

故障诊断系统手册

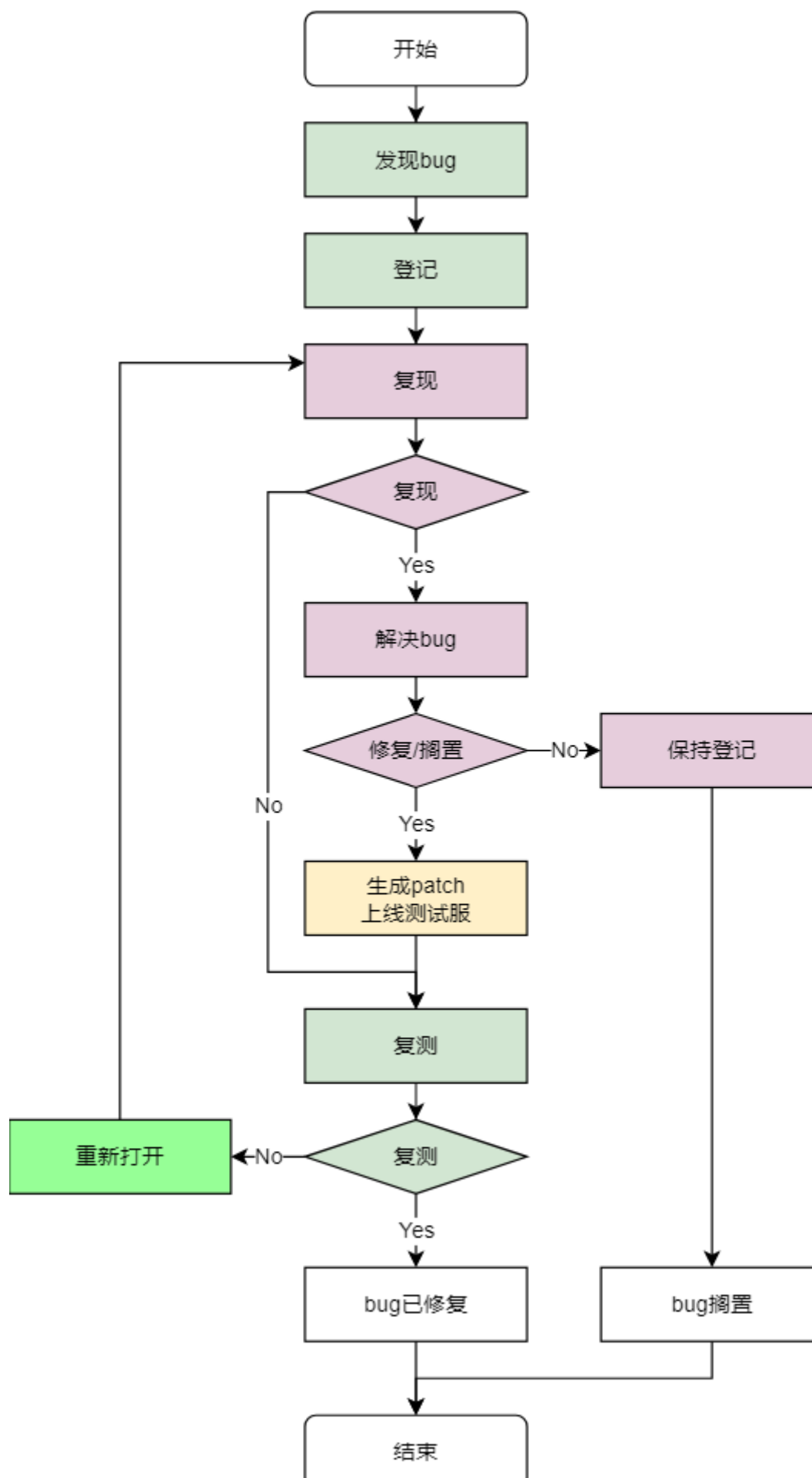
故障诊断系统手册

总体大纲

- 1. 开发
- 2. 测试
 - 2.1 测试大纲
 - 2.2 测试平台
 - 2.3 Bug管理
- 3. 部署
 - 3.1 部署准备
 - 3.2 部署方法
- 4. 使用情况
 - 4.1 华中数控使用情况
 - 4.2 使用案例
 - 4.2 大模型带来的帮助与不足
- 5. 后续工作
 - 5.1 目的
 - 5.2 计划

总体大纲

- 软件开发方法：快速原型
- 设计流程：
需求文档-设计文档-开发-测试-部署
- 迭代流程



要点说明

(1) 绿色为测试控制、紫色为开发控制、黄色运维控制

(2) Bug状态说明

- 1、打开 (open)
 - 2、已解决 (resolved)
 - 3、重新打开 (reopen)
 - 4、关闭 (closed)
- 开发只能对Bug状态进行“已解决”的修改
 - 测试对Bug状态进行打开与关闭的操作

(3) Bug的解决方案

- 1、已修复 (fixed)
 - 2、不予修复 (won't fix)
 - 3、推迟修复 (postpone)
 - 4、无法重现 (not repro)
 - 5、重复 (duplicate)
 - 6、设计如此 (by design)
 - 7、外部原因 (external)
 - 8、其他 (other, 如数据原因, 环境原因, 配置原因等)
- 只能由开发人员填写Bug的解决方案

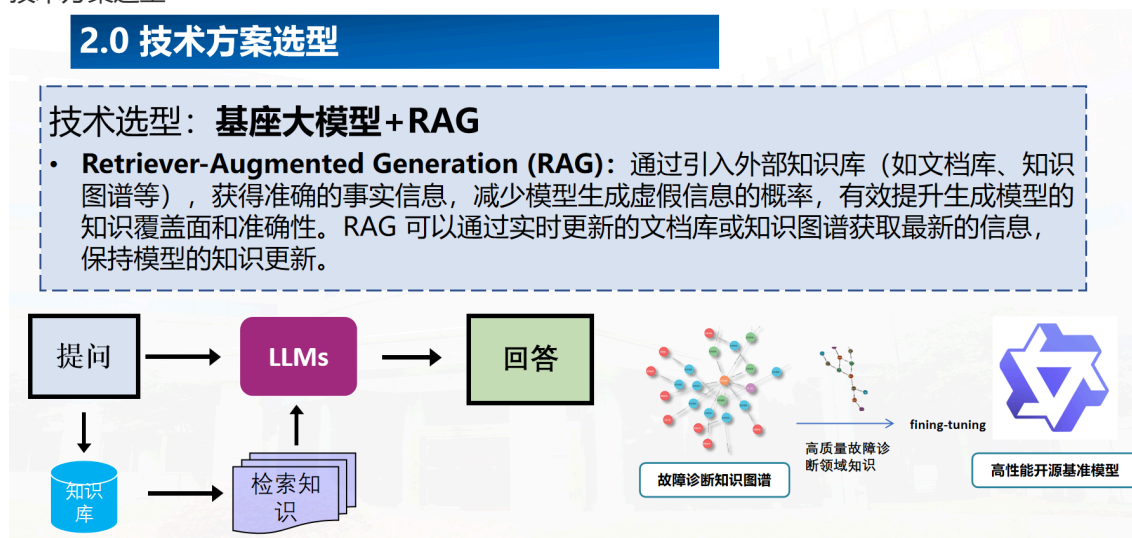
(4) 流程控制

当Bug状态变为已解决和关闭时, Bug的解决方案为必填字段, 否则不能改变状态
当Bug状态变为打开或重新打开时, 清空Bug的解决方案字段

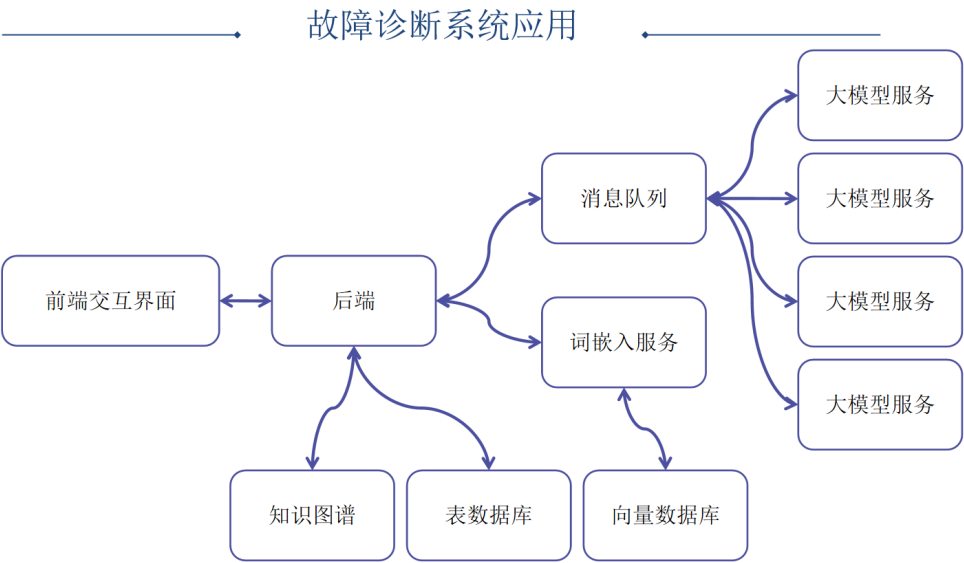
1. 开发

经过探索, 大模型在智能化数控系统多个场景得到应用, 如: 加工参数辅助生成、智能编程、人机交互、故障诊断、

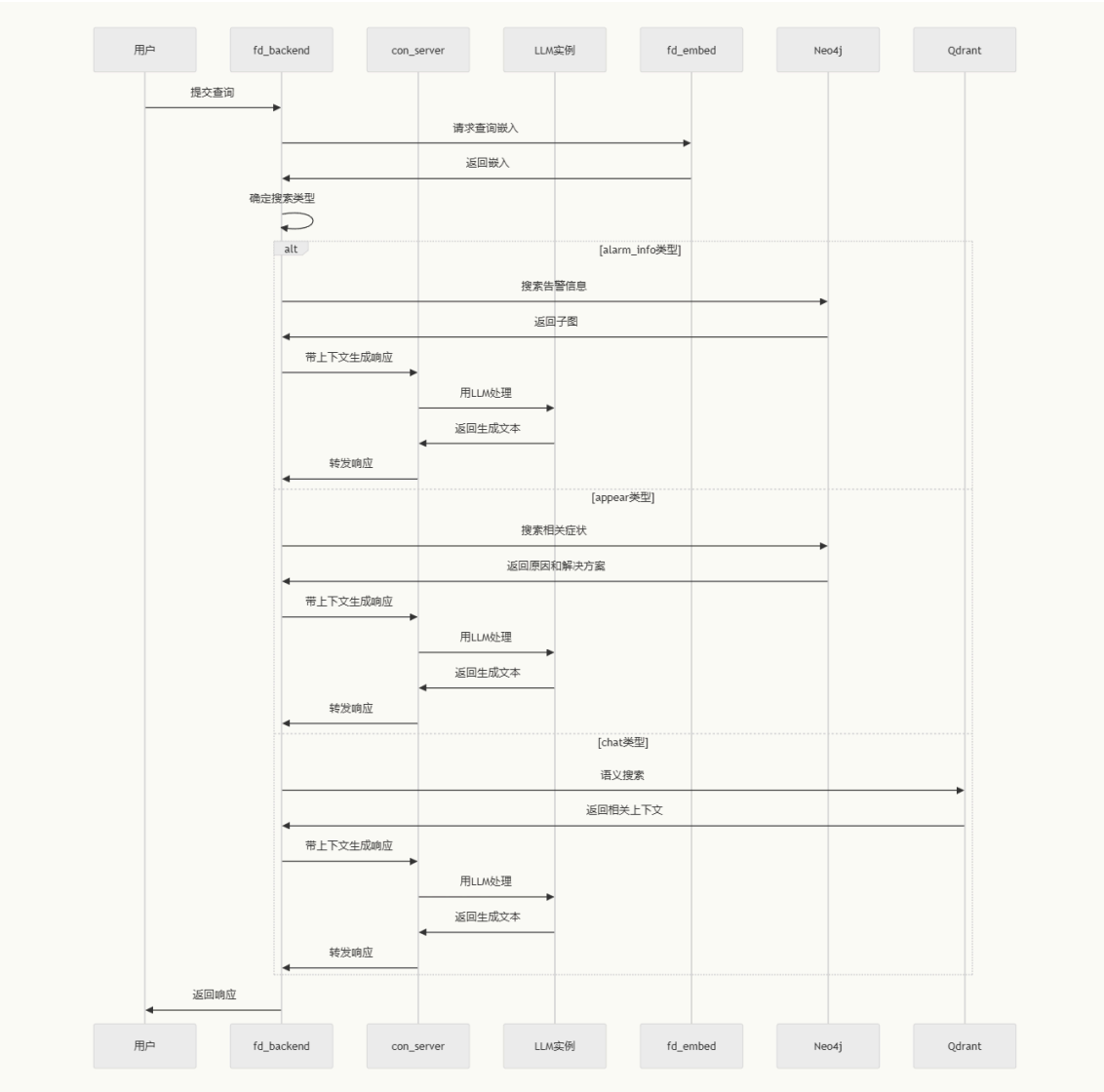
- 技术方案选型



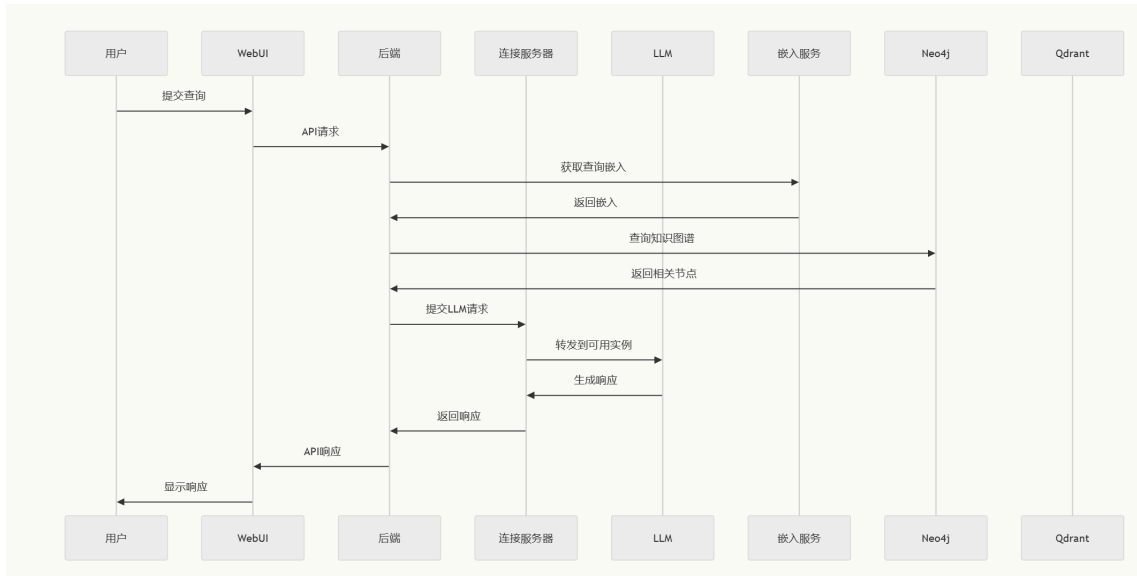
• 系统架构



• 数据流架构



- 通信流程



- 镜像 (部分)

- 镜像

- mysql**
 - 镜像:版本 - mysql:8.0
- neo4j**
 - 镜像:版本 - neo4j:5.17-community
- qdrant**
 - 镜像:版本 - qdrant/qdrant:v1.10.0
- redis**
 - 镜像:版本 - redis:6.0.16
- chat_server**
 - 镜像:版本 - chat_server:1.5_8.9/8.6
 - 简介: 用于运行大模型的环境, 其中包含大模型的模型文件
- embed_server**
 - 镜像:版本 - embed_server:v1.1
 - 简介: 用于运行数据向量化的环境, 其中包含模型文件
- baseenv**
 - 镜像:版本 - iss/baseenv:v1.10
 - 简介: 用于运行后端服务的环境

2. 测试

2.1 测试大纲

一、测试范围（优先级排序）

| 模块 | 二级模块 | 基准测试点 | 边界值测试点 |
|-------------|---------|--------------------------------|--|
| 基于知识图谱的故障诊断 | 对话 | 高频故障答案相关性（如何服过载/编码器异常） | （1）输入边界值 长度边界 内容边界 语言与编码边界 请求格式边界 （2）输出边界值 输出长度边界 错误处理与输出 |
| | 多轮对话 | 连续追问故障原因的逻辑连贯性 跨故障类型上下文保持能力 | 多轮对话边界 |
| 知识图谱管理 | 图谱概览 | 展示知识图谱的信息 动态关联展示正确性 | |
| | 数据上传 | 实现基于模板的数据上传 上传数据后按上传数据进行回答 | 上传文件边界 文件内容 表头测试 缺失值测试 |
| | 知识融合 | 机床厂私有知识与系统知识隔离性 | |
| 学习机制 | 反馈检测与处理 | 用户工程师实施反馈，系统自动检测、处理、并加入反馈数据库 | 反馈安全性检查 反馈的空置检查 |
| | 反馈更新 | 审查后知识同步更新 更新后诊断答案修正验证 | |

2.2 测试平台

| | | | | | | | | | |
|--|-----------------------------|----|------------------|----|------|------|------|--|--|
| 武汉智能设计... > 故障诊断系统测试库 > 计划 用例 基线 评审 报表 搜索 (Ctrl+K) + ? | | | | | | | | | |
| 用例 | 全部用例 | | | | | | | | |
| 全部用例 · 103 | Q 搜索 (Ctrl+G) 筛选 排序 103 个用例 | | | | | | | | |
| 用例模块 | <input type="checkbox"/> | 编号 | 标题 | 状态 | 重要程度 | 用例类型 | 执行结果 | | |
| 基于知识图谱的故障诊断模块 · 77 | | 1 | ISS-1 知识检索准确性 | 就绪 | P1 | 功能测试 | 受阻 | | |
| > 基准测试点 · 43 | | 2 | ISS-2 知识检索准确性1 | 就绪 | P1 | 功能测试 | 通过 | | |
| > 边界值测试点 · 34 | | 3 | ISS-3 知识检索准确性2 | 就绪 | P1 | 功能测试 | 失败 | | |
| 知识图谱管理模块 · 17 | | 4 | ISS-4 知识检索准确性3 | 就绪 | P1 | 功能测试 | 通过 | | |
| 图谱概览 · 2 | | 5 | ISS-5 知识检索准确性4 | 就绪 | P1 | 功能测试 | 受阻 | | |
| 数据上传 · 14 | | 6 | ISS-6 知识检索准确性5 | 就绪 | P1 | 功能测试 | 受阻 | | |
| 知识融合 · 1 | | 7 | ISS-7 知识检索准确性6 | 就绪 | P1 | 功能测试 | 受阻 | | |
| 学习机制模块 · 9 | | 8 | ISS-8 知识检索准确性7 | 就绪 | P1 | 功能测试 | 通过 | | |
| 反馈检测与处理 · 1 | | 9 | ISS-9 知识检索准确性8 | 就绪 | P1 | 功能测试 | 通过 | | |
| 反馈更新 · 8 | | 10 | ISS-10 知识检索准确性9 | 就绪 | P1 | 功能测试 | 通过 | | |
| 无模块用例 · 0 | | 11 | ISS-11 知识检索准确性10 | 就绪 | P1 | 功能测试 | 受阻 | | |
| | | 12 | ISS-12 知识检索准确性11 | 就绪 | P1 | 功能测试 | 通过 | | |
| | | 13 | ISS-13 知识检索准确性12 | 就绪 | P1 | 功能测试 | 通过 | | |
| | | 14 | ISS-14 知识检索准确性13 | 就绪 | P1 | 功能测试 | 通过 | | |
| | | 15 | ISS-15 知识检索准确性14 | 就绪 | P1 | 功能测试 | 通过 | | |
| | | 16 | ISS-16 知识检索准确性15 | 就绪 | P1 | 功能测试 | 通过 | | |
| | | 17 | ISS-17 知识检索准确性16 | 就绪 | P1 | 功能测试 | 通过 | | |
| | | 18 | ISS-18 知识检索准确性17 | 就绪 | P1 | 功能测试 | 通过 | | |

2.3 Bug管理

故障诊断系统

了解表格即文档

上次编辑 昨天 20:26 | 自动

搜索

全部数据 : (开发) 待解决Bug | (测试) 待复测视图 | 添加记录 | 字段管理 | 分组(1) | 筛选 | 排序 | 行高 | 填色

| Bug编号 | (测试) Bug内容 | (开发或测试) | (测试) |
|-------|--|-----------------|------|
| 1 | 提问前后包含回车向大模型提问显示接口异常 | 关闭 (closed) | 1 |
| 2 | 进行部分sql注入时，系统未能完全拦截，具有一定安全风险 | 打开 (open) | 2 |
| 51 | 提交工程师反馈时，若文本里还有诸如“关闭”等字样，会被系统判定为攻击行为，阻止加入反馈... | 打开 (open) | 2 |
| 50 | 工程师反馈为空时，仍能提取信息，在系统中收到待审核信息 | 打开 (open) | 2 |
| 47 | 反馈检测后，检查工程师反馈节点，标签不准确，from_feedback: false | 打开 (open) | 2 |
| 45 | 上传文件有缺失项时，没有提示 | 关闭 (closed) | 2 |
| 44 | 上传带有超长字符文本的文件，在回答时不能正常显示，系统提示超时报错 | 关闭 (closed) | 2 |
| 43 | 按正确格式上传文件后，不能通过“现象”得到确的“报警原因”及“解决方案” | 打开 (open) | 2 |
| 41 | 上传表头混乱的文件，未作区分，将“现象”与“报警信息”弄混 | 关闭 (closed) | 2 |
| 39 | 上传空文件时，仍能上传成功 | 关闭 (closed) | 2 |
| 38 | 数据更新后，根据现象提问后与系统原有节点相似度过高，返回答案不匹配 | 打开 (open) | 2 |
| 37 | 进行上下文记忆攻击，结果显示报错 | 关闭 (closed) | 2 |
| 36 | 进行上下文记忆攻击，系统执行了虚拟数据操作 | 重新打开 (reop... | 2 |
| 35 | 进行密码学攻击时，未能识别密码学攻击模式 | 关闭 (closed) | 2 |
| 34 | 输入大量不超过token值的并且参杂有关技术杂乱文本，但仍能得到答案 | 打开 (open) | 2 |
| 28 | 输入超长字母文本后，系统报错 | 关闭 (closed) | 2 |
| 21 | 提问某些问题后，未能正常解答 | 已解决 (resolve... | 2 |
| 15 | 反馈抽取出的路径不准确率 | 打开 (open) | 2 |
| 16 | 系统中存在未能正确的合并的节点 | 已解决 (resolve... | 2 |
| 17 | 反馈由已更新修改为未更新，系统无法处理 | 打开 (open) | 2 |

3. 部署

3.1 部署准备

1. 硬件资源

2. 软件环境

- 操作系统：ubuntu 22.04
- GPU驱动：nvidia-driver-550
- 容器化工具：Docker, Docker-compose, nvidia-container-runtime
- 网络环境：外网可访问

3. 镜像版本(所有镜像由HS提供)

- 镜像
 - mysql
 - 镜像:版本 - mysql:5.7
 - neo4j
 - 镜像:版本 - neo4j:5.17-community
 - qdrant
 - 镜像:版本 - qdrant/qdrant:v1.10.0
 - redis

- 镜像:版本 - redis:6.0.16
- **chat_server**
 - 镜像:版本 - chat_server:1.3
 - 简介：用于运行大模型的环境，其中包含大模型的模型文件
- **embed_server**
 - 镜像:版本 - embed_server:v1.0
 - 简介：用于运行数据向量化的环境，其中包含模型文件
- **baseenv**
 - 镜像:版本 - iss/baseenv:v1.9
 - 简介：用于运行后端服务的环境

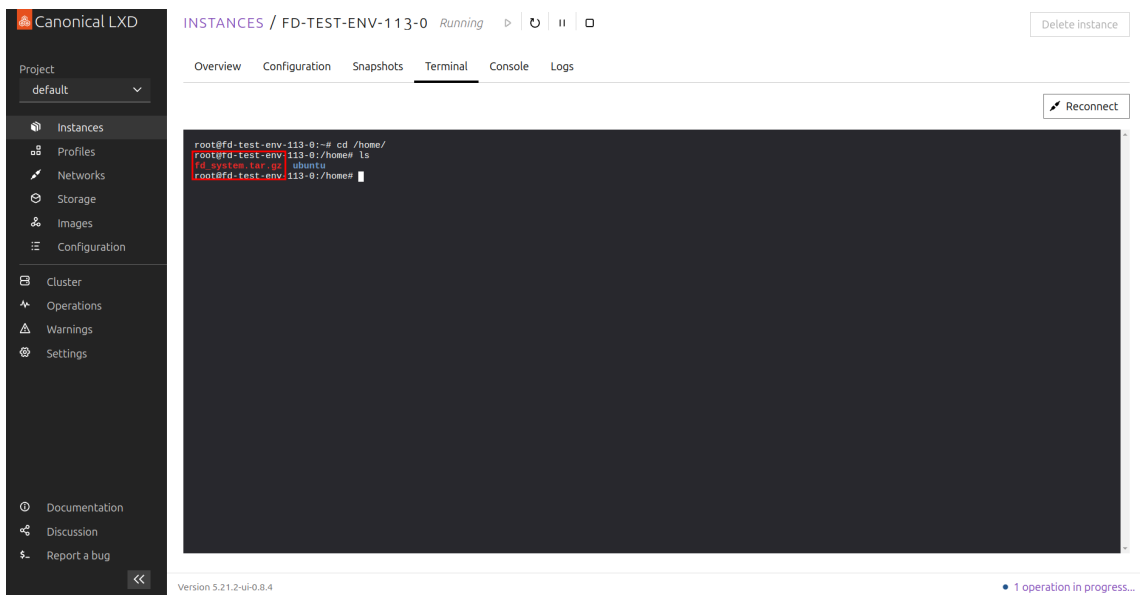
4. 部署地址

- 网络：wifi连接-NERC-INTERNAL wifi密码：Nerc@123
- 网址：<https://192.168.69.150>
- 账号：名字拼音 密码：Nerc@名字首字母小写 root密码：Nerc@名字首字母小写

3.2 部署方法

1. 将系统压缩包复制到/home/目录下

```
# path : /home/
# └─ fd_system.tar.gz
```



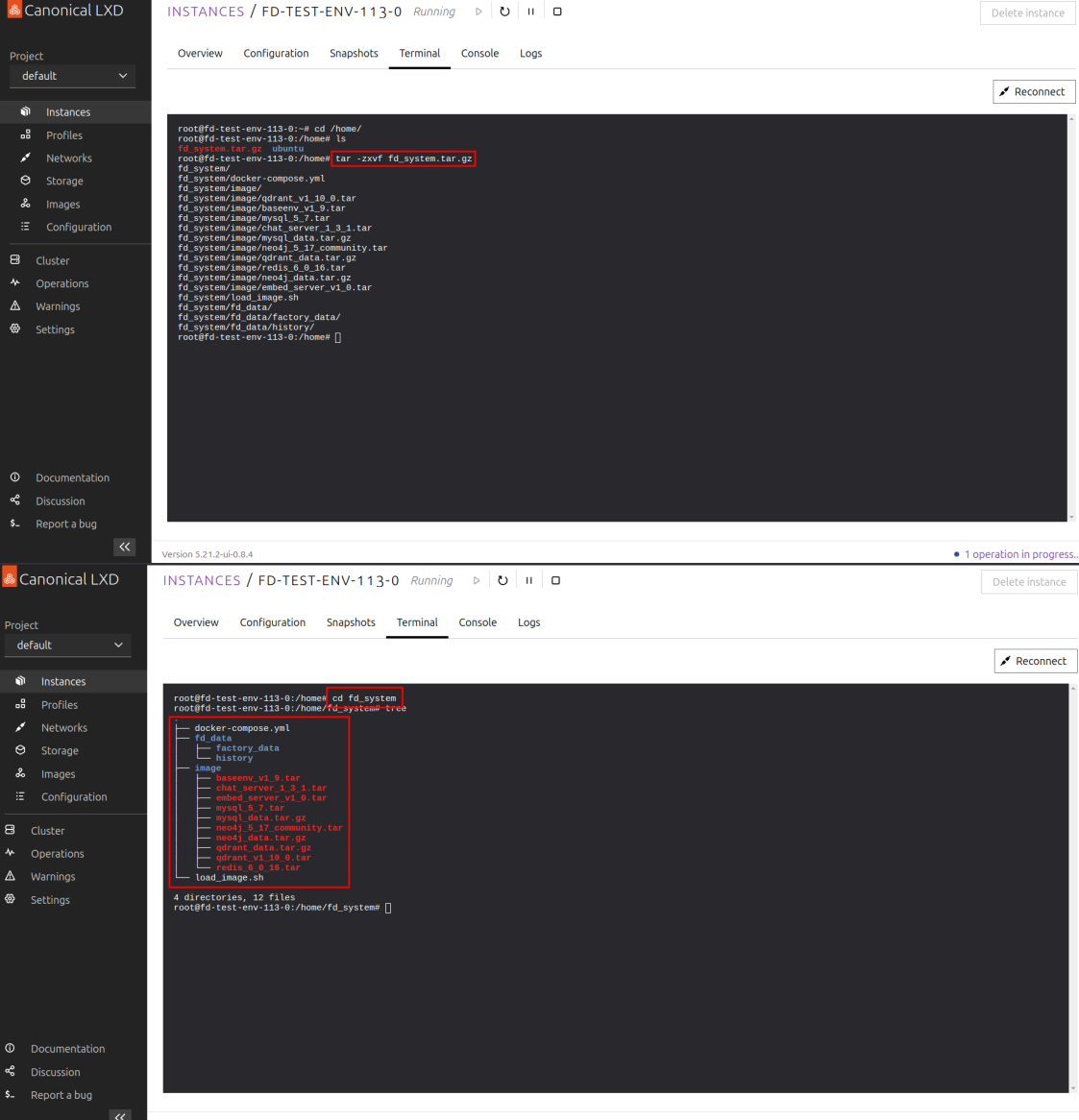
2. 解压系统压缩包

```
# path : /home/
tar -zxvf fd_system.tar.gz
cd fd_system
tree -L 2

# 解压后的目录:
# └─ fd_system
#     └─ chat_server
#     └─ con_server
```



```
# |— docker-compose.yml
# |— fd_backend
# |— fd_data
# |— fd_embed
# |— image
# |   |— mysql_5_7.tar
# |   |— neo4j_5_17_community.tar
# |   |— qdrant_v1_10_0.tar
# |   |— redis_6_0_16.tar
# |   |— chat_server_1_3.tar
# |   |— embed_server_v1_0.tar
# |   |— baseenv_v1_9.tar
# |— interaction
# |— kg
# |— webui
```



The screenshot shows the Canonical LXD web interface. On the left is a sidebar with navigation options like Project, Instances, Profiles, Networks, Storage, Images, Configuration, Cluster, Operations, Warnings, and Settings. The main panel displays the 'INSTANCES / FD-TEST-ENV-113-0' page, which is currently in the 'Terminal' tab. The terminal shows the following commands and output:

```
root@fd-test-env-113-0:~# cd /home/
root@fd-test-env-113-0:/home# ls
fd_system.tar.gz ubuntu
root@fd-test-env-113-0:/home# tar -zxvf fd_system.tar.gz
fd_system/
fd_system/docker-compose.yml
fd_system/image/
fd_system/image/qdrant_v1_10_0.tar
fd_system/image/baseenv_v1_9.tar
fd_system/image/mysql_5_7.tar
fd_system/image/chat_server_1_3_1.tar
fd_system/image/mysql_data.tar.gz
fd_system/image/neo4j_5_17_community.tar
fd_system/image/qdrant_data.tar.gz
fd_system/image/redis_6_0_16.tar
fd_system/image/neo4j_data.tar.gz
fd_system/image/embed_server_v1_0.tar
fd_system/load_image.sh
fd_system/fd_data/
fd_system/fd_data/factory_data/
fd_system/fd_data/history/
root@fd-test-env-113-0:/home#
```

The output of the 'tar' command is listed below:

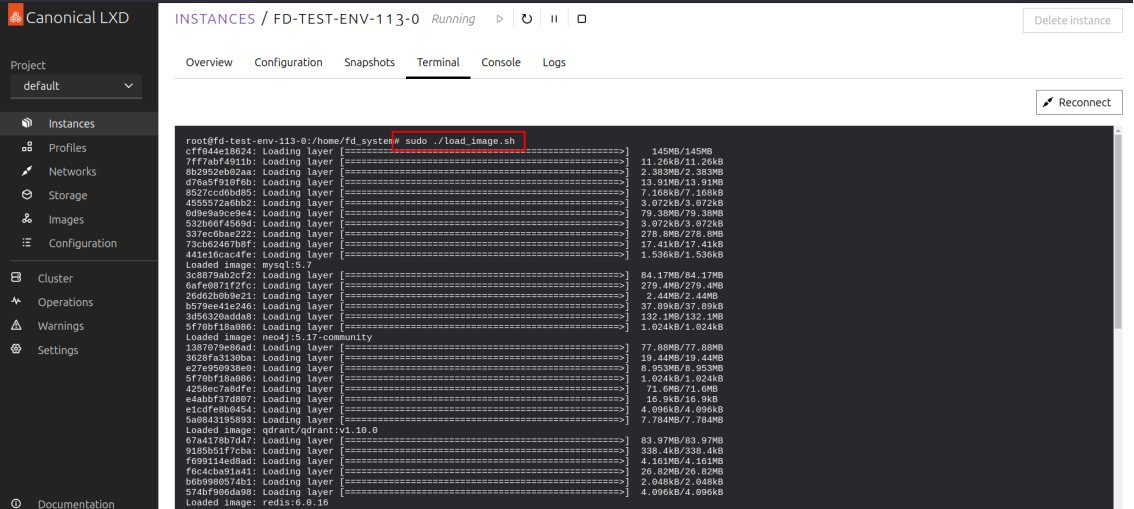
```
— docker-compose.yml
fd_data
  factory_data
  history
image
  baseenv_v1_9.tar
  chat_server_1_3_1.tar
  embed_server_v1_0.tar
  mysql_5_7.tar
  mysql_data.tar.gz
  neo4j_5_17_community.tar
  neo4j_data.tar.gz
  qdrant_data.tar.gz
  qdrant_v1_10_0.tar
  redis_6_0_16.tar
load_image.sh

4 directories, 12 files
root@fd-test-env-113-0:/home/fd_system#
```

3. 载入镜像及数据卷

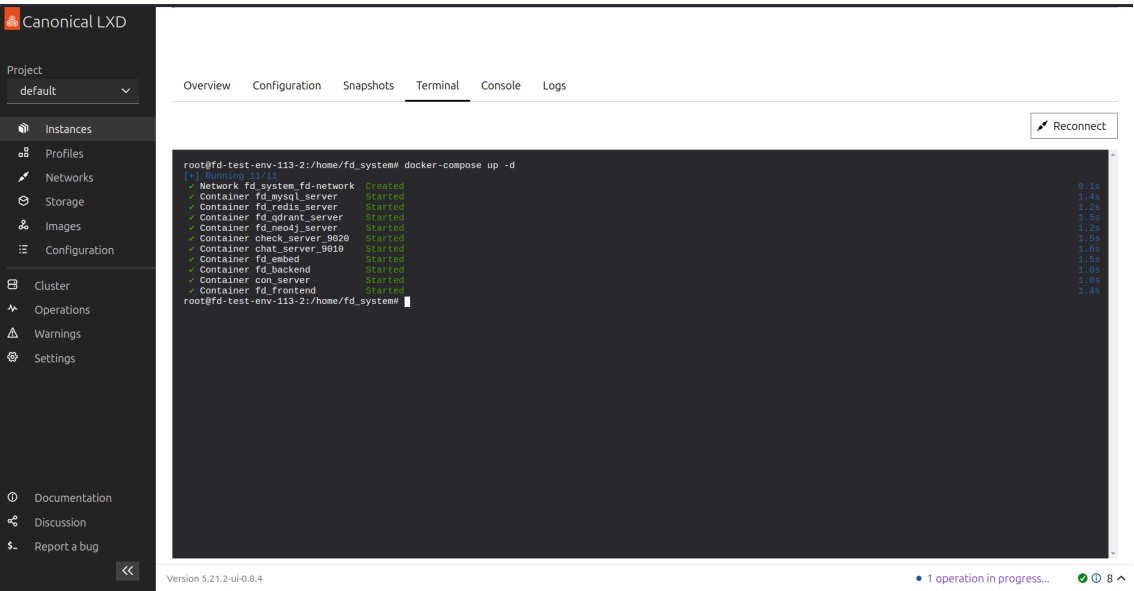
```
# path /home/fd_system
sudo ./load_image.sh
```

注意：载入镜像及数据卷需要一定时间，请耐心等待



4. 基于docker compose启动服务

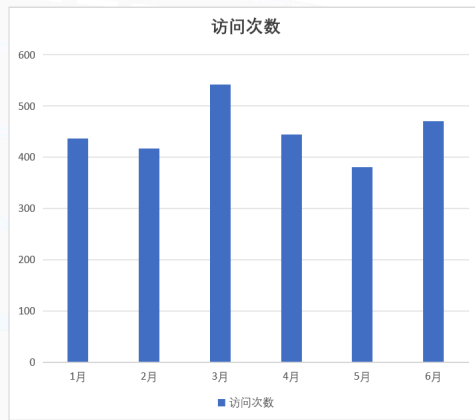
```
# path /home/fd_system
sudo docker-compose up -d
```



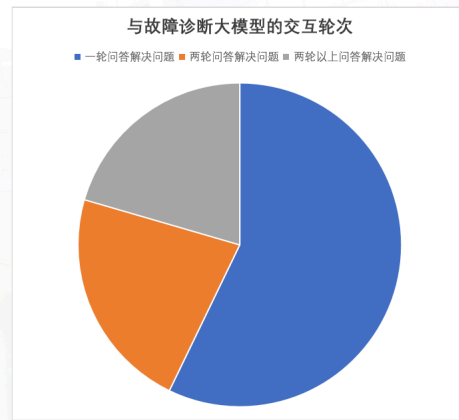
4. 使用情况

4.1 华中数控使用情况

日常使用数据，需要大模型补充



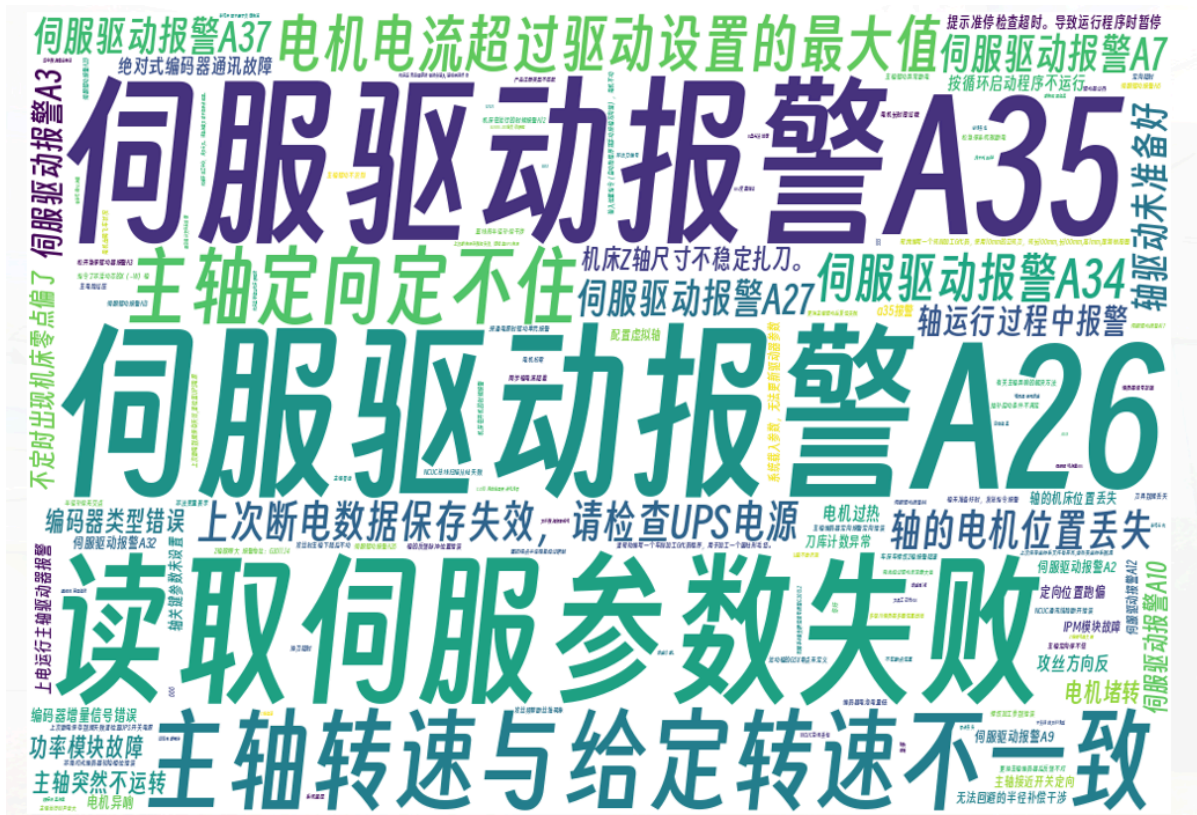
每月访问400+次



累计访问6693次

4

日常工程师咨询较多的案例



4.2 使用案例

远程桌面+大模型

- 大模型解决一些常见直观的问题

- 和远程桌面结合，解决一些常见参数优化和调整的问题
提供真正的远程故障处理手段。



4.2 大模型带来的帮助与不足

帮助

- 经验案例的共享，尤其对经验相对不丰富的人员可以通过平台获取他人经验参考。相同故障案例直接引用，提升现场故障判断的效率
- 对一些产品知识点可以快速的查询，免去查找说明书等复杂过程
- 结合远程桌面功能，更快提供问题处理

待改进

- 用户方使用积极性不高，传统直接打电话、微信和系统方沟通的形式还是主流，使用大模型的习惯还需要慢慢的培养。
- 习惯培养的过程，就需要大模型提供的答复越来越精准，这点大模型知识库还需要改进。

5. 后续工作

5.1 目的

现阶段故障诊断系统知识质量不足，针对这一问题计划分两个阶段，与华中数控工程师进行故障诊断数据审核。

第一阶段（25年7月底~25年9月底）：聚焦伺服驱动领域高频报警数据和常见故障案例

第二阶段（25年10月初~25年12月底）：聚焦其他工单数据和案例数据

5.2 计划

