ORACLE
FUSION MIDDLEWARE
GOLDENGATE

An Oracle White Paper
July 2013

Ensuring Data Consistency with Oracle GoldenGate Veridata

ORACLE

Executive Overview

The requirement for high data availability and the need to access data at or near 24/7/365 without performance degradation and service interruption has created the need for having redundant distributed copies of the data. However, in today's complex IT environment, maintaining data consistency across distributed copies of data is challenging and the possibility of data discrepancy is an unfortunate reality. If not discovered and addressed, bad data can lead to incorrect decision-making; failed service-level agreements; and ultimately, operational, financial, and legal risk.

Oracle GoldenGate Veridata provides companies with high confidence in achieving data consistency in their backup systems, reporting databases, active-active databases, and other types of redundant data systems. Throughout this paper we will commonly refer to these systems as target databases. The Veridata application enables periodic checks—as frequently as desired—between the source and the target databases without taking either system offline.

Oracle GoldenGate Veridata provides an easy-to-use yet powerful solution for identifying out-of-synch data before it negatively impacts the business. Deployed together with the Oracle GoldenGate real-time data replication product or separately, Oracle GoldenGate Veridata ensures data consistency is maintained across databases.

Challenges in Maintaining Data Consistency

Before we discuss the requirements for the solution that helps manage data consistency across databases, we need to understand the common causes of data discrepancies in an enterprise.

Data discrepancy occurs when the data in the target database deviates from the source database. The extent to which the data deviates depends on various factors, some of which may be intended and others unintended.

Even when using products that replicate data reliably, such as Oracle GoldenGate, there remain potential causes of data discrepancy. If the goal of the target database(s) is to be strictly consistent with the source database, then IT will need to put processes and policy in place to insure this outcome. Some of the potential causes of data discrepancy are described in the following sections.

Migration Errors

Different kinds of migration tools are employed to facilitate the initial load of the target databases before replication can begin. Differences in configuration for handling data by the migration tools and replication products can result in data discrepancies.

For example, a migration tool may use '?' and the replication product may use 'null' when the value of the column is unknown. When performing a migration, there could potentially be open transactions that were unaccounted for, causing missing data on the target.

Differences in Source and Target

Differences in source and target database configuration, for example different encodings, locales, endianess, or database versions can cause subtle discrepancies to happen during migration and replication. For example, incompatible character sets or date/time format and ranges can cause errors to creep into the target database.

Instantiation Errors

Before migration or replication can begin, the target database(s) will need to be instantiated with the correct schema and constraints. Failure to do so will result in the source and target being out of sync.

For example, failing to set a Primary Key/Unique Key can result in duplicate rows. Duplicate rows can be created even if the source database has no duplicate rows as there are no guarantees that the migration job completed without any failures. Other instantiation errors include improperly migrating jobs, scripts, and triggers, which can improperly modify data.

Configuration Errors

Improper and unintended configuration of replication products can cause discrepancies. This type of discrepancy doesn't show up in the replication logs, since from the replication product's perspective it is performing as configured. This may prevent QA tests from detecting the issues as well.

For example, some configured behavior is fine from a replication perspective but not acceptable from a data quality perspective. When using Oracle GoldenGate, for instance, a DBA could use the parameter 'REPLACEBADCHAR' which replaces an unprintable character with something more appropriate. While this permits replication to proceed without issue, it may not be acceptable to an enterprise from a data quality perspective without appropriate approvals.

As another example, Oracle GoldenGate provides options for suppressing the triggers and cascade deletes when applying transactions on the target database. Not using these options or improper usage of these options can result in invalid data.

Gaps in Replication

Although replication is enabled between source and target databases(s) and is working perfectly well, there are instances where data inserted on the source will not be replicated. For example, when data is inserted in bulk, users typically use options in the database (for example NOLOGGING in Oracle database) that would cause the replication system to avoid capturing this data.

Replication Latency

With asynchronous replication, there will be a short lag between changes to the source database and delivery of those changes to the target. Failure to meet the maximum latency requirement, however, can potentially violate service level agreement levels or data compliance requirements.

Infrastructure Failures

Infrastructure errors such as system failure, disk corruption, and network failure can potentially cause data discrepancy between the source and target. After recovering the failed systems, a key task is to ensure that the data consistency between the source and target systems is intact.

While replication or migration products such as Oracle GoldenGate typically have check pointing capabilities, they cannot guarantee the quality of data in the target database when the systems are brought back up after a crash.

User Errors

Often target databases are created to offload query processing from the source database. This enables rich operational reporting without impacting the applications running on the source database. Depending on the technology used, target databases may be open for writing as well as reading. If so, even despite IT policies, users/DBAs can modify data either unintentionally or maliciously.

Application Errors

Applications that use target databases can potentially change data due to faulty logic as well as, during application upgrades. Also, even if replication is working today and data is consistent, IT can potentially develop new applications that use a target database and may potentially modify the data at some point in the future.

Requirements for Managing Data Consistency

Now that we have a good understanding of the challenges of maintaining consistent data across an enterprise, we will discuss some key requirements that not only help to address the above challenges but also fit seamlessly into the IT organization. The technology requirements for managing data consistency between databases are outlined below.

- High speed, low impact data comparisons
- Support for heterogeneous databases.
- Capability for handling large data volumes
- Flexible options for managing data comparisons
- Support for live databases with constantly changing data
- Minimally intrusive

- Zero downtime of source and target systems
- Capability to identify data inconsistency
- Low impact on hardware and network resources
- Detailed and actionable reports about inconsistent data
- Flexible reporting for varying roles and access levels
- Data security
- Easy to use, understand, configure, deploy, and diagnose

Oracle GoldenGate Veridata

Oracle GoldenGate Veridata is an easy to use, high performance and minimally invasive product that helps manage data consistency across an enterprise. It can be used in multiple scenarios to ensure data consistency, including comparing databases with large data volumes or comparing databases with constantly changing data.

Oracle GoldenGate Veridata provides easy configuration through its user interface, offers flexible comparison options, and supports heterogeneous databases. Built on Java technology, it offers globalization and security support when working with data.

Oracle GoldenGate Veridata seamlessly works with products such as Oracle GoldenGate and other similar products by complementing their functionality and ensuring data consistency. Please refer to the Oracle GoldenGate Veridata web site for the latest product information.

How Oracle GoldenGate Veridata works

Before we get into the architectural, installation, and configuration details of Veridata, we will briefly describe how Veridata works.

Unlike an all or nothing approach, Veridata provides the capability of picking what to compare as well as offering flexibility on how to compare, so that only pertinent data is processed and relevant differences (also known as out-of-sync rows) are highlighted.

In the **initial comparison (or row hash) step**, rows are retrieved from the source and target tables with a query. If the source and target databases are of different types, the columns are converted to a standardized data type format for accurate comparison.

By default, Veridata compares rows by comparing all columns of the primary key, value-for-value, and by using a hash value for all non-key columns. The unique digital signature that is used to calculate the hash value shrinks the data to be transferred over the network for the comparison, while still providing an efficient and highly reliable, but not absolute, mechanism for determining whether two rows contain the same or different column values.

For complete assurance in discovering out-of-sync rows, Oracle GoldenGate Veridata can be configured to compare non-key rows column-for-column, instead of using a hash. Full-column comparisons reduce the processing performance in proportion to the number of columns, increases network usage, and are not recommended as best practice

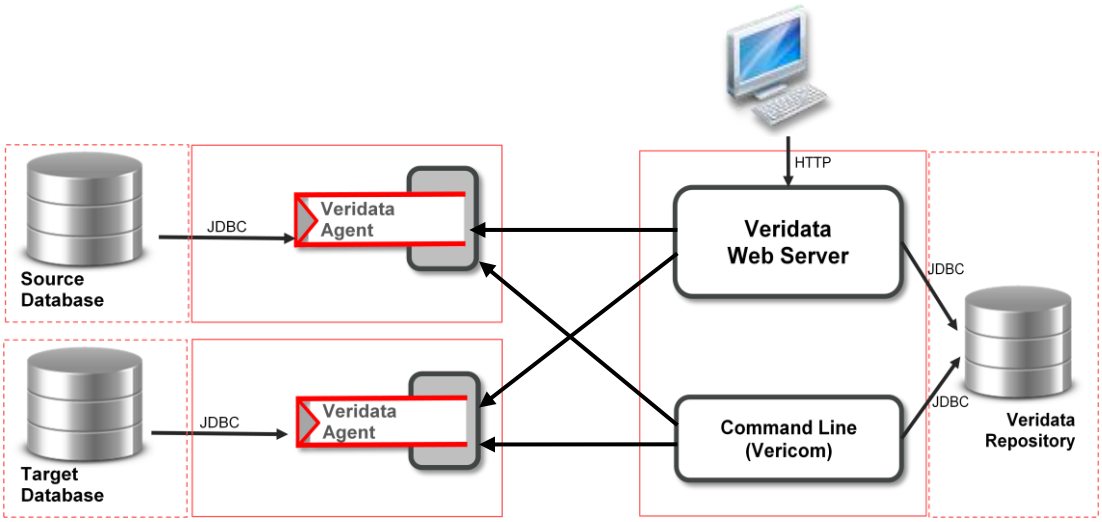
In live replication environments, after Veridata completes the initial comparison, it stores the rows that appear to be out-of-sync in a maybe out-of-sync (MOOS) queue. The cause of the uncertainty is that replication is working concurrently with comparison so the discrepancies may be the result of in-flight transactions captured on the source but not yet applied on the target.

In the **confirmation step also known as the confirm-out-of-sync (COOS) step**, Veridata ensures accurate results by confirming row status in a changing environment. By default, confirmation processing occurs in a thread that is parallel to the initial comparison step, but the confirmation of each row waits until after a specified replication latency threshold has expired. At the end of this step, rows in the MOOS queue are evaluated as follows:

- In-flight - Veridata was unable to confirm that the rows were out-of-sync as there was an update on the row after the initial comparison step.
- In-Sync – Veridata is able to confirm the rows are in-sync.
- Persistently out-of-sync – Veridata is able to confirm the row continues to be out-of-sync after the initial comparison step.

Upon completion of a job, the comparison reports and the out-of-sync reports can be viewed by using Oracle GoldenGate Veridata Web interface or by viewing the files directly.

Oracle GoldenGate Veridata Architecture



The above architecture diagram illustrates a typical architectural set up of the various Veridata components. Arrows indicate from where communication is initiated and all communications are 2 way. Dotted lines indicate that Veridata components can exist on the same machine or separate machine hosting the database.

Veridata Components

Veridata Server

Starting with version 11.2.1, the Veridata Server component is co-located with the Web Server, so we will refer to the pair as the Veridata Server. Using a web browser, users can connect to the Veridata server and configure Veridata jobs and their associated objects. The configuration is stored in a repository that resides in a database. After configuring the requisite Veridata meta-data, users can initiate comparisons and review reports.

Vericom

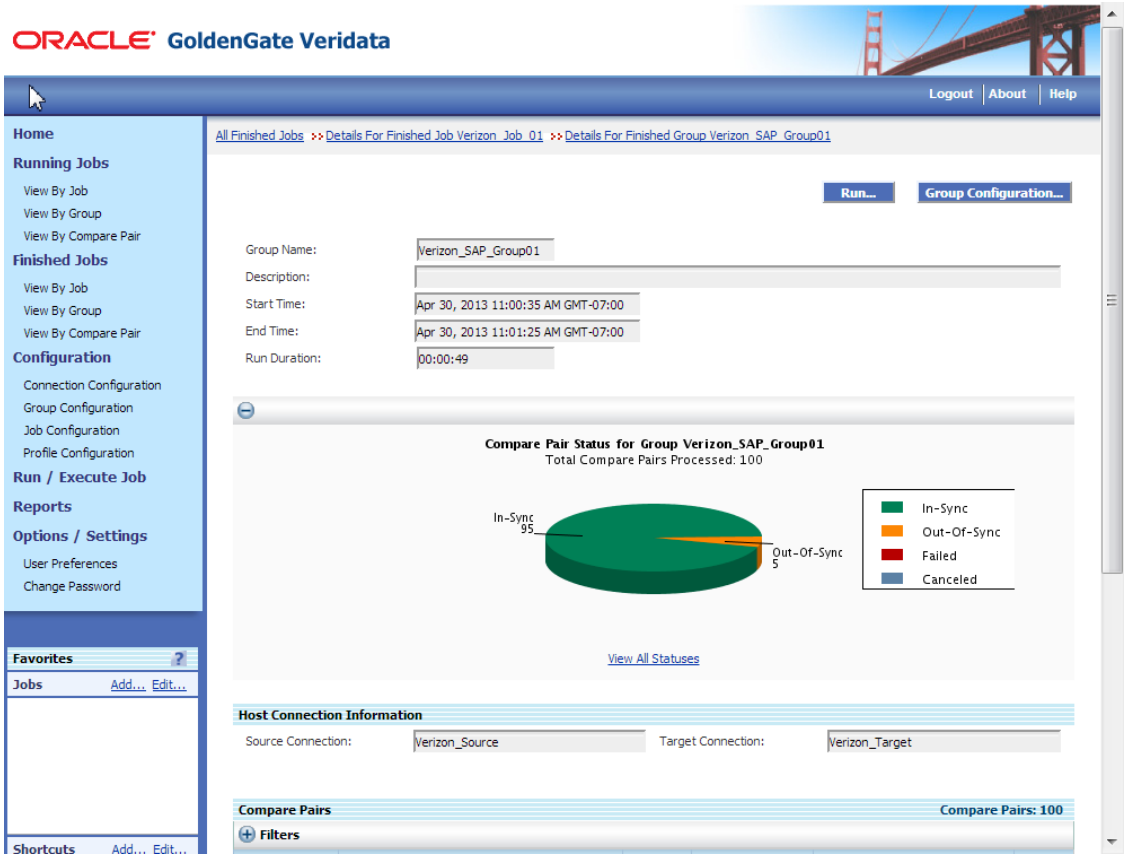
Vericom is a standalone Java program that is used to execute Veridata jobs. Vericom uses the configuration information from the repository to get the details about the Veridata job. Vericom can be invoked from the command line with appropriate options. Vericom can also be used to schedule comparison jobs. Veridata Server and Vericom are typically installed on a machine different from the source and target database.

Veridata Agents

Veridata agents are typically co-located on the machine where the source and target database instances are running. Agents are configured using a configuration file and are responsible for fetching rows from the database and forwarding them to the Veridata server or the Vericom process. Starting in 11.2.1 only Java based agents are supported. The C agent is only supported for NonStop SQL/MP and Enscribe databases running on the NonStop platform.

Veridata Web User Interface

The Veridata web user interface is used to configure Veridata artifacts, start Veridata Jobs and view reports. The following screen-shot illustrates the Veridata web user interface.



Customer Example

Thomson Reuters is the world's leading provider of intelligent information for businesses and professionals. The company combines industry expertise with innovative technology to deliver information—including news, legal briefings, and analyses—to decision makers in the financial, legal, tax and accounting, science, and media markets.

Challenges

The company faced several challenges in data management including ensuring data integrity for the company's flagship legal research products. These legal research products run on a multimaster, multidatabase configuration. With the goal to provide highly available, reliable services to assure user confidence, Thomson Reuters had to ensure that system users, including attorneys, paralegals, law clerks, and judges, have the most up-to-date and accurate information on legal cases and precedents.

Thomson Reuters had to use a solution that would identify discrepancies and resynch data between various databases without downtime to ensure a consistent user experience, and avoid revenue loss associated with compromised service levels.

Solution

Thomson Reuters deployed Oracle GoldenGate Veridata 11*g* and gained the ability to quickly and automatically identify when data is out of synch without interrupting database availability. The company eliminated the need to run complicated queries and spend significant staff resources to identify affected data—saving weeks of effort in some cases.

The solution enabled them to minimized downtime and reduced mean time to recovery when out-of-sync situations occur in the company's Oracle Database environments. Oracle GoldenGate Veridata facilitated the process of keeping databases online and in-sync, which is critical to ensuring service levels and user trust. Thomson Reuters' IT team extended Oracle GoldenGate Veridata's value by creating a PL/SQL procedure that parses Veridata, out-of-sync XML report files; targets out-of-sync rows; and automatically repairs the source—saving further time and resources.

Benefits

By using Oracle GoldenGate Veridata, Thomson Reuters eliminated the need to manually copy data from site to site and take down sites for extended periods of time to re-instantiate one site from another. Thus the company improved system availability, reduced risk of errors, and improved customer satisfaction.

Thomson Reuters leveraged the solution to efficiently detect possible errors and data omissions as the company migrates from third-party databases, including IBM DB2, to Oracle Database 11*g*, streamlining the process and avoiding potential data loss.

Thomson Reuters saved several hundred hours of downtime over a six-month period, thanks to the extended solution's advanced synching capabilities that eliminated the need to re-instantiate the databases—which process more than 72 million lines of data nightly.

Finally, the solution also improved their ability to health check databases in production and testing environments before placing them back in service after incident resolution.

Conclusion

In today's complex IT environments, the possibility of data discrepancies is an unfortunate reality. If not discovered and addressed, inconsistent data can lead to inaccurate decision-making; failed service-level agreements; and ultimately, operational, financial, and legal risk.

Oracle GoldenGate Veridata provides an easy-to-use yet powerful solution for identifying out-of-synch data before it negatively impacts the business. Deployed together, Oracle GoldenGate and Oracle GoldenGate Veridata enable real-time data integration and continuous availability solutions with validated data consistency.

Ensuring Data Consistency with Oracle GoldenGate Veridata
July 2013

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200

oracle.com

Oracle is committed to developing practices and products that help protect the environment

Hardware and Software, Engineered to Work Together