DVS Whitepaper

Spycer® – a distributed content management system for post production departments

2008

1. Preface

Digital media technology resulted in a lasting change in post production workflows. In digital post production physical "storage media" such as film rolls and video tapes recede in importance. Instead uncompressed image sequences in high resolution (e.g. in 2K or 4K) can be found as huge amounts of data stored on various storage solutions, and it is not uncommon that a blockbuster movie requires about 100 terabytes of storage. During post production a lot of different files exist, mostly distributed on several storages. Data management in such environments is a big issue when relying only on standard "tools" like notes, filing cards, labels, and hierarchical file structures with in-house directory naming conventions.

In the broadcast area the migration to HD has increasingly gained significance. At the same time, we see a migration to completely file based workflows. HD camcorders capture the material on various storage media. The video material is transferred faster than real time to large storage systems (e.g. SAN) which must handle not only video files with higher resolution than SD but also have to cope with a number of additional files, because the content will be provided not only for TV via cable or satellite but also for the internet and mobile applications with various resolutions and different data rates. These conditions result in multiplying the available and deployed content.

In principal an efficient file management is possible with metadata, referring to the stored content, and intelligent data management software. However, most of the available software solutions are not flexible enough to account for the different workflows in the film making industry. DVS presents an alternative approach with a distributed content management network and an easy-to-use software application called Spycer®. This whitepaper explains the concept of a pure distributed content management network with Spycer®.

Feel free to contact us in case of questions.

Kind regards,

The DVS Product Management

Contents

2. Situation in the post production industry

It is fair to say that the current post production process is based on files. But in the film making business this is mostly the case for the essence – i.e. audio and video data – only. During the content creation the creative staff nowadays has to handle a lot of files. However, browsing through big amounts of files and searching for the required ones should not be the major tasks of the staff. Unfortunately, in a standard network environment with several storages exactly these tasks are very time consuming because the standard search functions of current operating systems are not optimized for a post production workflow. Moreover, metadata mainly exist on paper or whiteboards only.

Image 2 (Copyright: DVS)

In order to overcome such problems various different Content Management Systems (CMS) are available. But such solutions are often developed for the special demands of the broadcast industry and not for movie post productions. Most CMS are intended to be long-term installations with a central database and a central storage, and the manufacturers of those systems rely on the assumption that workflows remain static and the maximum number of clients in the network is known.

Workflows in the post production process of a movie differ from the workflows of broadcasters. Often several departments use “their own” storage solutions. Workflows, and with them the number of workstations and storages in use, may even differ between the different creative phases. Thus, a very flexible CMS is required.

3. Requirements in the post production business

Users want to know at any time where an individual clip is located and they want to have easy access to its metadata. This makes efficient data tracking and data retrieval a must. Therefore, all typical file formats and their metadata must be supported in a CMS. Such a system has to support DPX file sequences with their header data, for example. Furthermore, a CMS should be flexible when dealing with different workflows and should not compel users to use only one specific workflow for the data management.

4. Situation in the broadcast industry

The migration to HD causes the broadcasters to enlarge the available storage space. This is especially true for the post production departments, which are using mostly uncompressed material for their high-quality composites. Moreover, the same finished material has to be reused for internet streaming and mobile TV platforms which is multiplying the amount of files on the various storage solutions.
For a fast relief additional storage solutions are added to the existing systems, building a confusing mass of different storage systems. Different storage solutions have to be managed and system administrators often wish to have a data management application which is running on all relevant operating systems. The desire to somehow connect all storage islands with each other and to gain more transparency over them is high. It becomes clear that gaining overview of all available material becomes more and more important when enlarging a facility's total amount of storage.

5. Requirements of post production departments in the broadcast business

Standard IT systems are used for the post production of broadcast content. Even if central storage systems are used in order to optimize day-to-day workflows and to keep the amount of copying processes low, there will always exist additional storage solutions and transfer processes between various systems.

If there are several storage solutions in a facility, it is extremely important to have data management and retrieval software at hand. Standard data management software for the IT business is not optimized for professional content creation. Search tools must provide typical metadata used for media files and search results should provide a preview picture.

Common copy processes of operating systems are not optimized for fast media file transfers and especially when using file sequences (e.g. results from a 3D rendering process) files get spread all over the storage slowing down the performance. Copy processes should not compromise the storage systems in that way.

6. Spycer® – a data management solution

With these requirements in mind DVS developed Spycer®, a software capable to find clips in several hundreds of terabytes or even petabytes of data, spread all over various storages. In contrast to a common CMS, which provides web-based clients only in order to browse a central database on a server, Spycer® is an active software that can be installed on every workstation involved in the post production process. It extracts metadata automatically as soon as a new file is stored on the local storage and provides this information to the SpycerNet. This distributed content management network operates without a central database or server.

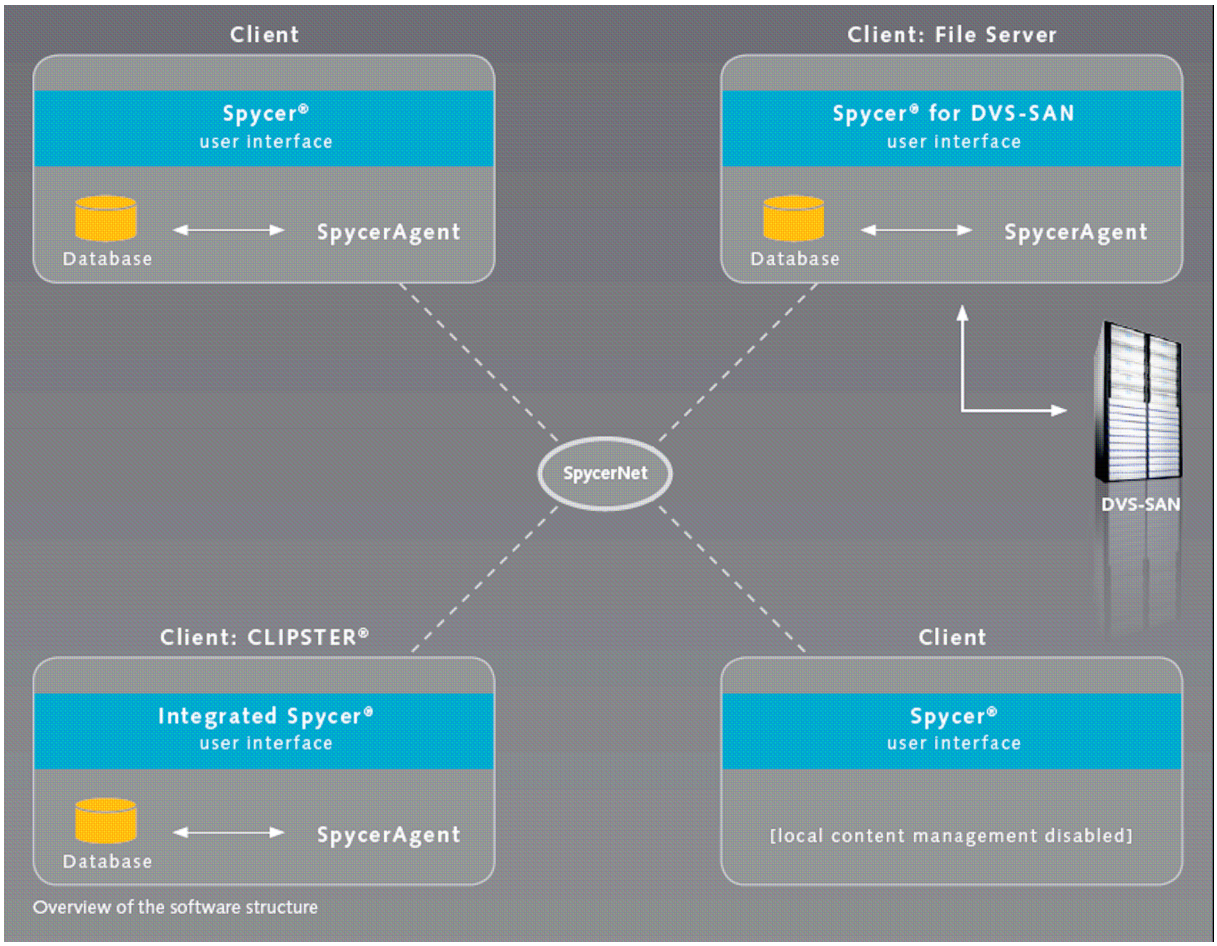


Image 3: SpycerNet with four Spycer® applications on different PCs. Every PC has a Spycer® user interface and the possibility to provide its own file information to the SpycerNet. (Copyright: DVS)

6.1 Network technology

It is hardly ever the case that a single central storage solution is the only storage where all data of a post house is stored and that it can be directly accessed from everyone involved in the project. In the majority of cases post production facilities deal with several islands of storage solutions for security, performance or economic reasons.

In a traditional IT infrastructure a central storage and database are used to administer all information. The problem is that many media files never get recorded in the database. This is especially true if an ordinary client-server structure forms the basis for the content management system (CMS). In such a case the client software provides just a Graphical User Interface (GUI) for the database of the CMS. The user has to enter all data manually. For automatic metadata extraction the files have to be stored only on the central storage system where background processes running around the clock can continuously update the database. These types of systems are often limited in scalability and performance due to a single server and database being responsible for all files. Additionally, you may get a more serious problem if the central database were suddenly offline. This single point of failure can not happen in a pure distributed solution.

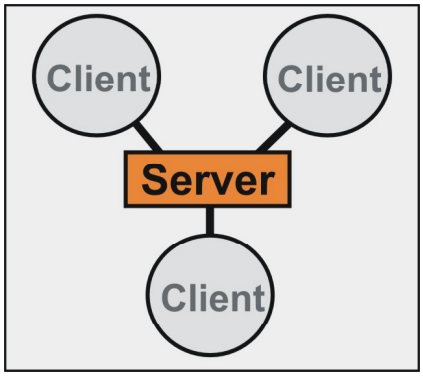


Image 8: Client-server system – a centralized approach (Copyright: DVS)

What if the database and active CMS applications were distributed on several workstations? Imagine a distributed content management solution where every workstation automatically provides its own database that can be shared with applications of the same kind in a network (aka peer-to-peer network). Every file gets a record in one of the databases, thus getting available to all others in the company network. This is how a scalable system should work. According to the rights assigned to an individual user, the user is able to find any media file in the network effortlessly by using common metadata information. Spycer® works exactly this way and thus grants complete content control.

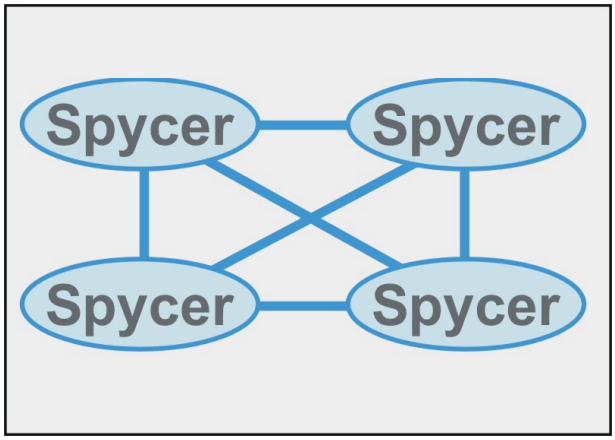


Image 9: The SpycerNet – a truly scalable distributed content management with several virtual connections (Copyright: DVS)

Via the SpycerNet several Spycer® applications communicate with each other for browsing, data retrieval and file handling purposes. Metadata is automatically extracted and can be used for searching and the retrieval of data in the network. The distributed content management system grows with every additional Spycer® application added to the network.

6.2 Fragmentation of image sequences

In a digital intermediate process or digital cinema post production, clips are stored as uncompressed file sequences. This means that large amounts of files are stored on any storage involved in the workflow. Most storage solutions do not provide efficient defragmentation software for these file sequences. Common defragmentation applications just defragment single files and are "not aware" of a possible relationship between files. For a real-time play-out it is inadequate to have the files individually defragmented only. Instead the whole sequence of files belonging to a clip has to be stored contiguously in large blocks on the storage and not scattered all over it.

File Format	Defrag	Size	Resolution	Color Depth
DPX	99%	1.02 GB	1920 x 1080 Pixel	10
DPX	0%	2.32 GB	1920 x 1080 Pixel	10
DPX	3%	3.56 GB	1920 x 1080 Pixel	10
DPX	100%	395.90 MB	1920 x 1080 Pixel	10
YUV8	100%	197.75 MB	1920 x 1080 Pixel	8
DPX	100%	1.37 GB	1920 x 1080 Pixel	8
DPX	100%	594.14 MB	1920 x 1080 Pixel	8
DPX	100%	594.14 MB	1920 x 1080 Pixel	8
DPX	100%	534.73 MB	1920 x 1080 Pixel	8
DPX	100%	475.08 MB	1920 x 1080 Pixel	10
BMP	100%	183.94 MB	1920 x 1080 Pixel	8
BMP	100%	282.75 MB	2048 x 1556 Pixel	8
DPX	100%	1.16 GB	1920 x 1080 Pixel	10
DPX	100%	1.36 GB	1920 x 1080 Pixel	10
BMP	100%	995.97 MB	1920 x 1080 Pixel	8
DPX	100%	942.24 MB	1920 x 1080 Pixel	10

Image 4: Spycer® shows the current defragmentation status of scattered sequences. (Copyright: DVS)

Spycer® is able to defragment any file and to analyze and de-scatter file sequences. A monitoring process updates the current defragmentation status of every sequence and keeps the user up-to-date.

6.3 File handling

Users require dependable copying and renaming processes for such masses of data. High-speed copying mechanisms should be easy to use for the creative staff. This means that drag'n'drop procedures must be available instead of complicated command line tools. DVSCopy, a fast copy tool that avoids fragmentation, is integrated in the application of Spycer®. Users will see two browsing panes which may show, for example, the source and destination storage of the files for easy drag'n'drop procedures.

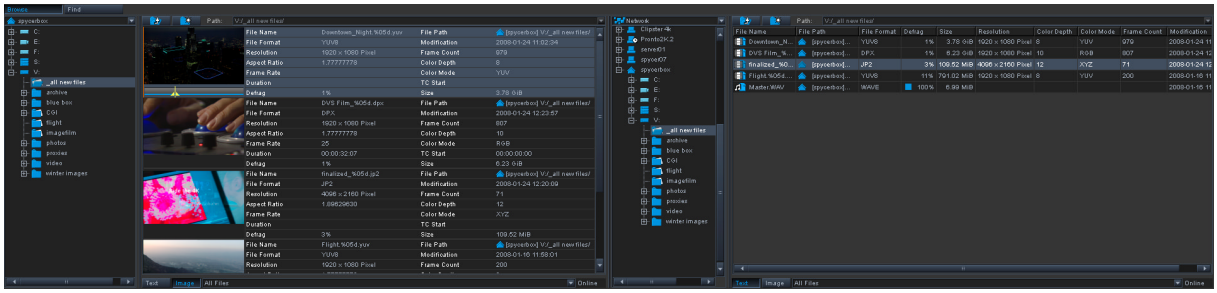


Image 5: Browsing panes with two directories in parallel enable easy drag'n'drop tasks. (Copyright: DVS)

With a built-in renaming function for file sequences it is very easy to change the index number or the whole naming pattern of the sequence. Spycer® goes through all files and renames them according to the given pattern automatically.

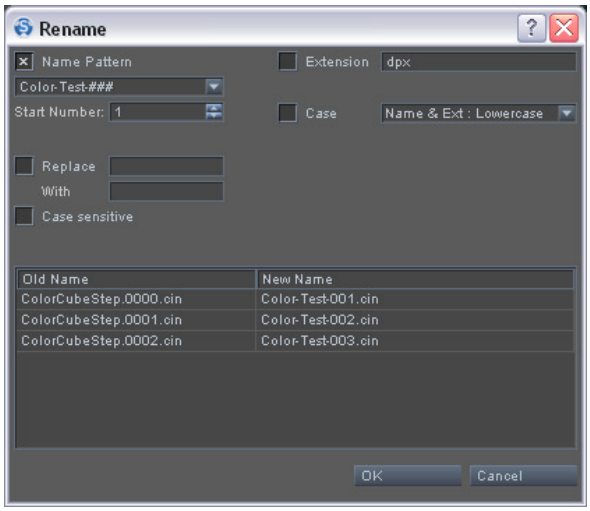


Image 6: File sequence requires a good and comfortable renaming function. (Copyright: DVS)

6.4 Finding the content

For data handling and tracking, users need intelligent software which monitors every change made in a directory. It is not always possible to use an ideal hierarchical directory structure for a project where everyone can find the required material on big central storages easily. In order to find a certain file, instead of browsing through several directories manually, it is faster to use a search tool which provides specific attributes for a search query. This is a new working paradigm. In Spycer® a powerful search engine is implemented where you can specify important metadata attributes for a search query to localize data fast in the whole network.

Every Spycer® application monitors its own pre-defined "watch folders" and provides automatically extracted metadata to the SpycerNet. A distributed content management network consisting of individually installed Spycer® applications is an ideal tool for post production where content is spread over different storage systems.

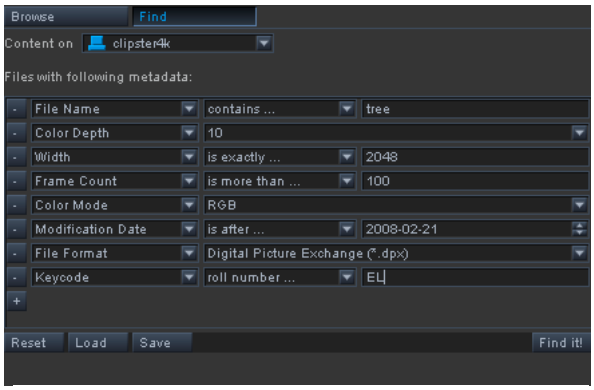
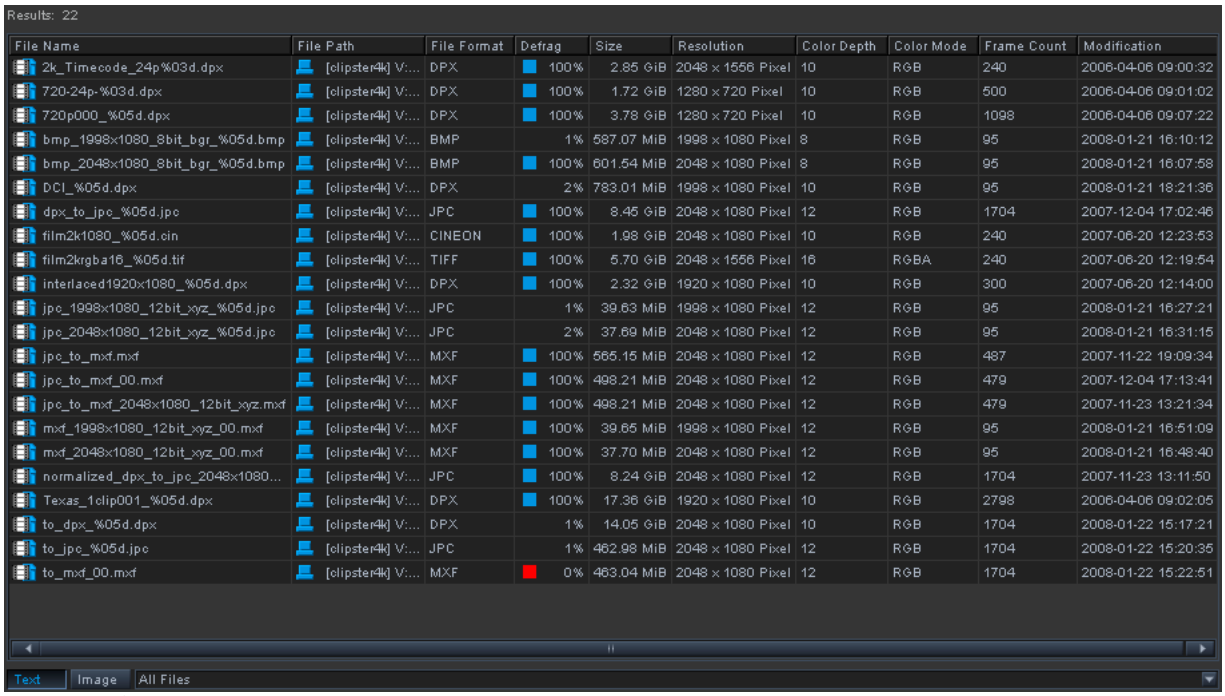


Image 7: Stored metadata help to find the right content. (Copyright: DVS)



Sorted List of the search results

6.5 Previewing the Content

Large lists of text are not enough for creative staff. Users in post production need thumbnails and proxy clips of the material which are stored everywhere in their facility.

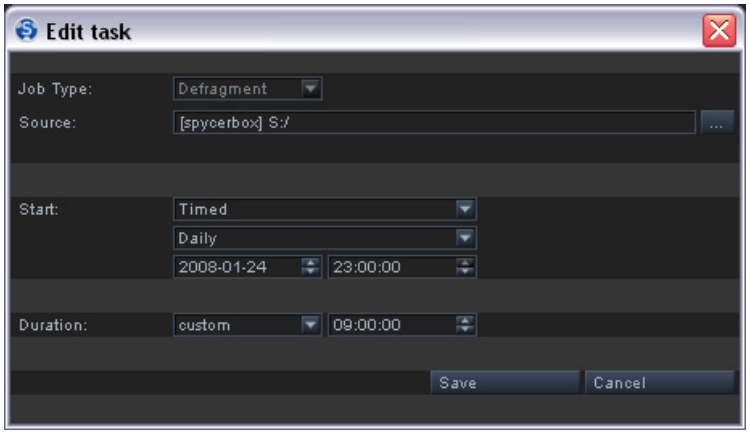
In the SpycerNet preview images are created on the fly, or low-res proxies are used for preview in the preview area. Frame number, timecode or keycode can be shown if available.

Scrub through clips or play back preview proxies with superimposed Keycode information via SpycerNet

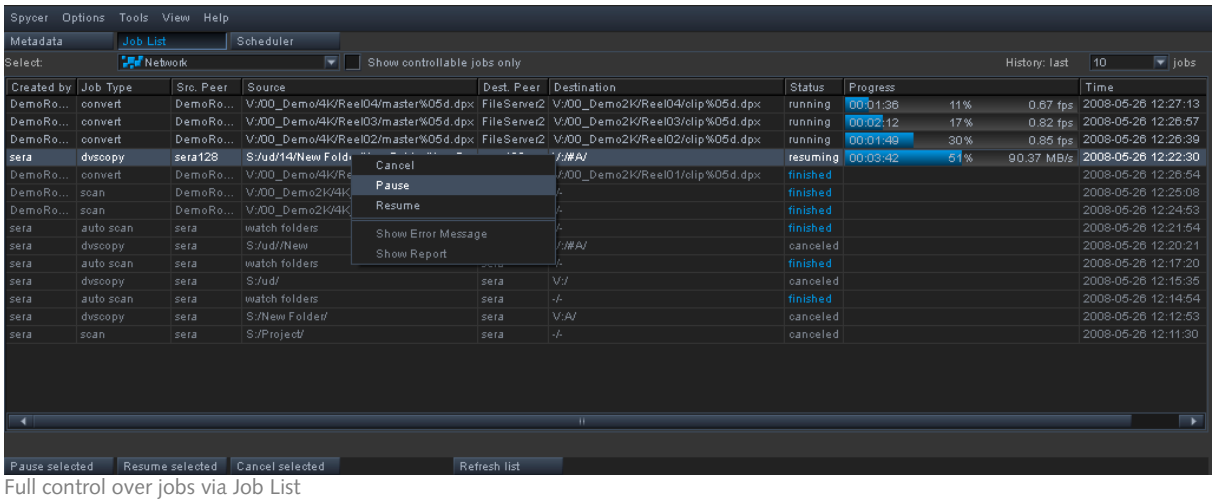
6.6 Administrators need control

In order to help users and administrators who are responsible for the post production infrastructure, an easy and fast access to processes in the content management network is needed.

Spycer® lists all current "jobs" in the SpycerNet. They can be paused, resumed or cancelled via user interface. With the built-in scheduler the administrator can easily set up periodical data management tasks on any peer in the SpycerNet.



Administrative tasks can be scheduled on any peer in the SpycerNet



Full control over jobs via Job List

7. About DVS

DVS Digital Video Systems AG is a premium manufacturer of high-quality hardware and software for professional film post production. Both CLIPSTER®, the company flagship, and Spycer®, the intelligent data manager, have won several prestigious awards. With its DI systems, DVS is the first company in the world that makes real-time 4K-processing possible. Thanks to DVS's long experience in the field of post production, the company is optimally prepared for the switchover to HDTV in the broadcast sector, and is equally active in the D-Cinema arena - since 2007, all DVS products meet the required standards. In addition to the company headquarters in Hanover, DVS has branch offices in Los Angeles, California, Miami, Florida and Paris, France as well as Singapore.

7. References

DVS: http://www.dvs.de

Spycer®: http://www.spycer.net/

8. Glossary

2K
Abbreviation for resolutions of image and video, e.g. 2048 x 1536 and 2048 x 1556 pixels. This resolution is recognized as the minimum resolution that is suitable for cinema presentations. In order to obtain other aspect ratios or to be compatible with technologies like HD-SDI, further formats like 2048 x 1080 and 2048 x 858 are common. As 2K is used for film production mainly, the color coding is mostly RGB 4:4:4, which gives a better quality than YUV 4:2:2.

4K
Abbreviation for an image and video resolution of 4096 x 3112. This resolution is recognized to be of the same quality like a high-quality cinema presentation. To obtain other aspect ratios or to be compatible with technologies like HD-SDI or DVI, further formats like 4096 x 2160 and 4096 x 1714 are common. The 4K resolution is higher than the one of a HDTV signal. As 4K is used for film production mainly the color coding is mostly RGB 4:4:4 which gives a better quality than YUV 4:2:2.

Client
A computer system that wants to access a service – sometimes a remote one – on another computer is called a client. Typically this happens within a network.

Client-Server Architecture
Network structure which separates server applications from client applications. A central server manages all data for different clients and provides them with the required data. The system's scalability depends on the server performance and the expandability of its hardware resources.

CLIPSTER®
CLIPSTER® is a turnkey solution by DVS. It is a one-stop Digital Intermediate solution for conforming and finishing uncompressed SD/HD/2K/4K data in any workflow. Moreover, this DI workstation carries out real-time effects, enables multi-resolution and is an open platform. CLIPSTER® offers stunning hardware power and innovative software for unrivaled flexibility and can be used in any video or film post-production environment. The high performance is to be seen in its real-time effects with up to 2 x 2K RGB 12 bit, its real-time playback of 4K RGB 10 bit DPX file sequences and its support of multiple video formats with real-time converting. Additionally, CLIPSTER® can handle real uncompressed video up to 4K RGB 16bit and runs real-time effects in 16 bit, with original native content being used for real-time processing. CLIPSTER® is an open platform: the Windows® XP workstation captures directly to NTFS and it possesses real-time support of graphic file sequences like DPX, TIFF, Cineon®, TGA, BMP, etc. Of course, an OpenFX plug-in interface is part of CLIPSTER® as well.

CMS
A Content Management System is a software that helps storing content and tracking changes made by users. It supports the organization of content via a database.

DPX
Abbreviation for Digital Picture Exchange. This file format can be found in digital film work and is considered an ANSI/SMPTE 268M standard. DPX files can store image data and additional metadata in their file header.

DVS-SAN
DVS-SAN is a high-performance central storage system for uncompressed video and digital film. It has been designed for use in the film and HD postproduction environment, where large amounts of data must be accessed in real time by multiple workstations. DVS-SAN meets the special demands of digital intermediate work and HD projects requiring ultrahigh data rates and fast access times. With DVS-SAN, several workstations can access the same data, concurrently and in real time, eliminating the need for copying and exporting. The system is upward-scalable to hundreds of terabytes and so offers enough storage capacity for several film projects.

Format
(1) The size, resolution, aspect ratio, color space, bit depth, format rate, etc. for a given image.
(2) The file format for a given image.
(3) The physical medium (such as film, video, etc.) used to capture or display an image sequence.
(4) A multitude of additional variations and subcategories of the first three definitions.

Fragmentation
(Data) fragmentation occurs when a piece of data in memory is divided into several parts being physically far apart. Generally, this is the result of attempting to insert a large block of data into several small free spaces on the storage.

GUI
Graphical User Interface. An interactive graphic displayed on a screen, being a means of operating a software.

Metadata
Data that describes other data. Generally, structured information that describes a (possibly unstructured) set of data. For example, a title can be a metadata item of a movie which is stored as a clip in a file. The frame rate and resolution of a clip are metadata items, too.

No single point of failure
Describes a configuration in which at least one of each component may fail without loosing the functionality or data of the system.

Operating system
Every computer needs a base program, the so-called operating system that manages the computer and grants control of various functions. Common examples are MS DOS and Windows® for PCs, Mac OS for Apple® Macintosh and UNIX for Linux®. On top of the operating system, specific applications are installed. General purpose operating systems allow a wide range of applications to be used, they do not necessarily allow the most efficient or fastest possible use of the hardware for the application.

Postproduction
The stage in the preparation of a film or video program after the original footage has been shot. Can include editing, encoding, computer program authoring, etc.

Real-time
Real-time is the idea or concept of a system that responds and reacts to signals as fast as they happen. One popular example is located in the games industry: when an operator moves a joystick and the video image on the screen seems to react at the exact same time, the processes that were needed to make the images move are said to be in “real-time”.

Resolution
In a reproduced image, a measure of the exact details can be seen or “resolved”. The number of the pixels in this image influences the display: high-definition is defined at approx. 2000 x 1000 pixels, SDTV at approx. 720 x 576 (PAL) or 720 x 487 (NTSC). Still, the number of pixels does not define the resolution itself, since the resolution is the result of the whole equipment interacting, i.e. the quality of the lens, the display tubes, film scanners, film processes, edit system, etc.

RGB
The abbreviation for the Red, Green and Blue signals, the primary colors of television. Cameras and telecines have red, blue and green receptors, the TV screen has red, green and blue phosphors illuminated by red, green and blue guns. Much of the image monitoring in a production center is in RGB. RGB is digitized with 4:4:4 sampling which occupies 50% more data than 4:2:2.

SAN
Storage Area Network.

Server
When a computer provides services to other computing systems (clients) over a network, it is defined as a server. Most complex computer systems today require a server, but the term can also refer to the software or hardware elements of such a system.

Spycer®
Innovative data management software from DVS that has been integrated in all DVS turnkey solutions. It provides a solution for dealing with large amounts of video data and its accompanying metadata. Spycer® provides editors, colorists and directors with a wide range of browsing, search and management tools. Spycer® offers more transparency in DI workflows since it assists by managing, searching and viewing content and metadata and also by browsing directories within the context of the current project. Fast data retrieval and high-speed copying are other key features of this application.

9. Contact

Feel free to contact us:

DVS Digital Video Solutions GmbH
Krepenstr. 8
30165 Hannover

Phone: +49 511 67 80 70
Fax: +49 511 63 00 70
e-mail: sales@dvs.de
Web: www.dvs.de

10. Notes