DataArtsFabric

Billing

Issue	01
Date	2025-07-11

HUAWEI CLOUD COMPUTING TECHNOLOGIES CO., LTD.

Trademarks and Permissions


and other Huawei trademarks are the property of Huawei Technologies Co., Ltd.
All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei Cloud and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Huawei Cloud Computing Technologies Co., Ltd.

Address: Huawei Cloud Data Center Jiaoxinggong Road
Qianzhong Avenue
Gui'an New District
Gui Zhou 550029
People's Republic of China

Website: https://www.huaweicloud.com/intl/en-us/

Contents

1 Overview

1.1 Ray and XDS Billing Overview

This document describes the billing modes, billing items, and renewal of Ray and XDS.

Billing Mode

DataArtsFabric offers two billing modes: yearly/monthly and pay-per-use.

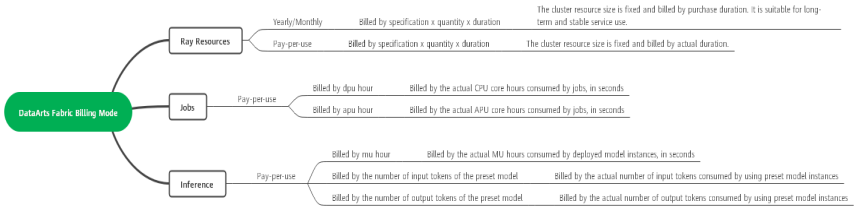
- Yearly/Monthly: This is a prepaid billing mode where you are charged based on the duration of your resource purchase. Longer purchase periods typically receive greater discounts. This mode is ideal for scenarios with stable compute resource demand throughout the purchase duration.
- Pay-per-use: With pay-per-use, you are billed after consuming resources, based on the actual usage duration of compute resources. Billing is precise to the second and settled hourly. This mode provides flexible resource usage aligned with service requirements, eliminating the need for pre-provisioning and preventing resource over- or under-allocation. It is suitable for scenarios where resource requirements fluctuate, ensuring you only pay for what you use.

Table 1-1 DataArtsFabric billing modes

Billing Mode	Payment Method	Billing Cycle	Applicable Billing Item
Yearly/ Monthly	Prepaid. Billed by the purchase period specified in your order.	Billed by the purchase period specified in the order	Ray resources

Billing Mode	Payment Method	Billing Cycle	Applicable Billing Item
Pay-per-use	Postpaid. Billed by usage duration.	Billed based on the actual usage duration of resources. A bill is generated every hour.	Ray resources, MU hours, SQL warm-up resources, and SQL compute unit hours

The following figure shows the billing modes of DataArtsFabric in different service scenarios.



Billing Items

DataArtsFabric implements distinct billing policies based on its Ray and inference service scenarios.

Table 1-2 Billing items

Billing Item	*Billing description*
Ray resources	You are billed based on the specifications and quantity of Ray resources provisioned. Pricing varies by Data Processing Unit (DPU) or AI Compute Unit (ACU) specifications. Both yearly/monthly and pay-per-use billing modes are available.
Model compute unit hours	Billing is based on the compute unit hours consumed by model instances deployed on inference endpoints. This item supports pay-per-use billing. The cost is calculated as: **(Number of model instances under an inference endpoint) × (Number of compute units) × (Usage duration reported in seconds)**. Refer to **Common Models** for specific compute unit requirements of different base models.

Bills

On the top navigation bar of the console, click **Billing**. In the left navigation pane of the Billing Center console, choose **Billing > Transaction and Detailed Bills** to view bill details. For details, see **Bills**.

Renewal/Arrears

For details, see **Expiration and Renewal**.

Stopping Billing

If you no longer need to use your cloud services, you can unsubscribe from or delete them to stop the billing. For details, see **Stopping Billing**.

Billing Examples

NOTICE

The following prices are for reference only. For the actual prices, see pricing details for each service.

1. **DataArtsFabric Ray resources**

   DataArtsFabric Ray resources are billed on a pay-per-use basis, with charges calculated per second. The hourly price is detailed in the DataArtsFabric product price list; divide this by 3600 to determine the per-second rate.

   Example: If a **fabric.ray.dpu.d1x** pay-per-use instance costs ¥0.2/hour, and you use 5 instances:

   - For 30 minutes of usage: (0.2/3600) x 5 x 30 x 60 = ¥0.5.
   - For 1 hour of usage: (0.2/3600) x 5 x 60 x 60 = ¥1.

2. **DataArtsFabric model computing unit (MU) hours**

   MU hours are billed on a pay-per-use basis, with charges calculated per second. The hourly price is specified in the DataArtsFabric product price details; divide this by 3600 to obtain the per-second rate.

   Example: A base model, such as LLAMA3_8B, consumes 2 MUs per inference endpoint instance. Assuming the MU hour price is ¥30.0/hour, the price is calculated based on the number of deployed model instances, multiplied by the MU consumption per instance, and then by the actual usage duration (seconds).

   - If one model service instance is used for 30 minutes (and the number of instances remains constant), the price is: (1/3600) x 1 x 2 x 30 x 60 = ¥30.
   - If, over 1 hour, two service instances are deployed for the first 15 minutes, and then one service instance for the remaining 45 minutes, the price is: (30/3600) x 1 x 2 x 45 x 60 + (30/3600) x 2 x 2 x 15 x 60 = ¥75.

1.2 DataArtsFabric SQL Billing Overview

This document describes the billing mode and billing items of DataArtsFabric SQL.

Billing Mode

DataArtsFabric SQL offers two serverless pay-per-use billing modes: Pay By Resource and Pay By Query.

- Pay By Resource: You are billed for **the number of warm-up resources** created for SQL endpoints, multiplied by their **running duration**. You can enable and disable the service at any time. Billing is precise to the second and settled hourly. A bill is generated every calendar hour, and fees are deducted from your account balance. This billing mode provides dedicated resources and a high SLA.
- Pay By Query: You are billed based on the **SQL computing unit hours** consumed by SQL statements running on public endpoints. Billing is precise to the second and settled hourly. A bill is generated every calendar hour, with fees deducted from your account balance. This billing mode uses shared resources and has a medium SLA.

For details about the billing differences, see **Table 1 Billing modes of DataArtsFabric SQL**.

Table 1-3 Billing modes of DataArtsFabric SQL

Billing Mode	Payment Method	Billing Cycle	Applicable Billing Item
Pay By Resource	Postpaid. Billed hourly, based on the running duration of purchased warm-up resources.	A CDR (Charge Detail Record) is generated every hour.	SQL warm-up resources
Pay By Query	Postpaid. Billed hourly, based on the SQL computing unit hours consumed by statements running on public endpoints.	A CDR is generated every hour.	SQL computing unit hours

Billing Items

DataArtsFabric SQL is billed based on the resources used in actual service scenarios. For details about the billing items, see **Table 1-4**.

Table 1-4 Billing items

Billing Item	Billing Description
SQL warm-up resources	Billed on a pay-per-use basis, this item charges for the running duration of warm-up resources created for SQL endpoints. Usage is calculated by multiplying the actual **number of warm-up resources** by their **running duration**, with usage reported per second.

Billing Item	Billing Description
SQL computing unit hours	Billed on a pay-per-use basis, this item charges for the **number of SQL computing unit resources** consumed by SQL statements running on public endpoints, multiplied by their **usage duration**. Usage is reported per second and aggregated every calendar hour.

Bills

On the top navigation bar of the console, click **Billing**. In the left navigation pane of the Billing Center console, choose **Billing > Transaction and Detailed Bills** to view bill details. For details, see **Bills**.

Arrears

For details, see **Expiration and Renewal**.

Stopping Billing

If you no longer need to use your cloud services, you can unsubscribe from or delete them to stop the billing. For details, see **Stopping Billing**.

Billing Examples

NOTICE

The following prices are for reference only. For the actual prices, see pricing details for each service.

1. **SQL warm-up resources**

   SQL warm-up resources are billed per second in pay-per-use mode. The hourly price is available in the DataArtsFabric product price details; divide this by 3600 to obtain the per-second rate.

   Example: Consider a **fabric.sql.dcu.std** flavor instance priced at 0.155 USD/core/hour. If you provision fifty warm-up resources, the cost is calculated by multiplying the **number of resources** by the actual **running duration** (billed to the second).

   - For 30 minutes of runtime: (0.155/3600)*50*30*60=3.875 USD.
   - For 1 hour of runtime: (0.155/3600)*50*60*60=7.75 USD.

2. **SQL Computing Unit (CU) hours**

   SQL computing unit hours are billed per second in pay-per-use mode. The hourly price is listed in the DataArtsFabric product price details; divide this by 3600 to obtain the per-second rate.

   Example: For a **fabric.sql.query** flavor instance priced at 0.597 USD/core/hour, the fee is calculated based on the DPU (Data Processing Unit) duration (core-seconds) consumed by running SQL statements, billed per second.

- From 9:00 to 10:00, three SQL statements are executed: Statement 1 (2.4 seconds), Statement 2 (3.5 seconds), and Statement 3 (6.3 seconds). The total DPU duration for this hour is 12.2 seconds. The fee is: (0.2015/3600)*12 =0.00067 USD.
- From 10:00 to 11:00, Statement 1 is executed but canceled mid-execution, consuming 18 seconds of DPU duration. Canceled SQL statements are also charged. The fee is: (0.2015/3600)*18=0.0010075 USD.

2 Billing Items

2.1 Ray and XDS Billing Items

Fabric implements distinct billing policies based on its Ray and inference service scenarios. For details, see **Table 2-1**.

Table 2-1 Billing items

Billing Item	*Billing description*
Ray resources	You are billed based on the specifications and quantity of Ray resources provisioned. Pricing varies by Data Processing Unit (DPU) or AI Compute Unit (ACU) specifications. Both yearly/monthly and pay-per-use billing modes are available.
Model compute unit hours	Billing is based on the compute unit hours consumed by model instances deployed on inference endpoints. This item supports pay-per-use billing. The cost is calculated as: **(Number of model instances under an inference endpoint) × (Number of compute units) × (Usage duration reported in seconds)**. Refer to **Common Models** for specific compute unit requirements of different base models.

2.2 DataArtsFabric SQL Billing Items

DataArtsFabric SQL is billed based on the resources used in actual service scenarios. For details about the billing items, see **Table 2-2**.

Table 2-2 SQL billing items

Billing Item	*Billing description*
SQL warm-up resources	Billed on a pay-per-use basis, this item charges for the running duration of warm-up resources created for SQL endpoints. Usage is calculated by multiplying the actual **number of warm-up resources** by their **running duration**, with usage reported per second.
SQL computing unit hours	Billed on a pay-per-use basis, this item charges for the **number of SQL computing unit resources** consumed by SQL statements running on public endpoints, multiplied by their **usage duration**. Usage is reported per second and aggregated every calendar hour.

3 Billing Mode

DataArtsFabric offers two billing modes: yearly/monthly and pay-per-use.

- Yearly/Monthly: This is a prepaid billing mode where you are charged based on the duration of your resource purchase. Longer purchase periods typically receive greater discounts. This mode is ideal for scenarios with stable compute resource demand throughout the purchase duration.
- Pay-per-use: With pay-per-use, you are billed after consuming resources, based on the actual usage duration of compute resources. Billing is precise to the second and settled hourly. This mode provides flexible resource usage aligned with service requirements, eliminating the need for pre-provisioning and preventing resource over- or under-allocation. It's best suited for scenarios where resource requirements fluctuate, ensuring you only pay for what you use.

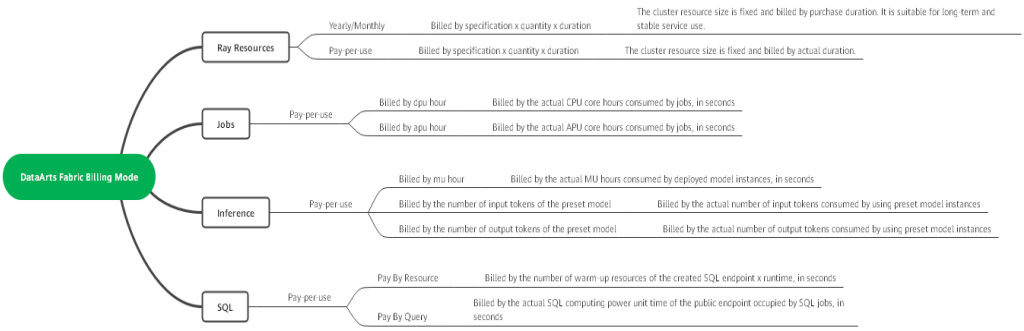
For details about the billing differences, see **Table 3-1**.

Table 3-1 DataArtsFabric billing modes

Billing Mode	Payment Method	Billing Cycle	Applicable Billing Item
Yearly/Monthly	Prepaid. Billed by the purchase period specified in your order.	Billed by the purchase period specified in the order	Ray resources
Pay-per-use	Postpaid. Billed by usage duration.	Billed based on the actual usage duration of resources. A bill is generated every hour.	Ray resources, MU hours, SQL warm-up resources, and SQL compute unit hours

Figure 3-1 shows the billing modes of DataArtsFabric in different service scenarios.

Figure 3-1 DataArtsFabric billing modes



4 Modifying Configurations

Currently, only Ray and SQL warm-up resources support configuration modifications. Other service scenarios utilize pay-per-use billing and do not involve order changes and billing mode changes.

Table 4-1 details the impact of modifying Ray or SQL warm-up resource sizes on billing.

Table 4-1 Impact on fees

Billing Mode	Change Scenario	Description
Pay-per-use	Changing the number of Ray resources (specifications upgrade/ downgrade)	The change will take effect immediately.
Pay-per-use	Changing the number of SQL warm-up resources (specifications upgrade/ downgrade)	The change will take effect immediately.
Yearly/ Monthly	Increasing the number of Ray resources (specification upgrade, with price difference charged)	The new resource quantity takes effect immediately upon specification upgrade and will be billed accordingly for the remainder of the original period.

Billing Mode	Change Scenario	Description
Yearly/ Monthly	Decreasing the number of Ray resources (specifications downgrade)	The new resource size takes effect immediately upon specification downgrade within the original period. The price difference will be refunded based on the used period.

5 Bills

- **Bill reporting period**

  The usage of pay-per-use resources is reported to the billing system at a fixed interval for settlement. These resources can be settled hourly, daily, or monthly, depending on the usage type.

  Example: If a cloud server, billed hourly, is deleted at 08:30, the fee for the 08:00 to 09:00 period will be deducted around 10:00. On the **Billing Center** > **Billing** > **Bills** > **Transaction Bills** page, **Expenditure Time** indicates the time when a pay-per-use product is used.

- **Viewing a complete bill**

  You can view both monthly summary bills and detailed bills on Huawei Cloud.

  - Summary bill: Summary data displays information such as payable amounts and fee deduction details across various dimensions. Only one summary record is shown per product. The final summary bill for the current month is generated on the third day of the following month and becomes viewable and exportable after 10:00 on the fourth day.
  - Bill details: You can view bill details in different dimensions, including transaction bills and custom bills. Custom bills can be filtered by usage, resource, or product.

- **Viewing the bill of a specified resource**
  - Querying Ray Resource Bills
    i. Obtain the resource ID from the Ray resource page in the cloud service console.
    ii. View the resource bill in Billing Center based on the resource ID. For details, see **How Do I Find Cloud Service Resources By Resource Names or IDs?** in Billing Center.
  - Querying Inference Endpoint CU Hour Bills

    The inference endpoint ID for DataArtsFabric services differs from the resource ID reported in the bill. The billing resource ID for an inference endpoint corresponds to the Model Unit (MU) and follows the format mu.{*Endpoint ID*}. For example, if the inference endpoint ID is **32de36ea-26c0-4876-ae48-fdbbb03cd455**, the resource ID reported to the bill is **mu.32de36ea-26c0-4876-ae48-fdbbb03cd455**.

    i. Obtain the endpoint ID from the inference endpoint page in the cloud service console.

ii. Construct the corresponding MU resource ID.

iii. View the resource bills in the Billing Center based on the resource ID. For details, see **How Do I Find Cloud Service Resources By Resource Names or IDs?** in Billing Center.

- Querying SQL Warm-Up Resource Bills

  i. Obtain the endpoint ID from the SQL endpoint page on the cloud service console. This endpoint ID is the resource ID in the bill.

  ii. View the resource bill in Billing Center based on the resource ID. For details, see **How Do I Find Cloud Service Resources By Resource Names or IDs?** in Billing Center.

- Querying SQL Compute Unit Bills for Public Endpoints

  When the public endpoint service is enabled for DataArtsFabric SQL, the SQL compute unit's resource ID is formatted as {*First six digits of the workspace ID*}_{*Public endpoint ID*}. For example, if the workspace ID is **6eeda84c-9ac9-4464-b1d8-122fe5c235c1** and the public endpoint ID is **0288b810-d119-4d51-b6d5-fccee06e209f**, the billed resource ID for the SQL computing unit is **6eeda8_0288b810-d119-4d51-b6d5-fccee06e209f**.

  i. From the DataArtsFabric homepage in the cloud service console, view and obtain the workspace ID.

  ii. Obtain the public endpoint ID from the SQL endpoint page in the cloud service console.

  iii. Combine these to form the resource ID. For details, see **How Do I Find Cloud Service Resources By Resource Names or IDs?** in Billing Center.

For more information about bills, see **Bill Management (Old Version)** or **Bill Management (New Version)**.

6 Stopping Billing

After reviewing your bill, you can stop charges for certain resources by following these steps:

1. Obtain resource information (e.g., IDs or names) from your bill.
2. Locate the corresponding cloud service resources within the cloud service console.
3. Stop billing for these resources.

The following is an example:

- Ray resources: To stop billing, you must delete or unsubscribe from the Ray resources. Be aware that deleting Ray resources may render existing Ray clusters unavailable.
- Job endpoints:
  a. Public endpoints: No charges are incurred unless jobs are actively executed using the public endpoint.
  b. Self-built endpoints: Warm-up resources are continuously charged after endpoint creation until the endpoint is deleted. Elastic resources are charged based on usage when the number of running jobs exceeds the warm-up resources. No fees are generated when no jobs are running.
- Inference MU: Billing stops when you delete the inference service instance under an inference endpoint or delete the inference endpoint itself.
- SQL resources:
  – SQL warm-up resources: Fees are continuously generated after endpoint creation. You must delete the endpoint to stop billing.
  – SQL computing units (public endpoints): No fees are generated unless SQL jobs are actively executed using the public endpoint.

7 Expiration and Renewal

If your account is in arrears, you can view the arrears details in the Billing Center. To prevent related resources from being stopped or released, you need to top up your account within the specified period. For details, see **Top-Up and Payment**.

If you do not renew or top up your account in time, your resources enter a grace period. If you still do not complete the payment or renewal after the grace period expires, you will enter a retention period. During this period, the resources will be suspended. If you still do not complete the payment or renewal after the retention period has ended, your data stored in the cloud service will be deleted and the resource will be released. For details, see **Resource Suspension and Release**.

Resource Expiration

If your account is in arrears, a retention period is provided based on your service tier. Once this retention period begins, your Ray resources, SQL resources, and model instances within the DataArtsFabric service are retained, but your account enters a restricted state. In the restricted state, you cannot create or use endpoints on the console, but you can still perform other operations. If the outstanding balance is not settled before the retention period expires, all data stored in DataArtsFabric will be permanently deleted and cannot be recovered.

For more information about the retention period, see **What Is a Retention Period of Huawei Cloud? How Long Is It?**

Renewal

Yearly/monthly Ray resources are renewable. You can renew them from the Renewal Management page on the management console. You can also set auto-renewal upon expiration. For more information, see **Renewal Management**.

