**EigenGWAS 云平台计划**

本数据分析平台拟依托阿里云平台，平台域名等由阿里云提供，发布前在我方实验室

linux服务器1(版本 Linux 200server 3.10.0-693.el7.x86\_64，kernal 80，memory 700Gb)

linux\_cluster服务器2（版本 Linux tc6000 3.10.0-693.el7.x86\_64）

测试。

本平台前端开发语言不限，团队之前使用R语言做了一个网页应用，源码见

https://github.com/gc5k/EigenGWASRDoc

**1 用户进入网页**

文本位（后期会放EigenGWAS介绍）

注册、登陆按钮

**2 注册&登陆**

**2.1 注册**

用户设置用户名、密码、邮箱，通过邮箱验证用户真伪，防止恶意用户注册行为（邮箱验证是否有效，是否有其他防止恶意注册的方法）。邮箱验证通过，在服务器开辟空间，建立用户个人数据库(什么数据库？MySQL)。

***p.s.*** 是否可以实现弹性空间分配，根据已上传文件大小实时扩容，尽量避免固定值空间分配。

个人数据库文件结构（拟）：1级结构：用户信息；上传数据；数据分析中间文件；数据分析结果；其他。

**2.2 登陆**

登陆页面：用户名、密码、忘记密码、注册

**3 个人界面**

*个人中心：*个人信息修改、完善，密码修改

*EigenGWAS：*数据上传

*任务：*进行中任务、全部任务；每个任务可查看基本信息（上传时间、开始时间、完成时间）、分析结果、操作（删除）。

*关于平台：*前后端开发者信息，联系方式

**4 用户功能**

**4.1 基本用户功能**

用户个人信息完善、修改，用户密码修改等...同步信息到数据库“用户信息”；非活跃用户清除服务器个人数据库（或者其他机制，允许新的空间需求可以占用非活跃用户空间）。

**4.2 EigenGWAS分析**

**4.2.1 数据上传页面**

数据、参数上传页面：数据文件（本地文件系统上传），染色体数目（数字输入），群体类型（给定类型选取），两个数值（给定范围数字选取）。

数据文件可能会非常大，这种情况下需要

1、文件大小超过某一阈值，限制上传，弹出文本信息（让用户找我们开辟更大的空间）；

2、文件大小很大但未超阈值，此时上传时间较长，页面最好能有一个上传时间指示器，同时要有防断流机制，允许用户断点重传

用户上传完数据并确认提交，数据与参数同步到数据库“上传数据”中。

**4.2.2 后台排队系统**

建立排队机制，排队系统可以根据当前系统实时可用计算资源，对新上传数据排队计算。前端显示目前队列位置，预估完成时间(时间复杂度和空间复杂度)。数据计算目前调用R语言脚本，计算产生的中间文件存储在“数据分析中间文件”中。

**4.2.3 结果通知机制**

数据分析结果存放在“数据分析结果”中，该结果同步到前端“分析结果”页面。给用户发邮件，邮件包含“任务”界面链接。系统发送的邮件由 [admin@XXX.com](mailto:admin@XXX.com)发出。

**4.2.4 数据保存机制**

分析结果（“数据分析结果”文件夹）永久保留；分析数据文件（“上传数据”文件夹）分析结束后移除（一个星期后；如果分析完移除，不能二次分析，因为有可能用户会二次登陆，调整参数继续分析数据。对于这类分析，可以适当后台降低优先级）。