



Goldengate从入门到精通 第1课

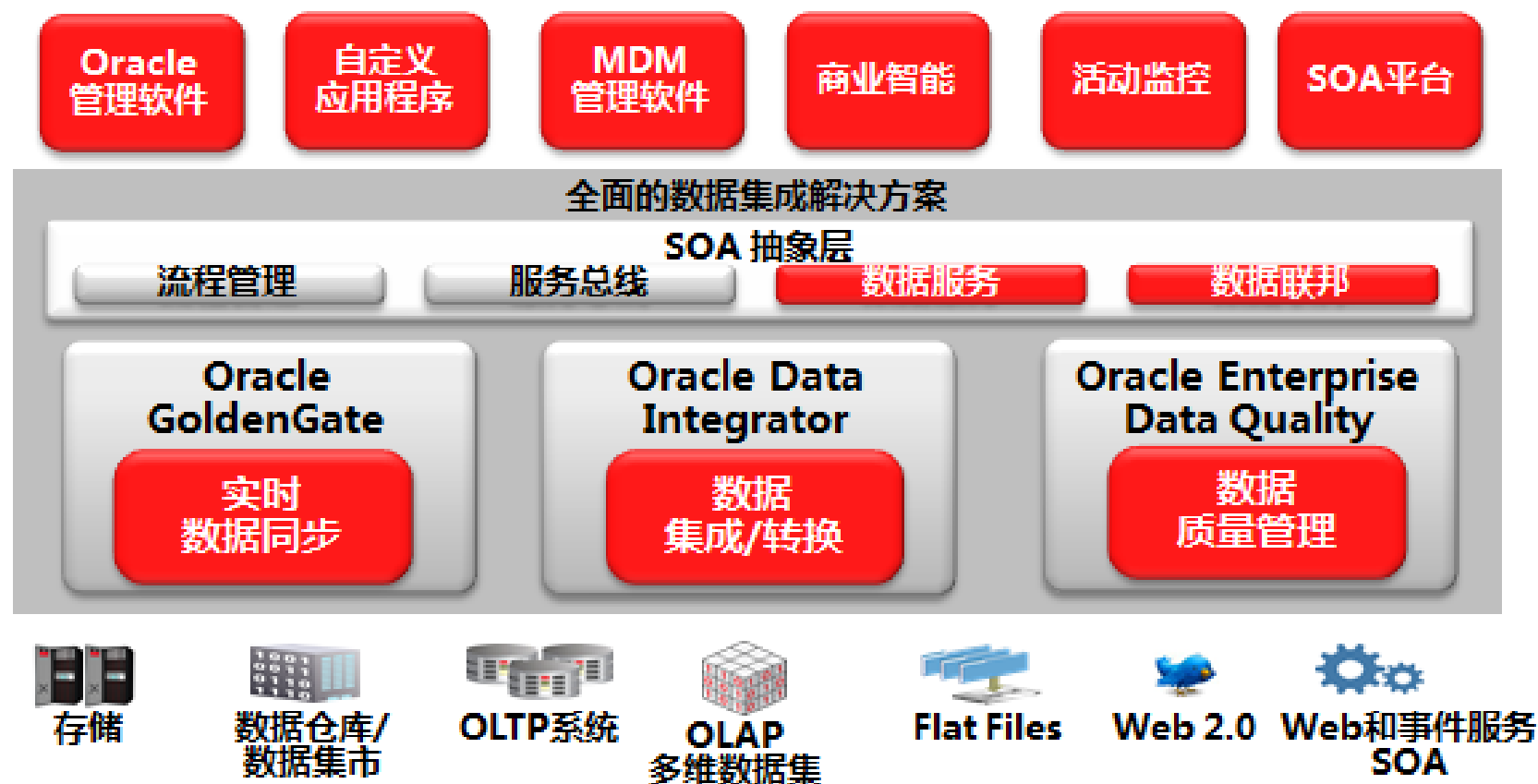
Goldengate概述

DATAGURU专业数据分析社区

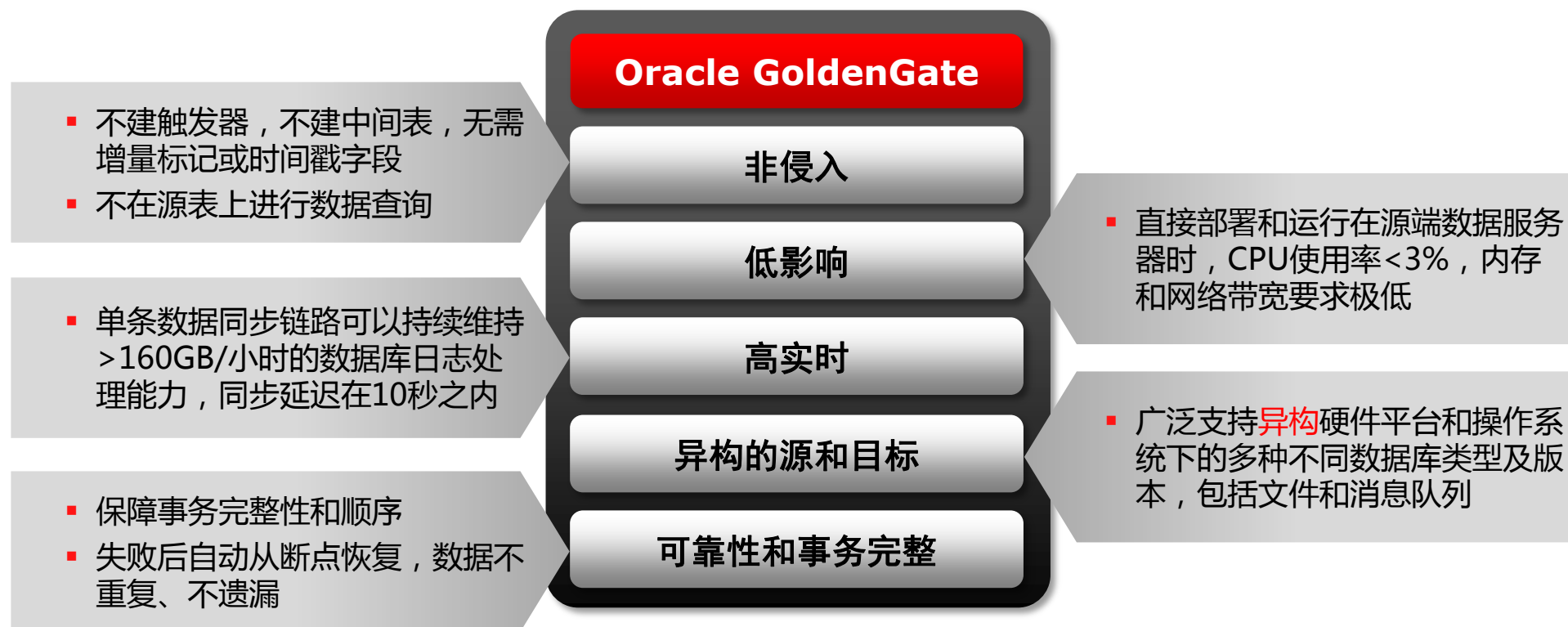
- Goldengate是什么？
- Goldengate适用的场景
- Goldengate基本原理
- Goldengate复制拓扑和异构
- Goldengate产品线介绍
- Goldengate产品的下载
- 快速的搭建一个Goldengate复制环境（非RAC环境）（linux和windows环境两套环境演示）

Goldengate是什么

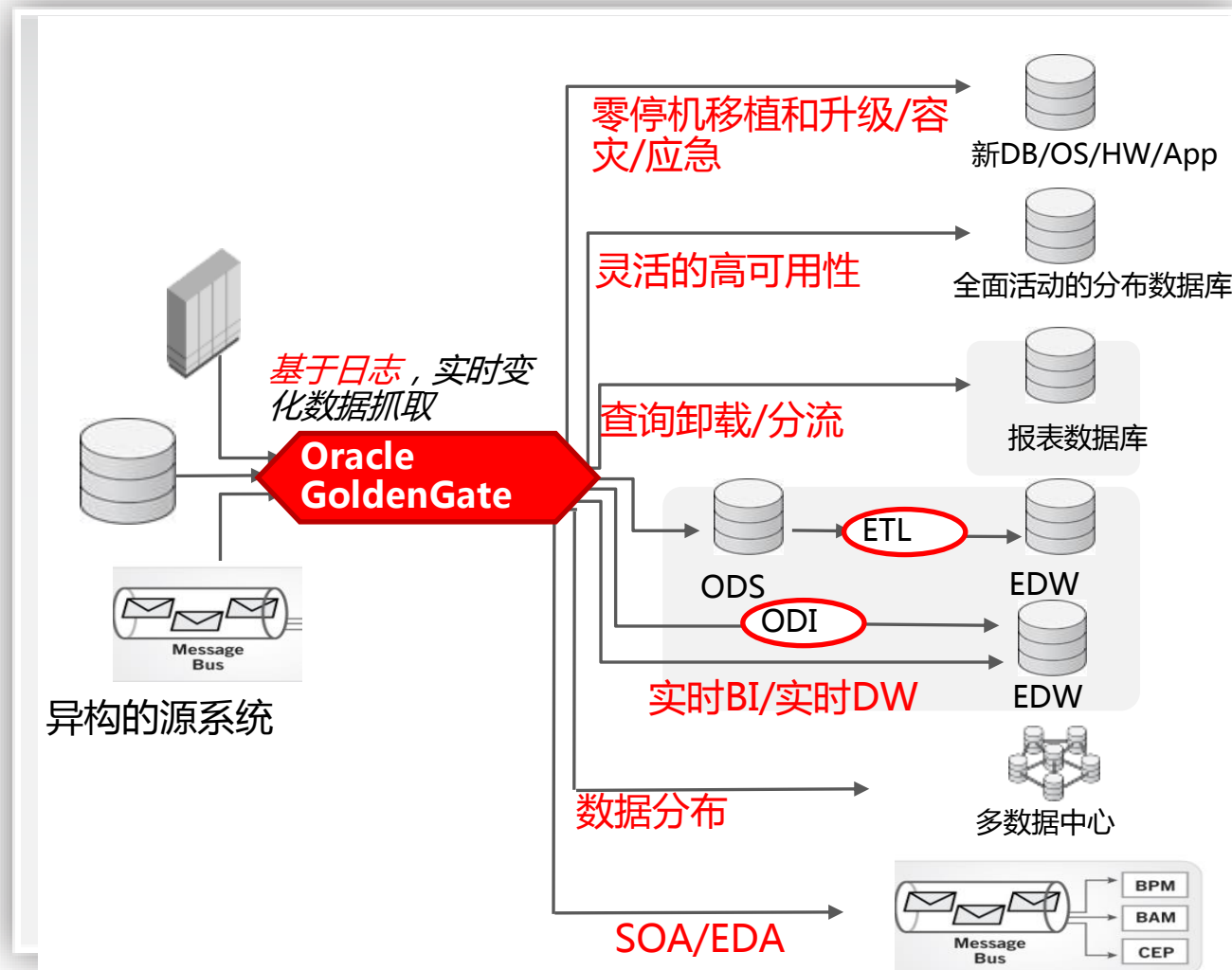
- Oracle的数据集成方案（Goldengate属于Oracle Fusion Middleware产品线，09年收购，Oracle stream的替代者）



- Oracle GoldenGate 提供**异构**环境间事务数据的**实时**、**低影响**的捕获、路由、转换和交付



Goldengate数据集成应用场景



- 标准化的单一的技术解决多种不同的需求
- 国内：主要场景用于系统的连续可用性和报表/BI系统的实时数据访问的要求

附加值高
灵活性高
好东西！

但是：经常
abend！经常初
始化！好累！！
好烂的产品！
维护成本高！

规范很重要！
基于对OGG的深入理解！

Goldengate基本原理

基于日志捕获技术的实时增量数据集成

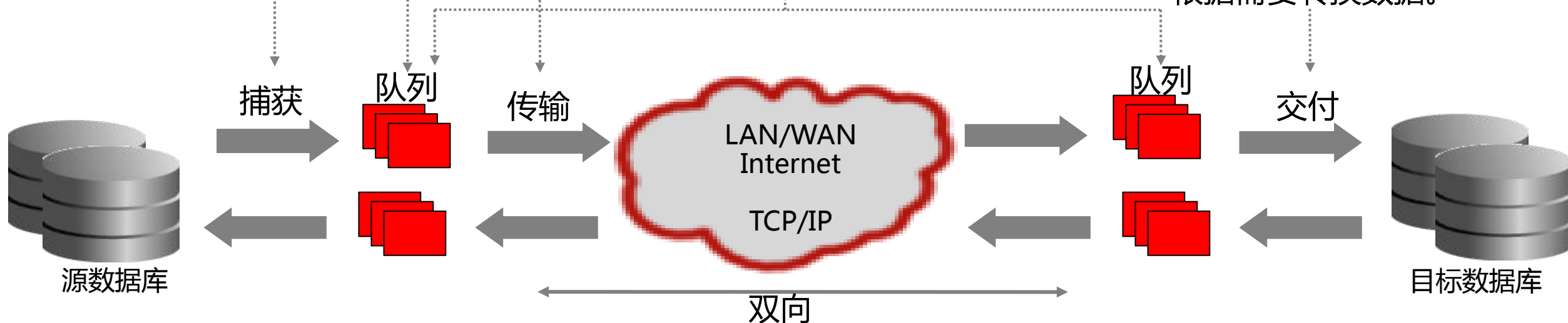
捕获：通过读取事务日志，在事务发生时捕获（也可过滤）提交的事务。任何非实际事务均不被复制到目标，保证不传播错误。

队列：准备数据，使数据排入队列以备路由。

传输：分发数据以便路由到目标，可实现一次捕获多次投递。

路由：压缩、加密数据以便路由到目标，最小化的带宽占用。

交付：保证事务数据完整，根据需要转换数据。



Goldengate基本原理 – 可靠性

确保事务完整性

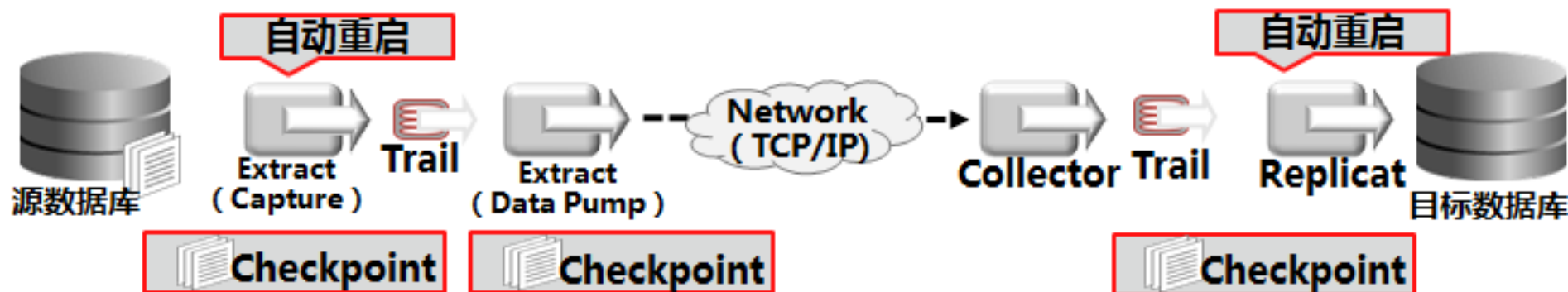
■ 各种故障的解决方法：每个进程的读、写检查点

➤ GoldeGate的故障

- ✓ 进程故障 -> 进程自动重启
- ✓ Trail文件故障 -> Checkpoint

➤ GoldenGate以外的故障

- ✓ 服务器/DB/网络的故障 -> Checkpoint



Goldengate基本原理 – 灵活性

数据的加工、拓扑的灵活、异构的支持、安装配置的灵活

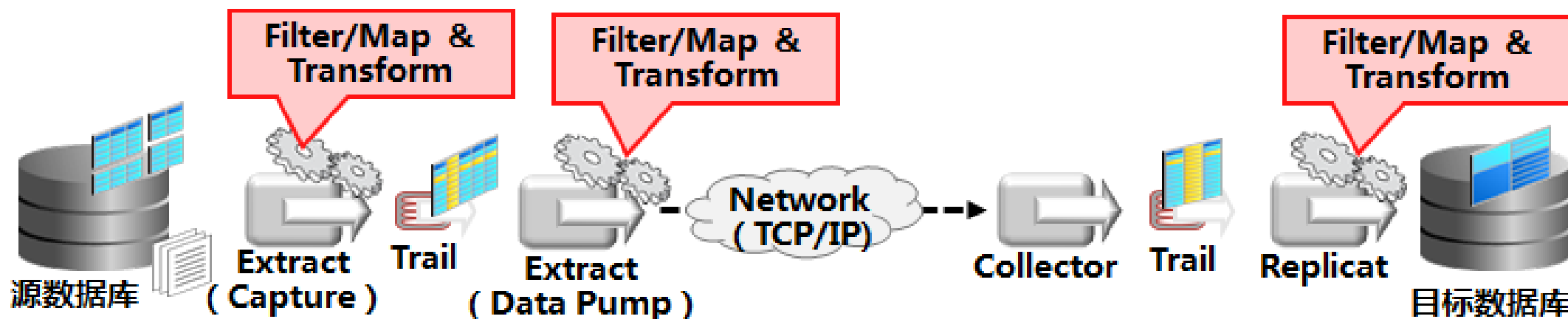


■ 可以配置Filter/Map以及转换规则的组件

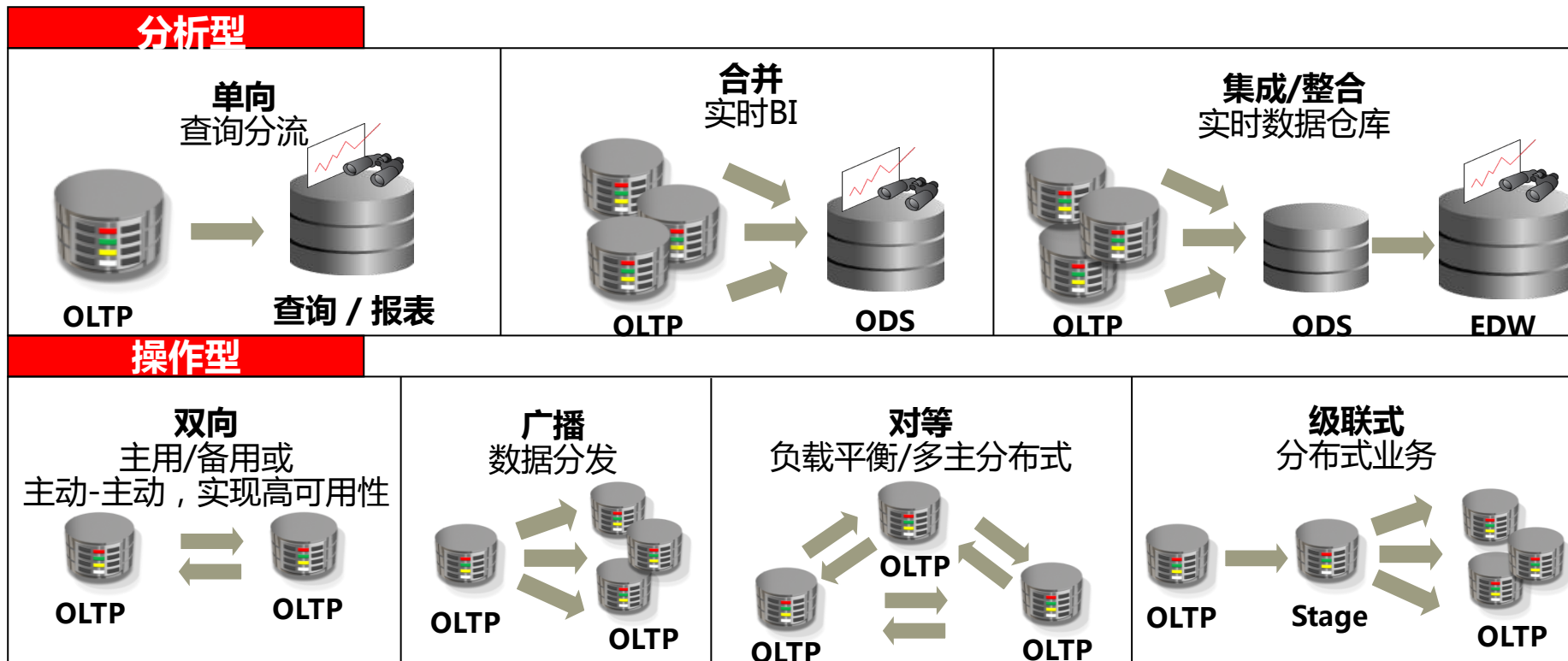
- Capture
- Data Pump
- Replicat

每个组件的安装时是灵活的，不同于其他产品；

每个组件的配置也是灵活的，只要trail文件、checkpoint信息在，想怎么玩就怎么玩！



Goldengate复制拓扑



捕获和交付

- Oracle
- IBM DB2 LUW
- IBM DB2 z 系列
- IBM DB2 i 系列/AS400
- Microsoft SQL Server
- Sybase ASE
- Teradata
- C-Tree
- Enscribe
- SQL/MP
- SQL/MX
- MySQL
- JMS 消息队列

交付

- Netezza
- TimesTen
- PostgreSQL
- GreenPlum
- HP Neoview
- 平面文件
- ETL 产品
- Coherence hotcache

GoldenGate 12c 新增支持

- Oracle 12c
- MySQL 5.6
- MySQL Cluster 7.1 和 7.2
- Sybase 15.7
- DB2 LUW 10.1
- Teradata 14.10
- SQL Server 2012

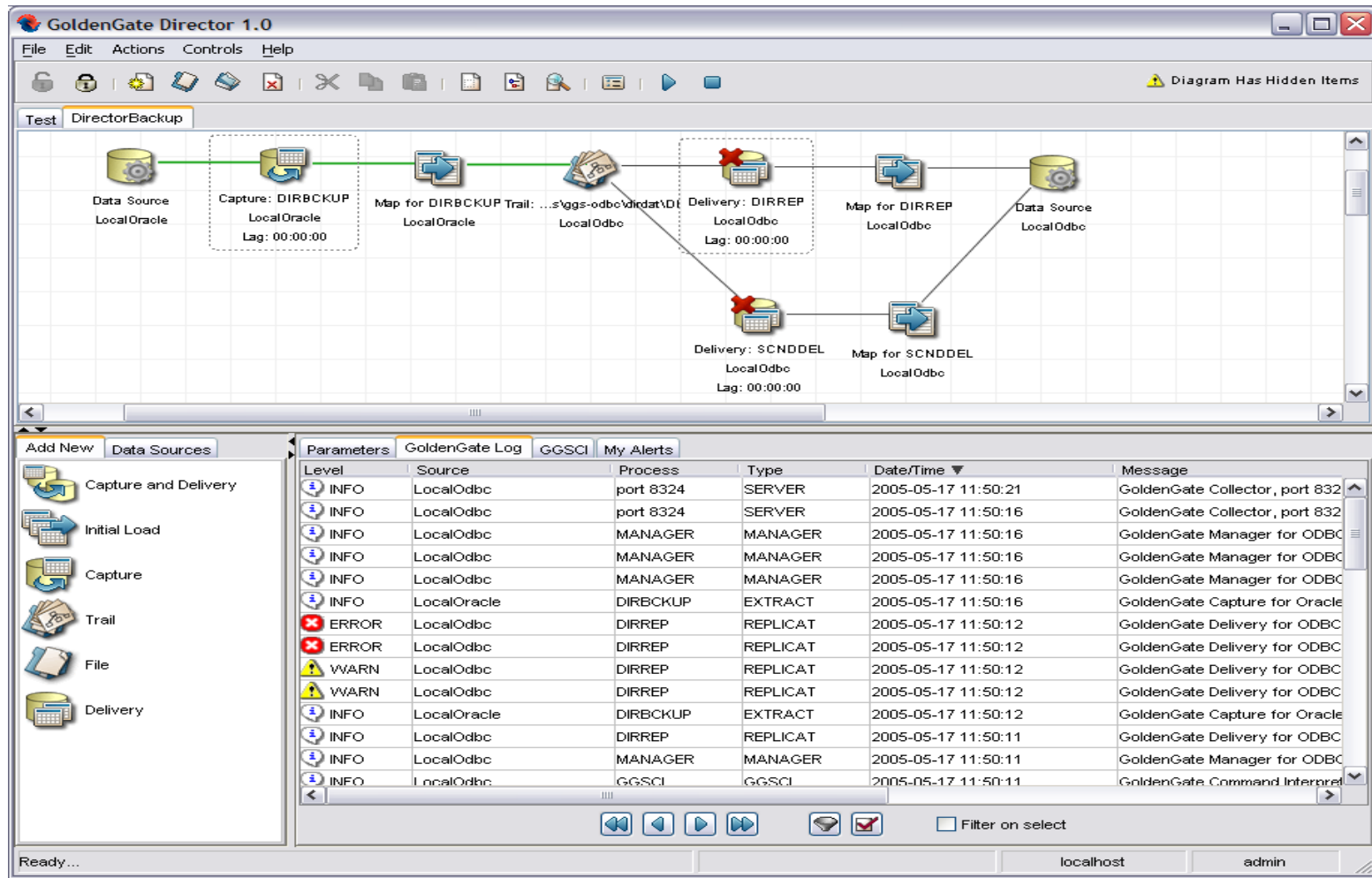
- Oracle Goldengate (凡是安装的都需要license , RAC每个节点都需要license)
- Management Pack for Goldengate
 - Director
 - Monitor
 - EM12c plug-in
- Goldengate Veridata (不能脱离OGG环境)
- Goldengate Application Adapter

OGG Management Pack

远程集中管理和监控



- 定义，配置，管理和报告
GoldenGate在业务系统中配置情况
- 使用GoldenGate管理服务体系架构
- 关键特性:
 - 集中化管理所有的GG进程
 - 丰富的基于web的客户端界面增强了可视性
 - 告警通知，Email方式或者第三方的监控产品集成
 - 实时反馈GG的进程运行情况
 - 无影响的部署不会影响系统性能

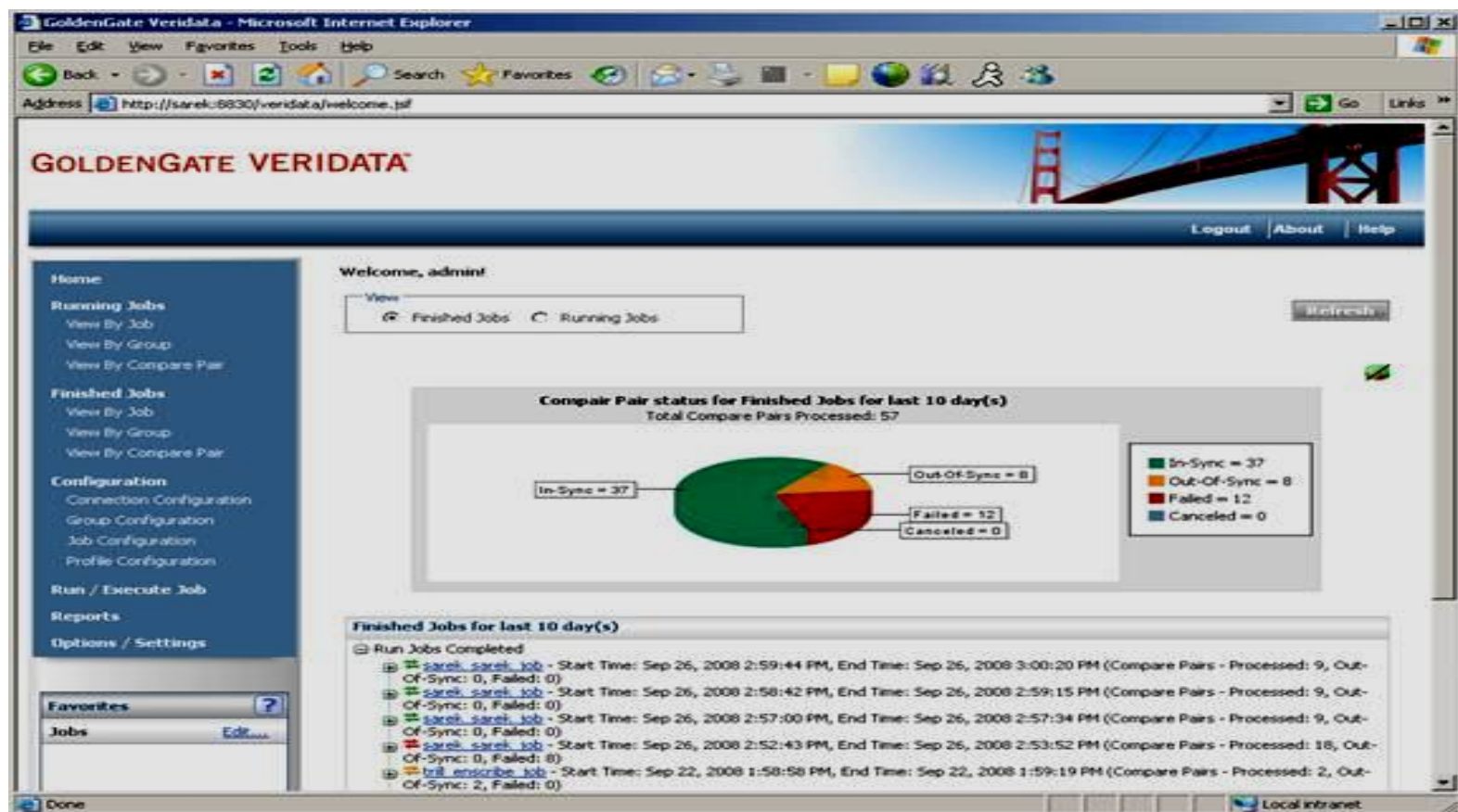


GoldenGate Veridata™

数据一致性比较



- 高速的数据比对解决方案
 - 在正在使用的数据库直接审计，识别和报告数据的差异性
- 优点:
 - 能快速地减少风险和对数据差异采取行动
 - 加快和简化数据比较
 - 不中断业务系统
 - 加强决策和报告的可信度
- 关键特性:
 - 支持大数据量比对
 - 当数据库在线的时候运行比对
 - 选择性数据比对
 - 灵活的报表机制
 - 支持异构数据库
 - 生成直观比较报告



Goldengate产品的下载

- <http://www.oracle.com/technetwork/middleware/goldengate/downloads/index.html>
- edelivery.oracle.com
- support.oracle.com - 11g的patch即安装包（12c不是）

快速的搭建一个Goldengate复制环境

- OGG11g - linux
- OGG12c - windows

首先检查认证列表！（同时注意APPS应用的适用场景）

建议相同版本OGG之间复制，实在不行，从逻辑上分析：建议源端低版本，目标端可以高版本。

安装步骤描述 – 准备操作系统用户和文件系统

- 准备安装和运行用户（操作系统用户）
 - 建议使用Oracle用户
 - 也可使用新建用户，但需做以下配置
 - 该用户必须加入到oinstall组
 - 该用户使用与oracle相同的profile文件
 - 操作系统必须为该用户开放一定的CPU、内存等资源限制，可以通过ulimit -a查看，建议在/etc/security/limits.conf中设置和oracle用户一致
- 准备安装位置和空间
 - OGG可以安装于任意位置
 - RAC下建议安装在存储阵列，可以在单点失败后切换到其它节点继续运行
 - 分配空间
 - 软件本身所需空间约为200M以下（OGG 12c需要约500M空间，建议1GB）
 - 为缓存数据建议保留相当于1-3天归档日志的存储空间（trial文件）
 - 将安装目录的owner修改为OGG运行用户

- 数据库必须打开归档模式
 - OGG重启时需要读取未提交交易开始时的日志
- 验证未使用并行日志（只针对9i）
 - Sql> show parameter LOG_PARALLELISM。如结果为1，无需修改；否则需要修改为1。
- 如果是OGG12c，需要设置(不管源和目标，只要运行12c的就要设置)：
alter system set enable_goldengate_replication=true;
- 打开数据库最小附加日志
SQL>select supplemental_log_data_min from v\$database;
结果为NO，表示没有打开附加日志。使用以下命令将数据库附加日志打开：
SQL>alter database add supplemental log data;
切换日志以使附加日志生效：
SQL>ALTER SYSTEM ARCHIVE LOG CURRENT;

- 建立OGG数据库用户
 - OGG需要从数据字典读取表结构
 - 如Oracle日志中没有足够信息，则需通过flash back或者直接读取数据库记录（例如，CLOB/BLOB/BINARY等）
- OGG数据库用户权限
 - CREATE USER *goldengate* IDENTIFIED BY *goldengate* DEFAULT TABLESPACE *USERS*;
 - GRANT CONNECT TO *goldengate*;
 - GRANT CREATE SESSION TO *goldengate*;
 - GRANT ALTER SESSION TO *goldengate*;
 - GRANT RESOURCE TO *goldengate*;
 - GRANT SELECT ANY DICTIONARY TO *goldengate*;
 - GRANT SELECT ANY TABLE TO *goldengate*;
 - GRANT FLASHBACK ANY TABLE TO *goldengate*;
 - GRANT ALTER ANY TABLE TO *goldengate*;

- 建立OGG数据库用户
 - 通过sql将数据写入数据库
- OGG目标数据库用户权限
 - CREATE USER *goldengate* IDENTIFIED BY *goldengate* DEFAULT TABLESPACE *USERS*;
 - GRANT ALTER SESSION TO goldengate;
 - GRANT CREATE SESSION TO goldengate;
 - GRANT CONNECT TO goldengate;
 - GRANT RESOURCE TO goldengate;
 - GRANT SELECT ANY DICTIONARY TO goldengate;
 - GRANT SELECT ANY TABLE TO goldengate;
 - GRANT FLASHBACK ANY TABLE TO goldengate;
 - GRANT INSERT ANY TABLE TO goldengate;
 - GRANT UPDATE ANY TABLE TO goldengate;
 - GRANT DELETE ANY TABLE TO goldengate;

安装步骤描述 – 软件安装（源和目标）

- 上传软件包到安装目录
- 以oracle用户登录
- 使用`unzip XXX.zip`解压缩；
- 使用`tar xvf XXX.tar`解开tar包；
- 在goldengate目录下，执行`./ggsci`进入命令行界面，左侧提示出现GGSCI (myhost) 1>，表示进入成功
如果ggsci进入出错，设置一下环境变量：
`export LD_LIBRARY_PATH=$ORACLE_HOME/lib`
- 在ggsci环境下创建GoldenGate子目录
`GGSCI>create subdirs`

该命令会在OGG安装目录下建立若干子目录，其中几个主要目录如下所示：

dirchk：用于存放各个进程的检查点

dirdat：用于存放数据队列文件

dirprm：用于存放各进程参数文件

dirrpt：用于存放各进程报告

dirpcs：存放各个正在运行的进程信息

安装步骤描述 – Mgr进程配置（源和目标）

- 以oracle用户登录，在goldengate目录下，执行./ggsci进入命令行界面
- 为mgr进程配置参数
GGSCI>edit param mgr
会打开一个编辑窗口，在其中输入以下内容然后保存退出。
port 7809
--DYNAMICPORTLIST 7820-7830, 7833, 7835
--autostart er *
--autorestart er *, retries 5, waitminutes 3
--purgeoldextracts /u01/oracle/goldengate/dirdat/*,usecheckpoints, minkeepdays 7
- 启动mgr进程
GGSCI>start mgr

说明：参数前面加入两个中划线表示该参数被注释掉

安装步骤描述 - 添加表级附加日志（源）

- 以oracle用户进入goldengate目录下执行./ggsci命令进入命令行界面；
- 在GGSCI命令行中登录数据库，为所有要复制的数据表添加trandata
ggsci>dblogin userid *goldengate*, password *goldengate*
ggsci>add trandata myschema.mytable1

.....

或者

ggsci>add trandata myschema.*

结尾无逗号，注意空格位置

安装步骤描述 – 添加抽取进程（源）

- 添加exta进程，从数据库日志中抓取有效数据

```
GGSCI> add ext exta, tranlog, begin now
```

exta为进程名，一般为ext开头表示是extract进程，后面可以加1-2位字符标识；

tranlog表示要抽取数据库日志；

begin now表示从当前时间点开始抽取

安装步骤描述 – 抽取进程直接到目标端（源）

- 添加远程队列ra，exta进程捕捉到的有效数据直接写入远程队列 **一般不会采用这样的方式**

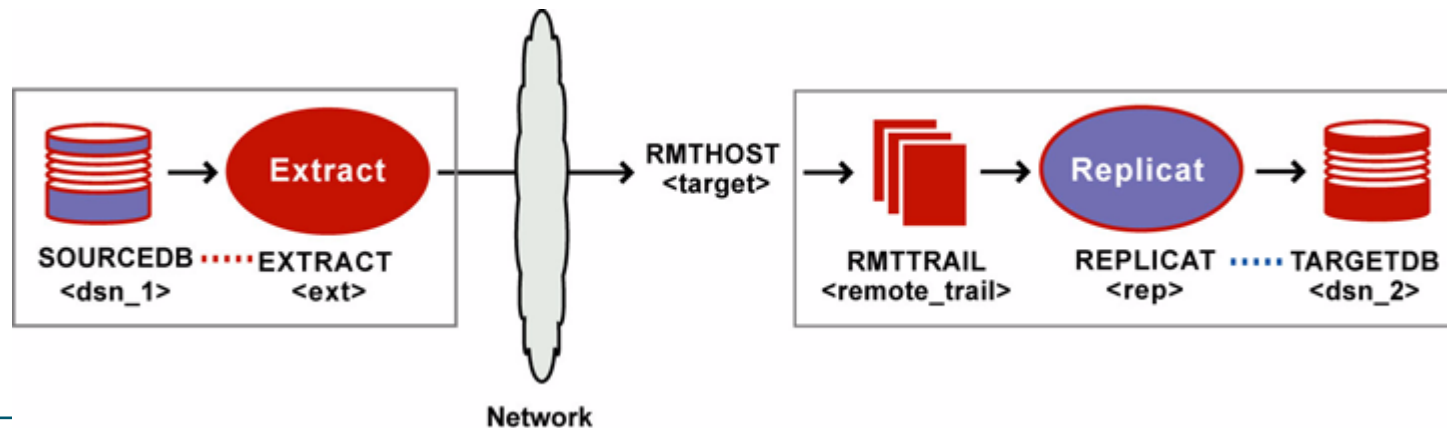
```
GGSCI> add rmttrail /u01/oracle/goldengate/dirdat/ra, ext exta, MEGABYTES 20
```

每一个ext进程都要对应一个本地或者远程队列，指明ext进程抽取的数据要写到何处。

Rmttrail参数指定的位置就是目标队列的位置（注意：不是在源主机上的位置），队列一般以路径加上两个标识如ra来表示，ext后面指定该队列是对应哪个ext进程。

Megabytes表示每个队列文件的大小限制，超过该大小则会滚动写下一个队列文件。

- 注意：如果网络中端，则extract会中止，如果长时间以后再启动则有可能日志已经被删除，因此建议在异地和系统经常要重启或停机的情况下采用添加一个传输进程的方式。



安装步骤描述 - 配置传输进程（源）

- 如果是异地复制或者网络不稳定，建议首先抽取到本地，然后再单独配一个进程传输到目标。

创建完exta进程后，使用本地队列时为exta进程配置队列如下：

```
GGSCI> add exttrail /u01/oracle/goldengate/dirdat/la, ext exta, MEGABYTES 20
```

- 由于此时exta复制到了本地，还需要配置一个传输进程负责将数据原封不动的搬运到目标端：

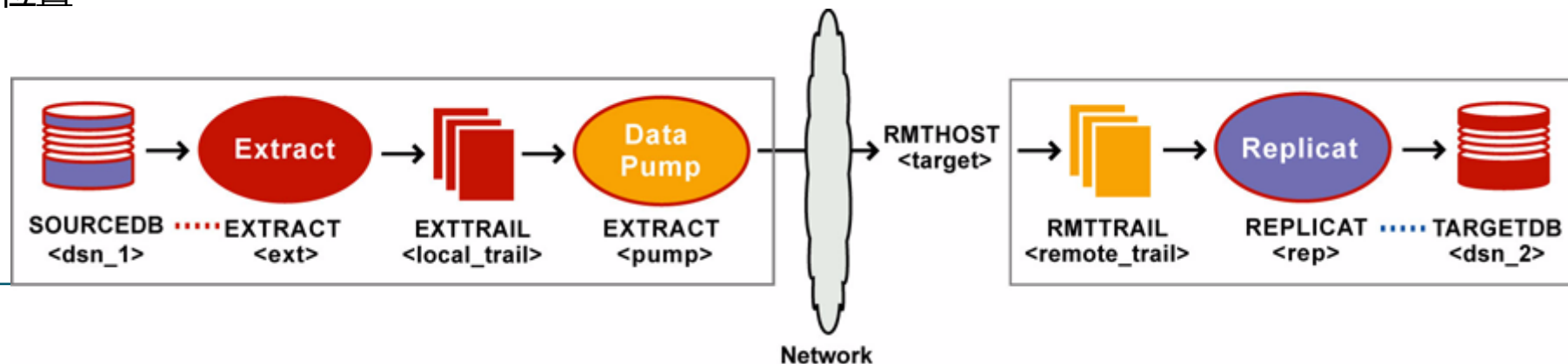
```
GGSCI> add extract dpea, EXTTRAILSOURCE /u01/oracle/goldengate/dirdat/la
```

注意此时exttrail指定的是本地队列位置。

然后为该传输进程配置远程队列：

```
GGSCI> add rmttrail /u01/oracle/goldengate/dirdat/ra, ext dpea, MEGABYTES 20
```

其中队列位置是在目标主机上的位置



安装步骤描述 - 配置抽取进程参数 (源)

- 如果是直接投递到目标端, 不使用data pump (对应前面进程的配置方法, 一般不这样用)

```
GGSCI>edit param exta
EXTRACT exta
setenv(NLS_LANG = AMERICAN_AMERICA.AL32UTF8)
setenv(ORACLE_SID = oggsource)
USERID goldengate, PASSWORD goldengate
rmthost oggtarget, mgrport 7809
rmtrail /u01/oracle/goldengate/dirdat/ra
dynamicresolution
--GETUPDATEBEFORES
--NOCOMPRESSDELETES
table myschema.maytable1;
table myschema.maytable2;
--table myschema.*;
--TABLE hq.acct, COLS (col1, col3);
--TABLE hq.acct, COLSEXCEPT (col4);
```

安装步骤描述 - 配置抽取进程参数（源）

- 如果是使用了 **data pump**，则抽取日志进程 **exta** 参数如下

```
GGSCI>edit param exta
EXTRACT exta
setenv ( NLS_LANG = AMERICAN_AMERICA.ZHS16GBK )
setenv (ORACLE_SID = oggsource)
USERID goldengate, PASSWORD goldengate
EXTTRAIL /u01/oracle/goldengate/dirdat/la
dynamicresolution
--GETUPDATEBEFORES
--NOCOMPRESSDELETES
table myschema.maytable1;
table myschema.maytable2;
--table myschema.*;
--TABLE hq.acct, COLS (col1, col3); //只复制该表的col1和col3列
--TABLE hq.acct, COLSEXCEPT (col4); //复制该表除col4外的所有列
```

相比不使用 **data pump**，其区别就是使用 **exttrail** 代替了 **rmttrail**，也无需再配置 **rmthost**。

安装步骤描述 - 配置抽取进程参数 (源)

■ 如果是使用了data pump, 还需要配置data pump进程dpea

extract dpea

setenv (NLS_LANG = AMERICAN_AMERICA.ZHS16GBK)

passthru

--REPORT AT 01:59

--reportrollover at 02:00

rmthost 99.16.1.11,mgrport 7809, compress

rmtrail /oradata/goldengate/dirdat/ra

dynamicresolution

table myschema.mytable1;

table myschema.mytable2;

TABLE hq.acct; //注意此时不用再过滤cols了。

其中：passthru表示本进程是一个传输进程data pump，无需跟数据库交互，只需要搬运数据即可；
因为data pump要传输数据到目标，所以需要配置rmthost和rmtrail指定目标主机和队列信息

安装步骤描述 - 启动源端进程（源）

■ 启动源端所有进程

GGSCI>start ext *

表示启动所有的extract进程。

也可以使用start直接加进程名，如 start dpea分别启动各个进程。

可以通过info all查看各进程状态，正常启动后应当是running状态。

注：此时应当保证目标端mgr已经启动，否则可能dpea进程无法启动。

安装步骤描述 – 投递进程配置（目标）初始化目标表



- GoldenGate的extract进程CDC工作方式不负责为目标端建立对应表和初始数据
- 可以通过exp/imp、RMAN、init load进行数据初始化
- 如果目标端不需要历史数据，可以只建立表结构。

投递进程配置（目标端）- 添加投递进程

- 添加repa进程，将队列中的数据应用到目标库

```
GGSCI> add rep repa, exttrail /u01/oracle/goldengate/dirdat/ra, nodbcheckpoint
```

repa为进程名，一般为rep开头表示是replicat进程，后面可以加1-2位字符标识，一般与ext进程对应；

exttrail表示要抽取的数据队列，注意是目标端的队列位置；

nodbcheckpoint表示不使用数据库检查点。

投递进程配置（目标端）- 配置投递进程参数

■ 为投递进程配置参数

```
GGSCI>edit param repa
repicat repa
setenv (NLS_LANG = 'SIMPLIFIED CHINESE_CHINA.AL32UTF8')
setenv (ORACLE_SID = GUTJ)
userid goldengate, password *****
--REPORT AT 01:59
--reportrollover at 02:00
repererror default,abend
discardfile /u01/oracle/goldengate/dirrpt/rep.dsc,append, megabytes 10
assumetargetdefs
--allownoopupdates
dynamicresolution
--INSERTALLRECORDS
map myschema.maytable, target myschema.maytable;
map gy25500.F_09B201A_APPEND, target odiu.F_B201A_APPEND, COLMAP ( usedefaults , etltime =
  @GETENV ("GGHEADER", "COMMITTIMESTAMP"));
```


投递进程配置（目标端）-启动目标端进程

■ 启动目标端投递进程

GGSCI>start rep *

表示启动所有的replicat进程。

也可以使用start直接加进程名，如 start repa分别启动各个进程。

可以通过info all查看各进程状态，正常启动后应当是running状态

注意：如果涉及到在线数据初始化过程，语法不一样

安装过程总结一下

- 分段配置
- 相关参数使用合理性
- 相对路径和绝对路径的一致性

问题？

- 在线数据初始化如何做？
- RAC和ASM环境怎么处理？
- 真实的生产环境是如何去实施的？

且听下回讲解！

Goldengate在容灾场景的优势

安全、低带宽要求、无距离限制、数据可靠



- GoldenGate数据同步所需的网络带宽是目前各种技术中最低的，支持跨大洲的广域网数据传输，可以节省大笔费用
 - 复制全库数据时，通常GoldenGate队列文件大小与数据库日志大小的比例为1:4，GoldenGate网络传输压缩比例为1:8，TCP/IP网络传输效率最高为0.7；
 - 根据以上数据估算，以业务高峰期每小时产生50GB数据库日志为例，GoldenGate维持秒级数据同步延迟的带宽要求为：

$$50 * 1024 \text{ (MB/小时)} * 8 \text{ (每字节8位)} / 3600 \text{ (秒/小时)} / 4 \text{ (队列/日志比例)} \\ / 8 \text{ (传输压缩比例)} / 0.7 \text{ (网络传输效率)} = 5.08 \text{ Mbps}$$

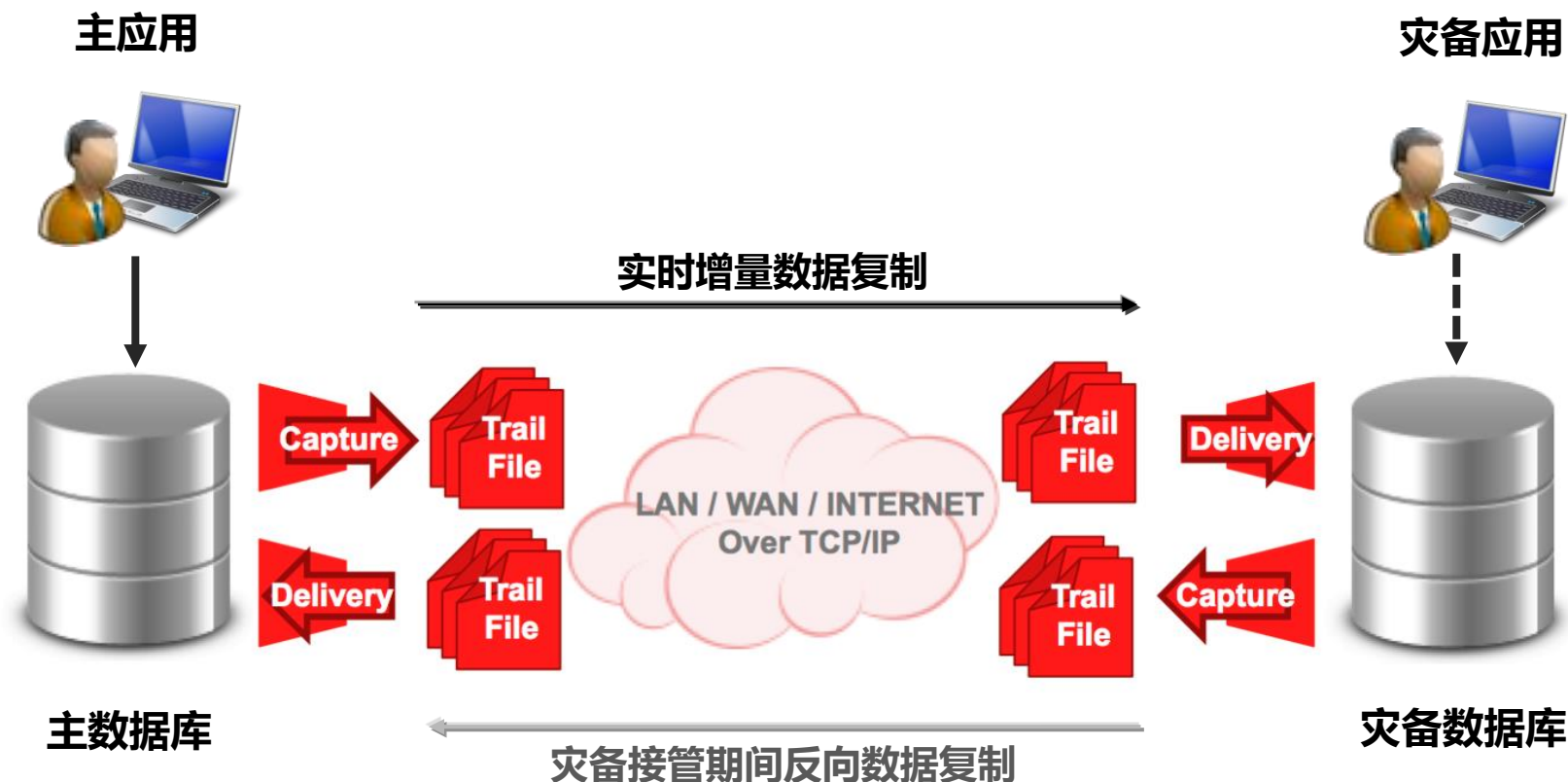
- 128~256位AES加密技术，在广域网链路中有效保护数据安全
- 各种故障或异常后的自动断点恢复和数据一致性保障
- 容灾演练
- RPO和RTO，涉及到方案选型

DATAGURU专业数据分析社区

- 1) 数据库容灾/应急
- 2) 系统迁移/升级/维护
- 3) 查询分载/集中查询
- 4) 双活/多主数据中心
- 5) 实时数据仓库/BI
- 6) 数据集中/分发/同步
- 7) 事件驱动集成
- 8) ...

应用场景一：数据库容灾/应急

低成本、低风险、高可用



方案特点：

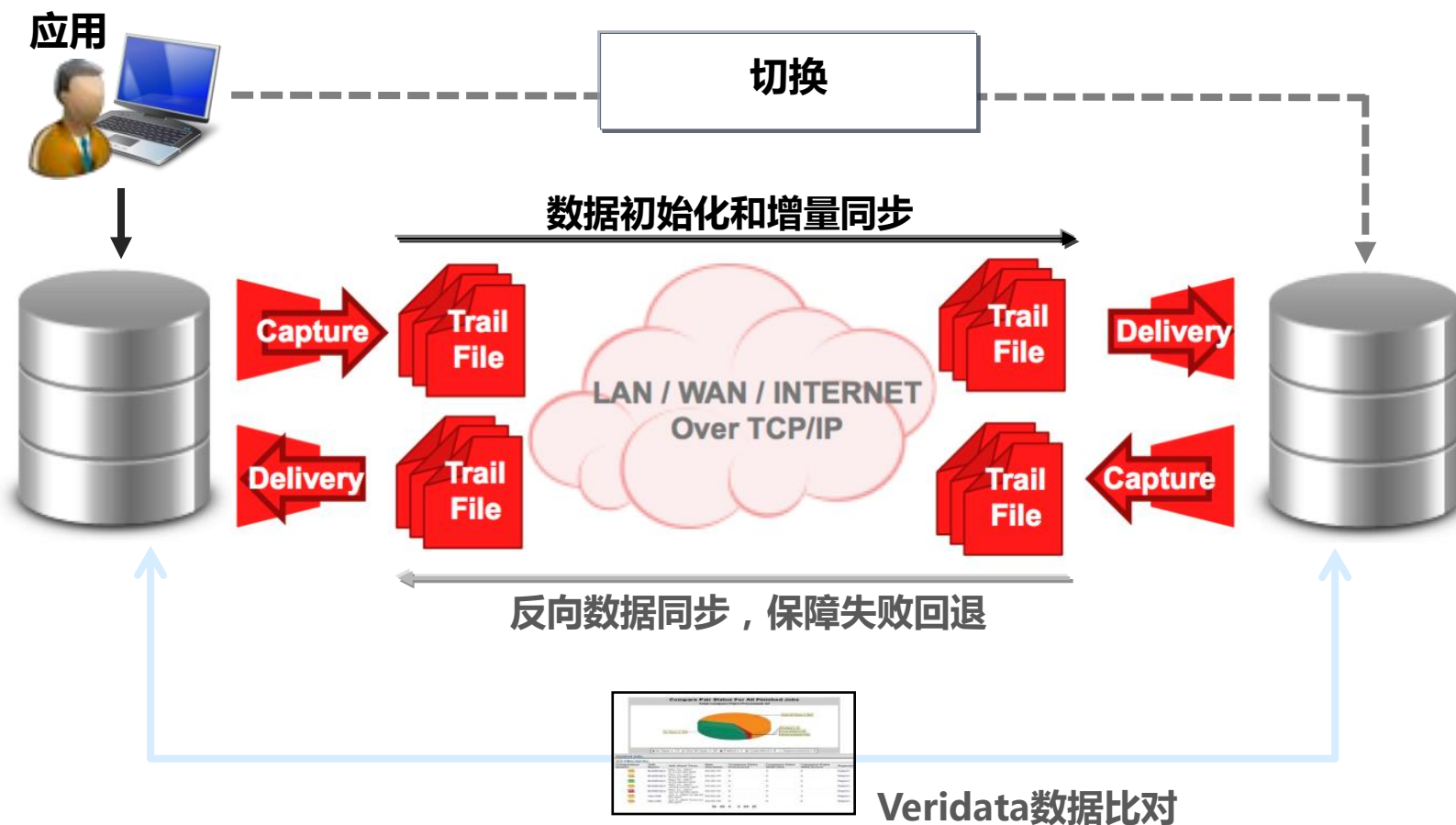
- 支持多种数据库
- 低延迟（**亚秒级RPO**）、低带宽要求，适合远程容灾
- 灾备端Active，实现快速接管（**最小化RTO**），消除切换风险
- 灾备端可灵活选择硬件、OS、数据库版本，支持利旧
- 支持部分核心数据应急、逻辑一备多、误删除保护

场景扩展：

- 灾备库用于查询分载
- 双活数据中心

应用场景二：系统迁移/升级/维护

最小化停机时间



方案特点：

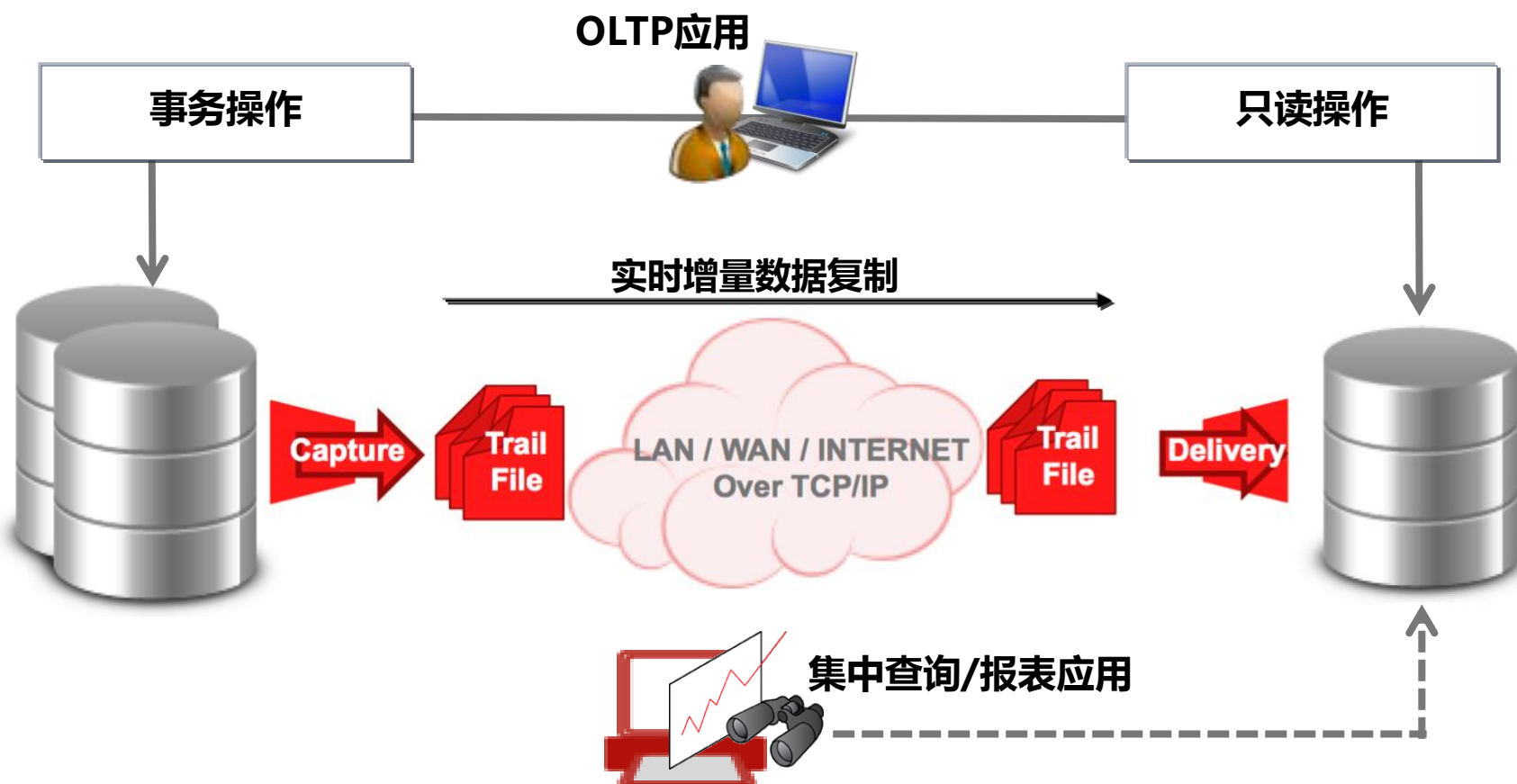
- 业务零中断或近似零中断
- 广泛适用于各种系统迁移/升级/维护需求
- 支持分段迁移
- 支持数据转换/映射
- 反向同步便于失败回退，降低风险
- 基于GoldenGate Veridata 实现数据比对

场景扩展：

- 旧库用于应急/查询分载

应用场景三：查询分载/集中查询

有效提高生产数据库性能

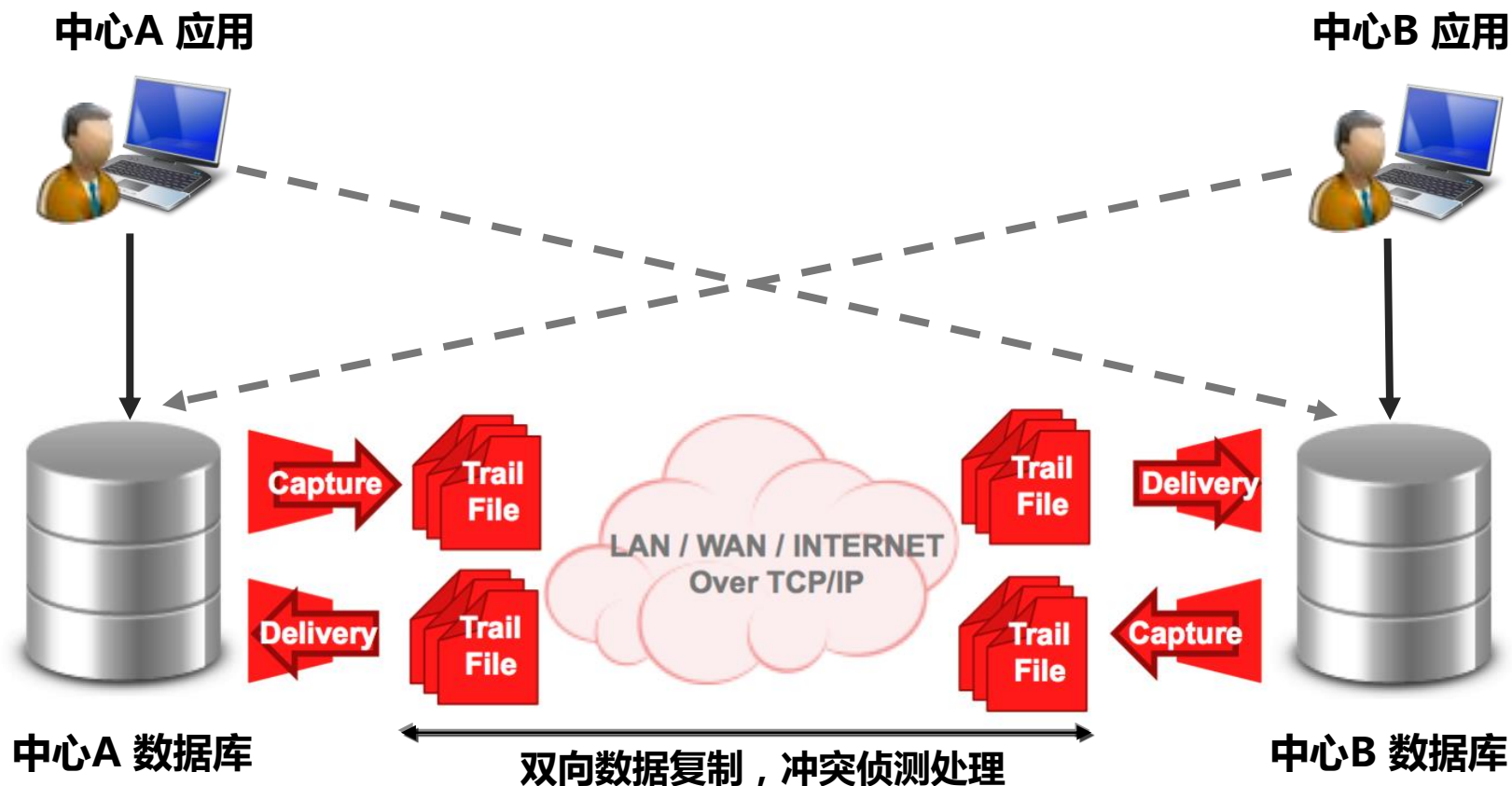


方案特点：

- 实时数据复制，支持查询/只读操作
- 降低/剥离OLTP应用后台数据库的查询负载，提高数据库性能
- 支持异构环境，以低成本、开放式平台分担昂贵生产系统或遗留系统的压力
- 可以集中/合并多个应用的数据，从而实现统一查询/报表

应用场景四：双活/多主数据中心

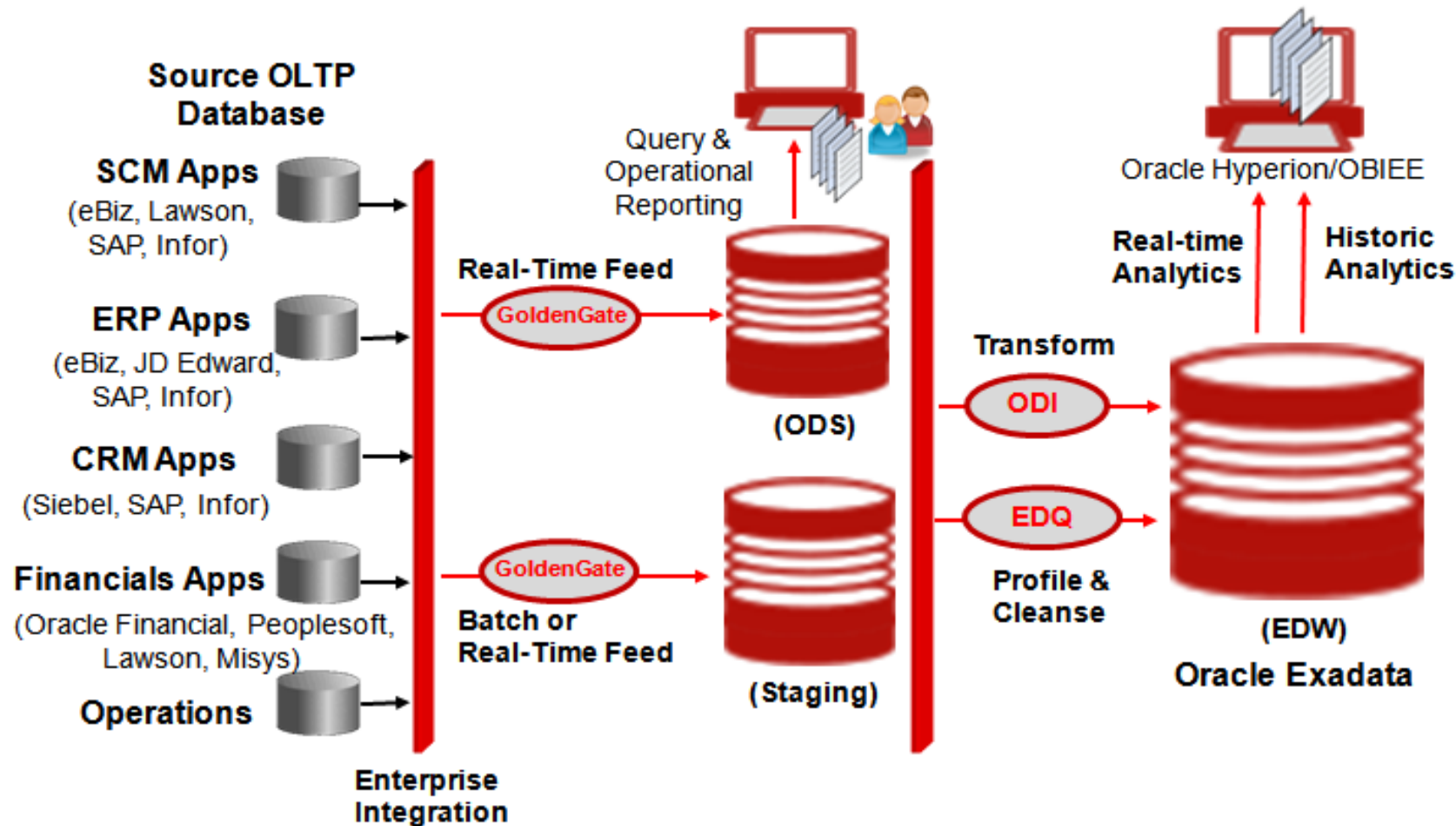
充分的调研才能实现好的效果



方案特点：

- 亚秒级延迟、低带宽要求、无距离限制的双向数据复制
- 分布式计算提高性能
- 任一中心计划/非计划停机时，保障业务的连续可用性
- 支持广泛异构环境
- 尽可能避免数据冲突和及时自动处理冲突是实现双活的关键：GoldenGate提供冲突侦测处理机制，发现和记录数据冲突，按照设定规则自动处理冲突

应用场景五：实时数据仓库/BI



解决方案：

- 提高业务分析能力
- 为决策提供实时数据
- 全天候业务运营，消除批处理时间窗口
- 分析和改善数据质量

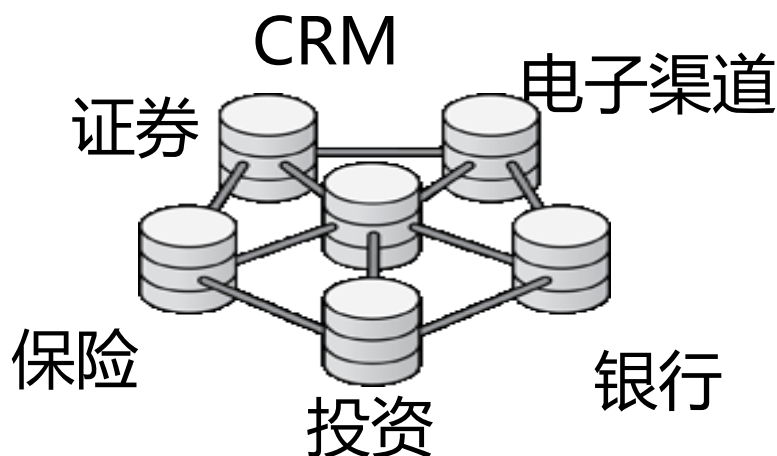
收益：

- 从OLTP系统获取最新的实时数据，更好运营于市场决策
- 提供实时数据，不需要批处理窗口
- 对OLTP的资源 and 性能影响最小
- 高效率的数据移动和转化
- 操作性和分析性BI同时进行

应用场景六：数据集中/分发/同步

业务挑战

- 某金融客户提供保险、银行、证券、投资等全金融服务，大量的业务应用、ERP应用、集团内部应用之间广泛存在实时数据集成需求
- 随着集团业务的扩展，实时数据集成新需求和需求变化频繁
- 对于业务交易繁忙的系统，从应用层进行数据集成很容易造成交易性能下降、集成延迟大、数据一致性难以保障等问题
- 应用后台数据库包括Oracle、Sybase ASE、DB2、MS SQLServer、MySQL多种类型



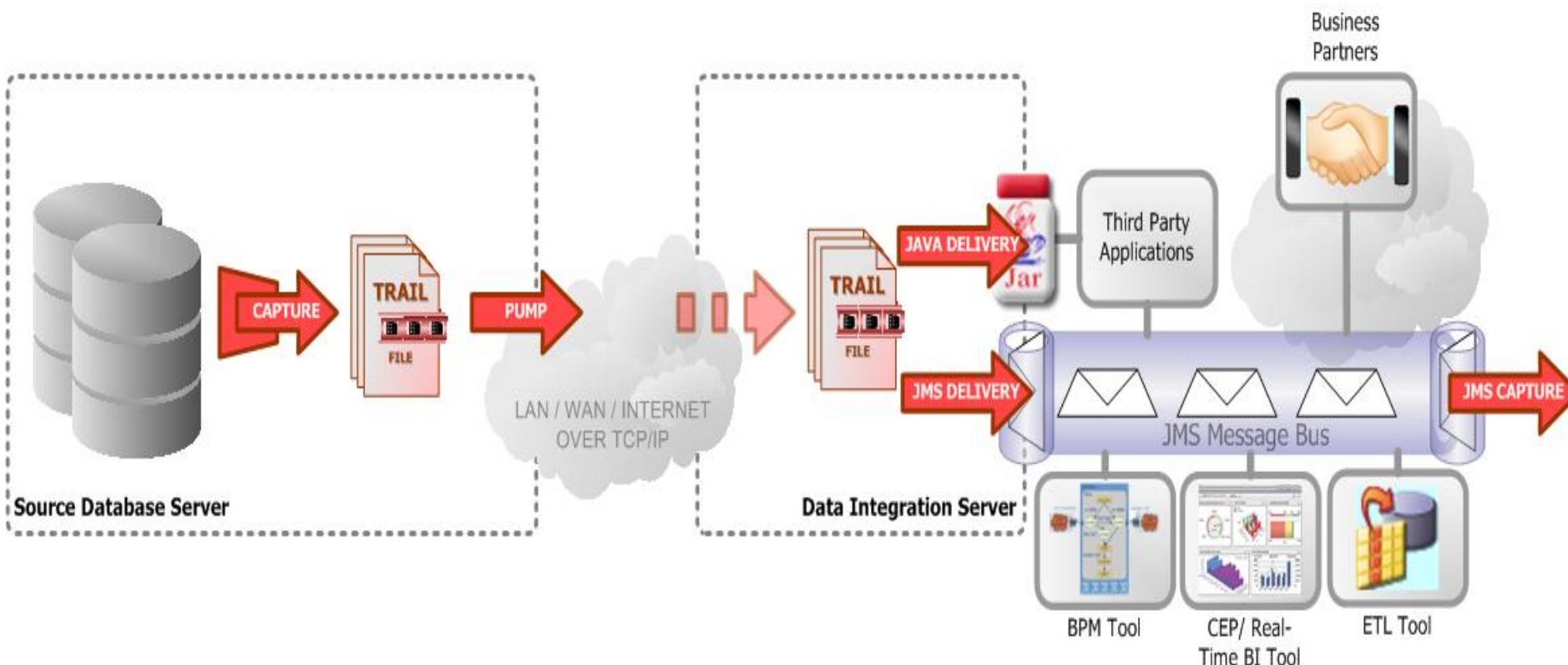
解决方案

- 选择GoldenGate作为集团内实时数据同步平台，从数据库层提供高效可靠的实时数据集成解决方案
- 数据管理部下设立专门团队，负责数据集成需求分析、方案设计、GoldenGate规范制定、实施维护等工作
- 制定了一套使用GoldenGate的规范，包括适宜场景、链路设计、部署实施、组件命名、参数设置、运维管理等
- 开发了GoldenGate配置管理web应用，根据模板自动生成GoldenGate配置文件，并提供配置版本管理，简化和加速数据集成实施

应用场景七：事件驱动集成

优势

- 非侵入式实时捕获源系统事件(实时数据变更)
- 支持广泛的异构事件源
- 以XML格式发布到JMS消息总线(发布-订阅集成)
- 直接调用Java 或C++ API 响应事件



低成本获取高端知识 技术成就梦想



炼数成金

edu.dataguru.cn

DATAGURU专业数据分析社区

Goldengate从入门到精通 讲师 顾铁军

Thanks

FAQ时间