Data Integration Overview

Phill Rizzo
Regional Manager, Data Integration Solutions
phillip.rizzo@oracle.com
Oracle’s Data Integration Solutions
Oracle Products for Data Movement
Comparing How They Work
Oracle Data Integration Solutions

Enterprise-Wide Solutions

Heterogeneous Source Systems

OGG Zero Downtime Migrations & Upgrades

OGG Active/Active Maximum Availability

OGG Query Off-Loading

OGG Real-Time & Batch BI

OGG Data Distribution

ADG New DB/HW/OS/APP

ADG Fully Active Distributed DB

ADG Reporting Database

ADG Data Warehouse

ODI ODS

ODI Global Data Centers

ODI SOA/EDA

Disaster Recovery

BPM

BPM

CEP
Support For Any Type of Data Integration

**Analytical**

OLTP → Query / Report → OLTP → ODS → OLTP → ODS → EDW

**Operational**

OLTP → OLTP → OLTP → OLTP → OLTP → OLTP → OLTP

Old

Heterogeneous

New

Heterogeneous

Heterogeneous

Heterogeneous
Maximizing Availability
## Maximizing Availability

### Increased Availability

<table>
<thead>
<tr>
<th>Component Redundancy</th>
<th>System/Database Site Redundancy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Focus on Data Protection</td>
</tr>
<tr>
<td>MAA Foundation</td>
<td>In Data Center or Across Data Centers</td>
</tr>
<tr>
<td>In Datacenter</td>
<td>Complete one-way physical replication</td>
</tr>
<tr>
<td>Server Failover</td>
<td>Support for up to 30 standby databases</td>
</tr>
<tr>
<td>Disk Striping</td>
<td>Unique corruption protection</td>
</tr>
<tr>
<td>Disk Mirroring</td>
<td>Oracle only with limited support for heterogeneous Platforms</td>
</tr>
<tr>
<td>Continuous Point-in-Time recovery</td>
<td>Read-Only Access (Continuous transmission while recovery is stopped)</td>
</tr>
<tr>
<td>Data Recovery Backups</td>
<td>Read-Only Access while replication is Active</td>
</tr>
<tr>
<td>RAC Rolling Maintenance</td>
<td>Database Rolling upgrade and Standby-first Patching</td>
</tr>
<tr>
<td>RAC, ASM, Flashback RMAN, OSB</td>
<td>Data Guard (Included in Oracle EE)</td>
</tr>
</tbody>
</table>

---

**MAA Foundation**
Focus on Data Protection

**In Datacenter**
In Data Center or Across Data Centers

**Server Failover**
Complete one-way physical replication
Support for up to 30 standby databases
Unique corruption protection
Oracle only with limited support for heterogeneous Platforms

**Disk Striping**
Read-Only Access (Continuous transmission while recovery is stopped)

**Disk Mirroring**
Read-Only Access while replication is Active

**Continuous Point-in-Time recovery**
Read-Write access while replication is active

**Data Recovery Backups**

**RAC Rolling Maintenance**
Database Rolling upgrade and Standby-first Patching

**RAC, ASM, Flashback RMAN, OSB**

**Data Guard** (Included in Oracle EE)

**Active Data Guard**

**GoldenGate**

---

**ORACLE**
Five Steps to Maximize Availability

**Oracle Maximum Availability Architecture**

**ASM, RMAN, OSB**
- Storage failure
- Data recovery
- Backups

**RAC**
- Instance failure
- Server failure
- RAC rolling maintenance
- Performance scale-out
- Consolidation

**Flashback**
- Continuous point-in-time recovery
- Granular repair of logical corruptions
- Transaction
- Table
- Database

**Active Data Guard**
- Database failure
- System failure
- Site failure
- Zero data loss
- Automatic database failover
- Lost-write protection
- Database rolling upgrade
- Offload read-only workload and backups
- Some migrations

**GoldenGate**
- Online platform and application upgrades
- Bi-directional and multi-master replication
- Distribute read-only & read-write workload
- An alternative to physical replication for site protection
- Flexible planned maintenance and heterogeneous migrations
- Zero downtime upgrades and migrations
GoldenGate
Oracle GoldenGate Technology Differentiators

Oracle GoldenGate provides low-impact capture, routing, transformation, and delivery of database transactions across heterogeneous environments in real-time.

Key Differentiators

- **Performance**: Non-intrusive, Low Impact, Outside DB, Sub-second to Seconds Latency
- **Extensible & Flexible**: Open, Modular, Asynchronous Architecture - Heterogeneous Sources & Targets
- **Reliable**: Maintains Transactional Integrity - Resilient Against Interruptions and Failures
Oracle GoldenGate Runtime Architecture

**Capture:** Committed changes are captured (and can be filtered) in Real Time by reading the transaction logs.

**Trail files:** Stages and queues data for routing. Binary file that can be encrypted.

**Pump:** OGG process to distribute data for routing to one or multiple targets.

**Route:** Data can be compressed and encrypted for routing over TCP/IP.

**Delivery:** Applies data with transaction integrity, transforming the data as required.

**Manager Process:** Start – Stop, Process Monitoring & Restart, Trail File Housekeeping
# GoldenGate: Heterogeneity

<table>
<thead>
<tr>
<th>Databases</th>
<th>O/S and Platforms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Oracle GoldenGate Capture and Delivery:</strong></td>
<td><strong>Linux</strong></td>
</tr>
<tr>
<td>- Oracle</td>
<td><strong>Solaris</strong></td>
</tr>
<tr>
<td>- DB2</td>
<td><strong>Windows 2000, 2003, XP</strong></td>
</tr>
<tr>
<td>- Microsoft SQL Server</td>
<td><strong>HP NonStop</strong></td>
</tr>
<tr>
<td>- Sybase ASE</td>
<td><strong>HP-UX</strong></td>
</tr>
<tr>
<td>- Teradata</td>
<td><strong>IBM AIX</strong></td>
</tr>
<tr>
<td>- Enscribe</td>
<td><strong>IBM zSeries</strong></td>
</tr>
<tr>
<td>- SQL/MP</td>
<td><strong>IBM iSeries (AS/400)</strong></td>
</tr>
<tr>
<td>- SQL/MX</td>
<td><strong>zLinux</strong></td>
</tr>
<tr>
<td>- MySQL</td>
<td></td>
</tr>
<tr>
<td>- JMS message queues</td>
<td></td>
</tr>
<tr>
<td><strong>Oracle GoldenGate Delivery:</strong></td>
<td></td>
</tr>
<tr>
<td>- All listed above, plus:</td>
<td></td>
</tr>
<tr>
<td>- TimesTen, IBM System i</td>
<td></td>
</tr>
<tr>
<td>- Netezza, Greenplum, HP Neoview</td>
<td></td>
</tr>
<tr>
<td>- ETL products</td>
<td></td>
</tr>
</tbody>
</table>
GoldenGate: Performance / Reliability

Performance

- Log-based change data capture
  - High volume
  - Low overhead
- Decoupled architecture
  - Multiple capture and delivery processes may be used to scale, but generally not required
  - Possible to split “hot” tables into a separate capture and delivery process
- Filtering and compression
- Transaction Grouping for Delivery
- Record Batching

Reliability

- Decoupled architecture
  - Individual processes can be restarted automatically
  - Tolerance to network outages (configurable)
- Recovery
  - Recovery ensures that no operations are skipped or duplicated after failure of any kind
- Checkpointing
  - In Capture and Delivery
- Transaction Integrity
GoldenGate & Active Data Guard
Active Data Guard and GoldenGate

Comparison

<table>
<thead>
<tr>
<th>Active Data Guard</th>
<th>Golden Gate</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Complete one-way physical replication</td>
<td>• Logical replication: Active-Active HA, one to many, many to one, subset replication, transformations, etc</td>
</tr>
<tr>
<td>• Standby database is an exact copy open read-only</td>
<td>• Target is a different database with the same data, it may have a different physical structure or indexing scheme</td>
</tr>
<tr>
<td>• Backups are interchangeable</td>
<td>• Extremely flexible, source and target are open read-write</td>
</tr>
<tr>
<td>• Integrated automatic database failover</td>
<td>• Trade-off flexibility for more deployment considerations: performance, management, some data type restrictions, application support</td>
</tr>
<tr>
<td>• Dual-purpose standby as test system open read-write</td>
<td>• Standard corruption protection integrated with Oracle Database</td>
</tr>
<tr>
<td>• Simplest to use, supports all applications and workloads</td>
<td>• Logical asynchronous replication</td>
</tr>
<tr>
<td>• Transparent to operate – no data type restrictions</td>
<td>• Flexible options for rolling maintenance</td>
</tr>
<tr>
<td>• Unique protection from silent corruption caused by lost-writes and automatic repair of corrupt data blocks</td>
<td>• Zero downtime migration and application upgrades</td>
</tr>
<tr>
<td>• Synchronous or asynchronous redo transmission</td>
<td>• Extensive cross-platform support</td>
</tr>
<tr>
<td>• Standby-first patching</td>
<td>• Limited cross-platform support</td>
</tr>
<tr>
<td>• Database rolling maintenance and upgrades</td>
<td>• Flexible options for rolling maintenance</td>
</tr>
<tr>
<td>• Limited cross-platform support</td>
<td>• Zero downtime migration and application upgrades</td>
</tr>
</tbody>
</table>

[Oracle Logo]
Oracle Active Data Guard
Real-Time Data Protection and Availability

Note: A DR copy may be multi-purposed for any combination of use cases described.
Oracle GoldenGate
Low-Impact, Real-Time Data Integration & Transactional Replication

- Log-based Changed Data
- New DB/HW/OS/APP
- Zero Downtime Upgrade & Migration
- Fully Active Distributed DB
- Active-Active for High Availability
- Exact Copy of Primary
- Disaster Recovery for Non-Oracle DB
- Reporting Database
- Query & Report Offloading
- Data Warehouse
- Real-time BI, Operational Reporting, MDM
- Global Data Centers
- Data Synchronization within the Enterprise
- ODS
- Event Driven Architecture, SOA
- Message Bus

Legacy Systems
Oracle & Non-Oracle Database(s)
Message Bus
GoldenGate and Active Data Guard

*In Concert*

- For Information Distribution & Consolidation, Application Upgrades & Changes
  - Use **GoldenGate** - heterogeneous, active-active, transformations, subsetting
- For Disaster Recovery / Data Protection / HA
  - Simple Full Oracle Database Protection
    - Use **Active Data Guard**
  - Application desiring flexible HA, active-active, schema changes, platform changes
    - Use **GoldenGate**
- Combine the two for full database protection and information distribution
Oracle Data Integrator
Oracle Data Integrator
Improved Performance, Increased Productivity and Lower TCO

- Declarative Set-based design
- Change Data Capture for Dynamic Updates
- Hot-pluggable Architecture
- Pluggable Knowledge Modules

Legacy Sources
Application Sources
OLTP DB Sources

Any Data Warehouse
Any Planning System
Optimized Data Loading through E-LT
The key to improved performance and reduced costs

Conventional ETL Architecture

- Extract
- Transform
- Load

Next Generation Architecture

- "E-LT"

Benefits:

- E-LT provides flexible architecture for optimized performance

- Leverage Set-based transformations

- Improved performance for loading, no network hop

- Takes advantage of existing hardware
Declarative Design
Improved Developer Productivity and Lower Maintenance Costs

Specify ETL Data Flow Graph

- Developer must define every step of Complex ETL Flow Logic
- Traditional approach requires specialized ETL skills
- And significant development and maintenance efforts

Declarative Set-based Design

- Simplifies the number of steps
- Automatically generates the Data Flow whatever the sources and target DB

Benefits

- Significantly reduce the learning curve
- Shorter implementation times
- Streamline access to non-IT pros
Pluggable Knowledge Modules Architecture

Key Architecture Benefits:

- Ease of maintenance,
- Flexibility, tailored to existing best practices,
- Reduces cost of ownership
Complementary and Used Together
Data Integrator and GoldenGate

- Fastest real-time solution
- Sub-second latency for real-time feeds
- Guarantee delivery eliminates data loss
- Eliminates down-time for migration and upgrades
- Least intrusive to source systems
- Leverage ELT/ETL for complex transformation

Oracle Data Integrator
Oracle GoldenGate
Oracle’s Data Integration Joint Solution
Best-of-Breed and Proven

Technology Differentiators:

Performance
- Lowest latency and highest throughput; non-invasive, low overhead

Extensible & Flexible
- De-coupled architecture; multiple deployment styles; open and extensible
- Maintain transactional integrity; resilient against interruptions and failures

Enterprise
- E-LT architecture for best performance of high data volume transformations
- Knowledge Module architecture for extensibility and flexible connectivity
- SOA-native, integrated with Fusion MW to fit future enterprise architectures
Oracle for Real-Time Business Intelligence
Best Practices Architecture

Oracle GoldenGate

Log-based, Real-time Data Feeds
trans3  trans2  trans1

Oracle Data Integrator

ODS Schema

Oracle Exadata

DW Schema

Source OLTP System

Oracle for Real-Time Business Intelligence

OBI EE Suite Plus

Real-time Analytics

Historic Analytics

Oracle Exadata

Oracle GoldenGate

Log-based, Real-time Data Feeds
trans3  trans2  trans1

Oracle Data Integrator

ODS Schema

Oracle Exadata

DW Schema
Real-Time CDC Integration for Data Warehouse
Best-in-class, integrated solution for Real-Time Data Warehouse

**Traditional ETL + CDC**
- Invasive Capture on OLTP systems using complex Adapters
- Batch window dependency
- Transformations in ETL engine on expensive middle tier servers
- Bulk load to the data warehouse with large nightly/daily batch

**ODI + Oracle GoldenGate**
- Continuous feeds from operational systems
- Maximum availability for sources and DW
- Non-invasive data capture
- Thin middle tier with transformations on the database platform (target)
- Mini-batches throughout the day or bulk processing nightly