

人工智能系统 System for AI

异构计算集群调度与资源管理系统

Scheduling and resource management system

课程概览

- 异构计算集群管理系统简介
- 作业, 镜像与容器
- 调度
- 代表性异构计算集群管理系统

在多租(Multi-Tenant) GPU 集群运行作业(Job)

拥有 10 GPU 配额



提交TF Job

拥有 20 GPU 配额

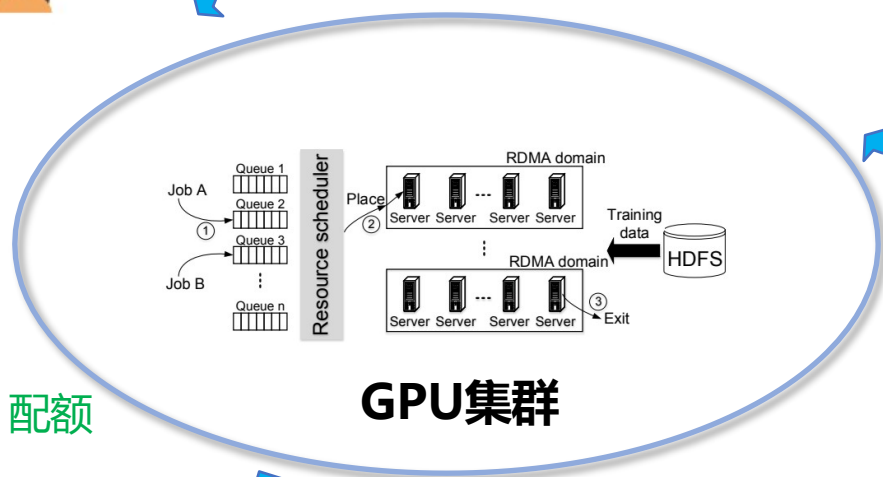


提交PyTorch Job

拥有 100 GPU 配额



提交MXNet Job



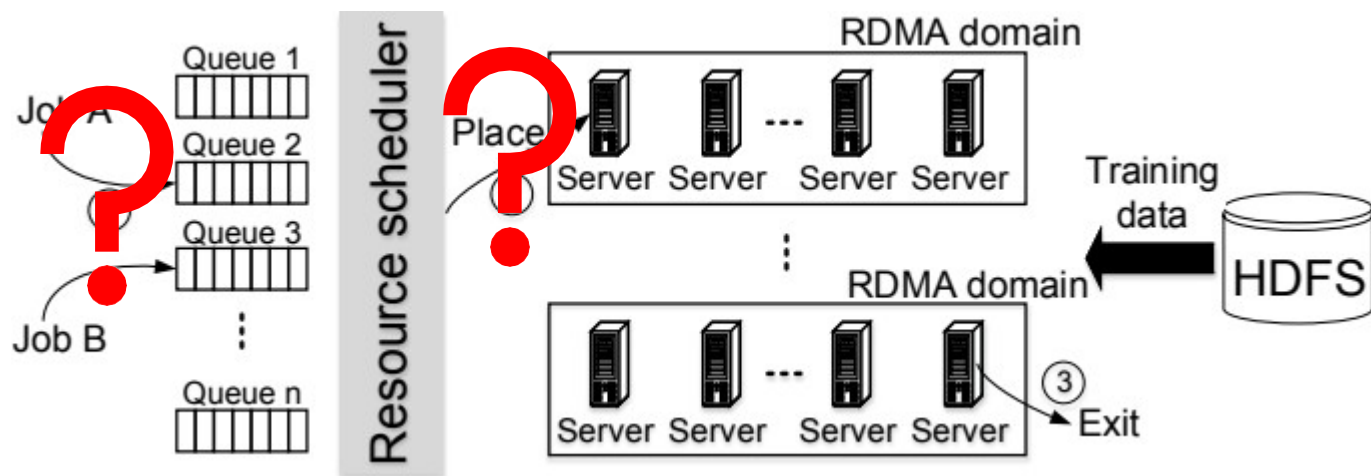
多用户共享多GPU服务器

- 多作业(Job), 多用户
- 作业环境需求多样
- 作业资源需求多样
- 服务器软件环境单一
- 服务器空闲资源多样

调度与资源管理系统重要性

- 提供人工智能基础架构支持
 - 深度学习作业(Job)调度(Scheduling)与管理
 - 异构硬件管理
- 提升生产力
 - 用户专注于模型创新，无需关注系统部署，管理
 - 模型，代码和数据共享，加速研究与创新

深度学习作业(Job)的生命周期



GPU集群

- 如何提交作业与解决环境依赖问题？
- 如何高效调度作业并分配资源？
- 如何将启动的作业运行时资源与命名空间隔离？

1. **作业提交与排队**
2. **作业资源分配与调度**
3. **作业执行完成与释放**

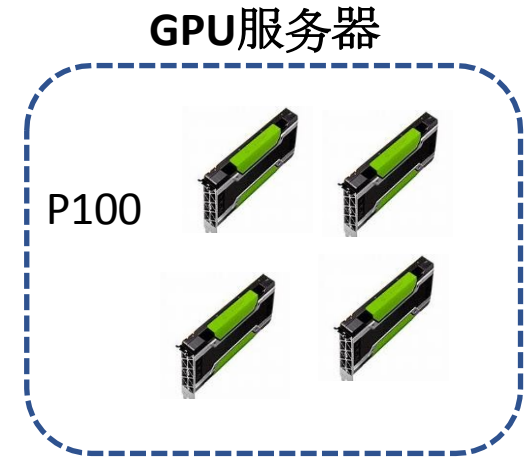
作业，镜像与容器

- 深度学习作业
- 镜像
- 容器

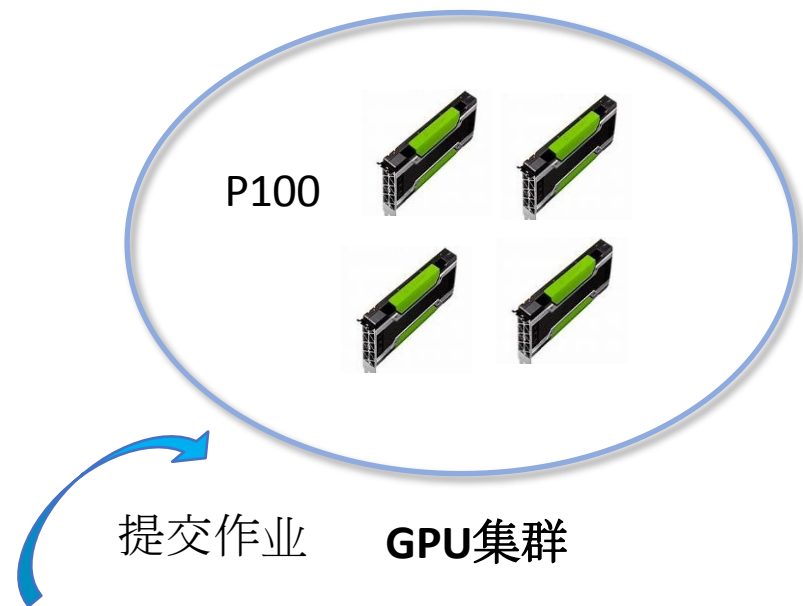
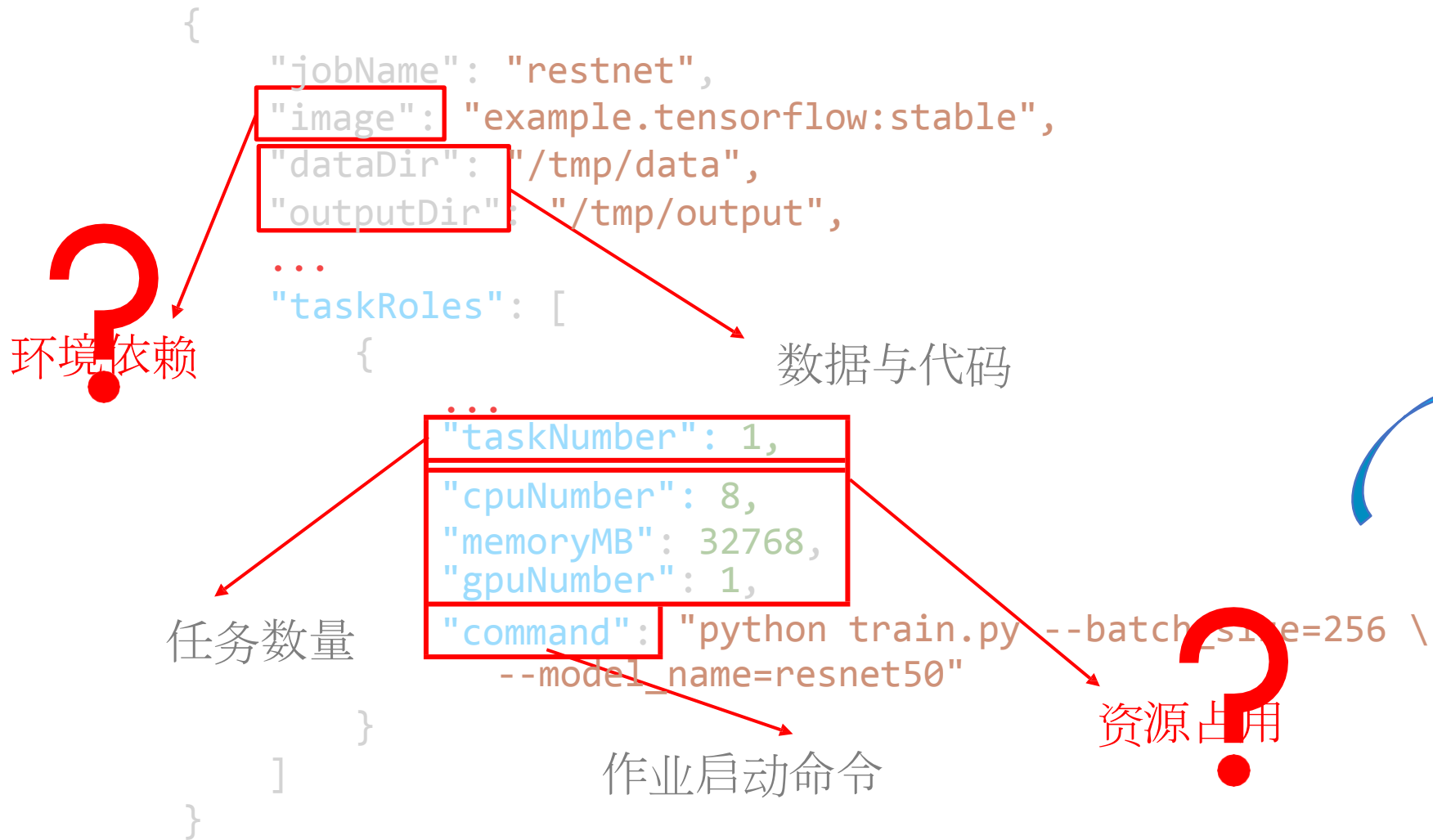
独占服务器执行深度学习作业

- 独占环境，无需考虑环境，资源隔离问题
- 环境依赖路径: 本地/anaconda3
- GPU环境依赖: 本地/usr/local/cuda
- 数据路径: 本地/data
- 直接执行启动脚本:

```
python train.py --batch_size=256 --model_name=resnet50
```



作业提交到平台



作业环境依赖问题

- 问题：
 - 服务器上没有预装好所需要的个性化环境?
 - 不同作业需要的框架，依赖和版本不同，安装繁琐且重复？
 - 部署服务器上可能会有大量重复安装的库，占用空间？
- 深度学习特有问题：
 - 深度学习作业需要安装CUDA依赖和深度学习框架等
- 目标：
 - 复用整体安装环境并创建新环境
 - 层级构建依赖，复用每一层级的依赖

作业运行时资源隔离问题

- 问题：
 - 集群资源被共享，如何保证作业互相之间不干扰和多占用资源？
 - 如何能够让不同作业可以运行不同的操作系统和命名空间？
 - 如何保证隔离的同时，作业启动的越快越好？
- 深度学习特有问题：
 - GPU和GPU memory如何隔离
- 目标：
 - 资源隔离
 - 轻量级启动

镜像与容器 - Docker

- Image

- 打包作业环境依赖

- Registry

- 共享镜像

- Container

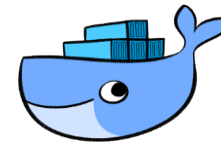
- 运行时作业资源管理与隔离

Docker Registry

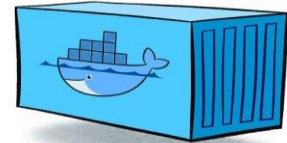


push ↑ pull ↓

build →



run →



```
FROM ubuntu:16.04
RUN apt-get update
RUN apt-get install -y vim
CMD ["vim"]
```

Dockerfile

Docker Image

Docker Container

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：

<https://d.book118.com/238111010011007030>