

# A New Approach to "Storage Management" Restrictions Using the "Data Quality" Concept

Koby Biller

Disklace Ltd., Rehovot, Israel

koby@disklace.com; <http://www.disklace.com>

## Abstract

Fragmentation of data in storage devices, a phenomenon that has attracted considerable research attention<sup>1-3</sup>, is a major problem underlying computer slowdown and other unpredictable storage-related symptoms. This poster presents the "Data Quality" approach, which includes an effective new method for measuring the quality of data affected by continuous fragmentation. It also incorporates a novel way to handle the affected data, with clear advantages over the current alternatives of replacing the device or upgrading the computer. As an added benefit, use of this approach allows a general picture to be obtained of the data quality throughout an organization.

## 1. Introduction

Computer disk fragmentation is an inevitable feature of the normal data life-cycle. It enables almost infinite usage of disk resources, until a point is reached where users feel that their computers have begun to work slowly, but without any indication to that effect from the operating system. Attempting to avoid this problem by manual maintenance is often futile. *Disklace Ltd.* has investigated the phenomenon of disk fragmentation and introduced a comprehensive solution that improves data life-cycle in both current and future storage devices. This is achieved by employing a novel technique that measures data quality and implements a concept of optimized preventive maintenance. The result is a prolonged computer lifespan, easier maintenance, and improved user satisfaction. After being thoroughly tested by the industry, this Data Quality model has been adopted by a large software vendor and is already operating in a laboratory at a well-known research institute.

## 2. Measurement method

The measurement formula shown in the poster is based on a statistical second-level moment calculating the fragmentation of the files and of the free space on the storage device. It is shown on a logarithmic scale, in which a higher number reflects a worse situation. The units are natural and are not related to the size of the storage device.

## 3. Distribution of the results

Analysis of the first measurements yielded a normal Gaussian distribution (Fig. 1), proving the initial hypothesis. The hat-shaped distribution persisted after

more than 100,000 measurements, supporting the notion that the population was not an incidental one.

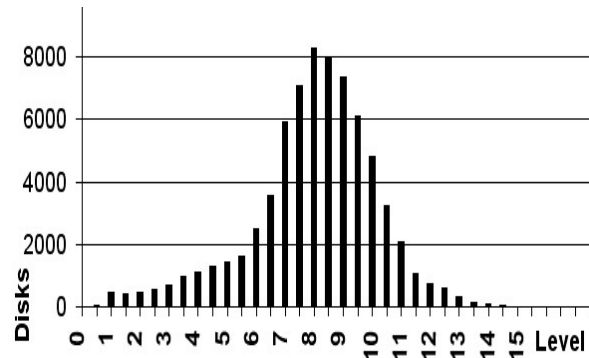


Fig. 1 Fragmentation measurement distribution.

## 4. Maintaining a constant fragmentation level

Maintenance of a storage device (defragmentation) with a high level of fragmentation yielded poor results. We therefore suggest that the data quality be maintained automatically once a defined threshold is exceeded. This stage provides an additional feature complementary to the current storage management methods, and completes the life-cycle of data management in the long run.

## 5. Implementation

**LaceWatcher PRO**, a product developed to apply the concept, enables the method to be implemented immediately and allows for continuous monitoring of the distribution of fragmentation results. This practice is especially important for the new solid-state devices and new technologies.

## References

- [1] BHADKAMKAR, M., FARFAN, F., HRISTIDIS, V., AND RANGASWAMI R. Storing semi-structured data on disk drives. *ACM Transactions on Storage* (2009) 5(2).
- [2] TABERO, J., SEPTIEN, J., MECHA, H., AND MOZOS, D. Allocation heuristics and defragmentation measures for reconfigurable systems management. *Integration, VLSI Journal* (2008) 41(2):281-296.
- [3] SHORE, J.E. On the external storage fragmentation produced by first-fit and best-fit allocation strategies. *Communications of the ACM* (1975) 18(8):433-440.

A full functioning demo of **Lacewatcher-PRO** for windows supporting 1TB storage and partitions up to 80GB is available at <http://www.disklace.com/SYSTOR10-demo.html>