Workflowers

Centralized storage and finishing workflow

Case study

April 2020

Introduction

This case study is based on a typical digital cinema finishing workflow. Workflowers has been involved in digital cinema workflows very early, when 2K was a challenge, for storage and processing power. A typical workflow optimization mission would start with a review the different processes and bottlenecks, and then proposition of changes and agile implementation.

On this document we focus on the image side of content, we may work on the sound aspect of production on future publications.

Authors: Cedric Lejeune, Benoît Ruiz

For any information please use contact@workflowers.net

Document version 1

Table of contents

Definitions

Workflow: exchange of informations between operators

Dataflow: exchange of data between machines

Offline editing: in a typical film or TV series workflow, the editing is made using lower resolution, compressed versions of the camera footage to reduce storage and processing power requirements

Online editing: in an offline/online workflow, an edit list

Conform artist: Artist in charge of managing the consistency between the offline editing and the online conform of the project. Depending on the size of the project, this role can be managed by one or a whole team of editors

Colorist: Artist in charge of the look of the project

Colorist assistant: this position can either be held by one (or several) junior colorists to assist in the management of the look aspect or by

Color scientist: Engineer in charge of the color consistency of image processing and display devices

IO operator: Operator in charge of receiving the files and communication with other vendors working on the project: editorial, visual effects, camera department

Data manager: Operator in charge of managing the storage capacity and the classification of files

EDL: Edit Decision List, a file that lists all the source footage starting and stopping time codes that constitute a movie timeline

AAF: Advanced Authoring Format, is a file that describes the complex transitions and compositions between shots in a timeline

XML: In an editorial workflow, it is often describing the EDL coming out of compatible systems (Final Cut Pro, Adobe)

SAN: Storage Area Network, block level centralised storage technology

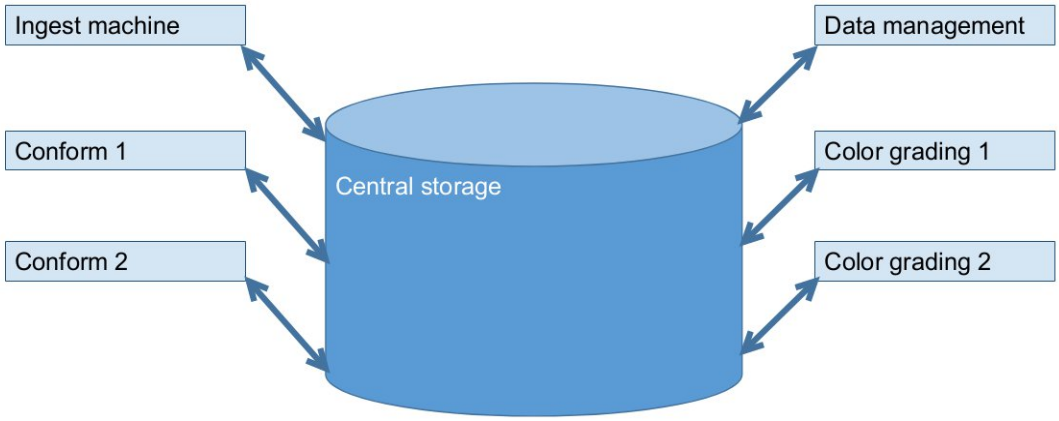
NAS: Network Attached Storage, network protocol centralised storage technology

TCO: Total Cost of Ownership

ROI: Return On Investment

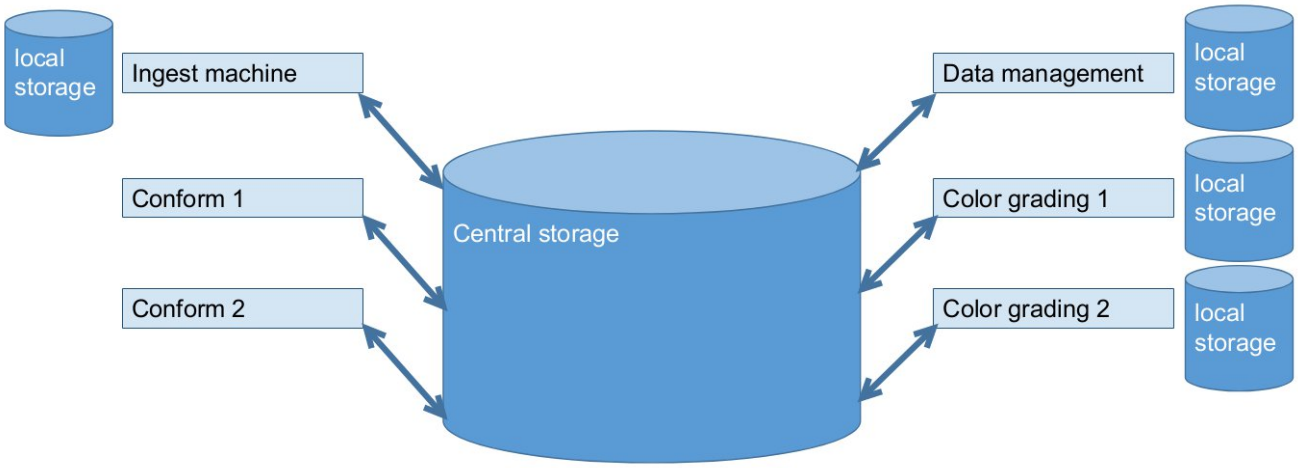
Centralized workflow issues

What is usually proposed by integrators and manufacturers for a centralized TV series or feature film color grading workflow looks like this:



This may look like common sense, as everything is centralized, one instance of each file, no risk of discrepancy in the versions as everyone is working on the same files

Unfortunately the reality of it when in production looks more like this:



Analysis

The second diagram doesn't show other problems generated by the multiplication of storage devices:

- Multiple versions of the same file
- Tracking of the versions made almost impossible
- Operators errors
- More complex processes
- More difficult control over TCO and ROI

This is usually an evolution over a couple of months:

1. Infrastructure sizing miscalculation

The original evaluation of system requirements may have been be forgetting some streams that happen in real life: backup/restoration of projects, exports, rendering, on top of the production flow. One of the problems with centralized storage is that it must combine 3 aspects:

- speed
- reliability
- capacity

When adding up those features, the cost generally doesn't grow linearly:

storage cost = speed cost **x** reliability cost **x** capacity cost

It is quite common to see some compromise made during the design phase because of the price of the "ideal" system.

2. Performance loss over time

Degradation of storage performance can occur pretty fast if proper maintenance of the file systems is not operated regularly .

3. Increasing in demand

Some technologies don't scale up nicely and show plateau effect when number of projects grows. Sometimes the addition of what could be considered minimal

4. failure of central storage during client attended session

This is a pretty common reason for adding local storage to a station as the whole facility gets a bad reputation when the image starts to stutter during a client attended session. Usually that extra storage is first used for renders to make sure that the client can see all the project in real time playback, but generally source footage is also copied later on to allow for more reliability and resilience from the central storage constraints, while generating synchronization issues at the same time.

5. data management unable to keep up with the data flow

The overflow of data and the requirement to move files quickly between the central storage and the archival device (typically LTO tape drive) makes some "buffer" storage necessary

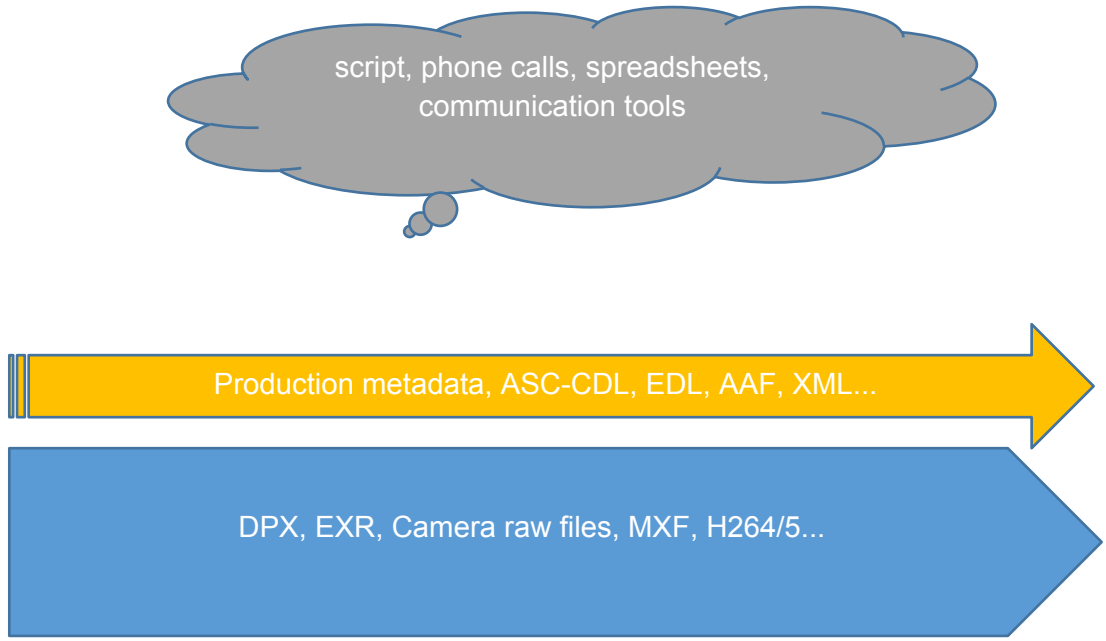
The highest costs though are not induced so much by the extra infrastructure : the added hardware also implies an increased level of complexity that leads to more complicated processes, a heavier workload and more human errors.

This very common situation is a direct consequence of a misconception that lies in the *workflow* word : in the diagram, what is described is not a work flow, but a data flow:

- a work flow is a sequence of tasks operated by humans and/or computers
- a data flow is the transmission of data from one machine to another

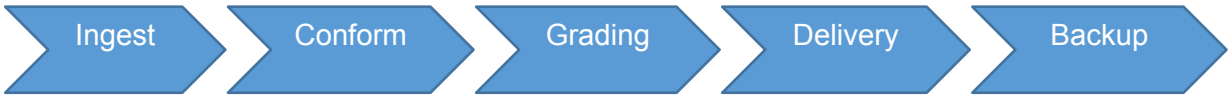
This means the mix of three very different data models :

1. the descriptive information model between humans , more or less formal
2. the descriptive information model between machines, leading to automation
3. the image material used by machines



Redesign

Let's start by expanding the different concepts. The workflow can be described in connected processes where each step can be considered a sub-system:



Ingest

This is where the elements coming from outside vendors are checked and validated for conformity.

Needs:

- validation tools
- naming convention
- fast storage for real time playback
- quality display device to check devices

Conform

This is where the movie gets assembled, from its different parts: original camera footage, special effects shots

Needs:

- proper quality proxies to check editorial and shot version changes
- sufficient bandwidth for real time playback of multiple tracks of proxies and offline reference file
- quality display device to check version differences

Grading/finishing

This is where the movie colors and effects are adjusted with the client

Needs:

- real time “rock solid” playback for full quality images with no slowing down during sessions
- reference display

Delivery

This is where the whole movie or some elements are prepared for delivering to other vendors

Needs:

-
- quality display if Quality Check is processed at that level

Backup

Needs:

- sufficient space to offload several projects from the production volume
- sufficient bandwidth to move in/out projects

(Render farm)

Servers that are used to perform calculation at different stage of the workflow. They may be handling different tasks or specialized for one particular software. It’s important to differentiate between rendering tasks as they can be significantly different in their need for computation power and storage speed requirements, so testing in production conditions is key to avoid network overload. For example CG renders load a heavy scene file and textures and output

very few frames per hour, whether color grading renders loads and render one picture at a time at high speeds, so the computing and network load models are very different.

Notes

Depending on the overall bandwidth of the infrastructure, even background copying of files can create slow down in the performance of both storage systems and workstations. Creating network regions (sub networks) may help in separating issues.

Most creative software solutions requiring real time playback propose the generation of “optimized” proxies. It’s worth considering if these can be generated as a background process or by another machine (rendering server).

An automation system for file transfers, potentially connected to a version/production tracking solution, can synchronize versions between different storage arrays, and reduce significantly human errors.

It’s interesting to see that image display consideration are linked to bandwidth, because content needs to be checked. It is also linked with the expertise of operators for efficiency.

Costs considerations

Capital expenditures

A typical virtualized SAN (Storage Area Network) system usually request machines to run the same version of the driver, which means being from the same generation, forcing the upgrade of all systems at the same time for most upgrades, being speed or capacity as in general current generation is more interesting than upgrading original generation.

A NAS based solution would be more flexible and allow for more heterogeneity between the systems, allowing also to scale up and down different components independently. It is also easier to repurpose older systems for other tasks.

A hybrid system can be designed, focusing on the requirements of each sub part of the workflow.

HR considerations

People involved in the transformation of the workflow will communicate differently, so those changes in the processes should be accompanied with training, and documented. An efficient workflow let people spend less time on boring tasks and more on higher value ones.

Ideally the evaluation of technical solutions should be managed by groups of users, experts or not, to collect feedback on test setups before getting in production. Agile methodology can be applied for testing.

Efficiency evaluation

If we consider the production chain as a system, we can consider what goes in and out:

In:

- Information from clients and other vendors
- Data from clients and other vendors
- Work time
- Planned and unplanned expenditures
- Devices
- Energy

Out:

- Information to clients and other vendors
- Data to clients and other vendors
- Technical quality of content
- Customer feedback
- Margin
- CO_2 emissions
- Fatal Heat
- Cooling liquids from cooling system leaks
- Waste of electrical and electronic equipment (WEEE)

Each point can be evaluated in quantity and quality. The methodology proposed is to decide which point we control on a test project, establish a protocol and a set of metrics that are used during the evolution steps.

Conclusion

This case study exposes some key elements in the concept of workflow, the most important one being that it's an organizational concept, not a technical concept. Technology will enable the tools for everyone involved to exchange data more efficiently, only if the people know what they are doing.

Using a systemic approach to determine the needs attached for each task can help not only determining a set of tools, but also potentially ways to reorganize the operators while they document the in and outs. It's also a great way to determine the potential indicators to measure performance increase. Getting operators involved, with a good animation of the work groups, combined with an agile approach, is a great opportunity for team building and reworking legacy inefficiencies.

These days, with technologies like the cloud, remote connections or artificial intelligence enabling new tools for the operations, it's a good time to rethink the organization for more efficiency.