ALIBABA CLOUD

阿里云

阿里云最佳实践

机器学习&人工智能

文档版本：20210105

阿里云

法律声明

阿里云提醒您在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**。
Courier字体	命令或代码。	执行 `cd /d C:/window` 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	`bae log list --instanceid` *Instance_ID*
[] 或者 [a|b]	表示可选项，至多选择一个。	`ipconfig [-all|-t]`
{} 或者 {a|b}	表示必选项，至多选择一个。	`switch {active|stand}`

目录

1.混合云使用飞天AI加速工具

混合云场景自建K8S集群使用飞天AI加速工具提升训练和推理的性能。

直达最佳实践

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

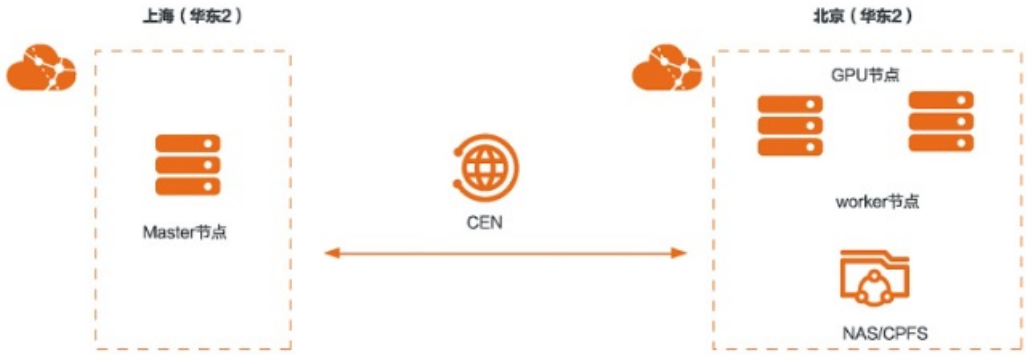
场景描述

本文介绍了混合云场景中，自建 Kubernetes服务，线下集群+云上弹性扩展阿里云GPU服务实例+飞天AI加速工具，并采用阿里云CPFS存储，运行AI训练+AI推理作业的操作步骤。

解决的问题

- 利用云企业网打通两个地域的VPC，自建Kubernetes集群。
- 使用飞天AI加速工具运行训练和推理作业。
- 使用CPFS存储共享数据。

部署架构图



2.智能媒体管理人脸语音识别

使用阿里云智能媒体管理服务(IMM)以及OSS进行在线教育视频智能分析。

直达最佳实践

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

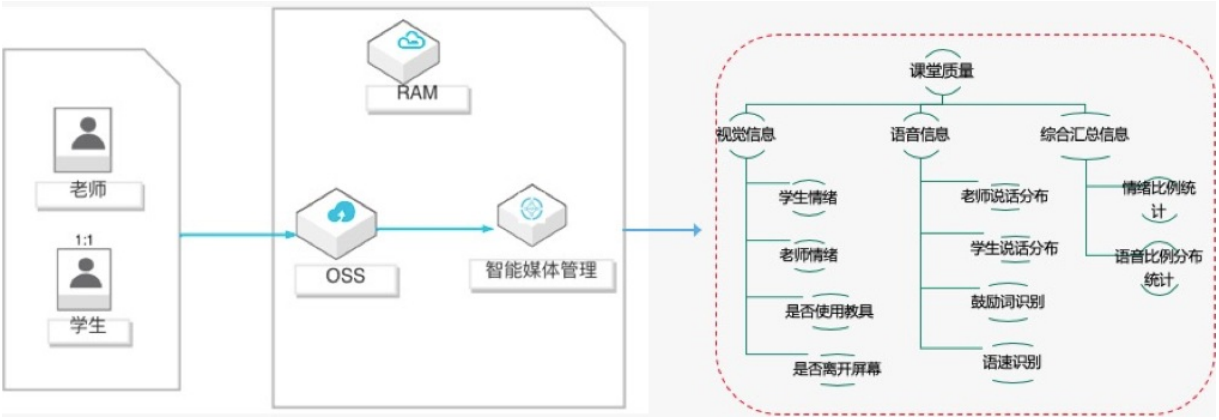
场景描述

阿里云智能媒体管理产品(IMM)及对象存储OSS以及访问控制RAM进行人脸语音识别等AI分析从而进行在线教育视频质量分析等AI智能分析场景。

解决的问题

- 视频智能分析用于不同业务目的如在线教育质量分析
- 智能表情识别分析
- 智能语音识别分析

部署架构图



3.GPU AI模型训练

搭建AI训练的容器环境，利用飞天AI加速工具进行AI模型训练加速。

直达最佳实践

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

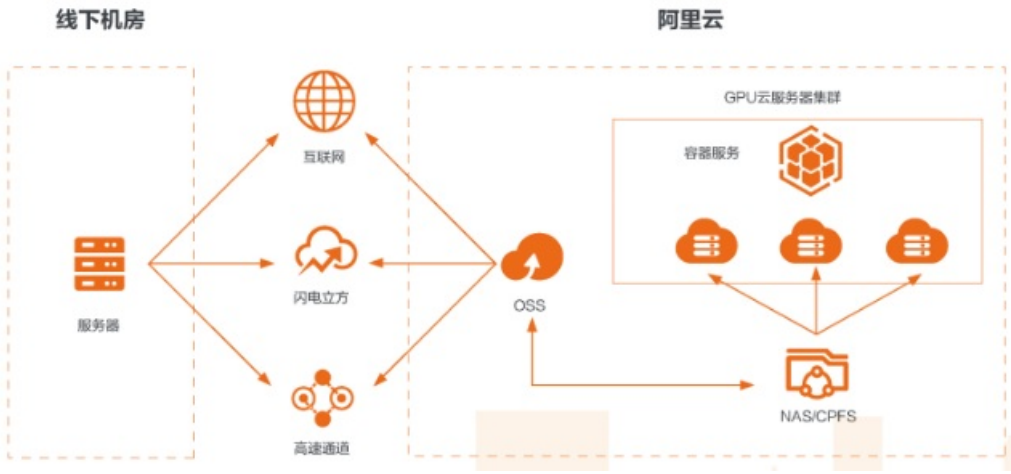
场景描述

本方案适用于AI图片训练场景，使用CPFS/NAS作为共享存储，利用容器服务Kubernetes版管理GPU云服务器集群进行图片AI训练。

解决的问题

- 搭建AI图片训练基础环境
- 使用CPFS存储训练数据
- 使用飞天AI加速训练服务加速训练
- 使用Arena一键提交作业

部署架构图



4.RAPIDS加速图像搜索

使用GPU实例+RAPIDS结合容器服务ACK实现明显的加速图像搜索任务的效果。

直达最佳实践

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

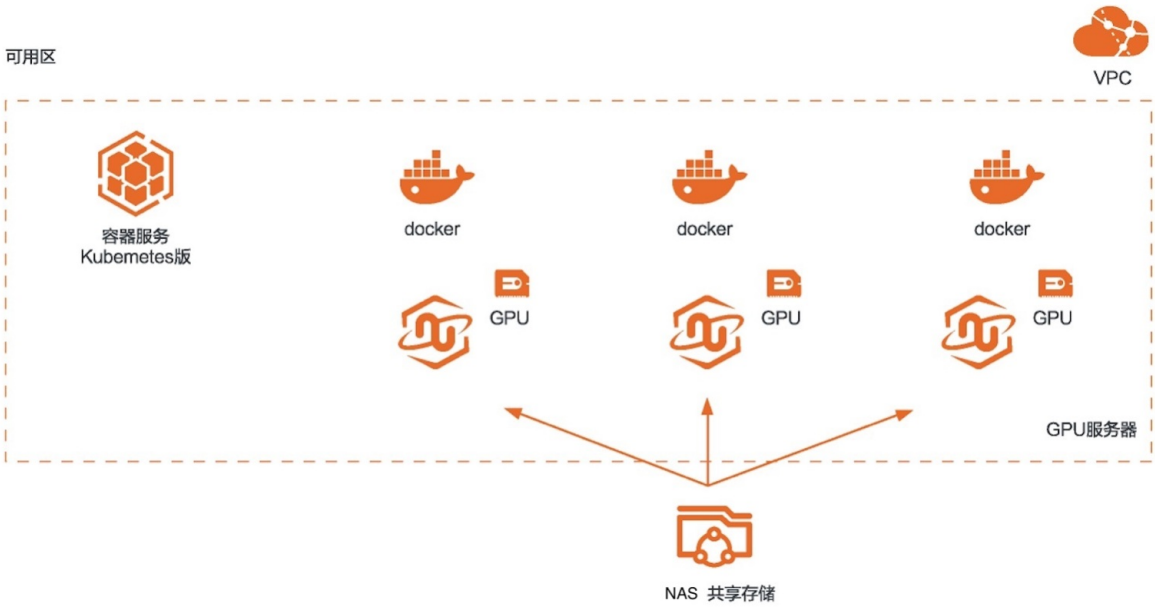
场景描述

本方案适用于使用RAPIDS加速平台+GPU云服务器来对图像搜索任务进行加速的场景。相比CPU，利用GPU+RAPIDS在图像搜索场景下可以取得非常明显的加速效果。

解决的问题

- 搭建RAPIDS加速图像搜索环境
- 使用容器服务Kubernetes版部署图像搜索环境
- 使用NAS存储计算数据

部署架构图



5.函数计算AI推理

本文介绍通过函数计算服务（Serverless服务）来部署AI推理（CPU密集型）服务的最佳实践。

直达最佳实践

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

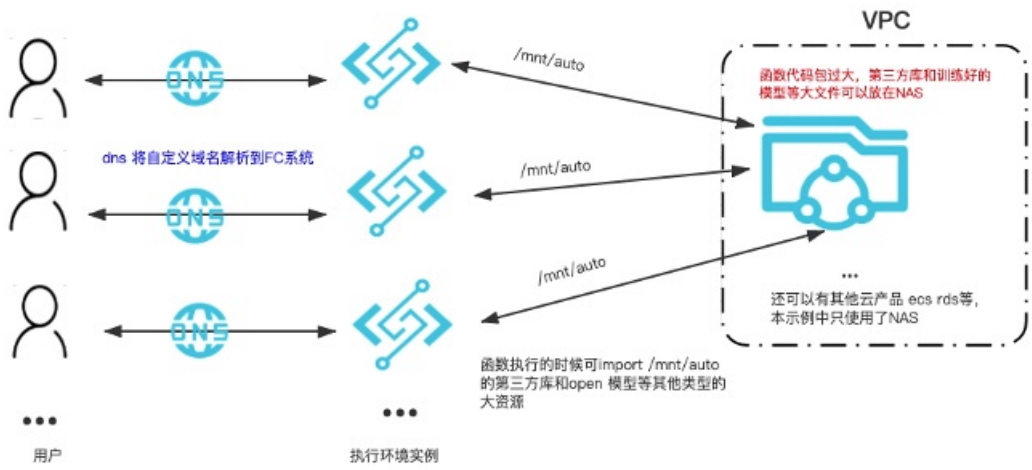
场景描述

通过函数计算服务（Serverless服务）来部署AI推理（CPU密集型）服务，达到快速部署、提升工程效率、弹性伸缩免运维、降低成本的目的。

解决的问题

- 如何使用函数计算部署AI推理服务
- 如何进行函数的压测

部署架构图



6.超级计算集群实现自然语言处理训练

使用裸金属GPU服务器、CPFS和Perseus框架，搭建NLP训练环境。

直达最佳实践

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。

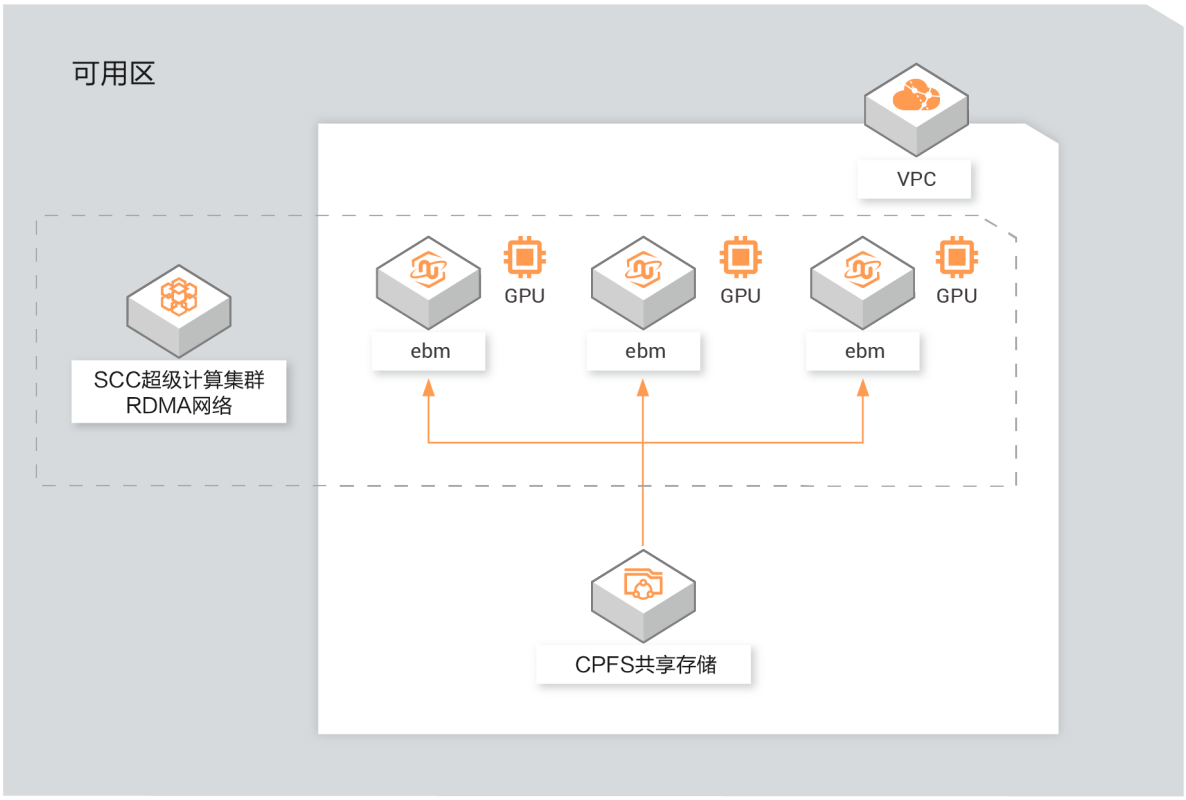
场景描述

本方案适用于自然语言训练场景，使用神龙GPU云服务器（SCCGN6）+CPFS进行NLP的训练，采用Bert模型。这里不使用容器，直接使用裸机进行NLP的Bert训练，使用飞天AI加速训练工具可以有效提升多机多卡的训练效率。

解决的问题

- 使用神龙GPU云服务器搭建NLP训练环境
- 使用SCC的RDMA网络
- 使用CPFS存储训练数据
- 使用飞天AI加速训练工具加速训练

部署架构图



7.弹性裸金属AI训练

使用弹性裸金属GPU服务器、CPFS和ACK服务，搭建高性能的AI训练架构。

直达最佳实践

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

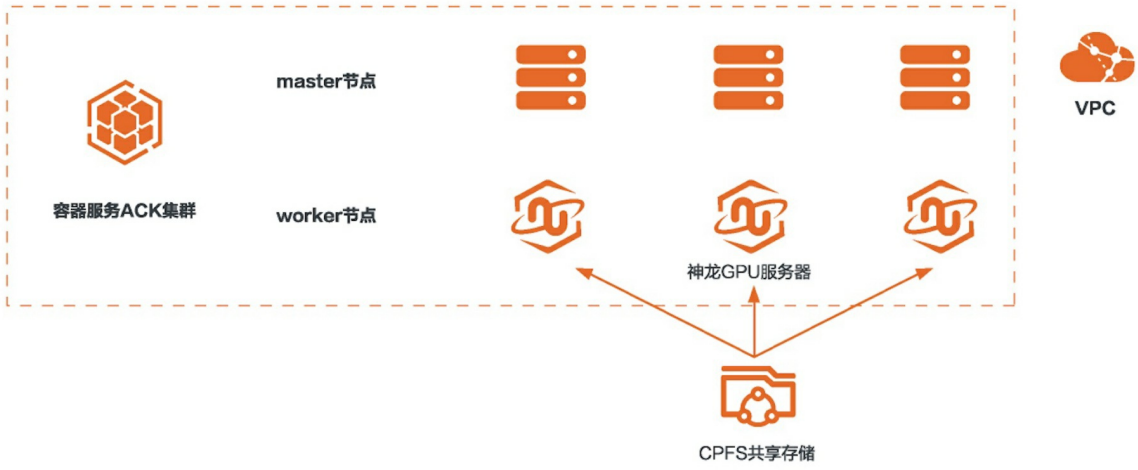
场景描述

本方案适用于AI图片训练场景，尤其是对性能要求苛刻，业务交付紧迫的场景。例如自动驾驶的模型训练（图片）等AI模型训练的场景。本方案使用了SCC超级计算集群，采用弹性裸金属GPU服务器+并行文件系统CPFS+RDMA网络+阿里云容器服务Kubernetes版+飞天AI加速训练工具，提供极致性能稳定的训练环境，保障业务能力。

解决的问题

- 搭建AI图片训练基础环境
- 使用CPFS存储训练数据
- 使用飞天AI加速训练工具加速训练
- 使用Arena一键提交作业

部署架构图



8.RAPIDS加速机器学习

使用GPU实例+RAPIDS结合ACK实现端到端加速机器学习任务加速。

直达最佳实践

点击查看最佳实践详情

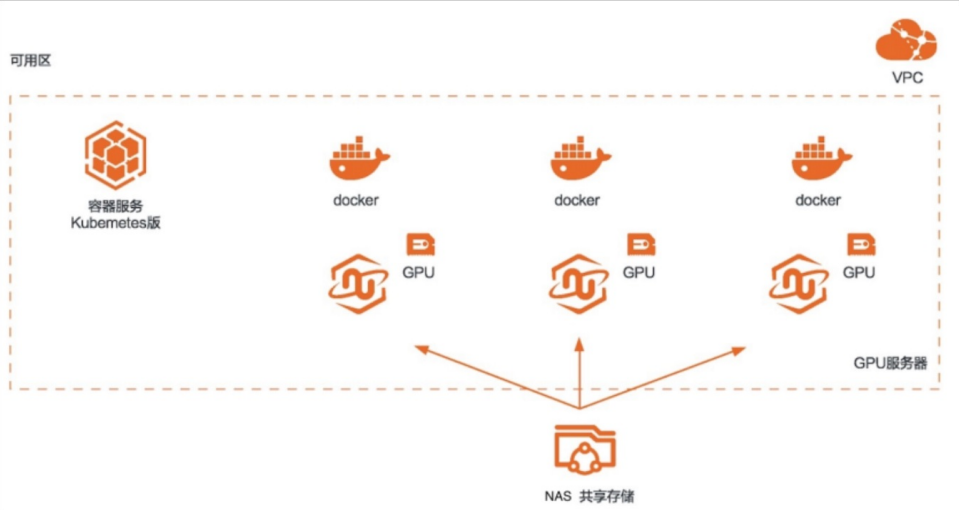
更多最佳实践

点击查看更多阿里云最佳实践

场景描述

本方案适用于使用RAPIDS加速库+GPU云服务器来对机器学习任务或者数据科学任务进行加速的场景。相比CPU，利用GPU+RAPIDS在某些场景下可以取得非常明显的加速效果。

解决架构



解决问题

- 搭建RAPIDS加速机器学习环境
- 使用容器服务Kubernetes版部署 RAPIDS环境
- 使用NAS存储计算数据

相关产品

- **容器服务 ACK**

  容器服务 Kubernetes 版（简称 ACK）提供高性能可伸缩的容器应用管理能力，支持企业级容器化应用的全生命周期管理。整合阿里云虚拟化、存储、网络和安全能力，打造云端最佳容器化应用运行环境。

  更多关于容器服务 ACK的介绍，参见容器服务 ACK产品详情页。

- **GPU服务器**

  GPU云服务器是基于GPU应用的计算服务，多适用于AI深度学习，视频处理，科学计算，图形可视化，等应用场景。

  更多关于GPU服务器的介绍，参见GPU服务器产品详情页。

- **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。