# HPE Shadowbase Compare –
Validate that a Source and Target Database Match

A Gravic, Inc. White Paper
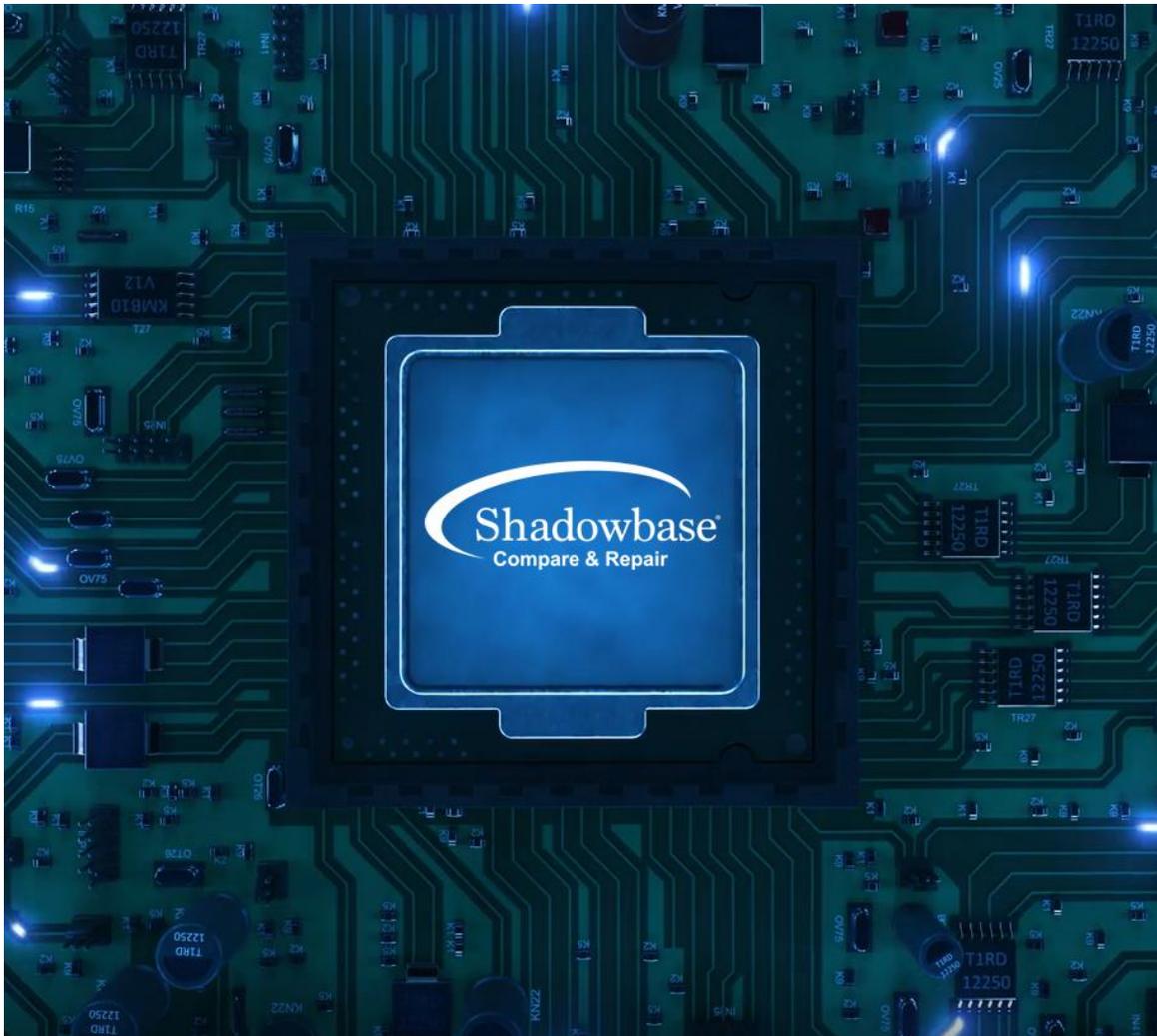
Executive Summary

How does an enterprise maintain up-to-date backup copies of a database? Online data replication is an excellent solution for its numerous advantages. However, it is of little value if the database copies are not consistent.

HPE Shadowbase Compare is a solution that validates two like HPE NonStop databases (Enscribe or SQL) and reports any differences found. Shadowbase Compare is excellent for audit compliance (validating that the source and target data matches), satisfying regulatory requirements, and for confirming data in-flight and at rest is not corrupted.

For example, if a disaster, a ransomware attack, or other such hacker activity corrupts one of the databases, or even if an authorized user misconfigures the application or replication environment, then Shadowbase Compare will highlight these differences. Applications and databases can remain online for updating on both the source and target while the comparison process occurs. *Shadowbase Compare is recommended in every business continuity environment.* Use Shadowbase Compare for absolute certainty that your replication environment is running correctly, and that your backup database is a true and accurate copy of the source database.

If you cannot prove your source and target databases match, you are risking your enterprise's survival. Why take unnecessary risks with data consistency? Use Shadowbase Compare to prove that your backup matches production, preparing for a successful failover at any time.

Table of Contents

Table of Figures

HPE Shadowbase Compare –
Validate that a Source and Target Database Match

The HPE Shadowbase Compare solution validates that a source and target database match. Shadowbase Compare is a utility that compares two HPE NonStop Enscribe or SQL databases and reports any differences between them.

Applications may remain active and online for updating while the comparison runs. Since Shadowbase Compare works while any (or no) data replication product is replicating between the databases, it can also be used to 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 (e.g., for Enscribe-to-Enscribe or SQL-to-SQL business continuity). Shadowbase Compare works across TCP/IP or Expand connected systems.

Summary of Benefits

1. Validate source and target database consistency
2. Prove to auditors and business stakeholders that databases match
3. Correct differences found between two databases
4. Applications can remain active and online, avoiding an outage while Shadowbase Compare runs
5. Works in any HPE NonStop environment, with any or no replication engine
6. Satisfies auditing best practices: replication is independent from Shadowbase Compare
7. Efficiently compares source and target files or tables
8. Numerous configuration parameters are available for summary and detailed reports

Enterprise Survey: To Compare or Not to Compare[1]

Why don't companies compare their databases? Because they:

1. don't think they need to and are convinced the source and target match (even without proof).
2. don't know how or have an adequate tool.
3. don't have the time, personnel, or project funding.
4. don't have management or key stakeholder support.
5. believe the comparison requires taking an application outage.

---

[1]These statistics were collected from a May, 2019 GoToWebinar poll conducted by Gravic, Inc.

In a Webinar, We Asked Participants:

1. Have you ever encountered a situation where two tables that are supposed to be the same do not match?

    A. Yes – 60%
    B. No – 40%

2. If your boss asked you to prove your Business Continuity solution was working, could you?

    A. Yes – 62%
    B. No – 38%

3. How would you correct a situation where the source and target didn't match?

    A. Use a third-party tool – 22%
    B. Write a script – 33%
    C. Don't know/other – 44%
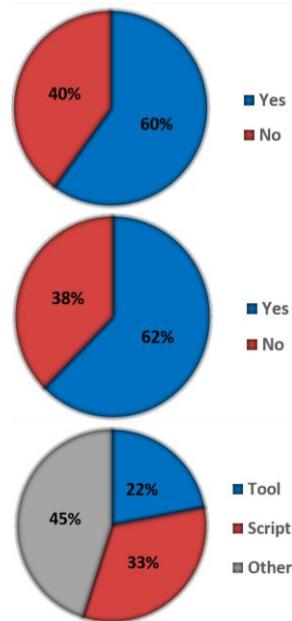
HPE Shadowbase Compare Architecture

HPE Shadowbase Compare is an easy-to-use, flexible, and powerful comparison tool. Use it while production continues: with open or closed databases, with or without replication. For the comparison, source and target tables can be located on the same, or different, HPE NonStop systems. For example, while replication occurs, use Shadowbase Compare to validate database consistency. Additionally, Shadowbase Compare is more efficient than other tools. It starts a process independent from replication, and then directly reads and compares files or tables in memory.

One set of files or tables (or key ranges) is compared at a time for each comparison run; however, many comparisons can be run in parallel against the same table set. For example, large comparisons can be broken down into smaller ones by using partition breaks. In some cases, the data to be compared can be compressed to lessen the network load.

User Requirements

Shadowbase Compare is easy to use. When running Compare, you should have read access to the source and target files, and only needs general knowledge on:
1. HPE NonStop servers
2. HPE NonStop SQL databases
3. HPE NonStop command line tools (like TACL, SQLCI, or OSS MXCI)

File/Table Requirements

HPE Shadowbase Compare compares the contents of two like HPE NonStop tables (SQL) or files (Enscribe) that are similar in schema structure.[2] Shadowbase Compare for SQL/MP compares two SQL/MP tables; Shadowbase Compare for SQL/MX compares two SQL/MX tables; and Shadowbase Compare for Enscribe compares two Enscribe files[3].

Source-side Comparison

Figure 1 depicts a sample architecture for a source-side comparison. In this case, Shadowbase Compare SQL is configured to run on the source system. In the figure, Shadowbase replication is shown as running, and replicating the changes (from the source system's TMF Audit Trail) to the target system and applying

---

[2] Enscribe files must be the same file type (key-sequenced, relative, or entry-sequenced), record length, and primary keys. SQL tables must have a *similar* structure (some column differences are allowed).

[3] Please note that heterogeneous Compare is not available at this time. Please contact Gravic if you are interested in a heterogeneous Compare solution.

them. Having the Shadowbase replication engine (or any vendor's engine) active at the time of the compare operation is optional, since it handles any replication-induced differences as described below. It accesses and reads the data to be compared from the source file/table and the target file/table, then generates a report.
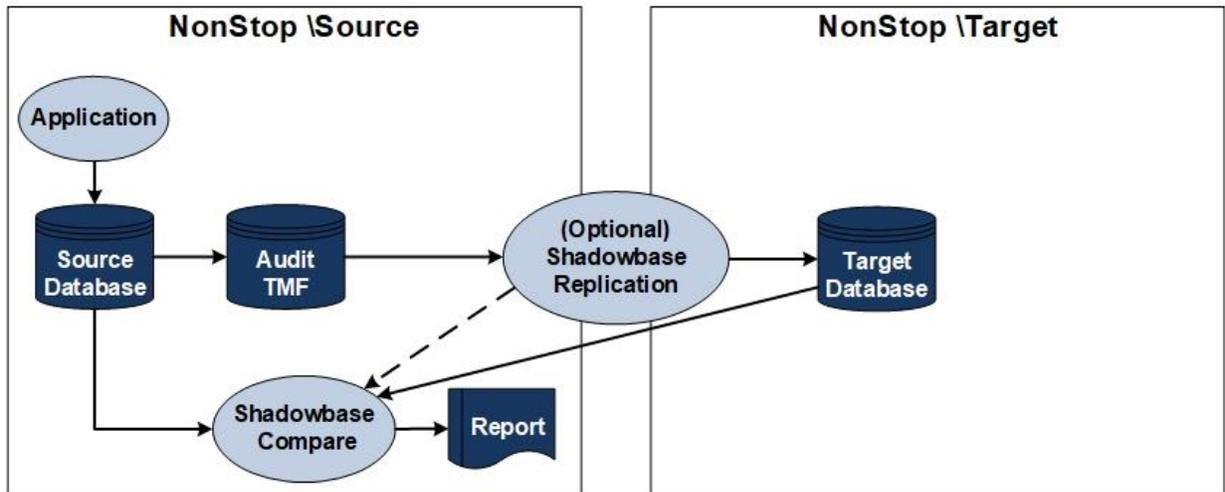
Figure 1 – HPE Shadowbase Compare: Source-side Comparison

Target-side Comparison

Figure 2 depicts a sample architecture for a target-side comparison. It is similar to Figure 1, except Compare is running on the target system. Shadowbase Compare reads and compares the source and target tables, then generates a report.
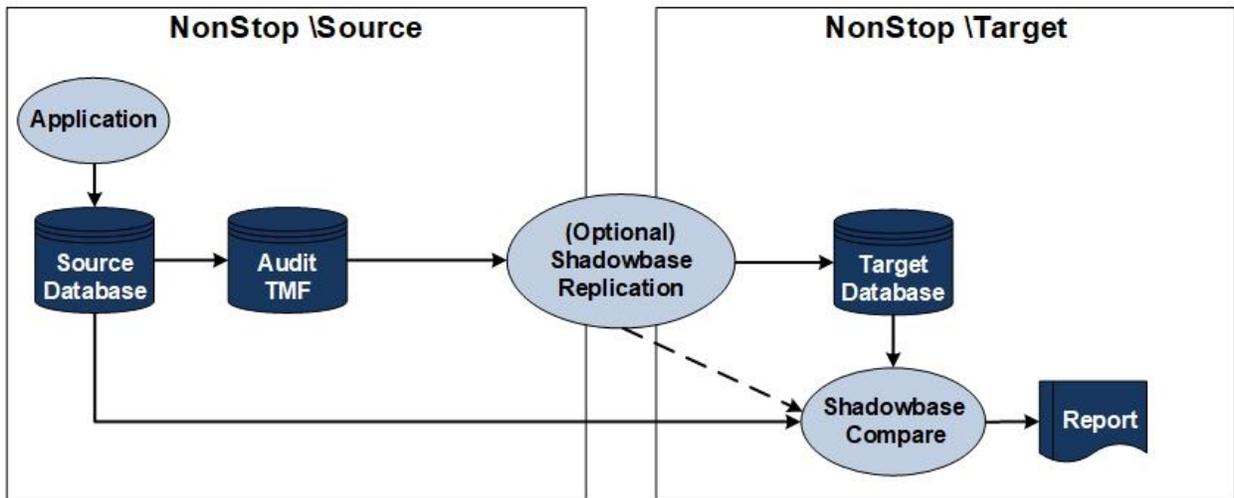
Figure 2 – HPE Shadowbase Compare: Target-side Comparison

In both figures, an optional data replication engine can be actively replicating between the source and target environments. Shadowbase Compare can detect differences caused by replication latency, and continually retries the comparison operation to avoid "false positives."

Note that HPE Shadowbase Compare Enscribe and SOLVUTIL (Compare for OSS Regular Files) processes run on both the source and target system.

Frame of Reference

Either side can be considered the "source," or correct version of the data, and the "target," or copy of the data. Changing this "polarity" alters the frame of reference for reporting what is extra data and what is missing data from the final discrepancy report.

Browse Access

Shadowbase Compare opens the source and target files/tables in shared, read-only mode. The solution never locks any rows in the source or target files/tables (it uses "BROWSE" access); the production environment is not affected by locked rows due to the comparison operation. When a comparison runs, the files/tables can be opened by applications in all modes except exclusive mode.

This method improves data concurrency; however, it also means that there can be read changes that later are undone in a TMF transactional environment (e.g., the source transaction aborts). Shadowbase Compare automatically handles this type of mismatch as described below.

In-flight, In-process, or Queued Differences

Shadowbase Compare can be used with an active, real-time replication engine. In a production environment, it is common for temporary differences to exist between source and target tables. The replication engine sees changes applied to a source table, and immediately creates a replica, sending and then applying it to a target table. The time this process takes is called "*replication latency*"[4] or "*lag.*" The not yet applied changes may be in-flight or in-process. If Shadowbase Compare compares source and target rows in such a situation, it finds a difference. Once the change has been applied to the target table this difference disappears.

Shadowbase Compare's Unique Iteration Algorithm (UIA)

Shadowbase Compare's UIA operates as follows:
1. It scans the rows for differences.
   A. If the rows match, it moves to the next set of rows.
   B. If they do not match (a difference is found), it creates a "difference" table, and adds the two rows' primary keys.
   C. This process continues until all rows are compared.
2. Immediately after step one, the comparison reviews the rows from the difference table. (Typically, a smaller number of rows are compared.)
   A. If the rows now match, it deletes them from the difference table, or it leaves the mismatch in the difference table.
   B. This process continues until no entries are left in the difference table or the max number of iterations is reached.
3. You can configure the number of iterations. After the algorithm finishes iterating through the difference table, it is assumed the remaining differences are "real." These differences are generated in the final report with statistics on the mismatches.

False Positives

This algorithm eliminates most cases of real differences due to conditions such as data replication latency. However, certain situations will report "false positives," for example, when the replication product is not running, or replication latency is longer than the algorithm's total time spent iterating through the difference table. To resolve these false positives, the comparison is simply rerun to see if the differences were resolved or not.

HPE Shadowbase Compare Products

HPE Shadowbase Compare runs in a batch-like manner from a command line user interface and includes:
- Shadowbase Compare for SQL (includes SQL/MP and SQL/MX)

---

[4] Replication latency cannot be eliminated. However, synchronous replication can be utilized to eliminate problems created by the latency such as data loss. Please reference: https://www.shadowbasesoftware.com/publications/2017/08/shadowbase-synchronous-replication-product-overview/

- Shadowbase Compare for Enscribe
- Shadowbase Compare for OSS Regular Files

HPE Shadowbase Compare for SQL

Network Requirements

Either Expand or TCP/IP connectivity can be used. If Expand is not available (or you desire to use TCP/IP for inter-node communication), Shadowbase Compare SQL will use a "Remote Agent" process to compare the read data.[5]

Comparison Processing

The comparison is done on a row-by-row and column-by-column basis. The range of rows or the range of columns to be compared is defined via user-supplied parameters (e.g., the entire table, or a key range supplied by primary key or index path). If differences are found, you have numerous Repair options to correct them. Actions are specific to the type of difference found (e.g., missing rows, extra rows, or different values between certain columns). The actions may include steps to assist in repairing the differences.

Using HPE Shadowbase Compare for SQL

A comparison is executed in a few simple commands:
1. "RUN SQLCMPE" (to access Shadowbase Compare from TACL)
2. COMPARE>"source $data02.yyydema0.exam11" (specifies the source table)
3. COMPARE>"target $data02.yyydema0.exam12" (specifies the target table)
4. COMPARE>"start"

Configurable parameters for Shadowbase Compare for SQL include:
- Names of the tables to compare
- Description of the row-to-row relationship between the tables
- Range of rows to compare (specified by a "WHERE" clause)
- Columns to compare (including data conversion rules/actions)
- Level of output detail required (granularity of comparison differences to provide in the listing, for example, difference counts versus individual column data values to show specific differences)

Final Difference Report

Color key (provided for clarity here, not in the final printout):

- Green – Data that matches
- Red – Data that *does not* match
- Yellow – Data exists on the *source* but not the *target*
- Blue – Data exists on the *target* but not the *source*

---

[5] For more information, please reference the HPE documentation, *HPE Shadowbase Compare over TCP/IP*.

Source Table (contents before comparison listed below)

```
>>select * from table-1;

      ONE             TWO                      THREE        FOUR
      -----------     --------------------     ----------   -----------
                1     1                        1                      1
                2     2                        2                      2
                3     3                        3                      3
                4     4                        4                      4
                5     5                        5                      5
                6     6                        6                      6
                7     7                        7                      7
                8     8                        8                      8
                9     9                        9                      9
               10     10                       10                    10

      --- 10 row(s) selected.
```

Target Table (contents before comparison listed below)

```
>>select * from table-2;

      ONE             TWO                      THREE        FOUR
      -----------     --------------------     ----------   -----------
                1     1                        1                      1
                2     2                        2                      2
                3     3                        3                      3
                4     4                        4                      4
                5     5                        5                      5
                6     six                      sicks                  6
                7     seven                    seven                  7
                8     eight                    ate                    8
               11     eleven                   11                    11
               12     twelve                   twelve                12

      --- 10 row(s) selected.
```

Compare Output

```
COMPARE>source djbdata.tsttabl
COMPARE>target djbdata.tsttab2
COMPARE>start
Start                          :              2019-12-11 11:09:04

---> Difference(s) found for key <--------------
ONE                                 : 6

Source Column TWO = 6
Target Column TWO = six

Source Column THREE = 6
Target Column THREE = sicks

---> Difference(s) found for key <--------------
ONE                                 : 7

Source Column TWO = 7
Target Column TWO = seven

Source Column THREE = 7
Target Column THREE = seven

---> Difference(s) found for key <--------------
ONE                                 : 8

Source Column TWO = 8
Target Column TWO = eight

Source Column THREE = 8
Target Column THREE = ate

---> Target record not found for key <--------------
ONE                                 : 9

---> Target record not found for key <--------------
ONE                                 : 10

---> Source record not found for key <--------------
ONE                                 : 11

---> Source record not found for key <--------------
ONE                                 : 12
-------- Primary compare run finished.
Source: \VIV1.$QA.DJBDATA.TSTTABL
Target: \VIV1.$QA.DJBDATA.TSTTAB2

Number of source records read:          10
Number of target records read:          10
Total differences          :            7
Column value differences   :            3
Source not found           :            2
Target not found           :            2
Tables are not equal!

Start                          :              2019-12-11 11:09:04
Finished                       :              2019-12-11 11:09:04
Duration                       :              00:00:00.020058
```

HPE Shadowbase Compare for Enscribe

Network Requirements

Either Expand or TCP/IP connectivity can be used.

Comparison Processing

The comparison is done on a record-by-record and field-by-field basis. The comparison starts a process in the source environment, which directly reads the source file, and sends the data that needs compared in large blocks to a user-defined system – typically the target system – to offload resources from main Production. This target-side comparison process reads the corresponding data from the target file and efficiently compares each record in-memory. This sequence continues until the end of either the file(s) or specified comparison key range is reached. If the comparison process runs on the source file system, then the above mechanism is reversed, with the target file contents sent by Shadowbase Compare Enscribe to the source file system.

The discrepancy report generated by Shadowbase Compare Enscribe lists the differences between the target and source, including:

- Records or rows in the source, but not in the target
    o Records that are in the source file, but not in the target file are reported as INSERTs.
    o The record in the source file needs to be written or INSERTed into the target file in order for the target file to match the source file
- Records or rows in the target, but not in the source
    o Records that are in the target file, but not in the source file are reported as DELETEs
    o The record in the target file needs to be removed in order for the target file to match the source file
- Records or rows in both whose keys match, but the data contents do not
    o Records that exist on both files, but the record data do not match are reported as UPDATEs
    o The record in the target file needs to be UPDATEd so its data content matches the source

Shadowbase Compare Enscribe always considers the user-defined SOURCE system as being the "record of truth" and ensures that the TARGET matches the source. To repair a source system's data, the system polarity is simply reversed.

Using HPE Shadowbase Compare for Enscribe

A comparison is executed in a few simple commands:
1. Create a SHADPARAM parameter file with the following parameters:
   A. VVSOURCEFILENAME=<location of source file>
   B. VVTARGETFILENAME=<location of target file>
2. "run solv -f<shadparm-filename> -s<shadparm-section-name>"

There are dozens of additional parameters to control various aspects of the comparison process.[6]

---

[6] Please reference the latest version of the Shadowbase manual on the HPE NonStop Technical Library for more information.

Final Difference Report

```
REPORT MODE: SUMMARY
--------------------------- FILE INFO ----------------------------
FILE             FILENAME           REC  BLK TYPE FMT ADT ADT COMP
---- ------------------------- ---- ---- ---- --- --- --------
SOURCE \H2.$QA.CJCDATA.KEYFILE  1000 4096    K   1 YES       NO
TARGET \H2.$QA.CJCDATA.KEYFILE2 1000 4096    K   1 YES       NO
--------------------------- KEY INFO ----------------------------
-----PRIMARY KEY-----
  FILE OFFSET LENGTH
------ ------ ------
SOURCE     0      5
TARGET     0      5
SUMMARY:
  FILE                              FILENAME
------   ---------------------------------
SOURCE              \H2.$QA.CJCDATA.KEYFILE
TARGET             \H2.$QA.CJCDATA.KEYFILE2
START KEY MODE: BEGINNING OF FILE
 STOP KEY MODE: END OF FILE


---Source---   ---------Rows---------   ---Target---
          0   Identical                            0
          1   Not in Target (Inserts)              -
          1   Mismatches    (Updates)              1
          -   Not in Source (Deletes)              1
          2   Fetched From Table                   2
              Resync Totals (I+U+D)                3
```

HPE Shadowbase Compare for OSS Regular Files

SOLVUTIL is an independent, standalone Guardian process that contains several different functions for copying (replicating) and comparing OSS Regular Files from a source to a target environment.

The method SOLVUTIL uses to copy the source file to the target file is referred to as "EOF Chasing," which means the SOLVUTIL process replicates data from the source to the target as new data is appended onto the end of the source file. SOLVUTIL does not detect, nor copy, source file update changes that were previously made in the already-replicated file.

When SOLVUTIL hits end-of-file during the copying process, it can be configured to delay and retry the copy operation that is waiting for more data to be appended into the source file.

Network Requirements

Expand connectivity is required if the files are located on different systems.[7]

Comparison Processing

When "compare" mode is turned on, SOLVUTIL compares the source data to the corresponding target data, ending the comparison at either EOF or at the first byte that mismatches.

Using HPE Shadowbase Compare for OSS

A comparison is executed in a few simple commands:

---

[7] Contact Gravic for the availability of Shadowbase SOLVUTIL for OSS Regular Files support over TCP/IP.

1. Create a SHADPARAM parameter file with the following parameters:
   A. SOURCEPATHNAME=<pathname>
   B. TARGETPATHNAME=<pathname>
   C. COMMAND = COMPARE
2. "run solvutil/name /-smysection -f$myvol.mysub.myconfig"
   A. "mysection" is the section of the parameter file, and "$myvol.mysub.myconfig" is the name of the parameter file.

There are dozens of additional parameters to control various aspects of the comparison process.[8]

Sample Final Difference Report

```
> param sourcepathname /home/osstest/gig
> param targetpathname /home/osstest/gigcopy
> param command compare
> run solvutil   -d -v

SOLVUTIL - T1122 - V6600L06 - (01JUN19), COPYRIGHT GRAVIC, INC.
1995-2019. PORTIONS COPYRIGHTED BY AND LICENSED FROM THIRD PARTIES. SEE
README FILE. USAGE SUBJECT TO THE TERMS OF A WRITTEN LICENSE AGREEMENT.
PATENTS: SEE WWW.GRAVIC.COM/GRAVICLABS/PATENTS/PRODUCTS.

Started \VIV1.$Z9YA:24988669 at 2019-06-25:17:29:13.237112
\VIV1.$GPJ.DEV6600.SOLVUTIL|CPU 0|Node 254|Pri 158|Build Jun 24 2019 16:58:31
-f=SHADPARM -s=SOLVUTIL -verbose=TRUE -disablelogprefix=TRUE
file:SHADPARM section:SOLVUTIL - file and section exist
CONSOLELOGGING = enable
FILELOGGING = '' (not specified, using default)
EMSLOGGING = '' (not specified, using default)
COMMAND = compare
LICENSEFNAME = 'SHADPASS' (not specified, using default)
SOURCEPATHNAME = /home/osstest/gig
TARGETPATHNAME = /home/osstest/gigcopy
SOURCEFILENAME = \VIV1.$OSS.ZYQ00001.Z0000728
TARGETFILENAME = \VIV1.$SAS4.ZYQ00001.Z00007FR
               \VIV1.$OSS.ZYQ00001.Z0000728        \VIV1.$SAS4.ZYQ00001.Z00007FR
Files compare for length of 1073741824
Program ending at 2019-06-25:17:29:29.239264
```

Summary

Enterprises around the globe use online data replication for its numerous advantages, including business continuity. Still, IT personnel must prove to auditors and the business itself that the databases are consistent. If a disaster or malware corrupts the production database, or a user misconfigures the application or replication environment, then the backup system and database may not function properly when called into service. Worse, the IT personnel may not even be aware of the corruption, causing additional damage to enterprise operations. Offline database validation methods are impractical because taking the application offline is inconvenient and is unable to prove data consistency when an outage is most likely to occur: while systems are online and actively running.

HPE Shadowbase Compare validates that two like HPE NonStop databases match, and highlights any differences found by creating a summary or detailed report. Applications and databases can remain online for updating on both the source and target while the comparison process occurs.

Shadowbase Compare includes Shadowbase Compare for SQL/MP, Shadowbase Compare for SQL/MX, and Shadowbase Compare for Enscribe. *Shadowbase Compare is recommended for every NonStop business continuity environment* to be absolutely certain that your replication environment is running correctly. It proves that the source and target database copies are consistent and helps satisfy regulatory and audit requirements for data verification.

---

[8] Please reference the latest version of the manual for more information.

International Partner Information

Global

Hewlett Packard Enterprise
6280 America Center Drive
San Jose, CA 95002
USA
Tel: +1.800.607.3567
www.hpe.com

Japan

High Availability Systems Co. Ltd
MS Shibaura Bldg.
4-13-23 Shibaura
Minato-ku, Tokyo 108-0023
Japan
Tel: +81 3 5730 8870
Fax: +81 3 5730 8629
www.ha-sys.co.jp

Gravic, Inc. Contact Information

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

Hewlett Packard Enterprise Business Partner Information

Hewlett Packard Enterprise directly sells and supports Shadowbase Solutions under the name **HPE Shadowbase**. For more information, please contact your local HPE account team or visit our website.

Copyright and Trademark Information