# Alibaba Cloud

Machine Learning Platform for Al Pricing

Document Version: 20211221

C-J Alibaba Cloud

## 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.

- You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloudauthorized 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 because of product version upgrade, adjustment, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and an updated version of this document will be released through Alibaba Cloud-authorized channels from time to time. You should 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 this document based on the "status quo", "being defective", and "existing functions" of its products and services. 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 take legal responsibility for any errors or lost profits incurred by any organization, company, or individual arising from download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, take responsibility for any indirect, consequential, punitive, contingent, 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 contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, 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 this document 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 and/or its affiliates 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 directly contact Alibaba Cloud for any errors of this document.

# **Document conventions**

Style	Description	Example		
<u>↑</u> Danger	A danger notice 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.		
O Warning	A warning notice 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 restart an instance.		
C) Notice	A caution notice indicates warning information, supplementary instructions, and other content that the user must understand.	Notice: If the weight is set to 0, the server no longer receives new requests.		
? Note	A note indicates supplemental instructions, best practices, tips, and other content.	Note: You can use Ctrl + A to select all files.		
>	Closing angle brackets are used to indicate a multi-level menu cascade.	Click Settings> Network> Set network type.		
Bold	Bold formatting is used for buttons , menus, page names, and other UI elements.	Click OK.		
Courier font	Courier font is used for commands	Run the cd /d C:/window command to enter the Windows system folder.		
Italic	Italic formatting is used for parameters and variables.	bae log listinstanceid Instance_ID		
[] or [a b]	This format is used for an optional value, where only one item can be selected.	ipconfig [-all -t]		
{} or {alb}	This format is used for a required value, where only one item can be selected.	switch {active stand}		

# Table of Contents

1.Purchase	05
2.Metering and billing	<mark>0</mark> 8
2.1. Billing of Machine Learning Studio	80
2.2. Billing examples of Machine Learning Studio	09
2.3. Billing of DSW	14
2.4. Billing of DLC	15
2.5. Billing of EAS	15
2.6. Billing of AutoLearning	21
3.View bills and usage details	22

# 1.Purchase

This topic describes how to purchase services provided by Machine Learning Platform for AI.

#### Procedure

- 1. Log on to Machine Learning Platform for AI.
- 2. On the homepage, click Buy Now.
- 3. On the buy page, set the following parameters.
  - If you purchase Pay-As-You-Go (PAI Studio, DSW, and EAS) instances, set parameters on the Machine Learning page.

Parameter	Description				
Region	The available regions are: China (Beijing) China (Shanghai) China (Hangzhou) China (Hangzhou) China (Shenzhen) Singapore (Singapore) Malaysia (Kuala Lumpur) Malaysia (Kuala Lumpur) Indonesia (Jakarta) China (Hong Kong) China (Hong Kong) India (Mumbai) Australia (Sydney) US (Silicon Valley) US (Virginia) Germany (Frankfurt) UAE (Dubai) Japan (Tokyo)				
Version	Only <b>V2.0</b> is supported.				
PAI-Studio	The system does not support purchasing Machine Learning Studie				
PAI-DSW	DSW, and EAS instances separately. You must purchase the services in				
PAI-EAS	שנווענפג.				

• If you purchase **Subscription (PAI DSW)** instances, set parameters on the Subscription (PAI DSW) page.

Parameter

Description

Parameter	Description
Region	The available regions are: China (Beijing) China (Shanghai)
Resource Type	<ul> <li>The system supports the following GPU types:</li> <li>P100</li> <li>M40</li> <li>⑦ Note Different regions support different GPU types. The pricing of DSW instances varies based on the supported GPU types in supported regions. After you purchase an instance, you can upgrade</li> </ul>
Subscription Period	and renew the instance, but cannot downgrade the instance. The subscription cycle. You can select a subscription cycle from 1 to 12
GPUs	months.
Subscription PeriodThe subscription cycle. You can select a subscription cycle from 1 months.GPUsValid values: 1 to 12.	

• If you purchase EAS Resource Group (subscription), set parameters on the EAS (Subscription) page.

Parameter	Description
Region	The available regions are: China China (Hangzhou) China (Shanghai) China (Shanghai) China (Beijing) China (Beijing) China (Shenzhen) China (Shenzhen) China (Hong Kong) Asia Pacific Singapore (Singapore) Indonesia (Jakarta) Europe & Americas: Germany (Frankfurt) Middle East & India: India (Mumbai)
Node Specifications	The GPU and CPU specifications supported by the system.          ⑦ Note       Supported node specifications vary in different regions.

Parameter	Description
Nodes	Valid values: 1 to 1000.
Order Time	Select a subscription cycle from 1 to 12 months.
Auto-renewal	Specify whether to automatically renew the subscription of an instance after it expires.

• If you purchase EAS Resource Group (Pay-as-you-go), set parameters on the EAS Resource Group (Pay-As-You-Go) page.

Parameter	Description				
	The GPU and CPU specifications supported by the system.				
Node Specification	<b>Note</b> Supported <b>node specifications</b> vary in different regions.				
	The available regions are:				
	<ul> <li>China (Hangzhou)</li> </ul>				
	China (Shanghai)				
	<ul> <li>China (Beijing)</li> </ul>				
Region	China (Shenzhen)				
Region	<ul> <li>China (Hong Kong)</li> </ul>				
	<ul> <li>Singapore (Singapore)</li> </ul>				
	Indonesia (Jakarta) India (Mumbai)				
	<ul> <li>India (Multipal)</li> <li>Germany (Frankfurt)</li> </ul>				
Number of Nodes	Valid values: 1 to 100.				

4. After you set the parameters, confirm the purchase details, and click **Buy Now**.

# 2.Metering and billing2.1. Billing of Machine Learning Studio

This topic describes the billing rules for each module in Machine Learning Studio.

## Billing

One compute hour is equivalent to the usage of one CPU core and 4 GB of memory in one hour.

The following formula describes how to calculate the number of compute hours.

Number of compute hours = Max (Number of CPU cores  $\times$  Usage duration, Memory size  $\times$  Usage du ration/4)

Onte The memory size is measured in GB and the usage duration is measured in hours.

The following formula shows how to calculate the number of compute hours when you use 2 CPU cores and 5 GB of memory in one hour.

Number of compute hours = Max (2  $\times$  1, 5  $\times$  1/4) = 2

The following formula describes how to calculate the bill amount.

Bill amount= Number of compute hours × Unit price

## Pricing

Machine Learning Studio is billed on the pay-as-you-go basis. The following table shows the prices of the commonly used modules in Machine Learning Studio.

#### Machine Learning Platform for AI

Module	Description	Price (USD/compute hour)	Region
Data processing	Includes components such as data preprocessing and feature engineering.	0.16	<ul> <li>China (Beijing)</li> <li>China (Shanghai)</li> <li>China (Hangzhou)</li> </ul>
Data analysis	Includes components such as statistical analysis, machine learning, time series, and financials.	0.21	<ul> <li>China (Shenzhen)</li> <li>China (Hong Kong)</li> <li>Singapore (Singapore)</li> </ul>
Text analysis	The algorithm for text analysis.	0.27	<ul> <li>Malaysia (Kuala Lumpur)</li> <li>Indonesia (Jakarta)</li> <li>India (Mumbai)</li> <li>Australia (Sydney)</li> <li>US (Silicon Valley)</li> <li>US (Virginia)</li> <li>Germany (Frankfurt)</li> <li>UAE (Dubai)</li> <li>Japan (Tokyo)</li> </ul>

**?** Note MaxCompute SQL statements are the basic units for executing SQL scripts, JOIN, UNION, and filtering and mapping components. Therefore, bills may be generated for MaxCompute when you use Machine Learning Studio. For more information, see Billing method.

#### The following table shows the prices of deep learning components.

Module	Description	Price (USD/GPU/hour)	Region
Deep learning (M40)	Includes deep learning frameworks	1.308	China (Shanghai)
Deep learning (P100)	MXNet.	1.872	China (Beijing)

## Examples

For more information about billing examples, see Billing examples of Machine Learning Studio.

# 2.2. Billing examples of Machine Learning Studio

In this topic, the Probabilistic Linear Discriminant Analysis (PLDA) component is used as an example to describe how to calculate the fees for experiments running in Machine Learning Studio.

## Context

An experiment in Machine Learning Studio consists of more than one algorithm component, while an algorithm component is composed of multiple subtasks. To calculate the fees for an experiment, you need to first calculate the costs of subtasks in each algorithm component, then sum up the fees of all components used in the experiment.

#### Procedure

- 1. Determine the category of an algorithm component.
  - i. Log on to the Machine Learning Platform for AI console.
  - ii. In the left-side navigation pane, choose **Model Training > Studio-Modeling Visualization**.
  - iii. On the PAI Visualization Modeling page, click Machine Learning in the Operation column.

Machine Learning Platform For Al		Machine Learning Play	tform for AL / Model Training / Studio-Mod	eling Visualization						
Overview										
Data Preprocessing Dataset Manager	^	Create Project	Project Name V Please Input	Q						c
Smart Labeling		Project Name	Alias	Region	Project Administrator	MaxCompute Resources	Create Time	Open GPU 😗	Operation	
Model Training	^	1000	1000	China East 2 (Shanghai)	101-101-000		Jun 6, 2020 11:37 AM		Machine Learning	
Studio-Modeling Vi	sualiz							Items per p	age: 10 V C Previous	1 Next >
DSW-Notebook Serv	ice									

- iv. On the Algo Platform page, click Components in the left-side navigation pane.
- v. In the Components list, find the **PLDA** component. The PLDA component belongs to the category of text analysis. The component is billed at a price of USD 0.27/compute hour.

ୖ୷	S	earcl	h			Q
Home 57	>		Ma	achine Learning		
A Experime	>		De	ep Learning	Open	
জ	>		Re	inforcement Learning		
Notebook	>		De	ep Learning(CPU)		
Data	>		Tir	me Series		
மீ	~	6	Te	xt Analysis		
Compone			诶	Word Splitting		
₩ Models			柒	Convert Row, Col		
කි			柒	String Similarity		
Settings			柒	String Similarity		
			柒	Deprecated Wor		
			柒	N-gram Counting		
			柒	Text Summarizati		
			诶	Keyword Extracti		
			诶	Sentence Splitting		
			诶	Semantic Vector		
			柒	Doc2Vec		
			柒	Conditional Rand		
			柒	Document Simila		
			柒	Pointwise Mutual		
			诶	Conditional Rand		
			诶	Word Splitting (		
			渫	BERT Embedding		
			柒	Word Frequency		
			诶	TF-IDF		
			诶	PLDA		
			議	Word2Vec		

- 2. View the resources consumed for running all jobs of a subtask.
  - i. On the Algo Platform page, click **Experiments** in the left-side navigation pane.
  - ii. In the **My experiments** list, click an experiment, and view the experiment flowchart on the canvas.
  - iii. On the canvas, right-click the PLDA component.
  - iv. In the menu that appears, click **View Log**.

v. In the View Log dialog box, click a hyperlink. Each hyperlink corresponds to a subtask.



#### vi. Click **AlgoTask**.

#### Log det ails

URL	Project	InstanceID	Own	1er	3	StartTime	EndTime	Latency	Status	Progress	SourceXML
http://service	-			-		03/06/2020, 23:26:18	03/06/2020, 23:26:23	00:00:05	Terminated	100%	RML
AlgoTask type_transform											
ODPS Tasks											
Name	Type	Status	Result	Detail	History	StartTime	EndTime	Latency	TimeLine		
					-						

vii. In the TaskPlan section, view CPU and Memory.

Node XML: [type_transform]	×
The second se	<b>^</b>
Terretory (	
and the second se	
stages \$1.00.00	
and the second sec	
"TaskPlan":	And the second
\"CPU\": 100 \n	
AND AND IN CONTRACTOR OF A DESCRIPTION O	
The second	1000
Concerned from the second states of the second	
	-

- The number of used CPU cores is calculated by dividing the value of CPU by 100. In this example, one CPU core is used for running the job.
- The unit of Memory is MB. In this example, 1024 MB (1 GB) of memory is used for running the job.
- viii. On the Log Details page, double-click the task below **ODPS Task**, as shown in the Log details figure.
- ix. On the **Fuxi Jobs** tab, click the subtask. In the section that appears, click the **Terminated** tab. Latency indicates the operation duration of each job.

Detail for [Ida]														×
🕼 refresh														
Fuxi Job DAG 🛛 🔍	Mai	n Content												
	Fu	xi Jobs 📔 Summ	ary JSONSun	mary										
	Fu	xi Job Name:		-		-		a second						*
		TaskName	Fatal/Finished	/TotalInstC	Count I/C	Records I	/O Bytes FinishedP	ercentage	Status	StartTime	EndTin	ne Latr	ency(s)	TimeLine
		1 MWorker	0/49/49		0/0	) (	/0 1	00%	Terminated	03/06/2020, 19:49:25	03/06/2020, 1	9:49:50	00:00:25	
MWorker	La constante													
(40/40)														
(49/49)	M	Worker 🗵												
	Sm	artFilter Failed(0)	Terminated(49		Long	Tails(0)	Latency chart			Latency	{"min"-"00-00-	19" "avo"·"00·f	00·22" "m:	av"·"00·00·22"3
	5	Sinarchiter Falled(0) Terminated(49) Ali(49) Long-Talis(0) Latency Chart				Ctart	Time	EndTime	Latency(c)	Timel ine	,	N . 00.00.22 J		
	0	MWorker#31_0	ME1URXVNak	Buout	F	Terminate	d 100%	03/06/20	20.19:49:28	03/06/2020 19:49:50	00:00:22	TimeEnte		
	1	MWorker#0_0	PU1URXVNak	6	6	Terminate	100%	03/06/20	20.19:49:28	03/06/2020 19:49:50	00:00:22			
	2	MWorker#11_0	PU1URXVNak	Ĩ	ī	Terminate	d 100%	03/06/20	20,19:49:28	03/06/2020,19:49:50	00:00:22			
	3	MWorker#12_0	PU1URXVNak	Ĩ	Ĩ	Terminate	d 100%	03/06/20	20,19:49:28	03/06/2020,19:49:50	00:00:22			
	4	MWorker#13_0	ME1URXVNak		Ē	Terminate	d 100%	03/06/20	20,19:49:28	03/06/2020,19:49:50	00:00:22			
	5	MWorker#14_0	PU1URXVNak	E	E	Terminate	d 100%	03/06/20	20,19:49:28	03/06/2020,19:49:50	00:00:22			
	6	MWorker#8_0	M01URXVNak		J	Terminate	d 100%	03/06/20	20,19:49:28	03/06/2020,19:49:50	00:00:22			
	7	MWorker#16_0	PU1URXVNakl	5	j -	Terminate	d 100%	03/06/20	20,19:49:28	03/06/2020,19:49:50	00:00:22			
	8	MWorker#17_0	PU1URXVNak	1	1	Terminate	d 100%	03/06/20	20,19:49:28	03/06/2020,19:49:50	00:00:22			
	9	MWorker#7_0	Mk1URXVNakl		1	Terminate	d 100%	03/06/20	20,19:49:28	03/06/2020,19:49:50	00:00:22			
	10	MWorker#19_0	d01URXVNakU		1	Terminate	d 100%	03/06/20	20,19:49:28	03/06/2020,19:49:50	00:00:22			
	11	MWorker#1_0	eE1URXVNakU	1	1	Terminate	d 100%	03/06/20	20,19:49:28	03/06/2020/19:49:50	00:00:22			
	12	MWorker#5 0	PU1URXVNak	120		Terminate	100%	03/06/20	120.19:49:28	03/06/202019:49:50	00:00:22			

In this example, the subtask has 49 jobs, and each job runs for about 22 seconds.

3. Calculate the fee of the subtask.

i. Calculate the number of compute hours used in the subtask. For more information about how to calculate the number of compute hours, see Billing of Machine Learning Studio.

```
The number of compute hours used in the subtask = Max (Number of CPU cores \times Usage duration, Memory size \times Usage duration/4) = Max [49 \times 1 \times (22/3,600), 49 \times 1 \times 22/3,600/4] \approx 0.30 compute hour
```

This means that the subtask consumes about 0.3 compute hour.

ii. Calculate the fee of the subtask.

Subtask fee = Number of compute hours  $\times$  Unit price  $\approx$  0.30  $\times$  0.27 = USD 0.081

This means that the subtask costs about USD 0.081.

- 4. Calculate the total fee of all subtasks running in the PLDA component.
- 5. Sum up the fees of all subtasks to calculate the costs of the PLDA component.
- 6. Repeat the preceding steps to calculate the fees of all components used in the experiment.
- 7. Sum up the fees of all components to calculate the costs of the experiment.

## 2.3. Billing of DSW

Machine Learning Platform for AI provides the following Data Science Workshop (DSW) editions: Individual Edition, GPU On-sale Edition, and Explorer Edition. This topic describes the billing rule for each DSW edition.

#### **Individual Edition**

DSW Individual Edition supports only the pay-as-you-go billing method. The following section describes the billing rule for this edition:

```
Bill amount = (Price/60) × Usage duration
```

```
Onte Usage duration is measured in minutes.
```

CPU resource type	Specification	Price (USD/hour)	Region
ecs.c6.large	2 vCPUs + 4 GB of memory	0.066	
ecs.g6.2xlarge	8 vCPUs + 32 GB of memory	0.336	
ecs.g6.4xlarge	16 vCPUs + 64 GB of memory	0.673	
ecs.g6.8xlarge	32 vCPUs + 128 GB of memory	1.346	
ecs.g6.large	2 vCPUs + 8 GB of memory	0.084	

The following table lists the prices of pay-as-you-go instances of DSW Individual Edition.

CPU resource type	Specification	Price (USD/hour)	<ul> <li>China (Beijing)</li> <li>Region</li> <li>China (Shanghai)</li> </ul>
ecs.g6.xlarge	4 vCPUs + 16 GB of memory	0.168	<ul><li>China (Hangzhou)</li><li>China (Shenzhen)</li></ul>
ecs.gn5-c28g1.7xlarge	28 vCPUs + 112 GB of memory	4.017	
ecs.gn5-c4g1.xlarge	4 vCPUs + 30 GB of memory	2.15	
ecs.gn5-c8g1.2xlarge	8 vCPUs + 60 GB of memory	2.589	
ecs.gn5-c8g1.4xlarge	16 vCPUs + 120 GB of memory	5.177	
ecs.gn6e- c12g1.12xlarge	48 vCPUs + 368 GB of memory	13.28	

## Billing rule for DSW GPU On-sale Edition

DSW GPU On-sale Edition supports only the subscription billing method. The following section describes the billing rule for this edition:

Specification	Price (USD/GPU/month)	Region
P100	308	China (Beijing)
M40	76.453	China (Shanghai)

# 2.4. Billing of DLC

This topic describes the billing rules of Deep Learning Containers (DLC).

DLC trains deep learning models in Alibaba Cloud Container Service for Kubernetes (ACK). Fees are charged for resources, networks, or storage services associated with ACK clusters. For more information, see Billing of ACK. No additional fee is charged for using DLC.

## 2.5. Billing of EAS

This topic describes how Elastic Algorithm Service (EAS) is charged.

You are charged EAS fees for the resources that are used to deploy model services.

- You can deploy model services in shared resource groups and dedicated resource groups. Bills are generated based on the used resources of resource groups. For more information about the differences between the two types of resource groups, see Dedicated resource groups.
- If you deploy a model service in a shared resource group and a dedicated resource group, you are charged for using the resources of both resource groups.

The following table lists the billing rules for shared resource groups and dedicated resource groups.

#### Pricing Metering and billing

Resource	Billing item	Billing method	Billing rule	Method to stop billing
Shared resource group	The duration that model services have been running (the amount of time for which the model services have occupied shared resources)	Pay-as-you-go	Bills are generated based on the amount of shared resources that are occupied by model services. The billing starts immediately after model services are created.	Stop model services
		Pay-as-you-go	Only dedicated resource groups are charged. Model	Stop dedicated resource groups
Dedicate d resource group	The duration that resource groups have been running	Subscription	dedicated resource groups are not charged. The billing starts immediately after pay-as-you-go dedicated resource groups are created.	No

## Billing rules for shared resource groups

1 compute hour is equivalent to the usage of 1 CPU core and 4 GB of memory.

Each model service is billed based on the following formula:

```
Bill amount of each model service = Number of compute hours \times (Unit price/60) \times Usage durat ion
```

Onte Usage duration is measured in minutes.

• The system starts to bill the resources that are used by a model service after the model service enters the Running state. The system stops billing the resources that are used by a model service after the model service enters the Stopped state.

Notice Stop unused model services to avoid unnecessary costs.

- Bills are generated from the time when a model service starts to run and consume resources. The system stops billing immediately after a model service is suspended and releases resources.
- If you scale out a model service, newly added resources are billed after the scale-out activity is complete. If you scale in a model service, released resources are no longer billed, and only the remaining resources are billed.
- Fees are charged on a per minute basis. No fee is charged if a model service runs for less than one minute.

## Prices of shared resource groups

The catalog prices of pay-as-you go shared resource groups are hourly prices. However, fees are charged on a per minute basis. You can calculate the unit prices per minute by dividing the prices listed in the following table by 60. The following table lists the unit prices per hour.

Resource type	Price (USD/compute hour/hour)	Region
CPU	0.06	<ul> <li>China (Beijing)</li> <li>China (Shanghai)</li> <li>China (Hangzhou)</li> <li>China (Shenzhen)</li> <li>China (Hong Kong)</li> <li>Singapore (Singapore)</li> <li>Indonesia (Jakarta)</li> <li>India (Mumbai)</li> <li>Germany (Frankfurt)</li> </ul>

**?** Note Shared resource groups do not provide GPU resources. If you require GPU resources, you must purchase dedicated resource groups.

#### Billing examples for shared resource groups

For example, you used CPU resources in the China (Hangzhou) region to deploy a model service. The model service occupies 2 compute hours (2 CPU cores + 8 GB), and entered the Running state at 09:00:00 (UTC+8) on June 3, 2019. You reduced the occupied resources to 1 compute hour (1 CPU core + 4 GB) at 10:00:00 (UTC+8) on June 3, 2019. Then, you increased the occupied resources to 4 compute hours (4 CPU cores + 16 GB) at 11:00:00 (UTC+8) on June 3, 2019. At 12:00:00 (UTC+8) on June 3, 2019, the model service stopped running. The total fee is calculated based on the following formula:

Bill amount = 2 × (0.06/60) × 60 + 1 × (0.06/60) × 60 + 4 × (0.06/60) × 60 = USD 0.42

## Billing rules for dedicated resource groups

• Subscription

The following formula describes how to calculate the subscription fee for each subscription dedicated resource group:

```
Bill amount of each resource group = Number of resources × Unit price × Subscription dur ation
```

- Valid range of subscription duration: 1 to 12 months. Each month contains 30 calendar days.
- After you purchase a dedicated resource group, the resource group is free of charge on the current day, and the subscription takes effect from the next day. For example, you purchased a dedicated resource group on July 31, 2019, and the subscription duration is one month. Then, the resource group expired at 00:00:00 (UTC+8) on August 31, 2019.
- Some resources may not be available in one or more regions for a short period of time. During the time period, you cannot purchase dedicated resource groups in the regions.
- Pay-as-you-go

Each pay-as-you-go dedicated resource group is billed based on the following formula:

```
Bill amount of each resource group = Number of resources \times (Unit price/60) \times Usage durati on
```

Onte Usage duration is measured in minutes.

• The billing starts immediately after a resource group enters the **Running** state. The billing stops immediately after a resource group enters the **No Node** state.

Notice Stop unused resource groups to avoid unnecessary costs.

- Bills are generated from the time when a resource group enters the **Running** state. The system stops billing immediately after a resource group releases all occupied resources and enters the **No Node** state.
- If you scale out a resource group, newly added resources are billed after the scale-out activity is complete. If you scale in a resource group, released resources are no longer billed, and only the remaining resources are billed.
- Fees are charged on a per minute basis. No fee is charged if a resource group runs for less than one minute.
- Some resources may not be available in one or more regions for a short period of time. During the time period, you cannot purchase dedicated resource groups in the regions.

#### Prices of dedicated resource groups

			Price (USD)					
Model	GPU specific ation	CPU specific ation	Mainlan d China	China (Hong Kong)	Singapo re (Singap ore)	lndonesi a (Jakarta)	lndia (Mumbai )	German y (Frankfu rt)
ecs.c5.6 xlarge	N/A	24 CPU Core+48 GB	365	366	596	569	492	535
ecs.g5.6 xlarge	N/A	24 CPU Core+96 GB	495	495	744	745	592	786
ecs.gn5i - c4g1.xla rge	1 NVIDIA Tesla P4	4 CPU Core+16 GB	452	N/A	N/A	N/A	N/A	N/A
ecs.gn5i - c8g1.2xl arge	1 NVIDIA Tesla P4	8 CPU Core+32 GB	543	N/A	N/A	N/A	N/A	N/A

The following table lists the prices of subscription dedicated resource groups.

			Price (USD					
Model	GPU specific ation	CPU specific ation	Mainlan d China	China (Hong Kong)	Singapo re (Singap ore)	Indonesi a (Jakarta)	India (Mumbai )	German y (Frankfu rt)
ecs.gn6i - c4g1.xla rge	1 Tesla T4	4 CPU Core+15 GB	570	691	727	727	776	722
ecs.gn6i - c8g1.2xl arge	1 Tesla T4	8 CPU Core+31 GB	686	819	862	862	908	870
ecs.gn6i - c16g1.4 xlarge	1 Tesla T4	16 CPU Core+62 GB	805	1076	1132	1132	1170	1166
ecs.gn6i - c24g1.6 xlarge	1 Tesla T4	24 CPU Core+93 GB	843	1349	1420	1420	1396	1472
ecs.gn5- c4g1.xla rge	1 NVIDIA P100	4 CPU Core+30 GB	627	928	977	977	929	1006
ecs.gn5- c8g1.2xl arge	1 NVIDIA P100	8 CPU Core+60 GB	755	1118	1177	1177	1118	1212
ecs.gn5- c28g1.7 xlarge	1 NVIDIA P100	28 CPU Core+11 2 GB	1171	1609	1693	1693	1734	1744
ecs.gn6v - c8g1.2xl arge	1 NVIDIA V100	8 CPU Core+32 GB	1297	N/A	2381	N/A	N/A	N/A

Note Subscription dedicated resource groups are available only in the China (Hangzhou), China (Shanghai), China (Beijing), and China (Shenzhen) regions.

The catalog prices of pay-as-you go dedicated resource groups are hourly prices. However, fees are charged on a per minute basis. You can calculate the unit prices per minute by dividing the prices listed in the following table by 60. The following table lists the unit prices per hour.

#### Pricing Metering and billing

			Price (USD)					
Model	GPU specific ation	CPU specific ation	Mainlan d China	China (Hong Kong)	Singapo re (Singap ore)	lndonesi a (Jakarta)	German y (Frankfu rt)	lndia (Mumbai )
ecs.g6.4 xlarge	N/A	16 CPU Core+64 GB	0.66	1.2	1.08	1.08	1.08	0.90
ecs.g6.6 xlarge	N/A	24 CPU Core+96 GB	1.02	1.8	1.68	1.68	1.62	1.38
ecs.c5.6 xlarge	N/A	24 CPU Core+48 GB	1.26	1.2	1.26	1.20	1.14	1.08
ecs.g5.6 xlarge	N/A	24 CPU Core+96 GB	1.8	1.68	1.74	1.62	1.56	1.32
ecs.gn5i - c4g1.xla rge	1 NVIDIA Tesla P4	4 CPU Core+16 GB	1.62	N/A	N/A	N/A	N/A	N/A
ecs.gn5i - c8g1.2xl arge	1 NVIDIA Tesla P4	8 CPU Core+32 GB	1.98	N/A	N/A	N/A	N/A	N/A
ecs.gn6i - c4g1.xla rge	1 Tesla T4	4 CPU Core+15 GB	1.98	1.32	1.50	1.44	1.38	1.50
ecs.gn6i - c8g1.2xl arge	1 Tesla T4	8 CPU Core+31 GB	2.40	1.56	1.80	1.68	1.68	1.74
ecs.gn6i - c16g1.4 xlarge	1 Tesla T4	16 CPU Core+62 GB	2.76	2.10	2.34	2.22	2.28	2.28
ecs.gn6i - c24g1.6 xlarge	1 Tesla T4	24 CPU Core+93 GB	2.94	2.64	2.94	2.76	2.88	2.70

Model	GPU specific ation	CPU specific ation	Price (USD/					
			Mainlan d China	China (Hong Kong)	Singapo re (Singap ore)	lndonesi a (Jakarta)	German y (Frankfu rt)	lndia (Mumbai )
ecs.gn5- c4g1.xla rge	1 NVIDIA P100	4 CPU Core+30 GB	2.16	2.04	2.16	2.16	1.92	2.04
ecs.gn5- c8g1.2xl arge	1 NVIDIA P100	8 CPU Core+60 GB	2.64	2.46	2.58	2.58	2.28	2.46
ecs.gn5- c28g1.7 xlarge	1 NVIDIA P100	28 CPU Core+11 2 GB	4.08	3.54	3.72	3.72	3.84	3.78
ecs.gn6v - c8g1.2xl arge	1 NVIDIA V100	8 CPU Core+32 GB	4.50	N/A	5.22	N/A	N/A	N/A

Note Pay-as-you-go dedicated resource groups are available only in the China (Hangzhou), China (Shanghai), China (Beijing), and China (Shenzhen) regions.

#### Billing examples for dedicated resource groups

For example, you subscribe to two T4 GPUs in the China (Hangzhou) region for three months. The specification of each GPU is 4 CPU cores and 15 GB of memory. The bill amount is calculated based on the following formula:

Bill amount =  $2 \times 570 \times 3 = USD 3,420$ 

If you purchase two pay-as-you-go ecs.g6.6xlarge (24 CPU cores + 96 GB of memory) ECS instances in the China (Hangzhou) region, and use the instances for 45 minutes, the bill amount is calculated based on the following formula:

Bill amount =  $2 \times (1.02/60) \times 45 = USD 1.53$ 

## 2.6. Billing of AutoLearning

AutoLearning is in public preview and is free of charge.

**Note** When models trained by AutoLearning are deployed to Elastic Algorithm Service (EAS) as online services, bills are generated for EAS. For more information, see **Billing of EAS**.

# 3.View bills and usage details

This topic describes how to view bills and usage details of the services provided by Machine Learning Platform for AI.

#### View bills

- 1. On the Alibaba Cloud homepage, move the pointer over the username in the upper-right corner, and click Bills.
- 2. On the Bills page, click the Bills tab.
- 3. In the **Billing Cycle** field, select a month. Then, click the ricon next to **Product Name** and select

UserCenter	Bilis								
	Ownwer         Bits         Details           Biting Cycket         2026/48         Account         ✓         Order/Bit No.1         Search								
III All Menus >	Product Name: Machine Learning × Cl	aar 🔴					Customize Column O	ptions 🚓 Export Billin	g Overview (CSV)
Fund Management	Billing Cycle Account Name	Product Name 🍸 Product Detail 🦷	Subscription Type	Payment Time	Order No./Bill No.	Item 😭	Pretax Gross Amount	Invoice Discount	Round Down Di
Income/Expense	2020-08 alidocs@test.aliyunid.com	Machine Learning PaiEasPostpay	Pay-As-You-Go	2020-08-09 20:00:00 ~ 2020-08-09 21:00:00	2020080417060166	Pay-As-You-Go	¥ 0.000000	¥ 0.000000	¥ 0.0
•	2020-08 alidocs@test.aliyunid.com	Machine Learning PalEasPostpay	Pay-As-You-Go	2020-08-09 19:00:00 ~ 2020-08-09 20:00:00	2020080417060156	Pay-As-You-Go	¥ 0.000000	¥ 0.000000	¥ 0.0
Bil	2020-08 alidocs@test.aliyunid.com	Machine Learning PaiEasPostpay	Pay-As-You-Go	2020-08-09 18:00:00 ~ 2020-08-09 19:00:00	2020080417060137	Pay-As-You-Go	¥ 0.000000	¥ 0.000000	¥ 0.0
	2020-08 alidocs@test.aliyunid.com	Machine Learning PaiEasPostpay	Pay-As-You-Go	2020-08-09 16:00:00 ~ 2020-08-09 17:00:00	2020080417060127	Pay-As-You-Go	¥ 0.000000	¥ 0.000000	¥ 0.0
Bill Management	2020-06 alidocs@test.aliyunid.com	Machine Learning PaiEasPostpay	Pay-As-You-Go	2020-08-09 15:00:00 ~ 2020-08-09 16:00:00	2020080417060117	Pay-As-You-Go	¥ 0.000000	¥ 0.000000	¥ 0.0

- 4. (Optional)Click the ricon next to **Product Detail** to display only the target service.
- 5. View bills of Machine Learning Platform for AI within the selected billing cycle.

#### View billing details

- 1. On the **Bills** page, click the **Details** tab.
- 2. In the Billing Cycle field, select a month.
- 3. (Optional)In the Instance Name field, select Instance ID, and enter an instance ID.

You can log on to the Machine Learning Platform for AI console, and navigate to the Notebook Models page and Elastic Algorithm Service page to view instance IDs. The IDs of instances in Machine Learning Studio indicate the types of algorithm components. The instance IDs include text\_analysis (a text analysis component), data\_analysis (a data analysis component), data\_manipulation (a data preprocessing component), deep\_learning (a deep learning component), and default (a default algorithm component).

- 4. In the Statistic Item section, select Billing Item. In the Statistic Period section, select Billing Period.
- 5. Click the  $\gamma$  icon next to **Product Name** and select **Machine Learning**.

User Center	Bills Overview Bills Critical	5	how Graph
Home Fund Management	Billing Cycler         2020-68         Resource Groups         Instance Name           Statistic Feature         Billing Cycler         Onstance         Instance         Instance Name           Statistic Feature         Billing Cycler         Onstance         Instance         Instance	Search by orderbill number. Search	
Income/Expense	Product Name: Machine Learning × Clear	② Customize Column Options 👌 Export Billing Ov	erview (CSV)
Bill	Billing Cycle Cost Center 🙄 Account Name 🙄 Product Name 😭 Product Detail 🙄 Subscript	tion Type 🖓 Payment Time Order No./Bill No. Instance ID Resou	urce Group
<b>BII 1</b>	2020-08 Not Allocated Machine Learning PalEasPostpay Pay-As-Y	00-60 2020-08-03 19:00:00 ~ 2020-08-03 20:00:00 -	
Bill Management	2020-08 Not Allocated Machine Learning Machine Learning Platform for Al Pay-As-Y	0u-Go 2020-07-31 00:00:00 ~ 2020-08-01 00:00:00 -	

 Bill Adalysis
 2020-06
 Not Allocated
 Machine Learning
 Machine Learning Platform for Al
 Pay-As-You-Gg
 2020-07-31 00 00:00 - 2020-08-01 00:00:00

- 6. (Optional)Click the ricon next to **Product Detail** to display only the target service.
- 7. View the **billing details** of Machine Learning Platform for AI within the selected **billing cycle**.

Onte The billing details are updated with a delay by one day.

#### View usage details

- 1. On the **Bills** page, choose **Bill > Usage Record** in the left-side navigation pane.
- 2. From the **Product** drop-down list, select learn.
- 3. Specify Billing Method and Time Period.

The **billing methods** of the services provided by Machine Learning Platform for AI are shown in the following list:

- Machine Learning Studio: PAI\_ALGO
- Data Science Workshop (DSW): PAI\_DSW\_PAY
- Elast ic Algorit hm Service (EAS): **PAI\_EAS\_POST PAID**

	Usage Records							
alidocs@test.aliyuni	Note: 1. The format of exported files is CSV. You can use spreadsheet software to view these files. 2. If an error occurs in an exported file, we recommend that you follow the instructions provided in the error message to export the file again. 3. If the size of an exported file exceeds the maximum size, the file may be separated into several files. We recommend that you change the previous search conditions and export the file again.							
III All Menus >								
Home	Product: leam							
Fund Management								
Income/Expense	* Billing Method: PALEAS_POSTPAID							
<b>•</b>	PAI_EAS_POSTPAID							
Bill	* Time Period: PAI_DSW_PAY_TEST ⑦							
Bill	PAI_EAS_POSTPAID_TEST							
Monthly Bill	Time Unit: PAI_DSW_PAY							
GAAP Bill	PAI_ALGO							
Usage Record	• verification Code: JFF 3 Get a new captcha							
Export History	Export CSV							

- 4. Select a Time Unit, and enter the Verification Code.
- 5. Click Export CSV, and navigate to the Export Record page.
- 6. When the status of the file displayed in the **Status** column changes from **Pending Download** to **Exported**, click **Download** in the **Actions** column to download usage details.

< Back Export Record					Refresh
Exported files will be saved on Allbaba Cloud for 3 days.					
File Name	Type 🖓	Format 🖓	Status	Created At	Actions
10000000000000000000000000000000000000	Usage Record	CSV	Exported	2020-08-10 11:09:24	Download