



高睿嘉讯

智能数据库交互分析系统产品介绍V1.0

智能数据库语音/自然语言查询分析，所思即所得

目录

CONTENTS

01 产品简介

02 产品UI及功能特性介绍

03 AI编排中台介绍

04 复杂场景落地介绍

05 整体技术架构

06 最优实践

01 产品简介



1. 产品简介



本产品旨在打破人与数据之间的技术壁垒，构建“**对话即分析**”的新一代交互范式：

语言解构：利用大模型（LLM）深度理解人类的口语、常用语、垂直业务领域术语，将其精准转化为结构化SQL查询指令，直接操作、查询数据库。

门槛消除：让不懂代码的业务、运营、销售、管理人员能够像聊天一样直接探索、分析业务数据，日常工作不再求人，工作效能直接拉满。

分析即时化：实现“所思即所得”，将长达数天的取数周期缩短至秒级，让数据流转速度跟上业务思考速度，数据价值即时兑现，决策、营销快人一步。

普惠数据价值：利用大模型、向量数据库等构建私有知识库，行业领域知识库chat式问答，人人都是业务专家。

02 产品UI及功能特性介绍

2、自然语言转SQL系统-前台web界面-1

登录 智能数据库交互分析平台

电子邮箱

zhaokaien@gmail.com

密码

.....



登录

没有账号? [注册](#)

2、自然语言转SQL系统-前台web界面-2

The screenshot displays the second version of the Natural Language to SQL System's front-end web interface. The interface is divided into several sections:

- 左侧栏 (Left Sidebar):** Contains a navigation menu with options: 新对话 (New Conversation), 搜索 (Search), 笔记 (Notes), and 工作空间 (Workplace). A red box highlights this area, with a red arrow pointing to it labeled "工作区" (Work Area).
- 对话历史记录区 (Conversation History Area):** A red box highlights this area, with a red arrow pointing to it labeled "对话历史记录区" (Conversation History Area). It lists recent conversations and usage statistics for different months.
- 工作区 (Work Area):** The main workspace where interactions with the AI model occur. It includes:
 - 可上传文件 (File Upload):** Buttons for "截图" (Screenshot) and "上传文件" (Upload File).
 - 可联网搜索 (Internet Search):** Buttons for "联网搜索" (Network Search) and "代码解释器" (Code Interpreter).
 - 对话提示区 (Conversation Prompt Area):** Displays user input prompts such as "Explain options trading", "Help me study", and "Show me a code snippet".
- 语音输入区 (Voice Input Area):** A red box highlights this area, with a red arrow pointing to it labeled "语音输入区" (Voice Input Area). It contains a microphone icon.
- 对话高级设置 (Conversation Advanced Settings):** A red box highlights this area, with a red arrow pointing to it labeled "高级参数设置区" (Advanced Parameters Setting Area). It shows a list of configuration parameters with their current values.

参数 (Parameter)	值 (Value)	说明 (Description)
流式对话响应 (Stream Chat Response)	默认	
流式增量输出分块大小	默认	
函数调用 (Function Calling)	默认	
种子 (Seed)	默认	
停止序列 (Stop Sequence)	默认	
温度 (Temperature)	默认	
推理努力 (Reasoning Effort)	默认	
logit_bias	默认	
max_tokens	默认	
top_k	默认	
top_p	默认	
min_p	默认	
frequency_penalty	默认	
presence_penalty	默认	
mirostat	默认	
mirostat_eta	默认	
mirostat_tau	默认	

2、自然语言转SQL系统-前台web界面-3

新对话 工作空间 搜索 对话 今天 Order Information Query ...

用户信息查询 七月订单数量统计 上月用户数查询 七月订单统计

过去 7 天 订单查询需求 请查询订单表 新对话

三月 你好 全方位帮助

z zhaokaien

gpt-5 v +

询问您要查询的用户名为 full_name 字段吗？还是 email、phone 或其他字段？请确认您要匹配的字段名称。

找下名字叫xuexue的用户信息

支持自然语言转 sql 查询数据库

my-custom-fastgpt

SELECT * FROM users WHERE full_name = 'xuexue'; 以下是您提供的JSON项目列表以表格形式展示：

ID	FULL_NAME	EMAIL	PHONE	BALANCE	POINTS	STATUS	IS_ACTIVE	IS_SUP
3d39cb9d-1eed-42cc-8605-73ecf6ec1905	xuexue	null	+8613811354369	99999999.95	1000000010	active	true	false

如需隐藏某些字段或调整表格内容，请告知！

Mujo Mujo Mujo

gpt-5

我不太确定您想要查询什么。“Mujo”是要当作关键词来查哪张表的哪一列呢？请告知表名和字段，或你的意图。

如果要按关键词“Mujo”模糊查询，示例：

- 查用户名包含“Mujo”的用户：SELECT * FROM users WHERE full_name ILIKE '%Mujo%';
- 查邮箱包含“Mujo”的用户：SELECT * FROM users WHERE email ILIKE '%Mujo%';

输入消息

联网搜索 代码解释器

正在倾听...

支持语音直接查询数据库

2、自然语言转SQL系统-前台web界面-4

左侧工作空间功能模块：

- 工作空间
- 搜索
- 对话：
 - 今天
 - Order Information Query
- 用户信息查询
- 七月订单数量统计
- 上月用户数查询
- 七月订单统计
- 过去 7 天
- 订单查询需求
- 请查询订单表
- 新对话
- 三月
- 你好
- 全方位帮助

右侧模型管理功能模块：

- 通用
- 外部连接
- 模型：
 - 竞技场评估
 - 文档
 - 联网搜索
 - 代码执行
 - 界面
 - 语音
 - 图像
 - Pipeline
 - 数据库

模型列表：

模型名	描述	操作
ark-deepseek-r1-250120	ark-deepseek-r1-250120	开关
ark-deepseek-r1-250528	ark-deepseek-r1-250528	开关
ark-deepseek-v3-241226	ark-deepseek-v3-241226	开关
ark-deepseek-v3-250324	ark-deepseek-v3-250324	开关
ark-doubao-seed-1.6-250615	ark-doubao-seed-1.6-250615	开关
ark-doubao-seed-1.6-flash-250615	ark-doubao-seed-1.6-flash-250615	开关
ark-doubao-seed-1.6-flash-250715	ark-doubao-seed-1.6-flash-250715	开关
ark-doubao-seed-1.6-thinking-250615	ark-doubao-seed-1.6-thinking-250615	开关
ark-kimi-k2-250711	ark-kimi-k2-250711	开关
azure-tts-1	azure-tts-1	开关
babbage-002		开关

功能说明：

- 支持一键切换市面上所有常见大模型（右侧模型列表上方）
- 支持专用工作空间、对话历史分类、对话历史查询等功能（左侧工作空间下方）

底部状态栏：

- zhaokaien

2、自然语言转SQL系统-前台web界面-5

The screenshot shows the configuration page of a Natural Language to SQL system. On the left is a sidebar with navigation links like '新对话' (New Conversation), '工作空间' (Workplace), '搜索' (Search), '对话' (Conversation), '今天' (Today), 'Order Information Query', '用户信息查询' (User Information Query), '七月订单数量统计' (July Order Quantity Statistics), '上月用户数查询' (Last Month User Count Query), '七月订单统计' (July Order Statistics), '过去 7 天' (Past 7 Days), '订单查询需求' (Order Query Requirements), '请查询订单表' (Please Query Order Table), '新对话' (New Conversation), '三月' (March), '你好' (Hello), and '全方位帮助' (All-round Help). At the bottom left is a user icon with 'zhaokaien' and a '保存' (Save) button at the bottom right.

设置 (Settings)

- 通用 (General):
 - 语音转文本设置 (Text-to-Speech Settings): Whisper (本地) (Local)
- 外部连接 (External Connections):
 - 语音转文本引擎 (Text-to-Speech Engine): whisper
- 模型 (Models):
 - 语音转文本模型 (Text-to-Speech Model): base
- 竞技场评估 (Arena Evaluation):
 - Open WebUI 使用内置 faster-whisper。点击此处了解更多关于faster-whisper的信息，并查看可用的模型。
- 联网搜索 (Network Search):
 - 文本转语音设置 (Text-to-Speech Settings): 网页 API (Web API)
- 代码执行 (Code Execution):
 - 文本转语音引擎 (Text-to-Speech Engine): whisper
- 界面 (Interface):
 - 文本转语音音色 (Text-to-Speech Voice Color): 默认 (Default)
- 语音 (Voice):
 - 拆分回复 (Split Response): Punctuation
- 图像 (Image):
- Pipeline:
- 数据库 (Database):

支持语音转文字引擎一键切换 (Supports one-click switching of speech-to-text engines)

支持文本转语音 (Supports text-to-speech)

2、自然语言转SQL系统-前台web界面-6

新对话

工作空间

搜索

对话

今天

Order Information Query

用户信息查询

七月订单数量统计

上月用户数查询

七月订单统计

过去 7 天

订单查询需求

请查询订单表

新对话

三月

你好

全方位帮助

gpt-5

2025
14:16:22
GMT

2b9d-4b03-
9b6c-
974d8be8d05d

Jul 2025
14:16:22
GMT

2025
14:16:22 GMT

eed-
8605-
73ecf

如需导出为Markdown表格或其他格式, 请告知!

如需导出为Markdown表格或Excel表格, 请告知。

支持联网搜索

支持代码解释

输入消息

+ 联网搜索 代码解释器

z zhaokaien

请查询订单信息表

CREATED_AT	ID	ORDER_TOTAL	ORDER_TOTAL_POINTS	PAID_AT	PAYMENT_METHOD	STATUS	UPDATED_AT	USER
Wed, 02 Jul 2025 14:27:36 GMT	8f087889-37ee-4fb1-8229-b97d46a31853	0.01	100	null	wxpay	pending	Wed, 02 Jul 2025 15:39:08 GMT	52f79-0672-bf0b-d04d
Fri, 11 Jul 2025 18:27:21 GMT	cd2ca5a0-b859-472b-b318-636573c03048	0.04	400	Fri, 11 Jul 2025 18:27:21 GMT	balance	paid	Fri, 11 Jul 2025 18:27:21 GMT	3d39-1eed-8605-73ecf
Mon, 14 Jul 2025 14:16:22 GMT	90abc41-2b9d-4b03-9b6c-974d8be8d05d	0.01	100	Mon, 14 Jul 2025 14:16:22 GMT	balance	paid	Mon, 14 Jul 2025 14:16:22 GMT	3d39-1eed-8605-73ecf

2、自然语言转SQL系统-前台web界面-7

The screenshot shows the configuration interface for the Natural Language to SQL System. It includes sections for Content Extraction Engine, Embedding, and Retrieval, each with specific settings and features highlighted by red arrows and annotations.

Content Extraction Engine

- PDF 图像处理 (使用 OCR)** (highlighted with a red box): 支持 PDF、图像 OCR 识别
- Bypass Embedding and Retrieval
- 文本分切器
- 块大小 (Chunk Size): 1000
- 块重叠 (Chunk Overlap): 100

Embedding

- 语义向量模型引擎: 默认 (SentenceTransformers)
- 语义向量模型: sentence-transformers/all-MiniLM-L6-v2
- 警告: 如果您修改了语义向量模型, 则需要重新导入所有文档。

Retrieval

- 混合搜索
- Top K

支持混合搜索 (highlighted with a red arrow pointing to the Top K section)

支持文本切分、向量引擎的一键切换 (highlighted with a red arrow pointing to the Embedding section)

支持 PDF、图像 OCR 识别 (highlighted with a red arrow pointing to the PDF 图像处理 section)

Page number: 3

Bottom panel:

```
### Task:  
Respond to the user query using the provided context, incorporating inline citations in the format [source_id] **only when the <source_id> tag is explicitly provided** in the context.  
  
### Guidelines:  
- If you don't know the answer, clearly state that.  
- If you can't find the answer, clearly state that.
```

zhaokaien

保存

2、自然语言转SQL系统-前台web界面-日志

The screenshot shows a web-based application interface. On the left, there's a sidebar with various icons and sections like 'Open WebUI', '用户' (User), '新对话' (New Conversation), '搜索' (Search), '笔记' (Notes), '工作空间' (Workspace), and a '对话' (Conversation) section for 'July Welding Material Leader' and '李东进三月份领了多少焊材'. The main area displays a log of interactions:

- 九月: 新对话, 李东进三月份领了多少焊材
- 八月: 新对话, 新对话, 新对话, 补充需求信息, 芯片不含 PSEN 引脚, Bootloader 读写确认, 新对话, MSC1211Y5 烧录时序修正, Welding Material Usage, Welding Material Usage

编辑用户组

通用 权限 用户 (0)

对话权限
允许上传文件
允许使用对话高级设置
允许设置 Valves 变量
允许使用对话系统提示词
允许设置模型高级参数
允许删除对话记录
允许编辑对话记录
允许分享对话
允许导出对话
允许语音转文本
允许文本转语音

删除 保存

编辑用户组

通用 权限 用户 (0)

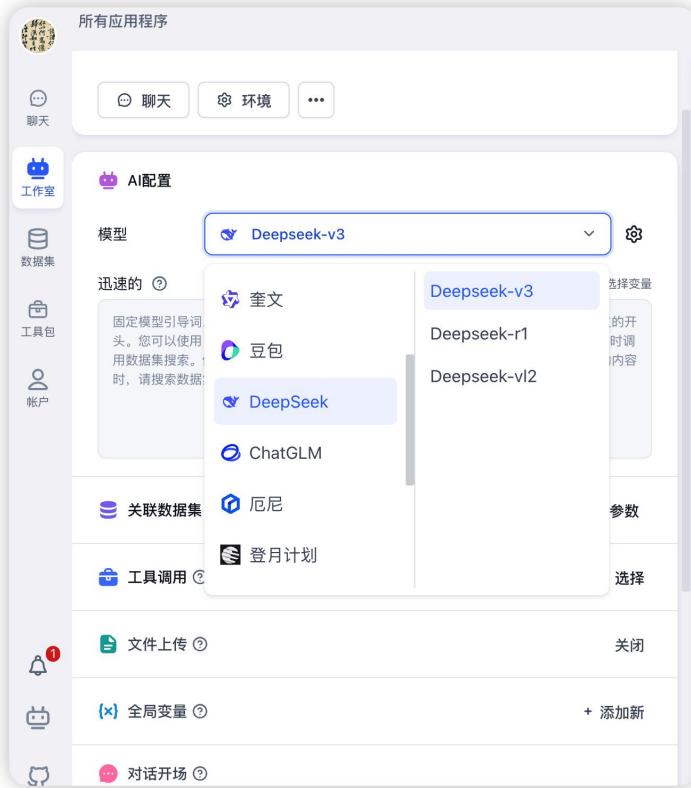
共享权限
模型公开共享
知识公开共享
提示词公开共享
工具公开共享
对话权限
允许上传文件

功能权限
直接连接工具服务器
联网搜索
图像生成
代码解释器
笔记

删除 保存

03 AI编排中台介绍

三、自然语言转SQL系统-AI编排平台-1



支持直接配置对接众多国内、外前沿大模型



数据的集成方式灵活多样

三、自然语言转SQL系统-AI编排平台-2

The screenshot displays the AI scheduling platform's dataset management and knowledge base configuration interface.

Left Sidebar: Includes icons for 聊天 (Chat), 工作台 (Workbench), 知识库 (Knowledge Base), 工具箱 (Toolbox), and 账号 (Account).

Top Bar: 显示根目录 / GeneralDatabase, 并包含 数据集 搜索测试 和 排队情况 按钮。

Dataset Management: 显示文件(1)列表，包含以下列：名称 (amazon)，处理模式 (-)，数据量 (-)，创建/更新时间 (2025-02-08 19:40, 2025-02-08 19:40)，状态 (已就绪)，启用 (开关开启)。右侧有新建/导入按钮和一个红色框选的数据集类型列表：文件夹 (文件夹图标)，手动数据集 (CSV图标)，文本数据集 (DOC图标)，表格数据集 (Excel图标)。

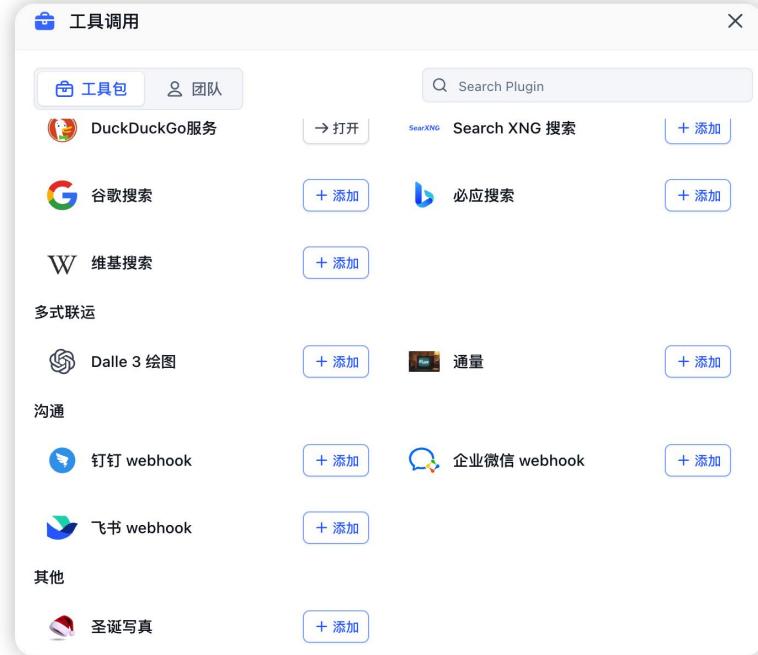
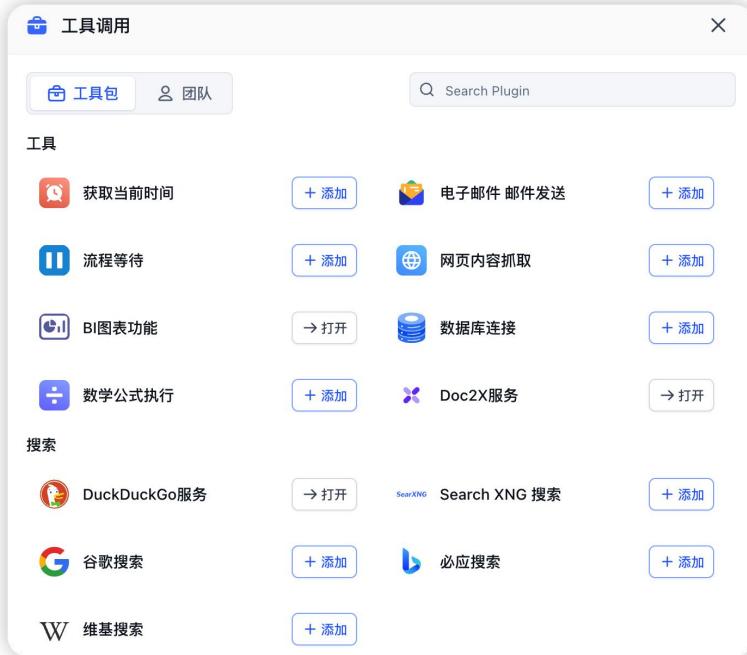
Knowledge Base Configuration: 显示通用知识库 GeneralDatabase 的配置，包括 知识库 ID (67595336b26652c19ee80d53)、索引模型 (m3e)、文本理解模型 (GPT-4o-mini)。

Text: 支持丰富的数据类型，可快速建立垂直领域知识库 (支持丰富的数据类型，可快速建立垂直领域知识库)

Text: 内置众多文本理解模型 可灵活选择 (内置众多文本理解模型 可灵活选择)

Bottom Text: 垂直领域知识库建立操作便捷

3、自然语言转SQL系统-AI编排平台-3



工具调用丰富多样

三、自然语言转SQL系统-AI编排平台-4



三、自然语言转SQL系统-AI编排平台-5

调用HttpAPI ⚙️ • 未保存

工作流程 发布 日志

免登录窗口
分享链接给其他用户，其他用户无需登录即可直接使用

API 请求
通过API集成到现有系统，或者企业微信、飞书等。

钉钉机器人
通过API直接连接钉钉机器人

微信公众号集成
通过API直接连接微信公众号

免登录窗口 ②

创建新链接

姓名	到期时间	积分消耗	返回报价	IP速率限制 (人/分钟)	身份验证	上次使用时间
----	------	------	------	---------------	------	--------

未创建分享链接

提供灵活易用的API接口，可以与众多主流应用进行集成对接

04 复杂场景落地介绍

4、NL2SQL解决方案-实现复杂场景-1

The screenshot shows the NL2SQL application interface. On the left, a sidebar lists navigation options: 简易配置 (Easy Configuration), 高级编排 (Advanced Arrangement), 发布应用 (Publish Application), 对话日志 (Conversation Log), and 立即对话 (Immediate Conversation). The main area is titled "私有" (Private) and contains a sub-section for "test-NL2SQL". It includes tabs for 对话 (Conversation), 发布 (Release), and 设置 (Settings). A central panel titled "应用配置" (Application Configuration) features an AI configuration section with an AI icon, an AI model selection dropdown set to "智谱AI_glm-4", and a text input field containing a placeholder message about being a database assistant. To the right, a "调试预览" (Debug Preview) section displays a query example and its SQL code.

使用rollup和cube语句，按半年度、季度、月度、周度、日度可上卷、下钻查询2024年交易流水成交金额的总量和明细

为了使用 `ROLLUP` 和 `CUBE` 语句实现上卷和下钻查询，我们可以根据 `tnv_dt` 字段来分组数据。以下是根据不同时间维度使用 `ROLLUP` 和 `CUBE` 的SQL示例：

半年度、季度、月度上卷（使用 ROLLUP）

```
sql 复制
SELECT
    TO_CHAR(tnv_dt, 'YYYY') AS year,
    TO_CHAR(tnv_dt, 'YYYY-MM') AS mon,
    TO_CHAR(tnv_dt, 'YYYY-Q') AS quar,
    SUM(tnv_val) AS total_value
FROM
```

在没有任何提示且存在语义二义性的情况下，准确根据自然语言描述生成sql代码。

4、NL2SQL解决方案-实现复杂场景-2

统计截至2024年3月7日我司所有受托类产品的管理规模

The screenshot shows a SQL code editor interface. At the top, there are icons for a brain, a search bar, and a refresh button. Below the title bar, the word "sql" is displayed. On the right side, there is a "复制" (Copy) button. The main area contains the following SQL query:

```
SELECT SUM(fpn.ttl_navl) AS total_management_
FROM fin_prd_nav fpn
JOIN pty_prd_base_info ppbi ON fpn.prd_code = ppbi.prd_code
WHERE ppbi.fofs_code = '000000'
AND (
    (ppbi.prd_code LIKE '1%' AND LENGTH(ppbi.prd_code) = 4)
    OR (ppbi.prd_code LIKE '2%' AND LENGTH(ppbi.prd_code) = 4)
    OR (ppbi.prd_code LIKE '4%' AND LENGTH(ppbi.prd_code) = 4)
    OR (ppbi.prd_code LIKE '5%' AND LENGTH(ppbi.prd_code) = 4)
    OR ppbi.prd_code IN ('7101', '7002', '7003', '7004')
)
AND ppbi.prd_typ != '4'
```

At the bottom of the editor, there are three buttons: "3条上下文" (3 context lines), "4.58s" (execution time), and "查看详情" (View details).

1、用户问题与AI生成的代码

The screenshot shows a database interface with a results window titled "结果 1 X". The query executed is the same as the one in the code editor. The result is a single row with the column name "total_management_scale" and the value "47,680,024,684.66". A red arrow points to this result row.

2、AI生成代码运行的结果

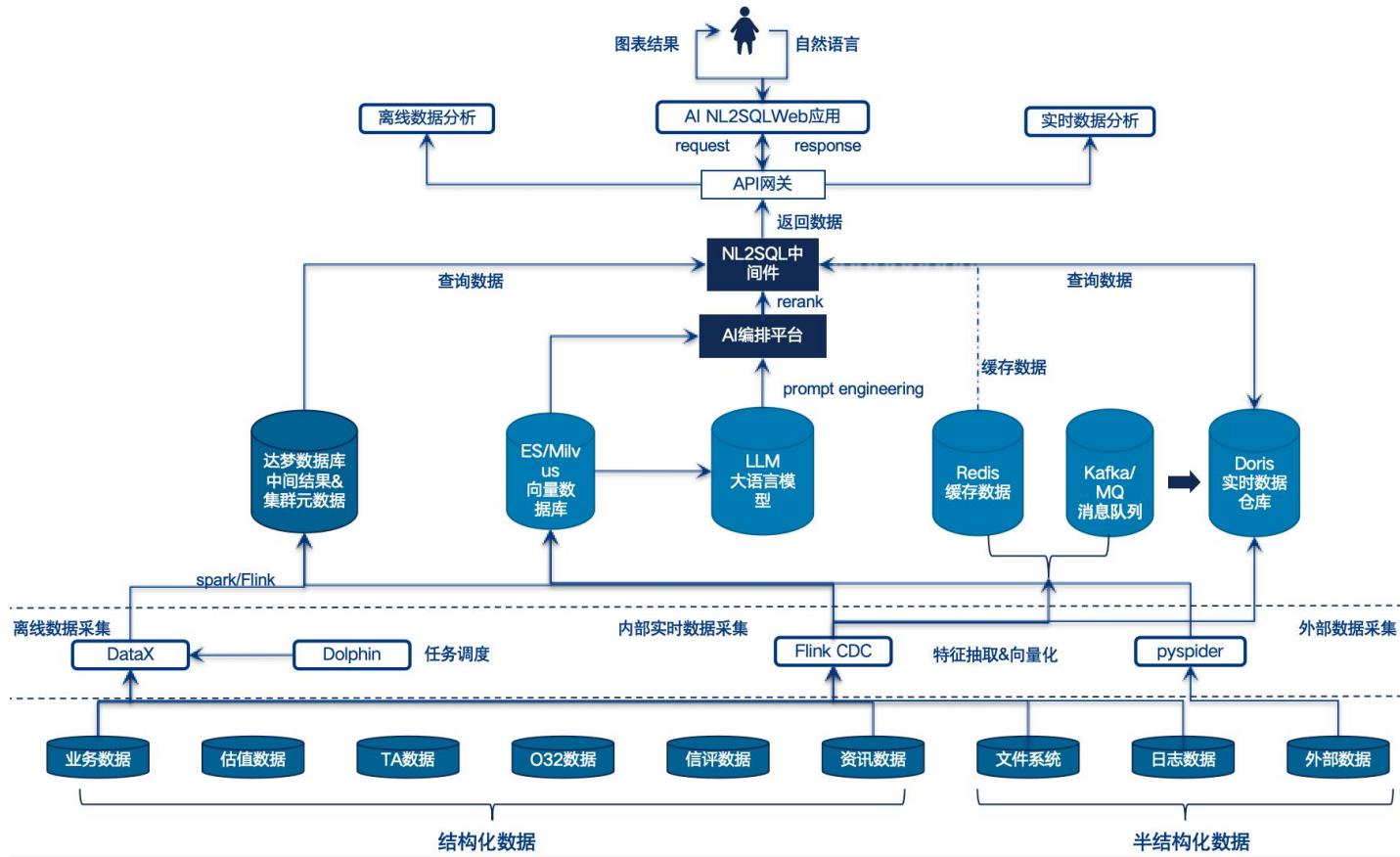
The screenshot shows a business report table with the following data:

统计日期: 2024/3/10			
账户	资产规模 (元)	客户/渠道	客户/渠道类型
总计	47,680,024,685		
non-UL	43,819,412,932	公司	关联方
UL	2,429,042,910	有限公司	关联方
	1,102,036,334	有限公司	关联方
	329,532,508	有限公司	关联方

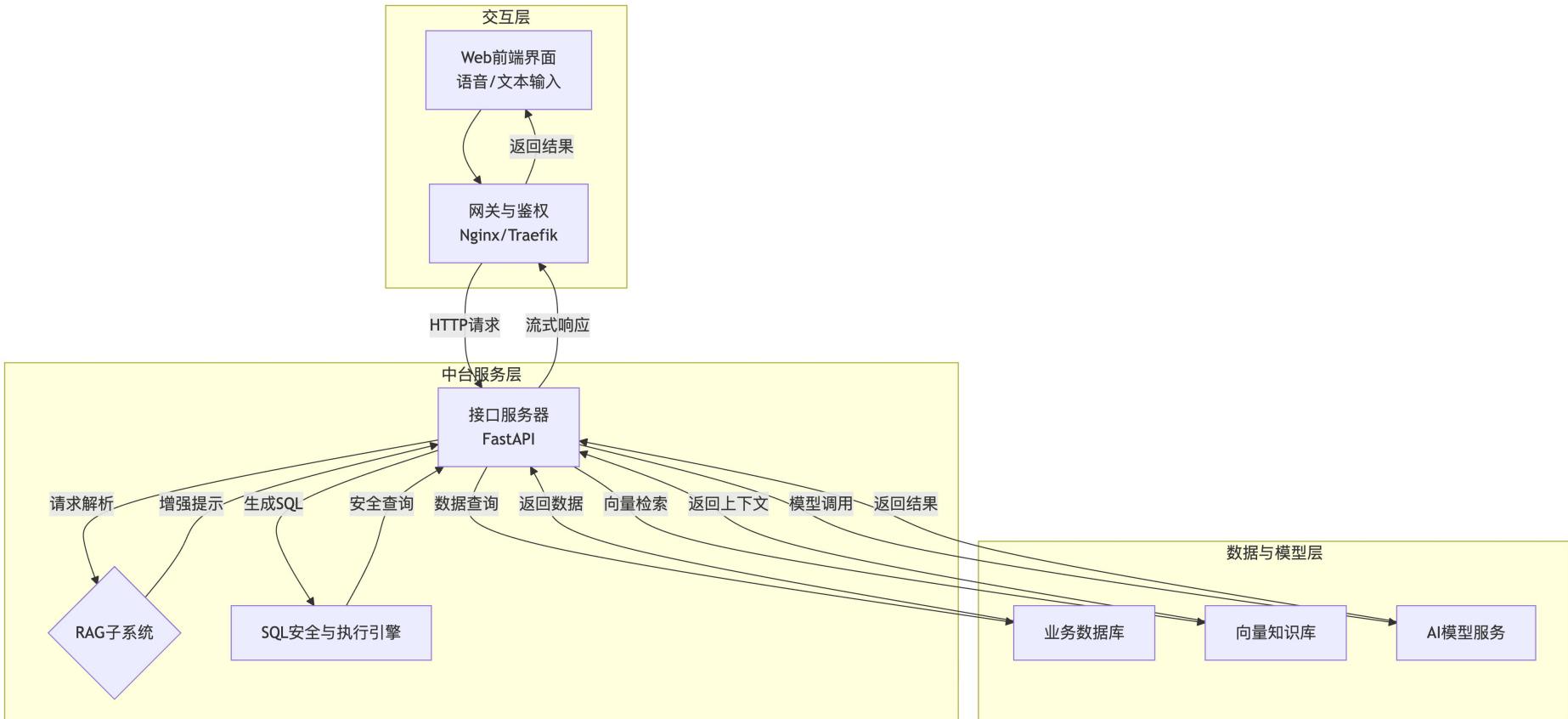
3、业务报送底稿数据

05 整体技术架构

5、整体技术架构



5、整体交互逻辑



5、软硬件采购建议-NVIDIA显卡配套软件方案

序号	大类	分类	名称	规格/版本	数量
1	软件	操作系统	统信UOS	统信 UOS Server V20 (1060e) x86_64 版	1
		数据库	达梦	DM8 官方 Linux 安装包	1
		容器	Docker	Community Edition 24.0.7	1
		显卡驱动	NVIDIA	NVIDIA Driver 535.161.07 (稳定长期支持版)	1
		推理引擎	vLLM	vLLM v0.6.1.post2	1

5、软硬件采购建议-NVIDIA显卡配套硬件方案

序号	大类	分类	名称	规格/版本	数量
2	硬件	服务器	浪潮	浪潮 NF5468M6 (AI 专用服务器),	2
		显卡	NVIDIA	2x NVIDIA A100 80GB PCIe	2
		CPU	Intel	Intel Xeon 8280L (28 核, 2.7 GHz)	2
		主机内存		512GB DDR4 (64GB x 8)	8
		硬盘	长鑫存储	2TB NVMe SSD 10TB SATA SSD, 根据数据量, 可选	1
		网络	H3C	2 x 100GbE 或 1 x 100GbE NIC (高速数据传输和模型分发) 1 x 10GbE 网络接口。	1

5、软硬件采购建议-华为显卡配套软件方案

序号	大类	分类	名称	规格/版本	数量
1	软件	操作系统	统信UOS	统信 UOS Server V20 (1060a) ARM64 版 (鲲鹏架构)	1
		数据库	达梦	DM8 官方鲲鹏 920 (ARM) 架构安装包	1
		容器	Docker	Docker: 24.0.7 (aarch64)	1
		固件驱动	华为	底层驱动 (NPU)CANN 8.0.RC2 .alpha002 (务必与 MindIE 版本对应) (适配 910B/C)	1
		推理引擎	华为	MindIE-Service 1.0.RC2(MindSpore Inference Engine)	1

5、软硬件采购建议-华为显卡配套硬件方案

序号	大类	分类	名称	规格/版本	数量
2	硬件	服务器	华鲲振宇	天宫 AT800 (Model 3000) A2 推理服务器	2
		显卡	华为	华为昇腾 910C显卡 (128GB, HBM3e) 或910B4显卡 (64GB, HBM3)	910C*2 910B*4
		CPU	海光	2 x 海光 64 核处理器 (符合信创架构要求)	2
		内存	长鑫存储	512GB DDR4 (64GB x 8)	8
		硬盘	华为	2TB NVMe SSD, 华为 NVMe SSD 8TB SATA SSD, 可选华为产品,	1
		网络	华为	2 x 100GbE 或 1 x 100GbE NIC。 1 x 10GbE 网络接口	3
		电源		2 x 2000W 电源	2

06 最优实践



6、最优实践-打好根基，规范化数据准备

知识库的质量始于源头数据的质量。所谓“Garbage In, Garbage Out”，混乱的数据源必然导致AI产生幻觉。企业应首先关注：

- **表结构规范**：确保数据遵循标准的二维表格格式，不含合并单元格或多级表头。避免使用交叉表，应将其转换为堆叠格式，这更利于机器解析。
- **表头与表值清晰**：列名应使用简洁明确的业务术语，避免使用内部编码或含糊的缩写。单元格内的数值应保持类型一致，不混入单位或特殊符号。
- **按主题创建数据集**：将描述同一业务主题（如“进销存”、“用户行为”）的相关数据表整合为逻辑上的“数据集”。这能极大降低AI在多表查询时错误理解字段意图的风险。

6. 最优实践-丰富语义，精确内容管理

在规范的数据基础上，需要为数据注入丰富的业务语义，让其“活”起来。

- **定义核心指标与维度**：明确定义企业范围内的核心业务指标（如“活跃用户数”、“客户生命周期价值”）及其计算口径。使用维度建模技术，将数据组织成事实表和维度表，清晰描述业务过程和分析视角。
- **建立业务术语词典（同义词管理）**：系统性地梳理并录入业务“行话”、“黑话”及其对应的标准数据字段。例如，将“客单价”映射为“平均订单金额”，将“新客”映射为“首次购买用户”。这是让ChatBI听懂业务语言的关键一步。
- **文档化与元数据管理**：为每个数据表、字段和指标提供清晰的描述和注释。这些元数据是RAG系统检索上下文、对抗幻觉的重要信息来源。

6、最优实践-注入智能，构建动态维护机制

知识库不是一个静态的资产，而是一个需要持续迭代和优化的生命体。

- **引入人机协作反馈闭环（Human-in-the-Loop）**：当AI的回答不准确时，应建立机制让用户能够轻松地反馈和修正。例如，DataFocus的“小慧点赞”功能，允许用户修改AI的关键词解析结果并“点赞”保存，系统会在后续相似查询中参考此次修正。
- **支持对话式知识沉淀**：允许用户在与ChatBI的对话中，通过简单的指令（如“记住：金牌产品是指年销量排名前10的产品”）动态地向个人或系统知识库中添加新知识，实现知识的即时扩充。
- **持续监控与审计**：定期审计知识库的使用情况和AI查询的准确率，识别出常见的错误模式和知识盲点，并进行针对性优化。



高睿嘉讯

谢谢观看



北京高睿嘉讯科技有限公司 | <https://www.beyondbrain.com.cn>

联系方式: 15652688286 15210341299 赵经理

邮箱: zhaokaien@gmail.com 欢迎垂询!