



# HPE Shadowbase Data Loading Methodologies

## Techniques for Online and Offline Database Loading

Before a database is useful, it must first be loaded with data (for example, when installing a new backup system for business continuity, or when performing a system/software upgrade or migration). The HPE Shadowbase Online Loader (SOLV) and HPE Shadowbase Extract, Transform and Load (ETL) Toolkit are utility products designed to help perform this task. However, there are alternatives, and various ways in which SOLV and ETL can be used to load a database:

- Bulk (Offline) Loading
- SOLV (Online) Loading
  - Merged Loading
  - Block loading



This technical brief provides an overview of these alternative methods, and their respective pros and cons.

### Bulk (Offline) Loading

This technique uses various disk management utilities (typically provided as part of operating system or database software, such as FUP/SQL DUP, Copy/Load, PAK/UNPAK, Backup/Restore, etc.) to bulk (point-in-time) extract data from a source database and load it into a target database. During the loading process the source database may be online, even for updating, however the target database remains offline and unavailable for use. Once the target database is loaded, before it can be used it must be brought into consistency with the source database (since data changes may have occurred on the source database since the extract was performed). In a Shadowbase data replication environment, this is achieved via replay of all transactions executed against the source database between the point at which the bulk extract was performed and the current time – at the completion of which the target database will be consistent with the target database and can be brought online. Figure 1 provides an overview of this process.

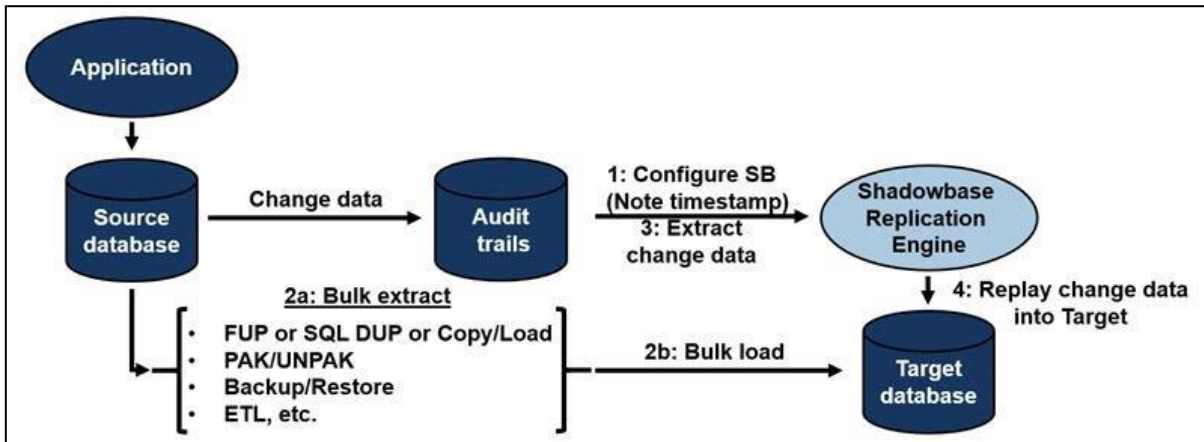


Figure 1: Bulk (Offline) Loading Method

There are also a few procedural details which can help facilitate this process:

1. A **timestamp needs to be noted** at the point the bulk extract is performed so that a time is known from which to begin the Shadowbase replay of queued transactions (step #1). This timestamp could simply be the wall clock time when the extract is performed. But a more reliable method is to have Shadowbase data replication up and configured between the source and target systems (but not actively replicating), and set a Shadowbase *restart point* at the time the extract is started. Then the transaction replay (in step #3) can simply be started from this Shadowbase restart point.
2. During the time the load is taking place (steps 2a and 2b) the source system is processing transactions, and the database audit trail (for example the TMF audit trail) is filling up with transaction entries (called “change data”). The audit trail disk space is not infinite, so it is possible that while the load is still in process, the audit trail may fill-

up and “roll,” and transactions will be lost. (They will be deleted from the audit trail; hence, they will not be found in the audit trail by Shadowbase software when the replay process begins, nor will they be applied to the target database.) To prevent this serious data loss situation from occurring, while the load is proceeding, HPE Shadowbase software can be configured to **replicate the source data changes from the audit trail into its own queue files**, using a [Shadowbase Queue Manager](#) from where they will be replayed later, thereby preventing the loss of any data if the audit trail rolls off the system and is scratched by TMF.

3. Once the load has completed (steps 2a and 2b), the transactions in the change log are replayed by Shadowbase to re-synchronize the target database with the source database using the transactions that occurred while the load occurred (step 3 and step 4).
4. A best practice is to **not** create indices nor alternate key files on the target database until *after* the transaction replay process is complete. The primary reason for this is performance, performing lots of change data replication with the indices in place can lead to locking issues on the indices, which slows down replay of the queued transactions.

### Bulk (Offline) Loading Pros and Cons

- Bulk Loading is very fast
  - It typically performs block-oriented operations (not record or row-oriented operations)
  - If source is active for updating during the load, source transactions need to queue and then be replayed into the target after the Bulk Load completes
- Target database is offline and unusable until the Bulk Load completes, and is typically inconsistent until the Change Data replay is complete
  - Often, indices or alternate keys are not created/loaded on the target until after the Bulk Load completes
- Source and target databases must be very similar
  - Typically, no (or minimal) data filtering, transformation, or cleansing is allowed
  - Often only available for homogeneous like-to-like source and target databases

### SOLV (Online) Merged Loading

This technique takes advantage of the benefits of the HPE Shadowbase SOLV utility, a key aspect of which is the ability to run the SOLV loading process in parallel with Shadowbase data replication from the source to the target database, enabling the target database to remain online and negating the need for transaction replay after the load completes. The procedure is straightforward; simply configure and start Shadowbase data replication between the source and target databases, and at the same time commence the SOLV loading process. Close integration between SOLV and Shadowbase data replication software handles the merging of load and change (replay) data into the target database, ensuring the target database remains consistent with the source database while the load occurs. Once the load is complete, there are no queued transactions to replay. Figure 2 provides an overview of this process.

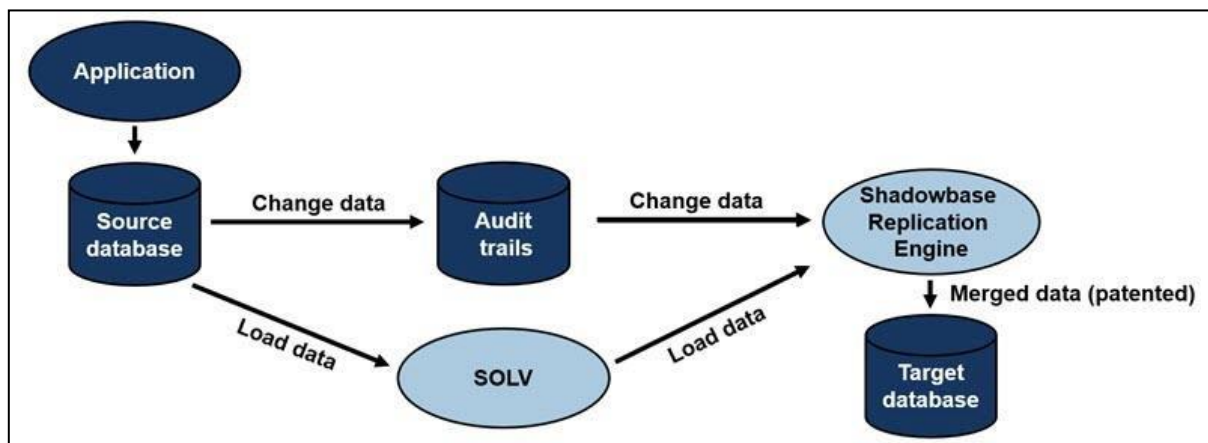


Figure 2: SOLV (Online) Merged Loading Method

### SOLV (Online) Merged Loading Pros and Cons

- Target database is online and consistent with the source database while the load operation occurs
  - Load and transaction replay in parallel
- No waiting for replay of queued transactions
  - Once load is finished, target database is immediately caught-up
- Source and target databases can be very different
  - Data filtering, transformation and cleansing is possible
  - Useful for homogeneous and heterogeneous source and target databases
- Slower than bulk (offline) loading as it inserts by record/row instead of block

## SOLV (Online) Loading

This technique is similar to SOLV merged loading as described above, except that the transaction replay and merging of the load data into the target database do not take place in parallel. Rather, this method is a hybrid of the bulk and SOLV loading methods, where SOLV (rather than some bulk loading utility) is used to first load the target database. Simultaneously, the source database changes are queued by Shadowbase software, and then replayed into the target database once the SOLV load is complete. The benefits of this hybrid approach are the flexibility of using SOLV (for heterogeneous data transformation, etc.), with the speed of bulk loading. As is the case for bulk loading, Shadowbase data replication can be configured to drain audit trail entries into its queue files to avoid the possibility of lost data as a result of audit trail rolls. Figure 3 provides an overview of this process.

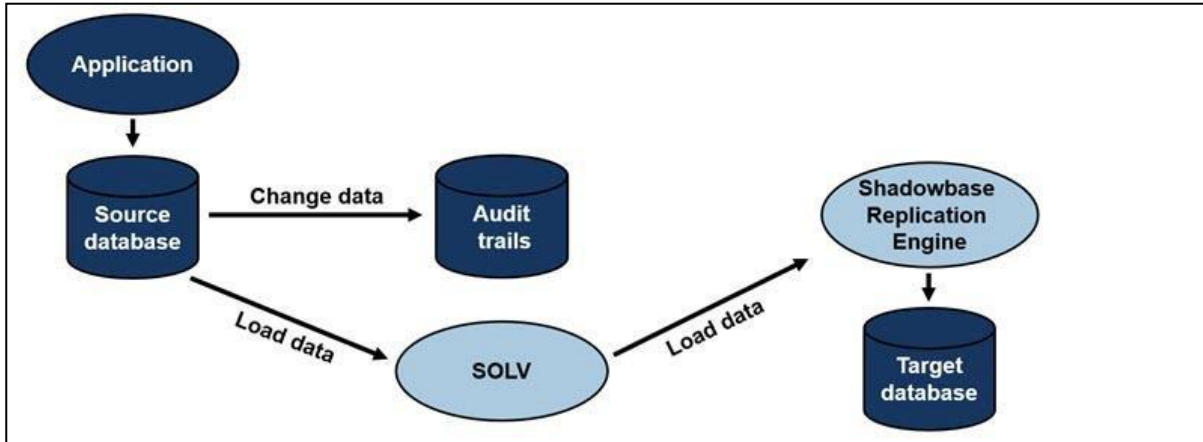


Figure 3: SOLV (Online) Block Loading Method

### SOLV (Online) Block Loading Pros and Cons

- Target database is online but inconsistent with the source database until transaction replay is complete
- Source and target databases can be similar or very different
  - Data filtering, transformation and cleansing is possible
  - Useful for homogeneous and heterogeneous source and target databases
- Slower than bulk (offline) loading, but faster than SOLV (online) merged loading
  - Target is inconsistent until the replay of queued transactions completes once the load completes

### Summary

A database must first be loaded before it is of any use. Different database loading techniques are possible using either off-the-shelf or HPE Shadowbase loading software (SOLV). Each method has its pluses and minuses, but one of them provides the best solution to satisfy the requirements of any particular database loading situation. 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.

#### Learn more:

[shadowbasesoftware.com](http://shadowbasesoftware.com)  
[hpe.com](http://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](mailto:shadowbase@gravic.com)  
Email Support: [sbsupport@gravic.com](mailto:sbsupport@gravic.com)

#### Please Follow:

