



GreatDB
万里数据库

技术白皮书

万里数据库评估迁移工具软件

20
25

稳定 · 性能 · 易用

北京万里开源软件有限公司

Beijing Great OpenSource Software Co., Ltd.



版权所有 北京万里开源软件有限公司 保留所有权利

目录

1. 产品介绍	2
2. 产品架构	2
2.1 产品架构说明	2
2.2 部署架构说明	3
3. 产品优势	3
3.1 快速部署	4
3.2 简单易用	4
3.3 安全高可靠	4
3.4 多种场景支持	5
4. 功能特性	5
4.1 迁移评估	7
4.2 数据迁移	7
4.3 对象迁移	7
4.4 断点续传	8
4.5 数据和结构校验	8
5. 性能	8
5.1 测试环境	8
5.2 测试结果	9
6. 应用场景	9
6.1 不停机迁移数据库	9
6.2 数据异地灾备	11
6.3 异地多活	11
6.4 横向扩展能力	12

1. 产品介绍

GreatDTS 是集评估、迁移、同步和校验功能于一体的数据迁移工具，GreatDTS 部署便捷、易于使用，提供高效的同构和异构数据复制服务。

GreatDTS 支持对象迁移评估、应用改造评估、全量数据迁移、增量数据同步和数据校验等功能，具备应用深度分析、语法树分析、断点续传、多线程多任务并行模式、事务级同步和全量数据一致性校验等特性。提供 GreatDB、Oracle、MySQL、Kafka、ClickHouse 等多种数据源之间的数据迁移和同步。

2. 产品架构

2.1 产品架构说明

GreatDTS 提供了一套分布式、易扩展、高可靠的产品架构，产品功能架构图如下。



整个产品功能架构分为三部分：

- 可视化控制台：用户交互层，可视化控制台提供数据管理的任务创建、管理、运维、安全操作审计等。
- 功能引擎层：负责具体的管理工作。包含评估引擎、迁移引擎、校验引擎
- 数据源连接层：负责为上层引擎提供数据源的连接访问。

2.2 部署架构说明

GreatDTS 由五类组件组成分别为 Master（可视化组件）、Worker（执行引擎）、Redis（三方缓存）、Minio（对象存储）、GreatDB/MySQL（元数据存储）。每个模块都可以高可靠部署易扩展，发生故障时可以实现自动切换。因为 Master 和 Worker 支持集群部署，Master 自动分布式调度到集群中任意一个 Worker 节点，可有效提升执行性能。如下为产品部署架构图。



Master: 高可靠的可视化配置管理控制台，负责任务的配置、管理和调度。一般采用多节点多实例部署，多节点互备，只会有一个被选为主节点，轻负载。

Worker: 高可靠的执行引擎，评估、迁移、校验等任务均运行在执行引擎节点上。

Redis: 自带高可靠的三方组件 Redis 作为缓存服务，主要用于存储增量对比的数据。另外在使用到 Oracle 增量复制时，需要额外部署 Kafka。

GreatDB/MySQL: 支持高可用部署的元数据存储，用于存储 Master 的元数据。

Minio: 支持分布式部署的对象存储，用于存储 GreatDTS 的文件，日志和报告等信息。

3. 产品优势

大型数据迁移无疑是一项复杂而艰巨的任务，当前市场上虽然存在多种解决方案，如备份/恢复方法、数据导入导出等，但针对迁移过程中的具体需求，仍需确保以下几点核心要素得到满足：

- **高效迁移大数据量：**在数据海洋般庞大的场景下，如何迅速且准确地完成

迁移，是首要挑战。

- 简化迁移流程：为了降低操作难度、减少出错率，迁移步骤的简化与智能化至关重要。
- 最小化对生产业务的影响：迁移过程中，如何确保生产环境的稳定运行，避免业务中断，是迁移成功的关键。
- 支持异构环境迁移：面对不同数据库系统间的差异，如何实现无缝迁移与适配，是迁移技术的难点。

针对上述挑战，万里公司精心打造了 GreatDTS 产品，这是一款稳定、高效、易用的数据库迁移与同步解决方案，能够完美应对上述难题，并展现出以下显著优势与特点。

3.1 快速部署

- Docker 化部署：利用 Docker 容器技术，实现 GreatDTS 的快速部署与资源隔离，降低部署复杂度，提升维护效率。

3.2 简单易用

- 可视化管理界面：提供直观、友好的控制台界面，用户可轻松创建、管理传输链路。
- 向导式链路创建：通过向导式流程，简化链路创建步骤，降低操作难度。
- 链路状态监控：新版控制台实时展示链路传输状态、进度及性能信息。

3.3 安全高可靠

- 安全传输协议与 Token 认证：各模块间采用安全传输协议及 Token 认证机制，保障数据传输的安全性。
- 断点续传与异常处理：具备自动断点续传功能，有效应对传输中断。同时，定期监测链路状态，确保异常时能及时响应与修复。
- 服务高可靠：GreatDTS 底层采用服务集群架构，确保单个节点故障时，任务能快速切换至其他节点，保障链路稳定可靠。

- 数据高可靠：对部分传输链路提供 7×24 小时的数据准确性校验，及时发现并纠正传输错误，确保数据可靠性。

3.4 多种场景支持

- 全生命周期覆盖：支持迁移前评估、数据和结构迁移、数据同步以及迁移后的数据和结构校验覆盖了数据库迁移的全生命周期。
- 业务不停机迁移：采用数据迁移方式，确保在最小化停机时间的前提下完成数据迁移。迁移过程中，源数据库可持续运行，保障业务连续性，实际停机时间缩短至分钟级别。
- 多端支持：支持 MySQL、Oracle 等多种数据库类型多个版本作为源和目标端。

4. 功能特性

GreatDTS 支持以下源和目标数据库：

源数据库	目标数据库	迁移评估	结构迁移	全量迁移	增量同步	结构对比	数据对比
GreatDB	GreatDB			✓	✓	✓	✓
	MySQL			✓	✓	✓	✓
	Oracle			✓	✓	✓	✓
GreatDB Cluster	GreatDB Cluster			✓	✓	✓	✓

	GreatDB			✓	✓	✓	✓
	MySQL			✓	✓	✓	✓
	Oracle			✓	✓	✓	✓
MySQL 5.7+	GreatDB	✓	✓	✓	✓	✓	✓
	MySQL	✓	✓	✓	✓	✓	✓
	Oracle			✓	✓	✓	✓
Oracle 11G +	GreatDB	✓	✓	✓	✓	✓	✓
	MySQL	✓	✓	✓	✓	✓	✓
	Oracle			✓	✓	✓	✓
ClickHouse	GreatDB		✓	✓			
	MySQL		✓	✓			
MyCat	GreatDB Cluster			✓	✓		

GreatDB			✓	✓		
---------	--	--	---	---	--	--

4.1 迁移评估

GreatDTS 提供了强大的迁移评估能力，能够深入分析源数据库的对象定义和数据内容。通过这一功能，用户可以获取详细的迁移建议，同时，系统还会采集并分析源数据库上的业务应用 SQL，生成支持的、需要修改后支持的以及不支持的 SQL 详细列表。这些建议和信息将极大地帮助业务团队进行必要的改造，从而确保迁移过程的平滑与高效。

4.2 数据迁移

GreatDTS 支持数据迁移，能够将数据库之间的数据进行高效迁移。通过标准化的数据模型和接口，确保数据在不同系统间无缝迁移，保障数据的一致性和完整性。同时 GreatDTS 提供多源数据库同步功能，能够同时对多个同构或异构数据源进行同步操作。采用增量和全量两种方式，确保各数据源之间数据的一致性和实时性，满足复杂的业务需求。

4.3 对象迁移

GreatDTS 支持数据库对象迁移，涵盖了表、视图、触发器、函数、存储过程等多种数据库对象，同时也支持分区表、内置函数、索引等复杂对象的迁移。在迁移过程中，系统还提供了日期格式转换和字符集转换的功能，以满足不同数据库之间的兼容性。

4.4 断点续传

为了应对大数据量迁移过程中系统或网络不稳定的情况，GreatDTS 的断点续传功能可以充分利用资源避免重复操作，有效应对突发状况减少业务中断时间。保证了迁移质量，大大提高了迁移的可靠性和稳定性，为用户提供了更好的使用体验。

4.5 数据和结构校验

数据迁移后，GreatDTS 提供了数据校验功能来验证迁移对象和数据的准确性。通过对数据结构执行完整性校验和对已迁移数据进行全量一致性校验，用户可以确认业务数据已经准确、有效、可靠地迁移到目标端。这一功能为迁移的准确性提供了有力保障。

5. 性能

GreatDTS 支持并行执行，可以充分利用资源提升迁移性能，如下为详细的性能数据。

5.1 测试环境

角色	ip	端口	处理器型号	物理cpu	逻辑cpu	内存	操作系统	系统类型
GreatDTS	172.17.130.41	8085	Intel(R) Xeon(R) Gold 6238 CPU @ 2.10GHz	4	176	251G	CentOS Linux release 7.6.1810 (Core)	x86_64 GNU/Linux
Oracle 11g	172.17.130.52	1521	Intel(R) Xeon(R) Gold 6238 CPU @ 2.10GHz	4	176	251G	CentOS Linux release 7.8.2003 (Core)	x86_64 GNU/Linux

角色	ip	端口	处理器型号	物理cpu	逻辑cpu	内存	操作系统	系统类型
mysql8.0	172.17.130.41	3306	Intel(R) Xeon(R) Gold 6238 CPU @ 2.10GHz	4	176	251G	CentOS Linux release 7.6.1810 (Core)	x86_64 GNU/Linux

5.2 测试结果

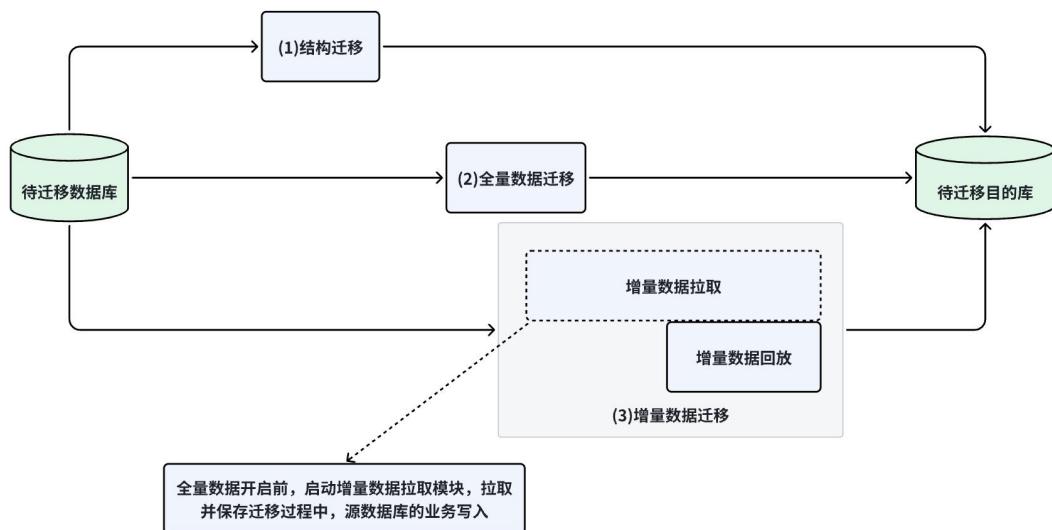
测试项	表数量/数据量	平均迁移速度	平均迁移数据量
100 张具有主键的表全量数据迁移速度	100 张*100 万	19,447 行/s	159.51M/s
100 张无主键的表全量数据迁移速度	100 张*100 万	23,208 行/s	190.35M/s
单张具有主键的大表全量数据迁移速度	1 张*1 亿	31,656 行/s	268M/s
单张无主键的大表全量数据迁移速度	1 张*1 亿	4,680 行/s	39.62M/s
单张具有主键的 lob (大对象字段) 表全量数据迁移速度	1 张*1 千万	3,424 行/s	210M/s
单张无主键的 lob (大对象字段) 表全量数据迁移速度	1 张*1 千万	662 行/s	40.64M/s
1.48T 数据全量迁移测试	101 张, 共 2 亿	41,650 行/s	322M/s

6. 应用场景

6.1 不停机迁移数据库

在传统的数据库迁移过程中，为确保数据的一致性，通常需要在迁移期间暂停向源数据库写入数据。这一步骤可能会因数据量和网络状况的不同而耗费数小时乃至数天，对业务连续性造成显著影响。

然而，借助 GreatDTS 可以实现数据库迁移过程中的最小化停机时间。在数据迁移的整个周期内，应用程序能够持续运行，无需中断。唯一的停机时间发生在将应用程序切换到目标数据库的瞬间，这个时间窗口通常可以缩短至分钟级别。



数据迁移过程包含以下三个关键阶段：

- **结构迁移**：此阶段涉及将源数据库的结构（如表、索引、视图等）复制到目标数据库。
- **全量数据迁移**：在结构迁移完成后，将源数据库中的所有数据复制到目标数据库。这一步骤确保了目标数据库在迁移开始时即拥有源数据库的完整数据副本。
- **增量数据迁移**：在全量数据迁移进行的同时或之后，DTS 会实时捕获源数据库中发生的数据变更（如插入、更新、删除操作），并将这些变更同步到目标数据库。这一步骤确保了迁移过程中源数据库和目标数据库之间的数据一致性。

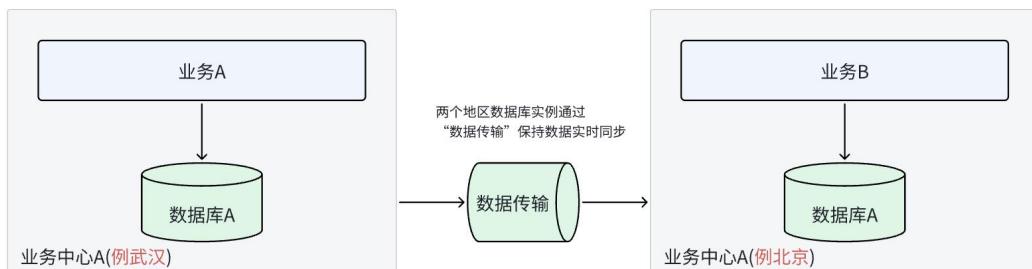
迁移完成后，您需要对目标数据库的数据和结构进行验证，以确保它们与您的应用程序完全兼容。验证通过后，您可以将应用程序切换到目标数据库，从而实现平滑且几乎无缝的迁移过程。

6.2 数据异地灾备

当您的应用程序仅部署在单一地域时，可能会面临断电、网络中断等不可预见的风险，这些风险有可能导致服务中断，进而影响用户体验和业务连续性。

为了提升服务的可用性和稳定性，可以考虑在另一个地域构建灾备中心。这一策略的核心在于利用 GreatDTS 实现业务中心和灾备中心之间的数据同步。DTS 会持续监控并同步两个地域之间的数据更新，确保地域间的数据始终保持一致。

在正常情况下，应用程序会继续使用业务中心的数据提供服务。一旦主地域（即业务中心所在地域）发生故障，可以迅速将用户请求切换到灾备地域。由于 GreatDTS 已经确保了数据的一致性，因此灾备地域可以无缝接管业务，继续为用户提供服务，从而最大限度地减少服务中断对业务的影响。



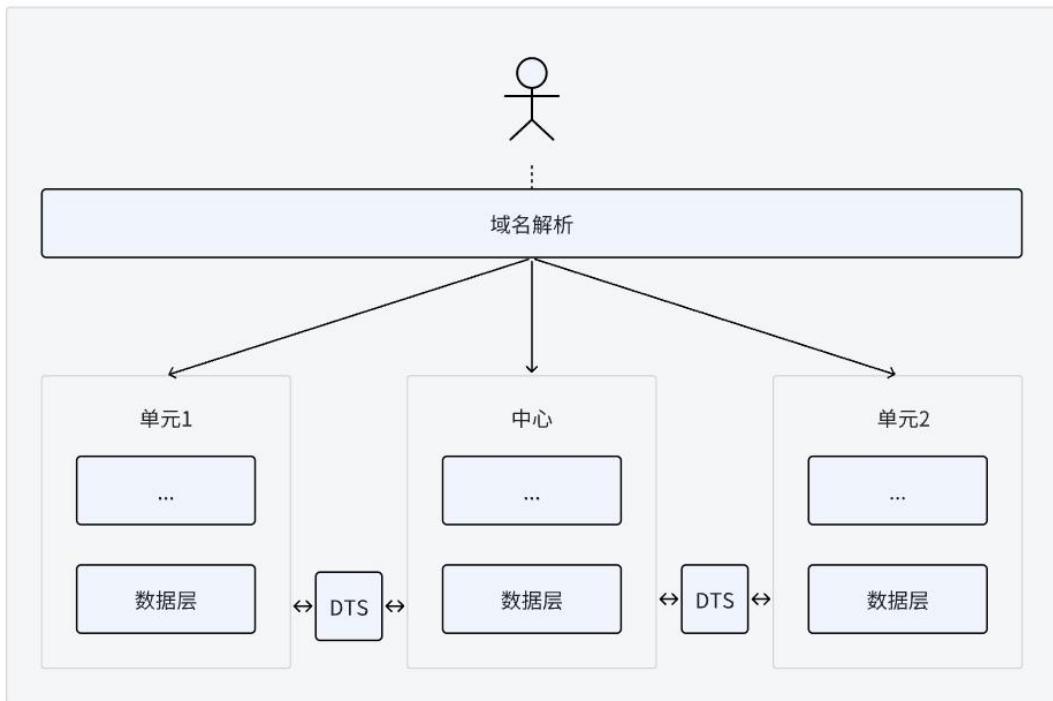
6.3 异地多活

随着业务的迅猛发展和用户数量的急剧增长，单一地域的业务部署可能会面临多重挑战：

- 用户体验受损：由于用户分布广泛，地理位置较远的用户在访问时可能会遇到较高的延迟，从而严重影响他们的使用体验。
- 业务扩展受限：单个地域的底层基础设施，如供电能力和网络带宽建设等，可能会成为业务进一步扩展的瓶颈。

为了应对这些挑战，可以考虑在同城或异地构建多个业务单元。这些业务单元之间通过 GreatDTS 实现数据的双向实时同步，确保全局数据的一致性和完整性。当任何一个业务单元发生故障时，可以迅速将故障单元的流量切换至其他正常运行的单元，实现业务的秒级恢复，从而有效保障服务的高可用性。

此外，还可以根据业务的某个特定维度（如用户所属区域）来分配和调度业务流量。通过将流量分流至各个业务单元，可以实现用户的就近访问，显著降低网络延迟，进一步提升用户体验。同时，多个业务单元分布在不同地域，也有助于打破单地域基础设施对业务扩展的限制，为业务的持续发展和扩展提供有力支持。



6.4 横向扩展能力

在面临大量读请求的应用场景中，单个数据库实例可能会因为承受不住全部的读取压力而出现性能瓶颈。为了有效缓解这一问题，可以借助 GreatDTS 的实时同步功能来构建只读实例。

通过 GreatDTS 的实时同步，您可以将主数据库实例的数据实时复制到这些只读实例中，确保数据的一致性和实时性。随后，可以将读请求分流至这些只读实例中，实现读能力的弹性扩展。这样，不仅可以分担主数据库实例的读取压力，还能提升整个系统的响应速度和稳定性。



联系我们 | Contact Us



地址：北京市朝阳区CBD国际大厦7层701B

电话：400-032-7868

邮箱：sales@greatdb.com

网站：<https://www.greatdb.com>

北京万里开源软件有限公司

稳定 · 性能 · 易用