

HOW QBR-S HANDLES COMPRESSION & DATA DEDUPLICATION

This blog post explains how the QBR-S provides Data Compression & Data Deduplication of backup data stored locally on the Primary Storage Volume of the device.

DEFINITIONS

Data Compression: Process of encoding information by using fewer bits than the original source of information.

Data Deduplication: Process of comparing two or more previously compressed data sets and removing duplicate chunks of data. Use of efficient Data Deduplication is dependent on also having Data Compression performed.

In-Line Data Deduplication: Application of data deduplication during the copy on write process performed during a backup.

EXAMPLE OF HOW TO CALCULATE RATIOS FOR DATA OF WHICH HAS ALREADY BEEN THROUGH DATA COMPRESSION & DATA DEDUPLICATION.

Most Data Compression & Data Deduplication processes use some form of the *LZ adaptive dictionary based algorithm*.

Creation of these dictionaries varies from one Data Compression process to another, however in this example we will pick out repeated words and put them into a numbered list which will allow us to list the words by their much smaller associated numeric value.

Example Quote

"He can compress the most words into the smallest idea of any man I know"

-Abraham Lincoln

Dictionary Created For This Example Quote

QBR *Knowledge base*

1. He
2. can
3. compress
4. the
5. most
6. words
7. into
- the (Dictionary value #2)
8. smallest
9. idea
10. of
11. any
12. man
13. I
14. know

New Sentence Created Using This Dictionary

"1 2 3 4 5 6 7 1 8 9 10 11 12 13 14"

The new compressed quote now uses 34 characters (including spaces).

The original quote contained 71 characters (including spaces)

Calculating The Data Compression Ratio of Abraham Lincoln's Quote

Formula:

Original Size ÷ Compressed = Compression Ratio

In Our Example:

71 ÷ 34 = 2.09x

Result Ratio of Our Example:

2.09x is The Data Compression Ratio