

基因容器

产品介绍

文档版本 01
发布日期 2023-04-30



版权所有 © 华为技术有限公司 2023。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。

目录

1 产品功能.....	1
2 产品优势.....	3
3 应用场景.....	4
4 使用限制.....	5
5 基本概念.....	6
6 权限管理.....	7
7 与其它云服务的关系.....	10

1 产品功能

📖 说明

基于华为云整体产品规划，基因领域将统一由EIHealth为入口承接，后续GCS会考虑逐渐下线。详见[最新动态](#)。

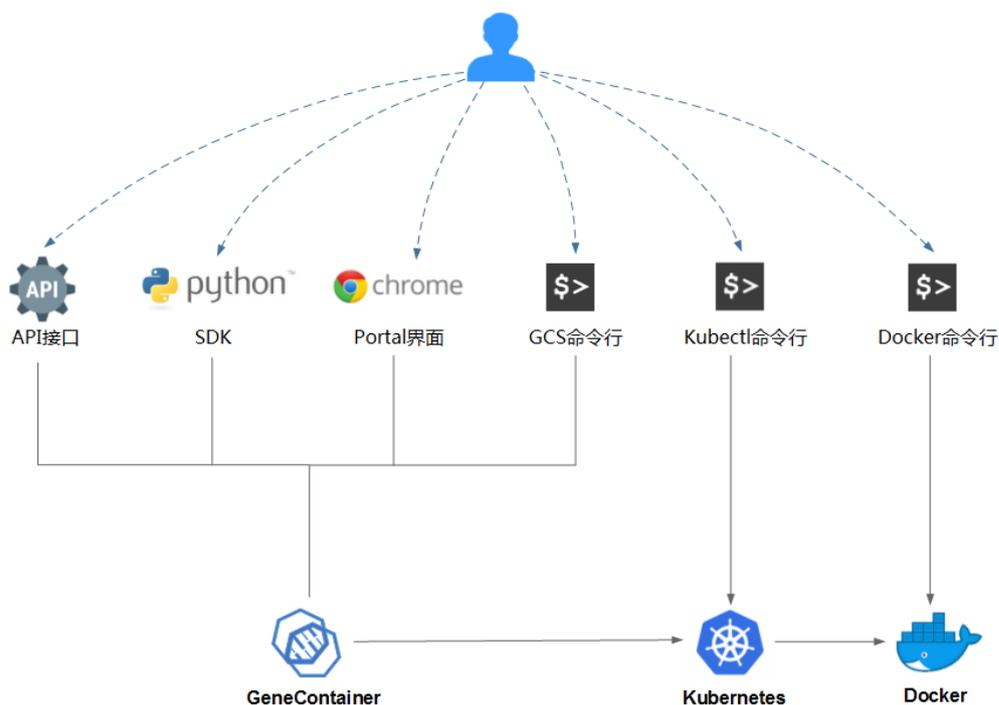
基因容器（GeneContainer Service，GCS）提供云端基因解决方案，支持DNA、RNA等主流生物基因测序场景。基因容器基于轻量级容器技术，结合大数据、深度学习算法，优化官方标准算法，为您提供灵活可定制的测序流程、秒级可伸缩的高可靠资源。

基因容器作为基因测序端到端完整解决方案，为您提供数据管理、测序工具平台、流程定义能力、运行流程以及查看流程执行结果的能力，同时以上能力均支持以CLI命令行方式执行。

基因容器的使用方式包括可视化界面、REST API、SDK、命令行，您可以选择任意一种或是搭配使用。

- 可视化界面使用方法请参见[快速入门](#)，其中GCS界面入口为<https://console.huaweicloud.com/gcs/>。
- REST API的使用方法请参见[API参考](#)。
- SDK使用方法请参见[SDK参考](#)，当前仅支持Python语言的SDK。
- 命令行使用方法请参见[CLI参考](#)。

图 1-1 基因容器的使用方式



PB 级数据管理

提供PB级存储单元，除了支持您对数据的增删改查等基本操作之外，还支持数据快速、冷备、复制、分享、加密等操作。

灵活的流程定制能力

内置多套已验证的业界主流流程，同时提供图形化流程设计器及手动模板编写两种方式供您选择，创建您的自定义测序流程。

快速迁移已有流程

基因容器支持多种流程描述语言，可以将您在第三方系统中的流程（如SJM、WDL、Snakemake、Batch等）快速迁移至基因容器上。

可视化交互式体验

您可以随时通过图形化界面查看流程执行的进度、状态、日志、中间文件以及测序结果。

丰富的计算资源

基因容器以华为云丰富计算资源为基础，支持包年包月及按需计费两种计费模式，依靠基因容器强大的调度能力，能实现秒级弹性伸缩，轻松应对海量测序任务。

友好的命令行工具

基因容器提供免安装命令行工具，你可以快速提交、管理任务，方便对接您的自动化系统。

2 产品优势

基因容器基于Kubernetes智能化基因计算任务调度和Spark等加速服务，为您提供低成本高性能的基因测序解决方案。支持对接深度学习框架，方便您深度解读报告。

秒级并发

基因容器利用容器技术的秒级并发能力，可将WGS从30小时缩短至5小时以内，对比同类竞品，使用相同样本的情况下，资源利用率大幅提升。

简单易用

不单独维护小资源池，使用华为云整个计算资源做为基础，保证资源充足，无需排队。为生物测序行业量身定做的业务流程，更贴近您的业务场景。

弹性伸缩

基因容器提供容器应用的秒级弹性伸缩能力，在流量突增时能快速弹性扩容，保障业务的连续性和高稳定性。当前支持按性能、时间、周期的弹性伸缩策略，您可以自由组合策略以应变业务峰值突发变化。

3 应用场景

通用场景

生物基因企业利用基因容器完成大规模基因组测序分析。

- **优势**
 - 极大提升计算速度，降低运营成本。
 - 流程设计门槛低、方式多样，可以通过设计器、模板甚至可以直接从其他格式模板直接转换
 - 不单独维护小资源池，可直接使用整个华为云计算资源作为基础。保证资源充足，无需排队
 - 为生物测序行业量身定做的业务流程，更贴近用户场景。

4 使用限制

集群限制

对于CCE环境和SEG环境，您的账户需要通过实名认证并且要有100元的余额，否则无法一键创建按量付费的“集群”和“弹性云服务器（下文中简称为ECS节点）”等资源。

区域限制

当前仅支持在华北区进行测序计算。

配额限制

使用基因容器执行测序时，需注意以下配额限制：

- 每个用户默认最多创建5个环境；
- 每个用户默认最多创建100个自定义流程，每个用户执行流程次数不能超过500次；
- 每个用户默认最多创建100个工具。

如果当前配额不能满足业务要求，请登录控制台，请[提交工单](#)申请调整配额。

使用基因容器时也会使用其他云服务，包括：云容器引擎、云容器实例、弹性云服务器、虚拟私有云、容器镜像服务、弹性文件服务、数据湖探索等。可登录控制台，到“资源 > 我的配额”页面查询这些云服务的具体配额信息。如果当前配额不能满足业务要求，请单击“申请扩大配额”，根据指导进行调整。

5 基本概念

流程 (Workflow)

基因测序流程包含测序过程所需工具的执行先后信息以及数据输入输出等定义。流程由至少一个工具组成。流程中的各个工具由其前后顺序关系形成数据流，前序工具为后序工具提供输入。基因容器为您提供了多套典型的测序流程，基于此流程，您可以快速完成测序任务。

工具 (Tool)

工具是生物信息软件的镜像封装，工具既可以编排入流程串联使用，也可以独立使用，基因容器提供了大量的公共工具，同时用户可以制作自定义工具，这些工具都存放在测序仓库中。

执行结果 (Execution)

基因计算流程的某次执行过程。

容器镜像

容器镜像是一个应用的快照。例如，一个容器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。容器镜像用于创建容器，您也可以下载其他人已经创建好的镜像来使用。

桶 (Bucket)

桶 (Bucket) 是对象存储 (简称OBS) 中存储对象的单位。对象存储提供了基于桶和对象的扁平化存储方式，桶中的所有对象都处于同一逻辑层级，去除了文件系统中的多层次树形目录结构。

6 权限管理

如果您需要对华为云上购买的GCS资源，为企业中的员工设置不同的访问权限，以达到不同员工之间的权限隔离，您可以使用统一身份认证服务（Identity and Access Management，简称IAM）进行精细的权限管理。该服务提供用户身份认证、权限分配、访问控制等功能，可以帮助您安全的控制华为云资源的访问。

通过IAM，您可以在华为云账号中给员工创建IAM用户，并授权控制他们对华为云资源的访问范围。例如您的员工中有负责软件开发的人员，您希望他们拥有GCS的使用权限，但是不希望他们拥有删除GCS等高危操作的权限，那么您可以使用IAM为开发人员创建用户，通过授予仅能使用GCS，但是不允许删除GCS的权限，控制他们对GCS资源的使用范围。

如果华为云账号已经能满足您的要求，不需要创建独立的IAM用户进行权限管理，您可以跳过本章节，不影响您使用GCS服务的其它功能。

IAM是华为云提供权限管理的基础服务，无需付费即可使用，您只需要为您账号中的资源进行付费。关于IAM的详细介绍，请参见[IAM产品介绍](#)。

GCS 权限

默认情况下，管理员创建的IAM用户没有任何权限，需要将其加入用户组，并给用户组授予策略或角色，才能使得用户组中的用户获得对应的权限，这一过程称为授权。授权后，用户就可以基于被授予的权限对云服务进行操作。

GCS部署时通过物理区域划分，为项目级服务。授权时，“作用范围”需要选择“区域级项目”，然后在指定区域（如华北-北京四）对应的项目（cn-north-4）中设置相关权限，并且该权限仅对此项目生效；如果在“所有项目”中设置权限，则该权限在所有区域项目中都生效。访问GCS时，需要先切换至授权区域。

如表6-1所示，包括了GCS的所有系统角色。由于华为云各服务之间存在业务交互关系，存储容灾服务的角色依赖其他服务的角色实现功能。因此给用户授予存储容灾服务的角色时，需要同时授予依赖的角色，存储容灾服务的权限才能生效。

表 6-1 GCS 系统策略

策略名称	描述	策略类型	依赖关系
GCS FullAccess	基因容器服务的管理员权限。拥有基因容器服务所有操作权限。	系统策略	依赖如下策略，请同时授予如下权限。 OBS OperateAccess：全局级策略，请在“全局区域 对象存储服务”中勾选。 对于开通企业项目的用户，如下权限由于不支持企业项目，需要在用户组下面配置： <ul style="list-style-type: none"> • AOM FullAccess：项目级策略 • EVS FullAccess：项目级权限
GCS CommonOperations	基因容器的普通用户。	系统策略	
GCS ReadOnlyAccess	基因容器服务的只读访问权限	系统策略	依赖如下策略，请同时授予如下权限。 OBS ReadOnlyAccess：全局级策略，请在“全局区域 对象存储服务”中勾选。 对于开通企业项目的用户，如下权限由于不支持企业项目，需要在用户组下面配置： <ul style="list-style-type: none"> • AOM ReadOnlyAccess：项目级策略 • EVS ReadOnlyAccess：项目级权限
GCS Administrator	基因容器服务的管理员权限。拥有基因容器服务所有操作权限。	系统角色	依赖如下策略，请同时授予如下权限。 <ul style="list-style-type: none"> • CCE FullAccess：项目级策略，在与“GCS Administrator”设置策略的同一项目中勾选。 • OBS OperateAccess：全局级策略，请在“全局区域 对象存储服务”中勾选。 • CCI FullAccess：项目级策略，在与“GCS Administrator”设置策略的同一项目中勾选。

使用 Tenant Administrator 权限说明

如果想要给GCS用户配置Tenant Administrator，需要配置项目级权限。

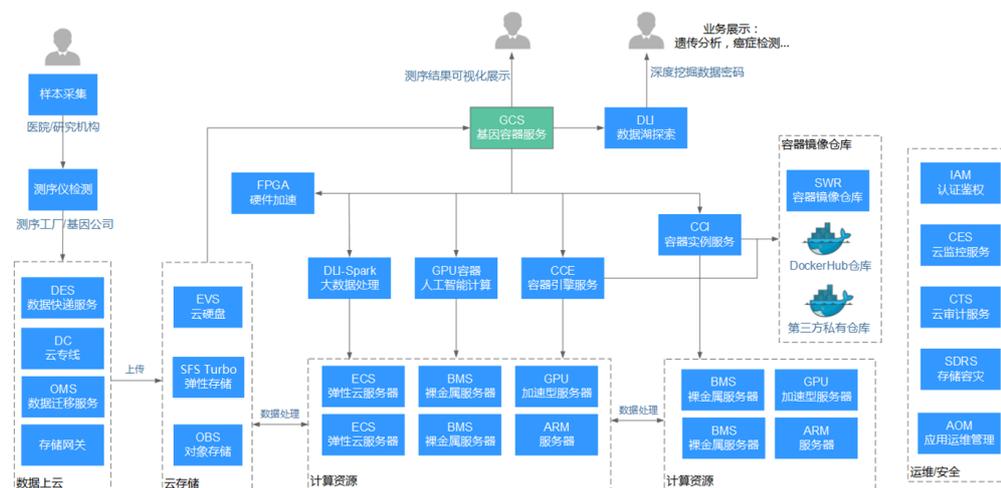
相关链接

- [IAM产品介绍](#)

7 与其它云服务的关系

基因容器服务与华为云其它服务的关系如图7-1所示。

图 7-1 基因容器与华为云服务的关系



云容器引擎 (CCE)

基因容器使用的底层资源均由CCE服务创建。云容器引擎 (Cloud Container Engine) 提供高可靠高性能的企业级容器应用管理服务，支持Kubernetes社区原生应用和工具，简化云上自动化容器运行环境搭建。应用编排服务提供编排云容器引擎的能力。

对象存储服务 (OBS)

基因容器使用对象存储服务存储您的原始数据、参考组数据、计算结果等。对象存储服务是一个基于对象的海量存储服务，为客户提供海量、安全、高可靠、低成本的数据存储能力，包括：创建、修改、删除桶，上传、下载、删除对象等。

容器镜像服务 (SWR)

基因容器使用容器镜像服务存储工具镜像，包括内置工具及自定义工具的存储。容器镜像服务 (Software Repository for Container) 是一种支持容器镜像全生命周期管理的的服务，提供简单易用、安全可靠的镜像管理功能，帮助用户快速部署容器化服务。

弹性文件服务（SFS）

基因容器服务使用弹性文件服务存放过程文件。弹性文件服务（Scalable File Service）为您的弹性云服务器（ECS）提供一个完全托管的共享文件存储，它提供标准文件协议（NFS），能够弹性伸缩至PB规模，具备可扩展的性能，为海量数据、高带宽型应用提供有力支持。

数据湖探索（DLI）

基因容器使用数据湖探索服务提供的Spark能力，对GATK、BWA等业务流程进行加速。数据湖探索（Data Lake Insight，简称DLI）是完全托管的数据分析服务，用户无需管理任何服务器，即开即用；本服务基于Spark 2.2的内存计算引擎，将支持跨源联邦查询能力，用户直接使用SQL，轻松实现对云上多数据源探索分析。

应用运维管理（AOM）

基因容器使用AOM服务提供的能力，简化环境运维，提高运维效率。华为云应用运维管理(Application Operations Management)是云上应用的一站式立体化运维管理平台，实时监控用户的应用及相关云资源，采集并关联资源的各项指标、日志及事件等数据共同分析应用健康状态，提供灵活的告警及丰富的数据可视化功能，帮助用户及时发现故障，全面掌握应用、资源及业务的实时运行状况。

统一身份认证（IAM）

统一身份认证（Identity and Access Management，简称IAM）是公有云系统（华为云）的安全管理服务，包括用户身份认证、权限分配、访问控制等功能。IAM为基因容器提供统一的身份认证、鉴权等服务。