定义标签

上传公文

### 删除文档

报告类型

月报

团队名称

系统维护团队

自定义标签 (可选)

留空则按报告类型和团队名称删除

删除提示:

此操作将永久删除所选条件匹配的所有文档, 请确认后继续。

删除匹配文档

## 核心价值指引

### 核心价值

在习近平新时代中国特色社会主义思想的指引下, 系统运维部门紧紧围绕党的十八大以来的政策方针, 积极推进银行数字化转型。在过去的一年中, 作为银行业务稳定运行的坚强后盾, 系统运维团队秉承科技引领、创新驱动的发展理念, 为保障金融科技应用的高效、安全和持续发展作出了重要贡献。

在党中央、国务院的正确领导下, 系统运维部门始终坚持以习近平新时代中国特色社会主义思想为指导, 全面贯彻落实党的十八大精神, 围绕银行数字化转型战略, 推动技术创新和维护工作稳

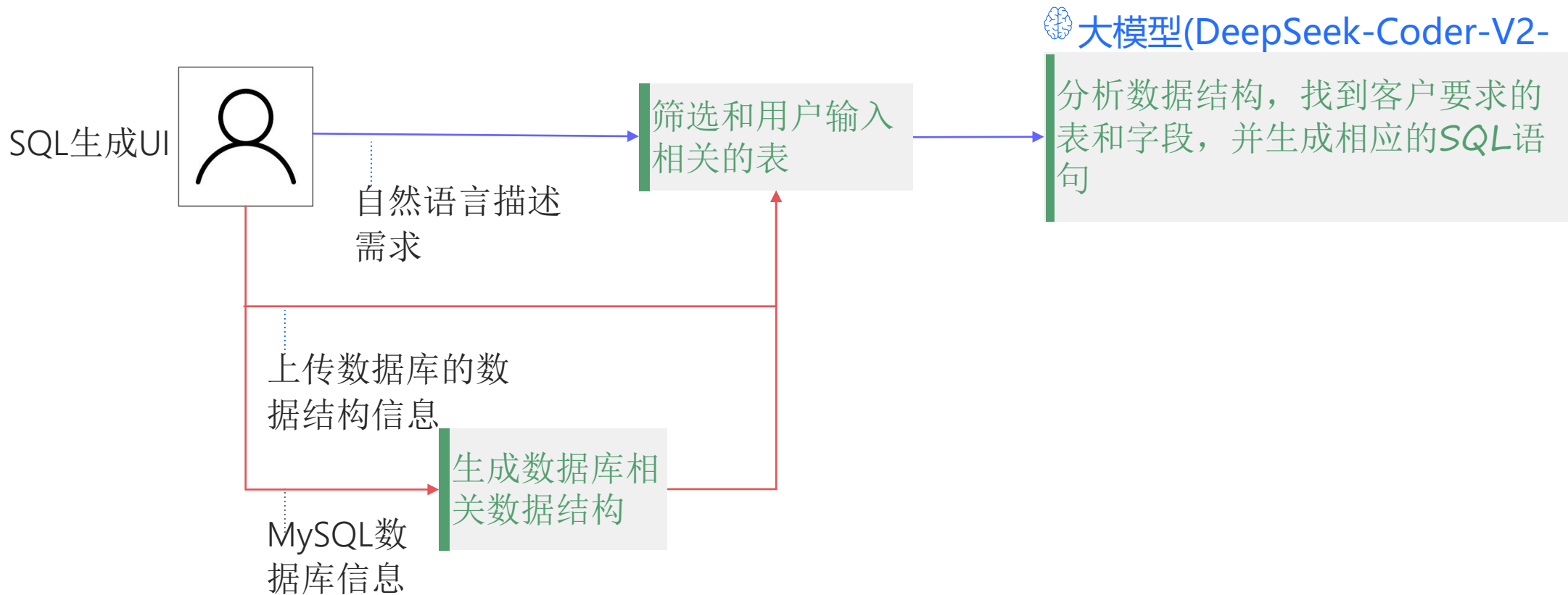
更新核心价值(每条核心价值用换行隔开, 上传文件限制2000个中文字符以内)

Choose File No file chosen

支持的格式: .txt

下载

# SQL生成功能实现路径



# SQL生成运行界面

## 1. MySQL 数据库连接

Host:

Port:

Username:

Password:

Database Name:

生成数据库Schema

## 2. 上传 Schema 文件

拖拽文件到这里或点击选择上传文件

prompt\_2.txt

Size: 62.9 KB | Modified: 12/11/2024, 2:14:41 PM

Delete

## 3. 自然语言查询

有哪些学生

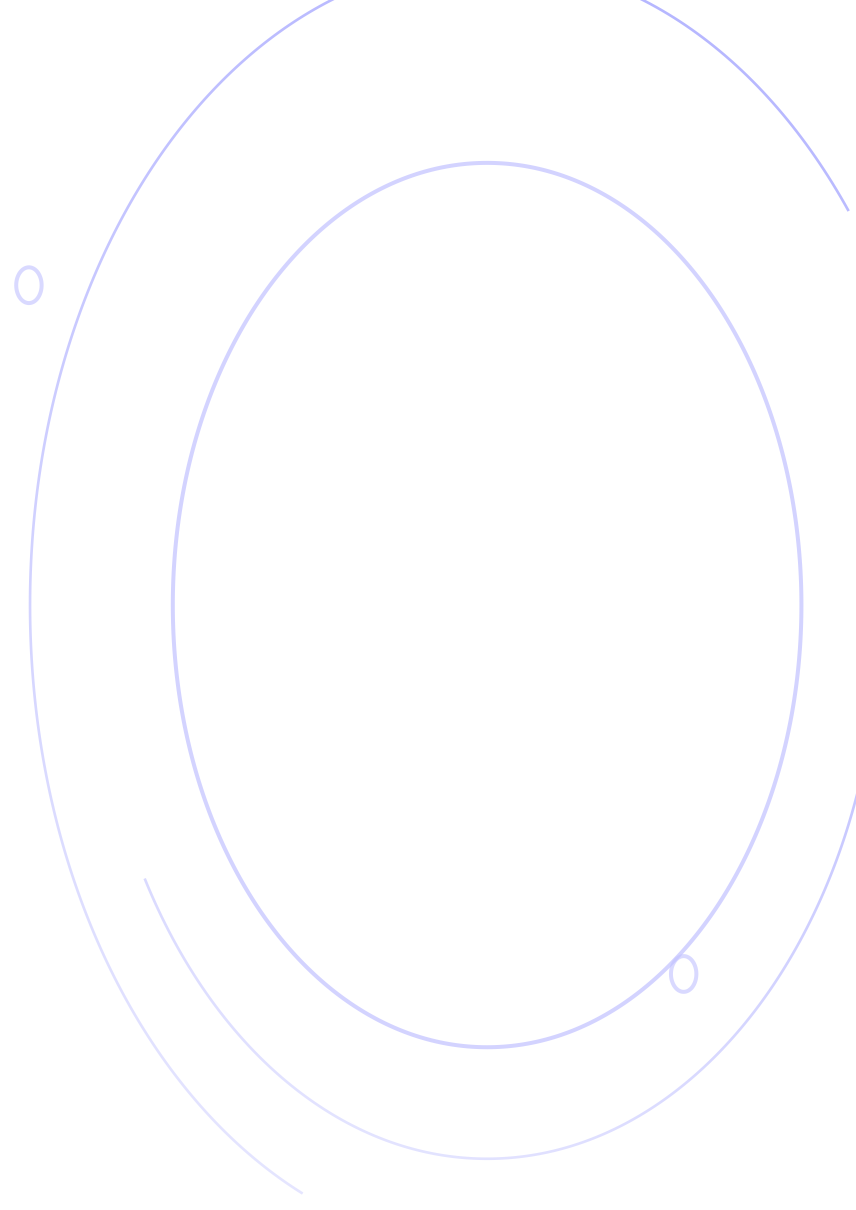
生成 SQL

## 4. 生成的 SQL 查询

```
```sql
SELECT s.first_name, s.last_name
FROM students s
JOIN enrollments e ON s.student_id = e.student_id
JOIN courses c ON e.course_id = c.course_id
WHERE c.course_name = 'Data Structures';
```
```

生成时间: 5秒左右

**感谢聆听!**  
**THANK YOU**



# 大模型驱动的业务智能化提升


## 2025 区域性银行数智化转型研讨会

4月25日 — 成都

主办单位:  金科创新社  
Fintech Innovation in China

合作单位:  **联想**

 **DCITS**  
神州信息

 **腾讯会议**

 **容联云**  
CLOOPEN

 **数势科技**  
DIGITFORCE