
Diaser

An open-source, data-vault application

Revision: 1.1



Creative Commons License

This work by Simon Hettrick and Damian Brasher is licensed under a Creative Commons Attribution-Share Alike 2.0 UK: England & Wales License.

Author: Simon Hettrick

List of Revisions

Revision No.	Author	Changes Made	Date Revised
1.0	Simon Hettrick	First draft	21/08/09
1.1	Damian Brasher	Review of draft	24/09/09

Title: Diaser	Version: 1.0
	Date: 21/08/09

Table of Contents

Diaser.....	1
An open-source, data-vault application.....	1
Introduction.....	2
<u>1.1 Overview of Diaser.....</u>	<u>2</u>
<u>1.2 Support.....</u>	<u>2</u>
<u>1.3 Licence.....</u>	<u>2</u>
<u>1.4 Installation.....</u>	<u>2</u>
Strengths.....	3
Title: Diaser.....	2
Operation.....	4
<u>1.5 An overview of Diaser's operation.....</u>	<u>4</u>
<u>1.6 Security.....</u>	<u>4</u>
<u>1.7 Archival procedure.....</u>	<u>4</u>
<u>1.8 Archival of full and differential volumes.....</u>	<u>5</u>
<u>1.9 Monitoring.....</u>	<u>5</u>
<u>1.10 Storage limits.....</u>	<u>5</u>
<u>1.11 Retrieval.....</u>	<u>5</u>

Title: Diaser	Version: 1.0
	Date: 21/08/09

Introduction

1.1 Overview of Diaser

Diaser archives the information stored on your system and ensures long-term data security at minimum cost.

Backing up the information stored on a system is an absolute necessity, but leads to a difficult problem. The long-term storage and organisation of back ups – a process known as *archiving* – is time-consuming, prone to mistakes and expensive. Diaser is the solution to these problems.

Diaser automates the archival process and allows the recall of back ups across your organisation's network, thus bypassing the labours involved in running a traditional tape archival system. Diaser automatically distributes archives across a three *nodes*. This multiple redundancy ensures that your archive is practically impervious to loss. Diaser uses off-the-shelf equipment, which minimises hardware costs and ensures that system maintenance is straightforward. Diaser is open source, so it is free to use, and uses open standards, so you can try Diaser without worrying about being locked in to its use in the long term. Unlike data centres, and other cloud-based archiving solutions, with Diaser your information is stored on resources that you own. This means that you never lose ownership of your information.

1.2 Support

Further information can be found on the Diaser website: www.diaser.org.uk.

Support with a specific issue can be requested through the support mailing list, which can be found at: <https://lists.sourceforge.net/lists/listinfo/diaser-support>

1.3 Licence

Diaser is distributed under the *Gnu General Public License, version 3.0*, the terms of which can be found at: www.gnu.org/licenses/gpl-3.0.html.

1.4 Installation

Diaser is easy to install. The installation procedure is covered in the quick-start documentation available on the Diaser website: <http://www.diaser.org.uk/about.html>.

Title: Diaser	Version: 1.0
	Date: 21/08/09

Strengths

- **Secure:** data transfer is encrypted and all nodes are secured with passwords and certificates
- **Reliable:** archiving occurs automatically and without user intervention so an archive can not be overlooked or forgotten
- **Low cost:** the software is free, uses standard hardware, requires a minimum of user interaction and archiving can be set to operate when network usage is low (typically at night)
- **Private:** information is stored on your own servers thus eliminating data-ownership issues to faced by other archival systems
- **Robust:** multiple redundancy ensures long-term data safety
- **Scalable:** simply upgrade the storage on each node to increase the volume of data that Diaser can archive
- **Easy:** Diaser is quick and easy to install and run

Title: Diaser	Version: 1.0
	Date: 21/08/09

Operation

1.5 An overview of Diaser's operation

To ensure long-term security of data, it is necessary to back up the files stored on an organisation's workstations and servers. Back ups are typically made at regular intervals – every day, every week or every month – causing a rapid growth in the volume of data that must be stored. Many organisations are required to store back ups for a long period of time – potentially many years. *Archiving* (the long-term storage and organisation of back ups) by conventional methods is often expensive and laborious. Diaser archives the information stored on your system and ensures long-term data security at minimum cost.

Diaser automatically takes the back-up file, known as a *volume*, which can be created by any of the standard back-up packages (Bacula, Microsoft Backup, etc.) and archives it onto three *nodes*. The same volume is stored on all three nodes, which means that the data is safe even in the very unlikely scenario where two of the nodes fail catastrophically. If data from a particular volume is required, a user simply connects to any of the three nodes and extracts the information that is required.

1.6 Security

Security is a fundamental concern of Diaser. It meets the most stringent security concerns of any organisation, namely:

- Encryption: Diaser operates over OpenSSH so all data transfer between nodes is encrypted
- Authentication: unauthorised access to the Diaser nodes is prevented through a system of passwords and certificates
- Emergency lock: Diaser features an emergency lock, which can be used to lock down the system if a node is compromised

Additional security measures can be added to the Diaser system. It is good practice to encrypt back-up volumes as they are produced, which prevents unauthorised access to the information held within the volume. Most back-up software has this facility. Diaser may also be implemented over a *Virtual Private Network* (VPN), which adds an additional layer of security to the Diaser system.

1.7 Archival procedure

The transfer between the back-up server and the three nodes is a two-phase process. In the first phase, a volume (i.e. the back-up file) is transferred from the back-up server to Node A, and then transferred from Node A to Node B. In the second phase, the archive is transferred from Node B to Node C. When phase two has completed, the same volume has been copied onto all three nodes. This multiple redundancy means that the system is practically impervious to data loss.

During installation, the user is asked to specify the times at which the first and second phases occur. The two phases can be selected to coincide with times of low network usage, which ensures that the maximum bandwidth is available on the network. With most organisations, the network typically experiences lowest usage at night when no one is using the network. Each node operates independently at specific times (the nodes are time-synchronised using NTP) as directed by its own set of instructions, known as the *hypervirtual autochanger* scripts, which are generated according to user-specified settings during installation of Diaser.

Title: Diaser	Version: 1.0
	Date: 21/08/09

1.8 Archival of full and differential volumes

There are two methods for backing up a system. The first is a *full* back up, which is a back up of everything stored on the system. The second is a *differential* (or *diff*) back up, which is a back up of only the files that have changed since the last back up. A full back up ensures that all files are safely backed up but is significantly larger than a differential back up. Most system administrators perform both full and differential back ups with a frequency to gain data safety without overtaxing the data storage they have available. Diaser can archive a mixture of full and differential volumes, or exclusively full volumes (as would be produced by a CCTV system, for example).

1.9 Monitoring

Each node generates a log file whenever it performs a process. This information can be used to monitor the Diaser system. In the beta release of Diaser, it is necessary for the user to either access the log files directly or write a script to extract data from them.

The first stable release of Diaser will include a web-based monitoring system that will automatically access the log files and prepare reports.

1.10 Storage limits

The maximum volume of data that can be stored by Diaser is a product of the bandwidth available on your network (i.e. the amount of data that your network can transmit simultaneously) and the *transfer time*, which is the time available to transfer data across the network for purposes. The table below provides an overview of the storage capacity available for common bandwidths and transfer times (abbreviated to *tt* in the table).

Bandwidth (Mbit/s)	Storage limit (Gbyte)					
	tt = 1 hour	tt = 2 hour	tt = 3 hour	tt = 4 hour	tt = 5 hour	tt = 6 hour
1	0.45	0.90	1.35	1.80	2.25	2.70
10	4.5	9.0	13.5	18.0	22.5	27.0
100	45	90	135	180	225	270
1000	450	900	1350	1800	2250	2700

Alternatively, the approximate storage available on a system can be calculate using the following equation:

$$\text{Storage Capacity (in Gbyte)} = \text{Bandwidth (Mbit/s)} \times \text{Time available for transfer (hours)} \times 0.45$$

Note that the actual storage value may be different to that calculated as the bandwidth of your network may fluctuate.

1.11 Retrieval

In the beta version of Diaser, retrieval of a volume from a node is performed using a provided tool. This requires a file-sharing tool, such as SCP, which is used to copy the required volume from any of the three nodes back to a user's computer. Once the volume has been retrieved, the back-up software that was used to generate the volume can be used to extract the data from within the volume.

The stable release of Diaser will include a retrieval client which can be used to access any of the three nodes and retrieve an archive.