

容器集群管理云平台 Cluster as a service

由美国Google+AWS原生容器集群cluster management团队打造

深度学习框架TensorFlow在Kubernetes上的实践

郑泽宇才云科技首席科学家 zeyu@caicloud.io



促进软件开发领域知识与创新的传播



关注InfoQ官方信息

及时获取QCon软件开发者 大会演讲视频信息



[北京站] 2016年12月2日-3日

咨询热线: 010-89880682



[北京站] 2017年4月16日-18日

咨询热线: 010-64738142

自我介绍













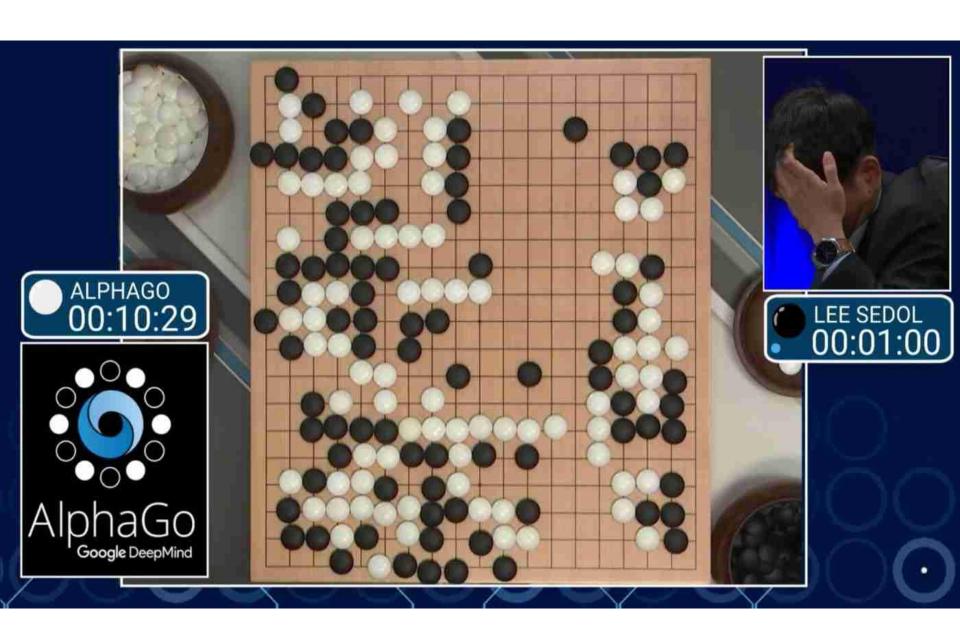
竞赛金牌

十佳论文 生

西贝尔奖学金

台歌电商

f 席科学家









郑泽宇才云科技 caicloud.io

提纲

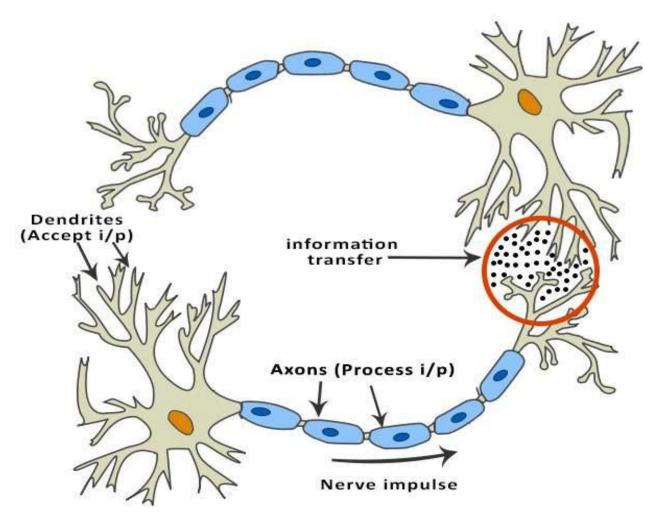


神经网络与深度学习简介

TensorFlow简介

TensorFlow on Kubernetes

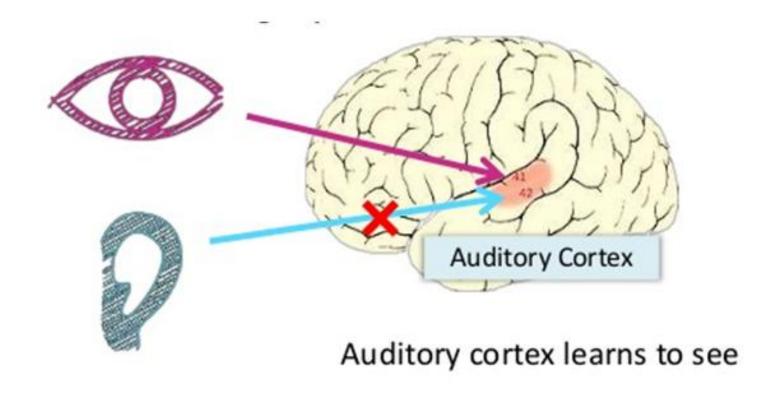




郑泽宇 才云科技 caicloud.io

通用学习模型假设





通用学习模型假设



BrainPort project —— 通过舌头"看"世界

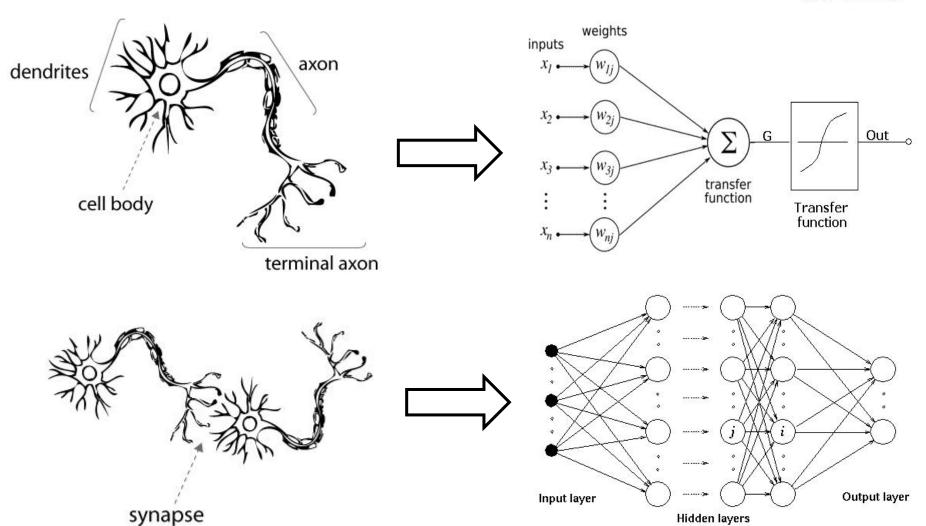






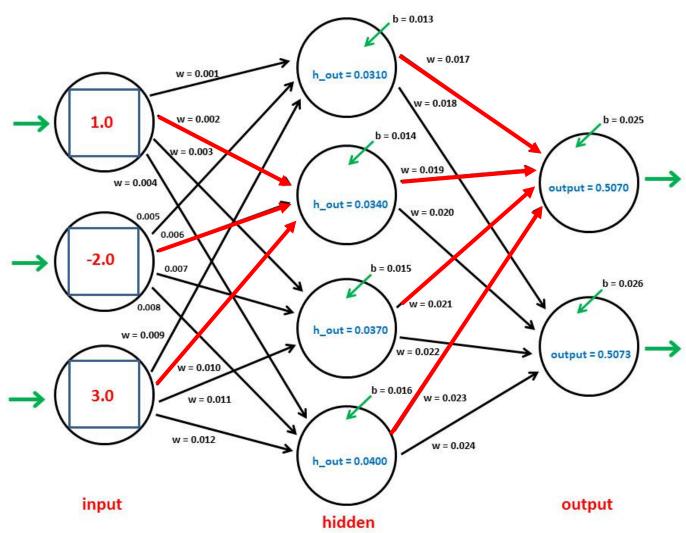






郑泽宇才云科技 caicloud.io

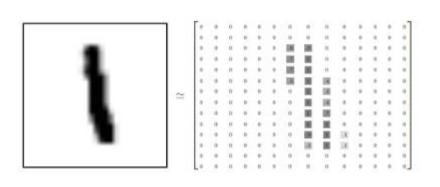


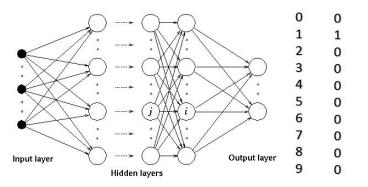


13

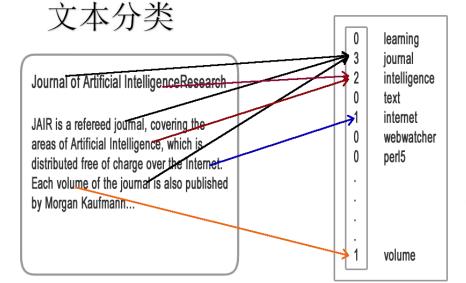


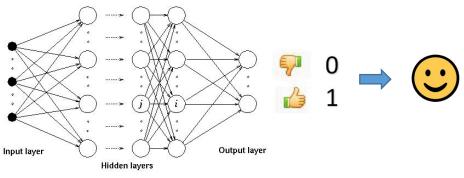
图像识别











深度学习——图像处理





深度学习——图像处理

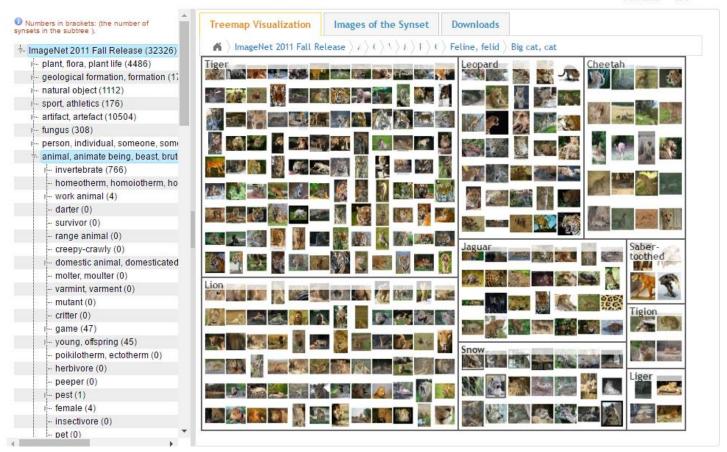


Big cat, cat

Any of several large cats typically able to roar and living in the wild

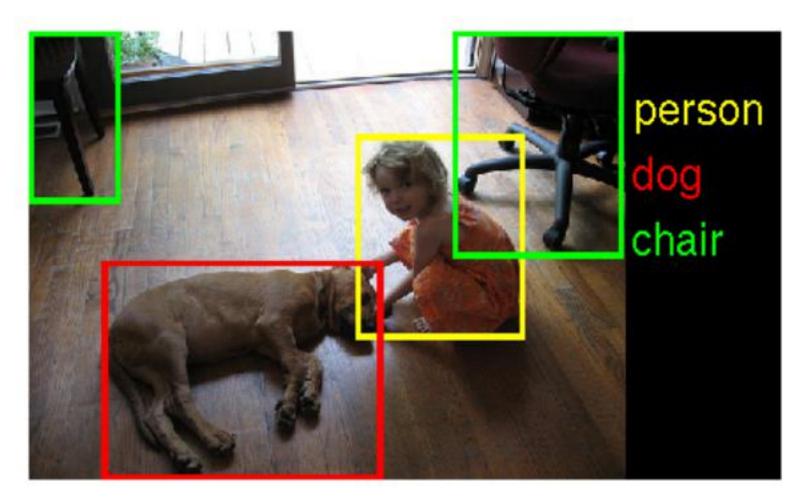
1404 pictures 93.35% Popularity Percentile





图像识别——图像分类





郑泽宇才云科技 caicloud.io

图像识别——图像分类





郑泽宇 才云科技 caicloud.io

提纲



神经网络与深度学习简介

TensorFlow简介

TensorFlow on Kubernetes

TensorFlow简介



灵活性

支持分布式

移植性

支持多语言

自动求导

兼顾科研和 产品

TensorFlow简介

























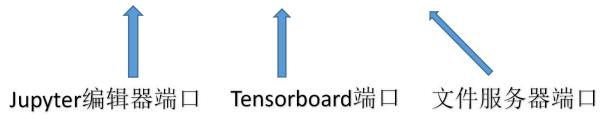


TensorFlow on Docker



• 运行镜像

docker run -p 8888:8888 -p 6006:6006 -p 8000:8000 index.caicloud.io/tensorflow:0.11.0



• 使用才云公有云服务





jupyter



TensorFlow Hello World



```
import tensorflow as tf

sess = tf.InteractiveSession()

with tf.name scope('input'):
    input1 = tf.constant([1.0, 2.0, 3.0], name="input1")

    input2 = tf.Variable(tf.random_uniform([3]), name="input2")

tf.initialize_all_variables().run()

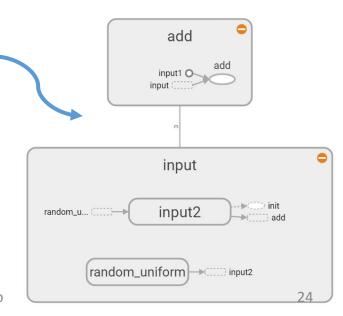
with tf.name_scope("add"):
    output = tf.add(input1, input2, name="add")

writer = tf.train.SummaryWriter("/log/demo-hello", sess.graph)

print output.eval()
```

所有结果都需要先运行才能获取

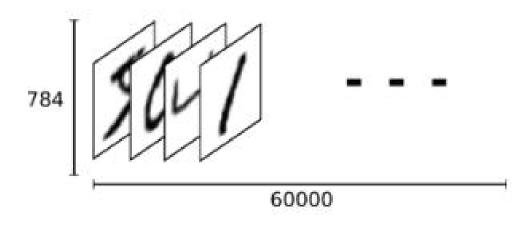
通过Sessions维护上下文, 所有执行都需要通过Session 数据都存储在"tensor"中 变量都存储在"Variable"中 在运行前所有变量都需要初始化



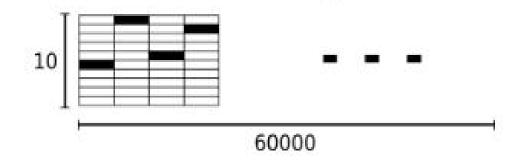
TensorFlow解决MNIST问题



mnist.train.images



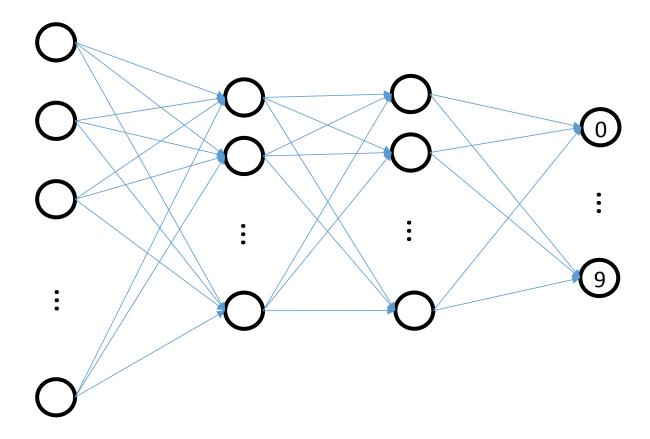
mnist.train.labels



TensorFlow解决MNIST问题



784个输入节点 200隐藏节点 200隐藏节点 10个输出节点



TensorFlow更高层封装



```
classifier = learn.TensorFlowDNNClassifier(
    hidden_units=[200, 200],
    n_classes=10, steps=5000)
```

classifier.fit(mnist.train.images, transform(mnist.train.labels))

```
score = metrics.accuracy_score(
    transform(mnist.test.labels),
    classifier.predict(mnist.test.images))
```

TensorFlow实现神经网络



- 全连接层
 - tf.nn.relu(tf.matmul(input, weights) + biases)
- 卷积层
 - tf.nn.conv2d(input, conv_weights, strides=[1, 1, 1, 1])
- 池化层
 - tf.nn.max_pool(input, ksize=[1, 2, 2, 1], strides=[1, 2, 2, 1], padding='SAME')
- LSTM结构 + 循环神经网络
 - lstm_cell = tf.nn.rnn_cell.BasicLSTMCell(hidden_nodes)
 - cell = tf.nn.rnn_cell.MultiRNNCell([lstm_cell] * layers)

TensorFlow Demo



https://github.com/caicloud/ tensorflow-demo

提纲



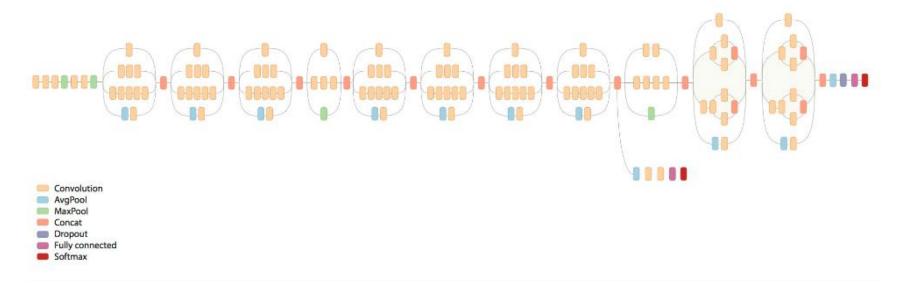
神经网络与深度学习简介

TensorFlow简介

TensorFlow on Kubernetes

没有免费午餐——计算量问题



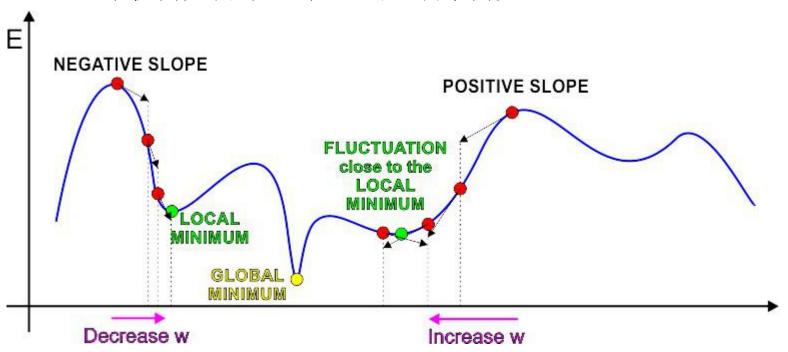


- Inception-v3 model for ImageNet
 - 2500万参数
 - 每次inference/forward prorogate 50亿 乘法/加法 操作

没有免费午餐——优化问题



- 神经网络结构复杂,难以直接求解
 - 迭代优化算法——梯度下降法
 - 需要海量数据以及海量计算
 - 单机需要训练达到78%的正确率需要? 5个月!



Kubernetes to the rescue



谷歌内部——Borg

- Google Brain跑在数十万台机器上
- 谷歌电商商品分类深度学习模型跑在1000+台机器上

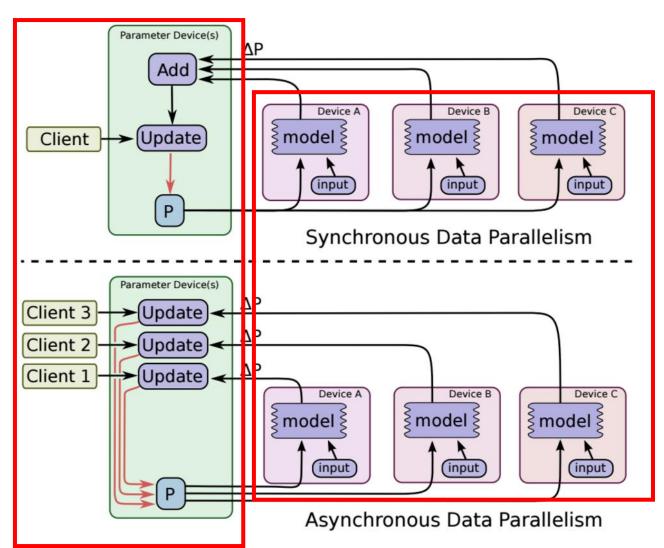
谷歌外部——Kubernetes

- Kubernetes为Borg的开源版,是一个容器集群管理系统
- Tensorflow原生态支持并行化的跑在kubernetes上

TensorFlow并行模式





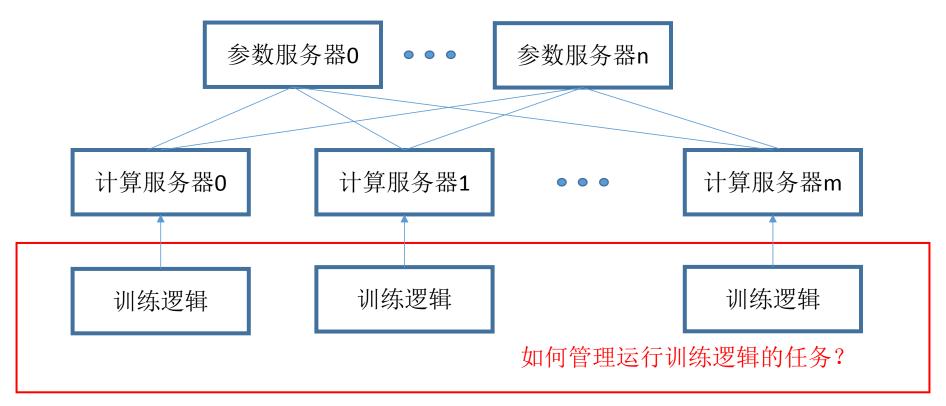


计算服务器

TensorFlow集群



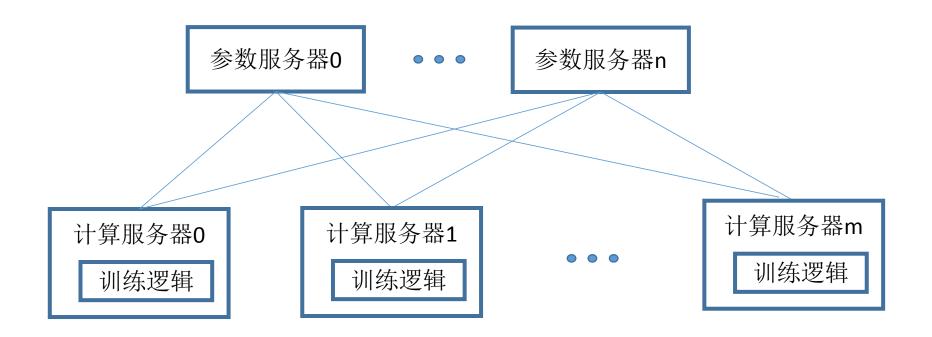
- 任务提交模式
 - 计算服务器一直等待任务提交



TensorFlow集群



• 将任务直接跑在计算服务器中



TensorFlow集群



• 配置TensorFlow集群

```
cluster = tf.train.ClusterSpec({
    "ps " : "tf-ps0:2222,tf-ps1:22222",
    "worker": "tf-worker0:2222, tf-worker2:2222"})
```

• 启动TensorFlow集群

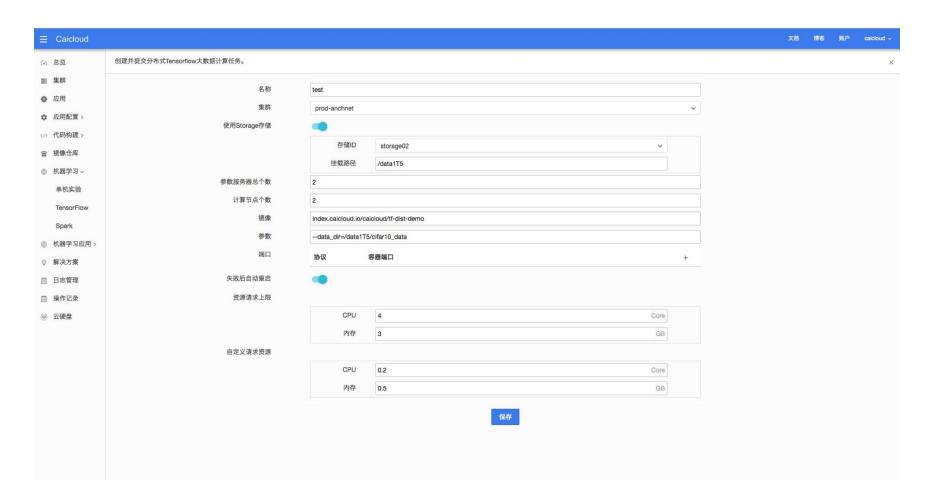
server = tf.train.Server(cluster, job_name='ps', task_index=0)

TensorFlow on Kubernetes

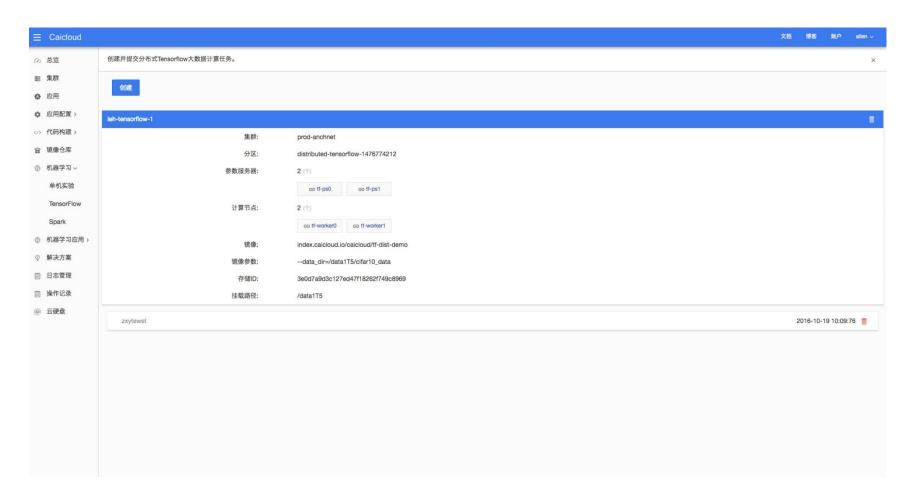


- 集群管理
 - 通过Kubernetes的DNS机制设置服务器地址
 - 通过replica controller控制失败重启
 - Kubernetes提供监控、调度等功能
- 生命周期管理
 - 目前不会自动结束
 - 难以区分正常结束还是异常退出
 - 手动管理
- 存储解决方案
 - 使用nfs、ceph等分布式存储
 - HDFS

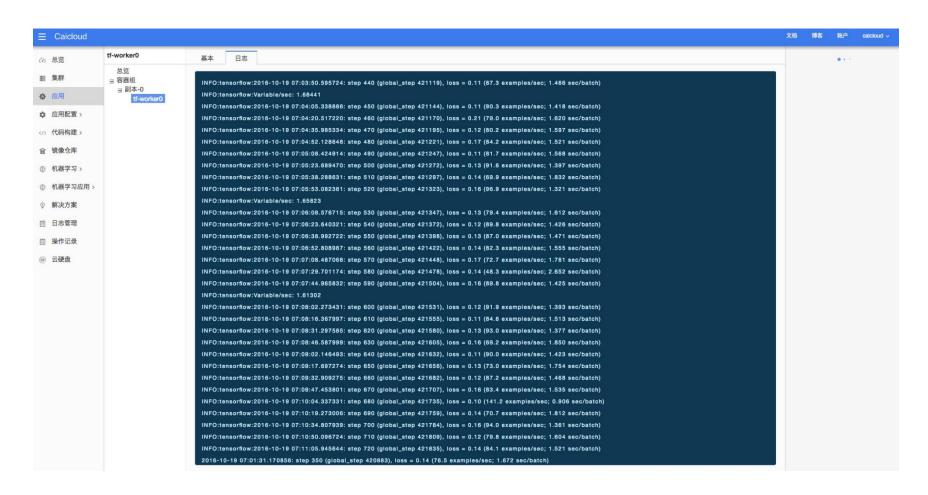




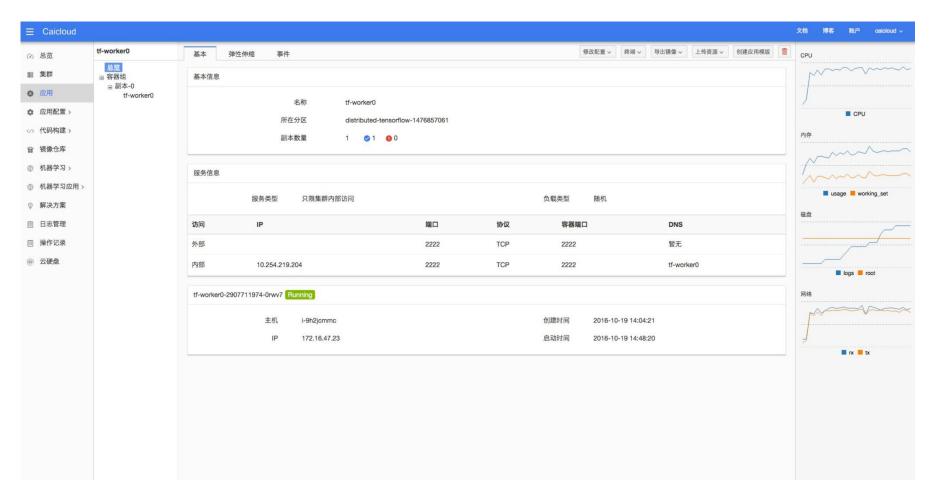












Question



