How Fast Is Our Data Volume Growing?

by
Steve Mackie

We all know the volume of data is growing. The question is how much data and how fast? IDC reports that enterprise data stores will grow an average of 60 percent annually. This number fluctuates dramatically depending on the enterprise and how business processes depend on and use data. Media intensive industries such as entertainment broadcasting, medical, legal, and insurance can expect to see data growth rates over 120 percent year on year while some brick and mortar industries may only see as much as 50 percent growth year on year. Knowing your organization’s growth rate is the important thing. All data storage management policy decisions and acquisition plans depend upon knowing how fast the data volume is growing.

It is also very important to understand the growth rate by type of data. Examination of volume growth trends by type can help storage managers identify possible system problems and anomalous user behavior. For instance, if the volume of system log files takes a sudden upward turn, it may indicate an application or system device (e.g. server or appliance) is malfunctioning. Many experienced system administrators have stories to tell about servers or workstations that crashed when all internal system disk space was filled up by a log file being written by a failing application or component. Likewise, a sudden upturn in database table and index volumes could indicate a database administrator is creating numerous unsafe duplicates of databases. This could be a perfectly normal part of testing and development or an unsafe unauthorized behavior in a production environment. While the storage administrators may not have direct responsibility over the issues their reports identify, they can be extremely helpful to other system administrators and IT management in maintaining a holistic awareness of the health of the data and application hosting environment and behavior of the users.

Data storage managers should collect and regularly report to management the following kinds of information: data volume by type/user/application/device, volume growth trends over time, and data aging (last accessed). Data volume growth trend charts are vital to management determining when additional capacity will need to be procured and implemented. Trend analysis of this data over the long term is vital to determining long term IT capital budget requirements.

Not all data is worth keeping. Data storage managers should keep a weather eye on opportunities to reduce the data volume through elimination of unnecessary or duplicate data and archiving. Investigation into utilization report anomalies and providing feedback to end users can help identify data that can be deleted or archived. Storage space reclamation is an important part of data storage management and should be addressed in a formal procedure if not policy. Procedures should ensure that storage managers never destroy data without consulting with affected end users and policies should require end users to provide feedback to data storage administrators when queried.

IT management can help abate the unstrained growth of data through user education and policy. For instance users should be trained and encouraged to leverage corporate shared directories, intranets, or collaborative work flow tools (e.g. MS SharePoint) rather than sending common use internal files as email attachments. Corporate policies restricting the use of the network for the proliferation of personal entertainment videos, pictures, and audio can significantly reduce the data storage burden, ease the strain on the email application, and most importantly help improve system performance.
IDC reports that data storage accounts for as much as 15 to 20 percent of IT capital spending in large enterprises. The chart below depicts the relative cost of data storage over the last ten years for industry types with various data growth profiles. Despite the fact that the relative cost of computer data storage media per unit volume has fallen 63 percent since 1998, the overwhelming data volume growth is causing storage costs to grow rapidly.

Industry has responded to the market’s need for more intelligent storage of data. Identification and elimination of duplicate data has entered the main stream and proven to be very helpful in economizing on storage infrastructure utilization rates. Vendors are claiming data reduction ratios anywhere from 5:1 to 90:1 depending on whether in primary or secondary storage mediums.

Significant savings can also be generated through employing a tiered storage model. Tiered storage involves categorizing data by criticality and establishing tiers of storage media to accommodate each level of data appropriate to each tier’s performance requirement. In order for IT management to make informed decisions about storage strategies and how to meet the growing demand for storage, they have to know how much data they have and how fast is it growing.

Storage Strategies, Inc. (SSI) is a data storage engineering and consulting firm that provides IT managers with subject matter expertise to help them define their true data storage requirements and craft IT capital investment strategies for meeting those requirements in an economical fashion.
How Fast is Our Data Volume Growing?