开发和编译系统要求

unbuntu 22.04+

cpu 需要支持 x86\_64v2

2核CPU 16G内存+

100G可以用硬盘空间+

~~网络支持访问谷歌~~

build proxy RUN go env -w GOPROXY=https://goproxy.cn

docker 22+

openjdk 17

pythone 3.10 +

node 21+

项目代码涉及模块

metadata-models 定义数据库字段，用于配置自动生成代码

metadata-service 自动生成的代码

datahub-graphql-core 后端 关键代码

datahub-frontend 前端和后端的中间层，基本不涉及修改

datahub-web-react 前端

开发运行

启动 ./gradlew quickstartDebuger <https://datahubproject.io/docs/docker/development>

修改 metadata-models 后 ./gradlew :metadata-service:war:build

修改 datahub-graphql-core 后 ./gradlew :metadata-service:war:build

修改 datahub-web-react 后 ./gradlew :datahub-frontend:build

重新启动 ./gradlew :docker:debugReload

启动前端实时看效果 yarn start

部署运行

代码格式化 ./gradlew :datahub-graphql-core:spotlessApply

启动 <https://datahubproject.io/docs/developers>

./gradlew build

./gradlew quickstart

隐藏 Explore your data&Platforms…

<https://github.com/imindzzz/datahub/commit/2f1c8b9832b6bbf5520901081eb3af1ee92b0b6c>

<https://github.com/imindzzz/datahub/commit/92566aed51bb76bf620685c0fab8171e4092268f>

<https://github.com/imindzzz/datahub/commit/4922a02fad164094fceb51c1084ade94979118e4>

const showExportYourData = me.platformPrivileges?.manageIngestion;

前端定位到显示位置，通过账号权限判断是否需要显示

Dataset Config & Remove

<https://github.com/imindzzz/datahub/commit/17a3f2ef387b6fcd9da6c6eab8be08a7ef446522>

Config功能，Dataset 本身有一个description字段，通过字符串搜索定位到有哪些地方相关代码

修改 metadata-models/src/main/.../dataset/EditableDatasetProperties.pdl

修改 datahub-graphql-core/src/main/resources/entity.graphql 给前端提供字段

~~重新编译生成service代码。为了能在后面修改代码是时有代码提示~~

修改datahub-graphql-core/src/main/..../DatasetUpdateInputMapper.java 增加getter setter

修改datahub-graphql-core/src/main/.../DatasetMapper.java 增加getter setter

重新编译生成service代码

修改 datahub-web-react/src/graphql/fragments.graphql 前端字段

实现前端页面操作

<https://github.com/imindzzz/datahub/commit/32ed4ead8852b5eb0a7e74d13e80a79cba205ff8>

Remove功能，把所有字段置为字符串

BusinessRelation BusinessDescription

<https://github.com/imindzzz/datahub/commit/2bf26fcd11b65e0d3b361236fa4a28088b5ba832>

Dataset Schema 本身有一个description字段，通过字符串搜索定位到有哪些地方相关代码

…..

Excel Preview

图形用户界面, 应用程序, 表格, Excel

描述已自动生成

容器外部文件目录 /home/tempuser/files

容器内部文件目录 /home/datahub/excels

<https://github.com/imindzzz/datahub/commit/4de1fb23ea139bb2cf170117d28ac6f8d8bbd11a>

第一版纯前端实现，读取浏览器上传文件后解析显示

<https://github.com/imindzzz/datahub/commit/aefe1c770fe2a13c53f251d2bcbcc8e8e19586a7>

第二版修改为读取后端文件，通过dataset名称匹配后端文件

datahub-frontend 添加路由，托管文件

所有docker-compose-\*.yaml 增加 volumes: /home/tempuser/files:/home/datahub/excels

<https://github.com/imindzzz/datahub/commit/9e661b783666765e09cf150f9bea8c72c7d7c84a>

修改缓存时间为1分钟,映射目录错误

Import Schema Business Description

导入模板，字段对应

图形用户界面, 应用程序

描述已自动生成

<https://github.com/imindzzz/datahub/commit/5e9fe93677c9dcd3df119e09ce3d1ea67ed175e5>

纯前端实现，读取excel后，循环调用编辑接口

其它参考资料

安装jdk <https://www.cnblogs.com/qinshengfei/p/18080460>

安装nvm nvm install 21 use 21配置 source <https://blog.csdn.net/nilm61/article/details/134266633>

安装python <https://cloud.tencent.com/developer/article/1835880> <https://blog.csdn.net/chiele/article/details/136395102>

安装docker 和comopse <https://cloud.tencent.com/developer/article/1701451> yum -y install docker-ce-26.1.2