

医疗智能体

快速入门

文档版本 01
发布日期 2024-06-07



版权所有 © 华为云计算技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为云计算技术有限公司

地址：贵州省贵安新区黔中大道交兴功路华为云数据中心 邮编：550029

网址：<https://www.huaweicloud.com/>

目录

1 登录平台	1
2 购买计算资源（主账号操作）	3
2.1 计算资源说明	3
2.2 资源看板	3
2.3 计算资源	4
2.4 存储资源	5
2.5 性能加速（可选）	5
2.6 数据库（可选）	6
3 创建项目	7
4 上传小于 1GB 数据	8
5 上传大于 1GB 数据	11
6 下载数据	14
7 运行作业	17
7.1 方式 1：使用预置的 NGS 流程	17
7.2 方式 2：使用预置应用搭建 NGS 流程	22
7.3 方式 3：自定义镜像运行 FastQC 流程	25
8 运行大规模虚拟药筛任务	32
8.1 药物数据输入格式说明	32
8.2 订阅 Docking Summary 流程	32
8.3 新建研究	33
8.4 查看药筛结果	34
8.5 查看药筛作业和结果下载	36
9 配置命令行工具	38
9.1 步骤 1：获取认证信息	38
9.2 步骤 2：获取命令行工具	39
9.3 步骤 3：初始化配置	42
10 入门实践	44

1 登录平台

主账号

1. 登录EIHealth**管理控制台**。
2. 选择华为账号登录。

图 1-1 主账号登录



子账号

1. 登录EIHealth**管理控制台**。
2. 选择IAM用户登录。

图 1-2 子账号登录



2 购买计算资源（主账号操作）

2.1 计算资源说明

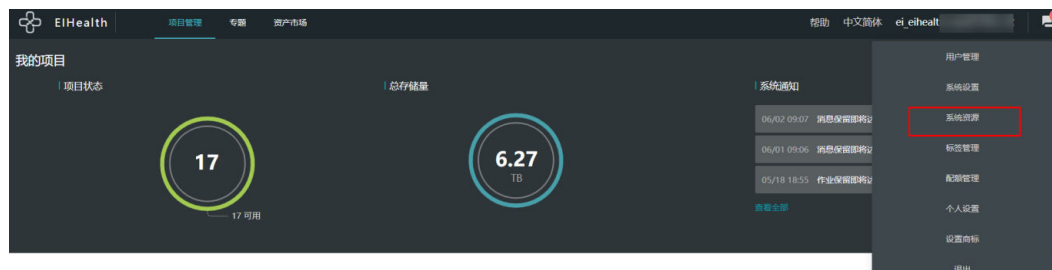
每个作业的运行依托的是后面的计算资源，所以需要管理员账号（主账号）提前购买计算资源。

说明

若平台无计算资源，则投递的作业会出现等待情况，直至购买了计算资源或者任务超时才会停止。

单击平台右上角的账号名称，会出现账号的相关操作，包括系统资源，系统配置等。注意这里只有管理员账号有系统资源，系统配置等操作，具体说明请参见[用户管理](#)。

图 2-1 系统资源



2.2 资源看板

在“资源看板”中，您可以实时监控计算资源、存储资源、性能加速、数据库的使用情况。

图 2-2 资源看板



2.3 计算资源

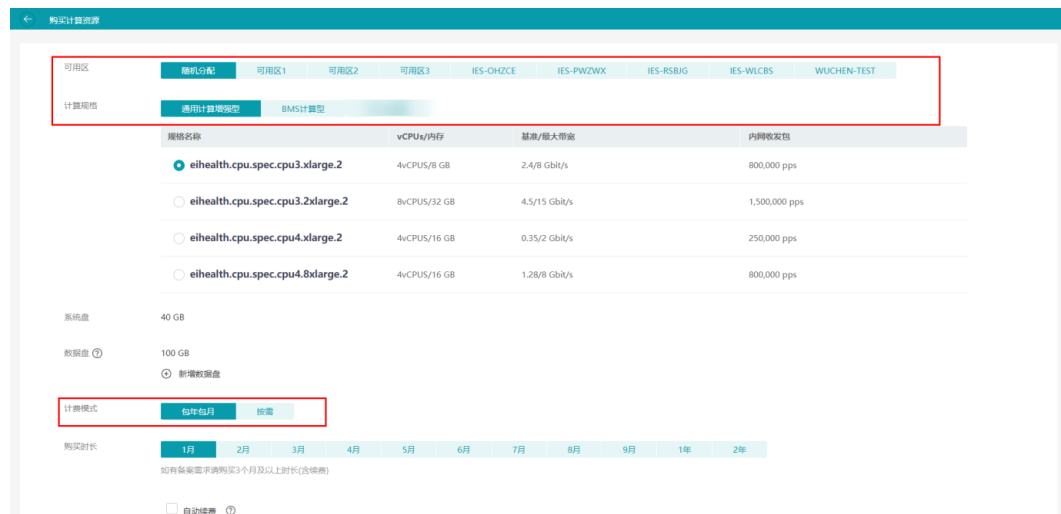
单击计算资源页面，列表中会显示已购买的计算资源，以及是否可用等信息，若新购买的账号则列表为空。

图 2-3 计算资源



单击“购买计算资源”按钮购买计算资源。可以根据业务需要选择不同的资源类型，其中可用区可以选择任意可用区即可。并支持包年包月购买，或者按需购买。

图 2-4 购买计算资源



2.4 存储资源

存储资源的计费模式是按需或者购买套餐包的形式，

- 按需计费可根据数据量的大小收费，故不需要提前进行购买
- 套餐包需提前购买，当存储用量超过套餐包规格时，超出部分将自动按量按需计费

图 2-5 存储资源

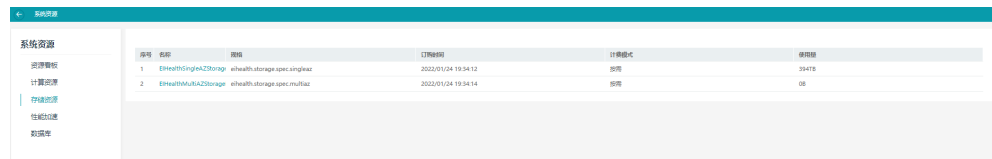
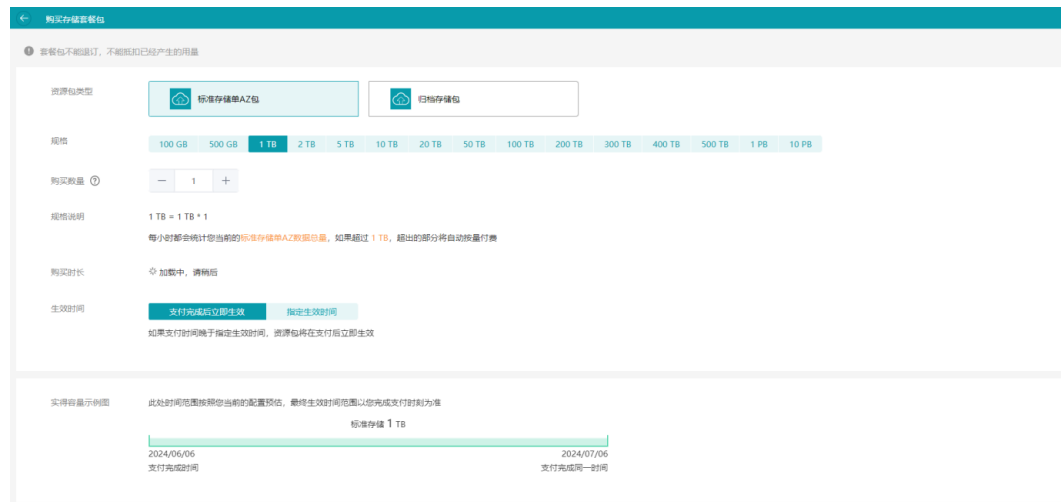


图 2-6 存储套餐包



图 2-7 购买存储套餐包



2.5 性能加速（可选）

当普通的计算资源不满足业务场景时，可以选择购买性能加速资源，加快算法的数据分析速度。

单击“购买性能加速”。选择包年包月或者按需，同时选择购买的性能盘大小。

图 2-8 购买性能加速



2.6 数据库（可选）

数据库对应的是“项目管理>数据库”功能，可按使用需要购买。

图 2-9 购买数据库



3 创建项目

新用户进入平台后需要自己创建项目，单击“创建项目”按钮，填写项目名称以及选择对应的数据保护策略，数据保护策略参考下面配置。

图 3-1 创建项目



单击项目名称可进入到对应的项目中。

图 3-2 进入项目



4 上传小于 1GB 数据

上传数据

上传数据是最常用的添加数据功能，但由于网页传输大文件时存在一定的限制，所以上传数据的大小不能超过1GB。

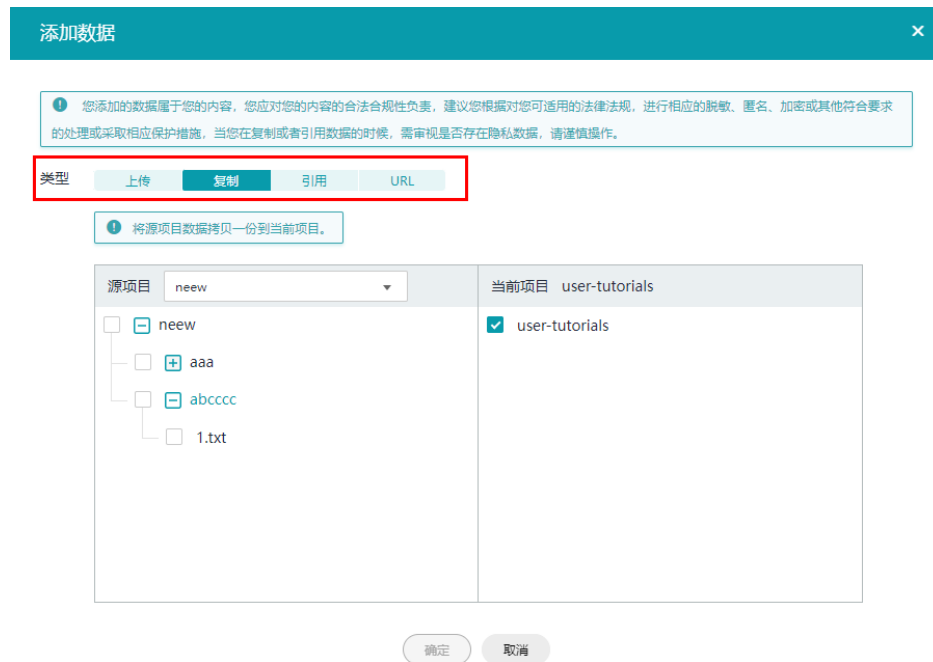
图 4-1 上传数据



复制数据

复制数据会从别的项目中复制一份数据到当前项目中，这里可选择的源项目需是当前账号为**管理员**的对应项目。如需设置项目成员，请参考用户指南中的[项目成员和权限](#)。

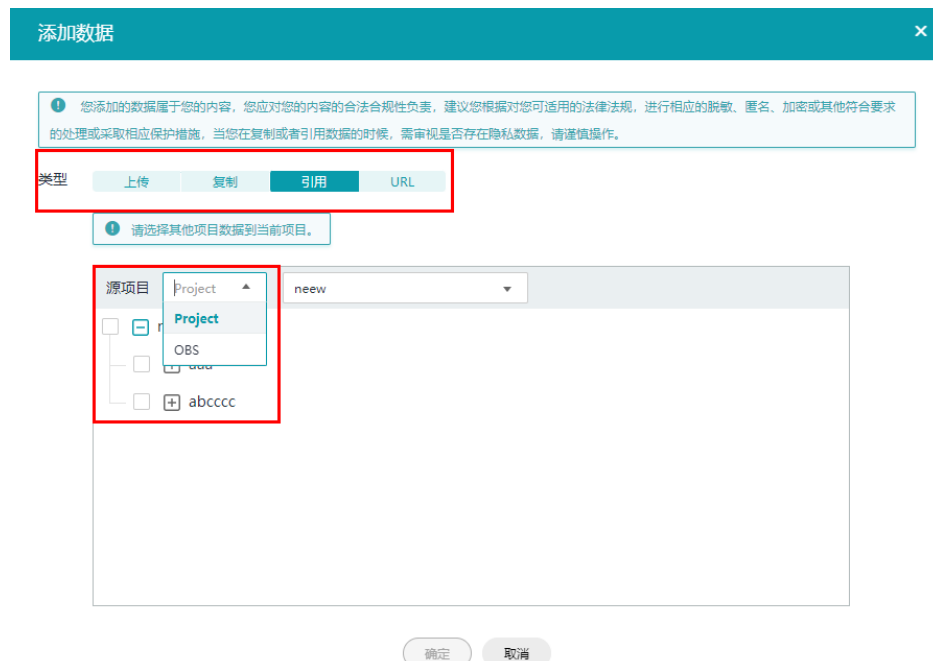
图 4-2 复制数据



引用数据

引用数据可以引用其他项目中或对应的obs桶的数据。引用的数据不能在该项目中进行更改。只能用来运行作业，导入数据到数据库等操作。

图 4-3 引用数据



URL 导入数据

以URL导入的方式添加数据类似于linux中的wget操作。单击“添加URL”按钮可以添加具体的链接，最多可添加10条链接。

图 4-4 URL 导入数据

The image shows a '添加数据' (Add Data) dialog box with a teal header and a close button. It contains two informational messages: one about data ownership and legal compliance, and another about content responsibility. Below these is a '类型' (Type) section with buttons for '上传' (Upload), '复制' (Copy), '引用' (Reference), and 'URL' (highlighted in red). A text field for '源数据URL地址' (Source Data URL Address) is present, with a note that it supports multiple URLs up to 10. A 'No1:' label is next to the input field. A '添加URL' (Add URL) button is highlighted in red below the input field. At the bottom, there is a list of items under '当前项目' (Current Project) with 'user-tutorials' selected. '确定' (Confirm) and '取消' (Cancel) buttons are at the very bottom.

添加数据

1 您添加的数据属于您的内容，您应对您的内容的合法合规性负责，建议您根据对您可适用的法律法规，进行相应的脱敏、匿名、加密或其他符合要求的处理或采取相应保护措施，当您在复制或引用数据的时候，需审视是否存在隐私数据，请谨慎操作。

类型 上传 复制 引用 URL

1 您应对下载数据内容的合法性、准确性、完整性和可靠性负责。您全权负责确保和维护与您内容中的任何机密信息或数据相关的任何所需通知、同意或授权。

源数据URL地址：支持多个URL同时导入，最多可添加10条URL

No1:

添加URL

当前项目 user-tutorials

- + user-tutorials

确定 取消

5 上传大于 1GB 数据

当数据过大，或者数据处于服务器上时，页面传输并不是很方便，所以可以利用 EIHealth 平台的命令行工具进行数据文件、文件夹传输。命令行配置方法请参见[配置命令行工具](#)。

上传数据

使用命令行工具 **upload** 命令，将本地数据上传到 EIHealth 平台中。该命令不支持将数据上传到引用目录。

最小可以上传 **0Byte** 的空文件或文件夹，最大可以上传 **48.8TB** 的单个文件。

数据在上传的过程中，受网络影响可能出现损坏，上传命令默认会在上传完成后，验证项目中数据的 MD5 值与本地数据的 MD5 值的一致性，以及验证项目中数据的大小与本地数据大小一致性。

- 命令结构
health upload <srcdir> <destdir> [flags]

表 5-1 参数说明

参数	简写	是否必选	说明
srcdir	无	是	源路径。
destdir	无	是	目的路径。
--rename	-e	否	重命名，上传文件时可选。
--recursive	-r	否	递归上传文件夹中所有的文件和子文件夹，上传文件夹时必选。
--force	-f	否	强制操作，不进行询问提示，上传文件夹时可选。
--flat	-l	否	上传文件夹时，只上传该文件夹下的所有内容，上传文件夹时可选。

参数	简写	是否必选	说明
--update	-u	否	增量上传操作，设置该参数后上传每个文件时会比对平台数据路径中的文件，仅在以下情况时上传数据： <ul style="list-style-type: none"> 文件不存在。 待上传文件大小与平台文件大小不一致。 文件的最后修改时间不一致。

● 命令示例

本节以Windows为例介绍eihealth-toolkit的使用过程，Linux和macOS环境使用方法基本相同，可参考。

- a. 上传数据前需要使用switch命令进入待操作的项目。例如，使用health switch project test命令进入到名为test的项目中。



```
C:\Users>health switch project test
switch to project test successfully!

C:\Users>
```

- b. 将本地数据上传到test项目中的src文件夹中。

详细的数据操作命令，如下载、删除、拷贝、切换路径等请参见“[命令行工具 > 数据管理命令](#)”章节“[命令行工具 > 数据管理命令](#)”章节。

- 将本地D:\local\data\test.txt路径下的txt文件上传至src文件夹中。

上传完成后，会在EIHealthG42 Health AI平台数据管理页面的src文件中，生成test.txt文件。

```
health upload D:\local\data\test.txt /src/
```

- 上传文件并重命名为abc.txt。

```
health upload D:\local\data\test.txt /src/ --rename abc.txt
```

- 递归上传本地D:\local路径下的所有文件和文件夹至项目src文件夹中（不上传local文件夹本身）。

```
health upload D:\local /src/ -r -f -l
```

- 上传本地D:\local文件夹至项目src文件夹中（上传local文件夹本身）。

```
health upload D:\local /src/ -r -f
```

- 增量上传文件，同步本地D:\local\test.txt文件至项目src文件夹中。

```
health upload D:\local\test.txt /src/ --update -r
```

说明

- 列举路径中的对象时，需使用/xxx/格式，如示例中的/src/。
- 列举本地路径中的文件夹对象时，需要使用“路径/文件名”或“路径\文件名”格式。请依据操作系统的路径规范使用，如示例中的D:\local。
- 如果路径中带有特殊字符比如（）之类的，运行的时候需要将整个路径用""括起来，或者用\将字符转义，例如：health cp "/src2/abs (1).txt" /src1/ 或者 health cp /src2/abs\(1\).txt /src1/ 。

6 下载数据

在 EIHealth 平台下载数据

须知

下载数据操作会产生流量费用，计费方式为按需计费，计费详情请参考[OBS数据下载费用](#)。

以下操作步骤是在“数据”页面下载数据至本地。您也可以使用命令行工具实现数据的下载。

1. 在EIHealth平台“项目 > 数据”页面，展开数据文件夹，选择待下载的数据。
2. 单击“操作”列“下载”。
3. 单击鼠标右键，选择“链接另存为”，下载数据。

图 6-1 下载数据



使用命令行工具下载数据

使用**download**命令将EIHealth平台的数据下载到本地，此命令不支持下载引用项目中的数据。

数据在下载的过程中，受网络影响可能出现损坏，下载命令默认会在下载完成后，验证项目中数据的MD5值与本地数据的MD5值的一致性，以及验证项目中数据的大小与本地数据大小一致性。

- 命令结构

```
health download <srcdir> <destdir> [flags]
```

表 6-1 参数说明

参数	简写	是否必选	描述
srcdir	无	是	源路径。
destdir	无	是	目的路径。
--rename	-e	否	重命名。
--recursive	-r	否	按指定的对象名前缀批量下载，批量下载时必须选。
--force	-f	否	强制操作，不进行询问提示，批量下载时可选。
--flat	-l	否	批量下载时，不包含上一级父对象名前缀，批量下载时可选。
--update	-u	否	增量下载操作，设置该参数后，下载每个文件时会比对本地的文件，仅在以下情况时下载数据： <ul style="list-style-type: none"> 文件不存在。 下载文件大小与本地文件大小不一致。 文件的最后修改时间不一致。

- 命令示例

本节以Windows为例介绍eihealth-toolkit的使用过程，Linux和macOS环境使用方法基本相同，可参考。

- 下载数据前需要使用**switch**命令进入待操作的项目。例如，使用**health switch project test**命令进入到名为test的项目中。
- 下载项目中的数据至本地。
 - 下载项目中src/abc.txt文件到本地的data路径中。
health download /src/abc.txt D:\local\data
 - 下载项目中的abc.txt文件到本地，重命名为abc1.txt。
health download /src/abc.txt D:\local\data --rename abc1.txt
 - 递归下载项目中src文件夹中的所有文件和文件夹（包含src文件夹本身）至本地data路径，且下载过程中不进行询问提示。
health download /src/ D:\local\data -r -f

- 递归下载项目中src文件夹中的所有文件和文件夹（不包含src文件夹本身）至本地data路径，且下载过程中不进行询问提示。
`health download /src/ D:\local\data -r -f --flat`
- 增量下载文件。同步项目中src文件夹下所有的文件和文件夹至本地data路径中。
`health download /src/ D:\local\data --update -r`

说明

列举路径中的对象时，需使用/xxx/格式，如示例中的/src/。

如果路径中带有特殊字符比如（ ）之类的，运行的时候需要将整个路径用""括起来，或者用\将字符转义，例如：`health cp "/src2/abs (1) .txt" /src1/` 或者 `health cp /src2/abs\(1)\.txt" /src1/`。

7 运行作业

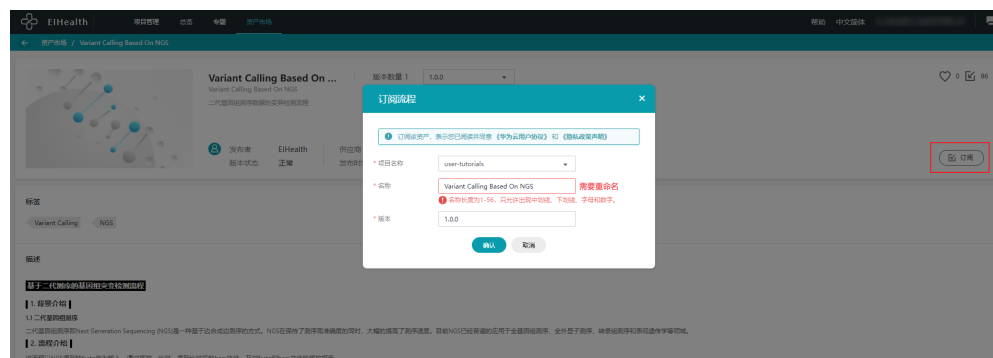
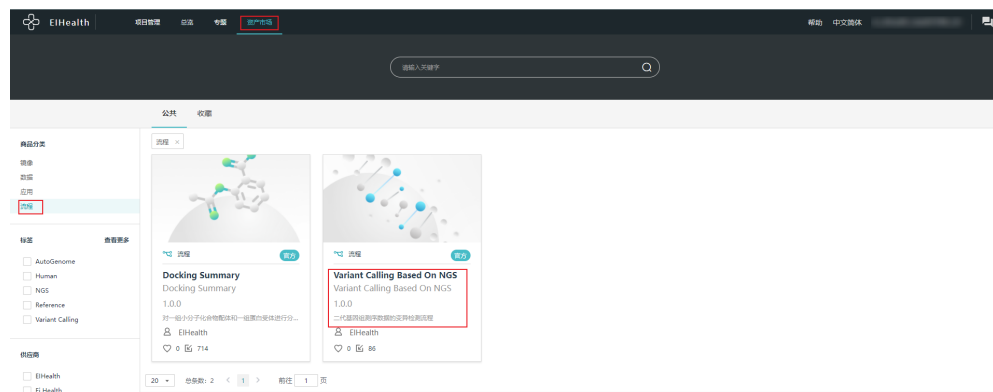
7.1 方式 1：使用预置的 NGS 流程

使用EIHealth平台预置的流程进行运行作业。

步骤 1：订阅流程

进入资产市场订阅已有的流程，以二代基因测序数据的变异检测流程为例。

图 7-1 订阅流程



可以在“工具 > 流程”列表可以看到刚刚订阅的流程。

图 7-2 查看订阅的流程

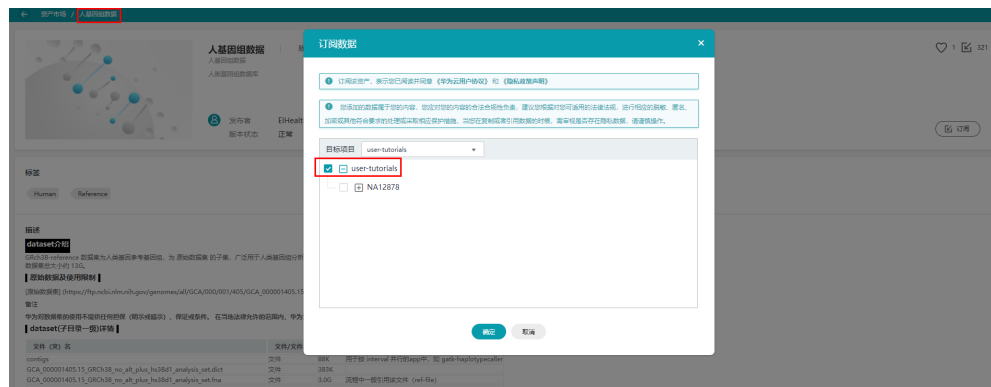
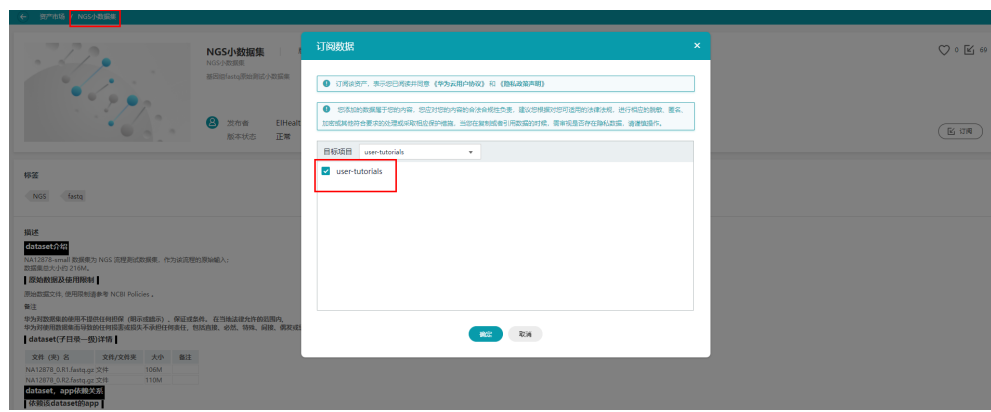


步骤 2: 订阅数据

若您本地有需要分析的二代基因组数据，则您可以用本地的数据。数据上传方法请参见[上传数据](#)。

若没有，可以先订阅资产市场里的示例数据进行分析，这里先用资产市场中的“人类基因组数据”和“NGS小数据集”进行分析。

图 7-3 订阅数据



可以在“数据”列表可以看到刚刚订阅的流程。

图 7-4 查看订阅的数据



步骤 3: 启动作业

1. 在流程页面，单击“启动作业”按钮，根据实际情况填写作业名称等信息。

图 7-5 启动作业

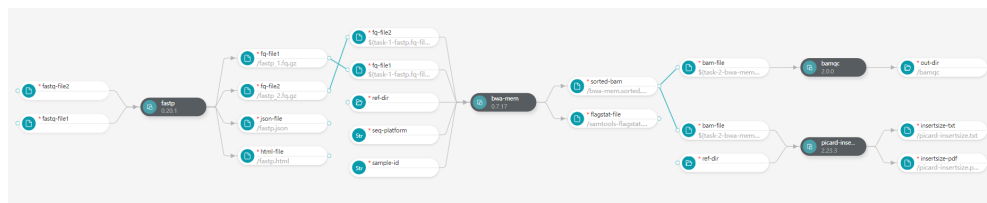


图 7-6 设置作业名称



2. 作业设置完成后，会显示流程中的各个应用信息，再此可以指定实际的运行参数。

图 7-7 NGS 流程



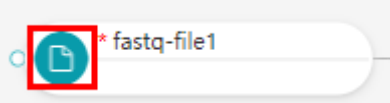
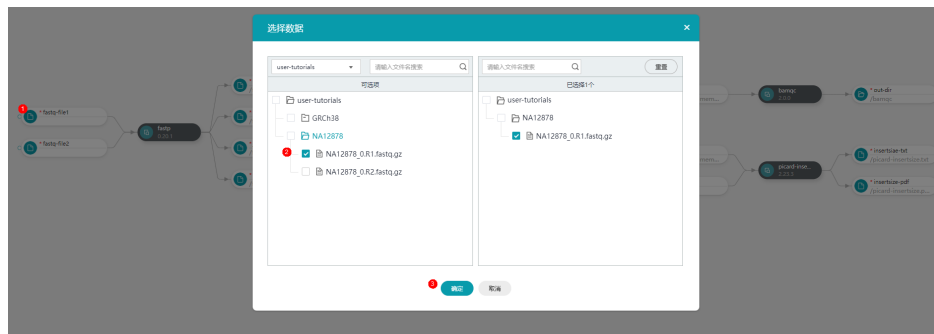
3. 单击  图标，可设置数据来源，可以指定File或者Directory类型参数。这里我们将NA12878目录下的两个fastq文件分别指定成fastq-file1和fastq-file2。

图 7-8 设置 fastp 输入数据



4. 将GRCh38目录指定成bwa-mem与picard-insertsize的输入。

图 7-9 设置 bwa-mem 输入数据

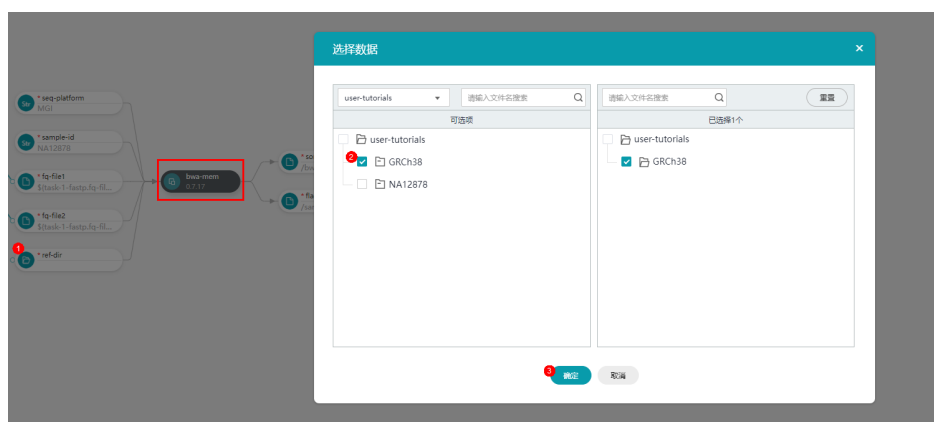
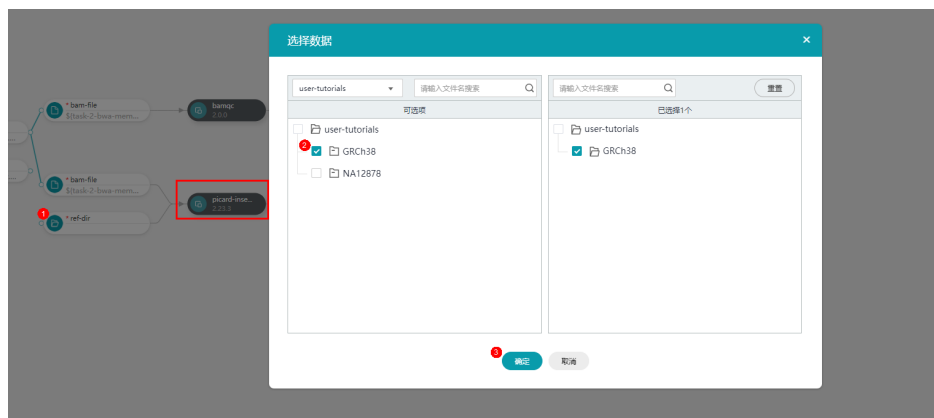
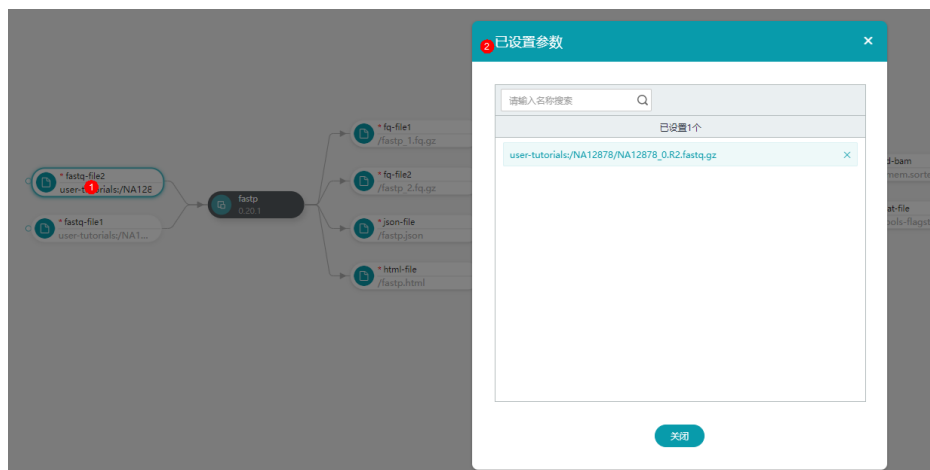


图 7-10 设置 picard-insertsize 输入数据



若是想查看已经设置好的参数，可以按照下图步骤进行查看，同时若发现指定的参数有问题，也可再此处进行删除参数。

图 7-11 删除参数



5. 配置好参数后可以单击右上角“启动作业”按钮进行启动作业，同时会跳转回作业列表。

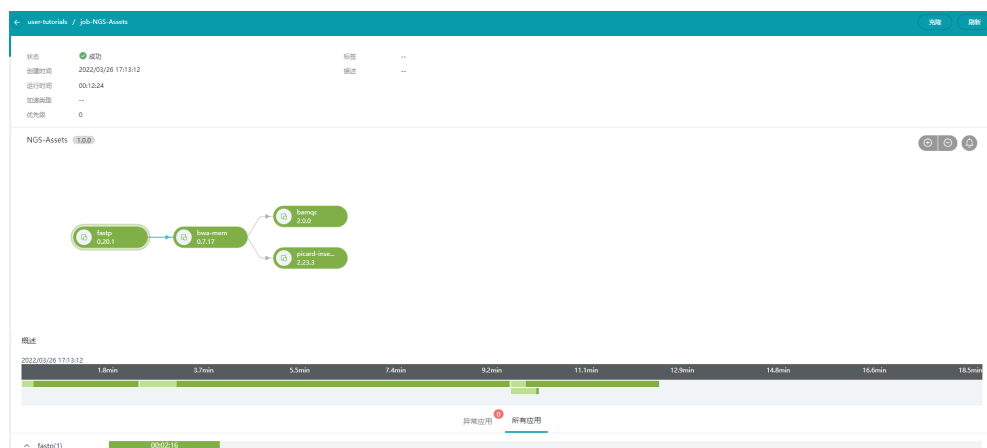
图 7-12 作业列表



步骤 4：查看作业运行状态、获取作业结果

在作业列表中，单击对应的作业名称，可查看作业运行状态，运行时间等信息。

图 7-13 作业运行状态



单击具体的每个应用，可以查看每个应用的运行信息：输入输出，节点参数，应用以及日志等信息。

单击具体的输出参数便可跳转到具体的文件位置。

图 7-14 查看应用信息

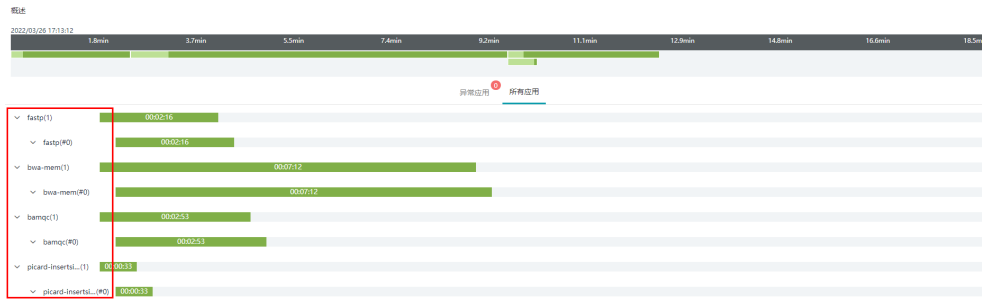


图 7-15 查看输入、输出数据位置



单击输出参数路径，可以跳转至输出文件位置。

图 7-16 获取输出数据



7.2 方式 2：使用预置应用搭建 NGS 流程

将EIHealth平台预置的应用构建成流程，并运行作业。以二代基因组分析流程：fastp, bwa-mem, bamqc, picard-insertsize两个算法为例。

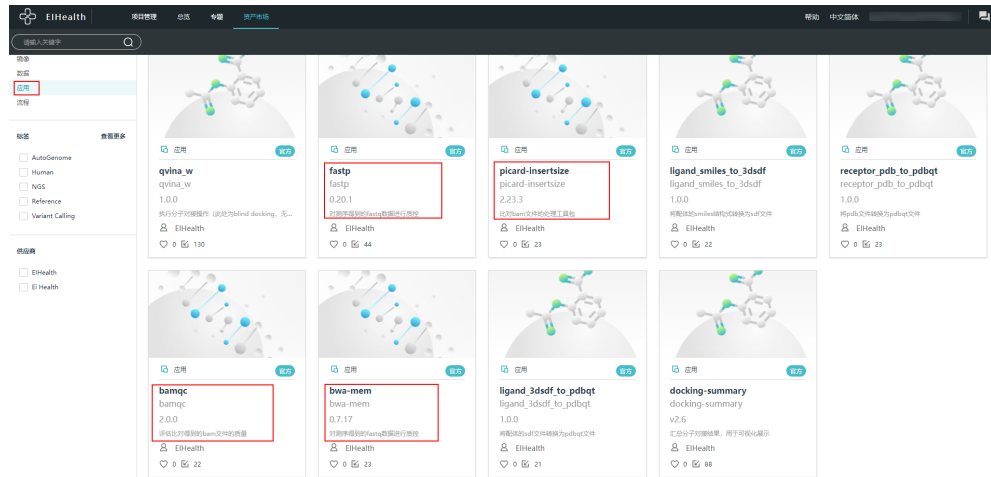
应用是对每个软件的镜像封装，将应用封装好后可以反复利用并也可以让其他人很容易的使用，不用担心复杂的开发环境问题。

流程是由1各或者N个应用串联构建而成，平台页面以简单拖拽、连接的方式将不同的应用连接起来构建成一个分析流程，避免了写繁杂的bash脚本。

步骤 1：订阅应用

进入资产市场，分别订阅fastp, bwa-mem, bamqc, picard-insertsize应用。

图 7-17 订阅应用



可以在“工具 > 应用”列表可以看到刚刚订阅的应用。

图 7-18 查看订阅的应用



步骤 2: 新建流程

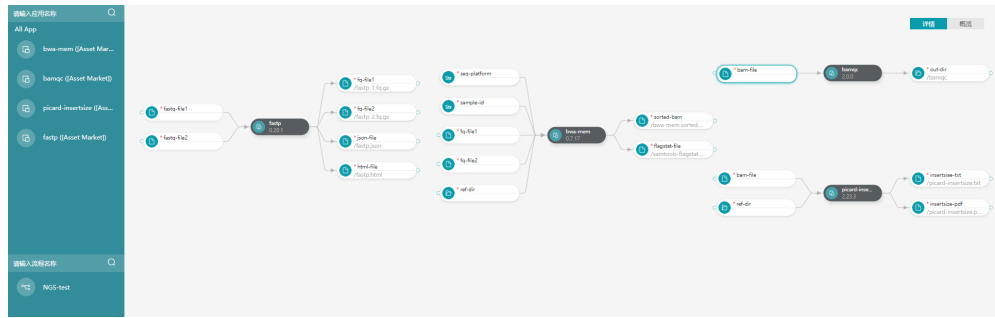
1. 在“工具 > 流程”页面，单击“新建流程”按钮，填写流程名称，版本等信息。

图 7-19 新建流程



2. 利用拖拽的方式将左侧边栏应用列表中的需要执行的应用拖拽到中间画布上，这里我们将fastp, bwa-mam, bamqc, picard-insertsize算法拖拽到中间画布上。

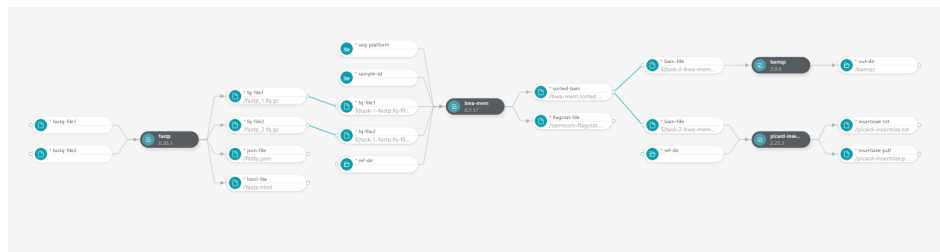
图 7-20 拖拽应用



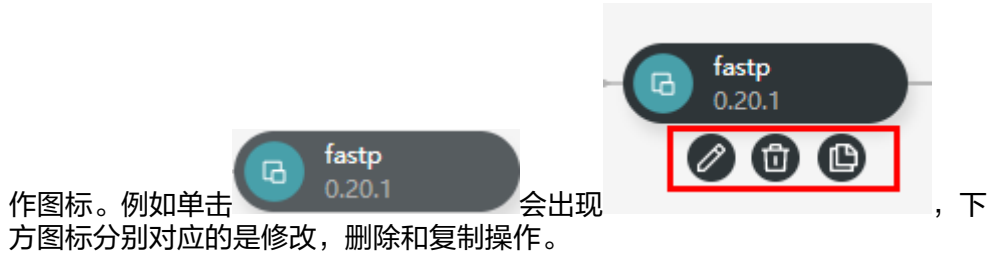
3. 设置输入、输出关系。

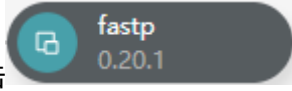

fastp的两个输出参数fq-file1和fq-file2是bwa-mem算法的输入参数，bwa-mem的sorted-bam参数是bamqc与picard-insertsize输入参数，参考图进行连接。

图 7-21 设置输入、输出关系



4. 若想修改应用的一些参数或删除某些应用，则可以点击每个应用会出现对应的操作图标。



例如单击  会出现  ，下方图标分别对应的是修改，删除和复制操作。

5. 构建好流程后单击右上角  图标保存，然后关闭当前页面。

步骤 3：新建作业

- 1. 单击作业页面的“新建作业”按钮，并选择需要运行的流程。这里作业名称可以用默认的名称，也可自己更改一个名称。同时，若想比较好的方式管理输出文件，也可以自行指定输出路径。

图 7-22 新建作业



2. 在作业页面添加好运行参数，添加方式与方式1相同。
3. 配置好参数后，单击右上角“启动作业”按钮启动作业，同时会跳转回作业列表。

图 7-23 作业列表



步骤 4：查看作业运行状态、获取作业结果

查看方法与方式1相同。

7.3 方式 3：自定义镜像运行 FastQC 流程

应用是运行作业的最小单位，每个应用依托于一个镜像进行创建。

方式2使用直接制作好的应用。用户也可以自己制作镜像，并基于镜像创建应用。

本示例中制作FastQC镜像，并基于镜像创建应用，运行分析作业。

镜像简介

由于生物信息学软件，往往由于不同的操作系统（Windows、Linux、Mac等）原因，无法实现统一的运维管理。同时，这些软件具有不同的版本和软件包，安装、使用过程复杂。

将生物信息学软件封装成Docker镜像，可以使程序在不同的环境中运行，并通过EIHealth平台的镜像管理，实现高效的调用，极大方便了软件的安装和运行。

Docker镜像是一个模板，是容器应用打包的标准格式，在部署容器化应用时可以指定镜像。例如一个Docker镜像可以包含一个完整的Ubuntu操作系统环境，里面仅安装了用户需要的应用程序及其依赖文件。

步骤 1：新建 fastqc 镜像

1. 在linux环境下配置docker环境。

```
curl -fsSL get.docker.com -o get-docker.sh
sh get-docker.sh
```

- 配置完成后，运行docker命令，可查看docker的信息。

图 7-24 docker 信息

```
(base) ~$ docker
Usage: docker [OPTIONS] COMMAND

A self-sufficient runtime for containers

Options:
  -c, --config string      Location of client config files (default "/home/wangbing/.docker")
  -D, --debug              Enable debug mode
  -H, --host list          Daemon socket(s) to connect to
  -l, --log-level string   Set the logging level ("debug"|"info"|"warn"|"error"|"fatal") (default "info")
  --tls                    Use TLS; implied by --tlsverify
  --tlscacert string       Trust certs signed only by this CA (default "/home/wangbing/.docker/ca.pem")
  --tlscert string         Path to TLS certificate file (default "/home/wangbing/.docker/cert.pem")
  --tlskey string          Path to TLS key file (default "/home/wangbing/.docker/key.pem")
  --tlsverify              Use TLS and verify the remote
  -v, --version            Print version information and quit

Management Commands:
  builder      Manage builds
  config       Manage Docker configs
  container    Manage containers
  engine       Manage the docker engine
  image        Manage images
  network      Manage networks
  node         Manage Swarm nodes
  plugin       Manage plugins
  secret       Manage Docker secrets
  service      Manage services
  stack        Manage Docker stacks
  swarm       Manage Swarm
```

- 构建fastqc镜像。
 - 可以在dockerhub官网上进行搜索，选择出对应的名称与版本。
 - 可以使用docker search fastqc命令进行搜索镜像。

```
(base) ~$ docker search fastqc
NAME                DESCRIPTION                               STARS     OFFICIAL    AUTOMATED
biocontainers/fastqc  fastqc                                     4         [OK]
fjukstad/fastqc     fastqc                                     1         [OK]
dcoy/fastqc         fastqc                                     1         [OK]
maxulyse/fastqc     Archived image for Sarek                  1         [OK]
pegi3s/fastqc       FastQC (https://www.bioinformatics.babrah... 1
staphb/fastqc       FastQC: A quality control analysis tool for... 0
davelabhub/fastqc   Mod of biocontainers version which doesn't s... 0
clinicalgenomics/fastqc  http://www.bioinformatics.babraham.ac.uk/pro... 0 [OK]
biowandrobe2/fastqc  jkkrc/fastqc                               0
nanozoo/fastqc       FROM: https://github.com/s-andrews/FastQC     0
4ndc1c/fastqc        stjudecloud/fastqc                         0
intellisqngs/fastqc  ummidock/fastqc                            0 [OK]
singlecellpipeline/fastqc  fastqdocker/fastqdocker                    0 [OK]
czbiohub/fastqc      Concatenate fastq files                     0
isaac1iao/fastqc     wgsipeline/fastqc                           0
choclab/fastqc-tiyo  umichbfxcore/fastqc                         0
sbrlab/fastqc        sbrlab/fastqc                               0
suplh/fastqc         suplh/fastqc                                0
```

- 若dockerhub上没有想要的镜像，可以利用基础的操作系统镜像进行构造，构造方式请参见[用户指南](#)。

- 选择和获取镜像。
搜索到的镜像列表选择一个fastqc镜像进行pull，这里我们选择pegi3s/fastqc。

```
(base) ~$ docker search fastqc
NAME                DESCRIPTION                               STARS     OFFICIAL    AUTOMATED
biocontainers/fastqc  fastqc                                     4         [OK]
fjukstad/fastqc     fastqc                                     1         [OK]
dcoy/fastqc         fastqc                                     1         [OK]
pegi3s/fastqc       FastQC (https://www.bioinformatics.babrah... 1 [OK]
maxulyse/fastqc     Archived image for Sarek                  1         [OK]
staphb/fastqc       FastQC: A quality control analysis tool for... 0
davelabhub/fastqc   Mod of biocontainers version which doesn't s... 0
clinicalgenomics/fastqc  http://www.bioinformatics.babraham.ac.uk/pro... 0 [OK]
biowandrobe2/fastqc  jkkrc/fastqc                               0
nanozoo/fastqc       FROM: https://github.com/s-andrews/FastQC     0
4ndc1c/fastqc        stjudecloud/fastqc                         0 [OK]
intellisqngs/fastqc  ummidock/fastqc                            0 [OK]
singlecellpipeline/fastqc  fastqdocker/fastqdocker                    0 [OK]
czbiohub/fastqc      Concatenate fastq files                     0
isaac1iao/fastqc     wgsipeline/fastqc                           0
choclab/fastqc-tiyo  umichbfxcore/fastqc                         0
sbrlab/fastqc        sbrlab/fastqc                               0
suplh/fastqc         suplh/fastqc                                0
```

- pull镜像。
若该镜像没有latest版本则需进入dockerhub官网进行查询。

```
(base) ~$ docker pull pegi3s/fastqc:latest
```

- 查看镜像是否pull成功。
用docker images命令查看镜像是否pull成功。如果您本地已经有很多镜像，防止查询出全部镜像，可在docker images命令后面添加 | grep pegi3s使查询更精准。

```
(base) ❯ docker images |grep pegi3s
pegi3s/fastqc
6d64070538e7 2 years ago 578MB latest
```

7. 查询fastqc软件命令。非常熟悉fastqc软件则可跳过此步骤。

复制上一步查找到的fastqc的id。利用**docker run imagesId fastqc --help**命令进行查询。**fastqc --help**可以替换成任何fastqc的命令，与linux类似。

```
(base) ❯ docker run 6d64070538e7 fastqc --help
FastQC - A high throughput sequence QC analysis tool

SYNOPSIS

fastqc seqfile1 seqfile2 .. seqfileN

fastqc [-o output dir] [--(no)extract] [-f fastq|bam|sam]
[-c contaminant file] seqfile1 .. seqfileN

DESCRIPTION

FastQC reads a set of sequence files and produces from each one a quality control report consisting of a number of different modules, each one of which will help to identify a different potential type of problem in your data.

If no files to process are specified on the command line then the program will start as an interactive graphical application. If files are provided on the command line then the program will run with no user interaction required. In this mode it is suitable for inclusion into a standardised analysis pipeline.

The options for the program as as follows:

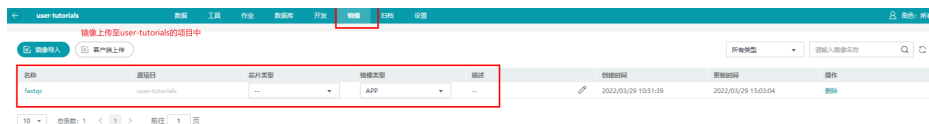
-h --help          Print this help file and exit
-v --version       Print the version of the program and exit
-o --outdir        Create all output files in the specified output directory.
```

步骤 2: 上传镜像至 EIHealth 平台

1. **配置命令行工具。**
若已完成配置，可跳过此步骤。
2. 在命令行工具所在的目录，使用switch命令进入EIHealth平台的项目中。
命令结构
health switch project <project-name>
命令示例，例如进入到名为user-tutorials的项目中
health switch project user-tutorials
3. 使用docker tag命令给要上传的镜像打标签
user-tutorial是项目名称
health docker tag pegi3s/fastqc:latest user-tutorials/fastqc:latest
4. 使用docker push命令将镜像上传至EIHealth平台。
-t APP 可不加，在平台上也可以对其进行设置
health docker push user-tutorials/fastqc:latest -t APP

```
(base) ❯ ./health docker tag pegi3s/fastqc:latest user-tutorials/fastqc:latest
(base) ❯ ./health docker push user-tutorials/fastqc:latest -t APP
The push refers to repository [user-tutorials/fastqc]
579602e3838e: Pushed
85d5b4a98733: Pushed
996d66246c8c: Pushed
0821f921bbf7: Pushed
311180c0cfcf: Pushed
8ac4be6e0e4b: Pushed
363295e4406a: Pushed
873cbe4bd21e: Pushed
latest: digest: sha256:56e96bf9cae61cd6dbab931c8e69e1340ef9abefbf4bddaf5d7b1a9abb9946eb size: 2001
```

上传成功后，可以转至平台查看已经上传的镜像。



步骤 3: 创建 fastqc 应用

上一步骤中，已经将fastqc镜像上传至平台，本步骤使用fastqc镜像创建应用。

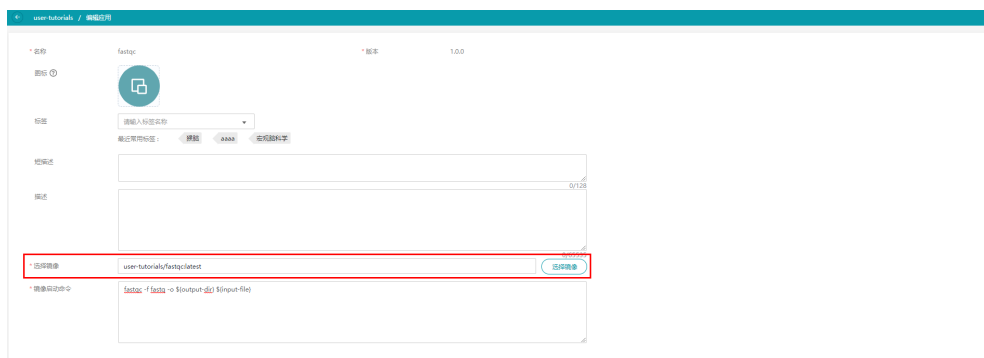
1. 进入“工具 > 应用”界面。

图 7-25 进入应用界面



2. 单击“新建应用”，创建fastqc应用。

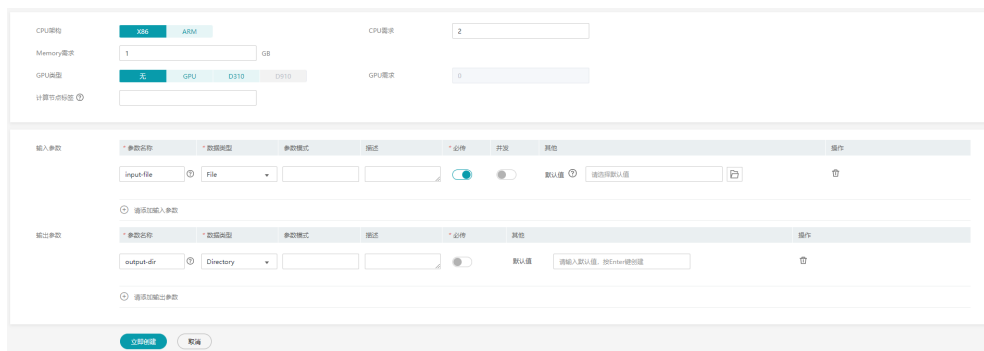
图 7-26 设置镜像名称、版本等信息



镜像启动命令如下所示。镜像启动命令需要引用输入、输出参数中的变量，并以大括号{}扩起，以\$符号进行引用。

```
fastqc -f fastq -o ${output-dir} ${input-file}
```

图 7-27 设置 CPU、内存、输入输出参数



3. 单击“立即创建”，完成fastqc应用的创建。

创建完成后的应用，将显示在应用列表中，您可以使用该应用创建分析作业。

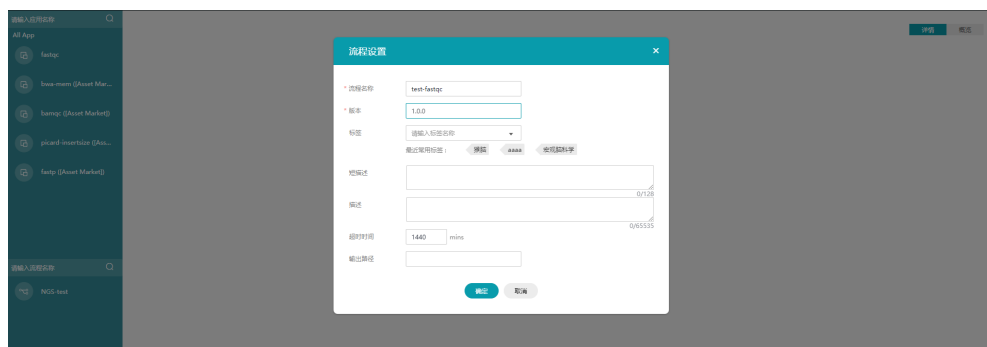
图 7-28 查看应用



步骤 4：搭建流程、运行作业

1. 进入“工具 > 流程”页面，点击“新建流程”进行新建流程，填写好流程名称，流程版本等信息。

图 7-29 新建流程



2. 将fastqc应用拖拽至画布中，并添加输入参数，操作方法请参见[方式2](#)。
3. 参数填写完成后，单击“保存”。并单击“新建作业”创建fastqc作业。

图 7-30 新建作业

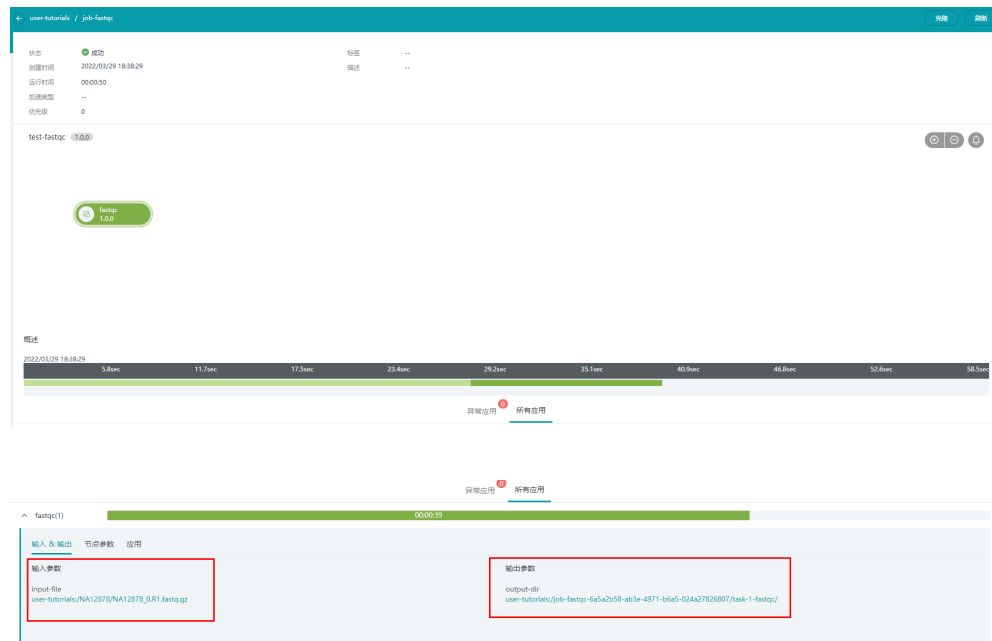


4. 启动作业。

步骤 5：获取分析结果

在“作业”列表中，可单击作业名称查看作业信息和输入输出数据。

图 7-31 查看作业信息



单击输出路径，可跳转至输出文件位置。html文件支持在线预览，zip文件可下载至本地查看。

图 7-32 输出数据



图 7-33 预览 html 文件

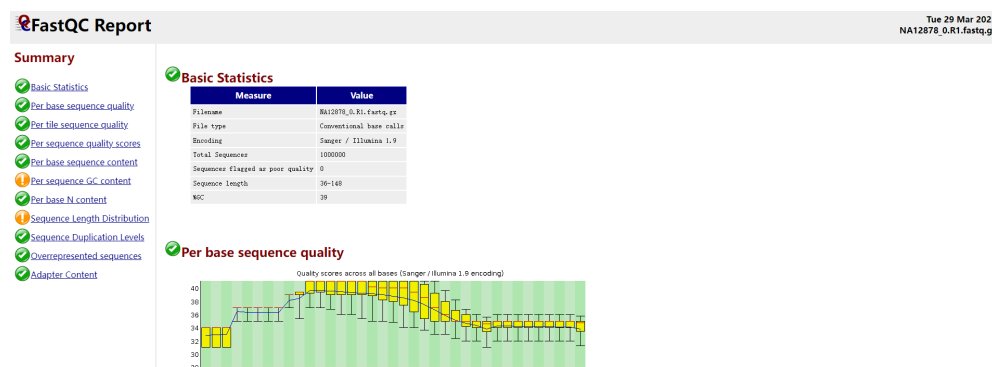


图 7-34 获取 zip 文件



8 运行大规模虚拟药筛任务

8.1 药物数据输入格式说明

药物虚拟筛选平台的输入数据，需要输入蛋白质和配体小分子数据。参考以下数据要求，将数据上传至项目中。

- 输入蛋白质，pdb格式，将要进行对接的所有靶标蛋白放在一个文件夹内。
- 输入配体小分子，配体小分子为如下格式的txt文件。
第一列为smiles字符串，第二列为smiles名称，中间用tab键分割。

ligand.txt - 记事本

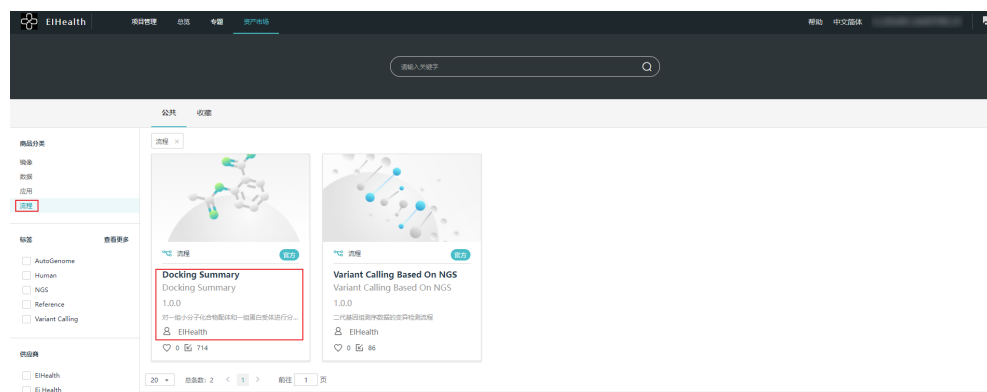
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

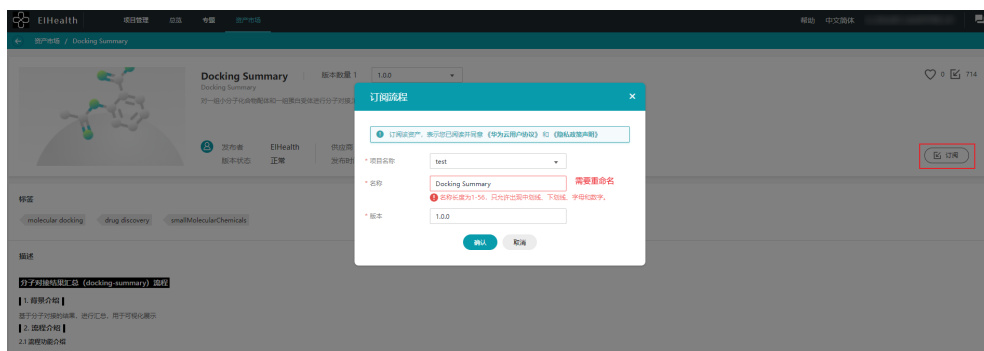
```
CCCCC smiles_name1
c1cccc1 smiles_name2
```

8.2 订阅 Docking Summary 流程

进入资产市场订阅Docking Summary流程。

图 8-1 订阅流程





可以在“工具 > 流程”列表可以看到刚刚订阅的流程。

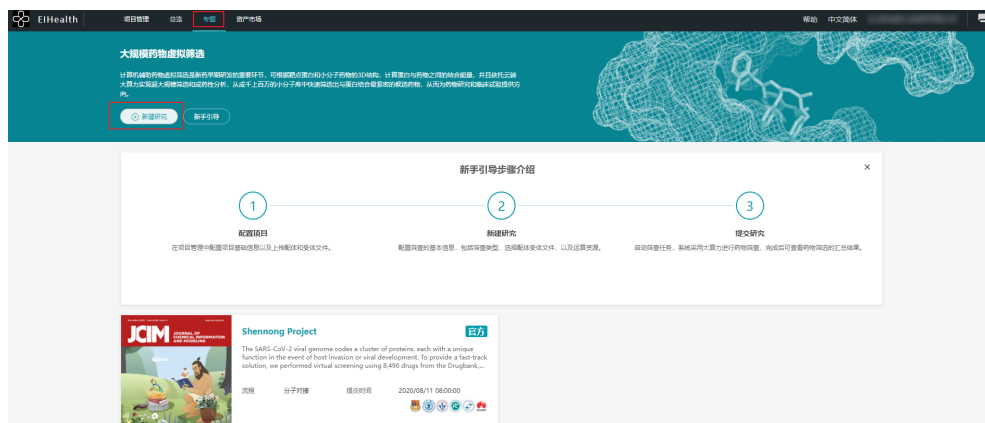
图 8-2 查看订阅的流程



8.3 新建研究

1. 进入“专题”页面，单击“新建研究”。

图 8-3 新建研究



2. 参考表8-1，设置运行信息。

表 8-1 参数说明

参数	说明
选择项目	选择创建好，并带有数据的项目。
研究名称	可自定义研究名称。
流程	选择资产市场中订阅的Docking Summary流程。
配体分子	选择上传的配体小分子文件。

参数	说明
受体蛋白	选择上传的受体蛋白文件夹。这里会默认用文件夹里面所有的蛋白质和配体小分子文件里面的小分子进行一一对接。
超时时间	根据受体配体对个数进行调整，一个受体配体对对接大约需要25s。

图 8-4 运行信息

3. 单击“提交”，运行作业。

8.4 查看药筛结果

药筛运行状态可在“专题”页面查看。

图 8-5 查看运行状态



运行完成后可单击任务名称，查看对接结合能的热图。

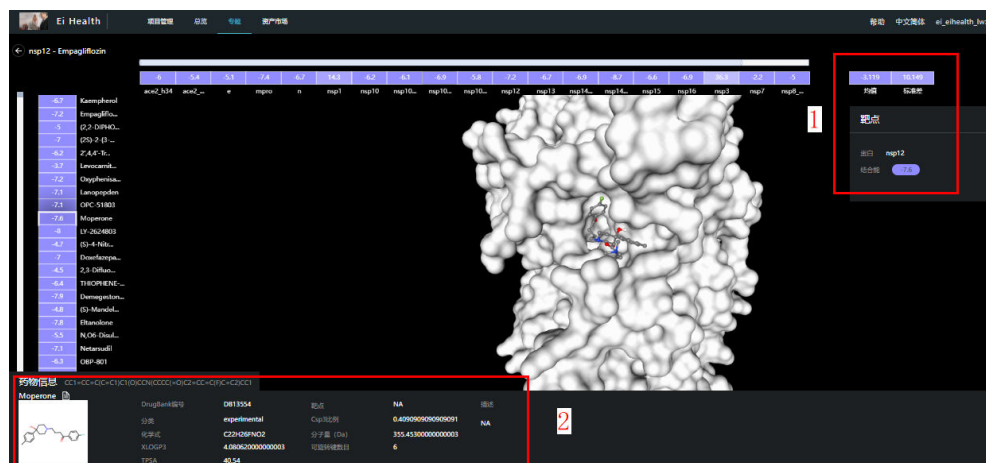
热图的横坐标为蛋白质名称，以及对接结合能的均值和标准差。纵坐标为配体小分子名称，数值为结合能的大小，结合能越小，颜色越偏紫色。

可以根据结合能的大小进行排序，结合能越小代表配体和受体结合越稳定。



单击结合能数值，可查看详细对接结果。

- 第1部分描述了蛋白质与配体小分子的结合能大小，以及该蛋白质在所有配体小分子的对接结果的均值和标准差。
- 第2部分描述了配体小分子信息，注释结果如下。
 - 分类：该小分子的在DrugBank中分类信息。
 - 化学式：小分子的化学式。
 - XLOGP3：小分子的脂溶性logP。
 - TPSA：小分子的拓扑极性表面积。
 - 靶点：小分子的靶点。
 - Csp3比例：小分子的碳饱和度。
 - 分子量 (Da)：小分子的分子量。
 - 可旋转键数目：小分子的可选择键数。

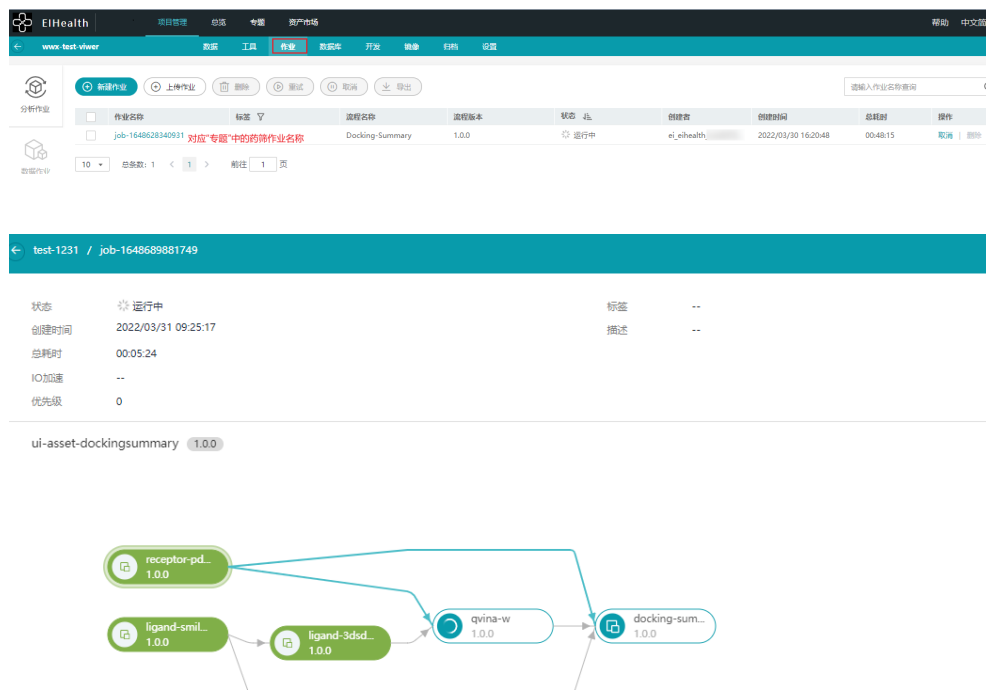


8.5 查看药筛作业和结果下载

查看作业进程

用户可以在“项目 > 作业”中查看药筛作业的后台进程。

图 8-6 查看作业进程

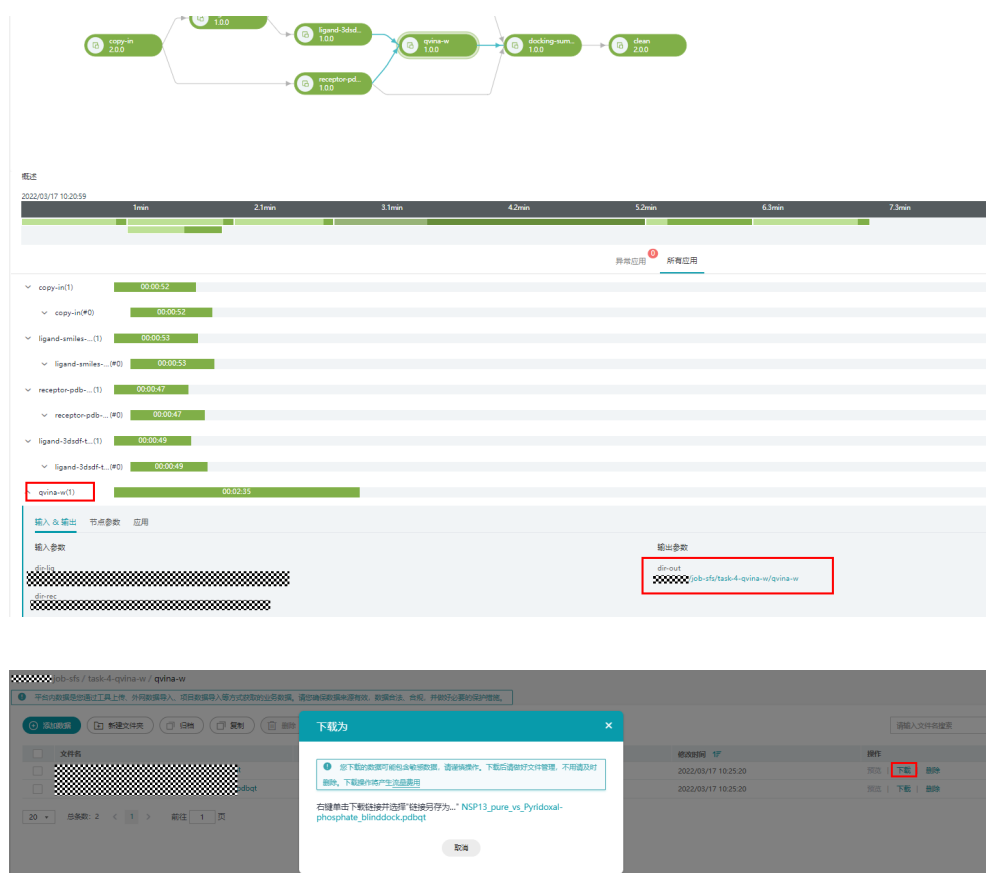


对接完成后的小分子构象下载

单击应用中的“qvina-w”，获取输出结果的路径。

文件的命名方式为：蛋白质名称_vs_配体名称_blinddock.pdbqt。

图 8-7 下载小分子构想数据



9 配置命令行工具

9.1 步骤 1：获取认证信息

获取 AK/SK

AK/SK (Access Key ID/Secret Access Key) 即访问密钥，包含访问密钥ID (AK) 和秘密访问密钥 (SK) 两部分，华为云通过AK识别用户的身份，通过SK对请求数据进行签名验证，用于确保请求的机密性、完整性和请求者身份的正确性。

1. 登录[华为云管理控制台](#)，鼠标指向页面右上角的用户名，在下拉列表中单击“我的凭证”。

图 9-1 我的凭证入口



2. 在“我的凭证”页面中选择“访问密钥”页签。单击“新增访问密钥”，按操作指引获取认证账号的AK/SK，请妥善保管AK/SK信息。

图 9-2 访问密钥



📖 说明

- 每个用户仅允许新增两个访问密钥。
- 为保证访问密钥的安全，访问密钥仅在初次生成时自动下载，后续不可再次通过管理控制台页面获取。请在生成后妥善保管。

获取区域名称

区域是一个地理区域的概念，由于带宽原因，会在多个地区建立数据中心，提供服务。

医疗智能体服务部署在华北-北京四，区域名称为cn-north-4；部署在华东-上海一，区域名称为cn-east-3。

获取平台 ID

平台ID与“命令行工具 > 步骤3 初始化配置”中的platform-id对应。

1. 登录医疗智能体平台。
2. 在右上角用户名中选择“个人设置”，获取平台ID。

图 9-3 个人设置

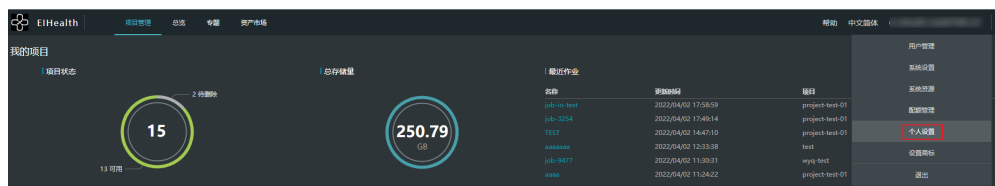


图 9-4 平台 ID

系统信息

平台 ID e53fb

9.2 步骤 2：获取命令行工具

下载命令行工具 eihealth-toolkit

针对不同操作系统，eihealth-toolkit下载地址如下所示。

表 9-1 下载列表

支持平台	下载地址
Windows 64位	health-windows-x86_64.zip 、 health-windows-x86_64.zip.sha256
Linux ARM 64位	health-linux-aarch64.tar 、 health-linux-aarch64.tar.sha256

支持平台	下载地址
Linux AMD 64位	health-linux-x86_64.tar 、 health-linux-x86_64.tar.sha256
macOS	health-macOS-x86_64.tar 、 health-macOS-x86_64.tar.sha256

📖 说明

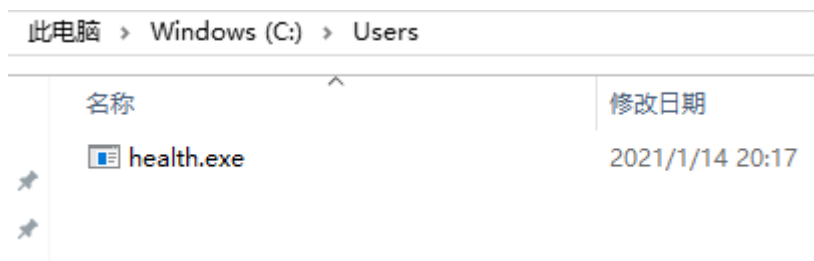
- 本页面命令行工具下载后，在使用时，需用到您注册华为账号并开通华为云时提供的用户名等信息，用于登录并操作EIHealth平台的项目、数据等资产。这些信息的处理将遵循您已接收的《[华为云用户协议](#)》及《[隐私政策声明](#)》约束。
- 下载地址中带有sha256后缀的链接，指的是对应软件包的校验文件。例如：Windows x64版本的下载链接是health-windows-x86_64，它的校验文件下载链接则是health-windows-x86_64.zip.sha256。

安装 eihealth-toolkit

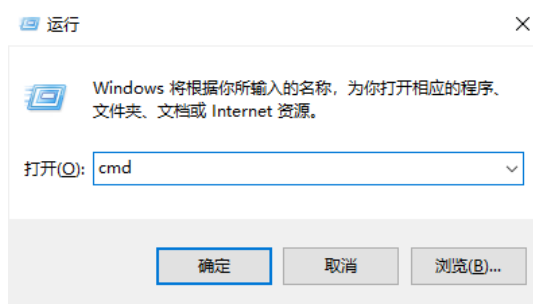
本示例中以Windows系统为例，介绍安装命令行工具的方法。

1. 获取Windows版本的命令行工具，得到health.exe文件，health文件无需安装，放置在任一文件夹中即可。例如，放到C盘Users文件夹中。

图 9-5 下载命令行工具



2. 使用键盘win键+R，输入cmd，并单击“确定”打开windows的cmd窗口。



3. 进入cmd窗口，输入health命令，即可使用。

如果命令行工具所在的文件夹和进入cmd的默认路径不一致，可以使用cd命令进入health文件所在目录。例如，命令行工具放在D盘demo文件夹中，可以使用如下命令进入。进入后，再执行health命令即可开始使用命令行工具。如果不清楚获取路径方法，请参考图9-7获取路径。

```
cd /d d:\demo
```

图 9-6 cmd 窗口输入 health 命令

```
C:\Users\>health
A simple command client for ei health.

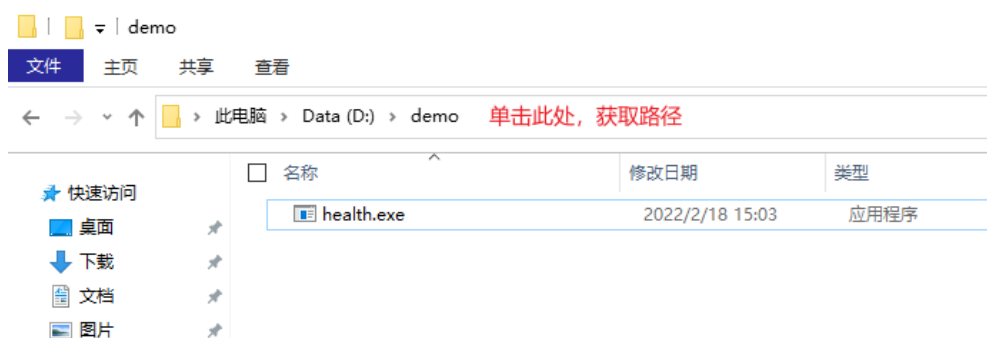
Usage:
  health [flags]
  health [command]

Available Commands:
  add          add a member...
  cancel      cancel or forcibly terminate the job
  cd          change project directory
  config      configure your config file ...
  cp          copy project objects
  create      create a app,workflow,job,template,database...
  delete      delete a app, workflow, job, template, database...
  docker      push, pull or list images...
  download    download objects from project directory to local directory
  edit        edit a app,workflow...
  get         get project, app, workflow, template, database ...
  help        Help about any command
  import      import template, data of database, app, workflow...
  ls          list project objects
  mkdir       create new project directory
  pwd         print current work directory
  retry       run the job or notebook again
  rm          delete project objects
  switch      switch project ...
  transfer    transfer project...
  upload      upload objects from local directory to project directory
  version     show cli version

Flags:
  -h, --help  help for health

Use "health [command] --help" for more information about a command.
```

图 9-7 获取路径



说明

使用Linux版本命令行工具时，您需要在本地搭建Linux环境，并将下载的health文件放至所需的目录下。macOS执行命令和linux一致，

- 如果当前目录为health所在目录，可以使用./health命令使用命令行工具。
- 如果当前目录不是health所在目录，需要使用绝对路径。如当前目录为/opt，假设health存放在/root/health-toolkit/下，需要指定/root/health-toolkit/health路径进行使用。
- 如果无法运行，提示Permission denied，使用**chmod 755 health**命令设置执行权限。

9.3 步骤 3：初始化配置

在使用命令行工具前，需要初始化配置信息，通过**config**命令对eihealth-toolkit进行初始化配置。本节以Windows为例介绍配置过程，Linux和macOS环境配置过程相同。

方法 1：使用账号、密码初始化

执行以下命令，进行初始化，命中的xxx和region信息请参考[表9-3](#)进行替换。

- Windows版命令行工具
health.exe config add --domain-name xxx --user-name xxx --password xxx --region cn-north-4 --platform-id xxx
- Linux、macOS版命令行工具
./health config add --domain-name xxx --user-name xxx --password xxx --region cn-north-4 --platform-id xxx

表 9-2 参数说明

参数	是否必填	说明
--domain-name	是	主账号名。
--user-name	是	子账号名。若是主账号则user-name与domain-name相同。
--password	是	密码。
--region	是	服务区域名称。请依据实际购买平台的区域填写。当前支持华北-北京四（cn-north-4）、华东-上海一（cn-east-3）。
--platform-id	是	平台ID，获取方法请参见 步骤1：获取认证信息 。

方法 2：使用 AK、SK 初始化

执行以下命令，进行初始化，命中的xxx和region信息请参考[表9-3](#)进行替换。

- Windows版命令行工具
health config add --ak xxx --sk xxx --region cn-north-4 --platform-id xxx
- Linux、macOS版命令行工具
./health config add --ak xxx --sk xxx --region cn-north-4 --platform-id xxx

表 9-3 参数说明

参数	是否必填	说明
--ak	是	AK(Access Key ID)：访问密钥ID。

参数	是否必填	说明
--sk	是	SK(Secret Access Key): 与访问密钥ID结合使用的密钥。AK、SK获取方法请参见 步骤1: 获取认证信息 。
--region	是	服务区域名称。请依据实际购买平台的区域填写。当前支持华北-北京四 (cn-north-4)、华东-上海一 (cn-east-3)。
--platform-id	是	平台ID, 获取方法请参见 步骤1: 获取认证信息 。

图 9-8 命令示例

```
C:\Users>health config add --ak B2[redacted] --sk MLT[redacted] --region cn-north-4 --platform-id ca[redacted]
add ak successfully!
add sk successfully!
add region successfully!
add platform-id successfully!
```

10 入门实践

表 10-1 常用最佳实践

实践	描述
基于二代测序的基因组突变检测	本最佳实践提供了通过命令行工具上传数据、上传镜像后，在医疗智能体平台搭建NGS流程，执行分析作业及批量执行NGS分析。
新型冠状病毒（COVID-19）虚拟药物筛选	本最佳实践介绍如何使用EIhealth平台虚拟药物筛选功能，通过获取示例数据，创建药物虚拟筛选任务并查看结果。