

Alibaba Cloud Container Service

Deep learning

Issue: 20190624

Legal disclaimer

Alibaba Cloud reminds you to carefully read and fully understand the terms and conditions of this legal disclaimer before you read or use this document. If you have read or used this document, it shall be deemed as your total acceptance of this legal disclaimer.

1. You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.
2. No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company, or individual in any form or by any means without the prior written consent of Alibaba Cloud.
3. The content of this document may be changed due to product version upgrades, adjustments, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and the updated versions of this document will be occasionally released through Alibaba Cloud-authorized channels. You shall pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.
4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides the document in the context that Alibaba Cloud products and services are provided on an "as is", "with all faults" and "as available" basis. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not bear any liability for any errors or financial losses incurred by any organizations, companies, or individuals arising from their download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, bear responsibility for any indirect, consequential, exemplary, incidental, special, or punitive damages, including lost profits arising from the use

or trust in this document, even if Alibaba Cloud has been notified of the possibility of such a loss.

5. By law, all the content of the Alibaba Cloud website, including but not limited to works, products, images, archives, information, materials, website architecture, website graphic layout, and webpage design, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of the Alibaba Cloud website, product programs, or content shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates).
6. Please contact Alibaba Cloud directly if you discover any errors in this document.

Generic conventions

Table -1: Style conventions

Style	Description	Example
	This warning information indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	 Danger: Resetting will result in the loss of user configuration data.
	This warning information indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.	 Warning: Restarting will cause business interruption. About 10 minutes are required to restore business.
	This indicates warning information, supplementary instructions, and other content that the user must understand.	 Notice: Take the necessary precautions to save exported data containing sensitive information.
	This indicates supplemental instructions, best practices, tips, and other content that is good to know for the user.	 Note: You can use Ctrl + A to select all files.
>	Multi-level menu cascade.	Settings > Network > Set network type
Bold	It is used for buttons, menus, page names, and other UI elements.	Click OK.
Courier font	It is used for commands.	Run the <code>cd / d C :/ windows</code> command to enter the Windows system folder.
<i>Italics</i>	It is used for parameters and variables.	<code>bae log list --instanceid Instance_ID</code>
[] or [a b]	It indicates that it is an optional value, and only one item can be selected.	<code>ipconfig [-all -t]</code>

Style	Description	Example
<code>{}</code> or <code>{a b}</code>	It indicates that it is a required value, and only one item can be selected.	<code>swich {stand slave}</code>

Contents

Legal disclaimer.....	I
Generic conventions.....	I
1 Overview.....	1
2 Environment preparations.....	2
2.1 Create a data volume.....	2
2.2 Create a container cluster.....	6
3 Prepare TensorFlow training data by using TFRecord and HDFS.....	12

1 Overview

Based on the powerful computing capability of Alibaba Cloud, the deep learning solution provides you with an easy, open, and end-to-end deep learning service platform. This solution enables data scientists and algorithm engineers to quickly use Alibaba Cloud resources (including Elastic Compute Service (ECS) instances, GPU instances, Alibaba Cloud HPC, Object Storage Service (OSS), Elastic MapReduce, and Server Load Balancer) to perform data preparation, model development, model training, evaluation, prediction, and other tasks. This solution also easily transfers the deep learning capability to service APIs, accelerating the integration with business applications.

The deep learning solution has the following features:

- **Simple:** Lowers the threshold for building and managing the deep learning platform.
- **Efficient:** Improves the efficiency of heterogeneous computing resources, such as CPU and GPU, and provides unified user experience.
- **Open:** Supports multiple mainstream deep learning frameworks, such as TensorFlow, Keras, and MXNet, and supports custom environments.
- **Full-cycle:** Provides best practices for building end-to-end deep learning task process based on the powerful service system of Alibaba Cloud.
- **Service-oriented:** Converts the deep learning capability to services, and easily integrates with applications on the cloud.

Start to use

1. Prepare the environment.

[Create a container cluster](#). To use OSS data volumes to store data, [Create an OSSFS data volume](#).

2. [Create a Jupyter environment](#) and [Use Git to manage codes in the Jupyter environment](#).
3. Run [Standalone model training](#) or [Distributed model training](#) to export the model.
4. Use the exported model to [Use TensorFlow Serving](#).

2 Environment preparations

2.1 Create a data volume

OSSFS is a FUSE-based file system officially provided by Alibaba Cloud (click <https://github.com/aliyun/ossfs> to view the project home page). OSSFS data volumes can package Object Storage Service (OSS) buckets as data volumes.

The performance and functions of OSSFS differ from those of local file systems because data must be synchronized to the cloud by means of network. We recommend that you do not run I/O-intensive applications such as databases or applications that require constantly rewriting files such as logs on OSSFS. OSSFS is applicable to scenarios such as sharing configuration files among containers and attachment upload that do not require rewriting.

OSSFS differs from local file systems in the following ways:

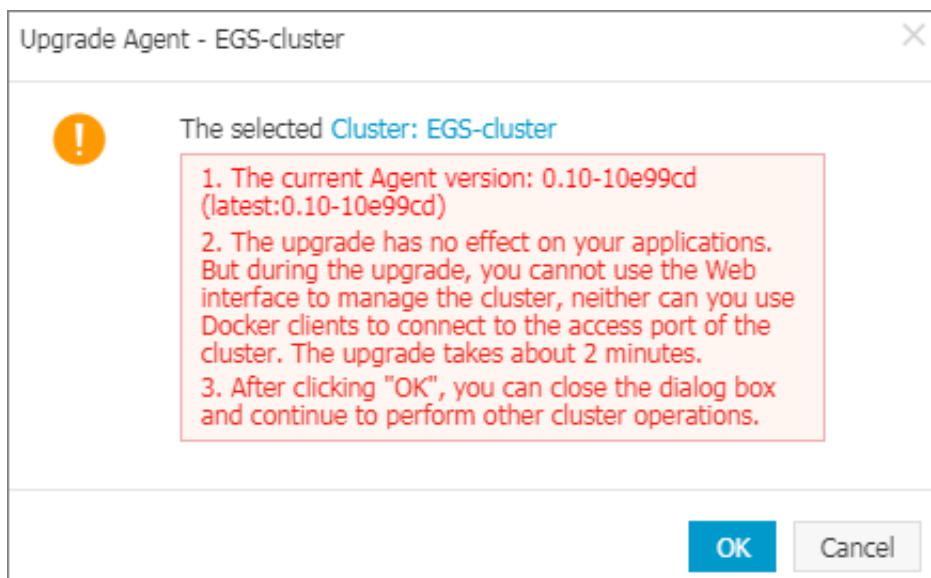
- Random write or append write leads to the entire file being overwritten.
- Metadata operations, such as list directory, provide poor performance because the system needs to remotely access the OSS server.
- The file/folder rename operation is not atomic.
- Coordinate the actions of each client on your own when multiple clients are mounted to the same OSS bucket. For example, avoid multiple clients from writing the same file.
- Hard links are not supported.

Prerequisites

To activate the data volume function, your cluster must meet the following two conditions:

- The cluster Agent is of version 0.6 or later.

You can view your Agent version on the Cluster List page. Click [More](#) > Upgrade Agent.



If your Agent version is earlier than 0.6, upgrade the Agent. For more information about how to upgrade the Agent, see [Upgrade Agent](#).

- The acsvolumedriver application is deployed in the cluster. We recommend that you upgrade to the latest version.

You can deploy and upgrade the acsvolumedriver application by upgrading system services. For more information, see [Upgrade system services](#).



Note:

When acsvolumedriver is upgraded or restarted, containers using OSSFS data volumes are restarted, and your services are also restarted.

Procedure

Step 1. Create an OSS bucket

Log on to the [OSS console](#) and create a bucket. For more information, see [Create a bucket](#).

In this example, a bucket located in China South 1 (Shenzhen) is created.

tensorflow-sample

TypeStandard StorageRegionChina South 1 (Shenzhen)Created At06/28/2017, 14:35Delete Bucket

OverviewFilesBasic SettingsDomain NamesImage ProcessingBasic DataHotspot StatisticsAPI StatisticsObject Access Statistics

Basic Data

Data in the Overview page and Bucket Overview page is not in real time. It is delayed for two to three hours.

Storage Used

Total Used

852.2 KB

Month-On-Month 0.00%Day-On-Day 0.00%

Internet Traffic This Month

Inbound

0 Byte

Internet Traffic Last Month 0Byte

Requests This Month

PUT

0

Requests Last Month 0

Files

2

File Fragments

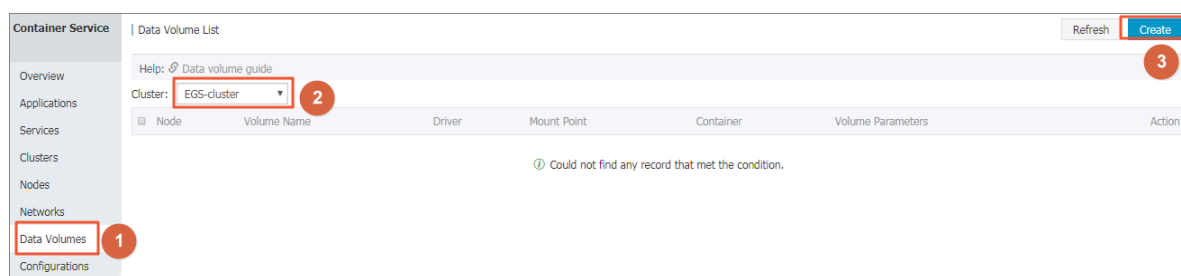
0

Access Domain Name

	Endpoint	Access Domain Name	HTTPS
Internet Access	oss-cn-shenzhen.aliyuncs.com	tensorflow-sample-oss-cn-shenzhen.aliyuncs.com	Support...
ECS Address for Classic Network Access (Intranet)	oss-cn-shenzhen-internal.aliyuncs.com	tensorflow-sample-oss-cn-shenzhen-internal.aliyuncs.com	Support...
ECS Address for VPC Network Access (Intranet)	oss-internal-cn-shenzhen.aliyuncs.com	tensorflow-sample-oss-internal-cn-shenzhen.aliyuncs.com	Support...

Step 2. Create an OSSFS data volume

1. Log on to the [Container Service console](#).
2. Click Data Volumes in the left-side navigation pane.
3. Select the cluster in which you want to create a data volume (tfoss in this example) from the Cluster drop-down list. Click Create in the upper-right corner.



4. The Create Data Volume dialog box appears. Select the Data Volume Type, as the OSS, set the data volume parameters and click Create. The Container Service creates data volumes with the same name on all nodes of the cluster.

Create Data Volume

Type: ☒ OSS ☐ Cloud Disk

Name:

Access Key ID:

Access Key Secret:

Optional Parameters: ☒ allow_other ☐ noxattr

Other Parameters:

For the formats of other parameters, refer to this document. Example: -o allow_other -o default_permission=666 -onoxattr

Note: Only clusters with volume driver version 0.7 or above support these parameters. You can go to the application list, find the acsvolumedriver application, and view the volumedriver service's image version in the service list on the application details page. If the image version is lower than 0.7, please upgrade the volumedriver.

Bucket ID: [Select Bucket](#)

Access Domain Name: ☐ Intranet ☐ Internet ☒ VPC

File Caching: ☐ Enable ☒ Disable

Create Cancel

- Name: The data volume name that must be unique in the cluster.
- Access Key ID/Access Key Secret: The AccessKey required to access OSS. You can obtain them from the [AccessKey console](#).
- Bucket ID: The name of the OSS bucket to be used. Click Select Bucket. Select the bucket (tensorflow-sample in this example) in the displayed dialog box and click Select.

- **Access Domain Name:** Select VPC.
- **File Caching:** Select Disable if you want to synchronize the modifications of the same file on multiple machines (for example, modify the file on machine A and read the modified contents on machine B).

**Note:**

Disabling the file caching slows down the ls folder, especially when many files are in the same folder. If you do not have the preceding requirement, enable the file cache to speed up `ls`.

Subsequent operations

After creating a data volume, you can use it in your application. For how to use data volumes in applications, see [Use third-party data volumes](#).

2.2 Create a container cluster

The deep learning solution supports container clusters with Elastic Compute Service (ECS) instances or GPU instances. This document uses container clusters with GPU instances as an example.

**Note:**

For how to create a container cluster with ECS instances, see [Create a cluster](#).

Limits

- Currently, Container Service only supports creating clusters with GN4 GPU instances in the following regions: China South 1 (Shenzhen), China East 2 (Shanghai), China North 2 (Beijing), and US West 1 (Silicon Valley).
- Currently, GN4 GPU instances only support Virtual Private Cloud (VPC).

Prerequisites

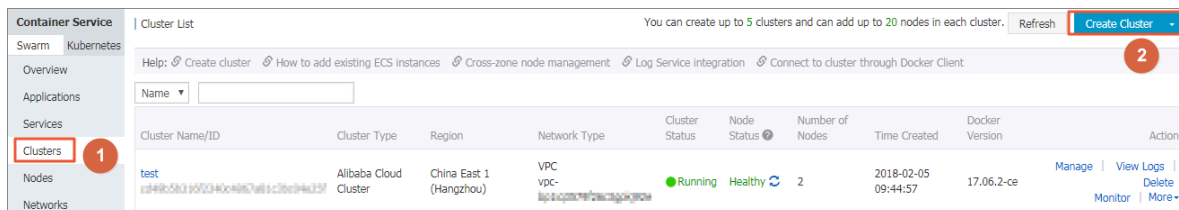
Currently, the Pay-As-You-Go GPU Compute Type GN4 instances need to be activated by [opening an ECS ticket](#) as follows:

I want to activate the Pay-As-You-Go GPU Compute Type GN4 instances. Thank you!

Procedure

1. Log on to the [Container Service console](#).

- Click **Swarm > Clusters** in the left-side navigation pane and click **Create Cluster** in the upper-right corner.



- Complete the following configurations. In this example, create a cluster named **EGS-cluster** in the region **China South 1 (Shenzhen)**.

* Cluster Name:

The cluster name should be 1-63 characters long, and can contain numbers, Chinese characters, English letters and hyphens.

Region:

China North 1 (Qingdao)	China North 2 (Beijing)	China East 1 (Hangzhou)	China East 2 (Shanghai)	China South 1 (Shenzhen)	Asia Pacific NE 1 (Tokyo)	US West 1 (Silicon Valley)	Asia Pacific SE 1 (Singapore)
Asia Pacific SE 2 (Sydney)	EU Central 1 (Frankfurt)	US East 1 (Virginia)	Hong Kong	China North 3 (Zhangjiakou)	Asia Pacific SE 3 (Kuala Lumpur)		

Zone:

- Cluster Name**: The name of the cluster to be created. It can be 1–64 characters long and contain numbers, Chinese characters, English letters, and hyphens (-).



Note:

The cluster name must be unique under the same account and the same region.

- Region**: Select the region in which the cluster will be deployed. Select **China South 1 (Shenzhen)**, **China East 2 (Shanghai)**, **China North 2 (Beijing)**, or **US West 1 (Silicon Valley)**.



Note:

Currently, Container Service only supports creating clusters with GN4 GPU instances in the following regions: **China South 1 (Shenzhen)**, **China East 2 (Shanghai)**, **China North 2 (Beijing)**, and **US West 1 (Silicon Valley)**.

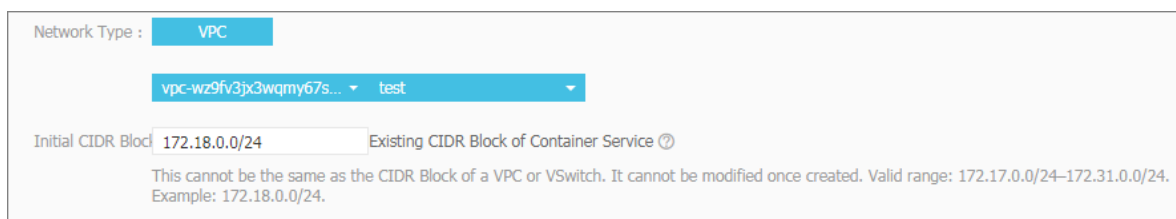
- Zone**: Select the zone for the cluster.



Note:

You can select the region and zone according to the distribution of your servers.

4. Select VPC as the Network Type and complete the configurations.



The screenshot shows a configuration form for a container network. At the top, 'Network Type' is set to 'VPC'. Below this, a dropdown menu shows 'vpc-wz9fv3jx3wqmy67s...' with a 'test' label. The 'Initial CIDR Block' is set to '172.18.0.0/24'. To the right, there is a label 'Existing CIDR Block of Container Service' with a help icon. Below these fields, a warning message states: 'This cannot be the same as the CIDR Block of a VPC or VSwitch. It cannot be modified once created. Valid range: 172.17.0.0/24–172.31.0.0/24. Example: 172.18.0.0/24.'

VPC enables you to build an isolated network environment based on Alibaba Cloud. You can have a full control over your own virtual network, including a free IP address range, Classless Inter-Domain Routing (CIDR) block division, and the configurations of route table and gateway.

Specify a VPC, a VSwitchId, and the initial CIDR block of a container (the subnet CIDR block where the Docker container belongs. For ease of IP management, each virtual machine container belongs to a different CIDR block, and container subnet CIDR block cannot conflict with virtual machine CIDR block).

We recommend that you build your own VPC/VSwitchId for the container cluster to prevent issues such as network conflicts.


5. Select whether to add nodes or not.

Add Node :

You can create a cluster with several new instances, or create a zero-node cluster and then add existing instances to the cluster. For how to add existing instances to the cluster, see [Add an existing instance](#).

· Add

a. Select the operating system for the node.

Operating System 

Currently, the supported operating systems include Ubuntu 14.04 64bit and CentOS 7.4 64bit.

b. Configure the instance specifications.

- Select Generation III as the Instance Generation, GPU Compute Type gn4 as the Instance Family,
- and 32-core, 48 GB (ecs.gn4.8xlarge) or 56-core, 96 GB (ecs.gn4.14xlarge) as the Instance Type.



Note:

If you have been approved to use the GN4 GPU instances but cannot find these two instance types, this is because no resource is currently available

for instances of these two types. We recommend that you purchase the instances again later or the next day.

The screenshot displays the ECS console configuration page. At the top, 'Instance Generation' is set to 'Generation III'. Under 'Instance Family', 'General Type n4' is selected. 'I/O Optimized' is set to 'IO optimized instance'. 'Instance Type' is '2-core, 4GB (ecs.n4.large)'. 'Instance Quantity' is '2 set(s)'. 'System Disk Type' and 'Data Disk Type' are both 'Ultra Cloud Disk'. 'Attach Data Disk' is unchecked. 'Login' is set to 'Password'. The 'Logon Password' field is filled with dots, and a confirmation field is also present. A note states: 'The password should be 8-30 characters long and contain three types of characters (uppercase/lowercase letters, numbers and special characters). During cluster creation, we will use this password for node configuration. The password will not be stored.'

You can configure the instance quantity, data disk capacity (the GPU instance has a 20 GB system disk by default), and logon password.



Note:

- The data disk is attached to the `/var/lib/docker` directory and used for the storage of Docker images and containers if you select the **Attach Data Disk** check box.
- In terms of performance and management, we recommend that you attach an independent data disk to the host and manage the persistent data in the container by using Docker volumes.

Do not Add

You can click **Add Existing Instance** to add existing instances to the cluster, or click **Add Existing Instances** on the **Cluster List** page to add existing instances to the cluster after the cluster is created.

6. Select whether to configure public Elastic IP (EIP) or not.

If you select VPC as the network type, Container Service configures an [EIP](#) for each instance under the VPC by default. If this is not required, select the **Do not Configure Public EIP** check box and then configure the SNAT gateway.

EIP : ☐ Do not Configure Public EIP

You must configure the SNAT (refer to the following documents) if a public EIP is not configured. Failure in configuring the SNAT will cause the VPC unable to access the public network. This will affect cluster creation and application deployment.
Documents for reference: [Configuring SNAT for Linux in a VPC environment to use a server proxy with EIP to access the Internet without a public network ECS instance](#)

7. Select whether to create a Server Load Balancer instance or not.

Server Load Balancer ☒ Automatically Create Server Load Balancer

A public network Server Load Balancer instance is created by default while a cluster is created. The billing method is [Pay-As-You-Go](#).

The Automatically Create Server Load Balancer check box is selected by default. With this check box selected, an Internet Server Load Balancer instance is created after the cluster is created. You can access the container applications in the cluster by using this Server Load Balancer instance. This is a Pay-As-You-Go Server Load Balancer instance.

8. Click Create Cluster.

Subsequent operations

On the Cluster List page, you can click **View Logs** at the right of the cluster to view the creation process logs of the cluster.

Cluster List									
You can create up to 5 clusters and can add up to 20 nodes in each cluster. Refresh Create Cluster									
Help: Create cluster How to add existing ECS instances Cross-zone node management Log Service integration Connect to cluster through Docker Client									
Name ▾									
Cluster Name/ID	Cluster Type	Region	Network Type	Cluster Status	Node Status	Number of Nodes	Time Created	Docker Version	Action
EGS-cluster c747b44aa29748a0963817e070e15a0d1	Alibaba Cloud Cluster	China South 1 (Shenzhen)	VPC vpc-wn9fv0jucwcmqy67am0k0	Running	Healthy	2	2018-01-17 11:17:41	17.06.2-ce	Manage View Logs Delete Monitor More ▾

3 Prepare TensorFlow training data by using TFRecord and HDFS

Data preparation and preprocessing play important roles in the deep learning and training process, and affect the speed and quality of model training.

TensorFlow supports the HDFS, integrates big data and deep learning, and completes the chain from data preparation to model training. The deep learning solution of Alibaba Cloud Container Service provides three distributed storage backends (Object Storage Service (OSS), NAS, and HDFS) to support TensorFlow.

This document describes how to convert data to the TFRecord format and store the generated TFRecord files to the HDFS. The HDFS of Alibaba Cloud Elastic MapReduce (E-MapReduce) is used in this example.

Why is TFRecord used

TFRecord is the unified standard data format defined in TensorFlow. It supports multithreading data read and uses the batch size and epoch parameters to control the size of a single batch and the number of iterations of the sample file during the training process. TFRecord can also make better use of the memory and easily perform data replication and movement. Therefore, it is the preferred option for TensorFlow to perform large-scale deep learning training.

Step 1 Create an E-MapReduce cluster

E-MapReduce is a big data processing system solution running on Alibaba Cloud platform. For more information, see [E-MapReduce overview](#).

Log on to the [E-MapReduce console](#) to create an E-MapReduce cluster. For how to create an E-MapReduce cluster, see [Create an E-MapReduce cluster](#).

In this example, a cluster located in China South 1 (Shenzhen) is created, and Network Type is set to VPC.

myEMR

Back to cluster list

Adjust size

Release

Cluster info

ID/Name

C-47BC06EEB3331F98 / myEMR

Region

cn-shenzhen

Start Time

2017/07/04 20:07:55

Log function

Open

Software configuration

High availability

No

Payment type

Pay-As-You-Go

Current status

Creating

Running time

11second(s)

Log path

oss://tensorflow-samples2

Bootstrap action/Software configuration

normal

ECS instance role

AliyunEmrEcsDefaultRole

Software information

Product version

EMR-3.2.0

Cluster type

HADOOP

Software information

hive 2.0.1, nginx 1.10.2, spark 2.1.1, ganglia 3.7.2, tez 0.8.4, hdfs 2.7.2, hue 3.11.0, zeppelin 0.7.1, sqoop 1.4.6, yarn 2.7.2, pig 0.14.0

Network information

Network type

VPC

Select security group

emr-default-securitygroup(sq-wz9ariwt962ng1zxfj)

Zone

cn-shenzhen-b

VPC/VSwitch

vpc-wz96ybdjior29djd4h / vsw-wz96z71x9k43b1z64mw8d

MasterNode information

Basic information

1

Bandwidth : 8M

CPU : 4Core

Memory : 16G

Data disk configuration : SSD Cloud Disk

80G X 1 disk(s)

ID

i-wz94hmr6mjy2k4x1lu

Status

Initializing

Public IP (?)

Private IP

192.168.1.102

Hardware configuration

CPU : 4Core | Memory : 16G | Data disk configuration : SSD Cloud Disk | 80G X 1 disk(s)

Step 2 Create container clusters and integrate networks between the two clusters

1. Log on to the [Container Service console](#) and create a GPU container cluster under the same Virtual Private Cloud (VPC).

Cluster List

You can create up to 5 clusters and can add up to 20 nodes in each cluster.

Refresh

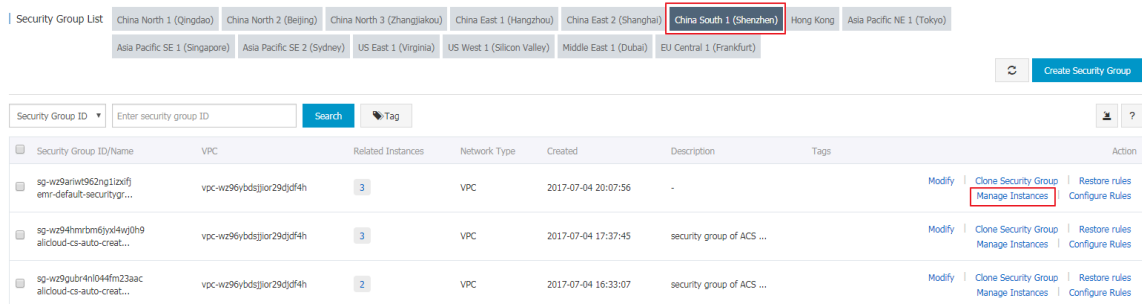
Create Cluster

Help: [Create cluster](#) [How to add existing ECS instances](#) [Cross-zone node management](#) [Log Service integration](#) [Connect to cluster through Docker Client](#)

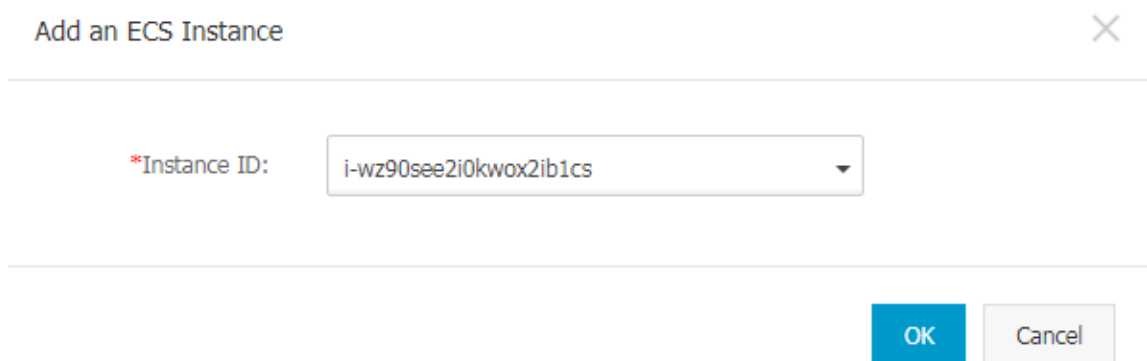
Name

Cluster Name/ID	Cluster Type	Region	Network Type	Cluster Status	Node Status	Number of Nodes	Time Created	Docker Version	Action
ElasticGPUService c-47bc06eeb3331f98	Alibaba Cloud Cluster	China South 1 (Shenzhen)	VPC vpc-wz96ybdjior29djd4h	Running	Healthy	2	2018-01-17 11:17:41	17.06.2-ce	Manage View Logs Delete Monitor More

2. Log on to the [ECS console](#) and add nodes of the Container Service cluster to the security group corresponding to the E-MapReduce cluster.
 - a. Select the region in which the security group resides (China South 1 (Shenzhen) in this example). Click Manage Instances at the right of the security group.



- b. Click Add an ECS Instance in the upper-right corner. Select a node in the container cluster and click OK.

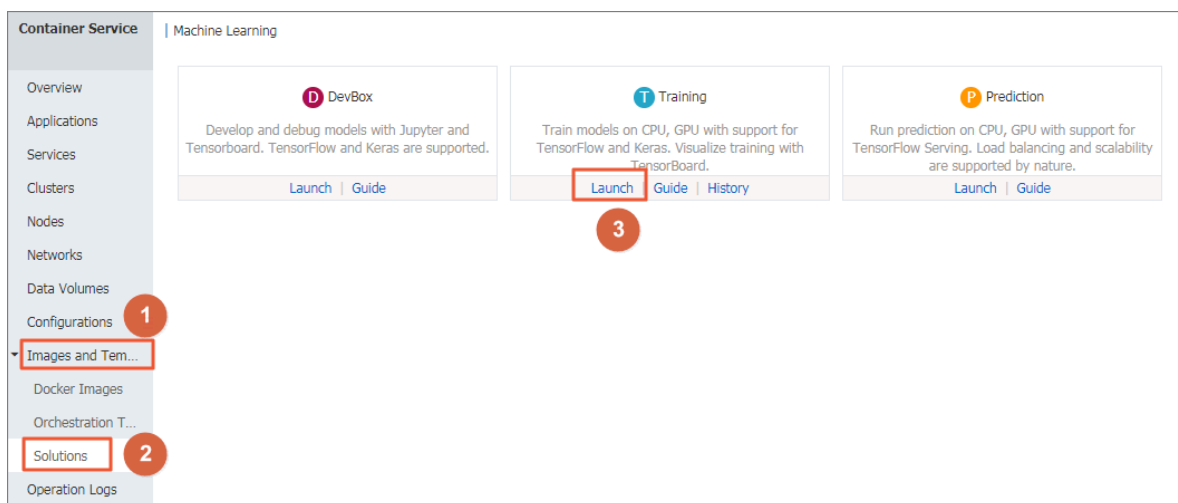


Step 3 Generate TFRecord data

In this example, the model training service provides a running environment to run `convert_to_records.py`, generate TFRecord data, and store the data in the HDFS.

1. Log on to the [Container Service console](#).
2. Click Images and Templates > > Solutions in the left-side navigation pane.

3. Click Launch in Training.



4. Configure the settings for model training and click OK.

Configurations in this example are as follows:

- **Framework:** Select tensorflow:1.0.0.
- **GPUs Per Worker:** Enter 0.
- **Data Source:** Select No Data Source.
- **Git URL:** Enter <https://code.aliyun.com/deeplearning/mnist-examples.git>.
- **Command:**

```
python neural_style.py -- iterations 50000 -- content
/ neural - style / examples / 1 - content . jpg --
styles / neural - style / examples / 1 - style . jpg
```

```
-- output / neural - style / output . jpg
```

Training
Back to Solution List

Cluster
ElasticGPUService

Application Name
prepare-data

The name should be 1-64 characters long, and can contain numbers, English letters and hyphens, but cannot start with a hyphen.

Framework
tensorflow:1.0.0

Distributed Training

GPUs Per Worker
0

Data Source
No Data Source

Git URL
https://code.aliyun.com/deeplearning/mnist-exan

Private Git Information

Command
python neural_style.py --iterations 50000 --content /neural-style/examples/1-content.jpg --styles /neural-style/examples/1-style.jpg --output /neural-style/output.jpg

Enable Monitor

OK

Then, the created application is displayed on the Application List page. Click the application name. Click the Logs tab and view the execution logs, which indicates the TFRecord files are stored in the HDFS.

Services
Containers
Logs
Events
Routes

Entrys Per Container: 100Items
Filter by Container Name: All
Filter by Start Time:
Download Logs

```

prepare-data-12_worker_1 | 2017-05-23T11:33:02.718478793Z Cloning training code from https://code.aliyun.com/deeplearning/mnist-examples.g
it
prepare-data-12_worker_1 | 2017-05-23T11:33:02.720364084Z Cloning into 'mnist-examples'...
prepare-data-12_worker_1 | 2017-05-23T11:33:07.340480568Z Done cloning code.
prepare-data-12_worker_1 | 2017-05-23T11:33:07.340598411Z Run training code under /starter/mnist-examples as: python convert_to_records.py
--directory hdfs://192.168.100.206:9000/mnist-tfrecord
prepare-data-12_worker_1 | 2017-05-23T11:34:05.508679080Z Extracting MNIST_data/train-images-idx3-ubyte.gz
prepare-data-12_worker_1 | 2017-05-23T11:34:05.508706069Z Extracting MNIST_data/train-labels-idx1-ubyte.gz
prepare-data-12_worker_1 | 2017-05-23T11:34:05.508710856Z Extracting MNIST_data/t10k-images-idx3-ubyte.gz
prepare-data-12_worker_1 | 2017-05-23T11:34:05.508714238Z Extracting MNIST_data/t10k-labels-idx1-ubyte.gz
prepare-data-12_worker_1 | 2017-05-23T11:34:05.508717590Z Writing hdfs://192.168.100.206:9000/mnist-tfrecord/train.tfrecords
prepare-data-12_worker_1 | 2017-05-23T11:34:05.508720874Z Writing hdfs://192.168.100.206:9000/mnist-tfrecord/validation.tfrecords
prepare-data-12_worker_1 | 2017-05-23T11:34:05.508724372Z Writing hdfs://192.168.100.206:9000/mnist-tfrecord/test.tfrecords
prepare-data-12_worker_1 | 2017-05-23T11:34:05.574277974Z Done running training code.
prepare-data-12_worker_1 | 2017-05-23T11:34:05.574318365Z Cannot find remote data volume , checkpoints are not persisted remotely.
prepare-data-12_worker_1 | 2017-05-23T11:34:05.574323498Z Done persisting checkpoints to remote storage.

```

You can log on to the E-MapReduce machine to check the generated TFRecord files.

```

# hdfs dfs -ls / mnist - tfrecord
SLF4J : Class path contains multiple SLF4J bindings .
SLF4J : Found binding in [ jar : file : / opt / apps / hadoop
- 2 . 7 . 2 / share / hadoop / common / lib / slf4j - log4j12 - 1 .
7 . 10 . jar ! / org / slf4j / impl / StaticLogg erBinder . class
]
SLF4J : Found binding in [ jar : file : / opt / apps / tez - 0
. 8 . 4 / lib / slf4j - log4j12 - 1 . 7 . 10 . jar ! / org / slf4j
/ impl / StaticLogg erBinder . class ]
SLF4J : See http : // www . slf4j . org / codes . html #
multiple_b indings for an explanatio n .
SLF4J : Actual binding is of type [ org . slf4j . impl .
Log4jLogge rFactory ]
Found 3 items
- rw - r -- r -- 3 root hadoop 8910000 2017 - 05 - 23 19
: 34 / mnist - tfrecord / test . tfrecords

```

```
- rw - r -- r -- 3   root   hadoop   49005000   2017 - 05 - 23   19
: 33 / mnist - tfrecord / train . tfrecords
- rw - r -- r -- 3   root   hadoop   4455000   2017 - 05 - 23   19
: 33 / mnist - tfrecord / validation . tfrecords
```