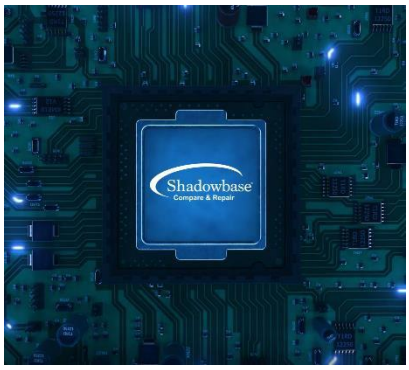
HPE Shadowbase Solution Brief

HPE Shadowbase Database Compare and Repair Technology

HPE Shadowbase Compare: Validate Whether Two Databases Match



HPE Shadowbase Compare is a utility that compares two HPE NonStop databases (Enscribe and/or SQL) and reports whether or not they match. Mismatches, called discrepancies, may occur for example, if either a disaster corrupts one of the databases, or a user misconfigures the application or replication environment.

Applications may remain active on both the source and target, even for updating, while the comparison takes place. Shadowbase Compare validatesthat a *target* database matches its *source* database, and may be used to helpsatisfy regulatory and audit requirements for data validation and verification.

Shadowbase Compare can compare both TMF-audited and non-audited files and tables. It can perform the comparison while *any* data replication product is replicating between the databases, for exampleto validate an online business continuity replication configuration. It is designed for comparing like-to-like environments, such as HPE NonStop to HPE NonStop source/target replication environments (e.g., for Enscribe-to-Enscribe or SQL-to- SQL business continuity). The files or tables can be located on the same system, or when on separate systems, Shadowbase Compare works with both TCP/IP and Expand connectivity.

HPE Shadowbase Repair: Correct Table Differences

HPE Shadowbase Repair is a function in the HPE Shadowbase Compare for SQL NonStop Utility to assist with correcting or repairing SQL/MP or SQL/MX database discrepancies between two tables. Shadowbase Repair works on any two tables in these databases, provided they are both of the same type (MP or MX). It does not require a Shadowbase replication configuration and can be used in any environment where a source and target table need to be verified. Repair is particularly useful in business continuity data replication environments to provide confidence (or certification) that the source and target data matches, and to repair the discrepancies when it does not.

HPE Shadowbase Compare Architecture

The HPE Shadowbase Compare utility consists of these products:

- **Shadowbase Compare for Enscribe**
- **Shadowbase Compare for SQL/MP**
- **Shadowbase Compare for SQL/MX**

The basic method of comparison is similar for each product. The source is always assumed to be correct. The contents of the specified target file or table (or a specified subset key range, either primary key or alternate/index) is read and compared against the contents of the specified source file or table. The “polarity” of which file or table is the source versus the target can be switched, and a comparison can be re-run if desired. See Figures 1 and 2 for sample configurations.

Shadowbase Compare then reports on the differences between the source and target, including records or rows in the source but not in the target, records or rows in the target but not in the source, and/or records or rows in both where keys match but the data contents do not. The results are summarized and optionally reported in various report output modes, from summary reports of difference counts to detailed reports of the actual field and column differences. One set of specific files or tables is individually compared for each comparison run, and many comparisons can be run in parallel against the same table set (such as along partition breaks).

The entire source record or row can be compared against the entire target record or row in order to determine and displayall discrepancies down to the byte level. For large tables that have small differences, Shadowbase Compare for SQL can compress many rows into a checksum and only compare the checksum, which will reduce overall network traffic and dramatically speed the comparison.

For each of the HPE Shadowbase Compare products, the source and the target environments should be exactly the same, or at least very similar to each other. Small differences, for example in the number of SQL columns to be compared or their data types, are allowed. Shadowbase Compare does not currently compare heterogeneous environments (such as when the databases are from different vendors), although this feature is under consideration for a future product release. If heterogeneous comparison technology is of interest, please contact us with the specific source and target platforms, databases, and versions.

The source and target file or table can have data replication actively running at the time of the comparison. Comparisons made when a record or row is being updated are flagged as potential mismatches. Shadowbase Compare has various features, including *iteration processing*, to eliminate or reduce “false positives” (records or rows that mismatch due to replication latency) in the final report.

Figure 1 depicts a sample architecture for a source side comparison. In this case, HPE Shadowbase Compare is configured to run on the source system (the source table is considered to be correct).

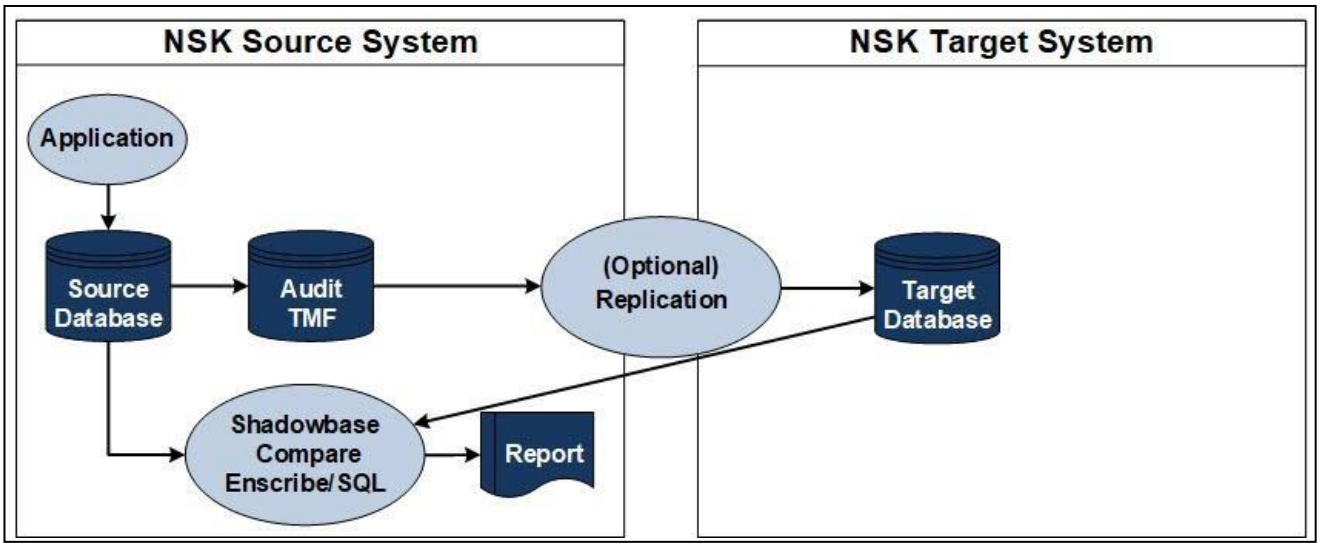


Figure 1 – HPE Shadowbase Compare Configuration – Source-side Compare

Figure 2 depicts a sample architecture for a target side comparison. In this case, HPE Shadowbase Compare is configured to run on the target system (the target table is considered to be correct).

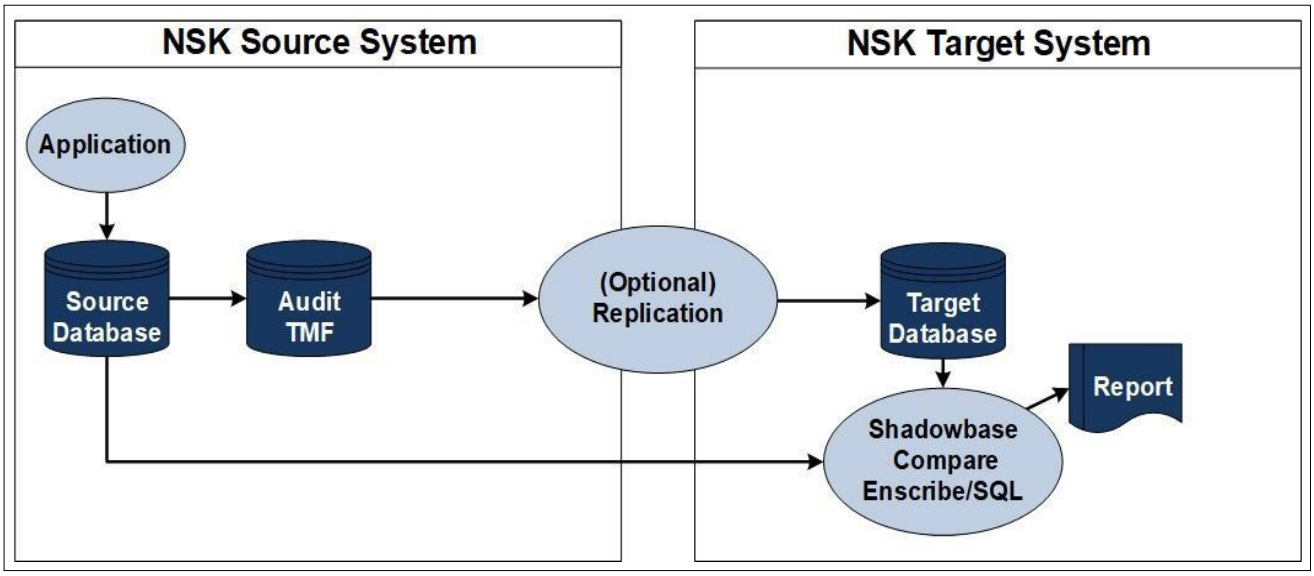


Figure 2 – HPE Shadowbase Compare Configuration – Target-side Compare

Compare Summary

The HPE Shadowbase Compare solution consists of three utility products: Shadowbase Compare for Enscribe, Shadowbase Compare for SQL/MP, and Shadowbase Compare for SQL/MX. These products allow for identifying differences between two files or two tables where the source and the target database schemas are exactly the same, or at least very similar to each other. When run in a data replication environment where the source database and/or target database is actively updated, Shadowbase Compare has various features to eliminate or reduce false positives in the final report. Shadowbase Compare is helpful for validating that a target database matches its source, and for satisfying regulatory and audit requirements for data verification.

HPE Shadowbase Repair Architecture

Mission-critical databases are often extremely large, consisting of millions, even billions, or more of data entries. Additionally, these databases are often changing very rapidly and replicated from a source to a target environment to improve availability, support OLAP, or to offload reporting.

It is important to know if there are discrepancies between the source and target environments. HPE Shadowbase Compare reports the discrepancies between two sets of SQL tables, which would be an error-prone and time-consuming task to correct if even only a handful of differences were found. With HPE Shadowbase Repair, the user can automatically and quickly correct these differences with just a few simple commands. The user has full control over selecting which table is considered the *source*, and which table is considered the *target*. Note, changing this frame of reference can be used to control where most of the comparison (and repair) work is performed, and can be set based on user preference, system loading, database security or access, or other factors.

The Shadowbase Compare Ledger Maps Source and Target Value Discrepancies

As shown in Figure 3, HPE Shadowbase Compare also supports using an extract process on each environment (one local to the source database, and one local to the target database), and then performs the compare on the target environment.

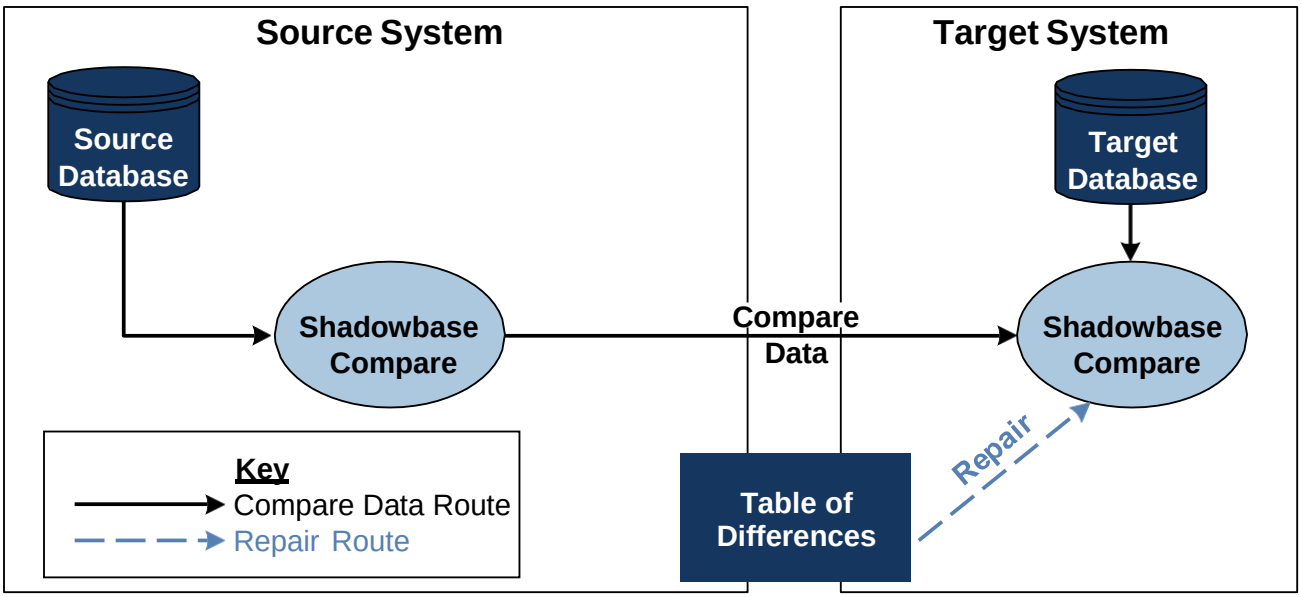


Figure 3 – HPE Shadowbase Compare

During processing, HPE Shadowbase Compare acts as a *ledger*. As it runs, it creates a table of differences (the ledger), which specifies where the differences are located in each table, and what are the value differences (as shown in Figure 3). Although a replication environment need not be present to run Compare and Repair, Repair usesthe frame of reference from the Compare run, noting which file or table was designated the source (and is assumed to be correct) and which is the target (that must be changed to match). Compare/Repair thus tracks the specific changes to be made to the target to make it match the *source*, as assumed in the following scenario. For example, data in the source that is not in the target is listed as "missing" from the target; therefore, an INSERT against the target table is required. Data in the target that is not in the source is listed as "extra" data in the target; therefore, a DELETE against the target table is required. Data in both the source and target that have different column values require an UPDATE for the target table. When a discrepancy is found, the user can specify how to repair the differing tables with a conditional action. Repair references the Compare ledger's information when repairing the data.

With Repair, the options for repairing the discrepancies (in order of suggested usage) include:

- **Option 1:** Update the source data with a "NO-OP" update, using the values in the source table (i.e., setting the source data to the same values). This approach relies on a replication engine to replicate the source update to the target to make the target consistent. Since the replication engine enforces serialization of changes at the target, the NO-OP update will be applied in the correct order when it reaches the target.
- **Option 2:** Update the target data directly, using the values in the source table. This approach directly updates the target to match the source.
- **Option 3:** Update the target data with a "NO-OP" update, using the values in the target table. This approach relies on a replication engine running in bi-directional mode to replicate the target update back to the source to make the source consistent.
- **Option 4:** Update the source data directly, using the values in the target table. This approach directly updates the source to match the target.

Option 1: Updating the Source Data with a NO-OP Update

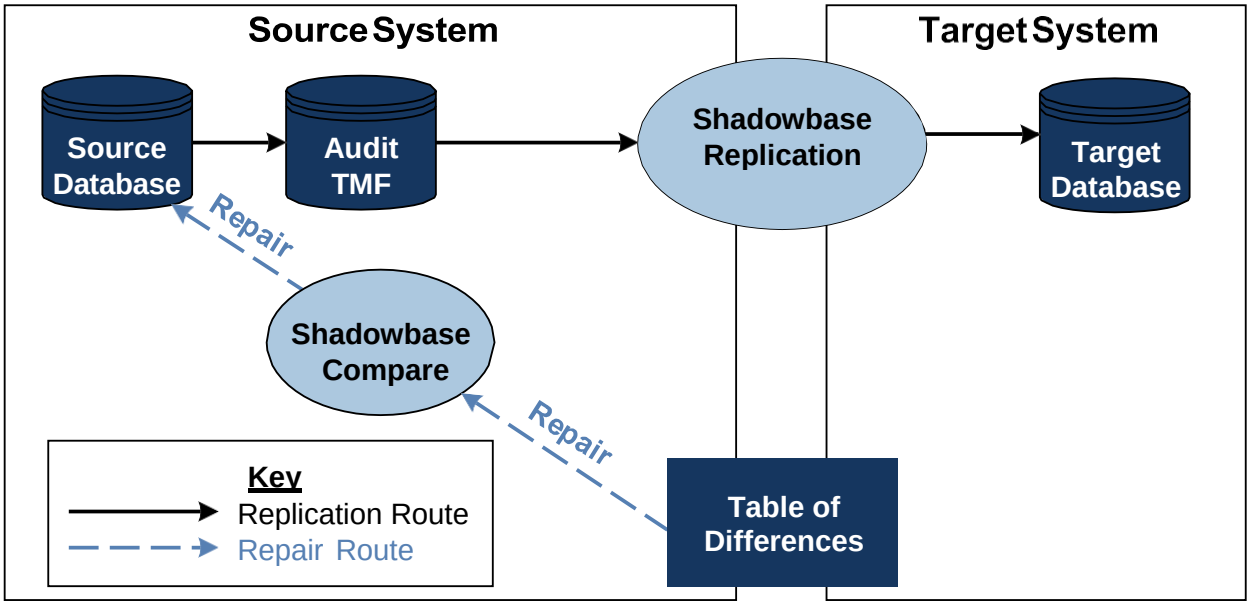


Figure 4 – Updating the Source Data with a NO-OP Update

This approach is suggested when the source and target databases have the same or different format and whendata replication is configured between them. The approach (as shown in Figure 4) assumes a replication product (in this case, Shadowbase software) is properly configured, running, and optionally performing data transformationsin-flight. (Note, a replication engine must be running in order for the replication of events from the source's audit trail to the target.) If a discrepancy is found between the source table and the target table while the Repair command is running, then it simply updates all of the columns of the source table with the already-set values (a "NO-OP update"). By updating the source columns with these values, the changes are captured and are replicated to the target table. This approach is the safest way to conduct the repair when the replication engine is active, and typically should be used if the replication engine is transforming the data in flight.

Option 2: Updating the Target Data Directly

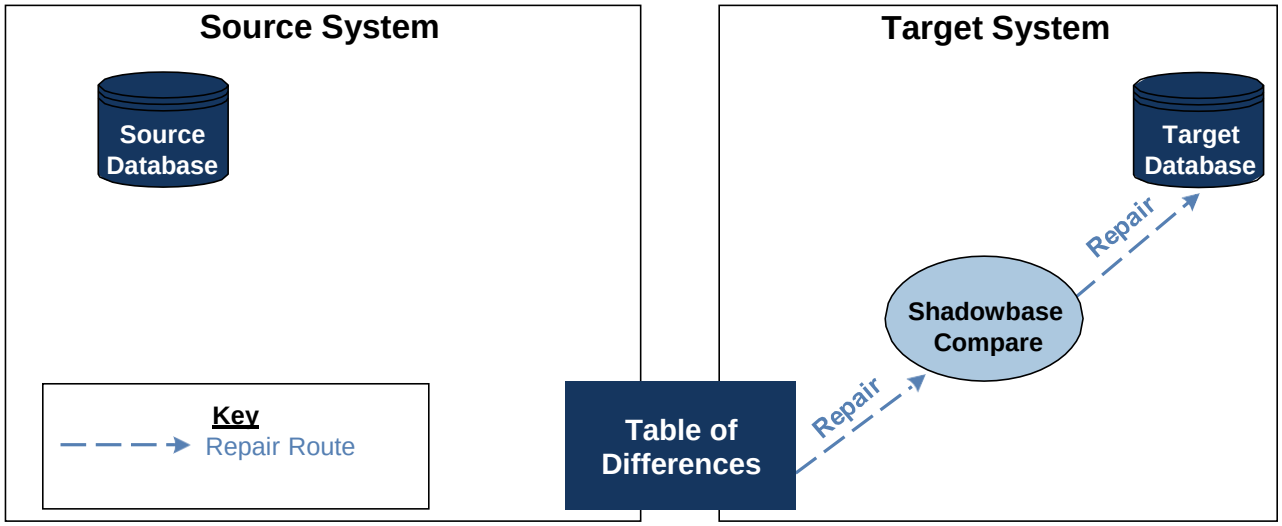


Figure 5 – Updating the Target Data Directly

This approach is suggested when the source and target databases have the same format and when replication is not running. If replication is active, this approach assumes Shadowbase uni-directional replication is properly configured, running, and not performing any form of data transformation, scrubbing, or filtering of the in-flight data, (because when Repair modifies the data, the data avoids this in-flight process).

If a discrepancy is found between the source table and the target table when Repair runs, it updates the current values of selected target columns using the values in the matching source columns (as shown in Figure 5). Optionally, the repair function can retrieve the current value of the source data before updating the target. If necessary, each updated column from source to target can be explicitly specified in order to avoid accidental damage to any additional data that exists in the target (but not the source).

Option 3: Updating the Target Data with a NO-OP Update

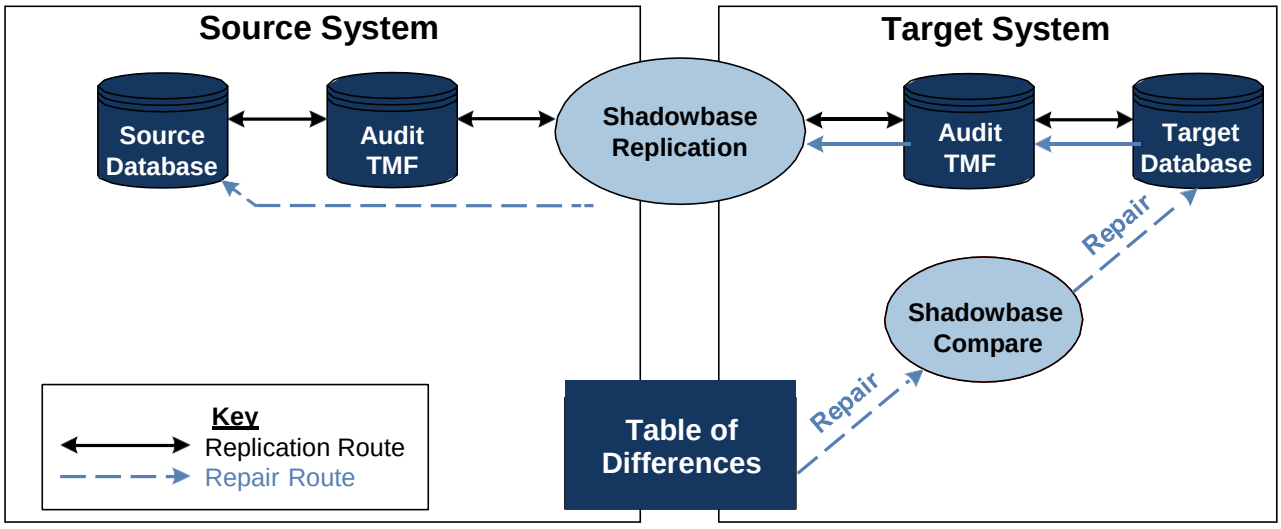


Figure 6 – Updating the Target Data with a NO-OP Update

This approach is similar to the NO-OP case against the source data described above; however, in this case it assumes replication (e.g., via Shadowbase software) is properly configured and running in a bi-directional environment (as shown in Figure 6). It creates a "NO-OP update," which replaces the data in the target table with a copy of itself to create an event in the audit trail and leverage the replication stream from the *target* table back to the *source* table to update the source table.

Option 4: Updating the Source Data Directly

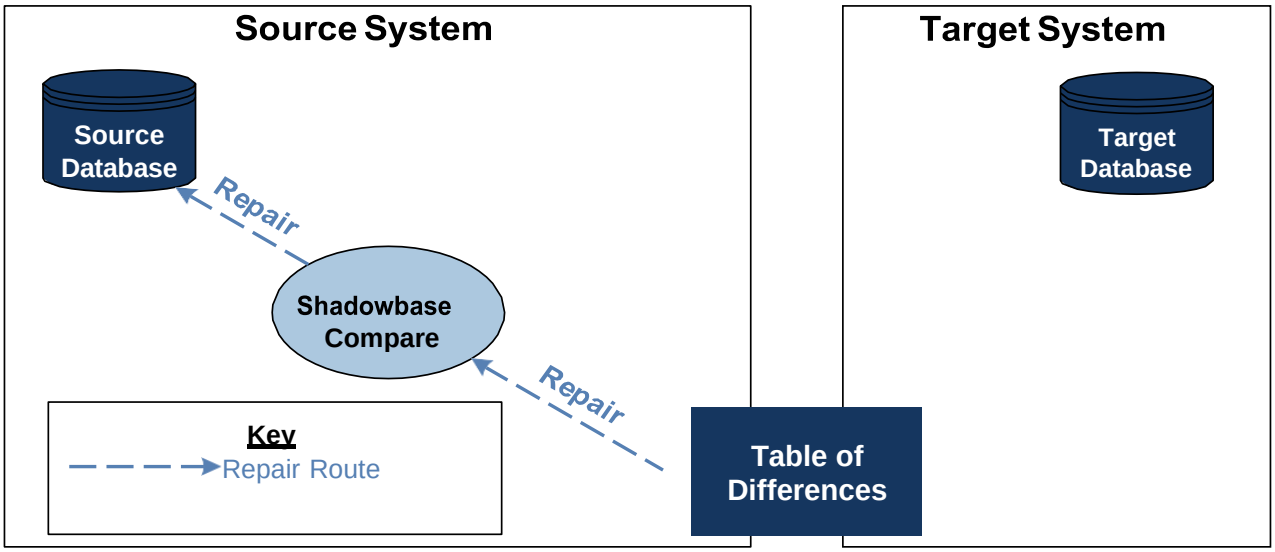


Figure 7 – Updating the Source Data Directly

This approach assumes there is no replication engine running and is symmetrical to "Updating the Target Data Directly." It directly updates the source table, using the values from the target table (as shown in Figure 7). This approach is least common, since the source data is often considered more *current* (consistent, or accurate) than the target data. It empowers the user with flexibility and numerous options for unique scenarios.

Insert and Delete

An insert can be specified if a source or target row is not found, and the missing row can be inserted from the other table. A delete can be specified if a source or target row is found, and the missing row can be deleted from the corresponding table.

Commands

There are three simple commands in order to run Repair. If Compare is run using these commands, then the differences are recorded and logged, and the user can then run Repair to correct these differences.

```
ON DIFF (*) REPAIR DO UPDATE TARGET (*);
ON SOURCE NOT FOUND REPAIR DO
DELETE; ON TARGET NOT FOUND REPAIR DO
INSERT ();
```

In practice, every repair operation should have at least three rules to handle each of these operations: whether toupdate the source or target table in case a discrepancy is found; what to do when a row exists on the target but not the source; and what to do when a row exists on the source but not the target.

Repair Summary

HPE Shadowbase Compare is a utility that finds the discrepancies between the user's source and target tables. The Repair command can then be executed to automatically repair any discrepancies between the tables. Before utilizing the Repair command, take caution to carefully consider how the various Repair options will correct the discrepancies. It is recommended that the user typically configure the commands so that the target table is properly updated to reflect the source table if any differences are found.

Hewlett Packard Enterprise globally sells and supports Shadowbase solutions under the name HPE Shadowbase. For more information, please contact your local HPE Shadowbase representative or visit our website. For additional information, please view our Shadowbase solution videos: Vimeo.com/ShadowbaseSoftware.

Learn more:
shadowbasesoftware.com
hpe.com

Contact us:
Gravic, Inc.
17 General Warren Blvd
Malvern, PA 19355-1245 USA
Tel: +1.610.647.6250
Fax: +1.610.647.7958
Email Sales: shadowbase@gravic.com
Email Support: sbsupport@gravic.com

Please Follow:

Copyright © 2018, 2021, 2022 by Gravic, Inc. Gravic, Shadowbase and Total Replication Solutions are registered trademarks of Gravic, Inc. All other brand and product names are the trademarks or registered trademarks of their respective owners. Specifications subject to change without notice.