THE SUITCASE

A CASE STUDY

NEW TECHNOLOGIES FOR MEDIA PRODUCTION IN THE CLOUD

By ERIK WEAVER & JOSHUA KOLDEN

Notes and Diagrams
By DANIEL DE LA ROSA

CONTACT

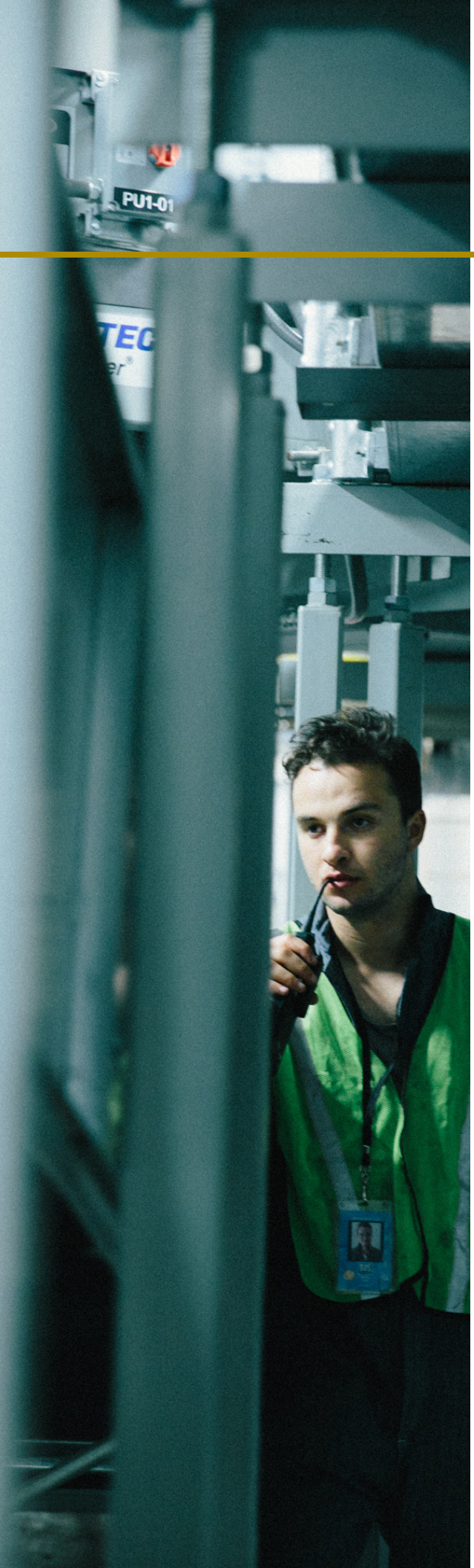
ERIK WEAVER, Program Director, Cloud Project
ENTERTAINMENT TECHNOLOGY CENTER@USC
3131 S Figueroa St, RZC 201, Los Angeles, CA 90089
eweaver@etcusc.org

TABLE OF CONTENTS

INTRO

HOW IT WORKS

DETAILED TEST CASES



ABSTRACT

"The Suitcase" is a short film produced by the Entertainment Technology Center at the University of Southern California (ETC), and directed by Abi Corbin. The film was commissioned to test the Cinema Content Creation Cloud (C4) framework. C4 is a set of open source tools and guidelines created to solve many of the problems that face media production in the cloud. This paper is concerned primarily with the results of "The Suitcase," and the cloud process employed during the film's production. It provides details on the framework's origin and design goals, its use on the "The Suitcase," and what was learned.

BACKGROUND ON "THE CLOUD" AND HOLLYWOOD

The phrase, “the cloud,” and images of clouds have been used as metaphors associated with various Internet-type technologies since ARPANET in the late 1970s. However, since the introduction of the Elastic Compute Cloud by Amazon in August 2006, the term has taken on a very specific meaning. The Cloud can be clearly differentiated from other things such as the INTERNET, Data Centers, or Mainframe Time Sharing by its economic model.

The Cloud refers to computing resources that can be utilized “on demand” resulting in usage-based billing. This allows computing power to be treated like a utility instead of like an infrastructure. The computer on your desk, or the data center your company owns and operates has a fixed amount of computing power that is purchased in advance. The utility model of the Cloud means that computing power is not purchased in advance, and the amount of computing power is not fixed.

This is exactly how Hollywood, and media production in general, has been working with production resources from the very beginning. On the face of it, there seems to be a great deal of value in cloud computing for the Media and Entertainment industry, and because of this, the Entertainment Technology Center launched “Project Cloud” to investigate the best ways to leverage the cloud for media production.

The single biggest challenge for the Media and Entertainment industry is understanding how to add the cloud to existing workflows. This understanding is hard to come by due to poorly understood terms like “the cloud” and the degree to which M&E utilizes computing resources differently than other industries.

For example, the “web services” business has a relatively light transport requirement as compared to media production. Even high-end HD video on the web is microscopic in size when compared to RAW production assets where 12MB files for single frames are common. Add to this, shooting ratios of 100:1 or more, multiple edits, post production, and visual effects, and it is easy to see that storage requirements in excess of 100TB of active storage per feature project for image data alone is not uncommon.

PROJECT CLOUD

The Entertainment Technology Center's "Project Cloud" led by Erik Weaver has endeavored to clarify the cloud and its use in media production. As with previous technology innovations, ETC helps to discover and encourage new technology to address fundamental industry needs. ETC has been very successful in this regard having been responsible both directly and indirectly for many technologies and standards being used in production and distribution throughout the world today.

The work supported by ETC is "pre" standardization, which encourages innovation and flexibility in the design and development process. The focus is to create pragmatic solutions that pay close attention to the needs of real world products, and the business of Media and Entertainment as a whole. After validating solutions in the real world, relevant technologies are then established as standards with the appropriate standards bodies for industry-wide adoption.

In an effort to be practical and realistic in its evaluations, ETC began a program of regular real world production tests. These tests are short films made by graduate students from the USC School of Cinematic Arts. So far, two films have been produced under this program, and a third is in development. The first film, "Luna," shot in 2014, focused on quantifying the "state of the art" in cloud production in order to provide a starting point for understanding the potentials and limitations of cloud technologies.

This paper is concerned primarily with the results of the second film, "The Suitcase," shot in 2015-2016, and the cloud process employed. It provides details on the framework's origin and design goals, its use on the "The Suitcase," and what was learned.

Abi Corbin, who was a USC Graduate Student at the time, directed both "Luna" and "The Suitcase." Aside from her talent and experience, it was decided that Abi Corbin should direct both of the films to provide continuity and the best feedback on how things improved from one production to the next.

OVERVIEW

Although by some definitions a student film, “The Suitcase,” was conceived at the outset to be a challenging professional production. The story is set at Boston’s Logan International Airport, and a significant amount of film was shot at an actual airport near Los Angeles.

The 26-minute film was assembled from 372 minutes of production footage resulting in a shooting ratio of 14 to 1. Three different professional cameras were used: the Arri Alexa, Arri Amira, and the Red Epic, producing three different types of camera original files. Roughly 2TB of data were produced per production day all of which was transferred nightly to cloud and local storage, for a total of around 25TB of production assets.

MOTIVATIONS

There are several key problems that anyone working with cloud resources eventually encounters. Likewise there are a number of industry-specific constraints to the types of solutions that can be considered.

Within the industry there is already a great deal of expertise in coordinating information systems. Interestingly, media production functions much like the cloud. Resources are typically rented instead of purchased, and data is passed between disparate systems, each managing data differently, while adapting to a constantly changing landscape of production departments, people, and organizations. This is usually managed gracefully and reliably and has been for many decades—the results of these chaotic, dynamic systems are nothing short of spectacular. It works.

Although the infrastructure components are different between feature film, telvision, music, broadcast, and other production pipelines, consistent, reliable coordination is the most fundamental challenge. Because of this, we've often seen media companies employ organizational consultants in an attempt to improve workflows. These imposed organizational systems, while often very well designed and successful in other industries, almost universally fail to take hold in this industry. They prove insufficient in comparison to the people who are already experts at solving these problems but in even more dynamic and flexible ways.

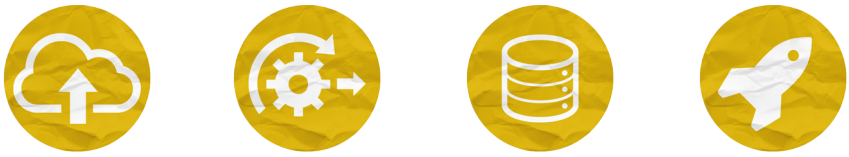
Coordinators, editors, cinematographers, or script supervisors are all experts with respect to their work. We refer to these people as "domain experts." Domain experts can be quite particular about the tools and techniques they employ, and to make the cloud useful, we must minimize the burden cloud workflows can cause. There may be disagreement between experts or different workflows may be necessary for different projects—even with the same expert. These workflows may not appear optimal in theoretical scenarios, but only the domain expert has the situational awareness and judgment to make the best choices for a given situation. This is why imposed systems typically fail.

THE FRAMEWORK

The problems that have been identified as core challenges for working with the cloud are the following:

1. Asynchronous cross-organization data management.
2. Data transport latency, bandwidth, and cost.
3. Cloud system integration and throughput.
4. Metadata management.
5. Security.
6. Interoperability between tools and between organizations.

THE WORKFLOW



For "The Suitcase," we outlined a number of tests to be conducted over the course of pre-production, production, and post-production project phases. During pre-production, the team evaluated a number of currently available "software as a service" production platforms. These included Sync-On-Set, Movie Magic, Scenechronize, and ScriptE. Though many of these tools do not currently have RESTful JSON APIs, as per ETC's recommendations, they are working in that direction.

Production consisted of an onset workflow that flowed through Technicolor's DIT Cart and On-Set Production services. In the first of three shoots, ARRIRAW was transferred via the onset cart shortly after capture to the cart's local RAID storage. From there it was transcoded to H.264 and DNxHD 115. The H.264 was transferred to PIX over onset wifi, or later in the evening, from Technicolor's Burbank office.

During production two G-Tech Studio drive systems were used. These were configured in RAID 5 and provided 24TB of storage and over 200MB/s of sustained throughput. These drives were alternated on each day of production, between the ETC offices at USC, and the production location. Throughout the day, production assets and the transcoded results were transferred to the G-Tech Studio drive on set, then sent to ETC at the end of day. The files that were not already in the cloud were then transferred to the cloud via ETC's connection provided by Sohonet.

The full Sohonet's connection bandwidth is 1GB/s, but ETC is configured to use only 1/3 the full line capacity. Nevertheless, using C4 optimized data transfer, the data was fully transferred to the cloud nightly leaving time for additional tests to be conducted.

A PostBox system was also tested in the role of onset DIT cart and as a studio side ingest tool. The PostBox system proved itself to be a very powerful alternative to a premium service provider. The PostBox is configured for portability and performance with an integrated Mac Pro, monitor, and carefully isolated Thunderbolt buses. The system adds additional expansion slots for the Mac Pro useful for expanding to 10GB/s networking and other specialized cards.

To provide enhanced file tracking each time data was copied from one location to another, C4 was used to identify the files and log transactional metadata. This provided searchable JSON format log files that associated C4 IDs with filename, dates, physical location, and drive names. This provided an unassailable proof of custody that could be used to resolve any missing file issue.

A number of different cloud workflows were tested including automated metadata extraction, transcoding and color correcting in the cloud, and file sharing with the team.

C4 ON SET

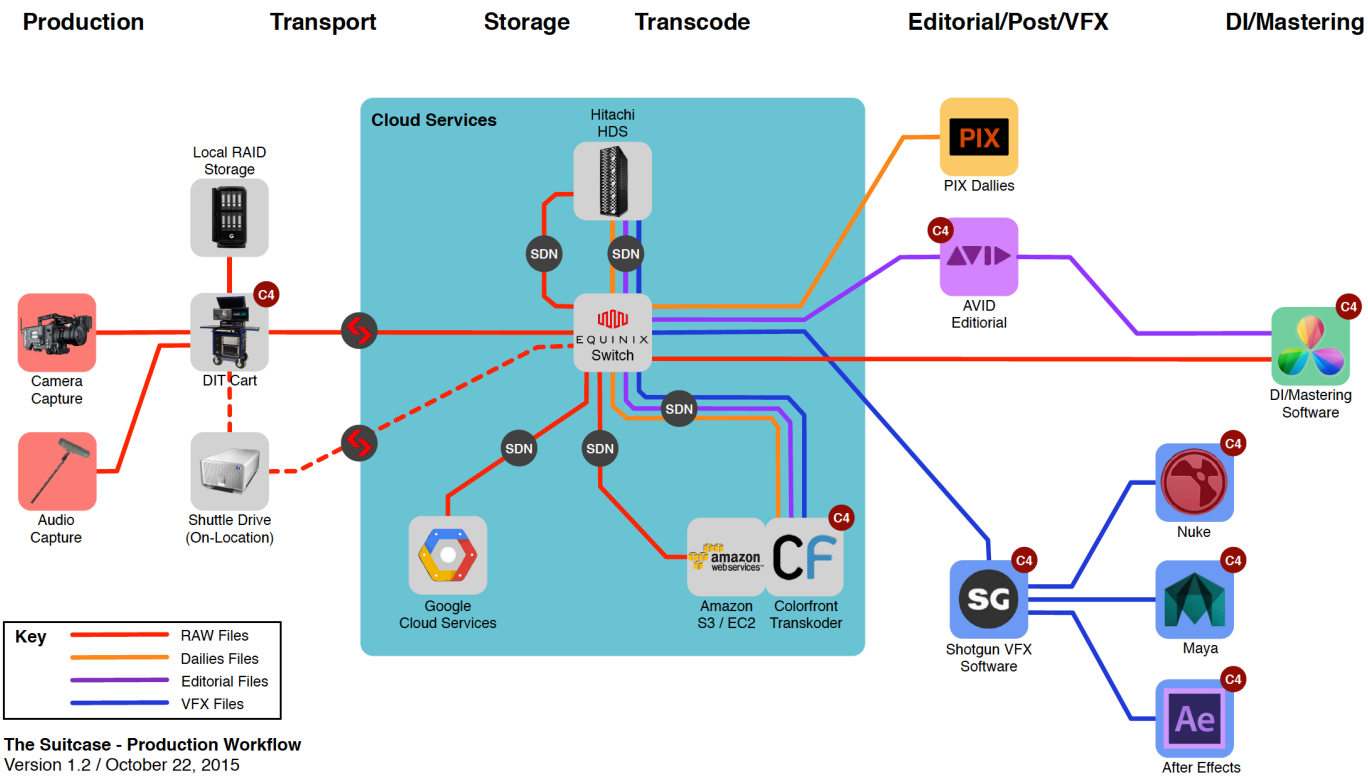
Allison Hettinger from On Location Services at Technicolor modified Technicolor's onset software and workflow to automatically capture the connections between assets by identifying the C4 IDs of the various assets, and associating them to each other. This included the identification of the source ARRIRAW footage, the H.264, the DNxHD footage, and the audio. The following is an example of the YAML-structured metadata and relationships that were captured.

```
/Volumes/es1/projects/suitcase/20151120/ARRIRAW/A013R0SW/A013C002_151120_R0SW:
  c4id: c43FctfME7ehwa1HZE9AXTJ7jfkZssaVouoaakoDm8gXALP1MJVRzbXGauLoQKyycxH-
in3gboL7PLXo71DUr7RJUd4
  filename: "A013C002_151120_R0SW.%07d.ari"
  camformat: "ARRIRAW"
  camroll: "A013"
  scene: "2"
  take: "2"
  shootdate: "11/20/2015"
  framerate: 23.976
  timecode:
    start: "00:23:40:21"
    end: "00:25:27:23"
  height: 2202
  width: 3424
---
/Volumes/es1/projects/suitcase/20151120/audio/DAY3/2T02.WAV:
  c4id: c41kjNQAjzZz4NhtNVFz7omWLq2Z9rC3rm5MiCBheLhAveiZXde8z9rD1r9ADFbN94m5G-
dW1UDeKLhf5MhGud2qFED
  source: c43FctfME7ehwa1HZE9AXTJ7jfkZssaVouoaakoDm8gXALP1MJVRzbXGauLoQKyycx-
Hin3gboL7PLXo71DUr7RJUd4
  filename: "2T02.WAV"
  notes: "Audio file"
---
/Volumes/es1/projects/suitcase/20151120/deliveries/RESULT1-mxf220-149/A013_2-
2.mxf:
  c4id: c45S284krEp8xLawko9hG2bkCMma3Eq14J6vKDNBzFqc9qqTKZoWxn-
4MPHrY48zv3kYjqDhzvE34UxoeeA1jLwRKkG
  source: c43FctfME7ehwa1HZE9AXTJ7jfkZssaVouoaakoDm8gXALP1MJVRzbXGauLoQKyycx-
Hin3gboL7PLXo71DUr7RJUd4
  filename: "A013_2-2.mxf"
  process: "Avid 115"
  notes: "Rendered via ExD"
---
/Volumes/es1/projects/suitcase/20151120/deliveries/RESULT2-pix-no-
roll720-149/2-2_A013.mov:
  c4id: c43LtUujt6UsQG6TB2pxWkSLBmnw4m63KXddPn8dhbK8kCR2vJ98eLke-
Lyi9Rdqv51aeZhvwo7StTSo9oTu6YPhvg7
  source: c43FctfME7ehwa1HZE9AXTJ7jfkZssaVouoaakoDm8gXALP1MJVRzbXGauLoQKyycx-
Hin3gboL7PLXo71DUr7RJUd4
```

Example Code 1, Excerpts of YAML relationship / metadata files from Technicolor.

Note how easy this metadata is to read and understand using the YAML language, yet it represents a complete and traceable record of the relationships between these files. Finding the C4 ID of any of these files (regardless of any path or filename change), allows one to quickly search and find the other related files. Prior to the use of C4 IDs, there was no way to do this without a centralized digital asset management solution, and even then it was prone to user error that could cause this information to be permanently lost.

We refer to this as “Indelible Metadata.” If you have the metadata in a text file, database, or other searchable form, and you have the asset, then C4 ensures that you can always connect the two together.



DETAILED TEST CASES

TAG DESCRIPTIONS

INFRASTRUCTURE

#MX #C4 #HDS #SOHONET #AWS #GC #SDN

TECH

#EDIT #VFX #AUTO

SOFTWARE & SERVICES

#ASPERA #PF #PIX #CF

PROCESSES

#ORCHESTRATION #BLOGS

PEOPLE

#DIT #DAM #VFX #TECH #COLORIST #ALL

PARTICIPANTS

TECHNICOLOR was a key partner in “The Suitcase.” They provided end-to-end onset to color and sound post, as well as creating an HDR Color and Master. C4 was integrated and tested on-set using carts which allowed new and unique workflow tests.

EQUINIX LA3 solution validation center was used to test principles of Interconnected Orientated Architectures (IOA) and multi-cloud private connections. We staged the HDS storage appliance at the colocation facility and created private direct connection to AWS & Google to leverage principles of IOA. Their solution utilized a data lake and cross-connects to help provide greater security and performance.

HGST/G-TECHNOLOGY was a key solution provider for data storage and workflow. They delivered over a 100 TB of various size storage appliances, including 1 TB rugged drives, 4 TB Dock stations, and 32 TB desktop raids.

GOOGLE GCP/ ZYNCRENDER provided cloud resources, including storage and compute for VFX and rendering. A direct connection was established at the Equinix data center to provide layer-2 connectivity from the HDS storage appliance.

SONY provided a BVM-X300 monitor for principle photography. This allowed the team accurate HDR onset viewing.

AMAZON WEB SERVICES provided cloud-computing resources, which included both storage S3 as well as computing for transcoding processes.

HITACHI DATA SYSTEMS supplied a 300 TB mixed block/object storage appliance that was placed in Equinix LA3, and allowed direct connections to both AWS & Google.

PARTICIPANTS

COLORFRONT provided ACES tools for transcoding in the cloud.

MARKLOGIC worked closely with Disney extracting authoritative structured metadata from the production process so that it did not have to be recreated downstream. This process was created through the implementation of semantic data modeling that is managed by the MarkLogic NoSQL semantic database platform.

The team also created a search application that leverages the tagged motion picture data, along with user profile information. This combination allows users to view the movie and the corresponding descriptive metadata at the same time.

GRAYMETA software tool allowed for the automated extraction of metadata from multiple software suites, and ingested both Technical and Descriptive metadata.

WEMERSIVE/VIDEO-STITCH delivered live 360-video capture and delivery with 15-frame delay to external viewers.

POSTBOX provided a portable DIT station for data workflow.

ARRI provided us with an ARRI Alexa XT and supporting camera package. To maximize HDR deliverable, we worked with ARRIRAW which produced data at a rate of _____ per hour.

PIX SYSTEMS provided a cloud-based digital dailies. PIX also integrated C4 into their system to help with data search and discovery.

FINAL DRAFT provided script writing software.

AVID

SYNCONSET

MOVIE MAGIC

SCENECHRONIZE

SCRIPTE