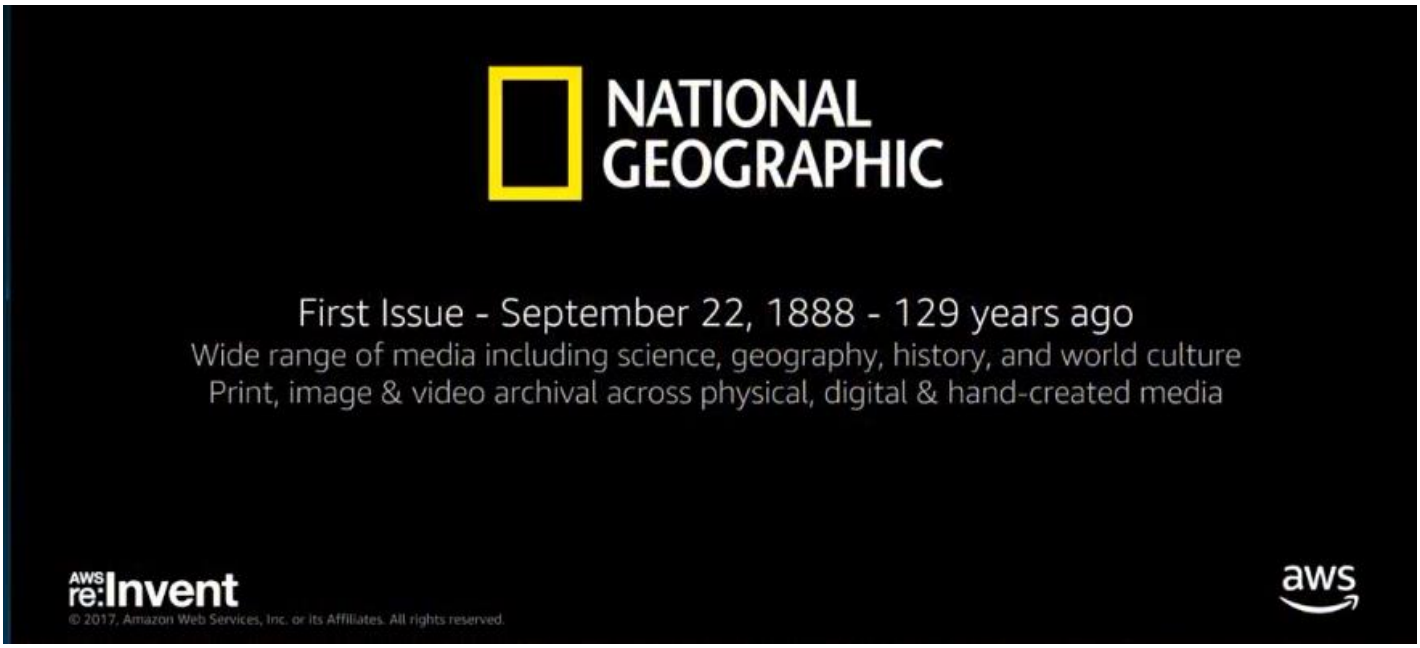
In this session, we take a pragmatic approach to enhancing ***common media workflows built around ingest, media asset management, live video, and OTT on-demand streaming***. We show ***how to extract metadata as an additional intelligence layer for video using Amazon AI services***, such as ***Amazon Rekognition***, in combination with turnkey architecture built around ***AWS Lambda***, ***Amazon ECS***, and ***Amazon EC2 Spot Instances***. The capabilities offered by Amazon AI services provide a unique opportunity to eliminate the traditional undifferentiated heavy lifting associated with ***contextual***, ***facial recognition*** and ***object-based media metadata creation***—that is, who is in with what, and where. We also discuss a large studio and broadcaster just starting to use these intelligent offerings from AWS as they change their method of how to best leverage the business value of their content. ***We can build Deep Learning pipelines in the cloud specifically enriching metadata and enriching the value of content***.



The Opportunity

- Petabytes of images
- 100+ years of content
- *How can we enrich our metadata in AWS?*
- *How can we unleash the value of content we already own once in AWS?*

The Challenge

- Niche Image Categories
- Low & Ultra High Resolutions
- Artifacts & Noise
- Black and White Footage
- Historical Context
- High Accuracy Required

Digital Transformation

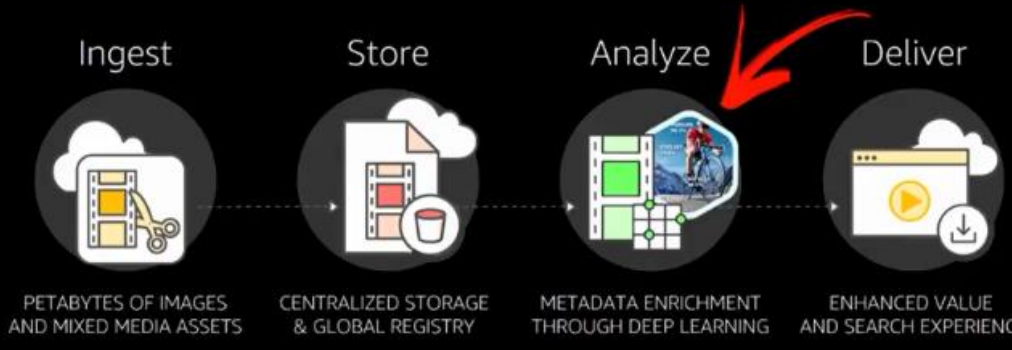
AWS Migration

- Storage / Archive
- Editing & Publishing
- Video Streaming
- Web Apps

Deep Learning-Based Image Analysis

Deep learning-based image recognition service
Search, verify, and organize millions of images

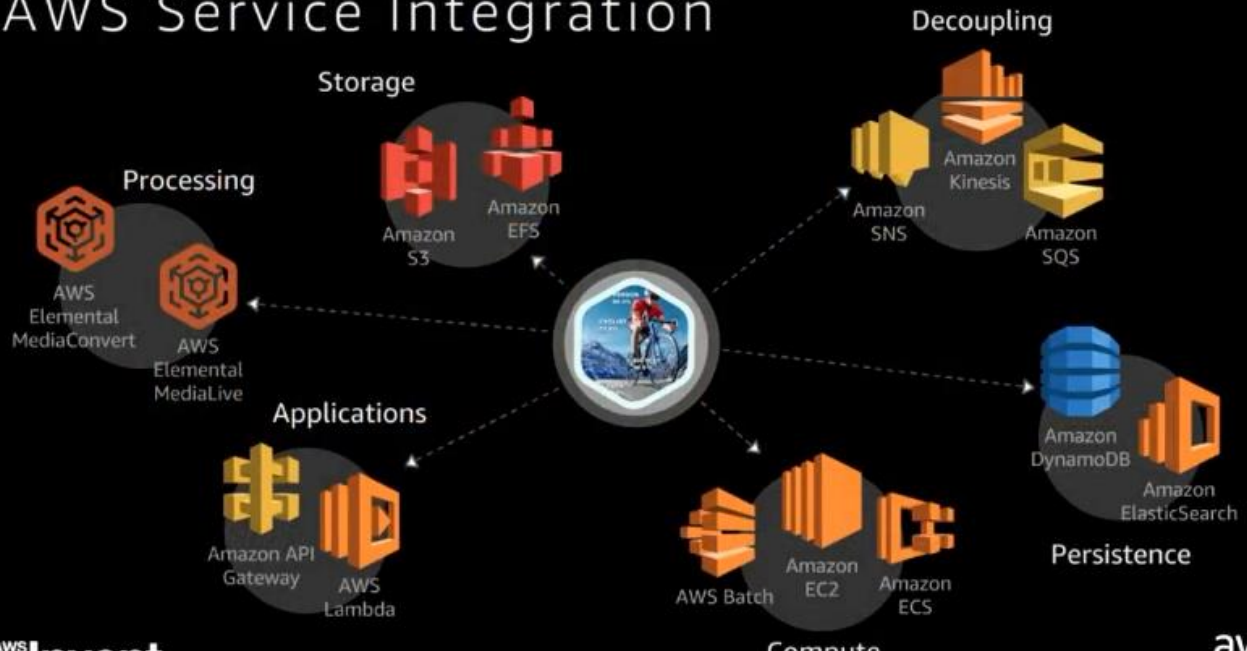
Media Intelligence Pipeline



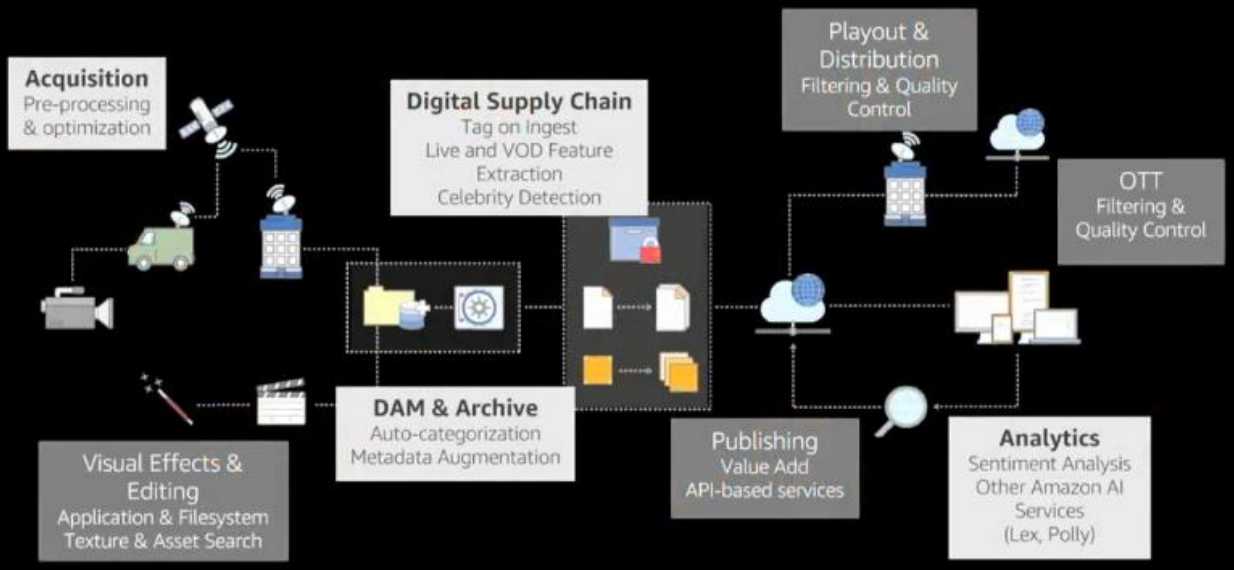
Object & Scene Detection

Identify objects, scenes & concepts, and provide confidence scores

AWS Service Integration



Wide Applicability across M&E Segments



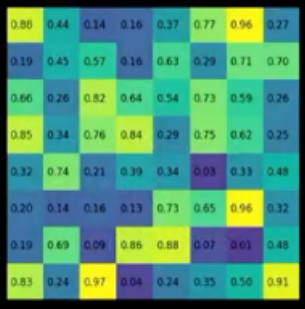
Integrating & Extending Rekognition

AWS Services	AWS Partners	3rd Party Software
• Amazon S3	• Asset Management	• AWS AI AMI
• AWS Lambda	• Media Workflow	• OpenCV
• AWS SQS & SNS	• Content Processing	• ImageMagick
• Amazon DynamoDB	• Image Optimization	• FFMPEG
• Amazon EC2 / SPOT	• Feature Extraction	• Many Others

BEAMR sdvi GRAYMETA vidispine Nventify Cloudinary moz://a

Why use Managed Deep Learning in the Cloud?

- Running deep learning infrastructure is hard
- Versioning deep learning models with zero downtime is even harder
- Built for scale - consistent response rate
- Native AWS service, security & event coupling
- Easily glued into existing pipelines – without disrupting production processes



Media Intelligence Pipeline

Design Requirements

- Self-service / multi-tenant for internal teams
- Stored metadata available via an API
- Automatic image resizing
- Unique ID creation for global ingest

Global Asset Ingest & Registration

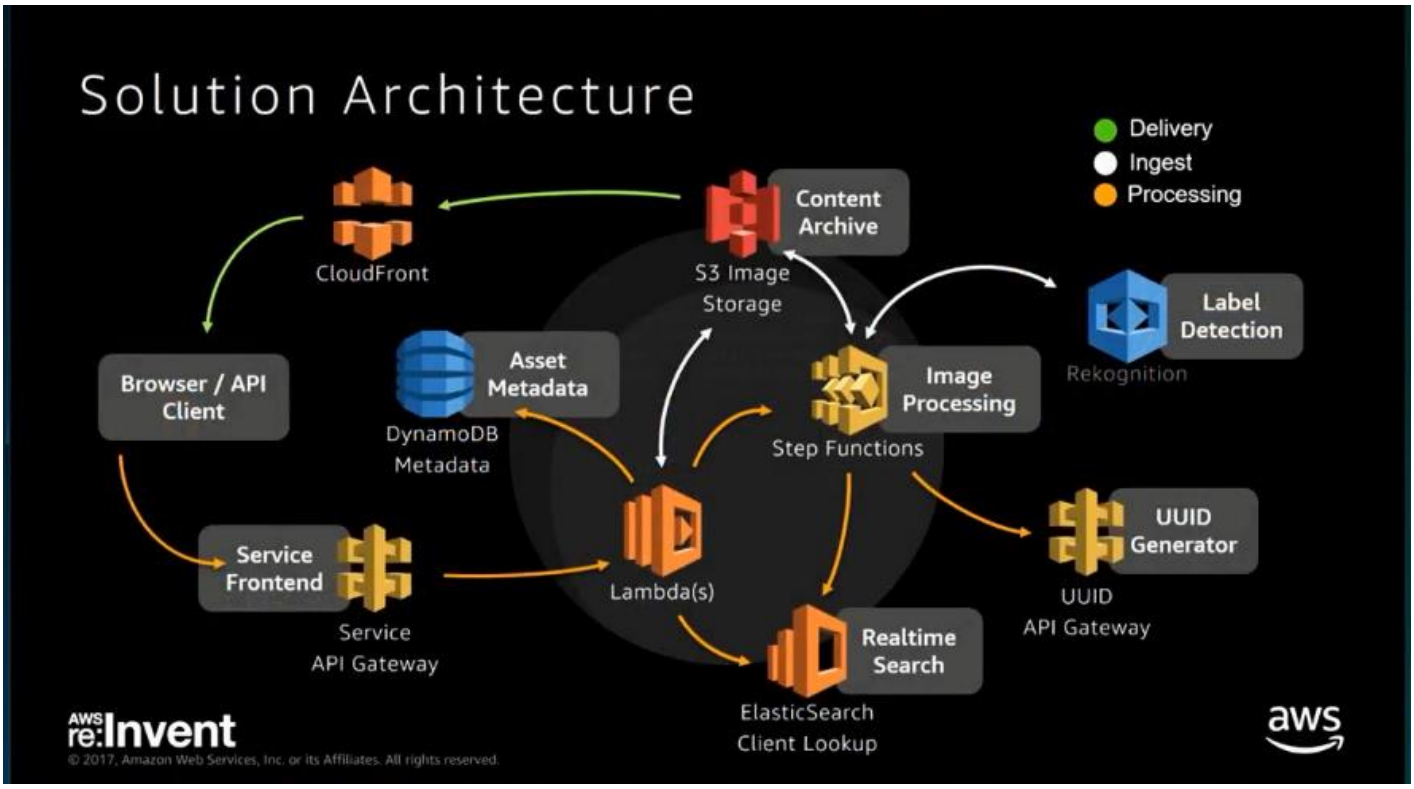
- New media may be shot for ingest anywhere on the planet (and beyond)
- Globally unique asset-ID registry which creates an ID for FOX media assets
- Service can handle parent-child relationships for asset versioning

Key AWS Components

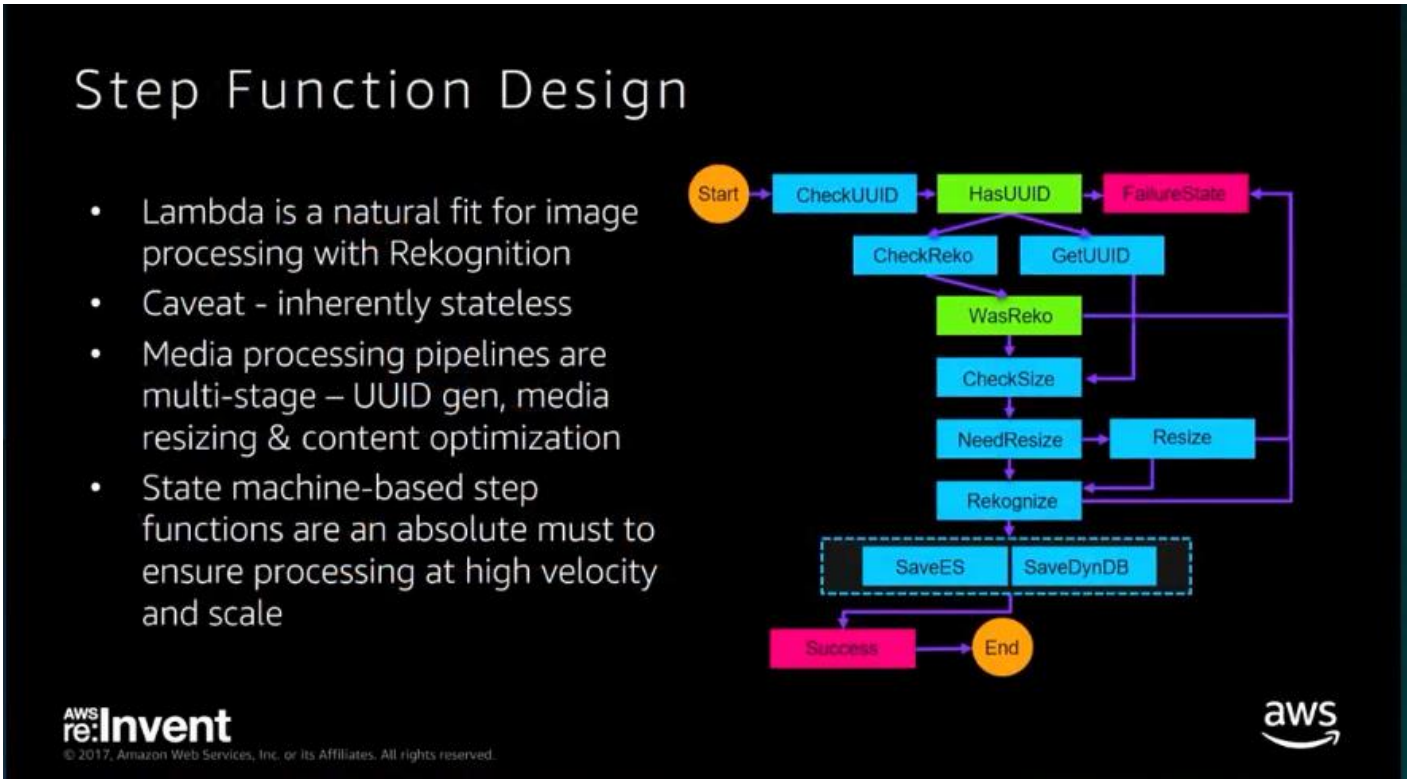
Lambda-Centric AWS Service Stack

Rekognition API Endpoints

We created a Step Function workflow that runs this entire process, this allows us to keep state and act as the single source of truth as we go through the system.



As images land in an S3 bucket, Lambda is watching for S3PubObject events and Lambda kicks off the Step Functions workflow. A request is sent to the UUID Generator to get the UUID for the current image, and then goes on.



Note that we are simultaneously writing to both DynamoDB and ElasticSearch at the same time and not replicating from one to another.

Recognition Sample Response

Hawkbill
Sea Turtle

```
"Labels": [ {
      "Confidence": 95.04956817626953,
      "Name": "Reptile" },{
      "Confidence": 95.04956817626953,
      "Name": "Sea Life" },{
      "Confidence": 95.04956817626953,
      "Name": "sea Turtle" },{
      "Confidence": 95.04956817626953,
      "Name": "Tortoise" },{
...
```

Mountain
Gorilla

```
"Labels": [ {
      "Confidence": 98.82418823242188,
      "Name": "Animal"},{
      "Confidence": 98.82418823242188,
      "Name": "Gorilla"},{
      "Confidence": 98.82418823242188,
      "Name": "Mammal"},{
      "Confidence": 98.82418823242188,
      "Name": "Monkey"},{
...
```

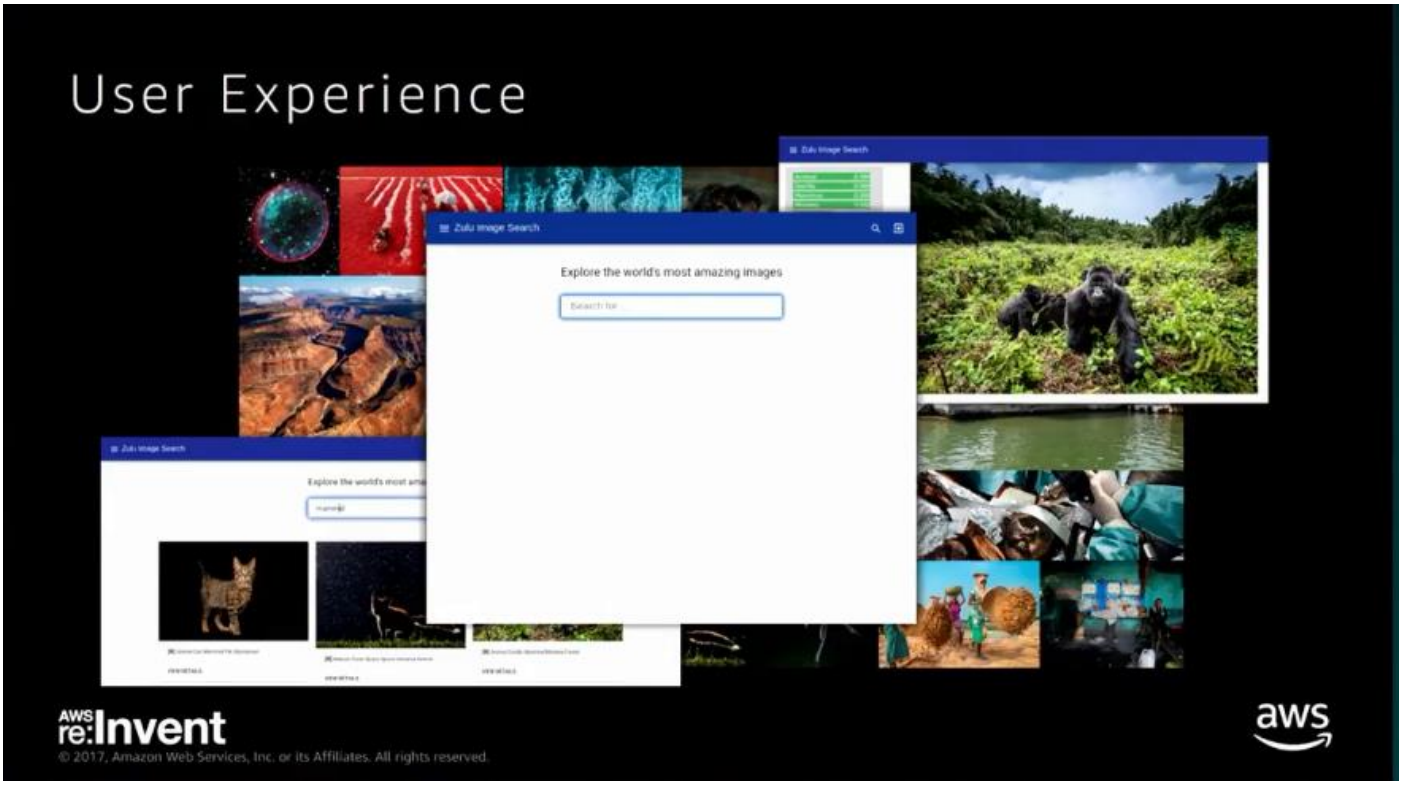
Label Data Storage

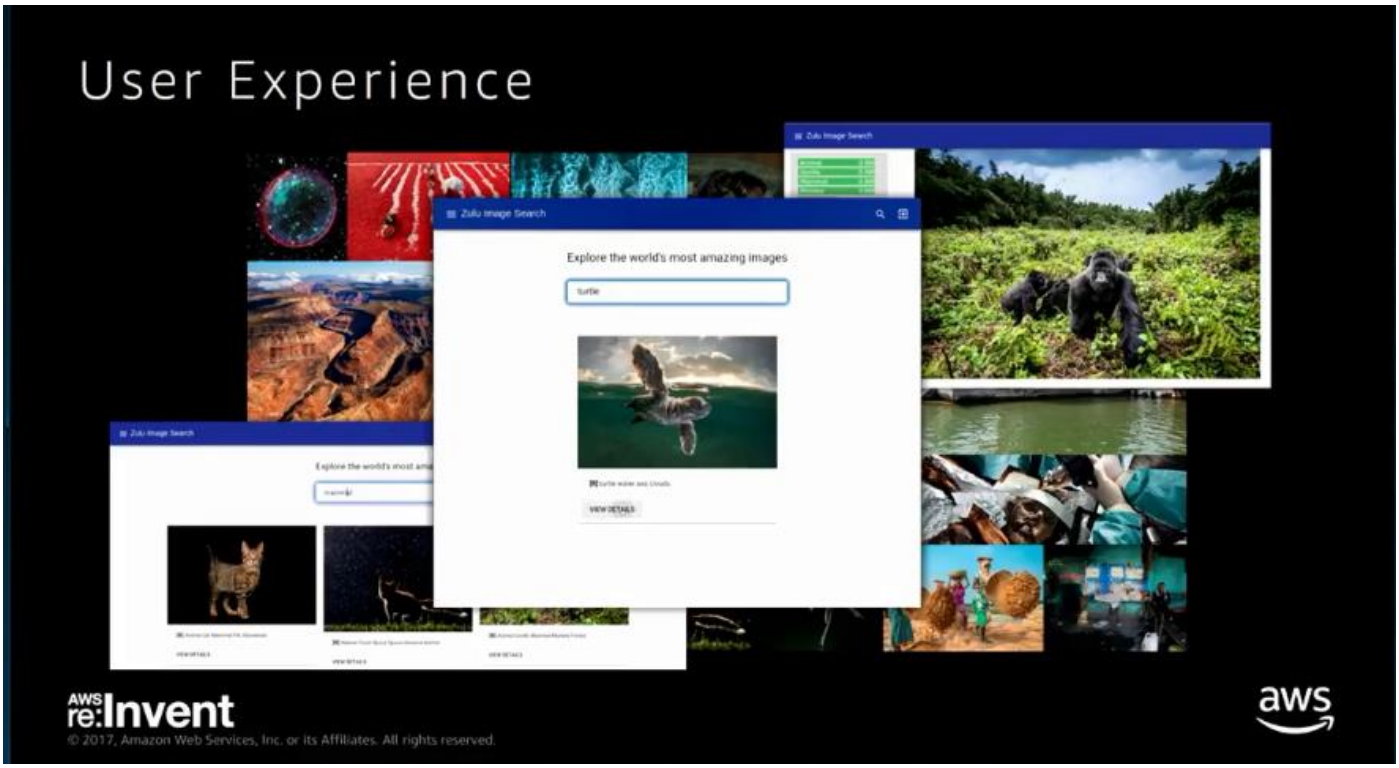
- JSON blobs well suited to unstructured ES search & NoSQL
- Multiple labels can be used to effectively widen ES search results
- Rekognition's MinConfidence threshold removes false positives; MaxLabels limits returned results
- Client-side filtering can be used to rank results by confidence score

```
[
  {
    "UUID": "<UUID>",
    "Bucket": "<bucket>",
    "Key": "<key>",
    "Labels": [
      {
        "Labels": [
          { "Name": "turtle", "Confidence": 98.4629 },
          { "Name": "water", "Confidence": 79.2097 },
          { "Name": "sea", "Confidence": 75.0611 },
          { "Name": "clouds", "Confidence": 50.5281 }
        ]
...
]
```

DynamoDB ElasticSearch

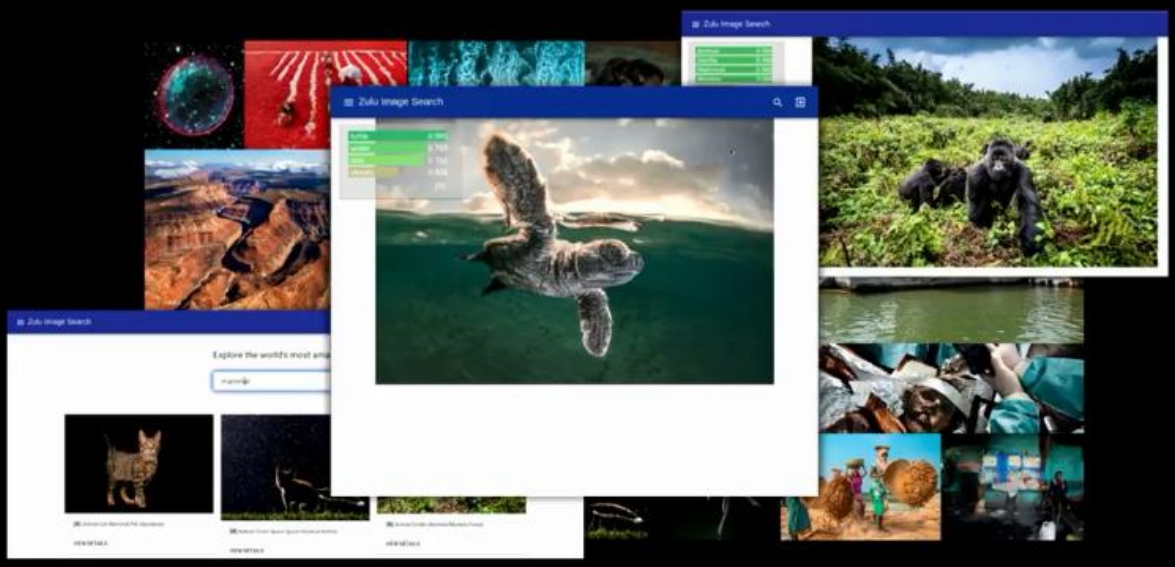
The Labels that come back from the ***AWS Rekognition***'s image recognition phase is then stored in both ***DynamoDB*** and ***ElasticSearch***, we can see the details of the image asset being stored. We can also add the ***workflowID*** of the ***Step Function workflow*** that generated this recognition details.





This is the API interface that we use to search for a photo and visualize the results metadata from ElasticSearch

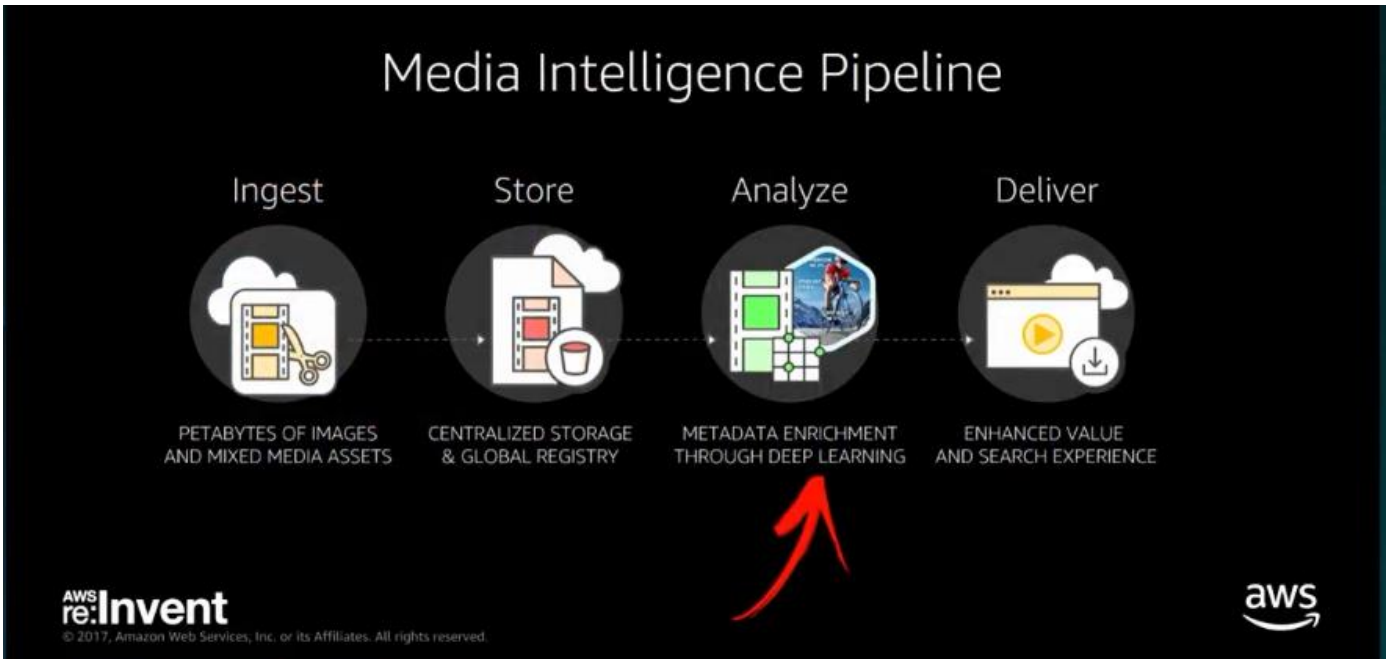
User Experience



Next Steps

- Video capability
- Metadata transformer for varying output req's
- Rekognition result differential tracking
- Integration with existing Web & Mobile apps

We can use ***Amazon Mechanical Turk*** to resolve edge cases that do not give a satisfactory recognition results, then add this back into the deep learning pipeline. We can use intelligent denoising deep learning for resolving images with a lot of dark or unclear spots.



Niche Context Deep Learning

Dog or Muffin?

This is a case where we need some ***Assisted Intelligence*** and use some service like ***Mechanical Turk***

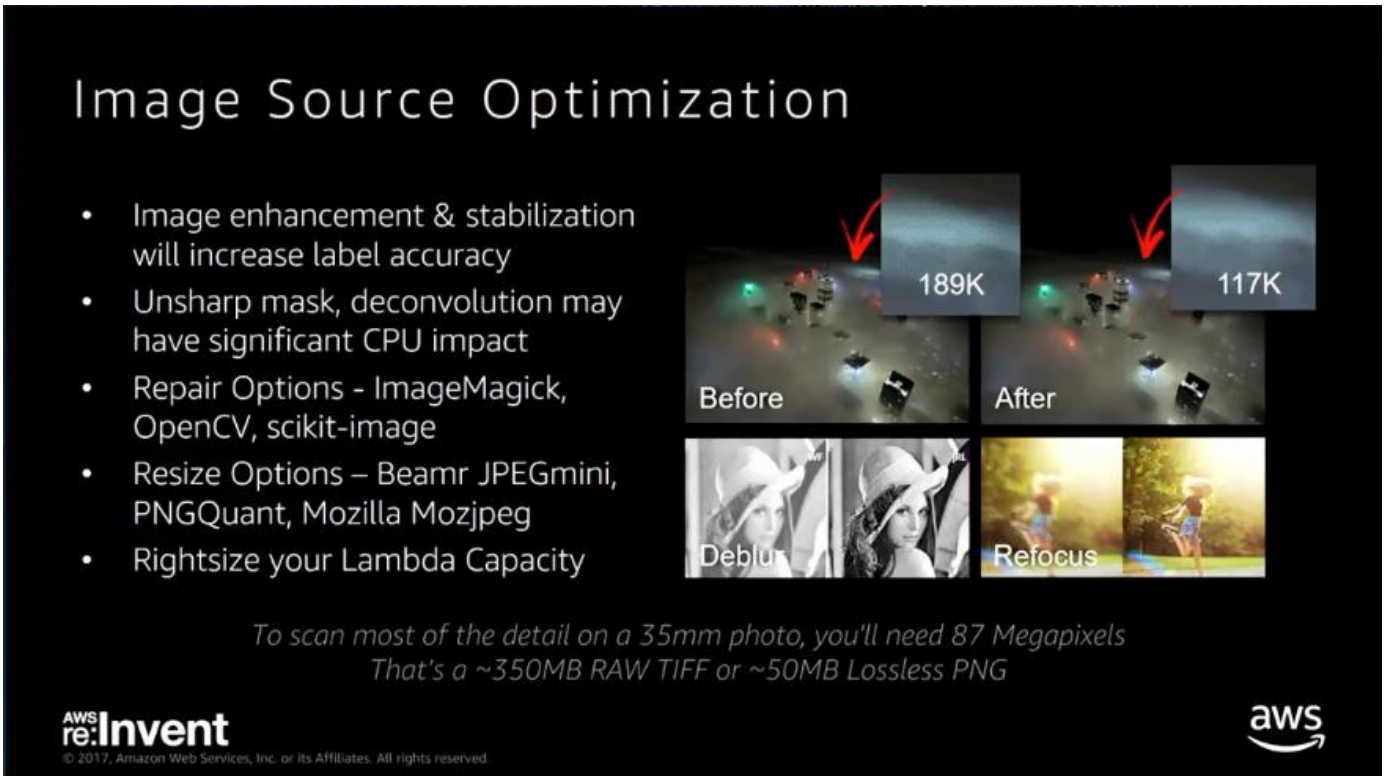
Niche Context Deep Learning

Feed Lot

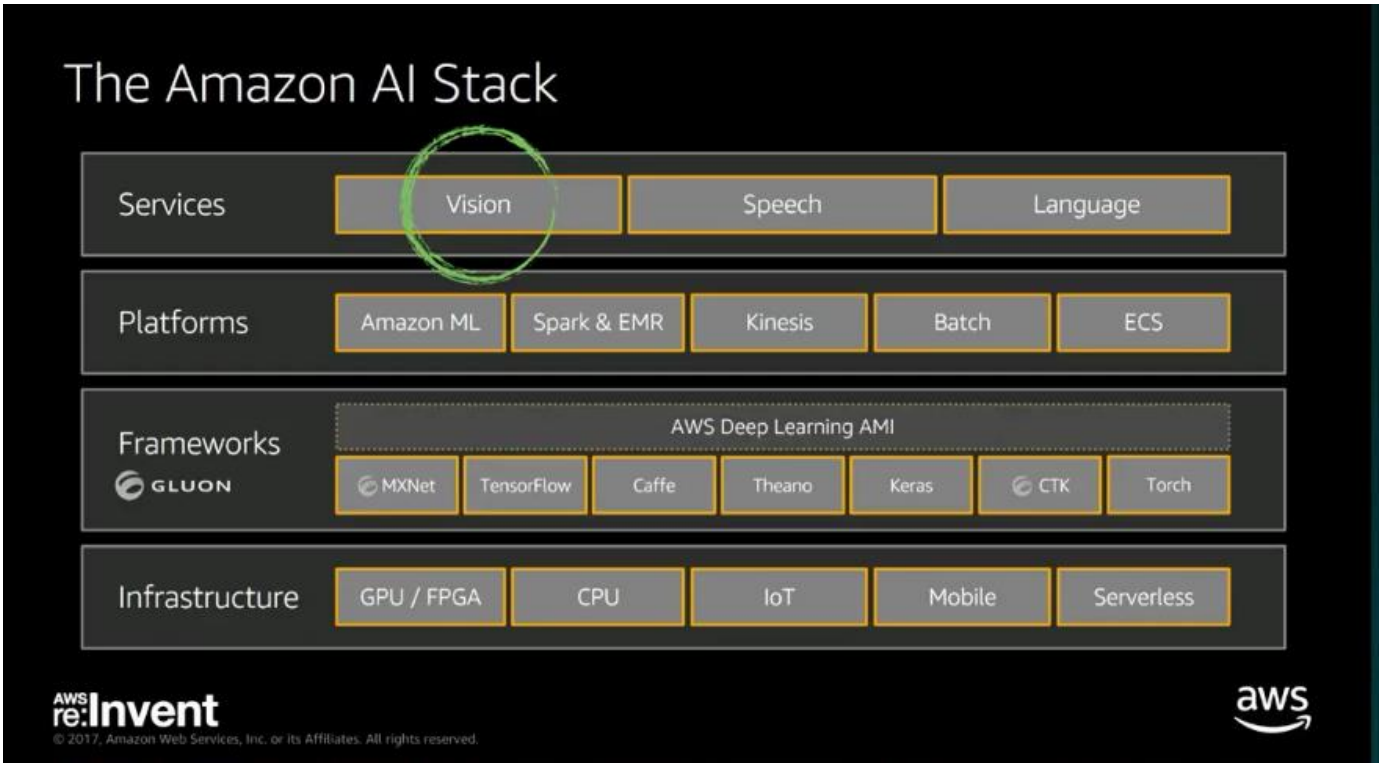
Regal Shoes

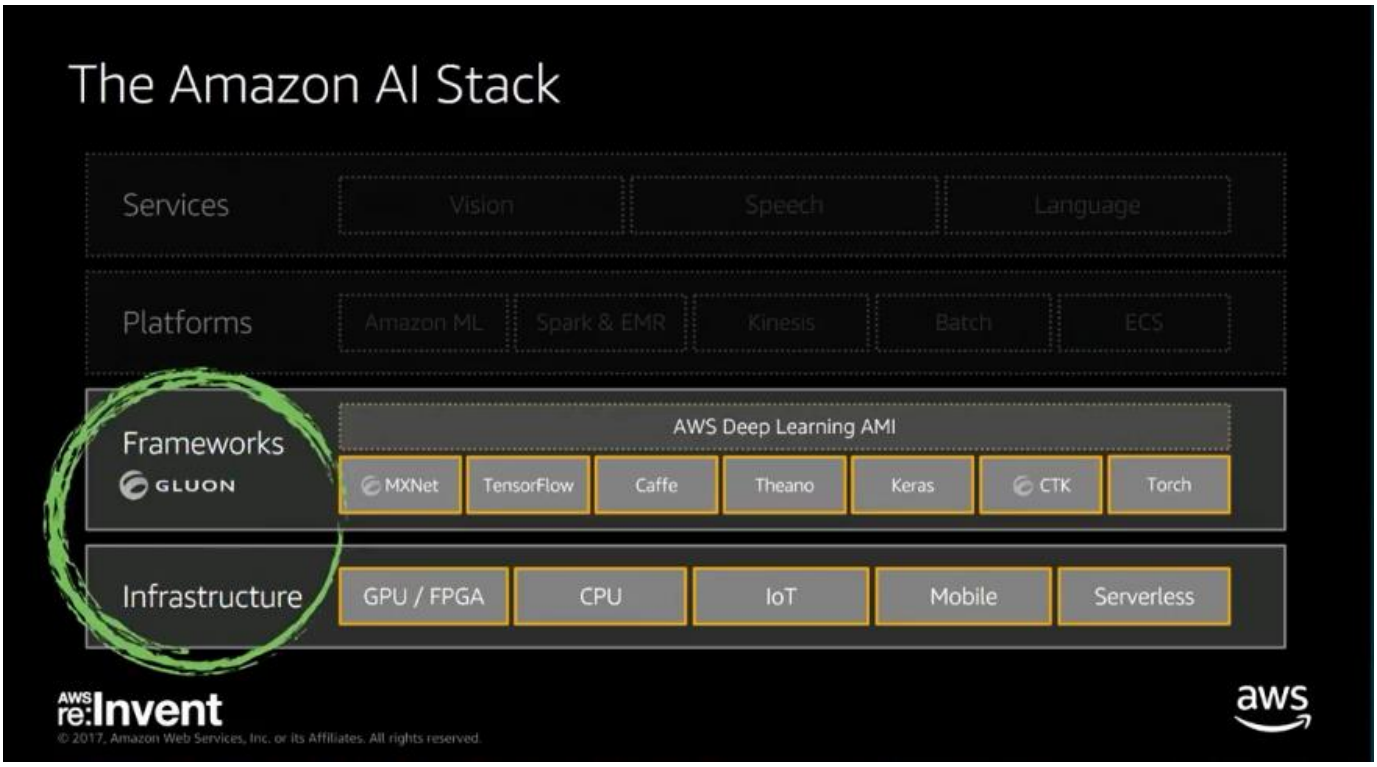
Sunbathers

Edge case deep learning engineering costs far outweigh that of a lower-cost human-aided task

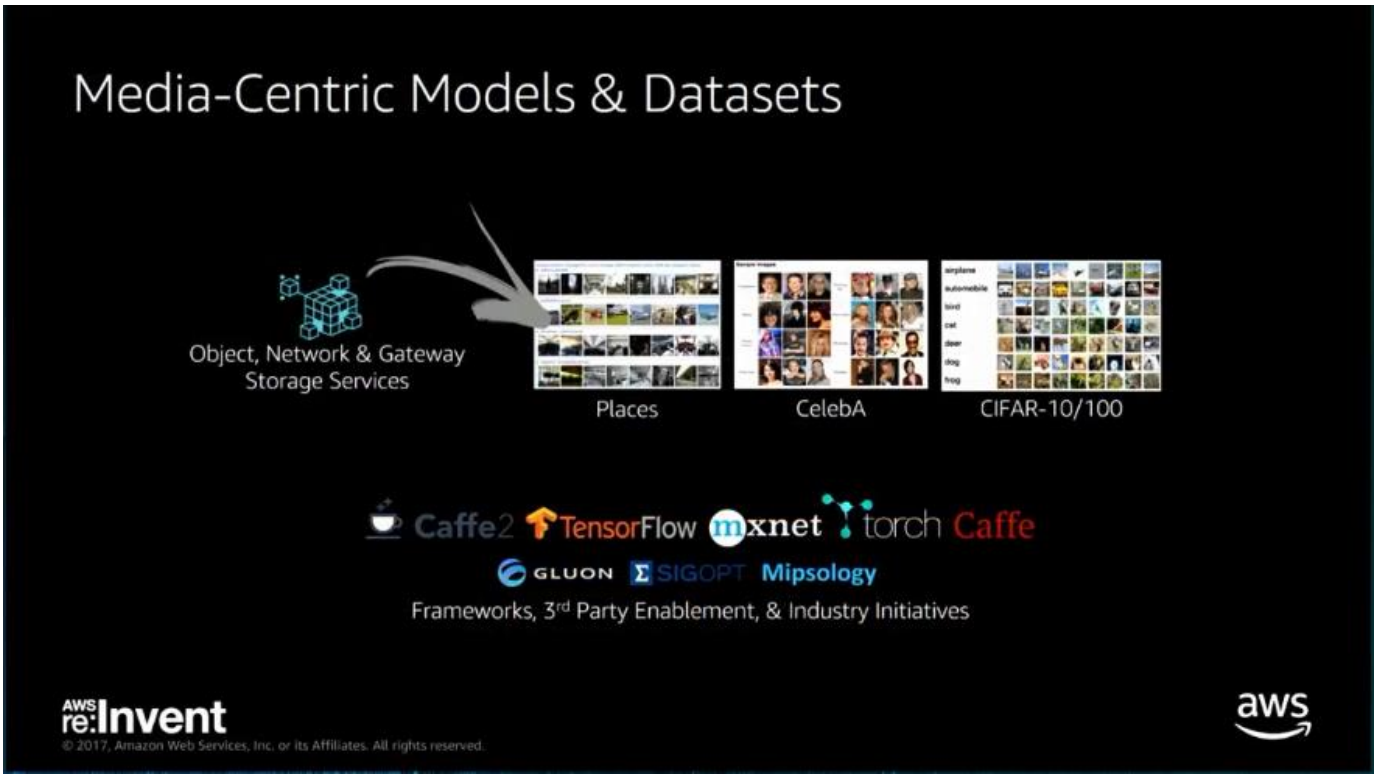


We can also have stages in the pipeline that can help improve the quality of the image and also some stages to reduce the size of the image while preserving quality. We can also have stages that can extract faces within an image and pass the face data to ***Amazon Rekognition*** for recognition and then add it to the original image metadata.





Here we are looking at deploying frameworks like TensorFlow, MXNet and so on top of infrastructure. We can run Caffe2 using lambdas too.



There are several datasets available in open source today that allows you to use them in your code.

GPU, CPU & FPGA Instances

P2 & P3: Distributed Training & Inference
Hyper-scale performance on NVIDIA V100s

G3: Multi-User Modeling
NVIDIA M60 GPUs, 16,384 cores

F1: High Speed Inference
Xilinx Ulstrascale Plus, 6,800 engines

X1: Specialized AI/ML/DL
128 vCPUs, 3,904 GiB RAM

Instance Name	GPU Count	vCPU Count	Memory	Network	EBS
p3.xlarge	1	8	61 GiB	~10Gbps	1.5 Gbps
p3.8xlarge	4	32	244 GiB	10Gbps	7Gbps
p3.16xlarge	8	64	488 GiB	25Gbps	14Gbps

*P3 Instances Provide up to 1 **Petaflop** of mixed precision performance, and 125 Teraflops of single precision floating point*

Artificial Artificial Intelligence

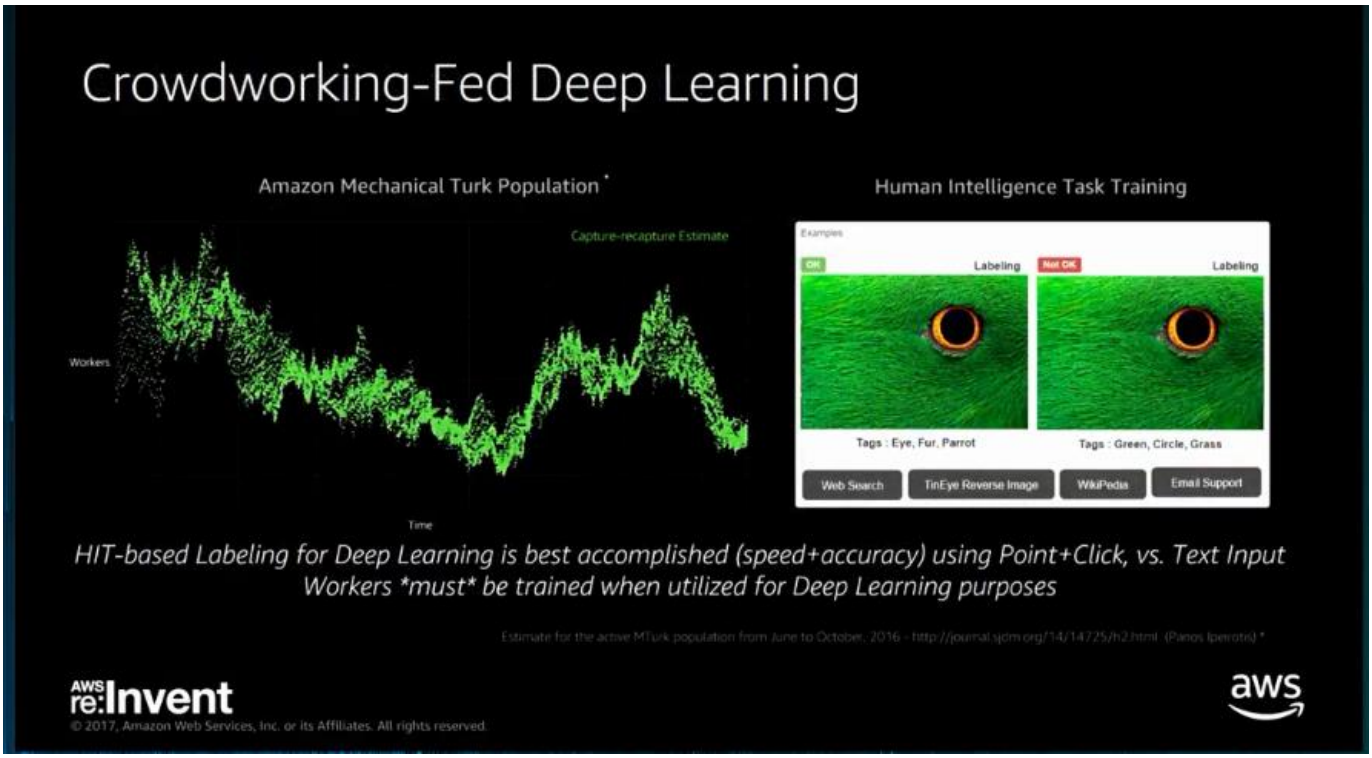
Ground Truth Generation & Niche Image Categorization at scale is time consuming & untenable

- Crowdworking to the rescue
- Deep Learning for Unlabeled Data
- Amazon Mechanical Turk – build machine learning datasets using HITs (human intelligence tasks), Requesters & Workers
- Human Inference can be as high as 100s of HITs/min (2/s @ ~200ms)

IMAGENET - ~50,000 AMT Workers

Price	Task	Images	Labels	Time	Workers
$0.01	Label at least one object in the image	237	678	13.5	37

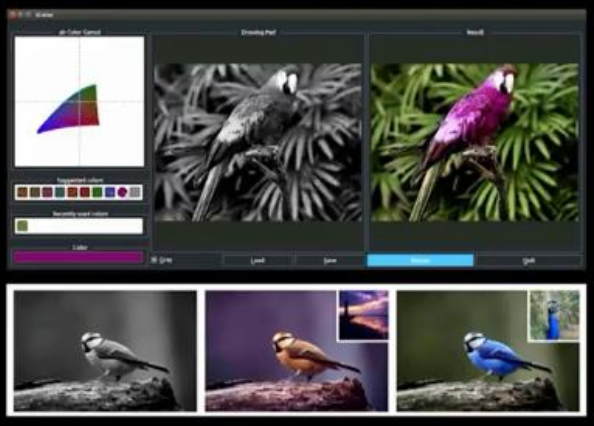
We can build a UI that presents this edge case images to a human user as a Human Intelligence Test and having them providing identification labels for each image displayed. They can either select labels from a set of images or attach tags that they know fits the image.



The left graph shows the population of available human workers while the brighter green dots are the returning people that have returned to help do more labelling. We do spot-checking QA of some of the tagging done by the human worker and we can increase their rating based on the quality of their work. We also encourage the workers to use Wikipedia and google searches to make their description and tagging better to achieve better rating scores.

Back to Our Image Archive

- Custom Concepts
  NLP – Rekognition + spaCy, Others
- Specialized Categories
  Transfer Learning w/ Finetuning
- Black & White Footage
  Deep Learning-based Colorization
- Low Resolutions
  Convolutional Neural Net Image Scaling
- Niche & Historical Context
  Crowdworking & OCR



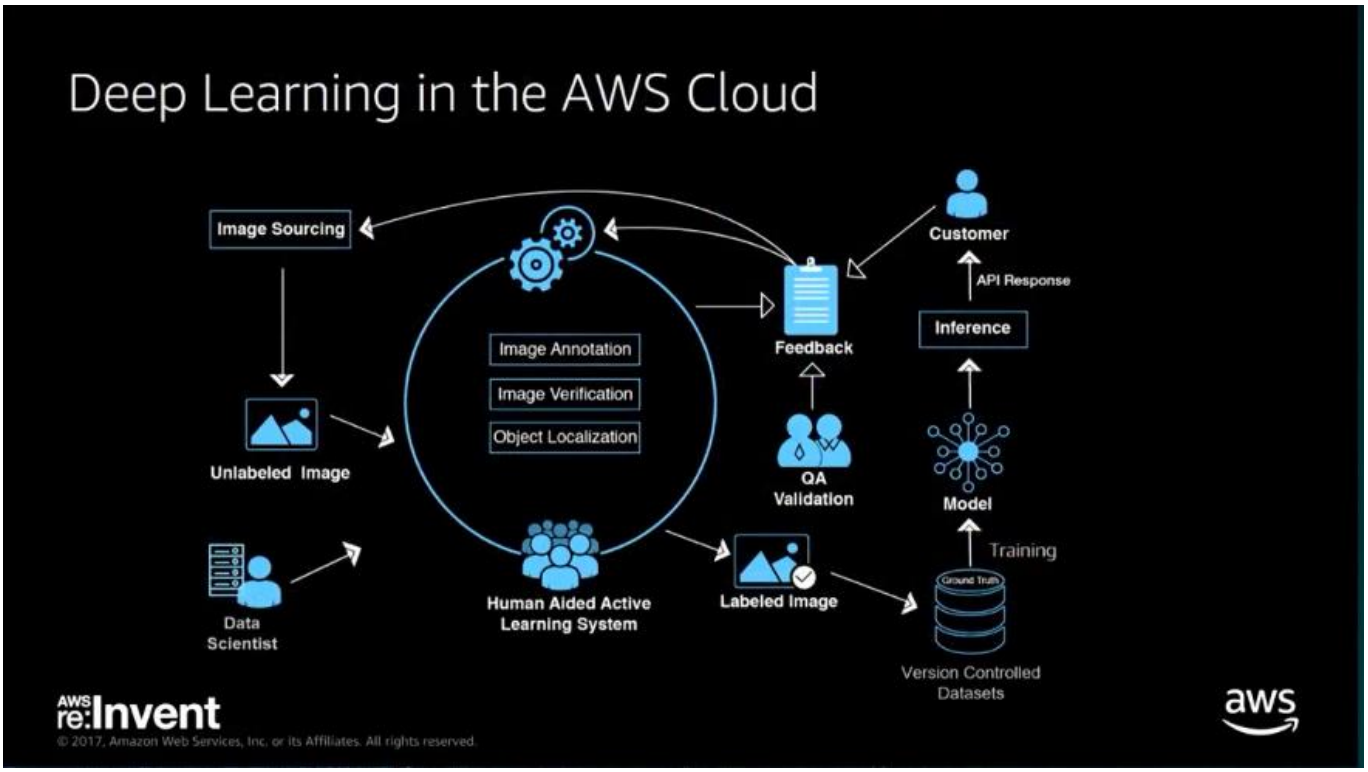
Real-Time User-Guided Image Colorization with Learned Deep Priors

Back to Our Image Archive

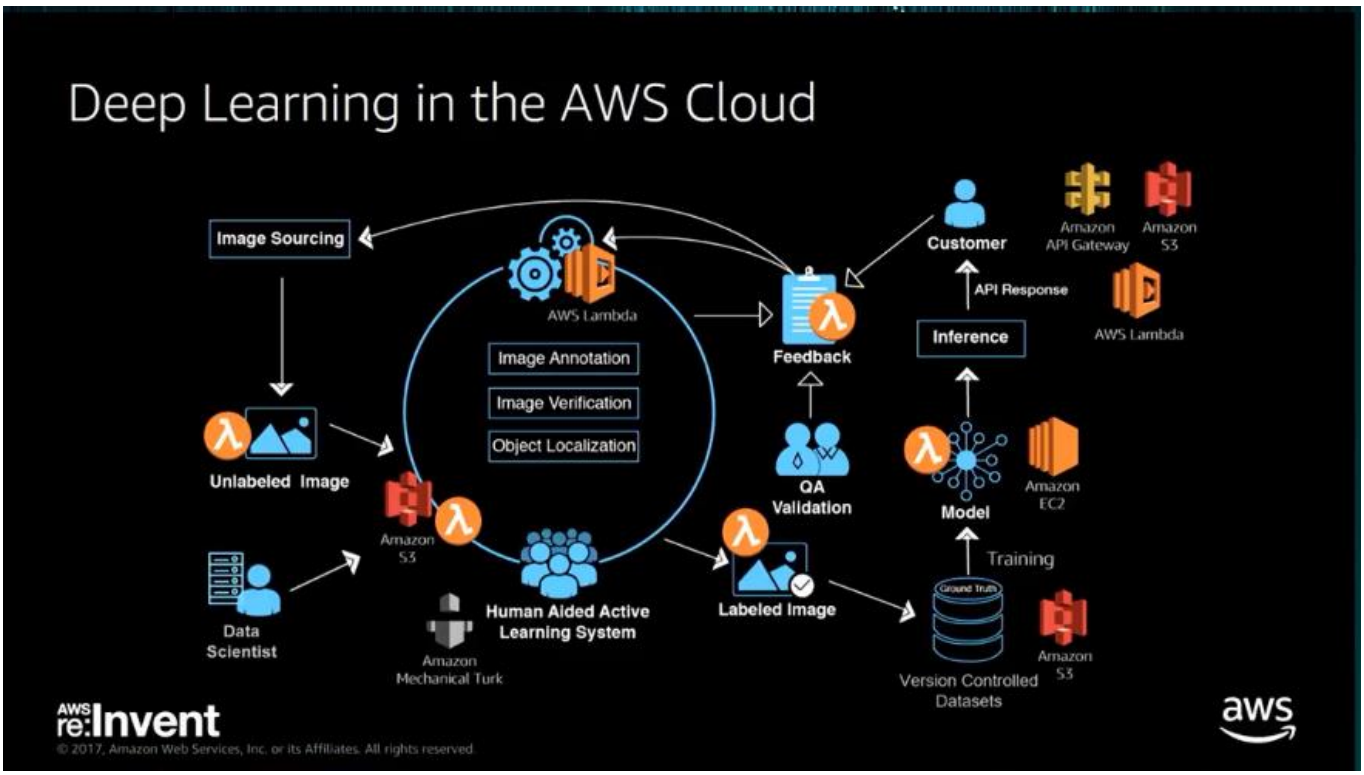
- Custom Concepts
  NLP – Rekognition + spaCy, Others
- Specialized Categories
  Transfer Learning w/ Finetuning
- Black & White Footage
  Deep Learning-based Colorization
- Low Resolutions
  Convolutional Neural Net Image Scaling
- Niche & Historical Context
  Crowdworking & OCR

Real-Time User-Guided Image Colorization with Learned Deep Priors

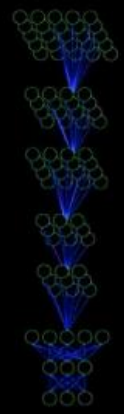
aws re:Invent
© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.
aws



The pipeline has a lot of lambdas and step functions. The main steps are content landing in S3, content gets picked up by a lambda and processed, then lands back on S3 again. We can use multiple buckets and auto-scaling groups with different deep learning algorithms to look at different bucket locations. We then use S3 triggers on the tags/labels to funnel the results into SNS queues for the appropriate deep learning infrastructure pipelines.



Key Takeaways & Wrap-Up



- There is no 'Magic Bullet'
- Highly specialized workloads still require cross-functional skills across DevOps, Development, Data Scientist & Media ecosystems
- Don't overcomplicate the pipeline & infrastructure
- Managed Deep Learning & AI Services are making more niche use cases viable via reduction of cost, development and time to market
- Target managed services to solve ' the 80% problem', then iterate
- When you have to, utilize compute diversification across GPU, CPU & FPGA, combined with Object Storage & Fractional Billing

AWS re:Invent

THANK YOU!

AWS re:Invent
© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

aws

Related Sessions

MAE403 – OTT State of Play: Innovation at Netflix, Hulu, Amazon Video and AWS Elemental

MAE404 – Enhanced Media Workflows on AWS using Amazon AI Services

MCL314 – Unlocking Media Workflows Using Amazon Rekognition

MCL318 – Deep Dive on Amazon Rekognition Architectures for Image Analysis