2.5.21 Data Reduction Pools

Data Reduction Pools (DRP) represent a significant enhancement to the storage pool concept. The reason is the virtualization layer is primarily a simple layer that runs the task of lookups between virtual and physical extents.

Data Reduction Pools is a new type of storage pool and implements techniques, such as thin-provisioning, compression, and deduplication to reduce the amount of physical capacity that is required to store data, and decrease the network infrastructure required. Savings in storage capacity requirements translate into reduction in the cost of storing the data.

The storage pools enable you to automatically de-allocate and reclaim capacity of thin-provisioned volumes that contain deleted data and for the first time, enable this reclaimed capacity to be reused by other volumes. Data reduction provides more performance from compressed volumes because of the new log structured pool that is implemented.

2.5.22 Deduplication

Data deduplication is one of the methods of reducing storage needs by eliminating redundant copies of a file. Data reduction is a way to decrease the storage disk and network infrastructure that is required, and optimize use of storage disks and improve data recovery infrastructure efficiency. Existing data or new data are standardized into chunks that are examined for redundancy. If data duplicates are detected, pointers are shifted to reference a single copy of the chunk and the extra duplicate data is then released.

Deduplication has benefits, such as storing more data per physical storage system, which saves energy by using fewer disk drives and decreasing the amount of data that must be sent across a network to another storage for backup replication and disaster recovery.

2.5.23 IP replication

IP replication was introduced in V7.2 and allows data replication between IBM Spectrum Virtualize family members. IP replication uses IP-based ports of the cluster node canisters. This function is transparent to servers and applications in the same way that traditional FC-based mirroring is. All remote mirroring modes (Metro Mirror, Global Mirror, and Global Mirror with changed volumes) are supported.

The configuration of the system is straightforward, and IBM Storwize family systems normally find each other in the network and can be selected from the GUI.

IP replication includes Bridgeworks SANSlide network optimization technology and is available at no extra charge. Remember, remote mirror is a chargeable option, but the price does not change with IP replication. Remote mirror users have access to the function at no extra charge.

IP connections that are used for replication can have long latency (the time to transmit a signal from one end to the other). This latency can be caused by distance or by many “hops” between switches and other appliances in the network. Traditional replication solutions transmit data, wait for a response, and then transmit more data, which can result in network utilization as low as 20% (based on IBM measurements).