导数工具 5.0.0 Release Note

导数工具 5.0.0 Release Note
- 版本信息
- 发版目的
- 性能提升
  - 整体提升概览
  - INSERT：普通导入吞吐最高达到原来的 3.75 倍
  - Load：20GB 旁路导入吞吐达到原来的 1.47 倍
  - Dump：CSV 导出吞吐最高达到原来的 2.18 倍
  - 小规格场景：解决内存限流瓶颈，吞吐最高达到原来的 38.73 倍
  - OB 侧累计 CPU 消耗
  - 核心升级：自适应增减 Reader/Writer，动态调整读写比例
  - 文件解析能力提升
- 新增特性
- 缺陷修复
  - 文件与脚本处理
  - 数据类型与对象兼容性
  - 异常处理与运行诊断
- 漏洞修复
- 依赖库升级

导数工具 5.0.0 Release Note

版本信息

项目	描述
发布日期	2026 年 11 月 3 日
版本号	5.0.0
下载地址	OceanBase 软件下载中心

发版目的

本次发版重点提升导入导出的开箱性能。相较 4.3.7，**已测常规场景吞吐最高达到原来的 3.75 倍；在 1G 堆内存的小规格场景下，针对内存资源不足、数据积压引发的持续限流瓶颈进行优化，吞吐可达到原**

来的 38.73 倍。通过引入自适应 Reader/Writer 调度、优化多格式文件解析与批量写入、复用导出缓冲，减少人工调参，提高 CPU 和内存使用效率，同时完善数据类型兼容性及异常处理能力。

性能提升

整体提升概览

分组	提升项	提升表现	测试条件
INSERT	普通导入吞吐	**最高达到原来的 3.75 倍**	ClickBench 100 万行，4G 堆内存
INSERT	小规格导入吞吐	**最高达到原来的 38.73 倍**	ClickBench 400 万行，1G 堆内存，解决持续内存限流瓶颈
INSERT	OB 侧累计 CPU 消耗	**减少 27.16%**	ClickBench 100 万行，4G 堆内存，完成同等导入任务
Load	旁路导入吞吐	**达到原来的 1.47 倍**	ClickBench 20GB，4G 堆内存
Load	OB 侧累计 CPU 消耗	**减少 7.84%**	ClickBench 20GB，4G 堆内存，完成同等导入任务
Dump	CSV 导出吞吐	**最高达到原来的 2.18 倍**	约 1GB 数据，4G 堆内存

INSERT：普通导入吞吐最高达到原来的 3.75 倍

优化读写调度、批次默认策略及批内绑定计划复用，减少读取等待、重复构造和写入开销，提高默认配置下的普通导入性能。

测试条件	4.3.7 耗时	5.0.0 耗时	加速比
100 万行，4G 堆内存	73.13 秒	**19.48 秒**	**3.75 倍**

Load：20GB 旁路导入吞吐达到原来的 1.47 倍

优化旁路导入的组批与在途数据预算，在大文件导入场景下降低数据积压和客户端内存占用。

测试条件	4.3.7 耗时	5.0.0 耗时	加速比

20GB，4G 堆内存	493.73 秒	**336.89 秒**	**1.47 倍**

资源指标	4.3.7	5.0.0	变化
客户端峰值 RSS	20.21 GiB	**3.47 GiB**	**约为原来的六分之一（17.17%）**

RSS 包含文件映射驻留页，不能等同于 Java 堆内存；累计 CPU 消耗下降也不代表导入期间的瞬时 CPU 负载降低。

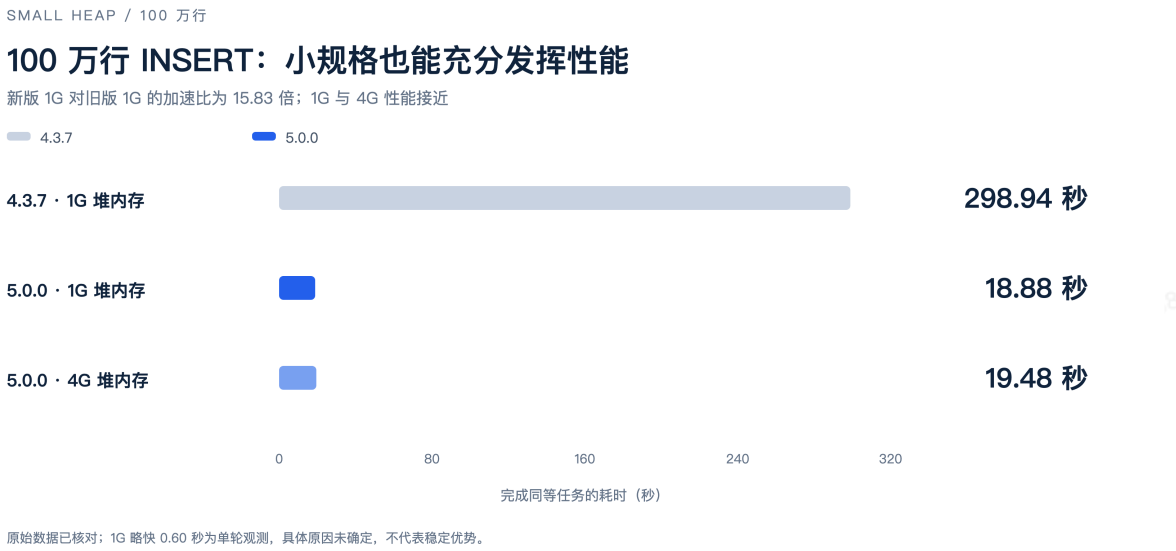
Dump：CSV 导出吞吐最高达到原来的 2.18 倍

通过复用行缓冲、减少初始化和对象分配开销，提高 CSV 等格式的导出效率。

源 CSV 规模	4.3.7 耗时	5.0.0 耗时	加速比
约 1GB（实测 0.99GB）	64.07 秒	**29.41 秒**	**2.18 倍**

小规格场景：解决内存限流瓶颈，吞吐最高达到原来的 38.73 倍

针对以往小规格环境下内存资源不足、在途数据积压导致频繁触发内存保护限流的问题，5.0.0 重构读写调度与任务级内存预算：按可用内存限制在途数据，在压力升高时收紧读取供给并优先排空已有批次，避免持续积压与限流反复叠加。在已测 1G 堆内存场景中，解决了旧版长期受内存限流制约的吞吐瓶颈，使小规格环境也能发挥并发写入能力。



1G 与 4G 相比，性能几乎无回退。 在同版本 400 万行测试中，1G 配置只使用 4G 配置**四分之一的堆内存，仍保持 97.46% 的吞吐**，大幅降低对堆内存大小的依赖。

OB 侧累计 CPU 消耗

除客户端处理能力提升外，完成同等导入任务时，OB 侧累计 CPU 消耗也更少：100 万行 INSERT **减少 27.16%**，20GB 旁路导入**减少 7.84%**。

测试条件	4.3.7 累计 CPU 消耗	5.0.0 累计 CPU 消耗	消耗减少
INSERT，100 万行，4G 堆内存	228.79 CPU 秒	**166.65 CPU 秒**	**27.16%**
Load 旁路导入，20GB，4G 堆内存	2471.13 CPU 秒	**2277.48 CPU 秒**	**7.84%**

核心升级：自适应增减 Reader/Writer，动态调整读写比例

5.0.0 默认启用自适应 Reader/Writer 调度和任务级在途数据预算。Reader、Writer 的数量可以随运行状态自动增加或减少，读写比例也随之动态调整，让文件读取供给与数据库写入能力相匹配。

- **动态调整读取与写入并发**：结合读取供给、写入吞吐、数据库响应、CPU 使用和堆内存压力，分别调整 Reader 与 Writer 数量，并重新平衡读写比例，减少读取不足或消费不及时造成的等待与积压。
- **内存压力下优先排空**：通过任务级预算控制在途批次占用，内存压力升高时收紧读取供给、按压力调整并发，优先完成已生成批次的写入，减少积压与内存保护限流反复叠加。
- **降低调参门槛**：常规 INSERT 场景无需反复尝试线程数、固定读写配比、缓冲容量和批次大小，让不同资源规格下的任务自动适配当前负载。

文件解析能力提升

优化多格式文件解析与读取路径，减少解析和对象分配开销，并增强单文件并行读取能力。

单线程解析处理能力提升 71.33%：CUT 单线程专项中，解析吞吐由 64.56 MiB/s 提高到 110.61 MiB/s。此处使用单线程解析实测结果，测试基线与下表一致。

格式	并发条件	解析吞吐提升
CSV	16 线程 → 16 线程	**155.77%**
CUT	单线程 → 单线程	**71.33%**
CUT	32 线程 → 32 线程	**56.76%**
SQL	16 线程 → 16 线程	**74.02%**
POS	16 路 → 16 路	**56.88%**
Avro	实际读取并发 1 路 → 16 路	**最高达到原来的 15.43 倍**
Parquet	实际读取并发 1 路 → 16 路	**最高达到原来的 13.40 倍**

新增特性

- CSV、CUT、POS 导入支持通过 `--skip-header N` 、 `--skip-footer N` 跳过每个原始文件首尾指定数量的记录。
- 增强导出并行能力，默认采用 SAMPLE 分片策略，并支持通过 `--split-type LEGACY` 使用原有策略。
- 增强 `secure-gen` 多账号凭据管理，支持在同一加密凭据文件中保存多组业务和 sys 账号，按用户、租户、集群及地址匹配凭据，简化多环境免密配置。
- 完善 OceanBase Oracle 模式对象定义导出，支持视图及视图列注释，并完善分区唯一索引信息保留。

缺陷修复

文件与脚本处理

- 修复 CSV 显式列映射、MySQL 列名匹配及控制文件查找问题，避免导入字段错位；修复首尾记录过滤场景下的分片性能回退。
- 修复 CUT 分片开头合法 U+FEFF 字符丢失的问题，以及 POS 短读、末行无换行、记录长度和多字节位置边界问题；非法编码进入坏记录处理。
- 修复 SQL 文件列清单变化、显式与隐式列混用及坏批次处理问题，避免影响后续合法记录。
- 完善 DDL/MIX 在注释、DELIMITER、Oracle 块及 q-quote 场景下的语句边界和 JDBC escape 改写处理，保持 MIX 文件内的会话状态和事务顺序。

数据类型与对象兼容性

- 修复 ORC 原生时间戳导入 OceanBase Oracle 时的 NLS 转换错误及 DECIMAL 精度保留问题；完善 Parquet/Avro 的时间、Decimal、NULL 和嵌套类型处理，修复 Avro union 合法数据被整批拒绝的问题。
- 修复 MySQL TIME 导出丢失微秒或不能保留超过 24 小时取值，以及参数绑定失败被计入成功行数的问题。
- 修复 Remap 导入时索引、约束名称冲突及特殊标识符改写错位；修复 OceanBase 4.x Oracle 普通 JDBC 导入父子表时的外键依赖顺序。
- 修复显式主键表仍导出隐藏主键的问题；默认关闭 LOB Locator V2，改善 LOB 导入导出兼容性。

异常处理与运行诊断

- 修复导出在磁盘满、文件滚动及收尾失败后任务卡住或错误标记完成的问题，完善并行单文件合并和 CSV 表头初始化的异常传递。
- 修复本地路径包含 `#` 、 `?` 、 `%` 等特殊字符时被错误解析，以及输出目录检查和报错不准确的问题。
- 恢复定期进度日志，修正吞吐、缓冲容量及等待时间的展示口径；修复连接初始化 SQL 失败后仍返回连接的问题。
- 修复 SSL 连接配置 `trustServerCertificate=true` 后仍强制要求 trustStore 的校验问题。

漏洞修复

- 升级 Guava，同步修复 Hadoop 重定位命名空间中的相关反序列化风险。
- 更新 Log4j、Netty 及 Parquet 内嵌 Jackson 等依赖，修复相关安全风险。
- 移除旧 SOFA Common 和 Commons Lang 2 运行时依赖。

依赖库升级

组件	4.3.7	5.0.0
Guava	32.0.1-jre	33.7.2-jre
Log4j	2.18.0	2.25.5
Netty	4.1.136.Final	4.1.137.Final
AWS SDK v2	2.35.4	2.54.14
obkv-table-client	2.4.0	2.4.1
parquet-jackson	1.15.1	1.18.1

统一 SQL/DDL 解析入口，使用 public-parser 2.0.10；Parquet 主体保持 1.15.1，以兼顾 Java 8 兼容性。