

# Alibaba Cloud Container Service

Deep learning

Issue: 20190627

## 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.	 <b>Danger:</b> 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.	 <b>Warning:</b> 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.	 <b>Notice:</b> 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.	 <b>Note:</b> You can use Ctrl + A to select all files.
>	Multi-level menu cascade.	Settings > Network > Set network type
<b>Bold</b>	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>switch {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 Model development.....	12
3.1 Create a Jupyter environment.....	12
3.2 Use Git to manage codes in the Jupyter environment.....	15
3.3 Access Jupyter services by using SSH.....	16
3.4 Access Jupyter services by means of port forwarding.....	20
4 Prepare TensorFlow training data by using TFRecord and HDFS.....	27



# 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:

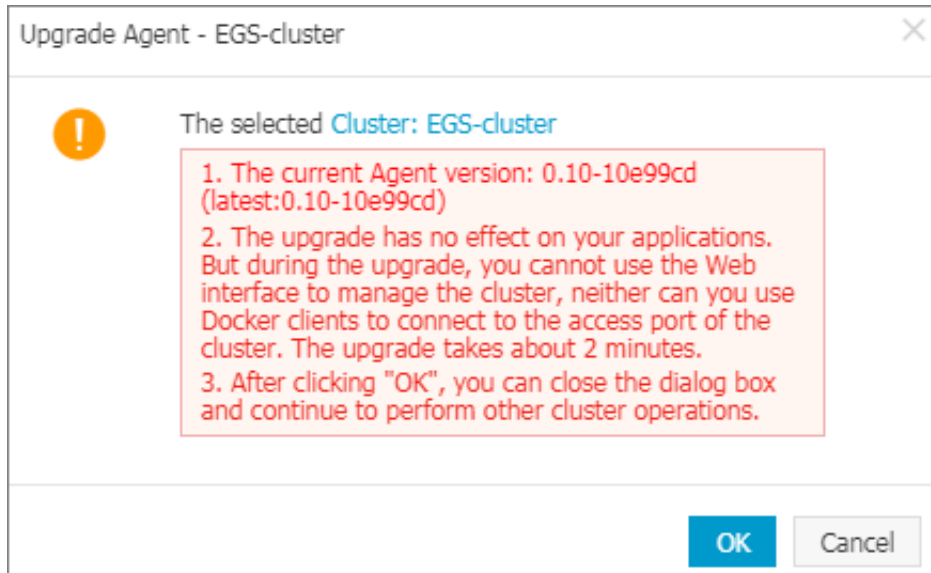
- Random write or append write leads to the entire file being overwritten.
- Metadata operations, such as list directory, provide poor performance because the system needs to remotely access the OSS server.
- The file/folder rename operation is not atomic.
- Coordinate the actions of each client on your 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

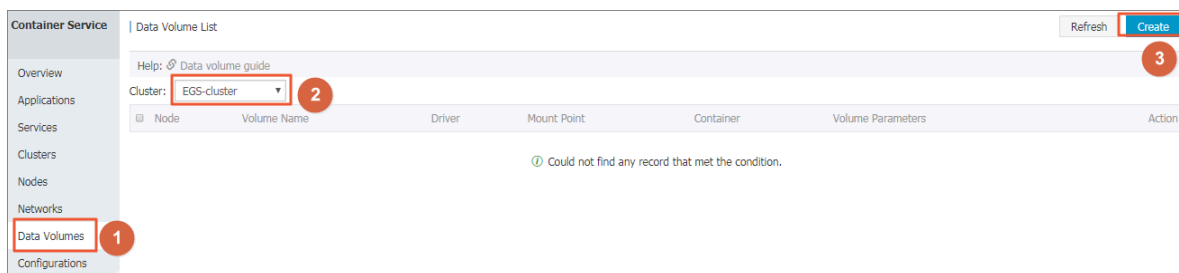
0

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

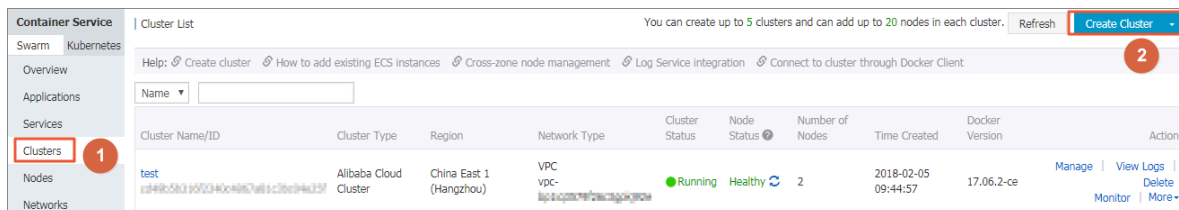
Currently, the Pay-As-You-Go GPU Compute Type GN4 instances need to be activated by [opening an ECS ticket](#) as follows:

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](#).

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



3. 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)	<b>China South 1 (Shenzhen)</b>	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)**.

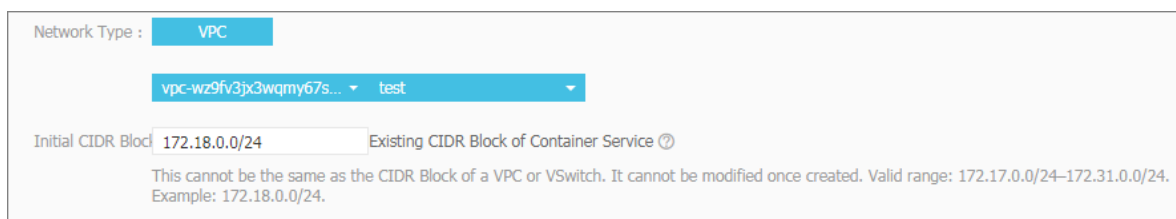
- **Zone**: Select the zone for the cluster.



**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 window 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 field for 'Existing CIDR Block of Container Service' with a help icon. A note at the bottom 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 :

Add

Do not Add

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

Ubuntu 14.04 64bit

?

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
<a href="#">EGS-cluster</a> <small>c747b44aa297448a09618177e72b115a0d1</small>	Alibaba Cloud Cluster	China South 1 (Shenzhen)	VPC vpc- <small>vnt6v0jucwcmqy67amk0t</small>	<span>Running</span> <span>Healthy</span>	2	2018-01-17 11:17:41	17.06.2-ce	<a href="#">Manage</a> <a href="#">View Logs</a> <a href="#">Delete</a> <a href="#">Monitor</a> <a href="#">More</a>	

## 3 Model development

---

### 3.1 Create a Jupyter environment

#### Prerequisites

Before running a model training task, make sure you have performed the following operations:

- Create a container cluster that contains a certain number of elastic computing resources (Elastic Compute Service (ECS) or EGS). For more information, see [Create a container cluster](#).
- To use Object Storage Service (OSS) to store data for model training, use the same account to create an OSS bucket, and create data volumes in the preceding container cluster to mount the OSS bucket as a local directory to the container in which you want to run the training task. For more information, see [Create a data volume](#).

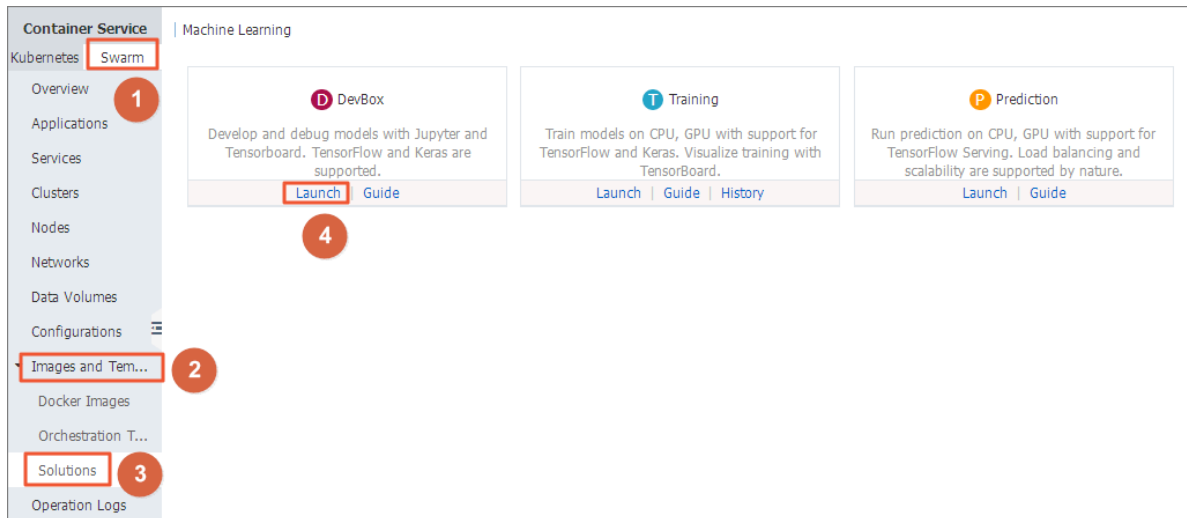
#### Conventions

To facilitate your application codes to read training data and output training logs, data in the training volume is stored in the `/ input` directory. Your codes read data from this directory.

#### Procedure

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

### 3. Click Launch in DevBox.



### 4. Configure the basic information for creating a Jupyter environment.

- **Cluster:** Select the cluster in which the created model development application is to be deployed. Select EGS-cluster in this example.
- **Application Name:** Name of the created application, which can be 1–64 characters long and contain numbers, English letters, and hyphens (-), but cannot start with a hyphen (-).
- **Framework:** The supported frameworks include TensorFlow, Keras, and Python.
- **GPUs:** The number of GPUs in use. If this field is set to 0, no GPU is used.
- **Data Source:** Select the data source used to store training data. Select the data volume created in the cluster by OSS or select Local Directory and then enter the absolute path. You can also select No Data Source.
- **Jupyter Password:** The password used to log on to Jupyter.
- **Enable Monitor:** Select whether or not to use TensorBoard to monitor the training status. With this check box selected, enter the path of the training logs in the Log Directory field and make sure that the path is the same as the log output path in the training codes.
- **Enable SSH:** Select whether or not to allow you to access services by using SSH. With this check box selected, enter your SSH Password.



**Note:**

For how to access services by using SSH, see [Access Jupyter services by using SSH](#).

Cluster	EGS-cluster
Application Name	mydevbox <small>The name should be 1-64 characters long, and can contain numbers, English letters and hyphens, but cannot start with a hyphen.</small>
Framework	tensorflow:1.1.0
GPUs	0
Data Source	Select Data Source
Jupyter Password	
	<input checked="" type="checkbox"/> <b>Enable Monitor</b>
Log Directory	/output/training_logs <small>Please ensure the same log file directory is used in your code.</small>
	<input checked="" type="checkbox"/> <b>Enable SSH</b>
SSH Password	
OK	

5. Click OK after completing the configurations.

6. On the Application List page, click the name of the created application.

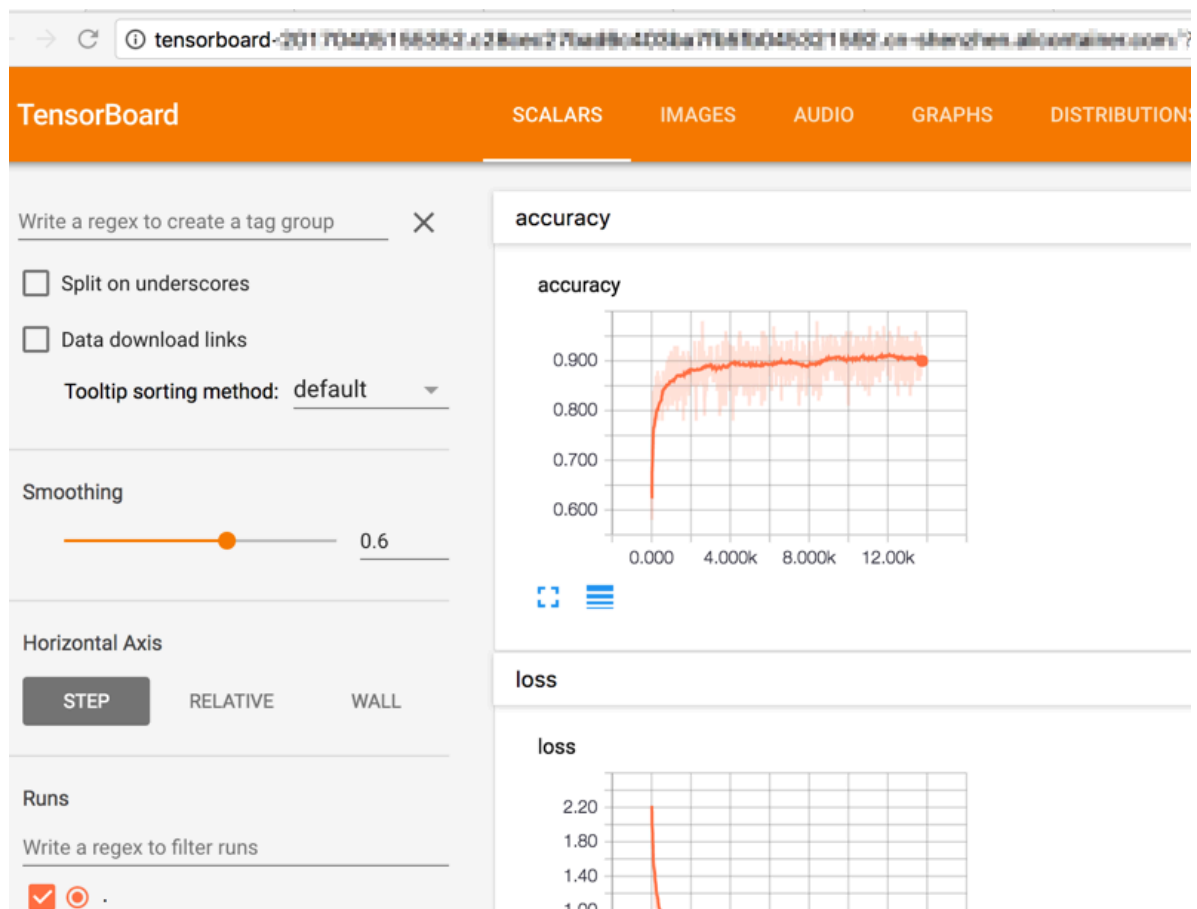
Cluster:	EGS-cluster	<input checked="" type="checkbox"/> Hide System Applications	<input type="checkbox"/> Hide Offline Applications	<input type="checkbox"/> Hide Online Applications	Name		Q	X
Name	Description	Status	Container Status	Time Created	Time Updated	Action		
mydevbox	eml solution application	Ready	Ready:2 Stop:0	05/19/2018,22:03:07	05/19/2018,22:05:05	Stop	Update Redeploy	Delete Events

7. Click the Routes tab. Two links starting with jupyter and tensorboard respectively are displayed.

Services	Containers	Logs	Events	Routes	
Route Address					Action
tensorboard-304808192300M5.c3M34.d4d444c144e813d4c13e4444d81.cn-hangzhou.aliyuncs.com					Set service weight
jupyter-304808192300M5.c3M34.d4d444c144e813d4c13e4444d81.cn-hangzhou.aliyuncs.com					Set service weight

8. Click the link starting with jupyter and enter the Jupyter password to access the Jupyter environment.

9. Click the link starting with tensorboard to view the training results.

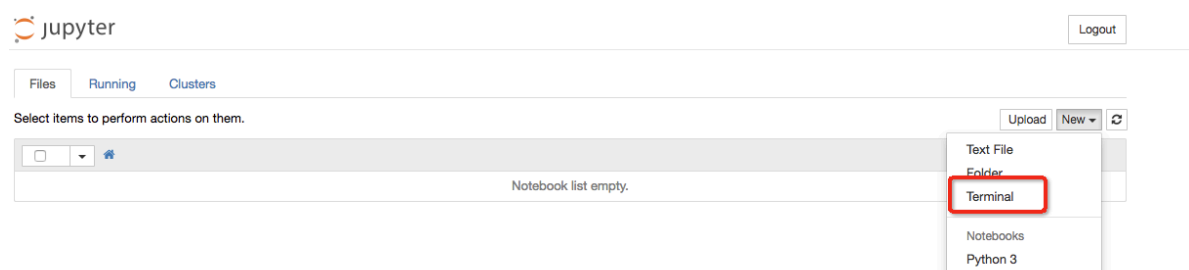


10. Training data in the distributed storage is stored in the local /input folder. You can read data from this folder.

## 3.2 Use Git to manage codes in the Jupyter environment

### Procedure

1. Create a terminal on the Jupyter home page.



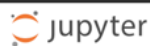
2. Run `git clone` in the terminal to download the application codes.

```
git clone https://[id]:[password]@github.com/[id]/test.git
Cloning into 'test'...
remote: Counting objects: 3, done.
```

```
remote : Total    3  ( delta  0 ), reused    0  ( delta  0 ),
pack - reused    0
Unpacking objects : 100 % ( 3 / 3 ), done .
Checking connecti ty ... done .
```

{ id } and { password } indicate the GitHub username and password respectively.

3. Return to the Jupyter home page. You can check the displayed application codes and use Jupyter to develop the corresponding codes.
4. You can return to the terminal and use Git to submit the codes.



```
# cd test
# git config --global user.email "eml@example.com"
# git config --global user.name "eml"
# git add --all
# git commit -m "commit"
[master 8d272d3] commit
1 file changed, 0 insertions(+), 0 deletions(-)
create mode 100644 test2
# git push
warning: push.default is unset; its implicit value is changing in
Git 2.0 from 'matching' to 'simple'. To squelch this message
and maintain the current behavior after the default changes, use:

    git config --global push.default matching

To squelch this message and adopt the new behavior now, use:

    git config --global push.default simple

When push.default is set to 'matching', git will push local branches
to the remote branches that already exist with the same name.

In Git 2.0, Git will default to the more conservative 'simple'
behavior, which only pushes the current branch to the corresponding
remote branch that 'git pull' uses to update the current branch.

See 'git help config' and search for 'push.default' for further information.
(the 'simple' mode was introduced in Git 1.7.11. Use the similar mode
'current' instead of 'simple' if you sometimes use older versions of Git)

Counting objects: 3, done.
Delta compression using up to 4 threads.
Compressing objects: 100% (2/2), done.
Writing objects: 100% (2/2), 257 bytes | 0 bytes/s, done.
Total 2 (delta 0), reused 0 (delta 0)
To https://cheyang:tang3zang@github.com/cheyang/test.git
   1cb83db..8d272d3  master -> master
#
```

### 3.3 Access Jupyter services by using SSH

If you select the Enable SSH check box when you [Create a Jupyter environment](#), you can follow the steps in this document to access your Jupyter services by using SSH.

In this example, the port mapping for SSH access is 192. 32769->22/tcp. 192. \*\* is the private IP address of the Elastic Compute Service (ECS) instance. You cannot use this IP address to access your Jupyter services. To access the Jupyter services, use the public Elastic IP (EIP) of the ECS instance to SSH access these services from outside.



Services	Containers	Logs	Events	Routes
----------	------------	------	--------	--------

Name/ID	Status	Health Check	Image	Port	Container IP	Node IP	Action
mydevbox_jupyter... ec398456113a4d27...	running	Normal	registry-vpc.cn-... sha256:fdeb9458...	192.168.1.1:32768->8888/tcp 6006/tcp 192.168.1.1:32769->22/tcp	172.17.0.1	192.168.1.1	<a href="#">Delete</a>   <a href="#">Stop</a>   <a href="#">Monitor</a>   <a href="#">Logs</a>   <a href="#">Web Terminal</a>
mydevbox_tensorb... 11ddef632053030...	running	Normal	registry-vpc.cn-... sha256:2b4c70877...	192.168.1.1:32770->6006/tcp	172.17.0.1	192.168.1.1	<a href="#">Delete</a>   <a href="#">Stop</a>   <a href="#">Monitor</a>   <a href="#">Logs</a>   <a href="#">Web Terminal</a>

You can log on to the [ECS console](#) to check the public EIP bound to the ECS instance. In this example, the public EIP is 39. 252.



#### Note:

If your ECS instance is not bound with a public EIP, bind one to your ECS instance. For more information, see [Manage a public EIP address](#).

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

Instance Name	Enter instance name (fuzzy search)	Search	Tag	Advanced Search	⌂	⚙	?
---------------	------------------------------------	--------	-----	-----------------	---	---	---

Instance ID/Name	Monitor	Zone	IP Address	Status(All)	Network Type(All)	Configuration	VPC Details	Billing Method(All)	Action
i-wz5etp6k2f51z4m5k c4bf7b732770f49489625...		China South 1 Zone B	39.252.152.252 (Elastic IP Address) 192.168.1.87 (Private IP Address)	Running	VPC	CPU: 4 Core(s) Memory: 30 GB (I/O Optimized) GPU : NVIDIA M40 100 Mbps ( peak value )	vpc-... vsw-...	Pay-As-You-Go 17-06-28 15:46 created	Manage   Connect   Change Configuration   More

In addition, to use the public EIP of the ECS instance to SSH access your Jupyter services from outside, configure the security group rules and open port 32769 first.

### Configure security group rules

1. Log on to the [ECS console](#).
2. Click Instances in the left-side navigation pane. Select the region (China South 1 (Shenzhen) in this example).

Elastic Computing Se...	Instance List	China North 1 (Qingdao)	China North 2 (Beijing)	China North 3 (Zhangjiakou)	China East 1 (Hangzhou)	China East 2 (Shanghai)	China South 1 (Shenzhen)	Hong Kong	Asia Pacific NE 1 (Tokyo)	⌂	Create Instance
		Asia Pacific SE 1 (Singapore)	Asia Pacific SE 2 (Sydney)	US East 1 (Virginia)	US West 1 (Silicon Valley)	Middle East 1 (Dubai)	EU Central 1 (Frankfurt)				

Overview	Instances	Block Storage	Cloud Disks	NAS	Snapshots & Images	Snapshots
----------	-----------	---------------	-------------	-----	--------------------	-----------

Instance Name	Enter instance name (fuzzy search)	Search	Tag	Advanced Search	Show All Resources	⌂	⚙	?
---------------	------------------------------------	--------	-----	-----------------	--------------------	---	---	---

Instance ID/Name	Monitor	Zone	IP Address	Status(All)	Network Type(All)	Configuration	Billing Method(All)	Action
i-wz5etp6k2f51z4m5k c4bf7b732770f49489625...		China South 1 Zone B	39.252.152.252 (Elastic IP Address) 192.168.1.87 (Private IP Address)	Running	VPC	CPU: 4 Core(s) Memory: 30 GB (I/O Optimized) GPU : NVIDIA M40 100 Mbps ( peak value )	Pay-As-You-Go 17-07-03 15:07 created	Manage   Connect   Change Configuration   More
i-wz5etp6k2f51z4m5k c3703f579999493484736...		China South 1 Zone B	39.252.152.252 (Elastic IP Address) 192.168.1.87 (Private IP Address)	Running	VPC	CPU: 4 Core(s) Memory: 30 GB (I/O Optimized) GPU : NVIDIA M40 100 Mbps ( peak value )	Pay-As-You-Go 17-06-28 15:46 created	Manage   Connect   Change Configuration   More

3. Click More at the right of the ECS instance that corresponds to the Container Service application mydevbox. > Select Security Group Configuration from the drop-down list.

- Click **Configure Rules** at the right of the security group that corresponds to the Container Service cluster.

Security Group ID/Name	Description	VPC	Actions
sg-bp1t2g3duo4qem7v4b6t1 ecs-test-vpc	-	vpc-bp1u6k1y9p0l1z9t0p0o0ll	<a href="#">Configure Rules</a>   <a href="#">Remove</a>

- Click **Add Security Group Rules**. The Add Security Group Rules dialog box appears. Enter the rule information and click **OK**.

### Add Security Group Rules

NIC:

Intranet

Rule Direction:

Inbound

Authorization Policy:

Allow

Protocol Type:

Custom TCP

\* Port Range:

32769/32769

Priority:

1

Authorization Type:

Address Field Access

\* Authorization Object:

0.0.0.0/0

Description:

It must contain 2-256 characters and it cannot begin with http:// or https://

OK

Cancel

## Access Jupyter services by using SSH

### In Linux

If you use a Linux machine, run the following command to access your Jupyter services by using SSH:

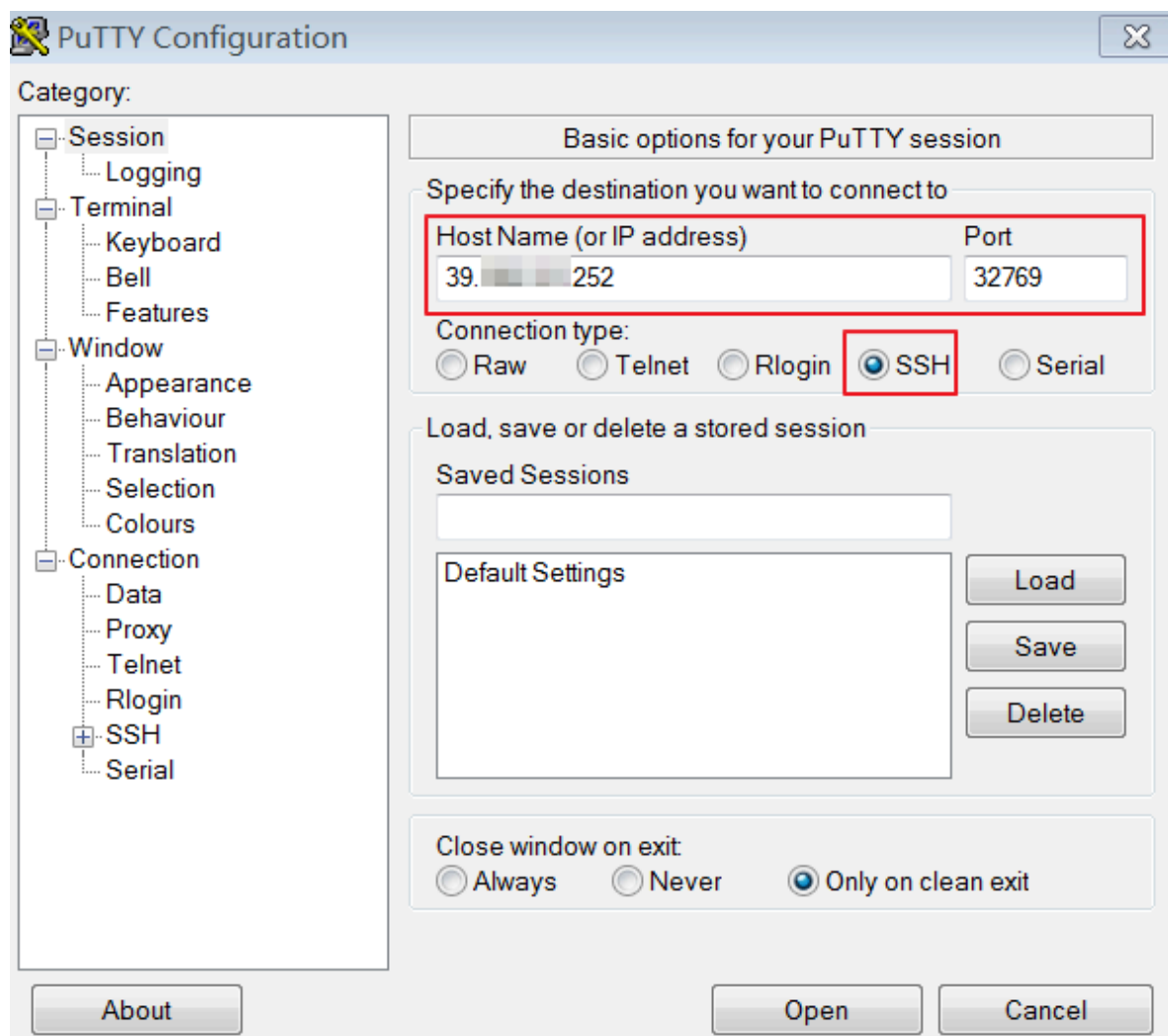
```
ssh -p 32769 root @ 39 . 252
```

32769 is the port to be accessed by SSH, and 39 . 252 is the public EIP bound to your ECS instance.

### In Windows

#### 1. Run PuTTY and configure the session.

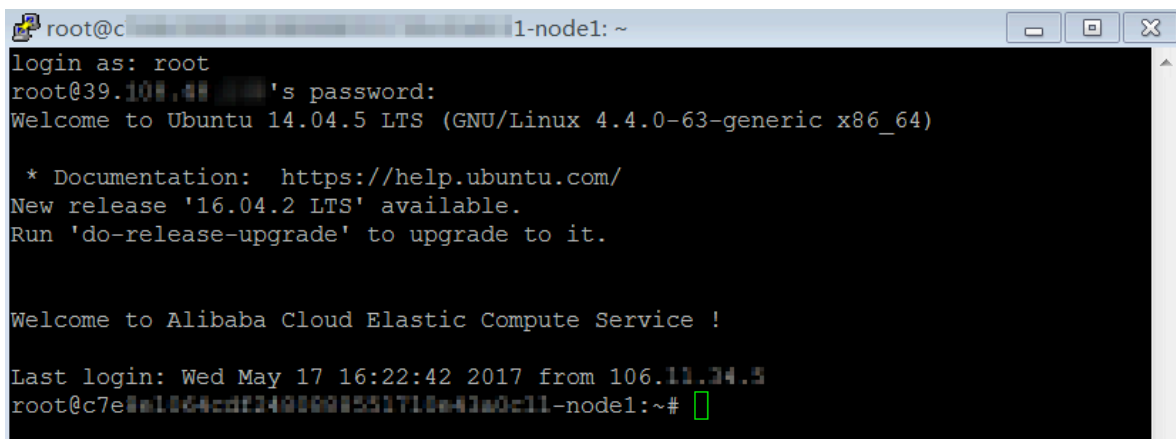
Set the IP address (public EIP of the ECS instance), the port to be accessed by SSH (32769 in this example), and select SSH as the connection type. Then, click Open.



2. In the displayed dialog box, log on to Jupyter and access services.

Enter the logon account `root` and the SSH Password configured for Enable SSH when you [Create a Jupyter environment](#).

You can access Jupyter services by using SSH.



```

root@c7e1a1884cdf248869551718a41a0c11-node1: ~
login as: root
root@39.101.41.141's password:
Welcome to Ubuntu 14.04.5 LTS (GNU/Linux 4.4.0-63-generic x86_64)

 * Documentation:  https://help.ubuntu.com/
New release '16.04.2 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

Welcome to Alibaba Cloud Elastic Compute Service !

Last login: Wed May 17 16:22:42 2017 from 106.11.34.5
root@c7e1a1884cdf248869551718a41a0c11-node1:~#

```

### 3.4 Access Jupyter services by means of port forwarding

Accessing Jupyter services by means of port forwarding has the following advantages

. However, the settings are complicated.

- Save your cost because no Server Load Balancer instance needs to be purchased.
- You can access Jupyter services from outside without opening a port on the official website.

In the following example, your Jupyter services are SSH accessed by forwarding the access request by means of the local port 12345.

Services	Containers	Logs	Events	Routes			
Name/ID	Status	Health Check	Image	Port	Container IP	Node IP	Action
mydevbox_jupyter... ec398456113a4d27...	running	Normal	registry-vpc.cn-... sha256:fdebb9458...	192.168.1.10:32768->8888/tcp 6006/tcp 192.168.1.10:32769->22/tcp	172.17.0.1	192.168.1.10	<a href="#">Delete</a>   <a href="#">Stop</a>   <a href="#">Monitor</a>   <a href="#">Logs</a>   <a href="#">Web Terminal</a>
mydevbox_tensorb... 11ddefd632053030...	running	Normal	registry-vpc.cn-... sha256:2b4c70877...	192.168.1.10:32770->6006/tcp	172.17.0.1	192.168.1.10	<a href="#">Delete</a>   <a href="#">Stop</a>   <a href="#">Monitor</a>   <a href="#">Logs</a>   <a href="#">Web Terminal</a>

#### Step 1 Set up the SSH tunnel

In MAC OS X and Linux

Run the following command to connect the local port with the Elastic Compute Service (ECS) instance.

```
ssh - ND 12345 root @ 39 . 252
```

`12345` is the local port to be used, which can be customized. `39 . 252` is the public Elastic IP (EIP) bound to your ECS instance.



**Note:**

You can log on to the [ECS console](#) to check the public EIP bound to the ECS instance. If your ECS instance is not bound with a public EIP, bind one to your ECS instance. For more information, see [../SP\\_73/DNEIP11899563/EN-US\\_TP\\_12805.dita#task\\_bh5\\_dll\\_vdb](#).

Instance List

China North 1 (Qingdao)

China North 2 (Beijing)

China North 3 (Zhangjiakou)

China East 1 (Hangzhou)

China East 2 (Shanghai)

China South 1 (Shenzhen)

Hong Kong

Asia Pacific NE 1 (Tokyo)

Asia Pacific SE 1 (Singapore)

Asia Pacific SE 2 (Sydney)

US East 1 (Virginia)

US West 1 (Silicon Valley)

Middle East 1 (Dubai)

EU Central 1 (Frankfurt)

Instance Name

Enter instance name (fuzzy search)

Search

Tag

Advanced Search

Instance ID/Name	Monitor	Zone	IP Address	Status(All)	Network Type(All)	Configuration	VPC Details	Billing Method(All)	Action
<div> <div></div> <div> <div> <div></div> <div></div> </div> <div> <div></div> <div></div> </div> </div> </div>	<div></div> <div></div>	China South 1 Zone B	<div>39.100.1252(Elastic IP Address)</div> <div>192.168.187(Private IP Address)</div>	<div>Running</div>	VPC	CPU: 4 Core(s) Memory: 30 GB (I/O Optimized) GPU : NVIDIA M40 100 Mbps ( peak value )	<div>vpc-</div> <div>vsw-</div>	Pay-As-You-Go 17-06-28 15:46 created	<a href="#">Manage</a>   <a href="#">Connect</a> <a href="#">Change Configuration</a>   <a href="#">More...</a>

## In Windows

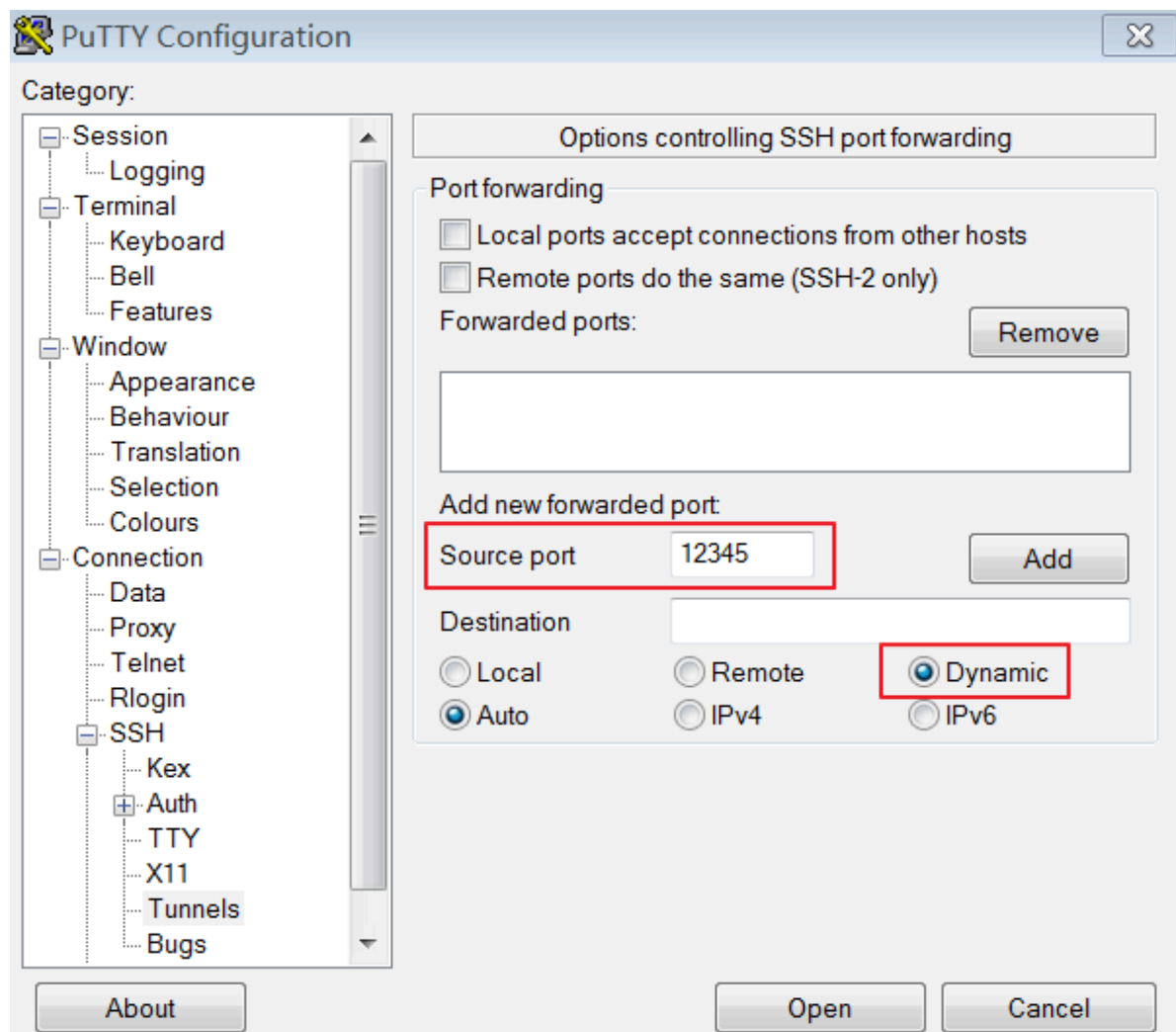
## 1. Configure local port forwarding.

Run PuTTY and configure the SSH Tunnels.

a. Set Source port. In this example, it is 12345.

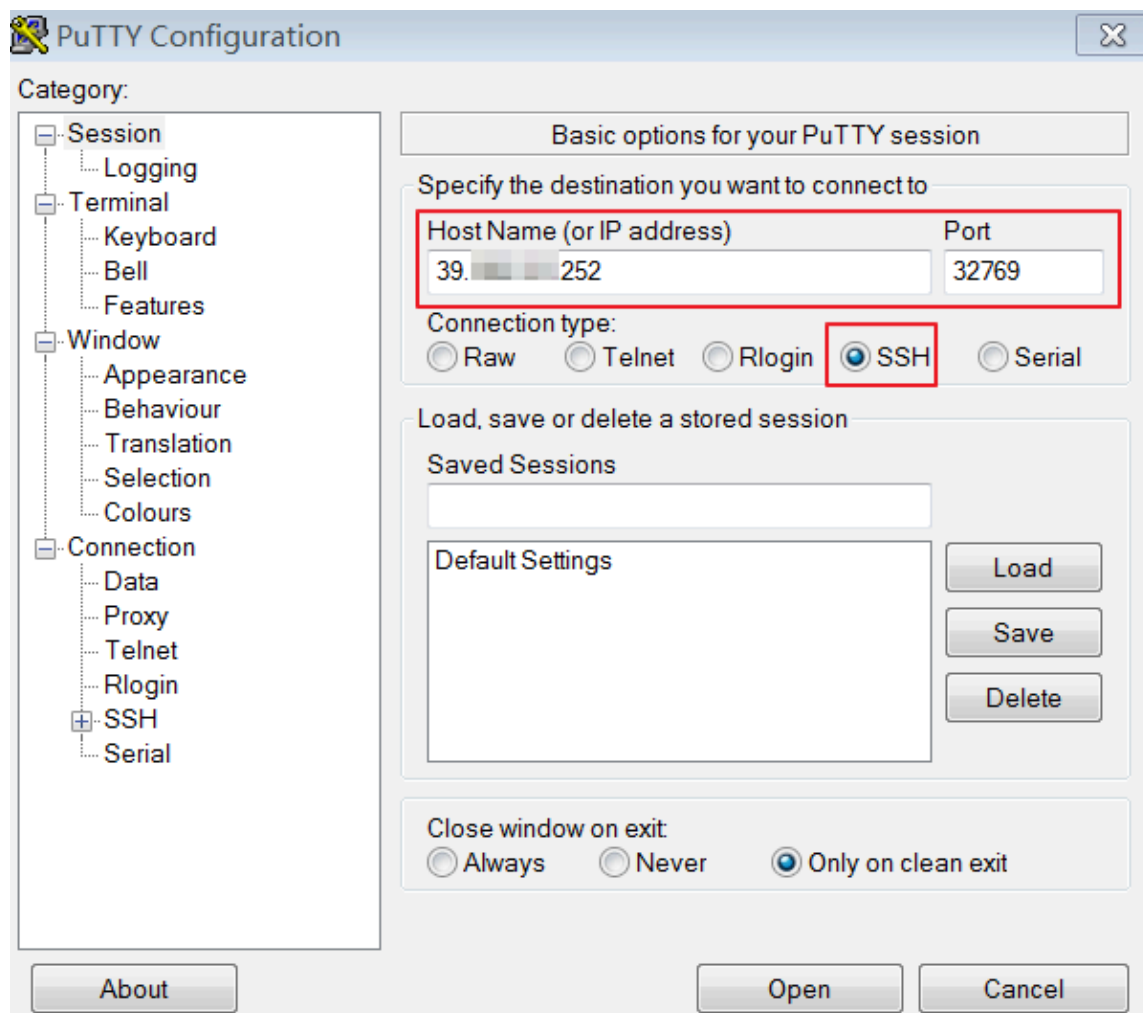
b. Select Dynamic.

c. Click Add.



2. Log on to the ECS instance.
  - a. Run PuTTY. Configure the Session, and then click Open.

Enter the IP address, namely, the public EIP of the ECS instance. In this example , it is 39. 252.



- b. In the displayed dialog box, enter your logon account and password for the ECS instance.

Then, you have successfully logged on to the ECS instance.

```
root@c 1-node1: ~
login as: root
root@39.100.11.111's password:
Welcome to Ubuntu 14.04.5 LTS (GNU/Linux 4.4.0-63-generic x86_64)

 * Documentation:  https://help.ubuntu.com/
New release '16.04.2 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

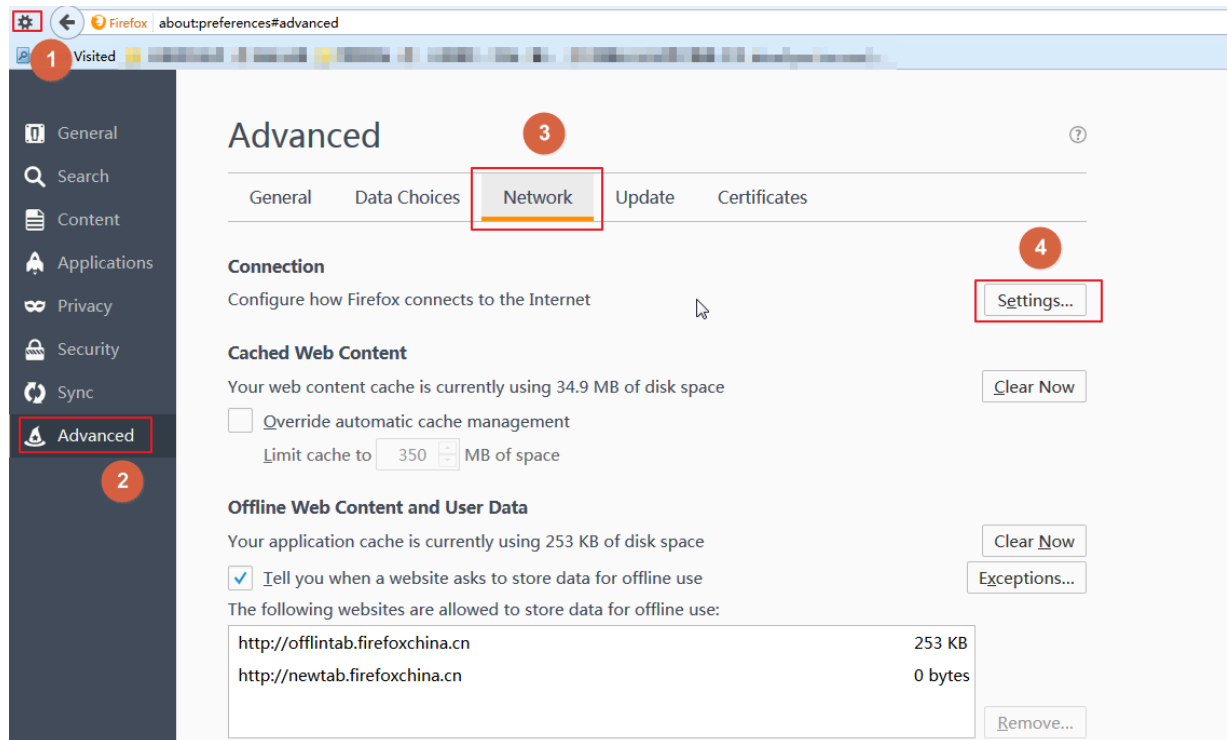
Welcome to Alibaba Cloud Elastic Compute Service !

Last login: Wed May 17 16:22:42 2017 from 106.11.34.3
root@c7e8e1864cdf1498698531718e43a0c11-node1:~#
```

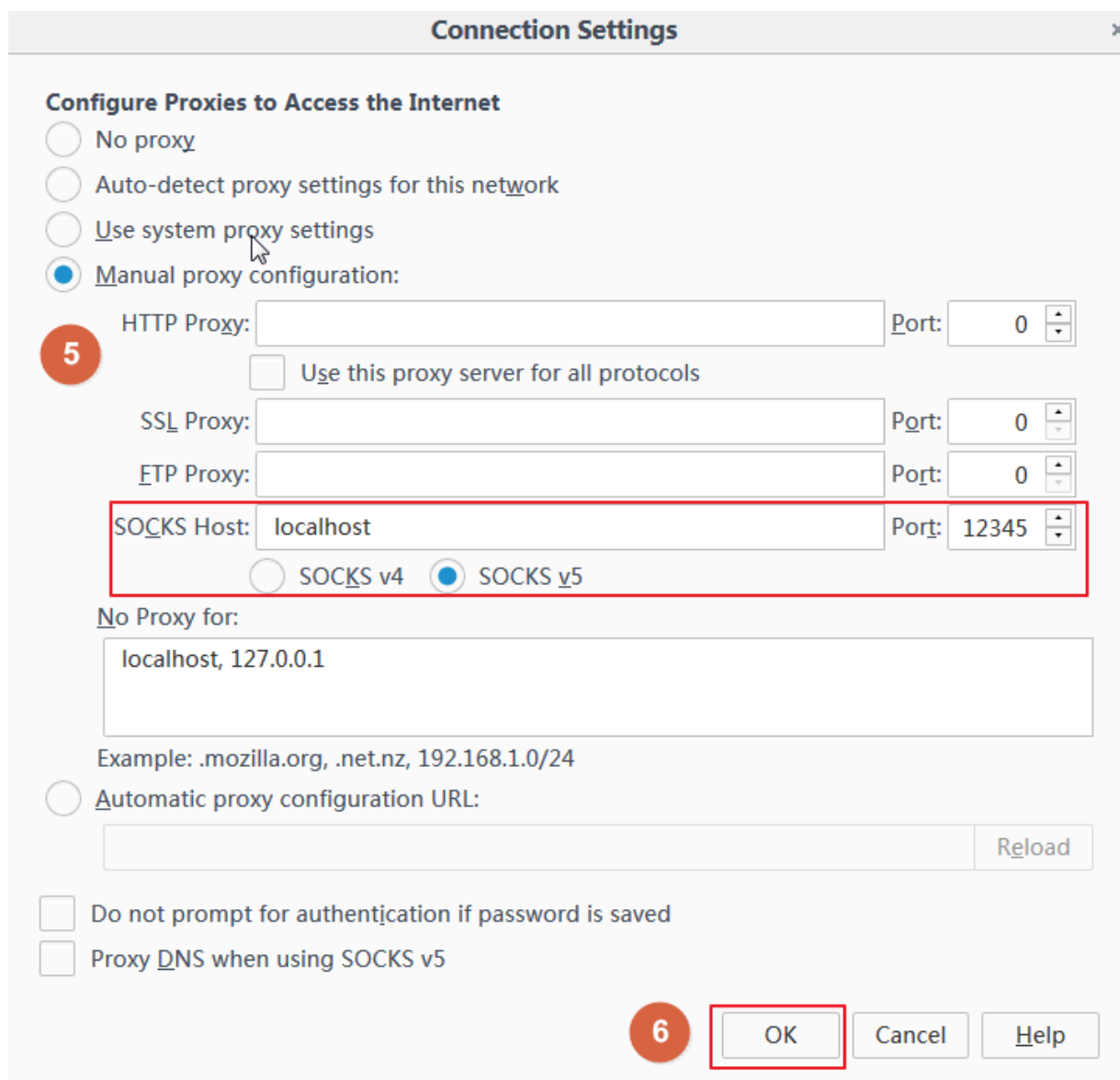
## Step 2 Configure network connection of the browser

### For Firefox

If you use the Firefox browser, open the browser, click Tools > Options > Advanced > Network, and click Settings in the Connection section. In the displayed window, set the SOCKS Host.







### For Chrome

Run the following command:

```
Chrome -- proxy - server =" socks5 :// localhost : 12345 " -- host -
resolver - rules =" MAP * 0 . 0 . 0 . 0 , EXCLUDE localhost "
-- user - data - dir =/ tmppath /
```

Wherein:

- 12345 is the local port in use.
- For Windows, / tmppath / can be written as %temp% or a similar path.  
For Linux or MAC OS X, /tmppath/ can be written as / tmp /.

The location of Chrome varies with operating systems, as described in the following table.

Operating system	Location of Chrome
Mac OS X	/ Applications / Google \ Chromapp / Contents / MacOS / Google \ Chrome
Linux	/ usr / bin / google - chrome
Windows	C : \ Program Files ( x86 ) \ Google \ Chrome \ Applicatio n \ chrome . exe

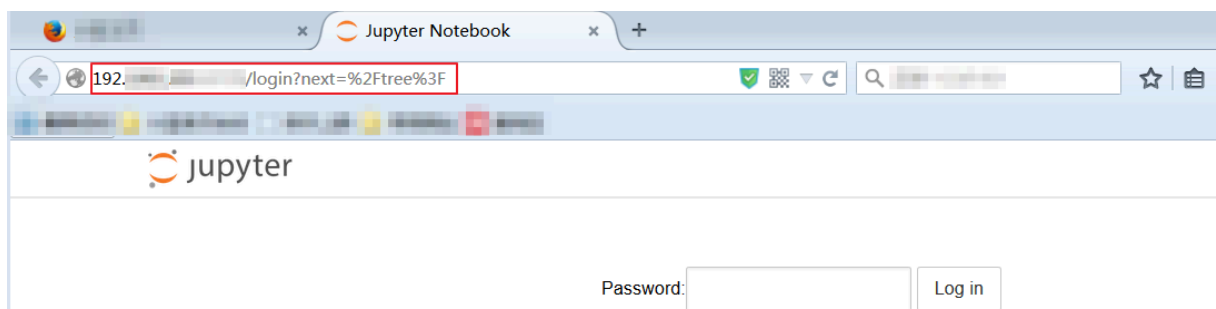
Take Windows as an example.

```
C:\Users\ >"C:\Program Files (x86)\Google\Chrome\Application\chrome.exe" --proxy-server="socks5://localhost:12345" --host-resolver-rules="MAP * 0.0.0.0 , EXCLUDE localhost" --user-data-dir=d:/tmp/path
C:\Users\ >
```

### Step 3 Access Jupyter services

In your browser, enter the access address of a Jupyter service. In this example, access the address 192. 83:32769. Then, you can access the Jupyter service by using SSH from outside.

Firefox:



Chrome:



## 4 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.

Back to cluster list

Adjust size

Release

Cluster info				
ID/Name C-47BC06EEB3331F98 / myEMR	Payment type Pay-As-You-Go			
Region cn-shenzhen	Current status Creating			
Start Time 2017/07/04 20:07:55	Running time 11second(s)			
Log function Open	Log path oss://tensorflow-samples2			
Software configuration	Bootstrap action/Software configuration normal			
High availability No	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-wz96ybd5jior29djd4h / 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	Status	Public IP (?)	Private IP	Hardware configuration
i-wz94hmr6mjy2k64x1lu	Initializing		192.168.1.102	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).

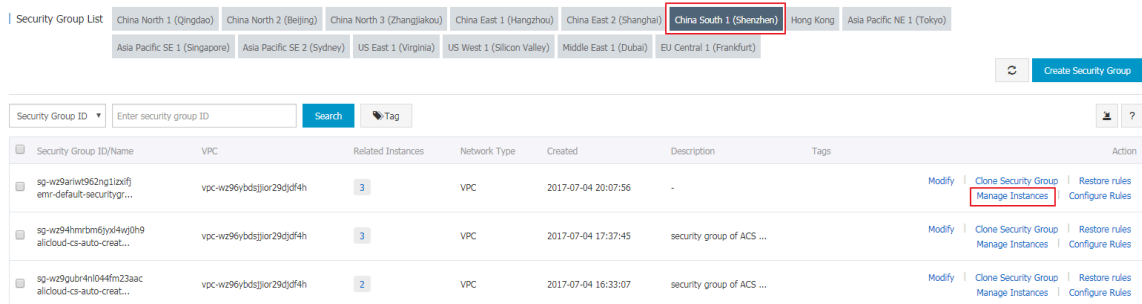
Cluster List
You can create up to 5 clusters and can add up to 20 nodes in each cluster.
[Refresh](#)
[Create Cluster](#)

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

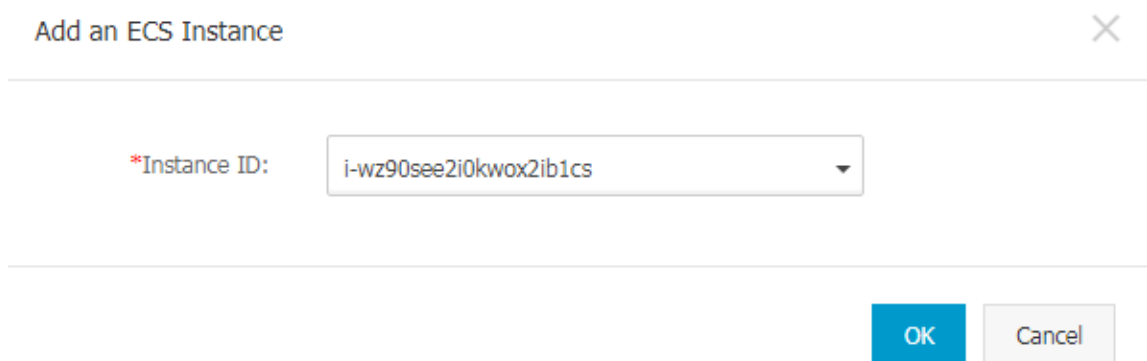
Name

Cluster Name/ID	Cluster Type	Region	Network Type	Cluster Status	Node Status	Number of Nodes	Time Created	Docker Version	Action
<a href="#">ElasticGPUService</a> C-47BC06EEB3331F98	Alibaba Cloud Cluster	China South 1 (Shenzhen)	VPC vpc-wz96ybd5jior29djd4h	Running	Healthy	2	2018-01-17 11:17:41	17.06.2-ce	<a href="#">Manage</a>   <a href="#">View Logs</a>   <a href="#">Delete</a>   <a href="#">Monitor</a>   <a href="#">More</a>

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.

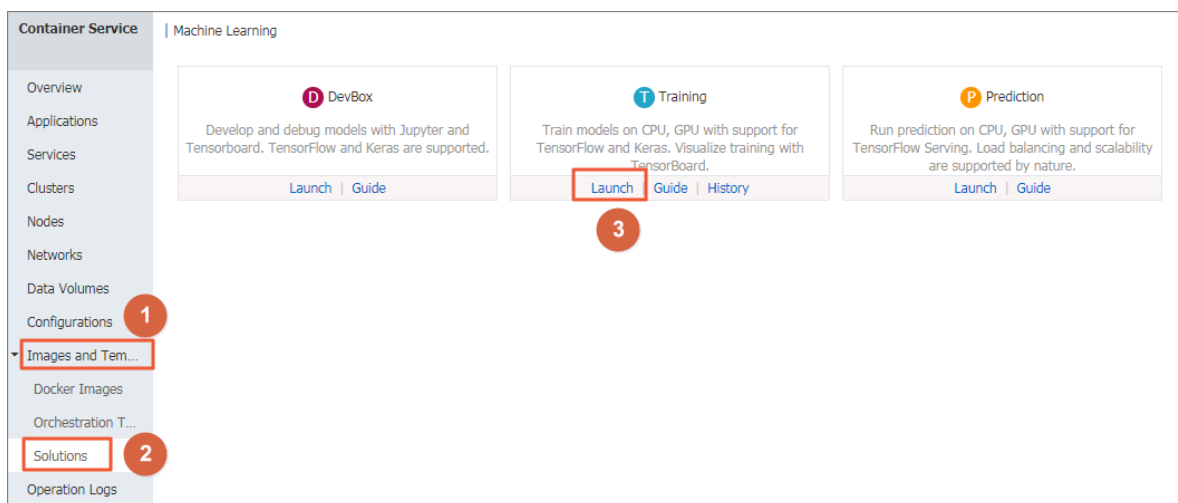


### Step 3 Generate TFRecord data

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

Training

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
```