

ALIBABA CLOUD

阿里云

阿里云最佳实践
生命科学

文档版本：20200817

 阿里云

法律声明

阿里云提醒您阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
 危险	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 危险 重置操作将丢失用户配置数据。
 警告	该类警示信息可能会导致系统重大变更甚至故障，或者导致人身伤害等结果。	 警告 重启操作将导致业务中断，恢复业务时间约十分钟。
 注意	用于警示信息、补充说明等，是用户必须了解的内容。	 注意 权重设置为0，该服务器不会再接受新请求。
 说明	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 说明 您也可以通过按Ctrl+A选中全部文件。
>	多级菜单递进。	单击设置> 网络> 设置网络类型。
粗体	表示按键、菜单、页面名称等UI元素。	在结果确认页面，单击确定。
<code>Courier</code> 字体	命令或代码。	执行 <code>cd /d C:/window</code> 命令，进入Windows系统文件夹。
<i>斜体</i>	表示参数、变量。	<code>bae log list --instanceid</code> <i>Instance_ID</i>
[] 或者 [a b]	表示可选项，至多选择一个。	<code>ipconfig [-all -t]</code>
{ } 或者 {a b}	表示必选项，至多选择一个。	<code>switch {active stand}</code>

目录

1. BCS Cromwell基因测序	05
2. BCS三代基因测序数据组装最佳实践	07

1.BCS Cromwell基因测序

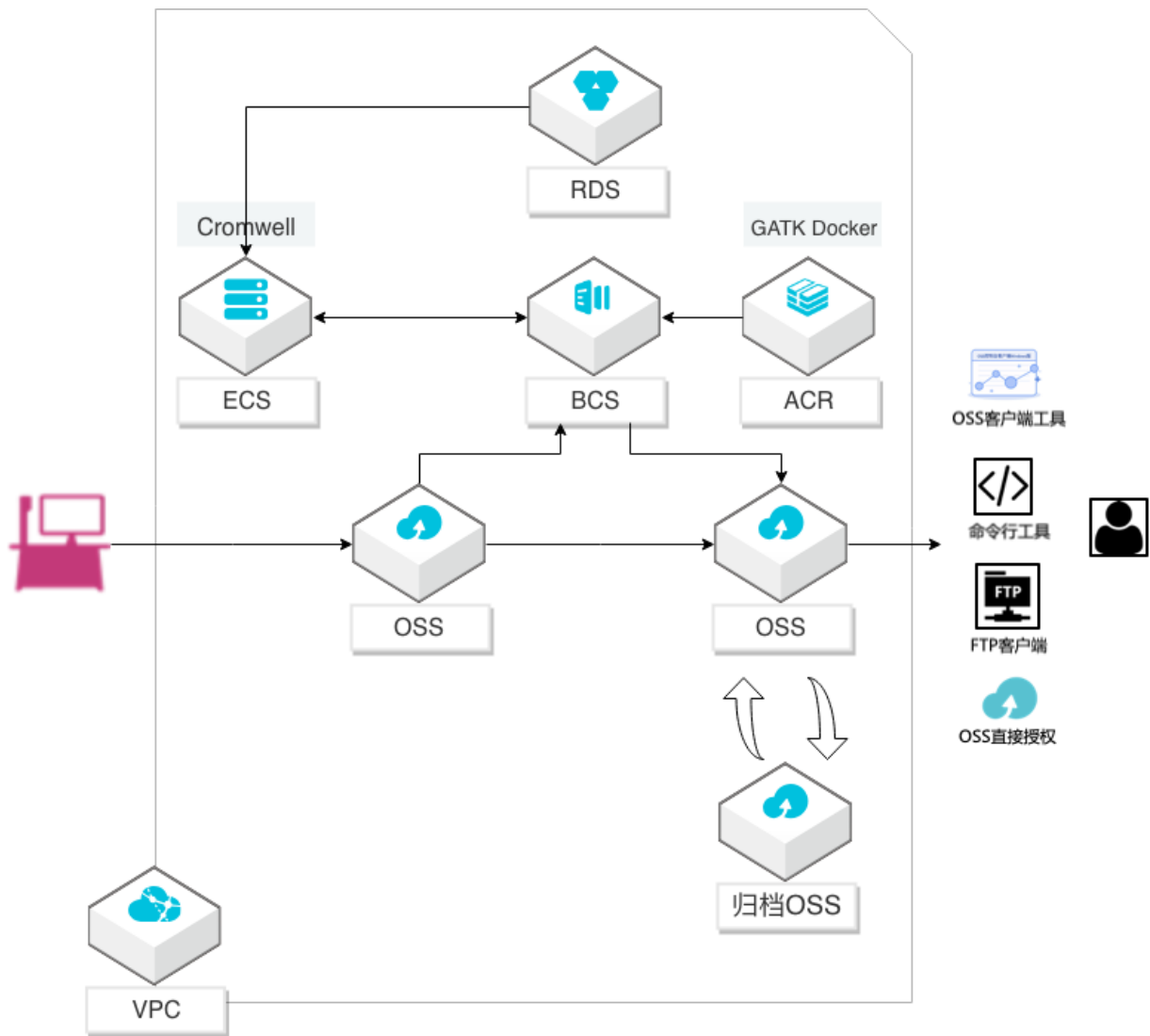
场景描述

本文介绍基于阿里云批量计算服务、Cromwell、GATK和工作流定义语言（WDL）描述的基因组数据分析流程，进行基因分析的最佳实践。

解决的问题

搭建以Cromwell为基因测序 workflow 管理平台，使用批量计算服务提供基因测序计算集群，运行以WDL定义的基因测序流程的云端基因测序平台。

部署架构图



选用的产品

- 访问控制

RAM 使您能够安全地集中管理对阿里云服务和资源的访问。您可以使用 RAM 创建和管理用户和组，并使用各种权限来允许或拒绝他们对云资源的访问。

更多关于访问控制的介绍，参见[访问控制产品详情页](#)。

- **云服务器ECS**

云服务器（Elastic Compute Service，简称ECS）是阿里云提供的性能卓越、稳定可靠、弹性扩展的IaaS（Infrastructure as a Service）级别云计算服务。云服务器ECS免去了您采购IT硬件的前期准备，让您像使用水、电、天然气等公共资源一样便捷、高效地使用服务器，实现计算资源的即开即用和弹性伸缩。阿里云ECS持续提供创新型服务器，解决多种业务需求，助力您的业务发展。

更多关于云服务器ECS的介绍，参见[云服务器ECS产品详情页](#)。

- **批量计算**

批量计算（BatchCompute）是一种适用于大规模并行批处理作业的分布式云服务。BatchCompute可支持海量作业并发规模，系统自动完成资源管理，作业调度和数据加载，并按实际使用量计费。BatchCompute广泛应用于电影动画渲染、生物数据分析、多媒体转码、金融保险分析、科学计算等领域。

更多关于批量计算的介绍，参见[批量计算产品详情页](#)。

- **专有网络VPC**

专有网络VPC帮助您基于阿里云构建出一个隔离的网络环境，并可以自定义IP地址范围、网段、路由表和网关等；此外，也可以通过专线/VPN/GRE等连接方式实现云上VPC与传统IDC的互联，构建混合云业务。

更多关于专有网络VPC的介绍，参见[专有网络VPC产品详情页](#)。

- **对象存储OSS**

阿里云对象存储服务（Object Storage Service，简称OSS），是阿里云提供的海量、安全、低成本、高可靠的云存储服务。其数据设计持久性不低于99.999999999%（12个9），服务设计可用性（或业务连续性）不低于99.995%。

更多关于对象存储OSS的介绍，参见[对象存储OSS产品详情页](#)。

- **云数据库RDS**

阿里云关系型数据库（Relational Database Service，简称RDS）是一种稳定可靠、可弹性伸缩的在线数据库服务。基于阿里云分布式文件系统和SSD盘高性能存储，RDS支持MySQL、SQL Server、PostgreSQL、PPAS（PostgreSQL Plus Advanced Server，高度兼容Oracle数据库）和MariaDB TX引擎，并且提供了容灾、备份、恢复、监控、迁移等方面的全套解决方案，彻底解决数据库运维的烦恼。

更多关于云数据库RDS的介绍，参见[云数据库RDS产品文档](#)。

- **容器镜像服务ACR**

提供安全的镜像托管能力，稳定的国内外镜像构建服务，便捷的镜像授权功能，方便用户进行镜像全生命周期管理。

更多关于容器镜像服务ACR的介绍，参见[容器服务ACR产品详情页](#)。

详细信息

[点击查看最佳实践详情页链接](#)

更多最佳实践

[点击查看更多阿里云最佳实践](#)

2. BCS三代基因测序数据组装最佳实践

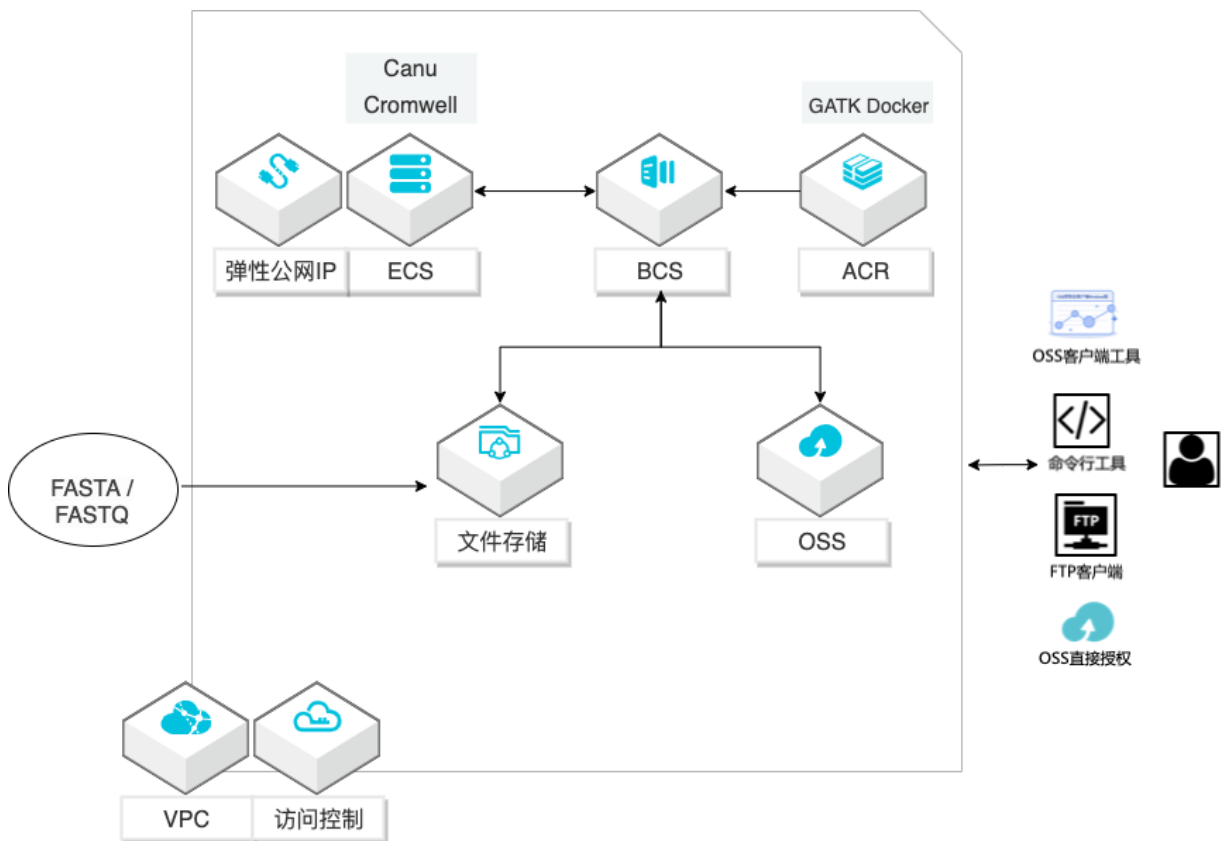
场景描述

本文介绍如何基于批量计算服务提供的WDL- Canu解决方案，进行三代基因组组装的最佳实践。

解决的问题

- 不更改Canu的使用习惯，基于云基础设施环境在云端进行三代组装。
- 充分利用云端基础设施弹性、灵活的优势。
- 以Workflow Description Language (WDL) 优化重构Canu工作流。
- 根据Canu全流程中的不同任务，批量计算启动适配任务需求的弹性伸缩集群大幅提高资源利用率及组装效率。

部署架构图



选用的产品

• 弹性公网IP (EIP)

独立的公网IP资源，可以绑定到阿里云专有网络VPC类型的ECS、NAT网关、私网负载均衡SLB上，并可以动态解绑，实现公网IP和ECS、NAT网关、SLB的解耦，满足灵活管理的要求。

更多关于弹性公网IP (EIP) 的介绍，参见[弹性公网IP \(EIP\) 产品详情页](#)。

• 云服务器ECS

云服务器（Elastic Compute Service，简称ECS）是阿里云提供的性能卓越、稳定可靠、弹性扩展的IaaS（Infrastructure as a Service）级别云计算服务。云服务器ECS免去了您采购IT硬件的前期准备，让您像使用水、电、天然气等公共资源一样便捷、高效地使用服务器，实现计算资源的即开即用和弹性伸缩。阿里云ECS持续提供创新型服务器，解决多种业务需求，助力您的业务发展。

更多关于云服务器ECS的介绍，参见[云服务器ECS产品详情页](#)。

- **专有网络VPC**

专有网络VPC帮助您基于阿里云构建出一个隔离的网络环境，并可以自定义IP地址范围、网段、路由表和网关等；此外，也可以通过专线/VPN/GRE等连接方式实现云上VPC与传统IDC的互联，构建混合云业务。

更多关于专有网络VPC的介绍，参见[专有网络VPC产品详情页](#)。

- **访问控制**

RAM使您能够安全地集中管理对阿里云服务和资源的访问。您可以使用RAM创建和管理用户和组，并使用各种权限来允许或拒绝他们对云资源的访问。

更多关于访问控制的介绍，参见[访问控制产品详情页](#)。

- **批量计算**

批量计算（BatchCompute）是一种适用于大规模并行批处理作业的分布式云服务。BatchCompute可支持海量作业并发规模，系统自动完成资源管理，作业调度和数据加载，并按实际使用量计费。BatchCompute广泛应用于电影动画渲染、生物数据分析、多媒体转码、金融保险分析、科学计算等领域。

更多关于批量计算的介绍，参见[批量计算产品详情页](#)。

- **文件存储NAS**

阿里云文件存储NAS是一个可共享访问，弹性扩展，高可靠，高性能的分布式文件系统。兼容POSIX文件接口，可支持数千台计算节点共享访问，可以挂载到弹性计算ECS、神龙裸金属、容器服务ACK、弹性容器ECI、批量计算BCS、高性能计算EHPC，AI训练PAI等计算业务上提供高性能的共享存储，用户无需修改应用程序，即可无缝迁移业务系统上云。

更多关于文件存储NAS的介绍，参见[文件存储NAS产品详情页](#)。

- **对象存储OSS**

阿里云对象存储服务（Object Storage Service，简称OSS），是阿里云提供的海量、安全、低成本、高可靠的云存储服务。其数据设计持久性不低于99.999999999%（12个9），服务设计可用性（或业务连续性）不低于99.995%。

更多关于对象存储OSS的介绍，参见[对象存储OSS产品详情页](#)。

详细信息

[点击查看最佳实践详情](#)

更多最佳实践

[点击查看更多阿里云最佳实践](#)