

Data Lake Insight

Service Bulletin

Issue 01
Date 2024-08-19



Copyright © Huawei Technologies Co., Ltd. 2024. All rights reserved.

No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of Huawei Technologies Co., Ltd.

Trademarks and Permissions



HUAWEI and other Huawei trademarks are trademarks 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 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.

Security Declaration

Vulnerability

Huawei's regulations on product vulnerability management are subject to the *Vul. Response Process*. For details about this process, visit the following web page:

<https://www.huawei.com/en/psirt/vul-response-process>

For vulnerability information, enterprise customers can visit the following web page:

<https://securitybulletin.huawei.com/enterprise/en/security-advisory>

Contents

1 Product Bulletin.....	1
1.1 EOL Announcement for DLI Yearly/Monthly and Pay-per-Use Queues as Well as Queue CUH Packages.....	1
1.2 EOS Announcement for DLI Flink 1.10 and Flink 1.11.....	4
1.3 EOS Announcement for DLI Spark 2.3.2.....	5
1.4 EOS Announcement for DLI Flink 1.7.....	6
2 Version Support Bulletin.....	8
2.1 Lifecycle of DLI Compute Engine Versions.....	8
2.2 What's New in Flink 1.15.....	9
2.3 What's New in Flink 1.12.....	10
2.4 What's New in Spark 3.3.1.....	10
2.5 What's New in Spark 3.1.1.....	12
2.6 What's New in Spark 2.4.5.....	13

1 Product Bulletin

1.1 EOL Announcement for DLI Yearly/Monthly and Pay-per-Use Queues as Well as Queue CUH Packages

Description

To improve resource sharing and increase the utilization of compute resources, the DLI team is upgrading yearly/monthly and pay-per-use queues to elastic resource pool queues. This means if you need to use DLI compute resources, you will need to buy an elastic resource pool and create queues within it.

- Huawei Cloud schedules an end of marketing (EOM) for DLI queues billed in yearly/monthly and pay-per-use modes and queue CUH packages on March 31, 2024, at 00:00:00 (GMT+08:00).
- Huawei Cloud schedules an end of life (EOL) for DLI queues billed in yearly/monthly and pay-per-use modes and queue CUH packages on June 30, 2025, at 00:00:00 (GMT+08:00).

Impact

- Once the EOM is reached, new DLI queues billed in yearly/monthly and pay-per-use modes and queue CUH packages, cannot be purchased.
Until June 30, 2024 at 00:00:00 (GMT+08:00), you can renew your queues for up to one year or modify their specifications to meet your service needs.
After June 30, 2024, at 00:00:00 (GMT+08:00), it will no longer be possible to renew or change your queues.
- Once the EOL is reached, queues will no longer be usable. Therefore, it is necessary to use an elastic resource pool or the **default** queue before the EOL. We recommend purchasing an elastic resource pool and creating queues within it to enjoy a wider range of DLI product capabilities.

Lifecycle of DLI Yearly/Monthly and Pay-per-Use Queues

Table 1-1 Lifecycle of DLI yearly/monthly and pay-per-use queues

Function	Status	EOM Date	EOL Date
DLI yearly/ monthly and pay-per-use queues	EOM	March 31, 2024	June 30, 2025

NOTE

- EOM: indicates that the sales of this function are stopped.
- EOL: indicates that all sales and service activities are stopped.

FAQ

- **What will happen to jobs running on DLI queues with yearly/monthly or pay-per-use billing after the EOM?**

Once the EOM is reached, new queues cannot be purchased.

 - Short-term solution: Until June 30, 2024 at 00:00:00 (GMT+08:00), you can renew your queues for up to one year or modify their specifications to meet your service needs.
 - Long-term solution: Use the elastic resource pool or **default** queue. You are advised to purchase an elastic resource pool and create queues in the pool.
- **What will happen to jobs running on DLI queues with yearly/monthly or pay-per-use billing after the EOL?**

Once the EOL is reached, queues billed in yearly/monthly or pay-per-use mode will no longer be able to execute jobs.

Before the EOL, you need to move your jobs from yearly/monthly and pay-per-use queues to queues in an elastic resource pool. To do this, you will need to purchase an elastic resource pool, create a queue within the pool, and then run your jobs on that queue.
- **What function can be used as a replacement after the EOM and EOL?**

If you are running jobs on queues billed in yearly/monthly or pay-per-use mode, use an elastic resource pool or the **default** queue as soon as possible.

 - For pay-per-use dedicated queues, they can be directly moved to newly purchased elastic resource pools.
 - For yearly/monthly or pay-per-use non-dedicated queues, you will need to unsubscribe from them first, purchase new elastic resource pools, and then execute jobs in those pools.
- **What are the advantages of elastic resource pool queues compared with yearly/monthly and pay-per-use queues?**
 - Yearly/Monthly and pay-per-use queues: Such queues have predetermined resource specifications. However, if a job's resource

requirements fluctuate, the queue resources may either go to waste or prove to be insufficient.

- Elastic resource pool queue: Dynamic scaling improves resource utilization.
- **Do I have to pay for an elastic resource pool? Is a queue created in an elastic resource pool charged separately?**

Elastic resource pools support the pay-per-use and package billing modes. For more information about the billing, see *Data Lake Insight Billing*.

Queues added to an elastic resource pool are not billed separately, but be included in the billing for the elastic resource pool.

- Pay-per-use: You are billed based on the actual CUs of the elastic resource pool.
- Yearly/Monthly: You are billed based on the actual CUs of the elastic resource pool, with the specification part billed yearly/monthly and any excess billed on a pay-per-use basis.
- Elastic resource pool CUH package: You are billed based on the price of the purchased package, with the specification within the package billed on a yearly/monthly basis. Any excess beyond the package specification will be billed on a pay-per-use basis.

For more billing information about elastic resource pools, see [Billing for Elastic Resource Pools](#).

- **How do I create an elastic resource pool queue?**
 - a. **Buy an elastic resource pool.**
 - i. On the DLI management console, choose **Resources > Resource Pool**.
 - ii. On the **Resource Pool** page, click **Buy Resource Pool** in the upper right corner.
Set parameters as instructed, click **Buy**, confirm the configuration, and click **Submit**.
 - b. **Create a queue in the elastic resource pool.**

Create one or more queues in the elastic resource pool to run jobs.

 - i. Switch to the **Resource Pool** page.
 - ii. Locate the target elastic resource pool and click **Add Queue** in the **Operation** column.
 - iii. On the **Add Queue** page, configure basic queue information. Click **OK**.
 - c. **Create a job.**

Create a job and run it on the queue you have created.

Announcement published on: November 21, 2023

1.2 EOS Announcement for DLI Flink 1.10 and Flink 1.11

Description

Huawei Cloud schedules an end of service (EOS) for DLI Flink 1.10 and Flink 1.11 at 00:00 (GMT+08:00) on December 31, 2023.

Impact

After the EOS, no technical support will be provided for DLI Flink 1.10 and Flink 1.11. You are advised to select the Flink engine of the latest version when executing jobs. DLI Flink 1.12 is recommended.

For jobs that are using Flink 1.10 or Flink 1.11, switch to the Flink engine of the latest version as soon as possible. Otherwise, no technical support will be provided if an error occurs during job execution.

FAQ

- **How does the EOS affect the jobs that are using Flink 1.10 or 1.11?**

If a queue is created after the EOS of Flink 1.10 or 1.11, the compute engine that has reached EOS cannot be selected during job execution.

Historical queues can still use Flink 1.10 or Flink 1.11 to execute jobs. However, if an error occurs during job execution, no technical support is provided. Replace the compute engine with a new version as soon as possible.
- **Which version can be used as a replacement after the EOS?**

DLI Flink 1.12 is recommended.
- **What are the advantages of Flink 1.12?**

Flink 1.12 supports DataGen, GaussDB(DWS), JDBC, MySQL CDC, Postgres CDC, Redis, Upsert Kafka, and HBase source tables.

For more advantages, see [Flink 1.12 Upgrade Guide](#).
- **Does the upgrade of Flink affect the DLI resource price?**

DLI bills you based on the amount of compute and storage resources consumed by jobs, regardless of the compute engine version.
- **How do I upgrade Flink to version 1.15?**
 - a. On the DLI management console, buy an elastic resource pool and create queues within the pool to provide compute resources required for job execution.
 - b. Log in to the DLI management console. In the navigation pane on the left, choose **Job Management** > **Flink Jobs**.
 - c. On the **Flink Jobs** page, locate the row that contains the target job and click **Edit** in the **Operation** column.
 - d. On the page displayed, select the latest Flink version. Flink 1.15 is recommended.

Announcement published on: July 6, 2023

1.3 EOS Announcement for DLI Spark 2.3.2

Description

Huawei Cloud schedules an end of service (EOS) for DLI Spark 2.3.2 at 00:00 (GMT+08:00) on December 31, 2023.

Impact

After the EOS, no technical support will be provided for DLI Spark 2.3.2. You are advised to select the Spark engine of the latest version when executing jobs. DLI Spark 2.4.5 is recommended.

For jobs that are using DLI Spark 2.3.2, switch to the Spark engine of the latest version as soon as possible. Otherwise, no technical support will be provided if an error occurs during job execution.

FAQ

- **How does the EOS affect the jobs that are using DLI Spark 2.3.2?**

If a queue is created after the EOS of Spark 2.3.2, the compute engine that has reached EOS cannot be selected during job execution.

Historical queues can still use Spark 2.3.2 to execute jobs. However, if an error occurs during job execution, no technical support is provided. Replace the compute engine with a new version as soon as possible.

- **Which version can be used as a replacement after the EOS?**

DLI Spark 2.4.5 is recommended.

- **What are the advantages of DLI Spark 2.4.5?**

Table 1-2 Advantages of Spark 2.4.5

Feature	Description
Merging small files	If a large number of small files are generated during SQL execution, job execution and table query will take a long time. In this case, you are advised to merge small files.
Modifying column comments of non-partitioned or partitioned tables	You can modify the column comments of non-partitioned or partitioned tables.
Collecting statistics on the CPU usage of SQL jobs	You can view the total CPU used on the console.
Viewing Spark logs of container clusters	You need to view logs in the container.

Feature	Description
Dynamic UDF loading (OBT)	The UDF takes effect without restarting the queue.
Supporting flame graphs on the Spark UI	Flame graphs can be created on the Spark UI.
Optimizing the query performance of the NOT IN statement for SQL jobs	The query performance of the NOT IN statement is improved.
Optimizing the query performance of the Multi-INSERT statement	The query performance of the Multi-INSERT statement is improved.

For more advantages, see [Spark SQL Upgrade Guide](#).

- **Does the upgrade affect the DLI resource price?**

DLI bills you based on the amount of compute and storage resources consumed by jobs, regardless of the compute engine version.

- **How do I upgrade DLI Spark to version 2.4.5?**

- a. On the DLI management console, buy an elastic resource pool and create queues within the pool to provide compute resources required for job execution.
- b. In the navigation pane on the left, choose **Job Management > Spark Jobs**. On the displayed page, click **Create Job** in the upper right corner.
- c. On the **Spark Jobs** page, locate the row that contains the target job and click **Edit** in the **Operation** column.
- d. On the page displayed, select the latest Spark version. Spark 2.4.5 is recommended.

Announcement published on: July 6, 2023

1.4 EOS Announcement for DLI Flink 1.7

Description

Huawei Cloud schedules an end of service (EOS) for DLI Flink 1.7 at 00:00 (GMT +08:00) on December 31, 2022.

Impact

After the EOS, no technical support will be provided for DLI Flink 1.7. You are advised to select the Flink engine of the latest version when executing jobs. DLI Flink 1.12 is recommended.

For jobs that are using Flink 1.7, switch to the Flink engine of the latest version as soon as possible. Otherwise, no technical support will be provided if an error occurs during job execution.

FAQ

- **Which functions of Flink 1.7 will not be evolved?**
 - The Flink Edge SQL function will no longer be supported for edge job processing after Flink 1.7 EOS, and subsequent versions of Flink will not support it either.
 - Similarly, the sensitive variable function will no longer be supported after Flink 1.7 EOS, and subsequent versions will not support it either.
- **How does the EOS affect the jobs that are using Flink 1.7?**

If a queue is created after the EOS of Flink 1.7, the compute engine that has reached EOS cannot be selected during job execution.

If you encounter any errors when Flink 1.7 is used to execute jobs on historical queues, note that this version will no longer receive any technical support. It is recommended that you switch to a later version of the compute engine as soon as possible.
- **Which version can be used as a replacement after the EOS?**

DLI Flink 1.12 is recommended.
- **What are the advantages of Flink 1.12?**

Flink 1.12 supports DataGen, GaussDB(DWS), JDBC, MySQL CDC, Postgres CDC, Redis, Upsert Kafka, and HBase source tables.

For more advantages, see [Flink 1.12 Upgrade Guide](#).
- **Does the upgrade of Flink affect the DLI resource price?**

DLI bills you based on the amount of compute and storage resources consumed by jobs, regardless of the compute engine version.
- **How do I upgrade Flink to version 1.12?**
 - a. On the DLI management console, buy an elastic resource pool and create queues within the pool to provide compute resources required for job execution.
 - b. Log in to the DLI management console. In the navigation pane on the left, choose **Job Management > Flink Jobs**.
 - c. On the **Spark Jobs** page, locate the row that contains the target job and click **Edit** in the **Operation** column.
 - d. On the page displayed, select the latest Flink version. Flink 1.12 is recommended.

Announcement published on: July 6, 2023

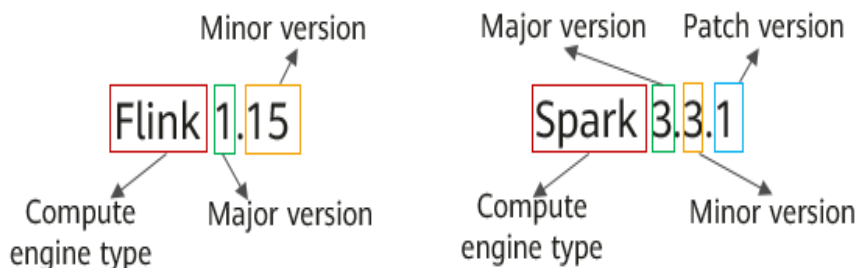
2 Version Support Bulletin

2.1 Lifecycle of DLI Compute Engine Versions

Version Description

DLI compute engine version is in *Compute engine name x.y.z* format. *Compute engine name* can be **Flink** or **Spark**. [Figure 2-1](#) describes the version.

Figure 2-1 DLI compute engine version description



Version Support

- Recommended Flink version: Flink 1.15
- Recommended Spark version: Spark 3.3.1

NOTE

You are not advised to use Spark/Flink engines of different versions for a long time.

- Doing so can lead to code incompatibility, which can negatively impact job execution efficiency.
- Doing so may result in job execution failures due to conflicts in dependencies. Jobs rely on specific versions of libraries or components.

Lifecycle of a Compute Engine Version

[Table 2-1](#) lists the lifecycle of DLI compute engine versions, based on which you can plan your version update pace.

Table 2-1 Lifecycle of DLI compute engine versions

Compute Engine Type	Version	Status	EOM Date	EOS Date
Flink	DLI Flink 1.15	Released	June 30, 2025	June 30, 2026
	DLI Flink 1.12	EOM	December 31, 2023	December 31, 2024
	DLI Flink 1.11	EOS	June 30, 2022	December 31, 2023
	DLI Flink 1.10	EOS	June 30, 2022	December 31, 2023
	DLI Flink 1.7	EOS	December 31, 2021	December 31, 2022
Spark	DLI Spark 3.3.1	Released	June 30, 2025	June 30, 2026
	DLI Spark 3.1.1	EOM	December 31, 2023	December 31, 2024
	DLI Spark 2.4.5	EOM	December 31, 2023	December 31, 2024
	DLI Spark 2.3.2	EOS	June 30, 2022	December 31, 2023

NOTE

- End of Marketing (EOM): indicates that the sales of this version are stopped.
- End of Service & Support (EOS): Services of this version are stopped. You are advised to use the engine of the latest version when running jobs. After this date, Huawei Cloud will no longer provide any technical support for the software version.

2.2 What's New in Flink 1.15

DLI complies with the release consistency of the open source Flink compute engine. This section describes the updates in Flink 1.15.

For details about Flink 1.15, see [Release Notes - Flink Jar 1.15](#) and .

Flink 1.15 Release Date

Version	Release Date	Status	EOM Date	EOS Date
DLI Flink 1.15	June 2023	Released	June 30, 2025	June 30, 2026

For more version support information, see [Lifecycle of DLI Compute Engine Versions](#).

Flink 1.15 Description

- The syntax design of Flink 1.15 has been improved to achieve higher compatibility and consistency with mainstream open-source technology standards.
- Flink 1.15 has added support for new connectors such as Hive and Hudi.
- Flink 1.15 now supports Mysql CDC 2.4.0.
- Flink 1.15 now supports integration with DEW-CSMS secret management, providing a privacy protection solution.

2.3 What's New in Flink 1.12

DLI complies with the release consistency of the open source Flink compute engine. This section describes the updates in Flink 1.12.

For more information about Flink 1.12, see [Release Notes - Flink 1.12](#).

Flink 1.12 Release Date

Version	Release Date	Status	EOM Date	EOS Date
DLI Flink 1.12	December 2021	Released	December 31, 2023	December 31, 2024

For more version support information, see [Lifecycle of DLI Compute Engine Versions](#).

Flink 1.12 Description

- Added support for DataGen, GaussDB(DWS), JDBC, MySQL CDC, Postgres CDC, Redis, Upsert Kafka, and HBase source tables.
- Added support for the merge of small files.
- Added support for Redis and RDS dimension tables.

2.4 What's New in Spark 3.3.1

DLI complies with the release consistency of the open source Spark compute engine. This section describes the updates in Spark 3.3.1.

For more information about Spark 3.3.1, see [Spark Release Notes](#).

Spark 3.3.1 Release Date

Version	Release Date	Status	EOM Date	EOS Date
DLI Spark 3.3.1	June 2023	Released	June 30, 2025	June 30, 2026

For more version support information, see [Lifecycle of DLI Compute Engine Versions](#).

Spark 3.3.1 Description

[Table 2-2](#) lists the main features of Spark 3.3.1.

For more information on new features and performance optimizations, see [Release Notes - Spark 3.3.1](#).

Table 2-2 Advantages of Spark 3.3.1

Feature	Description
Native performance acceleration	Improved the performance of Spark query statements.
Metadata access performance improvement	Improved Spark's metadata access performance for handling big data and enhanced data processing efficiency.
Improving the performance of OBS Committer when writing small files	Improved the performance of Object Storage Service (OBS) when writing small files, improving data transfer efficiency.
Dynamic executor shuffle data optimization	Improved the stability of resource scaling and cleaned up Executors when shuffle files are no longer needed.
Merging small files	If a large number of small files are generated during SQL execution, job execution and table query will take a long time. In this case, you are advised to merge small files.
Modifying column comments of non-partitioned or partitioned tables	You can modify the column comments of non-partitioned or partitioned tables.
Collecting statistics on the CPU usage of SQL jobs	You can view the total CPU used on the console.

Feature	Description
Viewing Spark logs of container clusters	You need to view logs in the container.
Dynamic UDF loading (OBT)	The UDF takes effect without restarting the queue.
Supporting flame graphs on the Spark UI	Flame graphs can be created on the Spark UI.
Optimizing the query performance of the NOT IN statement for SQL jobs	The query performance of the NOT IN statement is improved.
Optimizing the query performance of the Multi-INSERT statement	The query performance of the Multi-INSERT statement is improved.

2.5 What's New in Spark 3.1.1

DLI complies with the release consistency of the open source Spark compute engine. This section describes the updates in Spark 3.1.1.

For more information about Spark 3.1.1, see [Spark Release Notes](#).

Spark 3.1.1 Release Date

Version	Release Date	Status	EOM Date	EOS Date
DLI Spark 3.1.1	December 2021	Released	December 31, 2023	December 31, 2024

For more version support information, see [Lifecycle of DLI Compute Engine Versions](#).

Spark 3.1.1 Description

The following lists the main features of Spark 3.1.1.

For more new features, see [Release Notes - Spark 3.1.1](#).

- [SPARK-33050]: Upgraded Apache ORC to version 1.5.12.
- [SPARK-33092]: Improved subexpression elimination.
- [SPARK-33480]: Added support for the char/varchar data type.
- [SPARK-32302]: Optimized the pushdown of some predicates.
- [SPARK-30648]: Added support for the pushdown of predicates in JSON datasource tables.

- [SPARK-32346]: Added support for the pushdown of predicates in Avro datasource tables.
- [SPARK-32461]: Optimized the Shuffle Hash Join algorithm.
- [SPARK-32272]: Added the SQL-standard command **SET TIME ZONE**.
- [SPARK-21492]: Fixed memory leak caused by the sort-merge join algorithm.
- [SPARK-27812]: Upgraded the Kubernetes client to version 4.6.1.

 **NOTE**

DLI does not support built-in geospatial query functions since Spark 3.x.

2.6 What's New in Spark 2.4.5

DLI complies with the release consistency of the open source Spark compute engine. This document describes the updates in Spark 2.4.5.

For more information about Spark 2.4.5, see [Spark Release Notes](#).

Spark 2.4.5 Release Date

Version	Release Date	Status	EOM Date	EOS Date
DLI Spark 2.4.5	December 2021	Released	December 31, 2023	December 31, 2024

For more version support information, see [Lifecycle of DLI Compute Engine Versions](#).

Spark 2.4.5 Description

[Table 2-3](#) lists the main features of Spark 2.4.5.

For more new features, see [Release Notes - Spark 3.1.1](#).

Table 2-3 Advantages of Spark 2.4.5

Feature	Description
Merging small files	If a large number of small files are generated during SQL execution, job execution and table query will take a long time. In this case, you are advised to merge small files.
Modifying column comments of non-partitioned or partitioned tables	You can modify the column comments of non-partitioned or partitioned tables.
Collecting statistics on the CPU usage of SQL jobs	You can view the total CPU used on the console.

Feature	Description
Viewing Spark logs of container clusters	You need to view logs in the container.
Dynamic UDF loading (OBT)	The UDF takes effect without restarting the queue.
Supporting flame graphs on the Spark UI	Flame graphs can be created on the Spark UI.
Optimizing the query performance of the NOT IN statement for SQL jobs	The query performance of the NOT IN statement is improved.
Optimizing the query performance of the Multi-INSERT statement	The query performance of the Multi-INSERT statement is improved.