

高性能计算服务环境应用编程接口^①



和 荣^{1,2}, 肖海力¹, 王小宁¹, 卢莎莎¹, 迟学斌^{1,2}

¹(中国科学院 计算机网络信息中心, 北京 100083)

²(中国科学院大学, 北京 100049)

通信作者: 和 荣, E-mail: herong@sccas.cn

摘 要: 高性能计算服务环境主要面向用户、科研团队提供高性能计算服务. 随着环境接入的超算中心以及应用社区和业务平台越来越多, 超算中心以及社区和业务平台的用户希望能够使用原有账号登录高性能计算环境使用资源. 高性能计算服务环境目前提供的应用编程接口仅支持通过 LDAP 认证的网格账号. 为使得应用社区和业务平台用户使用自己原有的登录方式认证通过后就可访问高性能计算服务环境, 我们重新设计开发了高性能计算服务环境应用编程接口. 本文着重介绍新版应用编程接口的结构与部署实现, 并通过用例来说明如何调用新版接口. 新版接口为社区和业务平台接入高性能计算环境提供了更方便且安全地支撑.

关键词: 用户认证; 应用编程接口; 接口网关; 授权; 接口安全; 负载均衡

引用格式: 和荣, 肖海力, 王小宁, 卢莎莎, 迟学斌. 高性能计算服务环境应用编程接口. 计算机系统应用, 2022, 31(8): 184-191. <http://www.c-s-a.org.cn/1003-3254/8636.html>

Application Programming Interface for High Performance Computing Environment

HE Rong^{1,2}, XIAO Hai-Li¹, WANG Xiao-Ning¹, LU Sha-Sha¹, CHI Xue-Bin^{1,2}

¹(Computer Network Information Center, Chinese Academy of Sciences, Beijing 100083, China)

²(University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: The high-performance computing environment is designed to provide high-performance computing services for users and research teams. As more and more supercomputing centers, application communities, and service platforms have access to the environment, their users expect to use resources by logging in the high-performance computing environment with their original accounts. Now, only grid accounts authenticated by LDAP can access the application program interface (API) of the high-performance computing environment. To meet the expectation of users, we develop a new version of API for the high-performance computing environment. This study mainly introduces the structure and implementation of the new API, and the calling ways of the new API are exemplified. The new API can provide more convenient and secure services for the access of communities and service platforms to a high-performance computing environment.

Key words: user authentication; application program interface (API); API gateway; authorization; interface security; load balancing

高性能计算能力是国家综合实力和创新能力的重要体现, 是支撑我国科技持续发展的关键技术之一. 高性能计算服务环境^[1], 又称中国国家网格 (China national

grid, CNGrid), 在国家科技发展中具有重要推动作用. 高性能计算服务环境一直受到国家的高度重视与国家重点研发计划的持续支持. 历经 10 多年的发展, 高性

① 基金项目: 中国科学院战略性先导科技专项 (A 类)(XDA19020101)

收稿时间: 2021-11-17; 修改时间: 2021-12-14; 采用时间: 2021-12-21; csa 在线出版时间: 2022-05-30

能计算服务环境已经接入包括中国科学院计算机网络信息中心节点、国家超级计算天津中心节点、国家超级计算广州中心节点、国家超级计算深圳中心节点、国家超级计算无锡中心节点等在内的 22 个节点; 提供涉及量子化学、分子模拟、高能物理、生物科学等领域的开源软件、商业软件和自主研发软件. 高性能计算服务环境屏蔽了作业管理系统、接入方式以及管理制度等方面的异构性, 为科研人员提供具有统一访问入口、统一使用方法和用户技术支持的高水平高性能计算应用服务.

高性能计算服务环境作为一个服务环境, 其主要目标为吸引更多的用户和科研团队来使用. 目前, 高性能计算服务环境主要提供 3 类用户接口: 第 1 类命令行, 主要是面向传统的高性能计算用户, 支持程序编译等多种灵活的交互命令操作; 第 2 类高性能计算环境通用计算平台^[2], 主要面向领域用户, 通过计算门户即可提交作业并管理各类计算任务; 第 3 类应用编程接口, 主要面向应用社区和业务平台的建设提供支撑. 随着环境的发展, 接入的超算中心以及应用社区和业务平台越来越多, 希望超算中心以及社区和业务平台用户能够以原有账号登录高性能计算环境使用资源. 现有高性能计算环境仅支持通过 lightweight directory access protocol (LDAP)^[3] 认证的网格账号登录, 应用社区和业务平台都有自己的用户且认证方式各不相同, 同时集群用户也想直接访问环境中资源.

为了更好地支持社区与业务平台, 我们开发设计了新版高性能计算服务环境应用编程接口. 该接口支持环境认证与授权服务系统, 允许社区或业务平台用户通过认证授权后可直接调用接口访问国家高性能计算环境. 新版接口增加了网关模块来实现接口的访问控制、流量监测等, 增强了接口的易用和安全性.

本文围绕国家高性能计算服务环境应用编程接口展开叙述, 重点介绍接口的整体结构、关键技术、功能实现以及部署测试, 最后进行总结和展望.

1 整体结构

图 1 为接口的整体结构图, 外部应用社区和业务平台通过访问高性能计算服务环境应用编程接口网关与系统核心服务交互获取并使用高性能计算环境资源.

1.1 应用社区和业务平台

应用社区和业务平台是高性能计算专项或高性能

计算领域内的社区和业务平台, 包括生物医药社区、教育实践平台等. 他们发送 HTTP 请求到接口网关, 网关响应后返回数据给社区或业务平台.

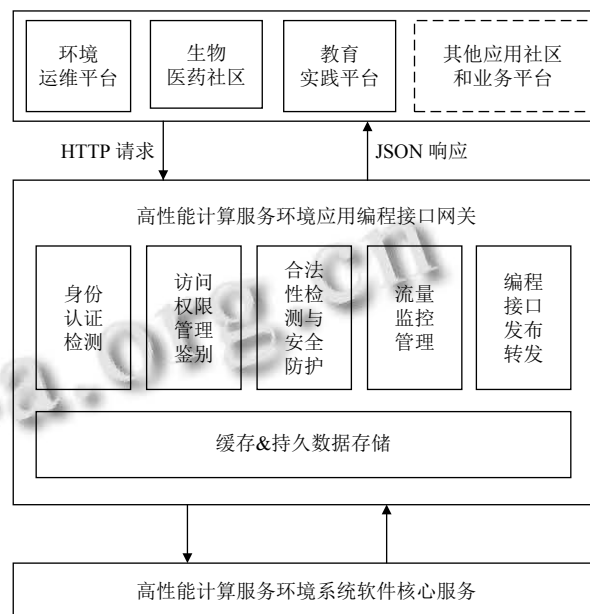


图 1 接口整体结构图

1.2 接口网关

接口网关的基本功能之一是转发客户端的请求至后台软件核心服务. 为了保证转发过程的安全性, 网关需要完成以下操作:

- 1) 身份认证检测, 确保用户身份信息的有效性.
- 2) 接口访问权限管理与鉴别, 判断用户是否具有访问该接口的权限.
- 3) 参数合法性检测和接口安全防护, 主要用于保护后端系统和数据不被恶意侵犯和篡改. 合法性主要是检测用户请求路径或参数是否有错误; 对接口调用进行完整性、有效性、唯一性鉴定, 包括对请求参数加密、参数携带的时间戳是否正确.
- 4) 对访问的请求进行流量控制与管理, 对最大连接数、HTTP 连接以及每秒请求数的限制. 设置一定的流量控制策略实现请求的最大化满足, 以最快的速度对用户请求做出响应.

1.3 系统核心软件服务

系统软件核心服务包括系统核心软件 SCE^[4,5]、SCEAPI 开发库^[6]和 SCE 转发服务.

SCE 实现了资源的统一管理和调度, 采用包括应用层、由中央服务 (center service, CS) 和前端服务 (front service, FS) 构成的服务层和由分布在全国各地

的高性能计算机 (high performance computer, HPC) 构成的资源层在内的单中心 3 层架构部署. FS 是 SCE 的前端服务, 主要用于资源的接入与监控. 通过定义 HPC 的各种驱动, 有效的屏蔽底层系统的差异; 系统服务层 CS 是 SCE 最核心的模块, 提供了用户使用超级计算环境所必需的最基本的功能以及若干扩展功能; 面向用户提供命令行、Web Portal、GUI 等多种使用方式.

SCEAPI 开发库是一套利用 Java 代码开发的代码库, 通过与 SCE 交互提供作业管理服务、文件传输服务、应用管理服务以及用户管理服务.

SCE 转发服务提供访问 SCE 的 API, 提供多线程、多用户的服务功能, 从而支持大量用户的访问. SCE 转发服务通过调用 SCEAPI 开发库完成具体的功能、组装 JSON 返回结果等信息.

2 关键技术

高性能计算服务环境应用编程接口主要面向应用社区和业务平台提供方便可用的接口访问高性能计算环境. 设计实现过程中主要涉及身份认证检测、接口访问控制、安全加固、流量控制 4 个方面的关键技术.

2.1 身份认证检测

用户身份认证检测是接口网关检测请求消息的第一道关卡, 负责检测用户通过认证的有效性. 客户端需接入高性能计算环境认证与授权服务^[7], 通过认证之后使用访问令牌 (后面简称 `access_token`) 访问接口网关服务. 接口网关对用户身份认证检测实际上就是判断 `access_token` 的有效性. 图 2 表示接口网关进行身份认证检测的流程.

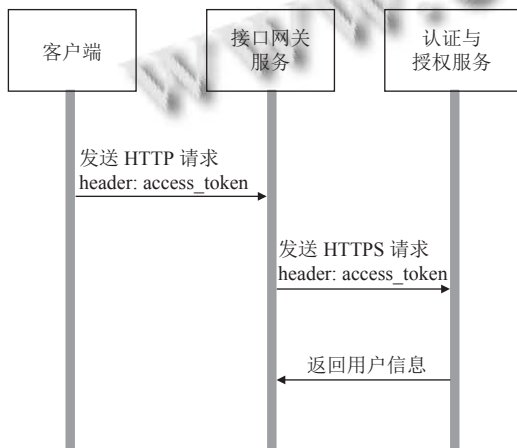


图 2 接口网关身份认证检测流程图

1) 判断消息头是否具有 `access_token`
网关接收到消息请求后判断消息头中是否含有 `access_token`, 没有则提示用户需要登录.

2) 检测数据库

根据获取到的 `access_token` 与数据库已经存储的信息进行比对, 若该 `access_token` 已有对应的用户信息且处于有效状态进行下一步权限判断; 否则与认证授权服务交互.

3) 与认证与授权服务交互

发送 `https` 请求给认证与授权服务, 通过返回的数据判断 `access_token` 是否有效并记录用户信息.

环境认证与授权服务系统主要用来实现标记不同用户的身份信息并根据用户需求授予权限. 认证授权系统基于 OAuth2、OpenID Connect (OIDC)^[8,9] 协议设计实现, 包括用户登录认证和授权两个功能模块. 具体结构图如图 3 所示. 目前已实现环境网格账号认证、集群账号中的“元”以及人工智能系统账号的认证, 授权主要是允许第三方客户端获取用户信息.

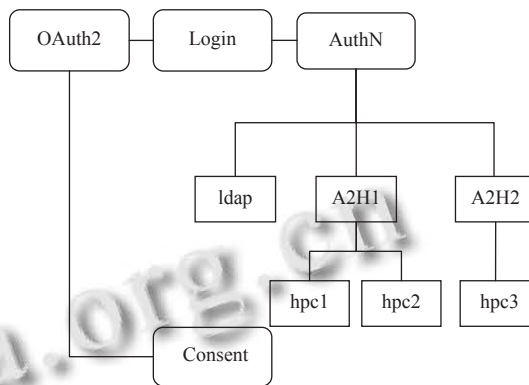


图 3 认证与授权系统结构图

第三方客户端通过认证与授权系统获取访问令牌具体流程图如图 4 所示.

1) 初始化一个 OAuth2 流程

客户端按照接入认证与授权服务的文档要求构建登录请求, 携带相关参数.

2) 认证

OAuth2 认证服务自动判断用户是否已经登录并跳转到相应的页面. 若未登录根据用户的选择进入登录页面, 用户在登录页面输入相应账号和密码进行验证.

3) 授权

用户登录成功后, 跳转至授权页面. 用户点击授予的权限发送给授权服务, 授权服务获取授权后客户端

可获取授权码。

4) 获取访问令牌

客户端利用返回的授权码从认证与授权服务获取访问令牌。

2.2 接口访问控制

用户身份认证检测之后是接口访问授权检测。高

性能计算环境中,不同的客户端不同的用户身份被授权访问不同的 API 组。网关系系统获取用户的认证身份信息之后,可以根据预设的分组信息检测对目标接口服务访问的权限。目前环境对接口的访问控制主要是基于角色的授权访问和第三方应用授权策略相结合来实现。

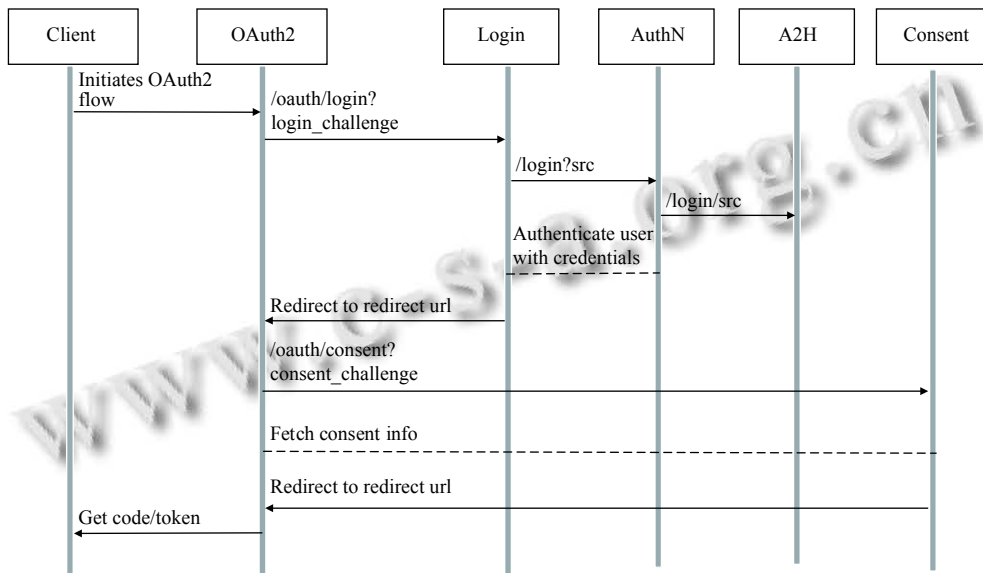


图4 认证授权流程图

1) 基于角色的授权访问

环境普通网格用户: 根据账号申请初期用户的需求给出所需权限。利用机器学习算法分析用户的使用习惯并适应性更改权限;

环境技术支持者: 除具备普通用户权限外,可查看所有用户的作业、修改作业状态、查看下载作业文件;

环境管理员: 可访问所有接口,包括修改用户信息等;

第三方应用用户: 根据用户初次通过认证与授权服务登录系统时所选择的权限为用户分配权限,初期用户可能只具有查看和个别集群使用权限,后续会根据用户的使用习惯动态调配。

2) 第三方应用授权策略

初步申请接入环境时给所有应用开通查看资源的权限,根据应用针对的用户常用的软件或资源开通某一集群的使用权限,保证用户可以访问环境并提交任务。后续根据应用用户自身的需求可增加权限。

用户发送访问请求,根据用户类型决定通过哪些访问策略为用户请求做出响应。环境账号发送的请求,

根据是否满足角色的访问控制做出拒绝或者满足的响应。第三方应用账号首先判断是否满足第三应用的授权,然后判断用户角色策略做出拒绝或者满足的响应。

2.3 安全加固

为了保证 API 请求的安全,客户端需要对发送的 API 进行加密,将加密后的签名串作为请求头的一部分传到接口网关。为防止请求的重放性,请求参数还加入了时间戳等。

接口网关安全检查流程如图 5 所示。网关首先对接收到的 API 请求进行跨域检查,避免跨域访问引起的安全问题。其次,检测请求所携带的应用标识 appid 是否是有效的。再次,对请求参数做加密签名,并判断与请求参数携带的签名串是否相同,如果签名串相同,则通过 API 完整性检查;否则提示用户签名错误。最后进行权限检查,结合应用标识 appid 和用户自己的权限做出判断。

网关安全检查可有效避免用户的恶意攻击,使得用户的请求在网关层就被阻断,有效提高了 SCE 转发服务的健壮性。

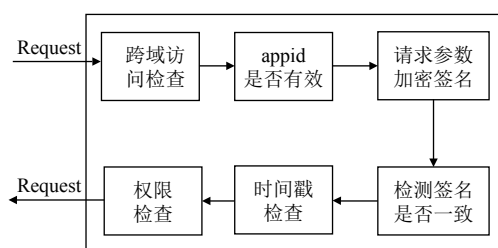


图5 接口网关安全检查流程图

2.4 流量控制

流量监控与管理是网关系统的常用功能,可根据运维需求对通过的请求流量做分流或熔断等处理.流量监控包括对最大连接数、HTTP 连接以及每秒请求数的限制.

根据以往接口运行经验,设置一定的流量控制策略实现请求的最大化满足,以尽快的速度对用户请求做出响应.

1) 用户流量限制

对用户单位时间内的调用次数设定最大访问值.避免用户频繁调用接口,设置用户每小时访问接口的最大值,每访问一次接口,访问次数累计,达到最大之后直接拒绝用户的请求.

2) 接口流量限制

对于作业提交接口,为避免用户批量提交作业,同一时间限制用户提交作业数目;对于需要与核心服务 SCE 交互才能获取数据的接口,受 SSH 最大连接数的限制,对接口设定最大访问次数.

3) 应用流量限制

对调用接口的应用(应用社区、业务平台)设置一个总的最大访问次数,避免应用被攻击后持续访问接口,造成接口服务无法正常服务.

4) 其他措施

对于统计类的接口,借助缓存等服务来存储数据,避免接口一直与后台服务交互;SCE 转发服务采用微服务的方式部署,避免相互之间的影响,对访问可做到分流限制.

3 功能实现

3.1 接口网关

1) 接口设计原则

为了尽量保持与现有的高性能计算环境应用接口设计一致,访问参数跟原有接口一样,每个接口都携带

应用标识字段;请求消息头增加两个字段:环境认证与授权系统签发的 `access_token`, 用户来源字段: `user-source`. 通过这两个字段可有效判断用户的登录信息并获取用户身份信息.

2) 合法性以及权限检测

通过过滤器实现合法性以及权限检测等功能,任何发送到网关的请求都要经过过滤器中定义的各种检测规则方可转发至 SCE 转发服务.如时间戳检查未通过时提示用户“时间戳错误”;签名验证未通过时提示“签名错误”;权限检测未通过时提示“API 授权错误”;跨域访问判断未通过时提示“访问被拒绝”.设置清晰的错误提示方便用户快速定位错误源头,尽快修复问题.

3) 接口转发模块

为方便网关层快速增加修改接口,所有后台接口通过配置文件来管理.后台接口地址需要存储在配置文件中,网关收到请求信息并经过安全诊断后按照配置中的路径进行转发.

当接口服务新增一个接口时,只需在配置文件中增加一条转发配置,网关会自动感知配置文件的变化,方便接口的快速发布.

3.2 SCE 转发服务

SCE 转发服务真正服务于用户请求提供返回数据.在实现上为了与核心软件 SCE 交互,需要每一个 API 接口都包括用户名和应用标识两个字段.下面就访问高性能计算环境资源相关的作业管理服务、文件传输服务和资源管理服务展开论述.

3.2.1 作业管理服务

作业管理服务主要提供用户查询、修改作业状态、提交作业.表 1 是作业管理服务包含的接口,用户可提交作业并根据自己的需求查询以及修改作业状态.

表 1 作业管理服务接口

方法	路径	功能
GET	/jobs	查询作业
POST	/jobs	提交作业
GET	/jobs/{ujid}/status	获取指定作业号的作业状态
PUT	/jobs/{ujid}/status	修改指定作业号的作业状态
POST	/jobs/recycleBin/restore	恢复还在回收站的作业文件或作业
DELETE	/jobs/recycleBin/empty	清空回收站的作业文件或作业

3.2.2 文件传输服务

文件传输服务主要用于上传、下载作业的文件信息,也包括查看作业文件目录信息、作业文件内容.接

口同时提供了可直接调用无需落地传输的数据中转传输服务 mcp^[10] 进行文件的上传和下载。

表 2 是文件传输服务所包含的接口, 目前用户比较常用的就是通过 mcp 服务实现的文件查看和上传。调用此接口时可不指定查看文件的位置, 服务器会首先查看集群服务器是否存在文件, 没有的话再去网格服务器查看文件。用户通过调用一次接口即可获取到自己的作业文件内容, 增强了用户的体验度。

表 2 文件传输服务接口

方法	路径	功能
GET	/data/jobs/{ujid}/cs	查询作业文件列表
POST	/data/jobs/{ujid}/hpc	查询作业在集群上的文件列表
GET	/data/jobs/{ujid}/hpc/{fileName}/view	查询集群上文件内容
GET	/data/jobs/{ujid}/view	查询集群或CS服务器上文件内容
PUT	/data/jobs/{ujid}/cs	上传文件至CS服务器
PUT	/data/jobs/{ujid}/mcp	通过mcp服务上传文件
PUT	/data/jobs/{ujid}/cs/{fileName}	上传文件至CS服务器
GET	/data/jobs/{ujid}/mcp/{fileName}	通过mcp服务查看文件
GET	/data/jobs/{ujid}/list	查看作业的目录信息
GET	/data/jobs/{ujid}/fileDownload	下载文件

3.2.3 资源管理服务

资源管理服务主要用于查询环境中所有的集群和应用信息。表 3 是资源管理服务的所有接口, 用户在提交作业之前需要先查看可提交此应用的集群信息, 然后选择可用的集群提交作业。

表 3 资源管理服务接口

方法	路径	功能
GET	/resources/hpcs	查询环境提供的集群列表
GET	/resources/queues	查询集群上的应用队列信息
GET	/resources/applications	查询环境提供的应用列表
GET	/resources/applications/{appName}	查询环境包含指定应用的集群

4 部署测试

高性能计算环境通用计算平台作为调用接口的第三方应用程序, 同时由我们自己研发实现。下面以高性能计算环境通用计算平台(简称 Portal 服务)为例来说明新版接口的工作流程。

图 6 是接口的具体部署图, 其中, 环境认证与授权服务、Portal 服务、接口网关服务和 SCE 转发服务都是以 Docker 容器的方式部署, 方便后期维护时服务之

间不会相互影响。接口网关收到请求后通过相关预处理转发至 SCE 转发服务中内部访问接口。

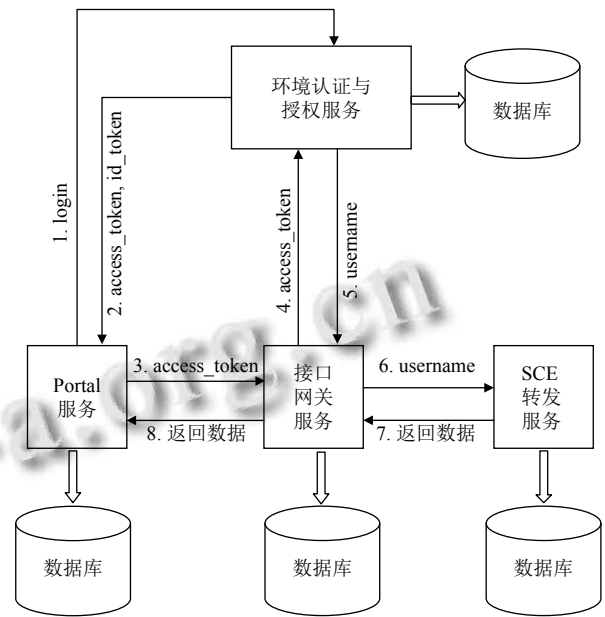


图 6 接口部署图

4.1 各项服务功能及对应数据库

表 4 介绍了接口部署时涉及到的各个服务以及对应的数据库的功能。

表 4 服务系统功能及对应数据库

服务名称	功能	数据库用途
环境认证与授权服务	标记不同用户的身份信息并根据用户需求授予权限	存放环境认证与授权服务的应用、用户登录信息等
Portal 服务	高性能计算环境通用计算平台	存放用户常用列表、修改颜色的标签以及用户登录信息
接口网关服务	外部调用的接口服务, 用于管理后台接口	接口权限管理、接口登录信息存储
SCE 转发服务	提供真正数据的后台接口	存储用户的作业信息、用户信息

1) 环境认证与授权服务

通过设置环境认证与授权服务对应的数据库可为接入到认证与授权服务端的应用和接口网关提供用户信息, 验证访问令牌 access_token 的有效性。

2) Portal 服务

通过 Portal 服务数据库存储用户的操作记录显示用户首页的常用列表、存储用户的登录信息可方便用户在登录有效期内直接利用接口获取数据。

3) 接口网关服务

接口网关服务通过数据存储的数据可用于判断用户

的权限、访问接口时携带的 `access_token` 信息有效性等,避免一直与认证与授权服务交互换取用户登录信息。

4) SCE 转发服务

SCE 转发服务数据库主要存储用户的作业信息、用户信息、环境中的资源信息。通过此数据库才能实现获取用户在环境中的作业数据。

4.2 用例

通过用例 Portal 服务介绍接入环境认证与授权服务以及调用接口的流程。

4.2.1 接入环境认证与授权服务

按照认证与授权服务颁发文档中的流程接入环境认证与授权服务系统,主要进行如下操作:

1) 申请接入账号

主要包括申请应用唯一标识、授权类型、授权范

围、回调地址。

2) 网站接入开发

构造登录认证请求、用户完成认证授权后浏览器跳转到回调地址,后台获取 `code` 码然后换取访问令牌和身份令牌。具体实现流程如前面图 3 所示,这里不再重复介绍。

3) 部署上线

Portal 服务接入环境认证与授权服务后,点击登录可跳转到认证与授权服务提供的登录页面,如图 7 所示,用户输入登录信息,验证通过后 Portal 后端通过获取到的 `code` 换取访问令牌和身份令牌,之后 Portal 端通过调用接口即可获取相应的数据。发送给网关的访问请求需要携带访问令牌和用户来源以便网关服务可以判断用户的有效性。



图 7 环境认证与授权服务登录页面

4.2.2 接口工作流程

用户从发送请求到获取到接口返回的数据需要经过以下 5 个步骤,图 8 为作业接口示例。

1) Portal 发送请求

Portal 按照对外公布的接口使用方法,拼接请求所需参数,对参数进行加密处理,加上当前时间戳信息,消息头添加认证与授权服务颁发的 `access_token`,给网关发送 HTTP 请求。

2) 网关进行身份认证检测

网关按照第 2.1 节身份认证检测流程检测此次请

求用户的身份,若未通过检测提示用户请先登录后访问。

3) 网关合法性及权限检测

网关对此请求的参数以及权限进行检测,保证接口的安全性。

4) 网关转发接口

网关把用户的请求信息转发至 SCE 转发服务。

5) 接口返回数据

接口返回数据给网关,网关对数据处理后转发给用户。

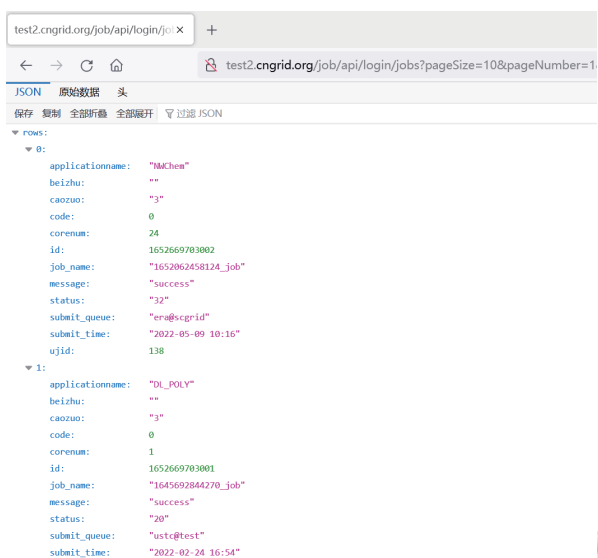


图8 作业接口示例

4.3 测试

接口部署完成后,我们对接口的正确性、响应时间等进行了测试。

1) 正确性测试

先后调用新版接口与原有接口返回数据对比,检测结果是否正确;不用户同时登录下调用接口检测是否出现数据混乱等错误现象。

2) 响应时间

图9为Portal调用查询应用、查询作业、提交作业、上传文件、查看文件、下载文件等常用接口的平均响应时间,横坐标为不同的接口、纵坐标为响应时间(s)。由图9可以看出接口基本上在0.2s内可返回结果,提交作业模块受作业参数、加载作业文件的影响响应时间略长,大约在0.5s左右。

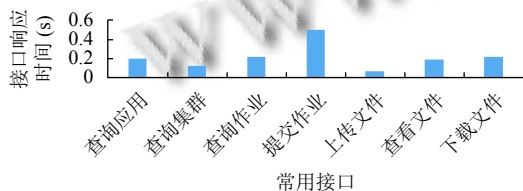


图9 常用接口响应时间

5 总结与展望

本文介绍了高性能计算服务环境应用编程接口的

总体结构图、关键技术以及如何设计实现和部署,并通过用例高性能计算环境通用计算平台Portal说明了如何调用接口访问国家高性能计算环境。重点突出了接口网关在用户身份认证检测、对接口的访问控制、安全性方面的加固以及流量控制方面的重要性。

应用社区和业务平台无须再重新申请网格账号,只需接入环境认证与授权服务即可通过新版接口访问高性能计算环境资源,大大提高了用户的使用便捷性。其他应用也可以通过接入环境认证与认证服务系统访问国家高性能计算环境,环境网格账号与应用社区、业务平台账号真正实现了互联互通并相互访问资源。

目前接口部署在测试环境中,后续会进一步对接口进行测试优化,并尽快部署到正式环境中推广给各大社区和业务平台使用。

参考文献

- 1 中国国家网格. <http://www.cnggrid.org>. [2021-11-15].
- 2 和荣,王小宁,卢莎莎,等.高性能计算环境通用计算平台.计算机系统应用,2019,28(12):55-62. [doi: 10.15888/j.cnki.csa.007176]
- 3 OpenLDAP. <https://www.openldap.org/doc/admin26>. [2021-11-15].
- 4 Dai ZH, Wu LJ, Xiao HL, et al. A lightweight grid middleware based on openssh—SCE. 6th International Conference on Grid and Cooperative Computing (GCC 2007). Urumchi: IEEE, 2007. 387-394.
- 5 龙斌,迟学斌,肖海力.基于命令行客户端的网格软件SCE设计与实现.计算机系统应用,2010,19(9):64-68. [doi: 10.3969/j.issn.1003-3254.2010.09.014]
- 6 曹荣强,肖海力,王小宁,等.基于REST风格的科学计算环境Web服务API.科研信息化技术与应用,2016,7(3):43-48.
- 7 环境认证与授权服务. <https://login.cnggrid.org/loginconsent/login>. (2020-12-24)[2021-11-15].
- 8 Hardt D. The oauth 2.0 authorization framework. <https://tools.ietf.org/html/rfc6749>. [2021-11-15].
- 9 Sakimura N, Bradley J, Jones M, et al. OpenID connect core 1.0 incorporating errata set 1. http://openid.net/specs/openid-connect-core-1_0.html. (2014-11-08).
- 10 刘玉环,王小宁,肖海力,等.面向科学计算云服务环境的数据中转传输.科研信息化技术与应用,2013,4(5):42-50.

(校对责编:孙君艳)