

Disk Space Capacity Planning

a report by

Gregory A Larsen

Database Administrator, Washington State Department of Health (DOH)



Gregory A Larsen is a database administrator (DBA) at the Washington State Department of Health (DOH). He is responsible for maintaining Structured Query Language (SQL) Server and other database management software and he works with customers and developers to design and implement database changes and solve database/application-related problems. During Mr Larsen's career as a DBA, he has managed databases on Windows, Unix and IBM mainframes. He has also been President of a local user group attended by application programmers and DBAs. Mr Larsen has a BSc in Computer Science, with a minor area of study in Mathematics.

Database administrators (DBAs) have many tasks, for example creating, backing up and restoring databases, developing a security and disaster recovery architecture, performance monitoring and tuning and user management, etc. Another task that is often overlooked is disk space capacity monitoring and planning. Without disk space capacity monitoring, a database might run out of disk space, which might cause it to crash or stop working.

Disk space capacity management is just as important as the other DBA tasks. If disk space capacity planning is not performed, there may not be enough disk space to handle database growth, or there may be more disk space than is needed. DBAs do not want to be asked questions about why people cannot use their application because there is no more disk space, or why all the disk capacity that was purchased last year is not being utilised. DBAs are busy people and do not want to be required to monitor disk space manually on a daily basis. Monitoring disk space is a time-consuming task. For these reasons, consideration should be given to having an automated tool to gather and report on disk space usage.

Most database management software does an excellent job of managing databases, but it does not do disk space capacity management. Disk space capacity management is the process of monitoring currently available disk space and planning accurately for future disk space requirements based on historical disk space usage.

There are daily, monthly and long-range aspects of disk space capacity management. The daily aspect deals with whether there is enough disk space to handle the database space needs of today. The monthly aspect reviews the monthly growth and notes any abnormal growth patterns. The long-range aspect of capacity planning is the science of planning accurately for quarterly and yearly disk space requirements.

Daily Capacity Management

On a daily basis, the DBA will need to know how much disk space is available. There might be various events or occurrences that would cause large amounts

of disk space to be consumed unexpectedly. A programmer might consume a large amount of disk space purposely or accidentally by adding millions of rows to a database table, for example. Events such as these might cause the database to grow and consume most, if not all, of the available free disk space. If the DBA was monitoring the disk space usage daily, then any process that uses up most of the available disk space would most likely not go unnoticed.

So, how should daily growth be monitored? There are two methods: checking the system manually every morning to determine how much disk space is available on each of the drives on every database server, or devising an automated method of reviewing all disks on the system to determine how much free disk space is available. DBAs are very busy individuals, and monitoring manually takes time. It is therefore advisable that DBAs build or acquire a tool that will monitor available disk space automatically.

As a minimum, this tool should traverse all the database servers daily and review each physical disk for available free space, then notify the DBA as to how much free space is available. This disk space tool could review the available disk space hourly and, if it dropped below a particular threshold, notify the DBAs. This way, as soon as the database server was running low on disk space (one gigabyte, for example), the DBAs would be notified (by e-mail). This would allow the DBAs to determine how best to handle the low disk space availability before some problem occurs as a result.

If appropriate capacity planning is being carried out, DBAs should rarely be notified that disks are running out of space. This normally only happens when some process has gone astray or a clean-up routine has been broken.

Long-range Capacity Planning

Another aspect of disk space capacity planning is the need to plan accurately for the disk space requirements to support the database for the next year or so. To do this, a DBA needs to gather disk usage information so that they can perform long-term disk space capacity planning.

If disk space usage has been tracked over time, it should be a fairly easy task to predict how much disk space the database might need over the next year. However, if no information has been collected about how much the databases have grown over the past few years, it will be more difficult to determine how much space might be needed in the future.

Once the decision has been made to track disk space consumption, it is necessary to determine exactly what to track – what important disk space information might be needed from a capacity planning perspective. Do you track the amount of space all the databases use, the amount of space a single database uses or the amount of space each table within a database uses? Do you track only the allocated space, or also the amount of used space? Is there disk space consumed by things such as raw text data and/or database back-ups that support the databases but are not actually part of the database that should also be monitored and tracked? By answering these questions, it should be possible to determine what information is important to be tracked for capacity management.

Once the decision has been made as to what should be tracked, the next consideration should be granularity – how often hourly, daily, weekly, etc. Ideally, information should be tracked at the smallest granularity possible for future reporting requirements. This allows for the most flexibility. Statistics can always be rounded up to a more summarised amount, but the reverse is not possible. The granularity should depend on the kind of database activity exhibited. If the database growth rate changes on a daily basis, then tracking disk space usage at daily intervals will show this daily activity. A frequency should be chosen that reflects the environment.

Building a home-grown tool can determine the amount of disk space used and allocated by each table in every database. Running this home-grown tool daily, the daily space usage information collected is then stored in a series of database tables. These tables can be used to develop personal trend reports to help with monthly reporting and long-range disk planning. This method allows the most flexibility and allows the user to track only what is needed and then store it in a format of their own design.

Monthly Disk Space Reporting

Disk space growth statistics can be used to produce monthly management reports. These reports are typically some form of chart that represents the visual aspect of disk space growth at the server level. By creating a chart for each server, it is easy to see the disk space growth rate of the servers. If, for a particular month, there is unusual growth that produces a peak or a valley on the chart, then the

cause of the spike should be documented on the graph. This documentation makes it possible to note, over time, the events that might lead to significant disk space growth. Documenting these occurrences assists DBAs with future planning for disk space.

Every time a database is reorganised, a significant amount of disk space is recovered. If the budget is small and space is running low, then performing a database reorganisation may free up some disk space. This kind of disk usage information is invaluable for making the right decision when trying to free up enough disk space to make it to the next disk purchase.

Other Benefits

Disk space usage information can be used for more than just space management. It might be valuable for a number of other uses, such as billing or planning the disk space requirements of future projects. When new projects are in the planning stages, it is important to ensure that consideration is given to how much disk space will be required initially and what the monthly growth rate might be. Planning for upcoming projects is extremely important so as to make sure that there is the available disk space when the project needs it. If a relationship can be established between the amount of data a new application will store, with the amount of disk space an existing application is currently using, then the disk space statistics gathered should help with projecting how much disk space a new project will require.

Another possible benefit for disk space usage statistics might be to compare the disk space usage between applications. Pie charts can be used to show management of the disk space usage relationship between applications. This information might be useful when determining how to fund the next disk space acquisition or how to handle monthly disk space billing.

Conclusion

Disk space capacity management should be considered just as important as all the other daily, monthly and yearly DBA tasks. Disk space statistics give the DBAs a good idea as to how the databases in an organisation consume valuable disk space resources. Whether a software package is purchased to gather space usage information, or a home-grown solution is built, it does not matter. All that is important is that a space usage tracking solution is found that meets the organisation's monitoring needs.

Without disk space capacity planning and monitoring, there is the risk of running out of disk space, and possible database outages, and no DBA wants to find themselves in that situation. ■