

数据湖探索

产品公告

文档版本 01
发布日期 2024-11-25



版权所有 © 华为技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

安全声明

漏洞处理流程

华为公司对产品漏洞管理的规定以“漏洞处理流程”为准，该流程的详细内容请参见如下网址：

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

如企业客户须获取漏洞信息，请参见如下网址：

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

目录

| | |
|--|----------|
| 1 产品公告 | 1 |
| 1.1 DLI “包年/包月”和“按需计费”计费模式队列和“队列 CU 时套餐包”下线 (EOL) 公告 | 1 |
| 1.2 DLI 经典型跨源连接下线 (EOL) 公告 | 3 |
| 1.3 DLI Flink 1.10、Flink1.11 版本停止服务 (EOS) 公告 | 4 |
| 1.4 DLI Spark 2.3.2 版本停止服务 (EOS) 公告 | 5 |
| 1.5 DLI Flink 1.7 版本停止服务 (EOS) 公告 | 7 |
| 2 版本支持公告 | 9 |
| 2.1 DLI 计算引擎版本生命周期 | 9 |
| 2.2 Flink 1.15 版本说明 | 10 |
| 2.3 Flink 1.12 版本说明 | 11 |
| 2.4 Spark 3.3.1 版本说明 | 12 |
| 2.5 Spark 3.1.1 版本说明 | 13 |
| 2.6 Spark 2.4.5 版本说明 | 14 |

1 产品公告

1.1 DLI “包年/包月”和“按需计费”计费模式队列和“队列CU时套餐包”下线（EOL）公告

公告说明

为了更好的实现资源共享，提高计算资源利用率，DLI将“包年/包月”和“按需计费”计费模式队列升级为“弹性资源池队列”。即使用DLI计算资源需先购买弹性资源池，并在弹性资源池中创建队列。

- DLI “包年/包月”和“按需计费”计费模式队列和“队列CU时套餐包”计划于2024年3月31日 00:00（北京时间）停止销售（EOM）。
- DLI “包年/包月”和“按需计费”计费模式队列和“队列CU时套餐包”计划于2025年6月30日 00:00（北京时间）下线（EOL）。

影响

- EOM后，不再支持新购队列资源，包括不支持新购“包年/包月”和“按需计费”计费模式队列和“队列CU时套餐包”。
为了满足部分业务的使用需求，在2024年6月30日 00:00:00（北京时间）前，您可以最大续订1年队列或者变更队列。
2024年6月30日 00:00:00（北京时间）之后将无法续订、变更队列。
- EOL后，队列资源将无法继续使用，请务必在下线前更换使用“弹性资源池”或“default队列”。推荐您购买弹性资源池，并在弹性资源池中创建队列，体验更丰富的DLI产品能力。

如您有任何问题，可随时通过[工单](#)或者服务热线（4000-955-988或950808）与我们联系。

DLI “包年/包月”和“按需计费”计费模式队列生命周期

表 1-1 DLI “包年/包月”和“按需计费”计费模式队列生命周期

| 功能 | 状态 | EOM时间 | EOL时间 |
|----------------------------|-----|------------|------------|
| DLI “包年/包月”和“按需计费”计费模式队列功能 | EOM | 2024年3月31日 | 2025年6月30日 |

说明

- EOM: End of Marketing, 停止该功能销售。
- EOL: End of life, 停止所有销售、服务活动。

常见问题

- **DLI “包年/包月”和“按需计费”计费模式队列停止销售（EOM）后，正在使用该功能的作业有哪些影响？**

EOM后停止队列资源的新购。

- 短期方案：为了满足部分业务的使用需求，在2024年6月30日 00:00:00（北京时间）前，您可以最大续订1年队列或者变更队列，在2024年6月30日 00:00:00（北京时间）后将无法续订、变更队列，
- 长期方案：更换使用“弹性资源池”或“default队列”，推荐您购买弹性资源池，并在弹性资源池中创建队列。

- **DLI “包年/包月”和“按需计费”计费模式队列下线（EOL）后，正在使用该功能的作业有哪些影响？**

EOL后，将无法使用“包年/包月”和“按需计费”计费模式队列执行作业。

请您务必在EOL前完成将“包年/包月”和“按需计费”计费模式队列上执行的作业迁移至弹性资源池队列。即用户重新购买弹性资源池，使用新购的弹性资源池队列执行作业。

- **DLI “包年/包月”和“按需计费”计费模式队列在EOM和EOL后，可以使用哪个功能替换？**

如果您正在使用“包年/包月”和“按需计费”计费模式队列运行作业，请尽快更换使用“弹性资源池”或“default队列”。

- 使用“按需专属”计费模式的队列可以直接迁移至新购的弹性资源池中。
- 使用“包年/包月”和“按需计费-非专属”的队列，请退订队列资源后，重新购买弹性资源池，使用新购的弹性资源池执行作业。

- **“弹性资源池队列”对比“包年/包月”和“按需计费”计费模式队列有哪些优势？**

- “包年/包月”和“按需计费”队列：固定资源规格，当作业任务对资源的请求量发生变化时，会导致队列资源浪费或者资源不足的问题。
- 弹性资源池队列：动态扩缩容，提高资源利用效率。

- **弹性资源池收费吗？弹性资源池中的队列收费吗？**

弹性资源池支持按需、包周期、套餐包的计费模式。您可以参考DLI计费说明了解更多弹性资源池计费信息。

添加到弹性资源池中的队列不再单独计费，以弹性资源池为计费项计费：

- 按需计费模式：按弹性资源池的实际CUs计费。
- 包年/包月计费模式：按弹性资源池的实际CUs计费，其中规格部分包年/包月，超出规格的部分按需计费。
- 弹性资源池CU时套餐包：按购买的套餐包周期价格扣费，其中套餐包内的规格包周期计费，超出套餐包规格的部分按需计费。

更多弹性资源池计费信息请参考[弹性资源池计费](#)。

● **怎样创建“弹性资源池队列”？**

a. **购买弹性资源池。**

- i. 在DLI管理控制台，单击“资源管理 > 弹性资源池”。
- ii. 在弹性资源池管理界面，单击界面右上角的“购买弹性资源池”。
按需配置弹性资源池的相关参数，完成弹性资源池的购买。

b. **在弹性资源池中创建队列。**

创建完弹性资源池后，弹性资源池可以添加一个或多个队列用于后续作业的运行。

- i. 在弹性资源池管理页面。
- ii. 选择要操作的弹性资源池，在“操作”列，单击“添加队列”。
- iii. 在“添加队列”界面，完成队列的基础信息配置。单击“确定”完成添加队列配置。

c. **创建作业。**

队列创建完成后，就可以使用弹性资源池中创建的队列进行作业配置。

公告发布日期：2023年11月21日

1.2 DLI 经典型跨源连接下线（EOL）公告

公告说明

华为云计划于2024年06月30 00:00（北京时间）将DLI经典型跨源连接的相关功能下线（EOL）。

影响

DLI经典型跨源连接的相关功能下线（EOL）后，不再提供该功能任何技术服务与支持。

如果您正在使用DLI经典型跨源连接的功能，请尽快更换使用[增强型跨源连接](#)，否则使用经典型跨源连接操作过程中出现的错误，不再提供该功能的任何技术服务支持。

如您有任何问题，可随时通过[工单](#)或者服务热线（4000-955-988或950808）与我们联系。

DLI 经典型跨源连接生命周期

表 1-2 DLI 经典型跨源连接生命周期

| 功能 | 状态 | EOM时间 | EOL时间 |
|------------|------------|-------------|------------|
| DLI经典型跨源连接 | 计划下线，不推荐使用 | 2023年12月30日 | 2024年6月30日 |

说明

- EOM: End of Marketing, 停止该功能销售。
- EOL: End of life, 停止所有销售、服务活动。

常见问题

- **DLI经典型跨源连接EOM和EOL后，正在使用该功能的作业有哪些影响？**
如果您正在使用DLI经典型跨源连接的功能，请尽快更换使用[增强型跨源连接](#)。
否则使用经典型跨源连接操作过程中出现的错误，不再提供该功能的任何技术服务支持。
- **DLI经典型跨源连接停止服务后，可以使用哪个功能替换？**
推荐使用[增强型跨源连接](#)。
DLI提供的增强型跨源连接功能，通过建立对等连接，实现DLI与数据源的VPC网络互通，为DLI的跨源分析提供了支持。
目前DLI支持跨源访问的数据源请参考[跨源分析开发方式](#)。
- **增强型跨源连接收费吗？**
增强型跨源连接不计费。
- **怎样创建增强型跨源连接？**
 - a. 在DLI管理控制台，单击“跨源管理 > 增强型跨源”。
 - b. 选择“增强型跨源”，单击“创建”。
配置增强型跨源连接的信息请参考创建[增强型跨源连接](#)。
创建完成后，增强型跨源连接的状态显示“已激活”，代表该连接创建成功。您还可以在队列管理页面测试数据源与DLI队列的连通性。

公告发布日期：2023年11月10日

1.3 DLI Flink 1.10、Flink1.11 版本停止服务（EOS）公告

公告说明

华为云计划于2023年12月31 00:00（北京时间）将DLI Flink 1.10、Flink1.11版本停止服务（EOS）。

影响

DLI Flink 1.10、Flink1.11版本停止服务（EOS）后，不再提供该软件版本的任何技术服务支持。建议您在执行作业时选择新版本的Flink引擎，推荐使用DLI Flink 1.15版本。

正在使用Flink 1.10、Flink1.11版本的作业也请您尽快切换至新版本的Flink引擎，否则作业执行过程中出现的错误，不再提供该版本的任何技术服务支持。

如您有任何问题，可随时通过[工单](#)或者服务热线（4000-955-988或950808）与我们联系。

常见问题

- **Flink 1.10、Flink1.11版本停止服务后，正在使用该版本的作业有哪些影响？**

在Flink 1.10、Flink1.11EOS时间后创建的对队列，在执行作业时不支持选择已经EOS的计算引擎。

历史创建的队列仍然可以使用Flink 1.10、Flink1.11执行作业，但作业执行过程中出现的错误，不再提供该版本的任何技术服务支持，请您尽快更换至新版本的计算引擎。

- **Flink 1.10、Flink1.11版本停止服务后，可以使用哪个版本替换？**

推荐使用DLI Flink 1.15版本。

- **Flink 1.12版本有哪些优势？**

Flink 1.12新增支持DataGen源表、DWS源表、JDBC源表、MySQL CDC源表、Postgres CDC源表、Redis源表、Upsert Kafka源表、Hbase源表

更多升级Flink 1.12版本的优势请参考[Flink 1.12升级指导](#)。

- **切换至新版本计算引擎对DLI资源价格是否有影响？**

DLI按作业运行所需的计算资源和存储资源计费，与计算引擎版本无关。

- **如何升级到Flink 1.15版本？**

a. 在DLI管理控制台，购买弹性资源池和队列资源，用于提供执行作业所需的计算资源。

b. 在DLI管理控制台，单击“作业管理 > Flink作业”。

c. 在作业管理页面，单击作业操作列的“编辑”，进入作业运行参数编辑页面。

d. 配置Flink版本，选择最新的Flink版本，推荐使用Flink 1.15。

Flink 1.15版本语法请参考[Flink OpenSource SQL1.15版本使用说明](#)。

公告发布日期：2023年7月6日

1.4 DLI Spark 2.3.2 版本停止服务（EOS）公告

公告说明

华为云计划于2023年12月31 00:00（北京时间）将Spark 2.3.2版本停止服务（EOS）。

影响

DLI Spark 2.3.2版本停止服务（EOS）后，不再提供该软件版本的任何技术服务支持。建议您在执行作业时选择新版本的Spark引擎，推荐使用DLI Spark 3.3.1版本。

正在使用DLI Spark 2.3.2版本的作业也请您尽快切换至新版本的Spark引擎，否则作业执行过程中出现的错误，不再提供该版本的任何技术服务支持。

如您有任何问题，可随时通过[工单](#)或者服务热线（4000-955-988或950808）与我们联系。

常见问题

- DLI Spark 2.3.2版本停止服务后，正在使用该版本的作业有哪些影响？**
在Spark 2.3.2 EOS时间后创建的队列，在执行作业时不支持选择已经EOS的计算引擎。
历史创建的队列仍然可以使用Spark 2.3.2执行作业，但作业执行过程中出现的错误，不再提供该版本的任何技术服务支持，请您尽快更换至新版本的计算引擎。
- DLI Spark 2.3.2版本停止服务后，可以使用哪个版本替换？**
推荐使用DLI Spark 3.3.1版本。
- DLI Spark 2.4.5版本有哪些优势？**

表 1-3 Spark 2.4.5 版本优势

| 特性 | 说明 |
|----------------------|---|
| 支持配置小文件合并 | 使用SQL过程中，生成的小文件过多时，会导致作业执行时间过长，且查询对应表时耗时增大，建议对小文件进行合并。 参考 如何合并小文件 完成合并小文件。 |
| 支持修改非分区表或分区表的列注释 | 修改非分区表或分区表的列注释。 |
| 支持统计SQL作业的CPU消耗 | 支持在控制台查看“CPU累计使用量”。 |
| 支持容器集群Spark日志跳转查看 | 需要在容器查看日志。 |
| 支持动态加载UDF（公测） | 无需重启队列UDF即可生效。 |
| Spark UI支持火焰图 | Spark UI支持绘制火焰图。 |
| 优化SQL作业NOT IN语句查询性能 | NOT IN语句查询性能提升。 |
| 优化Multi-INSERT语句查询性能 | Multi-INSERT语句查询性能提升。 |

- 切换至新版本对DLI资源价格是否有影响？**
DLI按作业运行所需的计算资源和存储资源计费，与计算引擎版本无关。

- **如何升级到DLI Spark 2.4.5版本?**
 - a. 在DLI管理控制台，购买弹性资源池和队列资源，用于提供执行作业所需的计算资源。
 - b. 单击“作业管理 > Spark作业”，新建Spark作业。
 - c. 在作业管理页面，单击作业操作列的“编辑”，进入作业运行参数编辑页面。
 - d. 配置Spark版本，选择最新的Spark引擎版本，推荐使用Spark 3.3.1。

公告发布日期：2023年7月6日

1.5 DLI Flink 1.7 版本停止服务（EOS）公告

公告说明

华为云于2022年12月31 00:00（北京时间）将DLI Flink 1.7版本停止服务（EOS）。

影响

DLI Flink 1.7版本停止服务（EOS）后，不再提供该软件版本相关的任何技术服务支持。建议您在执行作业时选择新版本的Flink引擎，推荐使用DLI Flink 1.15版本。

正在使用Flink 1.7版本的作业也请您尽快切换至新版本的Flink引擎，否则作业执行过程中出现的错误，不再提供该版本的任何技术服务支持。

如您有任何问题，可随时通过[工单](#)或者服务热线（4000-955-988或950808）与我们联系。

常见问题

- **Flink 1.7版本中有哪些重要功能不再演进?**
 - Flink Edge SQL功能随Flink1.7 EOS，后续版本的Flink均不再支持边缘场景的作业处理。
 - 敏感变量功能随Flink1.7 EOS，后续版本不再支持敏感变量。
- **Flink 1.7版本停止服务后，正在使用该版本的作业有哪些影响?**

在Flink 1.7 EOS后创建的队列，在执行作业时不支持选择已经EOS的计算引擎。
历史创建的队列使用Flink 1.7执行作业过程中出现的错误，不再提供该版本的任何技术服务支持，请您尽快更换至新版本的计算引擎。
- **Flink 1.7版本停止服务后，可以使用哪个版本替换?**

推荐使用DLI Flink 1.15版本。
- **Flink 1.12版本有哪些优势?**

Flink 1.12新增支持DataGen源表、DWS源表、JDBC源表、MySQL CDC源表、Postgres CDC源表、Redis源表、Upsert Kafka源表、Hbase源表
更多升级Flink 1.12版本的优势请参考[Flink 1.12升级指导](#)。
- **切换至新版本计算引擎对DLI资源价格是否有影响?**

DLI按作业运行所需的计算资源和存储资源计费，与计算引擎版本无关。
- **如何升级到Flink 1.12版本?**

- a. 在DLI管理控制台，购买弹性资源池和队列资源，用于提供执行作业所需的计算资源。
- b. 在DLI管理控制台，单击“作业管理 > Flink作业”。
- c. 在作业管理页面，单击作业操作列的“编辑”，进入作业运行参数编辑页面。
- d. 配置Flink版本，选择最新的Flink版本，推荐使用Flink 1.15。

公告发布日期：2023年7月6日

2 版本支持公告

2.1 DLI 计算引擎版本生命周期

版本号说明

DLI计算引擎版本号：格式为**计算引擎名称 x.y.z**，其中计算引擎分为Flink和Spark，版本号具体含义如**图2-1**所示。

图 2-1 DLI 计算引擎版本号



版本支持情况

- Flink计算引擎推荐版本：Flink 1.15。
- Spark计算引擎推荐版本：Spark 3.3.1。

📖 说明

不建议长期混用不同版本的Spark/Flink引擎。

- 长期混用不同版本的Spark/Flink引擎会导致代码在新旧版本之间不兼容，影响作业的执行效率。
- 当作业依赖于特定版本的库或组件，长期混用不同版本的Spark/Flink引擎可能会导致作业因依赖冲突而执行失败。

计算引擎版本生命周期

表2-1给出了DLI计算引擎版本生命周期，帮助您规划自己的版本更新节奏。

表 2-1 DLI 计算引擎版本生命周期

| 计算引擎类型 | 版本名称 | 状态 | EOM时间 | EOS时间 |
|--------|-----------------|-----|-------------|-------------|
| Flink | DLI Flink 1.15 | 已发布 | 2025年6月30日 | 2026年6月30日 |
| | DLI Flink 1.12 | EOM | 2023年12月31日 | 2024年12月31日 |
| | DLI Flink 1.11 | EOS | 2022年6月30日 | 2023年12月31日 |
| | DLI Flink 1.10 | EOS | 2022年6月30日 | 2023年12月31日 |
| | DLI Flink 1.7 | EOS | 2021年12月31日 | 2022年12月31日 |
| Spark | DLI Spark 3.3.1 | 已发布 | 2025年6月30日 | 2026年6月30日 |
| | DLI Spark 3.1.1 | EOM | 2023年12月31日 | 2024年12月31日 |
| | DLI Spark 2.4.5 | EOM | 2023年12月31日 | 2024年12月31日 |
| | DLI Spark 2.3.2 | EOS | 2022年6月30日 | 2023年12月31日 |

📖 说明

- EOM: End of Marketing, 停止该版本的销售。所有新购资源不再支持选择EOM的版本的引擎。
- EOS: End of Service & support, 停止该版本的服务, 建议您在执行作业时选择最新版本的引擎。在该日期之后, 不再提供该软件版本的任何技术服务支持。

2.2 Flink 1.15 版本说明

数据湖探索 (DLI) 遵循开源Flink计算引擎的发布一致性。本文介绍Flink 1.15版本所做的变更说明。

更多Flink 1.15版本说明请参考[Release Notes - Flink Jar 1.15](#)、[Flink OpenSource SQL1.15版本使用说明](#)。

Flink 1.15 版本发布时间

| 版本名称 | 发布时间 | 状态 | EOM时间 | EOS时间 |
|----------------|---------|-----|------------|------------|
| DLI Flink 1.15 | 2023年6月 | 已发布 | 2025年6月30日 | 2026年6月30日 |

更多版本支持信息请参考[DLI计算引擎版本生命周期](#)。

Flink 1.15 版本说明

- Flink 1.15版本在语法设计上实现了更高的兼容性，与主流开源技术标准保持一致。
- Flink 1.15版本新增读写Hive、Hudi等Connector。
- Flink 1.15版本数据同步迁移场景，优先推荐使用DataArts的数据集成。
- Flink 1.15版本支持集成DEW-CSMS凭证管理，提供隐私保护方案。
- Flink 1.15版本支持Flink Jar作业最小化提交。

📖 说明

Flink Jar作业最小化提交是指Flink仅提交作业必须的依赖项，而不是整个Flink环境。通过将非Connector的Flink依赖项（以flink-开头）和第三方库（如Hadoop、Hive、Hudi、Mysql-cdc）的作用域设置为provided，可以确保这些依赖项不会被包含在Jar作业中，从而实现最小化提交，避免依赖包与flink内核中依赖包冲突：

- 仅Flink 1.15版本支持Flink Jar作业最小化提交，通过在运行优化参数中配置 `flink.dli.job.jar.minimize-submission.enabled=true` 可以开启最小化提交。
- Flink相关依赖作用域请使用provided，即在依赖中添加 `<scope>provided</scope>`。主要包含org.apache.flink组下以flink-开头的非Connector依赖。
- Hadoop、Hive、Hudi、Mysql-cdc相关依赖，作用域请使用provided，即在依赖中添加 `<scope>provided</scope>`。
- Flink源代码中只有明确标注了@Public或者@PublicEvolving的才是公开供用户调用的方法，DLI只对这些方法的兼容性做出产品保证。

2.3 Flink 1.12 版本说明

数据湖探索（DLI）遵循开源Flink计算引擎的发布一致性。本文介绍Flink 1.12版本所做的变更说明。

更多Flink 1.12版本说明请参考[Release Notes - Flink 1.12](#)。

Flink 1.12 版本发布时间

| 版本名称 | 发布时间 | 状态 | EOM时间 | EOS时间 |
|----------------|----------|-----|-------------|-------------|
| DLI Flink 1.12 | 2021年12月 | 已发布 | 2023年12月31日 | 2024年12月31日 |

更多版本支持信息请参考[DLI计算引擎版本生命周期](#)。

Flink 1.12 版本说明

- Flink 1.12新增支持DataGen源表、DWS源表、JDBC源表、MySQL CDC源表、Postgres CDC源表、Redis源表、Upsert Kafka源表、Hbase源表。
- Flink 1.12新增支持小文件合并功能。
- Flink 1.12新增支持Redis维表、RDS维表。

2.4 Spark 3.3.1 版本说明

数据湖探索（DLI）遵循开源Spark计算引擎的发布一致性。本文介绍Spark 3.3.1版本所做的变更说明。

更多Spark 3.3.1版本说明请参考[Spark Release Notes](#)。

Spark 3.3.1 版本发布时间

| 版本名称 | 发布时间 | 状态 | EOM时间 | EOS时间 |
|-----------------|---------|-----|------------|------------|
| DLI Spark 3.3.1 | 2023年6月 | 已发布 | 2025年6月30日 | 2026年6月30日 |

更多版本支持信息请参考[DLI计算引擎版本生命周期](#)。

Spark 3.3.1 版本说明

[表2-2](#)列举了Spark 3.3.1 版本主要的功能特性。

更多版本新特性及性能优化请参考[Release Notes - Spark 3.3.1](#)。

表 2-2 Spark 3.3.1 版本优势

| 特性 | 说明 |
|------------------------|---|
| Native性能加速 | Spark查询语句性能提升。 |
| 元数据访问性能提升 | 提升Spark在处理大数据时的元数据访问性能，提高数据处理流程效率。 |
| 提升OBS committer小文件写性能 | 提升对象存储服务（OBS）在处理小文件写入时的性能，提高数据传输效率。 |
| 动态Executor shuffle数据优化 | 提升资源扩缩容的稳定性，当shuffle文件不需要时清理Executor。 |
| 支持配置小文件合并 | 使用SQL过程中，生成的小文件过多时，会导致作业执行时间过长，且查询对应表时耗时增大，建议对小文件进行合并。 参考 如何合并小文件 完成合并小文件。 |
| 支持修改非分区表或分区表的列注释 | 修改非分区表或分区表的列注释。 |
| 支持统计SQL作业的CPU消耗 | 支持在控制台查看“CPU累计使用量”。 |
| 支持容器集群Spark日志跳转查看 | 需要在容器查看日志。 |

| 特性 | 说明 |
|----------------------|-----------------------|
| 支持动态加载UDF（公测） | 无需重启队列UDF即可生效。 |
| Spark UI支持火焰图 | Spark UI支持绘制火焰图。 |
| 优化SQL作业NOT IN语句查询性能 | NOT IN语句查询性能提升。 |
| 优化Multi-INSERT语句查询性能 | Multi-INSERT语句查询性能提升。 |

2.5 Spark 3.1.1 版本说明

数据湖探索（DLI）遵循开源Spark计算引擎的发布一致性。本文介绍Spark 3.1.1版本所做的变更说明。

更多Spark 3.1.1版本说明请参考[Spark Release Notes](#)。

Spark 3.1.1 版本发布时间

| 版本名称 | 发布时间 | 状态 | EOM时间 | EOS时间 |
|-----------------|----------|-----|-------------|-------------|
| DLI Spark 3.1.1 | 2021年12月 | 已发布 | 2023年12月31日 | 2024年12月31日 |

更多版本支持信息请参考[DLI计算引擎版本生命周期](#)。

Spark 3.1.1 版本说明

下表列举了Spark 3.1.1 版本主要的功能特性。

更多版本新特性请参考[Release Notes - Spark 3.1.1](#)。

- 【SPARK-33050】：Apache ORC升级到1.5.12。
- 【SPARK-33092】：增强子表达式消减。
- 【SPARK-33480】：支持char/varchar数据类型。
- 【SPARK-32302】：部分谓词下推优化。
- 【SPARK-30648】：支持JSON datasource表谓词下推。
- 【SPARK-32346】：支持avro datasource表谓词下推。
- 【SPARK-32461】：shuffle hash join优化。
- 【SPARK-32272】：添加SQL标准命令SET TIME_ZONE。
- 【SPARK-21492】：修复排序合并加入中的内存泄漏。
- 【SPARK-27812】：K8S客户端版本提升到4.6.1。

 说明

DLI从Spark 3.x版本开始不支持内置地理空间查询函数。

2.6 Spark 2.4.5 版本说明

数据湖探索（DLI）遵循开源Spark计算引擎的发布一致性。本文介绍Spark 2.4.5版本所做的变更说明。

更多Spark 2.4.5版本说明请参考[Spark Release Notes](#)。

Spark 2.4.5 版本发布时间

| 版本名称 | 发布时间 | 状态 | EOM时间 | EOS时间 |
|-----------------|----------|-----|-------------|-------------|
| DLI Spark 2.4.5 | 2021年12月 | 已发布 | 2023年12月31日 | 2024年12月31日 |

更多版本支持信息请参考[DLI计算引擎版本生命周期](#)。

Spark 2.4.5 版本说明

[表2-3](#)列举了Spark 2.4.5 版本主要的功能特性。

更多版本新特性请参考[Release Notes - Spark 2.4.5](#)。

表 2-3 Spark 2.4.5 版本优势

| 特性 | 说明 |
|---------------------|---|
| 支持配置小文件合并 | 使用SQL过程中，生成的小文件过多时，会导致作业执行时间过长，且查询对应表时耗时增大，建议对小文件进行合并。 参考 如何合并小文件 完成合并小文件。 |
| 支持修改非分区表或分区表的列注释 | 修改非分区表或分区表的列注释。 |
| 支持统计SQL作业的CPU消耗 | 支持在控制台查看“CPU累计使用量”。 |
| 支持容器集群Spark日志跳转查看 | 需要在容器查看日志。 |
| 支持动态加载UDF（公测） | 无需重启队列UDF即可生效。 |
| Spark UI支持火焰图 | Spark UI支持绘制火焰图。 |
| 优化SQL作业NOT IN语句查询性能 | NOT IN语句查询性能提升。 |

| 特性 | 说明 |
|----------------------|-----------------------|
| 优化Multi-INSERT语句查询性能 | Multi-INSERT语句查询性能提升。 |