


 **核心数据处理流程**


 **完整处理流程图**

第一阶段：需求理解与

 用户输入


自然语言查询

例如：生成逾期率分析报告

 意图识别

分析类型判断


- 逾期率分析
- 客群分析
- 财务分析

 Schema检索

RAG技术

- 向量化表结构
- 相关表检索
- 字段映射

第二阶段：SQL生成与

 阶段检查点

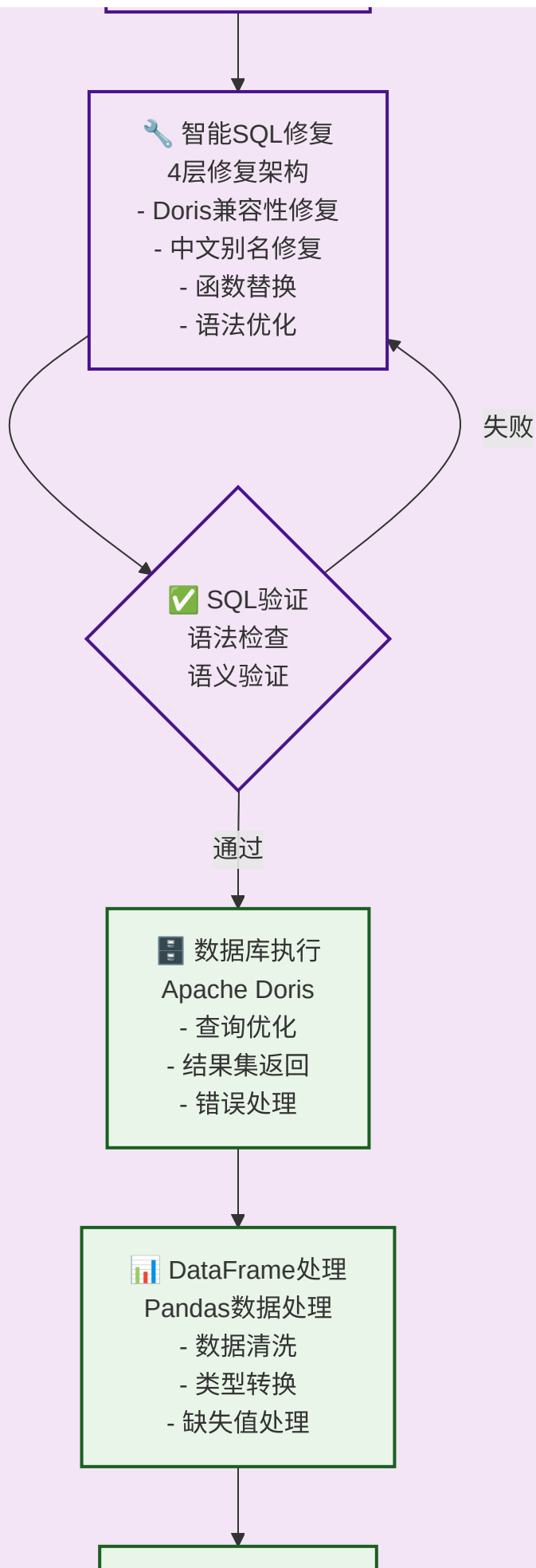
SQL生成准备就绪

- 意图已识别
- Schema已检索
- 准备进入执行阶段

 SQL生成

LLM驱动

- 提示工程
- 模板选择
- 复杂查询构建





统计计算

Python计算引擎

- 基础统计量
- 趋势分析
- 业务指标计算



场景识别

业务场景判断

- 关键词检测
- 数据特征分析
- 分析类型确定



数据驱动分析

DataDrivenAnalyzer

- 统计特征提取
- 趋势识别
- 异常检测



AI洞察生成

LLM分析引擎

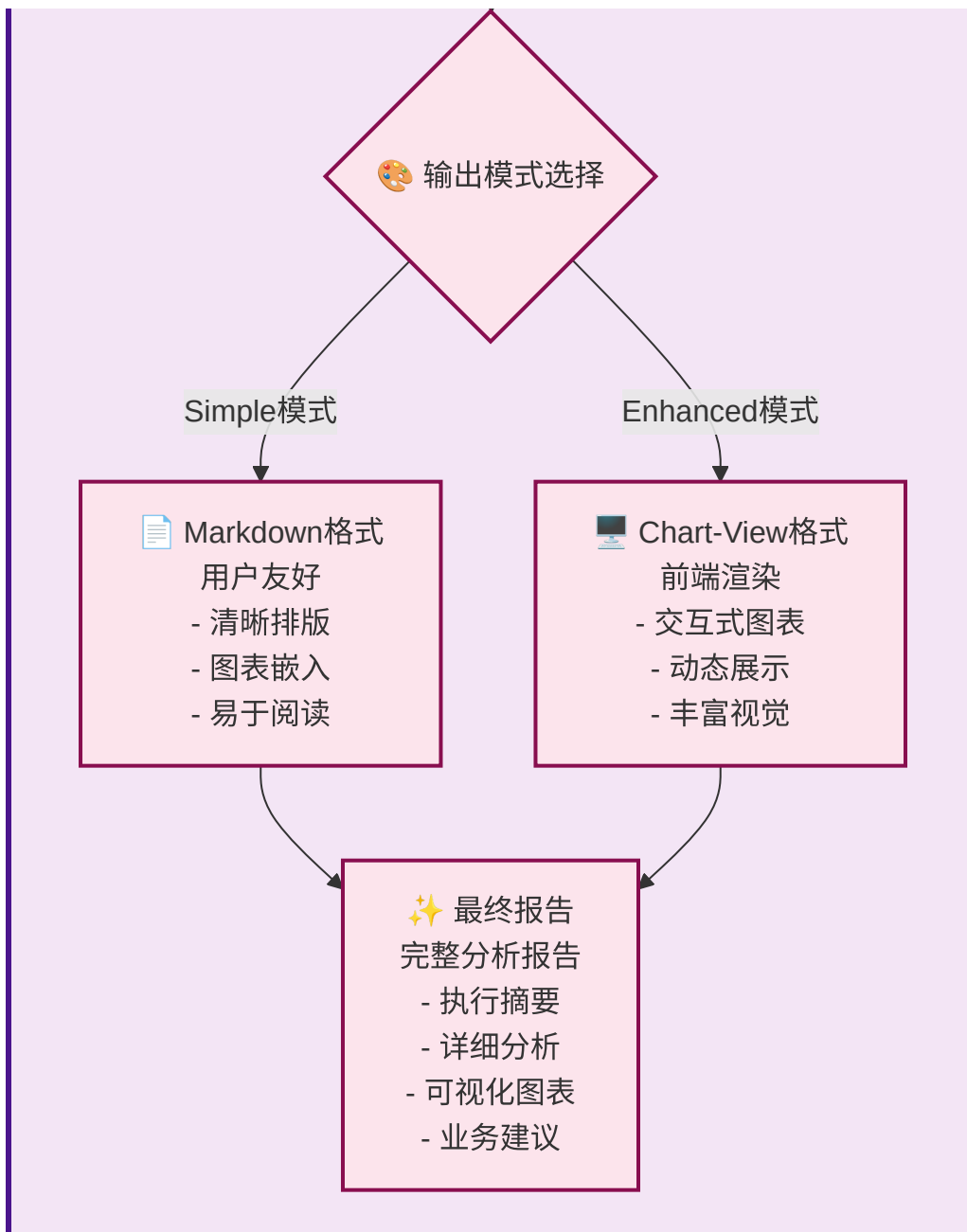
- 业务知识库
- 上下文理解
- 专业洞察



报告结构化

智能报告生成

- 关键发现
- 业务建议
- 风险提示



🎯 两阶段处理架构说明

🎯 第一阶段：需求理解与SQL准备

- **核心任务:** 理解用户需求，准备SQL生成所需的上下文
- **具体步骤:**
 - **用户输入:** 接收自然语言查询
 - **意图识别:** 判断分析类型（逾期率、客群、财务等）
 - **Schema检索:** 使用RAG技术检索相关表结构和字段信息
- **输出:** SQL生成所需的完整上下文信息
- **关键技术:** 自然语言理解 + RAG检索技术
- **阶段特点:** 轻量级处理，主要为第二阶段做准备

⚡ 第二阶段：SQL生成与智能分析

- **核心任务:** 从SQL生成到最终报告的完整数据处理流程
- **具体步骤:**
 - i. **SQL生成:** LLM驱动的智能SQL构建
 - ii. **SQL修复:** 4层修复架构解决兼容性问题
 - iii. **数据查询:** Apache Doris数据库执行
 - iv. **数据处理:** Pandas数据清洗和统计计算
 - v. **AI分析:** 数据驱动的智能洞察生成
 - vi. **报告生成:** 结构化专业分析报告输出
- **输出:** 完整的数据分析报告
- **关键技术:** LLM生成 + 智能修复 + 数据驱动分析
- **阶段特点:** 重量级处理，包含所有核心计算和分析逻辑

🔧 执行要点

1. **自动化程度高:** 用户只需输入自然语言，系统自动完成所有处理
 2. **错误自愈能力:** SQL修复器自动解决95%的兼容性问题
 3. **数据驱动:** 100%基于真实数据计算，确保分析准确性
 4. **双模式输出:** 同时满足简单阅读和复杂展示需求
-