

**Note:** Comprestimator can run for a long period (a few hours) when it is scanning a relatively empty device. The utility randomly selects and reads 256 KB samples from the device. If the sample is empty (that is, full of null values), it is skipped. A minimum number of samples with data is required to provide an accurate estimation. When a device is mostly empty, many random samples are empty. As a result, the utility runs for a longer time as it tries to gather enough non-empty samples that are required for an accurate estimate. The scan is stopped if the number of empty samples is over 95%.

## 10.6.2 Evaluating compression and deduplication

To help with the profiling and analysis of user workloads that must be migrated to the new system, IBM provides a highly accurate data reduction estimation tool that supports both deduplication and compression. The tool operates by scanning target workloads on any legacy array (from IBM or third party) and then merging all scan results to provide an integrated system level data reduction estimate.

The Data Reduction Estimator Tool (DRET) utility uses advanced mathematical and statistical algorithms to perform an analysis with low memory footprint. The utility runs on a host that can access the devices to be analyzed. It performs only read operations so it has no effect on the data stored on the device.

The following sections provide information about installing DRET on a host and using it to analyze devices on it. Depending on the environment configuration, in many cases DRET is used on more than one host to analyze more data types.

When DRET is used to analyze a block device that is used by a file system, all underlying data in the device is analyzed, regardless of whether this data belongs to files that were deleted from the file system. For example, you can fill a 100 GB file system and make it 100% used, and then, delete all the files in the file system to make it 0% used. When scanning the block device that is used for storing the file system in this example, the DRET accesses the data that belongs to the files that are deleted.

**Important:** The preferred method of using DRET is to analyze volumes that contain as much active data as possible rather than volumes that are mostly empty of data. This increases the accuracy level and reduces the risk of analyzing old data that is deleted, but might still have traces on the device.

For more information and the latest version of this utility, see this IBM Support [web page](#).

## 10.7 Data deduplication and compression on external storage

Starting from IBM Spectrum Virtualize V8.1.x, it supports over-provisioning on selected back-end controllers. This means that if back-end storage performs data deduplication or data compression on LUs provisioned from it, they still can be used as external MDisks on IBM Storwize V7000.

Thin-provisioned MDisks from controllers that are supported by this feature can be used as managed mode MDisks in IBM Storwize V7000 and added to storage pools (including DRPs).

Implementation steps for thin-provisioned MDisks are same as for fully allocated storage controllers. Extra caution is used when planning capacity for such configurations.