

# DLA迁移ADB企业版指南

## DLA停服下线

云原生数据湖分析DLA已于**2023年06月1日**起停止售卖，且功能不再继续演进，计划于**2025年10月31日**正式停服。

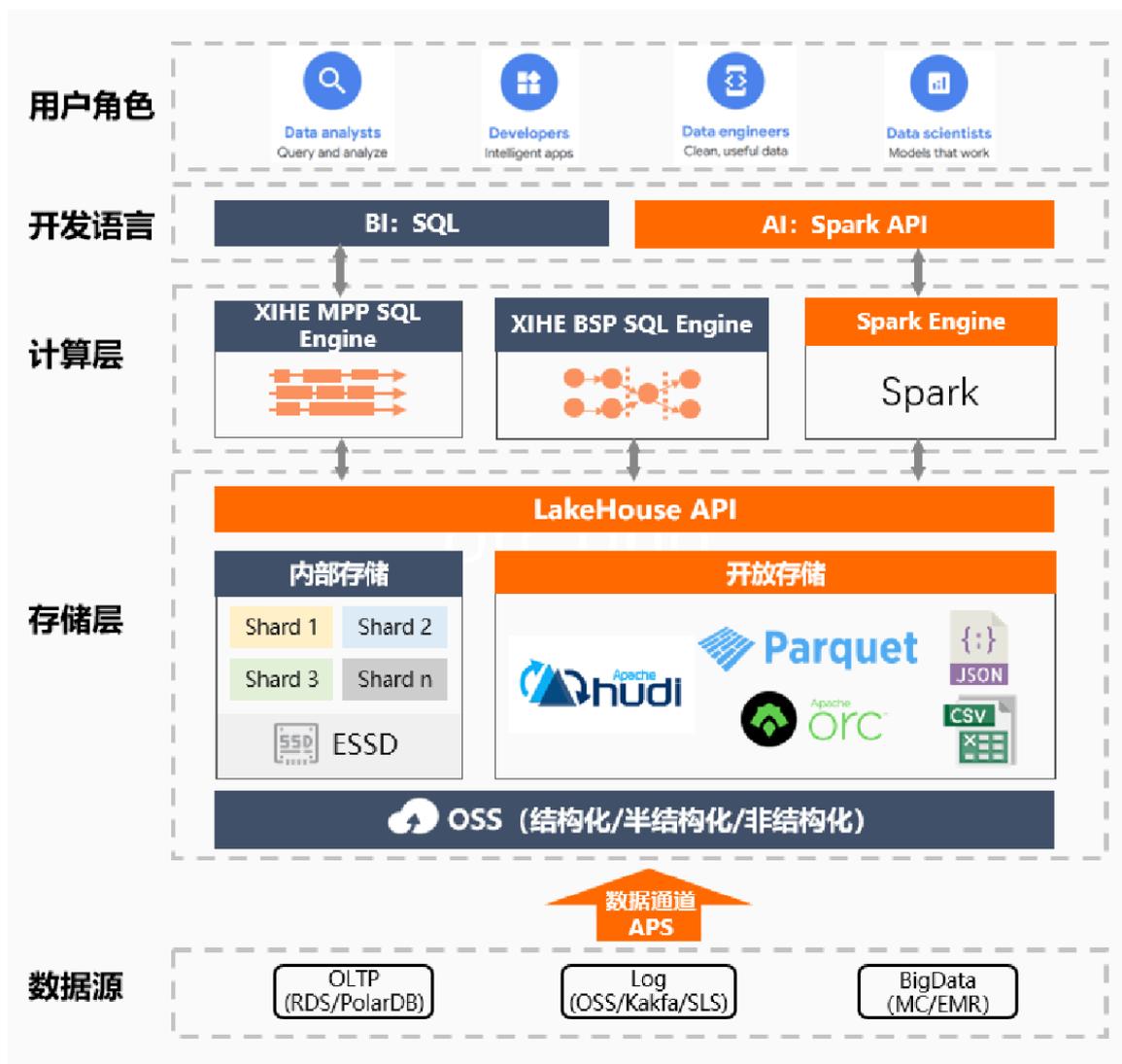
云原生数据仓库 AnalyticDB MySQL 版（以下简称ADB）企业版已正式商业化，且企业版整合了**ADB实时数仓与DLA**的核心功能。因此建议将DLA迁移至ADB企业版，以获得更优越的性能与功能体验。

**注意：**企业版其功能和价格与湖仓版完全一致，属于湖仓版的演进形态，创建ADB实例时，产品系列选择企业版即可。关于ADB企业版的更多信息，请参见 [产品系列](#)。

## ADB企业版优势

云原生数据仓库AnalyticDB MySQL版是一种支持高并发低延时查询的新一代云原生数据仓库，高度兼容MySQL协议以及SQL:92、SQL:99、SQL:2003标准，可以对海量数据进行即时的多维分析透视和业务探索,快速构建企业云上数据仓库。

新推出的AnalyticDB MySQL企业版可以帮助企业快速构建从湖到仓一体化技术架构，使用低成本离线处理能力完成数据的清洗加工，使用高性能在线分析能力完成数据的洞察探索。企业版架构图如下。



## ADB企业版与DLA功能兼容性

### ADB企业版与DLA兼容的功能

- 兼容MySQL语法和协议，可以通过JDBC或控制台同步或异步提交SQL。
- 支持查询多种类型外表，包括OSS(Hive/HUDI)、MySQL、ODPS、OTS、MongoDB、PgSQL, SQLServer、ADB等外部存储。
- 通过入湖实现SLS数据同步到HUDI外表。
- 集群支持在线扩缩容、分时弹性等功能。

### ADB企业版其他功能

- 支持高吞吐的数据实时增删改、低延时的实时分析和复杂ETL。
  - DLA只适合批量数据的写入、更新和删除，一般情况下，QPS较低，Query延迟为秒级或者更高。

- ADB支持高吞吐的数据实时写入、更新和删除，支持低延迟的点查、实时分析，也支持复杂的ETL。
- 支持Job型资源组，Job型资源组支持以更少的资源消耗运行数据量更大的SQL。

## ADB企业版与DLA对比

以TPC-H测试集为例，DLA集群和ADB企业版实例均为32核：

- DLA：该查询会因内存不足而失败。
- ADB企业版：该查询以交互式方式执行，会因查询内存不足而失败。若采用批处理方式（Batch）执行，即使ADB企业版为16核，也可以执行成功。更适用于对延迟不敏感的ETL类型任务或希望使用较少资源执行数据量更大的查询的场景。

ADB企业版在JOIN、AGG等较重要的SQL实现上做了较多的优化，例如向量化、hashTable优化等，SQL的整体性能有较大的提升。以TPC-H测试数据集（100 GB数据）为例，对比了相同计算资源下的ADB企业版和DLA的性能，ADB企业版性能提升20%左右。

更多ADB企业版功能介绍，请参见 [ADB产品文档](#)。

## DLA迁移方案

根据DLA中客户的使用情况不同,我们推出了多种迁移方案。如下表所示：

迁移方案	适用用户	费用预估（以杭州地域为例）	迁移方案	推荐指数
迁移DLA至ADB企业版预留资源池	希望SQL的延迟低，或希望继续使用类似DLA的服务，实现平滑迁移。	ADB企业版预留资源为24ACU（预留资源规格为8ACU，节点数为3），每月费用为3360元。 详情：ADB企业版预留资源费用单价为140元/ACU/月，费用=24 ACU*140元/ACU/月=3360元。	<a href="#">通过外表导入OSS数据</a>	五星级

迁移DLA至ADB企业版Job资源组	DLA扫描量较少，对SQL延迟要求不高。	每天两个小时执行SQL，且SQL平均消耗32 ACU，每月费用为576元。 详情：ADB企业版Job资源费用单价为0.3元/ACU/小时，费用=2小时*0.3元/ACU/小时*32 ACU*30天=576元。	<a href="#">通过外表导入OSS数据</a>	四颗星
迁移DLA至其他产品	希望使用其他开源大数据产品。	需自行评估。	需自行迁移。	三颗星

注意：具体费用以ADB控制台为准。

## 迁移最佳实践

ADB企业版由ADB数仓版和湖仓版演进而来，其功能与DLA有些许不同，内容如下所示：

- 写入OTS外表时，DLA使用hint `/+ ots-insert-as-update=true/` 把INSERT语义改为UPDATE语义。在ADB企业版中，可以直接使用REPLACE INTO语法实现。
- 若在DLA使用一键建仓，推荐根据上游数据源选择合适的同步方式，将数据写入到ADB湖存储或者仓存储。相关文档，请参见[数据导入](#)。
- 按扫描量付费不再支持。迁移到ADB企业版可以使用预留资源池，性能更优，但价格更高；如果对费用比较敏感，建议使用Job资源池，资源按需弹性，费用相对DLA扫描量略高，相对ADB企业版预留资源费用更低。

## ADB与DLA的SQL行为差异

- 更严格的类型转换检查。

例如，表的列数据类型为INT，查询中希望以BIGINT方式写入该列，默认不支持转换，会导致报错，如果您希望转换，可以在SQL增加CAST函数显式转换数据类型。

- ADB企业版中date为关键字，如果使用date作为列名,需要增加分号 (;) ，即 `date;` 。

- ADB企业版中，now函数和current\_date函数使用需要加上括号()。即 `now()` 、 `current_date()` 。
- Map类型输出结果格式不同，ADB企业版中使用更加通用的JSON格式显示MAP类型数据。
- 若SQL中同时包含GROUP BY和ORDER BY，DLA中该SQL可以执行成功，但是在ADB企业版中ORDER BY的列必须包含在聚合函数或GROUP BY列中。否则会出现类似下述报错：

```

    复制代码
    ``t`.`name`' must be an aggregate expression or appear in GROUP BY clause
  
```

- ADB企业版中，两张表在做Join操作时，如果有相同名字的列,需要为列名增加表名前缀。例如：

```

SELECT r.*, c.info FROM tmp_table1 AS r LEFT JOIN tmp_table2 AS c ON r.pid=
c.pid ORDER BY(这里需要指定表名)r.time DESC LIMIT 1000;
  
```

```

    复制代码
    WITH tmp_table1 AS (
      SELECT pid, time
      FROM table1
    ),
    tmp_table2 AS (
      SELECT pid, time, info
      FROM table2
    )
  
```

- DLA支持使用 `show query_task where..` 语法查询提交的SQL信息。ADB企业版支持 `SHOW [FULL] PROCESSLIST` 语法查询正在执行的任务，且在ADB企业版控制台的SQL诊断页面可以查看SQL的执行历史详情。
- 查询Parquet格式文件时，如果Parquet文件中数据类型为INT64，但建表语句指定的对应列的类型为INT，DLA不会报错，但此时查询结果可能不准确；ADB企业版则会提示如下错误：

```

    复制代码
    The column new_author_cnt is declared as type int, but the Parquet file
    declares the column as type INT64
  
```

- ADB企业版和DLA对于 `SELECT function(a) as a FROM t GROUP BY a` 这一类查询，SQL解析语义不同：
  - 在ADB企业版：SQL解析语义为 `SELECT function(a) as a FROM t GROUP BY a` 。
  - 在DLA：SQL解析语义为 `SELECT function(a) as a FROM t GROUP BY function(a)` 。

- SQL语义解析不同，将导致GROUP BY列不同，此时可以在ADB企业版中将SQL改写为 `SELECT function(a) as a_0 FROM TABLE t GROUP BY a_0` (t表中不包含a\_0列) 或者 `SELECT function(a) as a FROM TABLE t GROUP BY 1`。
- ADB企业版和DLA对于 `SELECT COUNT(DISTINCT a, b) FROM t` 这一类查询，SQL解析语义不同：
  - 在ADB企业版：COUNT DISTINCT不会计算带空值 (NULL) 的行。
  - 在DLA：会计算带空值 (NULL) 的列。
  - SQL语义解析不同，此时可以在ADB企业版中将SQL改写为 `SELECT COUNT(*) FROM (SELECT DISTINCT a, b FROM t)`。
- ADB企业版和DLA对于 `SELECT* FROM a left join b ORDER BY dt` (查询涉及多个表，且表中有相同命名的属性列) 这一类查询，SQL解析语义不同，在ADB企业版中会报错 `d t is ambiguous`，此时需要在ORDER BY子句中明确指定使用的是哪个表的列，例如 `tb l. dt`。
- ADB企业版和DLA对于 `SELECT IF(...) as"礼物名" FROM t GROUP BY"礼物名"` 这一类查询，SQL解析语义不同。在ADB企业版中会报错 `CASE WHEN THEN ELSE must be an aggregate expression or appear in GROUP BY clause`，此时需要将GROUP BY列中的 `"礼物名"` 改为 `礼物名/礼物名`，即去掉双引号/改成转译符。
- ADB中 `=null` 的结果与 `is null` 的结果不同,但DLA中 `=null` 的结果与 `is null` 的结果相同，因此在ADB企业版中需要把`"=null"`语法改为 `is null`。

## 结果类型差异

1、date\_format()函数返回结果的数据类型不同，DLA中为INT,ADB企业版中为DOUBLE。

例如：date\_format('20230108',1%u)-35在DLA的返回结果是-34，在ADB的返回结果是-34.0。建议在ADB企业版中修改SQL为 `cast(date_format('20230108', '%u') as int)`。

2、ceiling/ceil(1/4)函数返回结果的数据类型不同。在DLA的ceiling/ceil函数返回值都是整数，在ADB的ceiling/ceil函数在输入值类型为DOUBLE或FLOAT时,返回值类型为DOUBLE。

例如：ceiling/ceil(1/4)在DLA的结果是1，在ADB的结果是1.0。建议在ADB企业版中修改SQL为 `cast(ceil(1/4) as int)`。

3、date\_diff(datetime1, datetime2)函数,, 如果datetime1, datetime2数据类型是Date类型，DLA的结果是datetime2- datetime1，而ADB的结果是datetime1- datetime2，如果有使用到，需要调换datetime1,datetime2顺序。