

CA ARCserve Backup r12.5 - Data Deduplication

Today, as organizations continue to grow, there is an exponential increase in the data that is generated; Industry analysts predict that the data storage requirements would increase 30 to 50 percent each year. Business policies including compliance, disaster recovery, backup, approval, and delivery require you to save multiple copies of your data, adding additional maintenance overheads. Managing huge volumes of data has always been one of the major challenges for organizations. It is important for organizations to protect data, and have a cost-effective storage solution that would minimize the overheads and enable to focus on its core business growth.

Data Deduplication is one of the fastest growing backup technologies that can help companies reduce the amount of disk space required for backup. The significant reductions in backup space requirements enable you to fit more backups on the same physical media and retain media for longer periods of time.

CA ARCserve backup r12.5 provides an efficient data deduplication technology that can help eliminate redundant data, extend storage resources, speed backup and recovery times. CA ARCserve Backup data deduplication feature examines data for redundancy, stores unique data blocks to disk, and replaces the identified duplicate data block with a marker that points back to the unique storage location of each block.

Overview

CA ARCserve Backup provides an integrated data deduplication feature that helps you reduce your storage needs by eliminating redundant data while protecting the entire data. Compares backups with previously stored data at the block-level instead of comparing duplicates on a file-by-file comparison basis

During Backup, CA ARCserve Backup data deduplication feature analyzes the data streams looking for duplicate blocks, saves only unique blocks to disk. Creates three files (hash, reference and Data files) for every backup session and tracks the duplicate entries in special index files.

During restore, CA ARCserve Backup refers to the index files to identify the data needed to reassemble the original data stream.

Benefits

- Reduces the amount of physical storage required, manifests efficient usage of disks for backups.
- Easy to manage, configure, and integrate into existing backup environments.
- Provides data deduplication efficiency reports: Deduplication Promotion Report and Deduplication Status Report, which display the estimated savings of backup space, and the number of nodes backed up using a deduplication device.
- Optimization feature, for better throughputs and decreased CPU usage.
- Global deduplication feature identifies redundancy between the backup jobs performed on the root directories of two different computers.

Install/Configure

The Data Deduplication feature is installed as part of the CA ARCserve Backup Server r12.5 installation. However, if you are using any of the older versions of CA ARCserve agents (except for Netware, AS400 and Open VMS agents), you need to upgrade to CA ARCserve r12.5 to avail the benefits of data deduplication.

To use the CA ARCserve Backup data deduplication feature, you must create and configure data deduplication device. The deduplication device configuration process is similar to creating a file system device (FSD).

Configure Groups for Data Deduplication

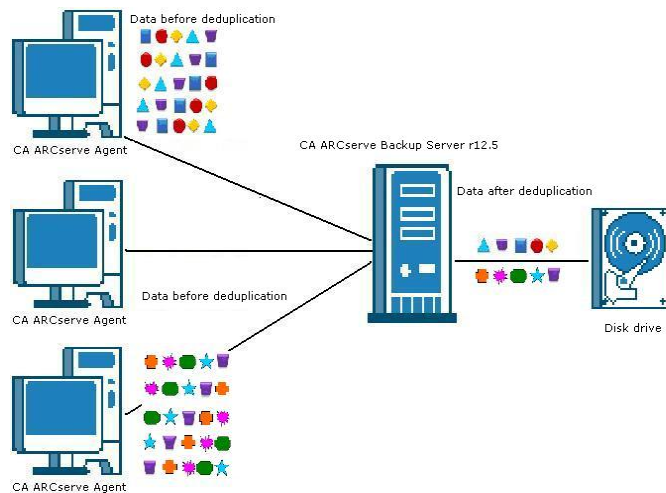
CA ARCserve Backup lets you configure deduplication groups using Device Manager. You may specify a Group name while creating a deduplication device, if you do not specify a group, CA ARCserve Backup automatically assigns it to a new deduplication device group. You may not assign more than one deduplication device to the same group, however, you can add a new (empty) group and assign a device to it later.

Optimization

CA ARCserve Backup provides an optimization feature for better throughputs. Select the Allow optimization in Deduplication Backups option located on the Deduplication Group Configuration screen. Optimization is supported on Windows volumes only. It is not supported for stream-based backups, such as SQL VDI, Exchange DB level, Oracle RMAN or VMware Image level backups.

How Data Deduplication works

Data Deduplication backup option operates like a normal backup job. To deduplicate data during a backup job, choose a deduplication device as the backup destination.



During the first backup, CA ARCserve Backup scans incoming data and segments it into data blocks, executes a hashing algorithm that assigns a unique value to each block of data and saves those values to a hash file. CA ARCserve Backup compares hash values. When duplicates are found, data is written to disk only once, and a reference is added to a reference file pointing back to the storage location of the first identified instance of that data block.

During subsequent backups, CA ARCserve backup scans incoming data, breaks it into range of blocks and executes the hashing algorithm to assign hash values. CA ARCserve Backup compares new hash values to previous values, looking for duplicates. When duplicates are found, data is not written to disk.

Restoring data that has been deduplicated follows the same procedure as a normal restore job.

Deduplication Device Management

CA ARCserve Backup enables you to create data deduplication devices, remove, or change the properties on an existing device using Device Configuration, however, you cannot modify the tape name. You must have proper security access to create a data deduplication device on a remote server.

Protect Deduplication Devices

When deduplication devices are created locally, the deduplication data files are excluded from CA ARCserve Backup jobs. You can protect the deduplication device by enabling the "Backup deduplication device data", option in global settings, which enables VSS option and take backup of the deduplication drive using VSS.

Frequently Asked Questions

Q: Can Data Deduplication Device (DDD) be used in staging and final destination?
A: Yes, you can submit a backup job to deduplication device by selecting deduplication device in final destination tab. When you choose a DDD device in the final destination tab, specify a purge policy (no copy policy). You can also submit a backup job by choosing the DDD device in staging tab, specify the copy and purge policy similar to any staging device.

Q: How many files are created for each data deduplication session?
A: 3 files are created for each backup- a hash file, a reference file, and a data file.

The hash file contains a series of hashes which were calculated in the order in which the data was backed up. Each entry also points to an offset in a specific reference file. The reference file contains offsets to data file. It also maintains a count for each hash. The data file stores the actual data.

Summary

CA ARCserve Backup r12.5 provides an intelligent data deduplication feature that reduces storage needs by eliminating redundant data. Only unique instance of the data is retained on storage media, redundant data is replaced with a pointer to the unique data copy. The significant reductions in backup space requirements made possible by data deduplication enable you to fit more backups on the same physical media and retain media for longer time.

To get maximum value from your investment in CA software, CA Education can help you get the right training for your unique challenges. Visit <http://ca.com/education> for additional information.

Copyright © 2009 CA. All rights reserved. All trademarks, trade names, service marks and logos referenced herein belong to their respective companies.