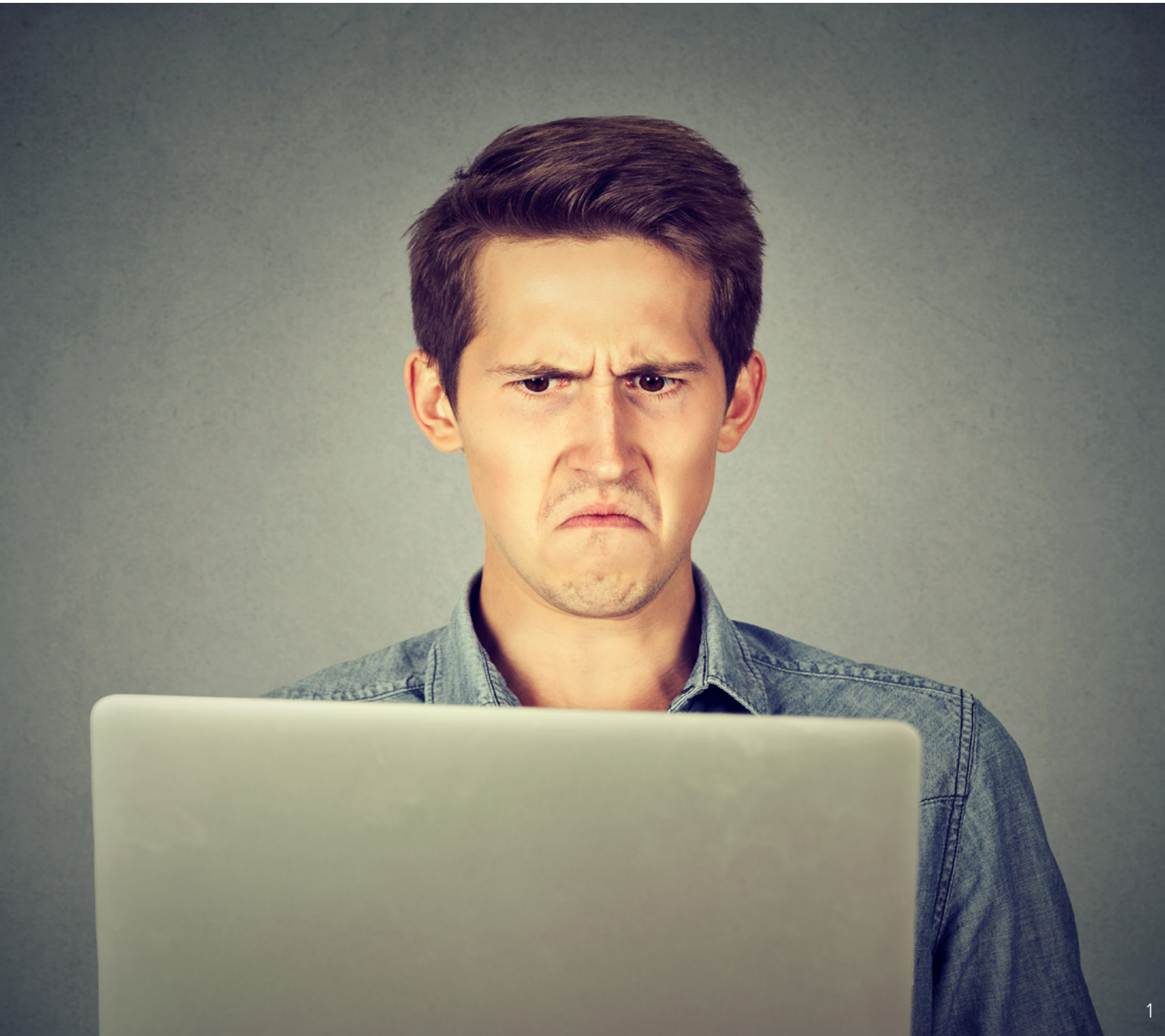
IDERA® | SQL Server Whitepaper

DATABASE BACKUP COMPRESSION VERSUS FILE DEDUPLICATION

BY STAN GEIGER

PROBLEM STATEMENT

Database Administrators often argue with their Storage Administrators about compressing database backups. Storage Administrators state that file deduplication from storage devices such as EMC Data Domain is a better strategy. Storage Administrators often tell Database Administrators not to compress their database backups in order to get the best file compression rates. This whitepaper discusses why not compressing database backups is likely a bad idea.



WHAT IS FILE DEDUPLICATION?

Appliances such as EMC Data Domain sit between the server and the storage medium. The network connects the server, appliance, and storage. When data is saved to the storage appliance, the file is sent from the server to the appliance. The appliance then deduplicates the data and stores the reduced file. File deduplication works, in most cases, by identifying any duplicate blocks within a file. The difference between compression and deduplication is that the file deduplication process is transparent such that the type of file does not matter.

Importantly, the data file has to be transferred over the network to the appliance and then file deduplication occurs. Consider, for example, a database backup file that has a size of 800 megabytes. The backup file is first sent over the network to the appliance. The file is then deduplicated down to a size of perhaps 50 megabytes. The file deduplication saved storage space, however, there is still the penalty of transferring the large file across the network to the appliance.

Let us say that compressing the database backup file results in a file size of 80 megabytes. Because the file is compressed before being sent to the appliance, a much smaller file is transferred across the network.. Deduplicating the backup file does not reduce it any further so that its size remains 80 megabytes. The EMC Data Domain appliance can reduce the uncompressed backup file, but the amount of data transferred through the network is an order of magnitude more.

FILE DEDUPLICATION SLOWS BACKUPS

Most of the time the bottleneck in a database backup is either transferring the file over the network or in writing the file to storage. Database backup software often uses technology that takes advantage of idle CPU cycles to compress database backups. This compression reduces the size of the database backup considerably when saving the file. The result is faster transfer rates over the network as well as faster disk writes.

For file deduplication, the entire file is written to disk after transferring over the network. It is then deduplicated on the storage appliance. The result is a reduction in storage at the cost of the file transfer.

FILE DEDUPLICATION ALSO SLOWS DATABASE RESTORES

Consider restoring the database from a database backup file on EMC Data Domain. The file is “re-hydrated”, meaning reduplicated to its original state when accessed. The entire database backup file with a size of 800 megabytes is transferred across the network and restored. If the file had been compressed using database backup software, the backup file could be as small as 80 megabytes. This small file size means that the network transfer rate is an order of magnitude faster. Summarizing, when verifying database backups the penalty of passing large files through the network is considerably higher.

Log shipping provides a unique performance hit in that it often performs restores across multiple servers using multiple files. This composite restoration means that there is a continuous back and forth of data transferring through the network.

THE STORAGE ADMINISTRATOR IS WRONG

File deduplication works well with uncompressed data. Uncompressed data typically occurs with file servers. File deduplication can save a significant amount of storage area network (SAN) space for these types of files. Storage Administrators look at database backups as nothing more than writing data files to the storage area network (SAN). Storage Administrators get frustrated by the lack of compression rate for database files. If you use uncompressed database backups, you may get an 80% file deduplication rate, however, there is a price to pay as was mentioned earlier. Storage Administrators also do not realize that they could be storing an order of magnitude more data as well as extending restore times, which could be a critical side effect.

The easiest way to prove the point is to write a compressed database backup to regular file storage and write an uncompressed file to file deduplication storage. Measure the transfer times as well as the time it takes to restore from the database backups. You will quickly prove your point.

DEDUPLICATION OF COMPRESSED DATA

People often ask why file deduplication of compressed data does not yield the desired result of decreased file size on the storage appliance. Consider the process as being the equivalent of compressing a compressed file. Compression removes duplicate bits of data. Consequently, trying to find duplicate file blocks with file deduplication is not going to yield much in file size savings.

DEDUPLICATION OF ENCRYPTED DATA

Another question arises around using encrypted files with file deduplication. File deduplication rates are not good with encrypted data. The reason is that encryption works by using a mask and rearranging bytes using a key. This rearranging of bytes means that file deduplication is not going to find many if any duplicate blocks of data.

ABOUT THE AUTHOR

Stan Geiger is Director, Product Management, Multi-Platform Tools at Idera, Inc. He has over 25 years of experience using Microsoft SQL Server. He has worked in various industries including finance, healthcare, energy, and the US Department of Defense. Stan has held several positions including Database Developer, Database Administrator, and Business Intelligence Architect. His experience lies in building Data Warehouse and ETL platforms, Business Intelligence Analytics, and OLTP systems.

IDERA'S SOLUTION

IDERA's SQL Safe Backup creates backups faster and saves space via dynamic compression with encryption. Advanced compression, disk writing, and multi-threading technologies significantly increase the backup speed. The IntelliCompress technology continually samples the system resource usage and automatically adjusts the compression level to ensure the smallest backup files in the fastest time given the state of the environment. Secure backups by employing encryption via 128-bit and 256-bit advanced encryption standard (AES) with a performance degradation of less than 0.5%. Backup and restore with low impact by running SQL Safe Backup as a separate process outside of the SQL Server process space. Backup to and restore from EMC Data Domain servers, and permanently store backup files to IBM Tivoli Storage Manager (TSM) to adhere to corporate storage standards.

Start a free, fully-functional, 14-day trial today!

Start for FREE