



目 录

目 录.....	2
第 1 章 GoldenGate 概述.....	3
1.1 GoldenGate 的历史.....	3
1.2 GoldenGate 家族.....	3
1.3 GoldenGate 及 Oracle 产品战略.....	3
1.4 GoldenGate 支持的平台以及数据库.....	4
1.5 GoldenGate 的定位.....	5
1.6 GoldenGate 技术架构.....	5
1.6.1 Manager 进程.....	5
1.6.2 Extract 进程.....	6
1.6.3 Pump 进程.....	6
1.6.4 Trail 文件.....	6
1.6.5 Relicat 进程.....	6
1.6.6 GGSCI.....	7
1.7 GoldenGate 的复制模式.....	7
1.8 各种应用模式展示.....	8
1.8.1 高可用性: Active-Active.....	8
1.8.2 零宕机升降和数据迁移.....	8
1.8.3 数据集成: 活动备份.....	9
1.8.4 商业智能.....	9
1.8.5 事务性数据集成.....	10
1.9 GoldenGate 拓扑结构.....	10
1.10 GoldenGate 10g/11g 新特性.....	11
1.11 GoldenGate 的优势.....	12
1.12 GoldenGate 应用情况.....	12

第 1 章 GoldenGate 概述

GoldenGate 现在是业内成熟的数据容灾与复制产品，经过多年的发展和完善，现在已经成为业内事实上的标准之一。

1.1 GoldenGate 的历史

GoldenGate 公司成立于 1995 年的美国加州旧金山，它的名称源自旧金山闻名于世的金门大桥。两位创始人 Eric Fish 和 Todd Davidson 最初旨在为 Tandem 计算机公司设计一个容错系统，由于 GoldenGate 的健壮性和出色的数据复制功能，银行用它来把 ATM 网络的交易数据发送到 IBM 大型机。后来广泛地应用到金融行业及要求数据复制高效、健壮的行业，全球 licences 数量超过 4000。

该公司于 2009 年 9 月被 Oracle 公司收购，在此之前 Oracle 和 GoldenGate 公司就有了长达超过 10 年的合作关系。Oracle 收购 GoldenGate 以后，按照 Oracle 公司的一贯策略，迅速把它和自己的数据库、中间件、以及应用集成，依托 Oracle 公司研发技术的优势，并对其做了大量的更新和改进。截止到目前为止，已经有全球有五百多家大客户使用 GoldenGate 作为其容灾、复制的解决方案。作为一个企业级的成熟产品，因为其快速、易用、灵活、健壮等特性，越来越多的用户把它作为关键业务系统容灾、复制、同步的首选。随着 Oracle 公司对其不断改进，这个产品会越来越稳定、成熟、可靠，同时也会有更多人去学习它，使用它。

目前 GoldenGate 的最新版本为 11.1，为了和 Oracle 的数据库、中间件产品的称谓保持一致，Oracle 称之为 11g，目前 Oracle 公司把它归到 Fusion Middleware 也就是融合中间件产品线中。但实际上它和数据库的联系更加紧密，有经验的 DBA 通过花较短的时间就可以迅速地掌握它。

1.2 GoldenGate 家族

Oracle GoldenGate 最为常见的家族成员包括 GoldenGate、GoldenGate Director（现更名为 GoldenGate Management Pack，但是绝大多数熟悉 GoldenGate 的人还是习惯性地称其为 GoldenGate Director，为了保证上下文的一致性，我这里统一称其为 GoldenGate Director）、GoldenGate Veridata。另外，也有我们不大常见的 GoldenGate for Mainframe，和 GoldenGate Adapters。

GoldenGate 产品是核心产品，GoldenGate Director 为 GoldenGate 提供友好的 GUI 配置管理界面，GoldenGate Veridata 为 GoldenGate 源端和目标端提供数据比对和校验的功能。

注意这三者并不是一个打包的产品，比如如果您购买了 GoldenGate 软件，如果需要图形界面或需要数据校验的功能，就需要额外购买 GoldenGate Director 或者 GoldenGate Veridata。

1.3 GoldenGate 及 Oracle 产品战略

在 Oracle 收购 GoldenGate 以后，与 Oracle 原有的 Data Guard、ODI 互为补充，共同为企业提供跨平台实时数据同步的解决方案。与此同时，又可以与 Oracle Real Application Cluster、Data Guard 一起为用户提供丰富、灵活的容灾方案及高可用性，这既是 Oracle 公司当前也是其未来的产品发展策略与方向。

另外有的读者可能注意到了 Oracle 公司在其集成在数据库中的 Streams 的态度的一些变化。Oracle 公司声称对其产品 Streams 将不再做任何重大改进，而是会把 Streams 的一些优秀特性集成到 GoldenGate 产品上来，同时 Oracle 也承诺继续对现有的使用 Streams 的客户提供技术支持。与此同时，集成在 Oracle 数据库中的另外一项逻辑复制产品——Data Guard Logical Standby 也遭受了同样的命运。Oracle 对其产品线的整合能力时见一斑。

1.4 GoldenGate 支持的平台以及数据库

截止到笔者撰写本书为止，GoldenGate 几乎支持市面上流行的所有主流的操作系统平台和数据库。

在不同的平台生成的 core 文件，有不同的本地堆栈跟踪工具来分析：

数据库产品		操作系统 及平台
源端	目标端	
Oracle Database	所有源端支持的数据库	Windows 2000, 2003, XP Server 2008
MySQL	HP Neoview	Linux
IBM DB2	ETL products	Oracle Sparc
Microsoft SQL Server	JMS message queues	HP NoStop
Sybase ASE	Netezza	HP-UX
Ingres	及任何 ODBC 兼容的数据库	HP-TRU64
Timesten		HP-OpenVMS
Teradata		IBM AIX
Enscribe		IBM z/OS
SQL/MP		
SQL/MX		

表 1-1

目前经过 GoldenGate 11.1 认证的主流数据库版本包括：

- Oracle 8i (只支持 DML)；
- Oracle 9i 以上所有的数据库版本 (支持 DML 和 DDL)；
- IBM DB2 UDB 8.1、8.2、9.1、9.5 以上版本 (只支持 DML)；
- Microsoft SQL Server 2000、2005、2008 (只支持 DML)；

- MySQL 4.1、5.0（只支持 DML）；
- Sybase ASE 12.5.4、15.0（只支持 DML）；

除 Oracle 数据库 9i 以上版本支持 DDL 以外，8i 及其它数据库均不支持 DDL。其它经认证的操作系统及数据库版本请参见资料篇的列表。

1.5 GoldenGate 的定位

- 零宕机时间数据库升级和迁移
- 满足用户亚秒级实时数据的需求；
- 可持续的数据高可用性和实时商务智能；
- 异构平台及跨操作系统实时数据同步；
- 对源系统和目标系统是非侵入式的。

1.6 GoldenGate 技术架构

和传统的逻辑复制一样，Oracle GoldenGate 实现原理是通过抽取源端的 redo log 或者 archive log，然后通过 TCP/IP 投递到目标端，最后解析还原应用到目标端，使目标端实现同源端数据同步。以下是 Oracle GoldenGate 的技术架构：

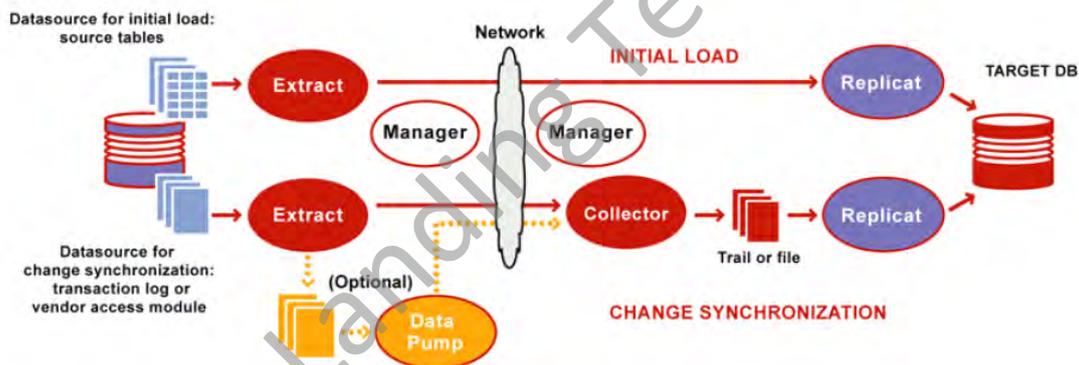


图 1-1

1.6.1 Manager 进程

Manager 进程是 GoldenGate 的控制进程。如果把所有的 Oracle 进程比喻为军队，那么 Manager 就相当于司令。Manager 进程运行在源端和目标端上，它主要作用有以下几个方面：启动、监控、重启 GoldenGate 的其它进程，报告错误及事件，分配数据存储空间，发布阈值报告等。

每个源端或者目标端有且只能存在一个 Manager 进程。其运行状态有两种 RUNNING（正在运行）和 STOPPED（已经停止）。

在 Windows 系统上，Manager 进程是作为一个服务来启动的，而在类 Unix 系统中，Manager 则是一个操作系统进程。

1.6.2 Extract 进程

Extract 运行在数据库源端，负责从源端数据表或者日志中捕获数据。在早期的 GoldenGate 版本中，它通常被称为 Collect 进程。按照其所处的阶段不同，Extract 的作用可以按照时间来划分：

初始数据装载阶段：在初始数据装载阶段，Extract 进程直接从源端的数据表中抽取数据。

同步变化捕获阶段：初始数据同步完成以后，Extract 进程负责捕获源端数据的变化（DML 和 DDL）

Extract 进程利用其内在的 checkpoint 机制，周期性地检查并记录其读写的位置，通常是写入到一个本地的 trail 文件。这种机制是为了保证如果 Extract 进程终止或者操作系统宕机，我们重新启动 Extract 进程后，GoldenGate 能够恢复到以前的状态，从上一个断点处继续往下运行，而不会有任何数据损失。

Extract 进程的状态包括 STOPPED（正常停止）、STARTING（正在启动）、RUNNING（正在运行）、ABENDED（Abnormal End 的缩写，表示异常结束）。

1.6.3 Pump 进程

Pump 进程运行在数据库源端，其作用非常简单。如果源端使用了本地的 trail 文件，那么 Pump 进程就会把 trail 以数据块的形式通过 TCP/IP 协议发送到目标端，这通常也是推荐的方式。Pump 进程本质是 Extract 进程的一种特殊形式，如果不使用 trails 文件，那么就是 Extract 进程在抽取完数据以后，直接投递到目标端。

与 Pump 进程相对应的叫 Server Collector 进程，这个进程不需要引起我们的关注，因为在实际操作过程中无需我们对其进行任何配置，所以对我们来说它是透明的。它运行在目标端，其任务就是把 Extract/Pump 投递过来的数据块重新组装成 trail 文件，我们称之为远程 trail 文件。

1.6.4 Trail 文件

为了更有效、更安全的把数据库事务信息从源端投递到目标端。GoldenGate 引进 trail 文件的概念。前面提到 Extract 抽取完数据以后 GoldenGate 会将抽取的事务信息转化为一种 GoldenGate 专有格式的文件。然后 Pump 负责把源端的 trail 文件投递到目标端，所以源、目标两端都会存在这种文件，源端存放的 trail 文件叫本地 trail 文件，目标端存放的 trail 文件叫远程 trail 文件。trail 文件存在的目的旨在防止单点故障，将事务信息持久化，并且使用 checkpoint 机制来记录其读写位置，如果故障发生，则数据可以根据 checkpoint 记录的位置来重传。

值得一提的是，trail 文件并不是总是必须的。我们可以在配置 Extract 进程的时候通过 TCP/IP 协议直接把日志的信息投递到目标端。但通常并不推荐这么做，因为一旦发生系统宕机或者网络故障，则有可能造成数据的丢失。

1.6.5 Relicat 进程

Relicat 进程，通常我们也把它叫做应用进程。运行在目标端，是数据传递的最后一站，负责读取目标端 trail 文件中的内容，并将其解析为 DML 或 DDL 语句，然后应用到目标数据库中。

和 Extract 进程一样，Relicat 也有其内部的 checkpoint 机制，保证进程重启启动后可以从上次记录的位置开始恢复而无数据损失的风险。

它的运行状态和 Extract 进程一致，包括 STOPPED，STARTING，RUNNING，ABENDED。

1.6.6 GGSCI

GGSCI 是 GoldenGate Software Command Interface 的缩写，它提供了十分丰富的命令来对 GoldenGate 进行各种操作，例如创建、修改、监控 GoldenGate 进程等等。

我们绝大部分的操作都是通过它来完成的。当然如果您需要 GUI 方式的图形界面来管理，则需要购买 GoldenGate Director。

1.7 GoldenGate 的复制模式

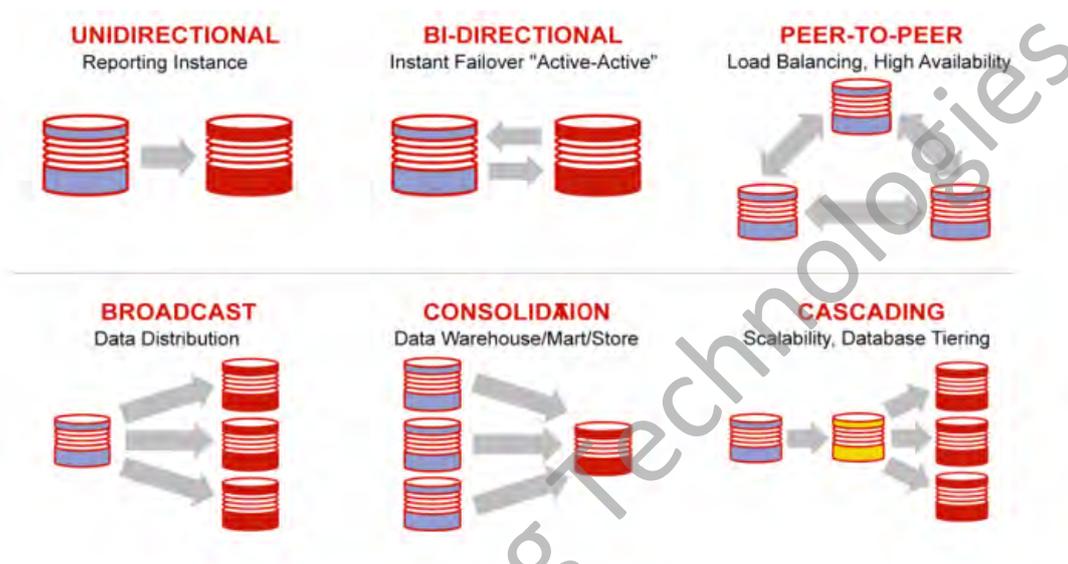


图 1-2

GoldenGate 的模式包括上图中描述的几种，其中“一对一”是 GoldenGate 最简单的一种模式，也是最常用的模式。

这种模式的一种典型应用就是用于数据容灾，通常源端数据库为生产端，目标端数据库为容灾端。

另外一种应用场景是把源端的 OLTP 系统产生的交易日志传送到目标端使用 BI 数据仓库或者是 OLAP。

1.8 各种应用模式展示

1.8.1 高可用性：Active-Active

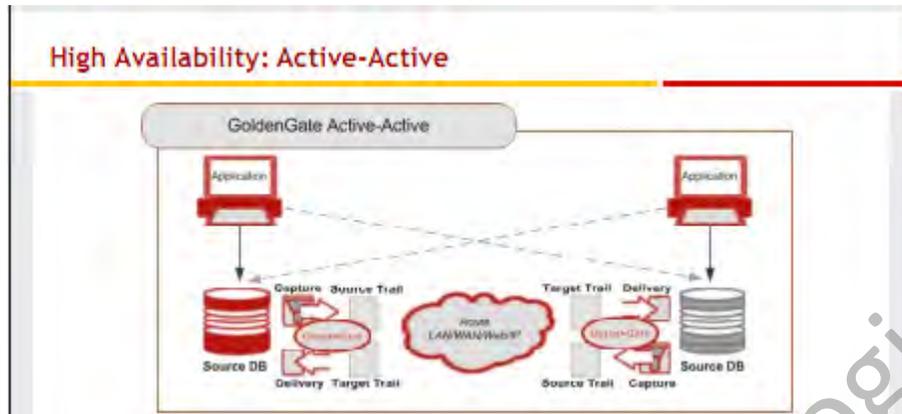


图 1-3

优点:

- 实现连续可用性
- 实现事务加载和分布式（用内嵌的冲突检测）
- 提高性能
- 降低 TCO

1.8.2 零宕机升降和数据迁移

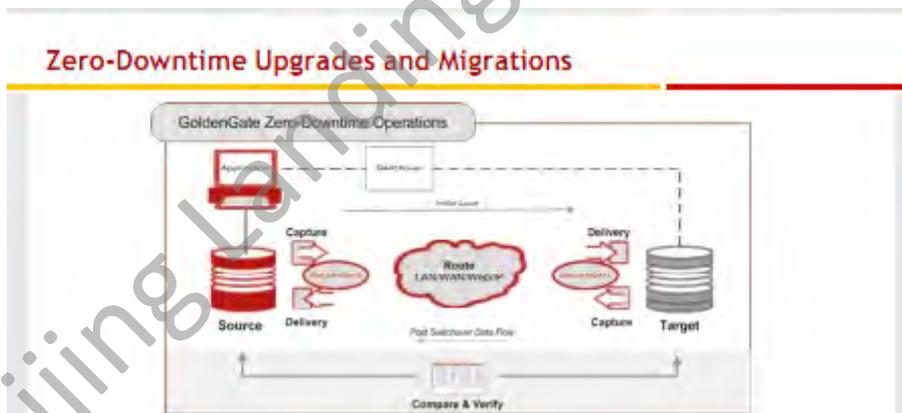


图 1-4

优点:

- 在硬件，数据库，操作系统或是应用程序升级和数据迁移中消除计划宕机
- 减小故障恢复应急风险
- 提高用户迁移数据成功率

1.8.3 数据集成：活动备份

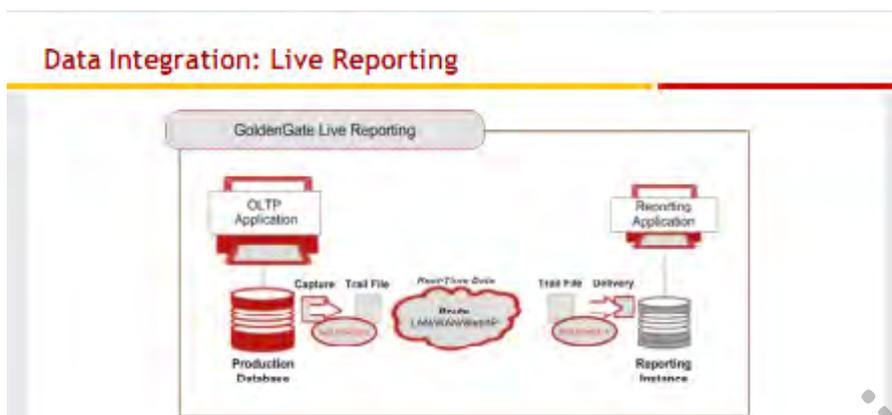


图 1-5

优点：

- 使用实时数据更好更快的决策
- 删除源端报告开销
- 减少用户的需求和数据量的成本和规模的增长
- 报表需求方面利用有效制度

1.8.4 商业智能

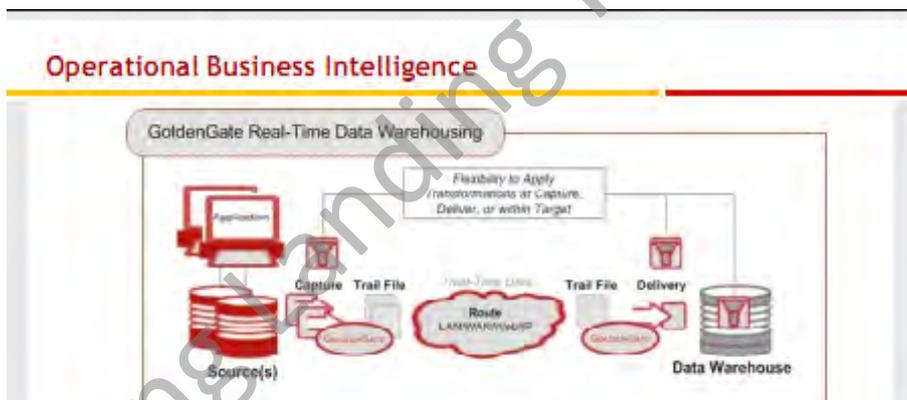


图 1-6

优点：

- 使用实时数据更好更快的决策
- 消除批量窗口依赖关系
- 减少源端开销
- 维护数据质量参照完整性
- 利用其灵具有活性转换和集成的 ETL

1.8.5 事务性数据集成

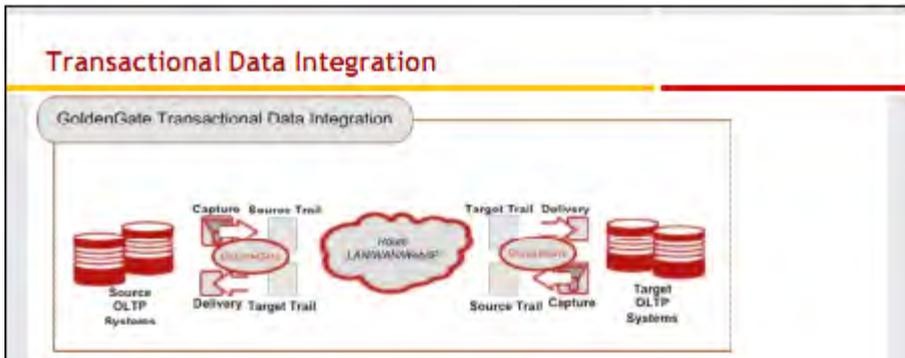


图 1-7

优点:

- 在事务处理系统很容易集成大量实时数据量
- 减少开销，消除批量窗口
- 提高可扩展性
- 增强 SOA 和 EDA 环境（投递基于 JMS 的消息传递系统）

1.9 GoldenGate 拓扑结构

GoldenGate 有着比较灵活的拓扑结构，如下所示：

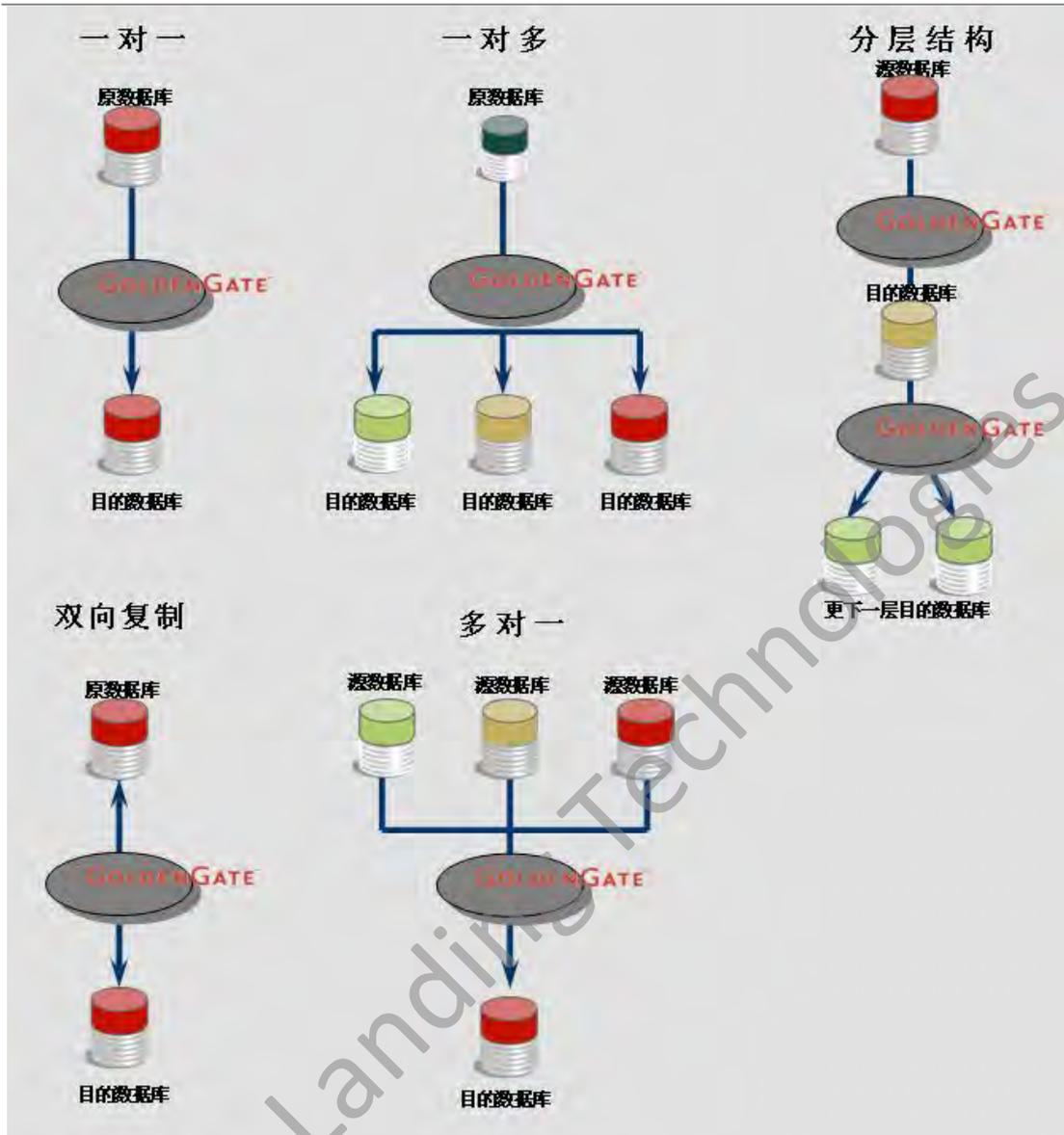


图 1-8

1.10 GoldenGate 10g/11g 新特性

1. 添加了解决方案在 Oracle E-Business Suite, Oracle PeopleSoft, and Oracle JD Edwards 等方面;
2. 支持更多的数据类型和导入模式在 Oracle Exadata 数据库上;
3. 扩展了对异构的支持:
 - 基于日志的捕获, 投递到 IBM DB2 v9.7;
 - 具有投递到 TimesTen databases 的天赋;
 - 以 JMS 为基础的消息系统中捕获;
 - 自动投递到 IBM 的 DB2。

4. 降低了在长事务交易中中断所影响的恢复时间；
5. 增强了追踪事务的能力，更容易去排除性能瓶颈。

1.11 GoldenGate 的优势

目前在数据同步/数据复制市场中，除 GoldenGate 外，不乏优秀的产品，他们大部分都是基于 redo 或归档日志进行提取记录实现与源同步。

以下是 Oracle GoldenGate 白皮书中列出的 GoldenGate 的优势：

移动	管理	集成
Speed -Subsecond Latency	Transaction integrity	Heterogeneous Data Sources
Volume -Thousands of TPS	Transaction Capture	Mapping
Log-based Capture	Guaranteed Delivery	Transformation
Native, Local Apply	Conflict Detection, Resolution	Enrichment
Efficient I/O and Band with Usage	Dynamic Rollback	Decoupled Architecture
Bidirectional	Incremental TDM	Table, Row, . Column Filtering
Group Transactions	Initial Data Load	XML, ASCII, SQL Formats
Bulk Operations	GUI-based Monitoring and Configuration	Queue Interface
Compression	Proactive Alerts	Stored Procedures
One-to-Many, Many-to-One	Encryption	User Exits
Cascade	Real-Time Deferred or Batch	ETL integration
	Event Markers	Java/JMS Integration

表 1-2

1.12 GoldenGate 应用情况

1. 高可用性与容灾
 - 容灾与应急备份

- 消除计划内停机
 - 双业务中心
 - OLTP 和 OLAP 分开
2. 主数据 数据库移植、升级
 3. 实时数据集成
 - 数据仓库实时供给
 - 实时报表
 - 政府、企业垂直部门级数据同步
 - 主数据

Beijing Landing Technologies