Welcome to:

AWS Step Functions in the wild!

Tim Bray, Senior Principal Technologist, AWS (@timbray, https://www.tbray.org/)

Patrick Brandt, Coca-Cola, IT Director and Solutions Strategist, Marketing Operations, E-Commerce, and R&D

AWS re:Invent

aws

This time last year:

Since then: **Regions**: Northern Virginia, Ohio, Oregon, Ireland, Frankfurt, London, Sydney, Tokyo

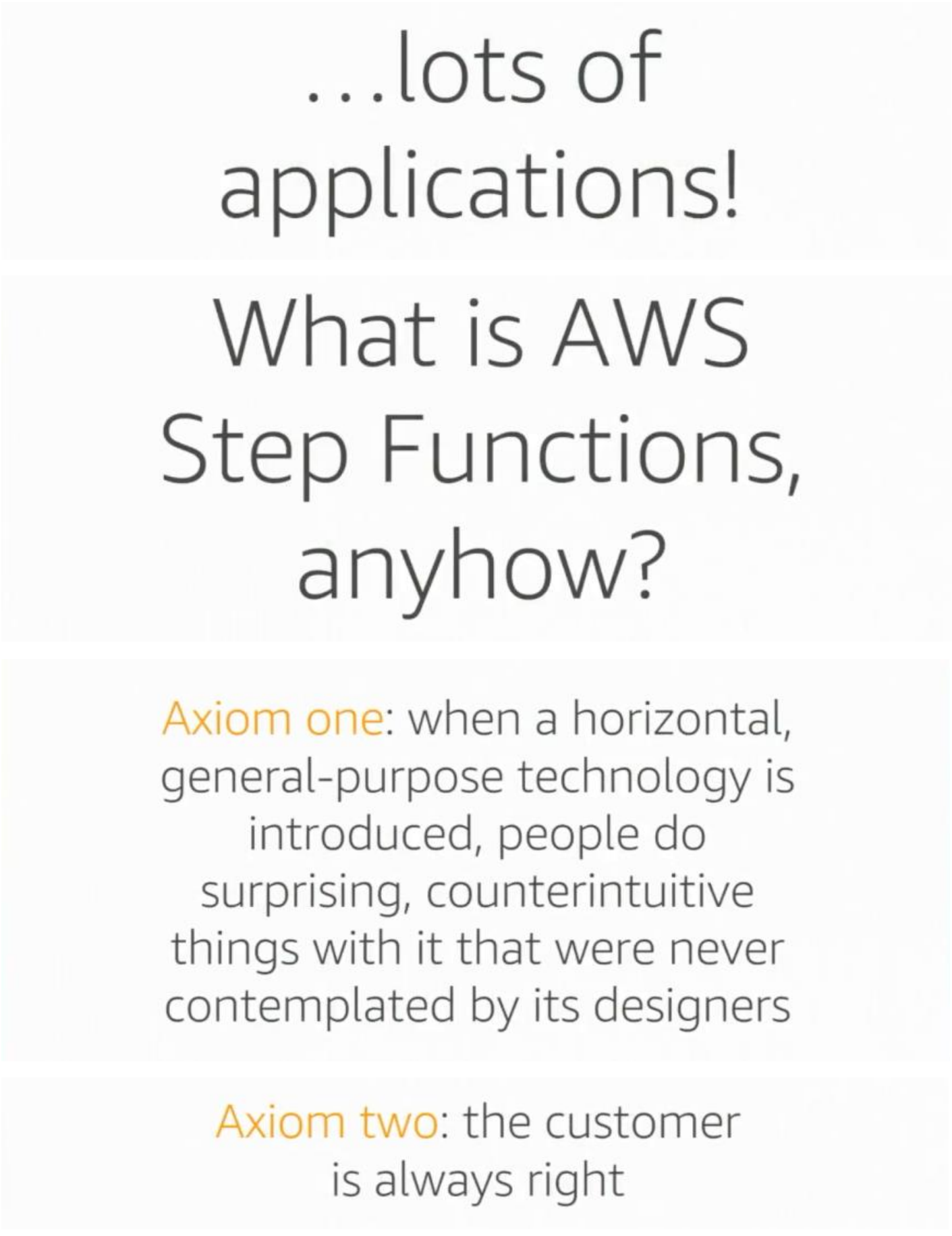
Integrations: AWS CloudFormation, Amazon CloudWatch Events, Amazon API Gateway

Aaaaand...

AWS re:Invent

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

aws



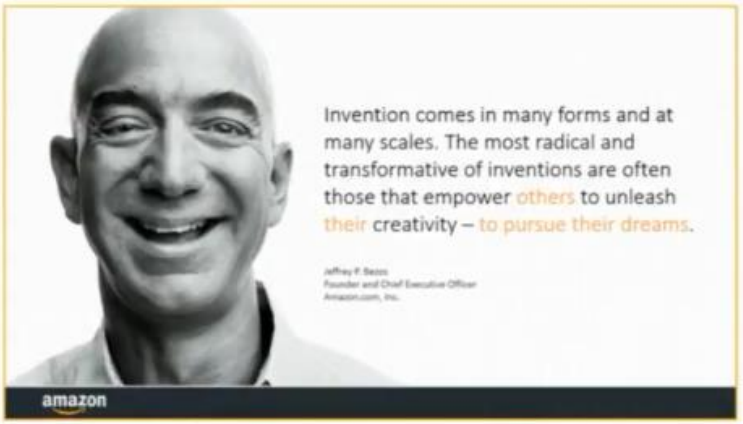
Step Functions are very useful for replacing your bash scripts, while also providing you with logging, audit trail, retries and concurrency than bash is capable of.

AWS Blog

Things Go Better With Step Functions

by Jeff Barr | on 05 OCT 2017 | in AWS Step Functions, Customer Success | Permalink | Comments | Share

I often give presentations on Amazon's culture of innovation, and start out with a slide that features a revealing quote from Amazon founder Jeff Bezos:



I love to sit down with our customers and to learn how we have empowered their creativity and to pursue their dreams. Earlier this year I chatted with Patrick Brandt from The Coca-Cola Company in order to learn how they used AWS Step Functions and other AWS services to support the Coke.com Vending Pass program. This program includes drink rewards earned by purchasing products at vending

EARN A DRINK REWARD AFTER 10 PURCHASES!

Vending Pass problem

"I'm stoked! I've got enough credits on my Vending Pass for a free drink!"

Vending Pass problem

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

Vending Pass problem

"I changed my mind, I'm not going to pay with my Vending Pass - I'm going to pay cash."

Later that day…

Vending Pass problem

!

"Uh Oh! I didn't pay with my Vending Pass, but the balance in my mobile wallet makes it look like I got charged the credit anyway!"

Vending Pass problem

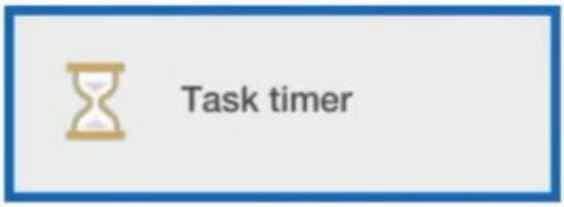
1. Thirsty consumer presents card, can buy a drink with Vending Pass or debit
2. But mobile wallet display can get out of sync under certain conditions
3. 

Vending Pass problem

- Just wait for backend to catch up
- Create a two-step state machine:
  - Wait (90 seconds)
  - Update mobile wallet
- Cheap and simple!

Task timer blueprint

Scheduling future work with Step Functions



docs.aws.amazon.com/step-functions/latest/dg/tutorial-creating-task-timer.html

FDA nutrition label reform

1990

Nutrition Facts
Serving Size 2/3 cup (55g)
Servings Per Container About 8

Amount Per Serving

Calories 230	Calories from Fat 72
	% Daily Value*
Total Fat 8g	**12%**
Saturated Fat 1g	**5%**
Trans Fat 0g	
Cholesterol 0mg	**0%**
Sodium 160mg	**7%**
Total Carbohydrate 37g	**12%**
Dietary Fiber 4g	**16%**
Sugars 1g	
Protein 3g	
Vitamin A	10%
Vitamin C	8%
Calcium	20%
Iron	45%

* Percent Daily Values are based on a 2,000 calorie diet. Your daily value may be higher or lower depending on your calorie needs.

	Calories:	2,000	2,500
Total Fat	Less than	65g	80g
Sat Fat	Less than	20g	25g
Cholesterol	Less than	300mg	300mg
Sodium	Less than	2,400mg	2,400mg
Total Carbohydrate		300g	375g
Dietary Fiber		25g	30g

2016

Nutrition Facts
8 servings per container
Serving size **2/3 cup (55g)**

Amount per serving
Calories **230**

	% Daily Value*
Total Fat 8g	**10%**
Saturated Fat 1g	**5%**
Trans Fat 0g	
Cholesterol 0mg	**0%**
Sodium 160mg	**7%**
Total Carbohydrate 37g	**13%**
Dietary Fiber 4g	**14%**
Total Sugars 12g	
Includes 10g Added Sugars	**20%**
Protein 3g	
Vitamin D 2mcg	10%
Calcium 260mg	20%
Iron 8mg	45%
Potassium 235mg	6%

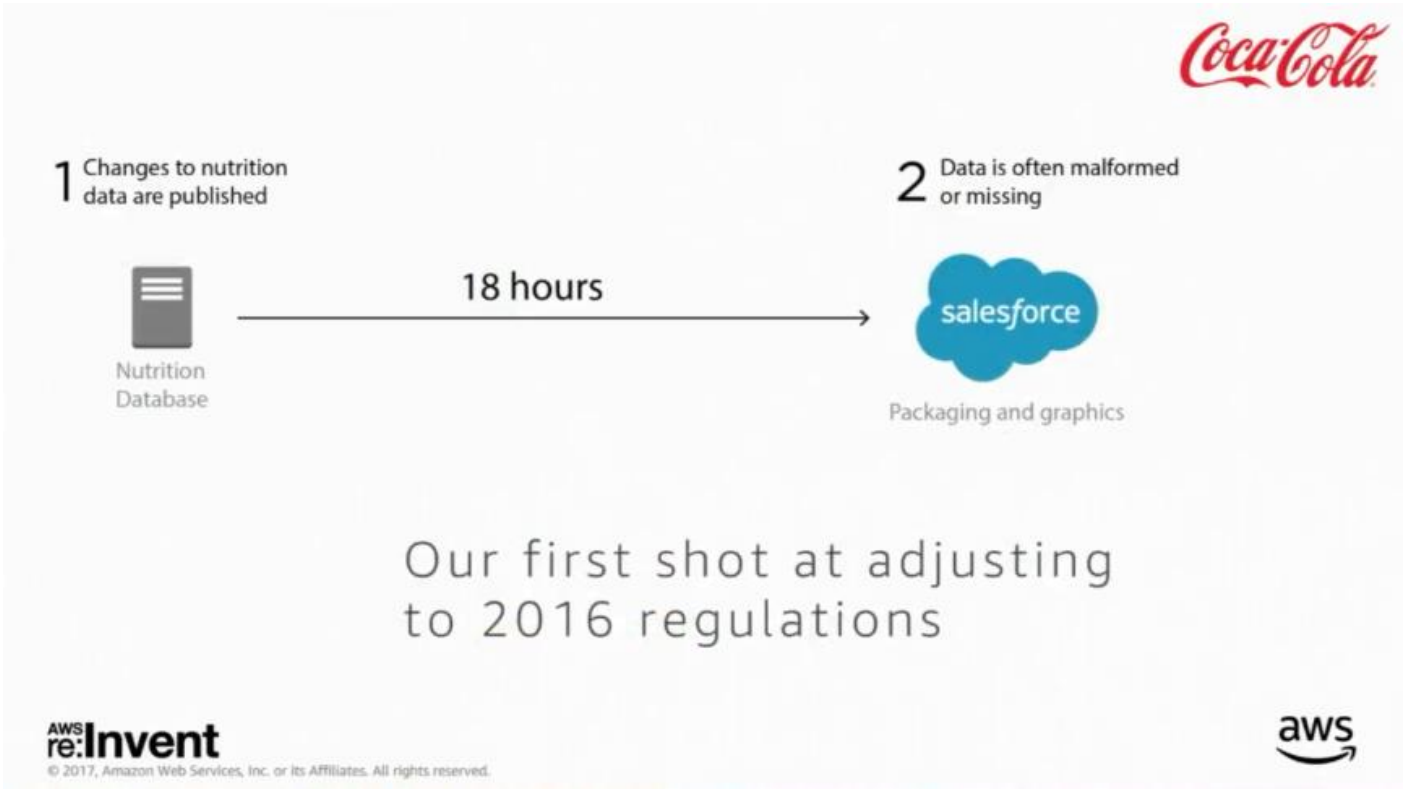
* The % Daily Value (DV) tells you how much a nutrient in a serving of food contributes to a daily diet. 2,000 calories a day is used for general nutrition advice.

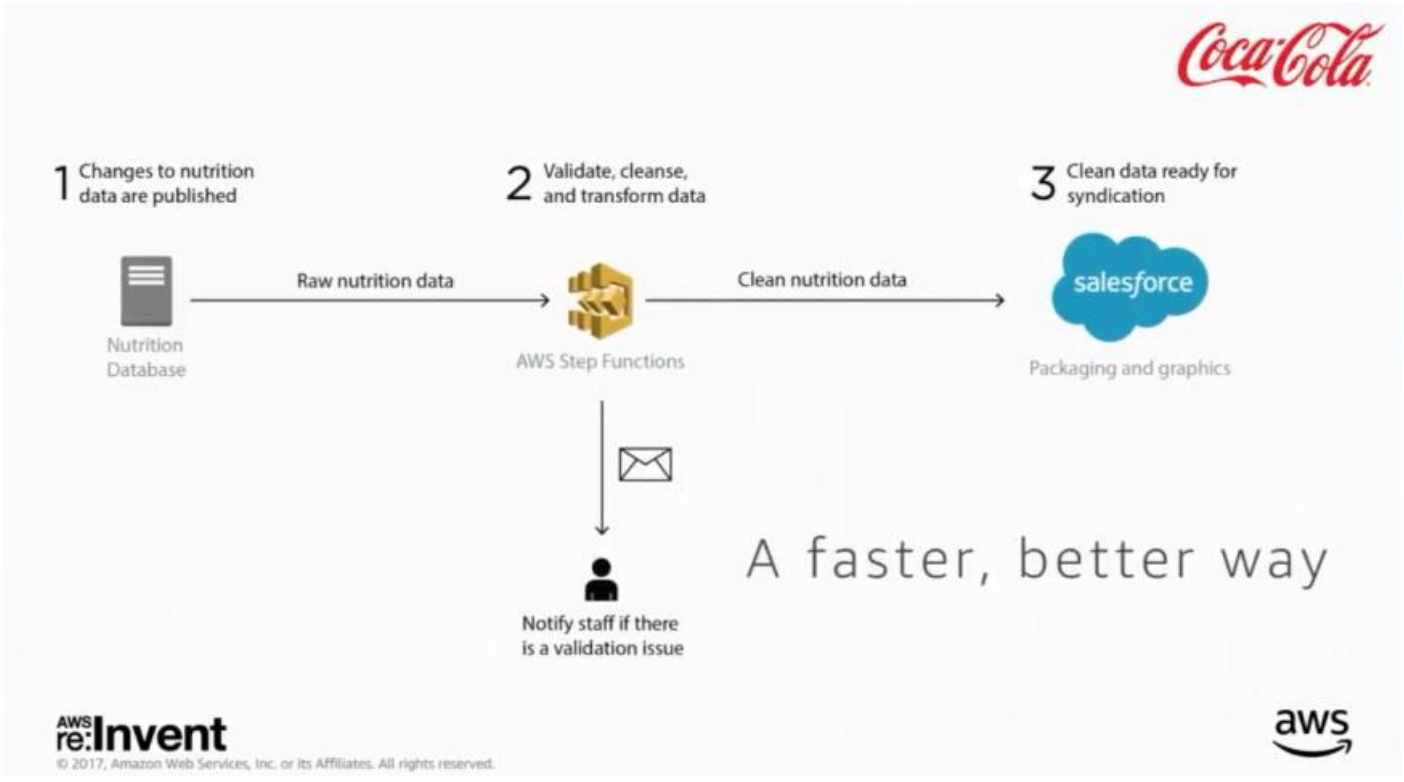
This is a big deal for The Coca-Cola Company!

Coca-Cola · Coca-Cola zero SUGAR · POWERADE · Diet Coke · Fanta
FRESCA · Barq's · DASANI · DASANI SPARKLING · GOLD PEAK TEA
Mello Yello · Sprite · Pibb Xtra · Minute Maid · Seagram's

Our food service partners need nutrition data too

Coca-Cola · Coca-Cola zero SUGAR · POWERADE · Diet Coke · Fanta
FRESCA · Barq's · DASANI · DASANI SPARKLING · GOLD PEAK TEA
Mello Yello · Sprite · Pibb Xtra · Minute Maid · Seagram's


1 Changes to nutrition data are published
2 Data is often malformed or missing
18 hours
Nutrition Database
salesforce
Packaging and graphics
Our first shot at adjusting to 2016 regulations
AWS re:Invent
© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.
aws

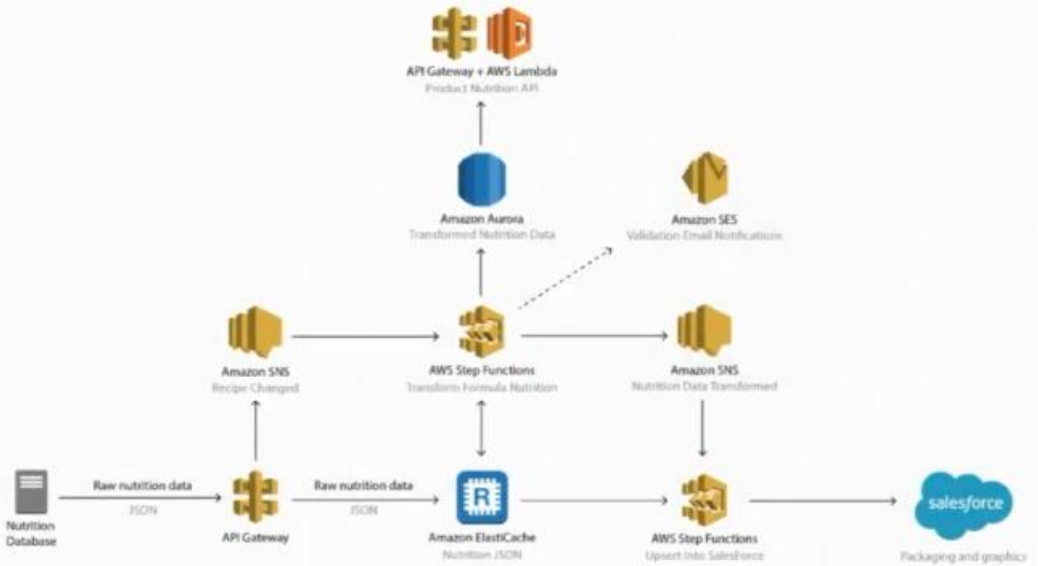

1 Changes to nutrition data are published
2 Validate, cleanse, and transform data
3 Clean data ready for syndication
Raw nutrition data
Clean nutrition data
Nutrition Database
AWS Step Functions
salesforce
Packaging and graphics
Notify staff if there is a validation issue
A faster, better way
AWS re:Invent
© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.
aws

Collaboration through Step Functions

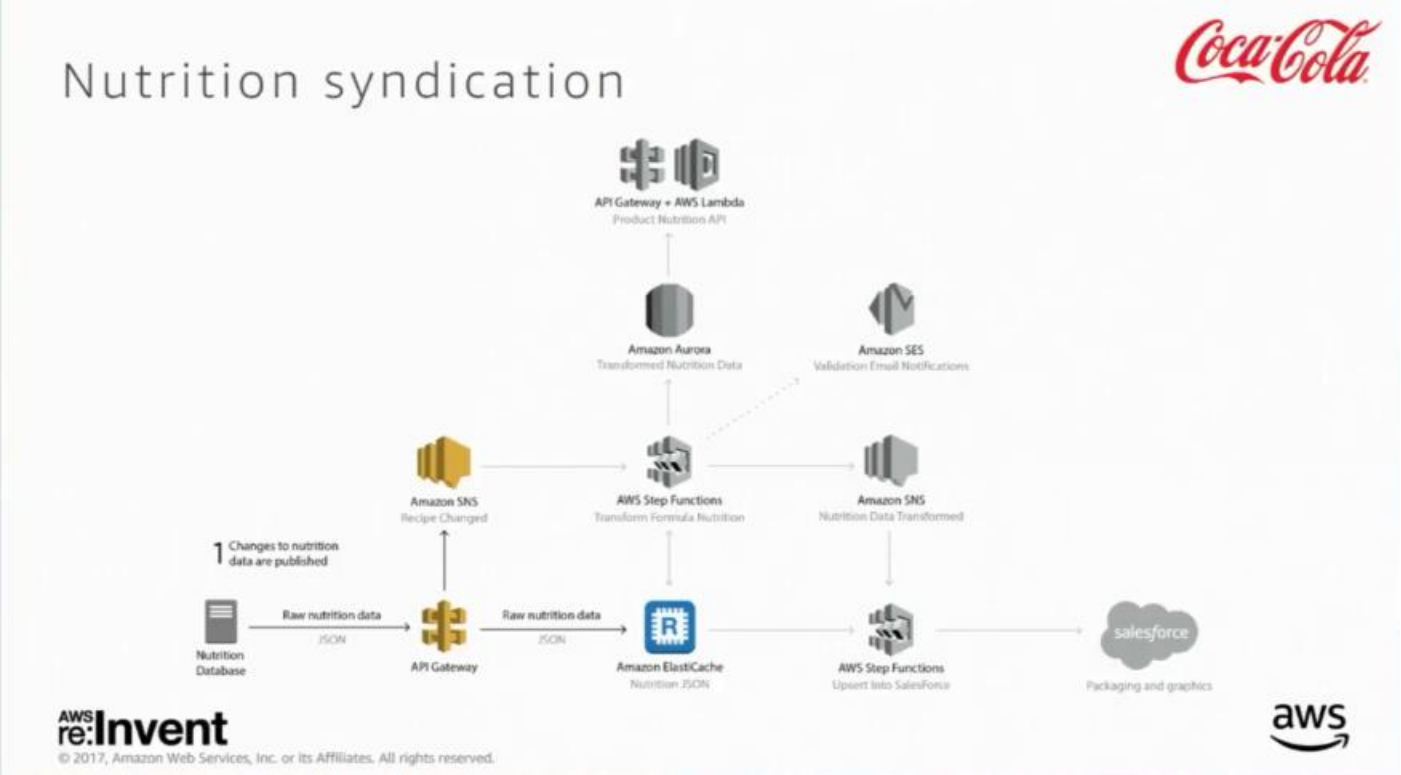
- Data validation and transformation steps are designed visually, directly with SRA personnel
- Validation and transformation steps verified in real-time as nutrition data flows through the state machine
- Process optimizations are identified and implemented on the spot

*Dramatic re-enactment

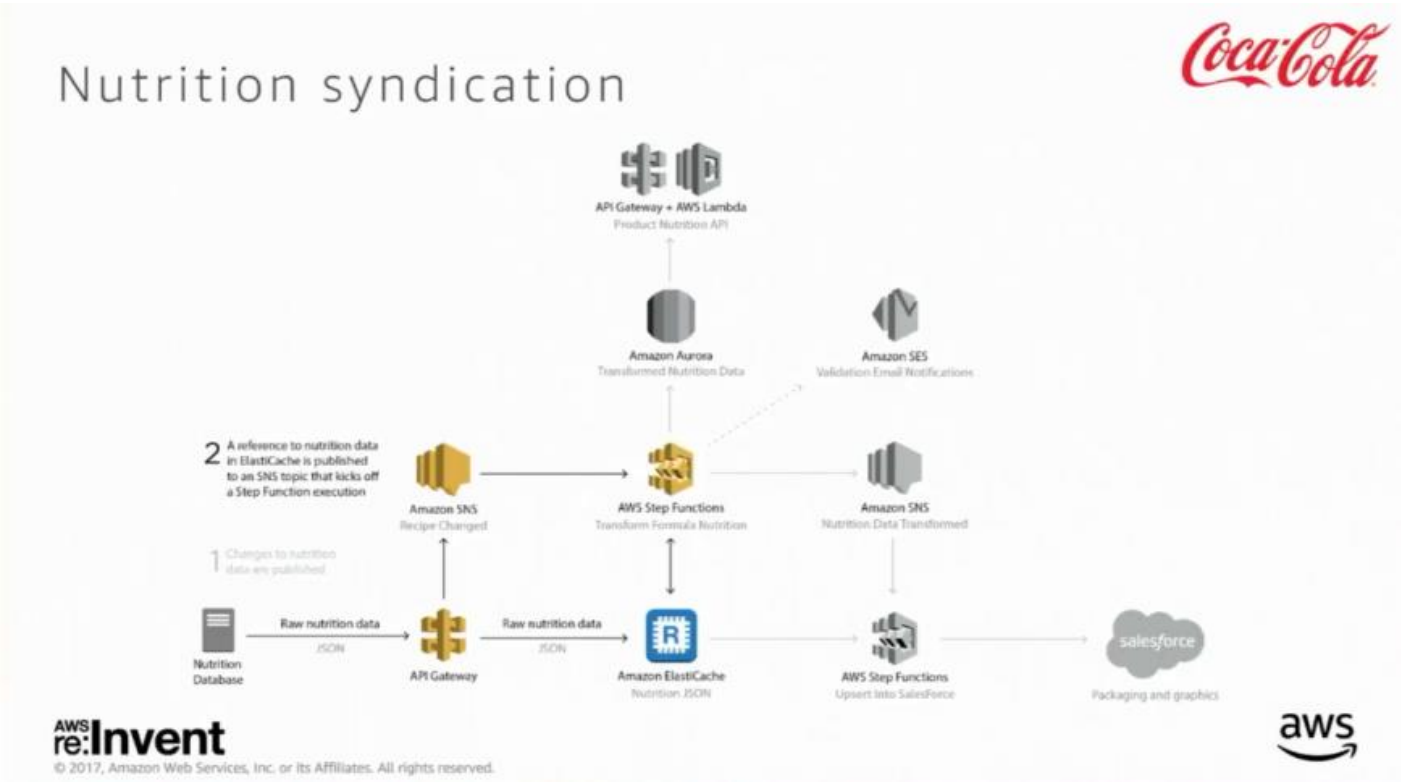
Nutrition syndication



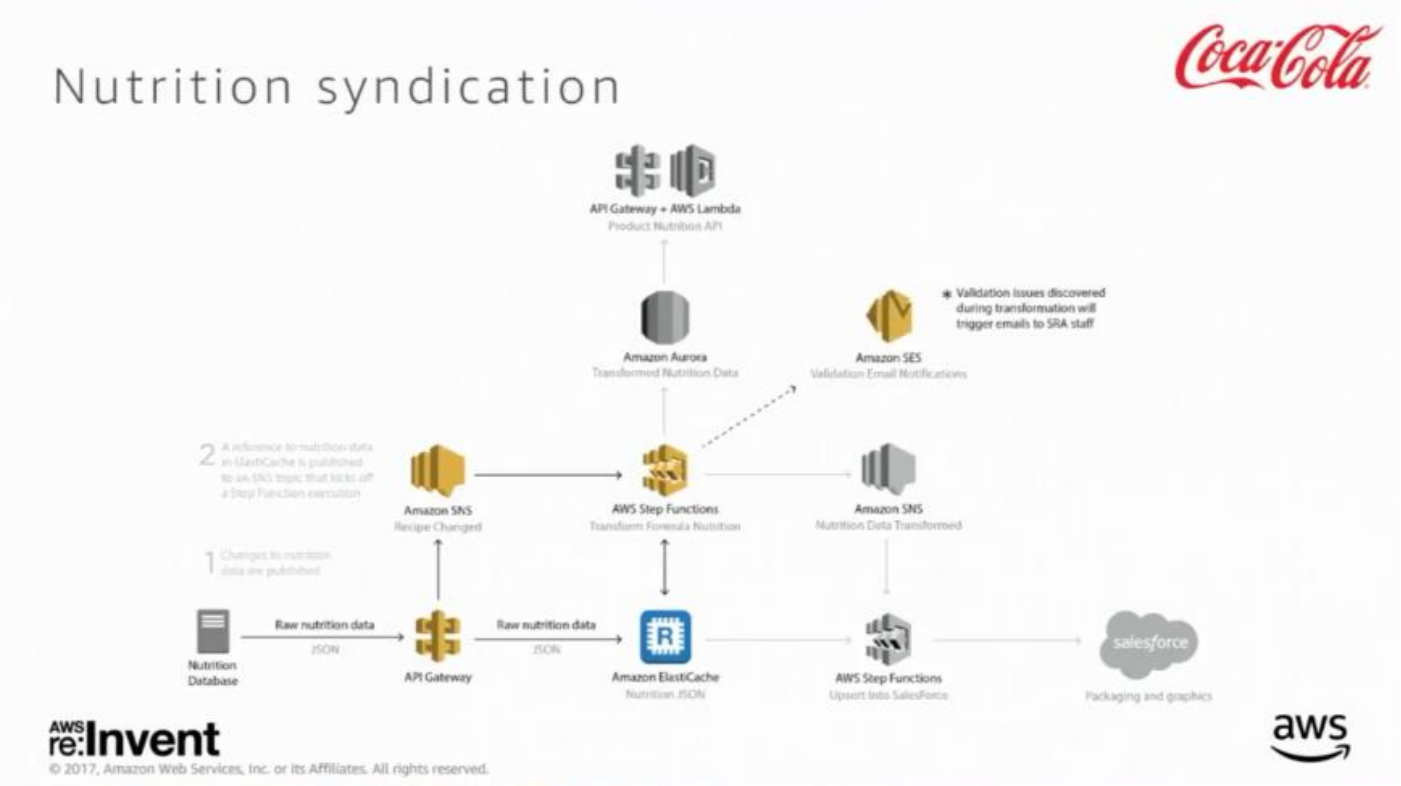
This is the entire architecture of the system.



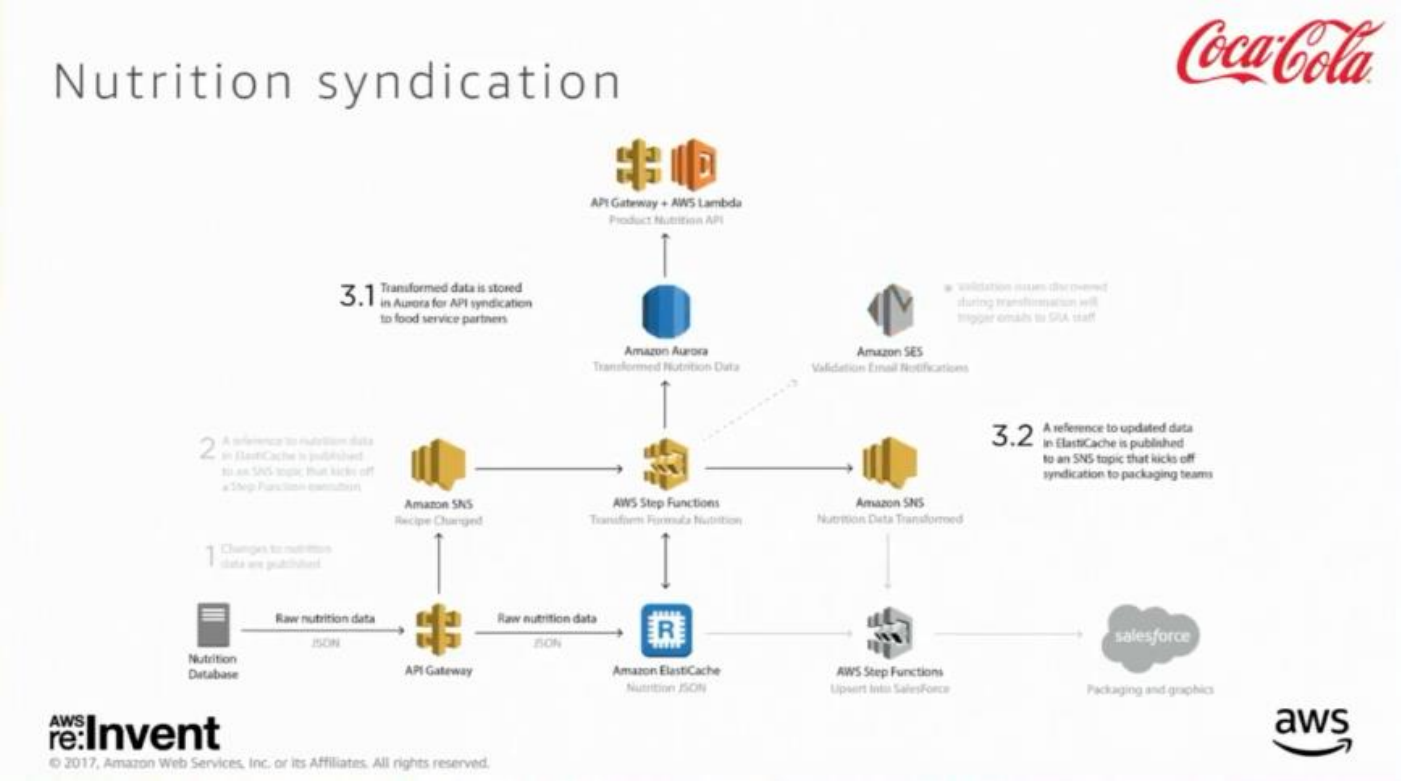
Step 1: changes in nutrition data are published and stored in the Nutrition Database, this sends a raw JSON payload request to the API Gateway endpoint, that JSON is stored in ElasticCache as a key/value pair, the key is published to an SNS topic.



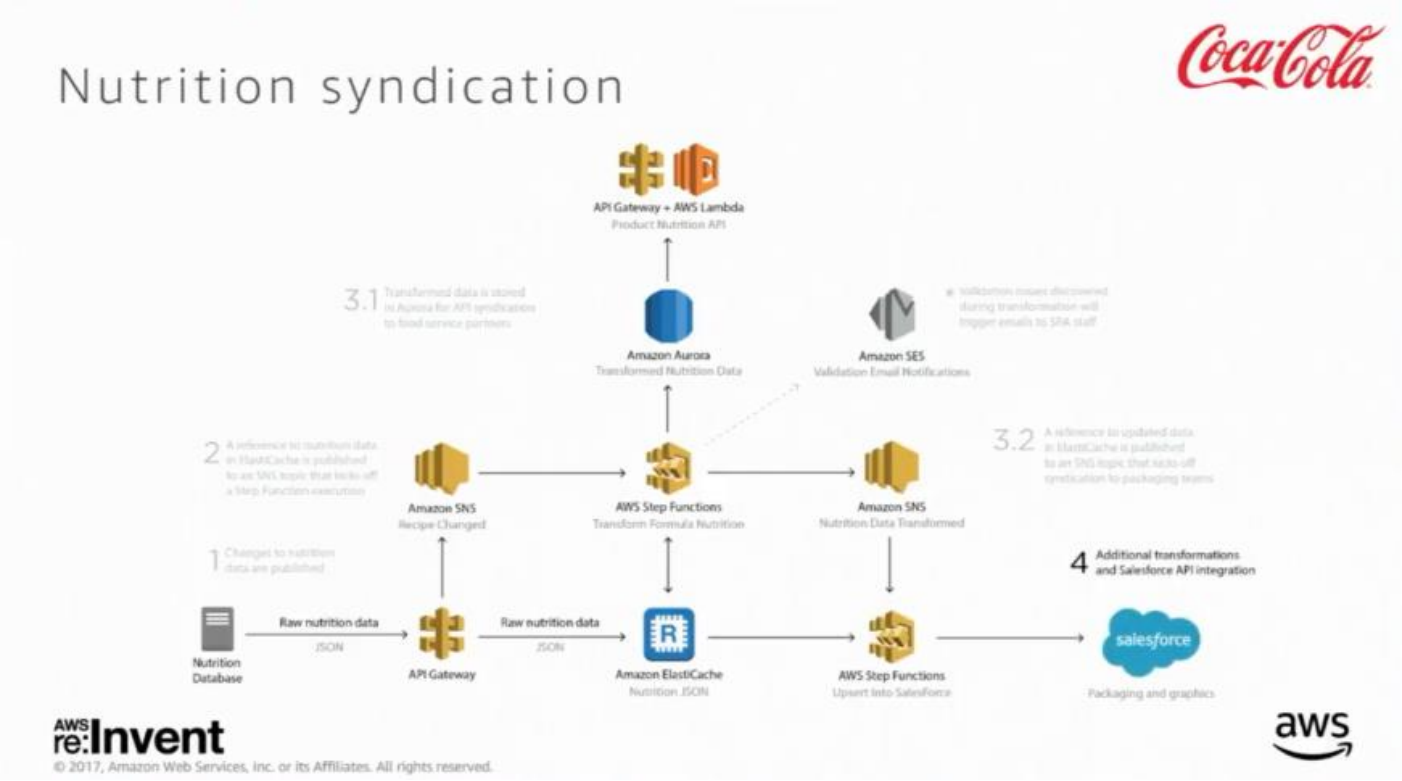
The publishing to an SNS topic then initiates a Step Function which operates against the data that was stored in ElasticCache.



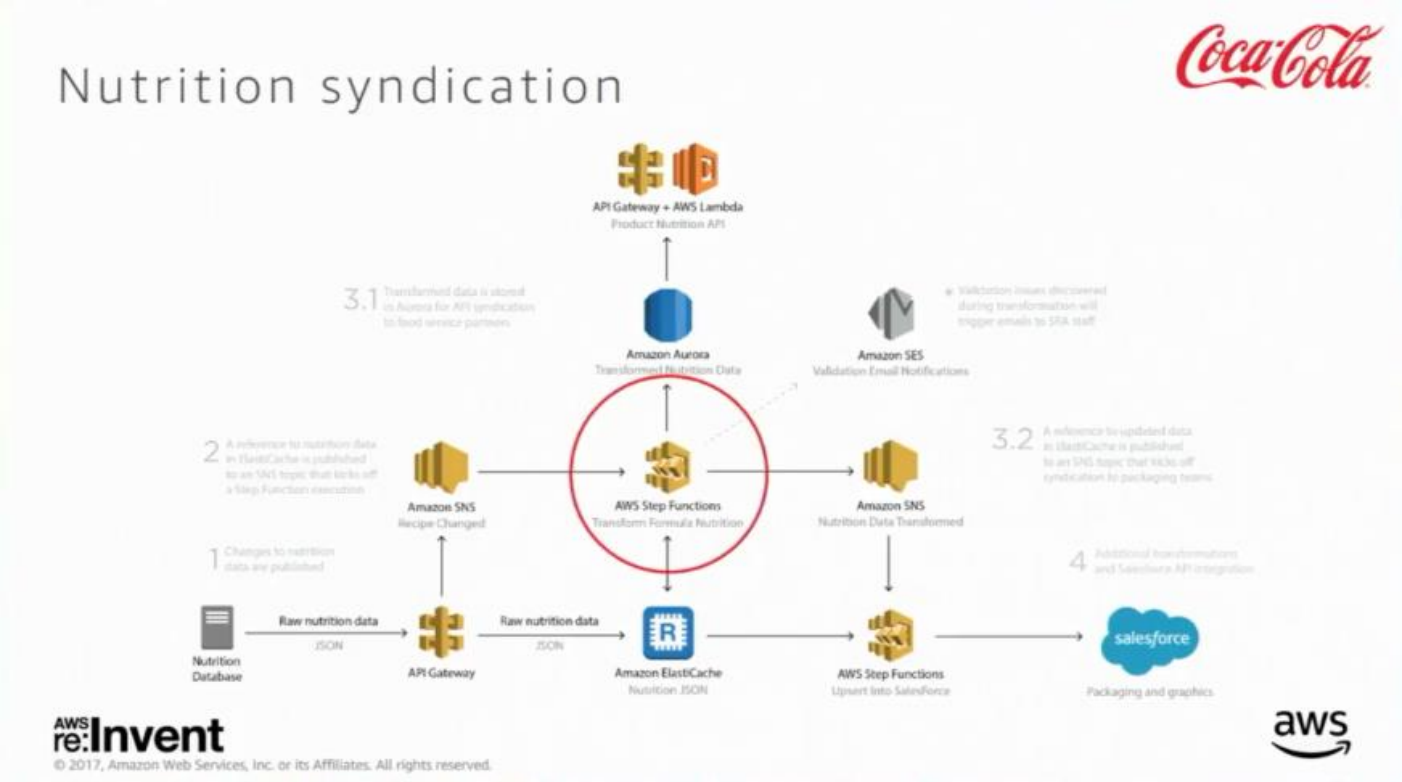
If an issue with the data is detected, an email is sent via SES to someone in the SRA group that can fix the issue.



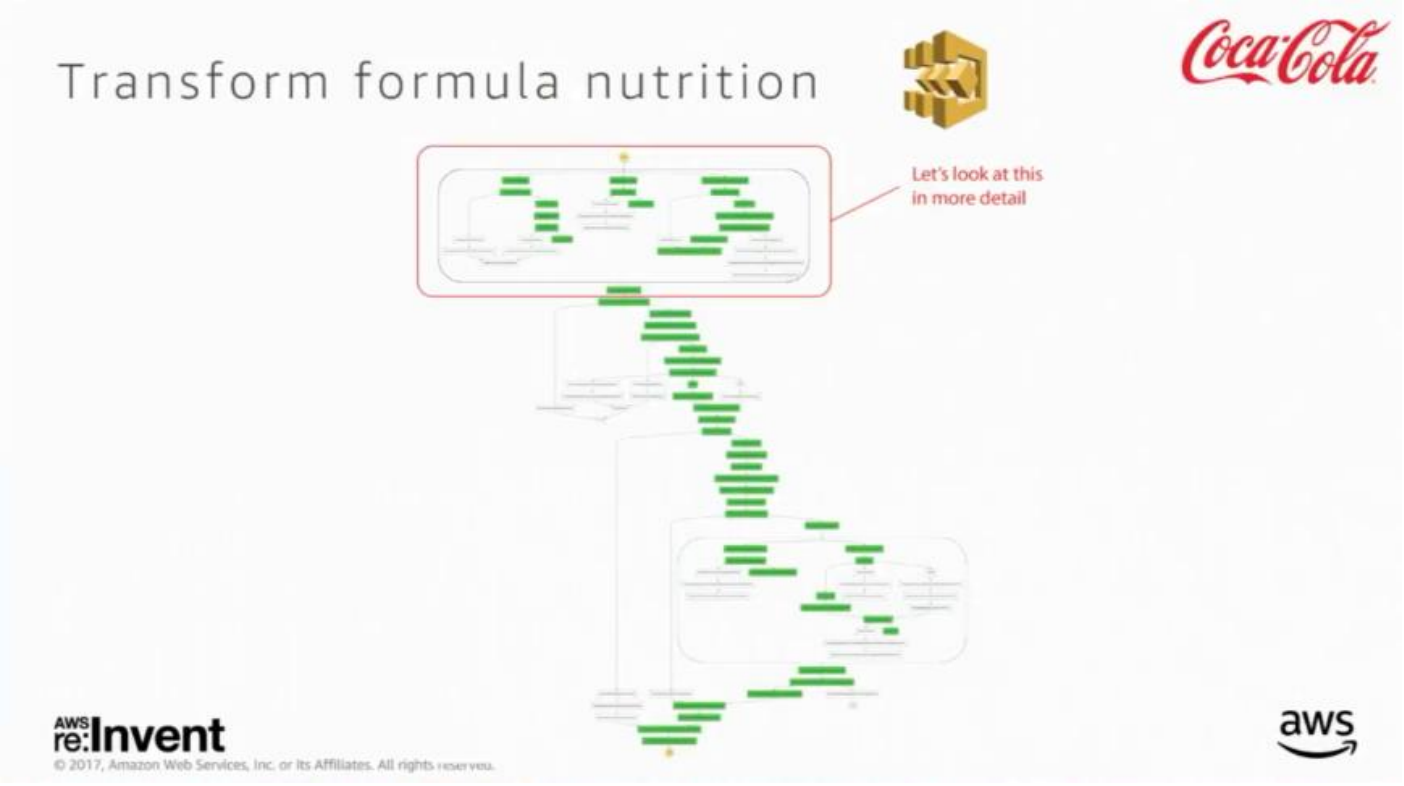
Provided everything is okay, the data flow forks out into 2. On one part, the data is stored in Amazon Aurora that effectively publishes the data through an API Gateway and Lambda so that our food service partners can pick up our nutrition data. The other part is that a message is published through another SNS topic that triggers the execution of another step function as below



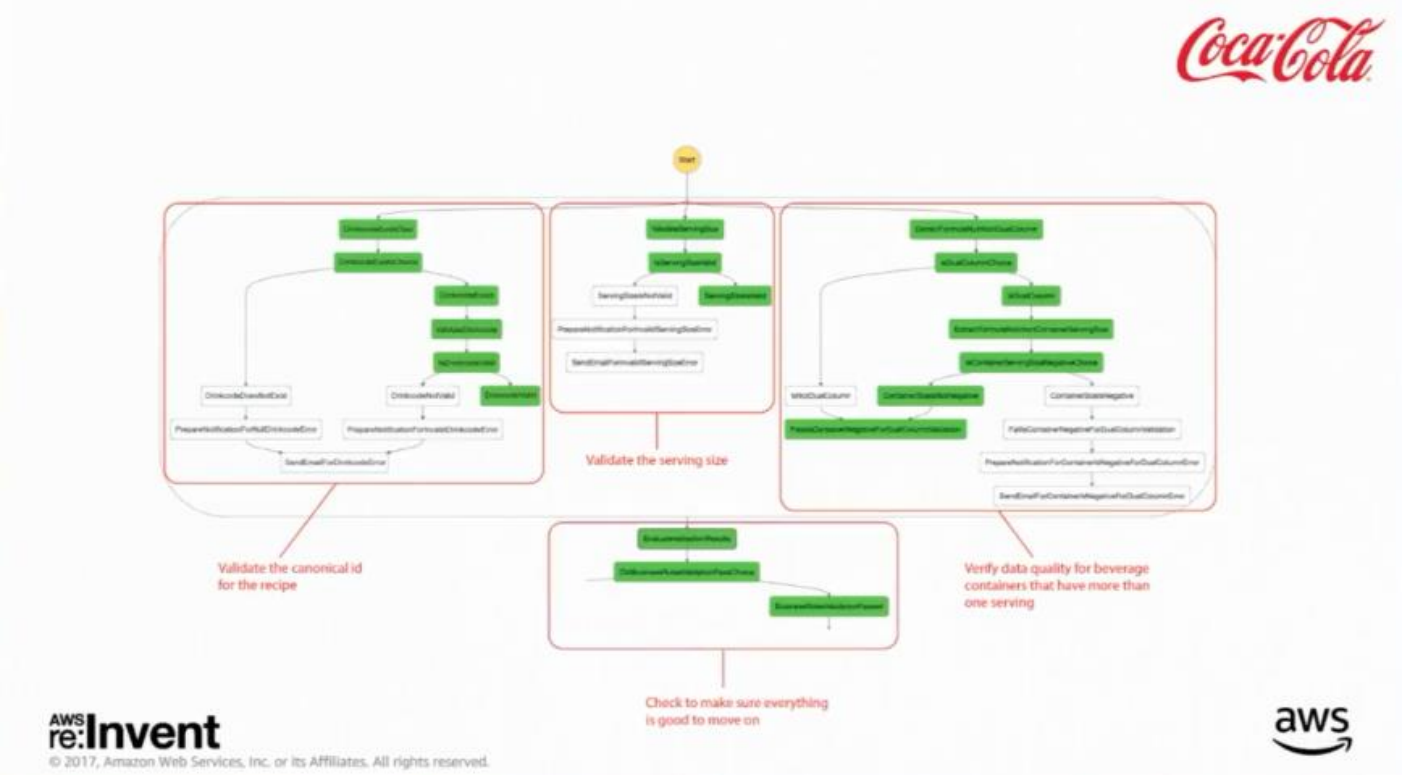
The step function takes the data and molds it into a format that it can be easily ingested into Salesforce



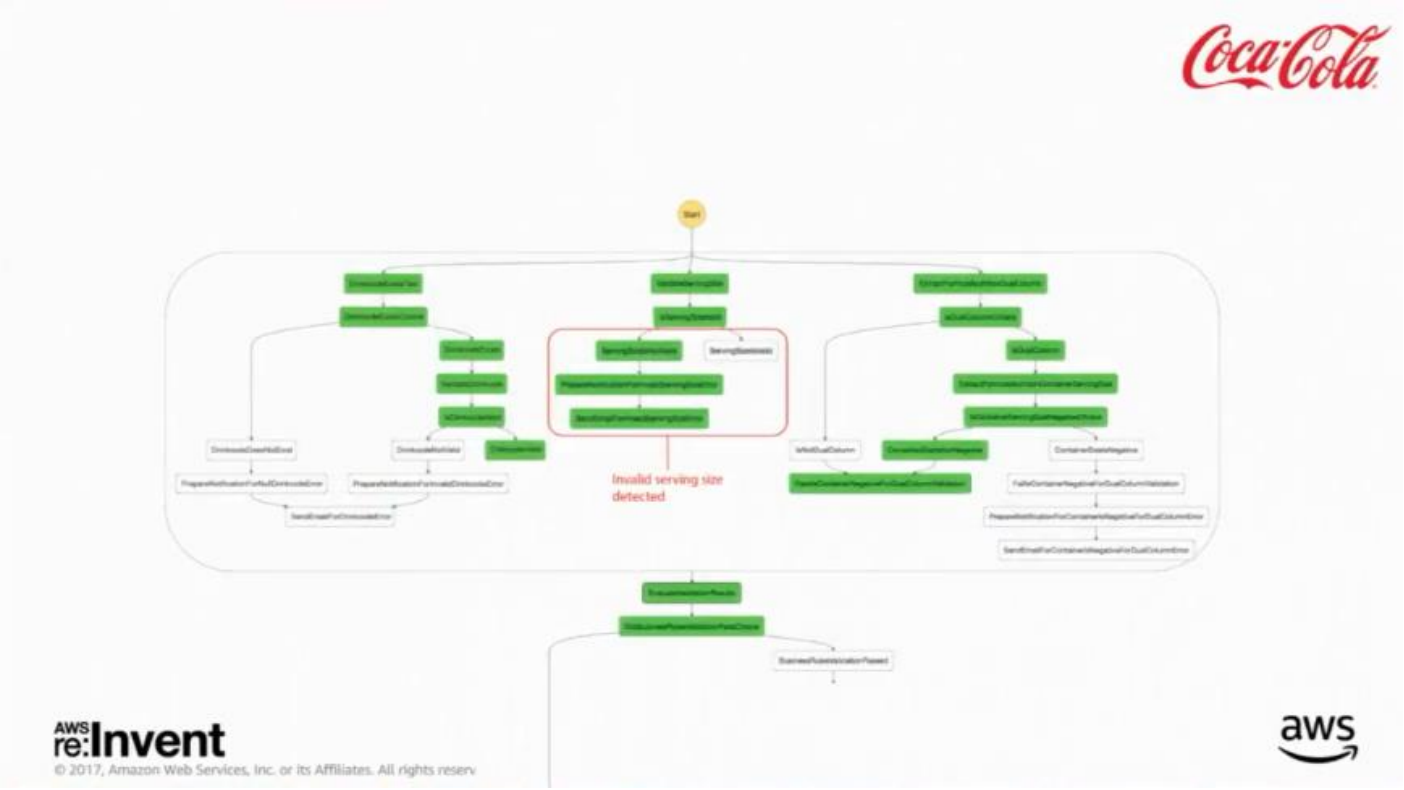
This particular ***Transform Formula Nutrition step function*** is the hub of this entire architecture.



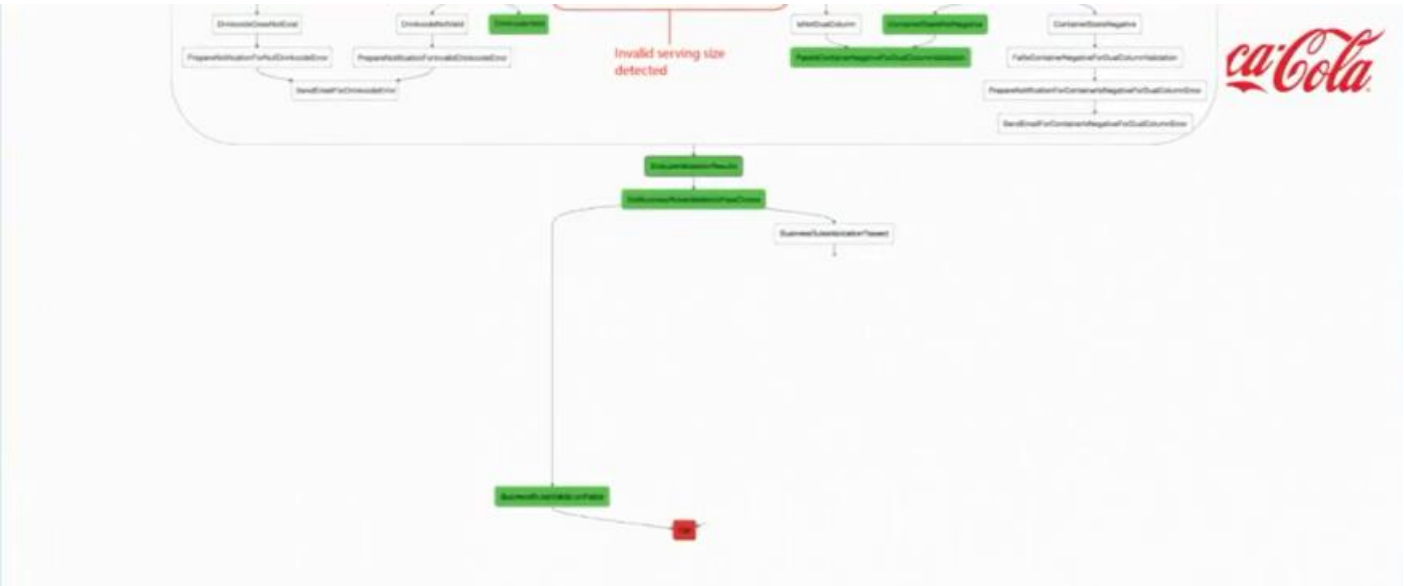
This is a huge state machine with 94 separate tasks.



These are the first set of parallel mini-state machines within the larger state machine. Each one of these smaller state machines result in a Boolean value, true or false. These Boolean values are then aggregated by a set of tasks further downstream to determine which branch should be executed



If this task gives a false value, execution takes the path of getting the specific error condition into a state ready to be emailed, then another step function sends out the email as a task.



This Boolean value then causes the entire state machine to flow into a failed state that ends the entire step function execution.

Takeaways

Reducing the distance between business language and developer language: win

Validation and transformation in real-time: win

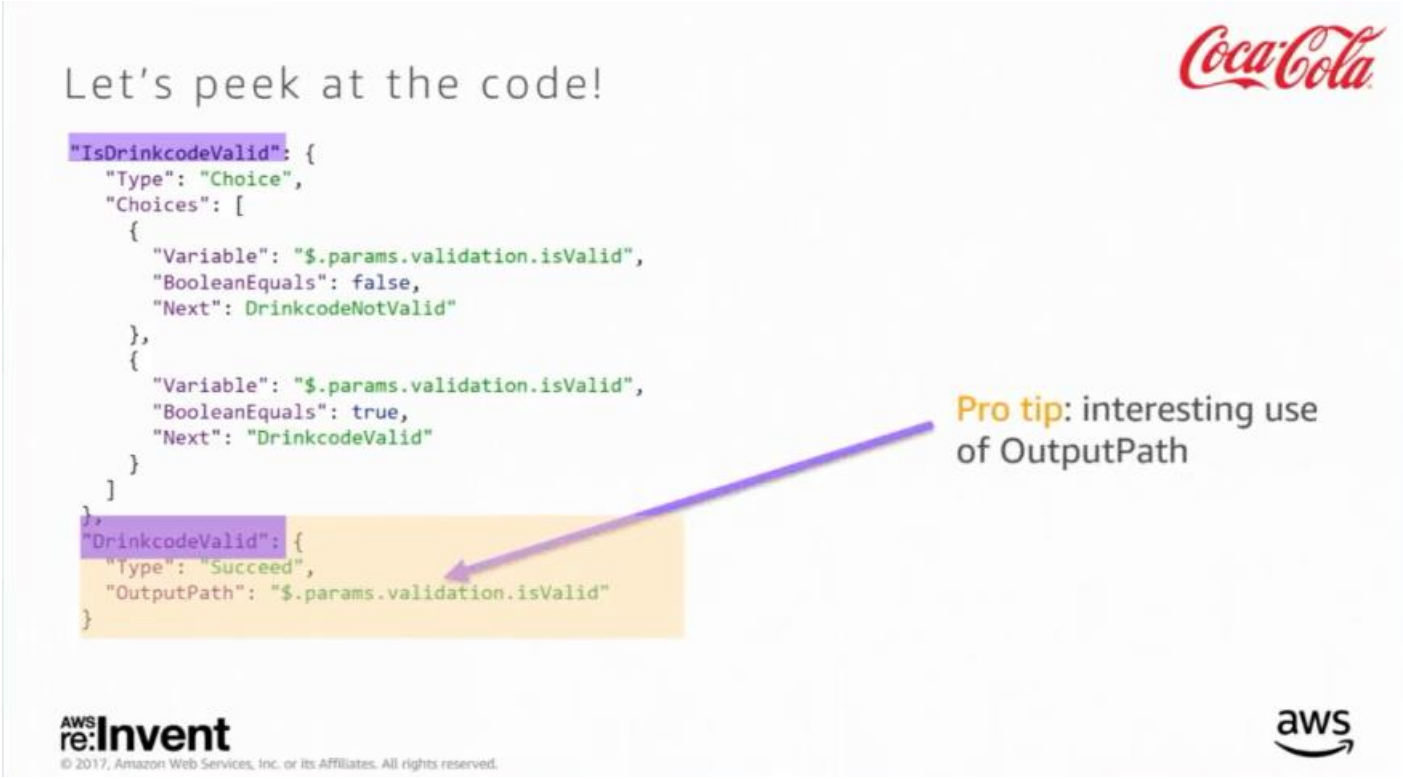
Managing large, complex state machines: open problem

Let's peek at the code!

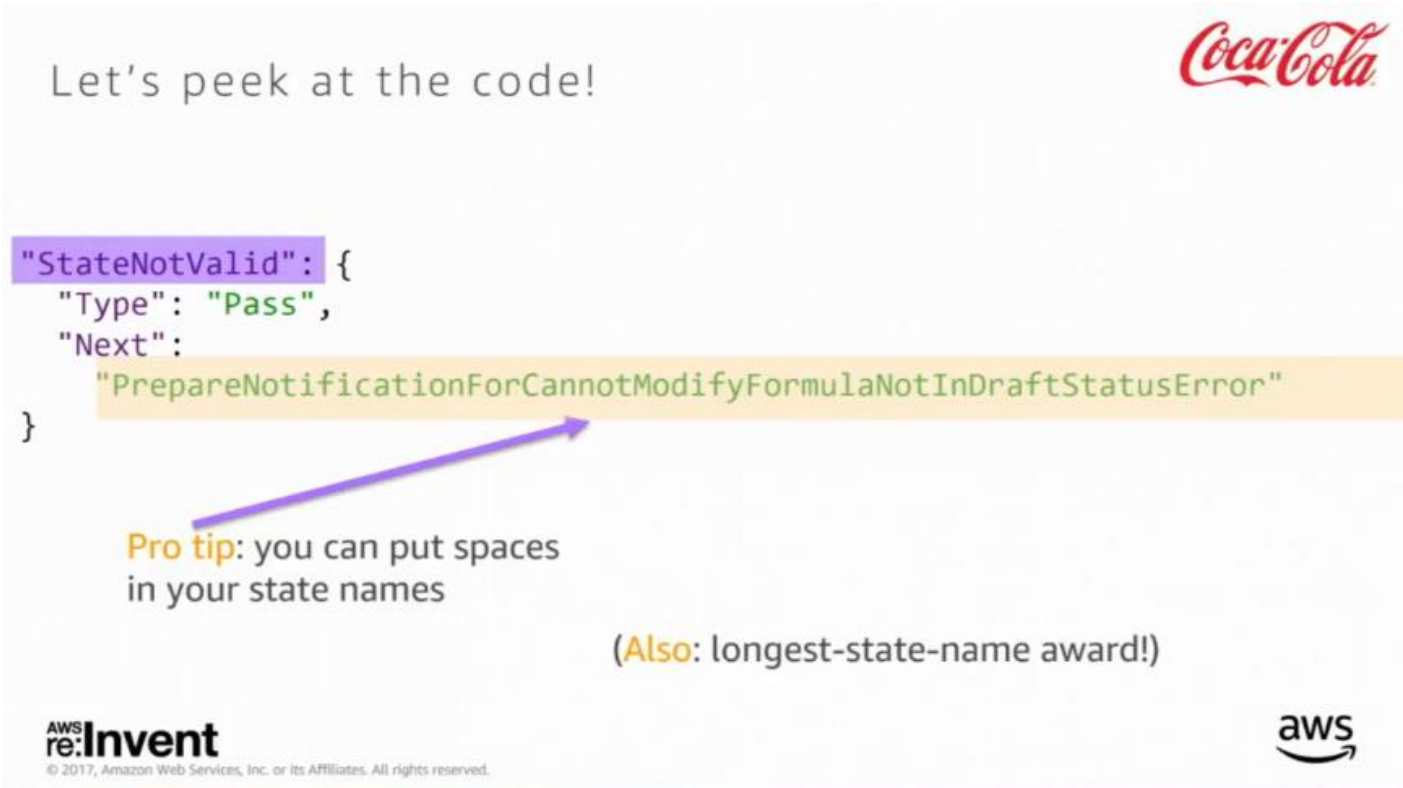
```
"Comment": "Process to transform a FormulaNutrition change",
"StartAt": "ValidateFormulaNutritionAgainstBusinessRule",
"States": {
  "ValidateFormulaNutritionAgainstBusinessRules": {
    "Type": "Parallel"
    "Branches": [
      {
        "StartAt": "DrinkcodeExistsTest",
        "States": {
          "DrinkcodeExistsTest": {
            "Type": "Task",
            "Resource": "arn:aws:lambda:xxx:xxx:function:DrinkcodeExistsTest",
            "ResultPath": "$.params.validation.isValid",
            "Next": "DrinkcodeExistsChoice"
          }
      }, . . .
```

Note: 39 distinct Lambda functions!

Each task is handled by a granular separate lambda



This helps to end the execution of the running state machine while passing out the result of the state machine



WHOSAY

About WHOSAY

- WHOSAY is the largest and most trusted influence marketer in the world. Founded in 2010, WHOSAY is built from the best of entertainment, technology and advertising. WHOSAY powers influence marketing campaigns across all verticals and utilizes every level of celebrity and influencer, delivering measurably superior results to other social and mobile advertising.
- Step Functions is a part of the WHOSAY MATCH application for searching and identifying influencers that match with brand campaigns

Analytics data collection problem

WHOSAY

- When a celeb social media post happens…
- Kick off a scheduled sequence of analytics runs

Analytics data collection solution

WHOSAY

They use lambda to run their analytics in a fully stateless manner as a serverless application. They have 1 parallel state with 11 wait states that ends up running the lambda 11 times.

Takeaways

"We can set it and forget it, with no maintenance, and it is easy to support. It was very easy to get going."

"Step Functions and Lambda are a perfect combination for building event-driven and delayed applications, even when tasks need to run longer than 5 minutes."

About Thomson Reuters

- Global organization, HQ in Toronto, 5,000+ employees
- Preparing news video clips for global broadcast and online delivery

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

Video processing problem

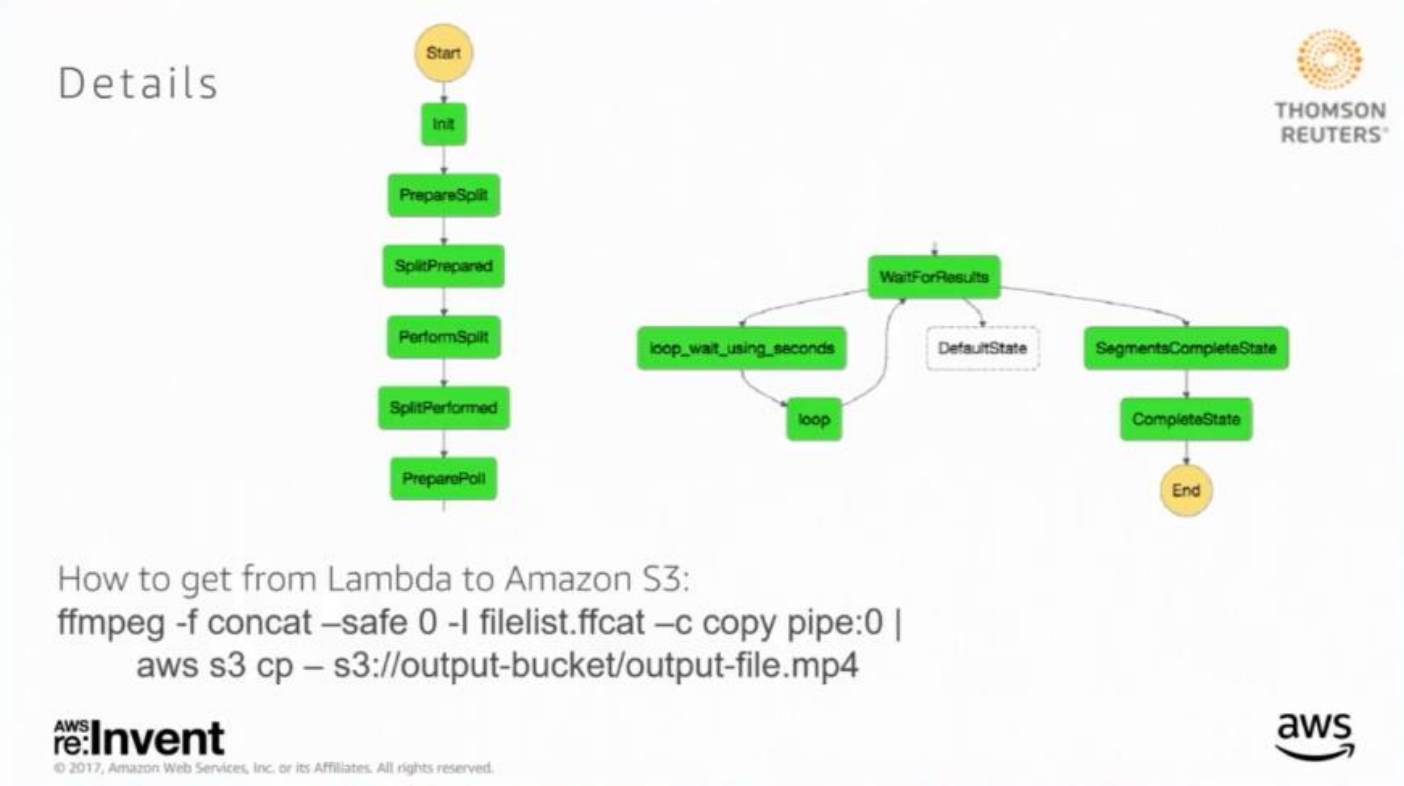
- Transcode 350 clips/day into 14 formats, fast
- It's all done with FFmpeg. The processing time is just about 100% of the video length
- Aargh!

The issue is that FFmpeg is single threaded and can take 30 min to transcode a 30-min video clip.

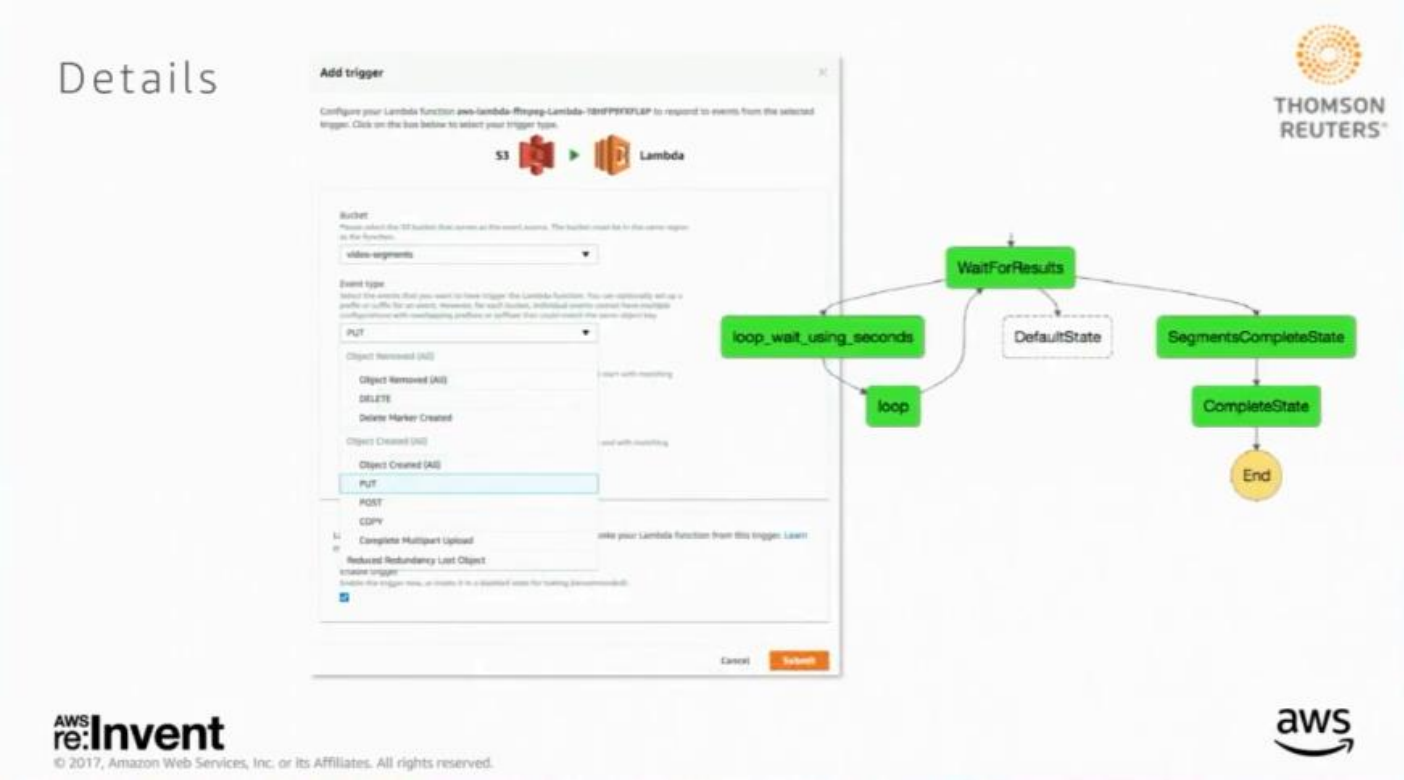
Video processing solution

- Derive keyframe locations within the source
- Split the source at the keyframes
- Process segments (typically 0.5 sec per) in parallel
- Concatenate segments
- Elapsed time: ~20 min down to ~2 minutes

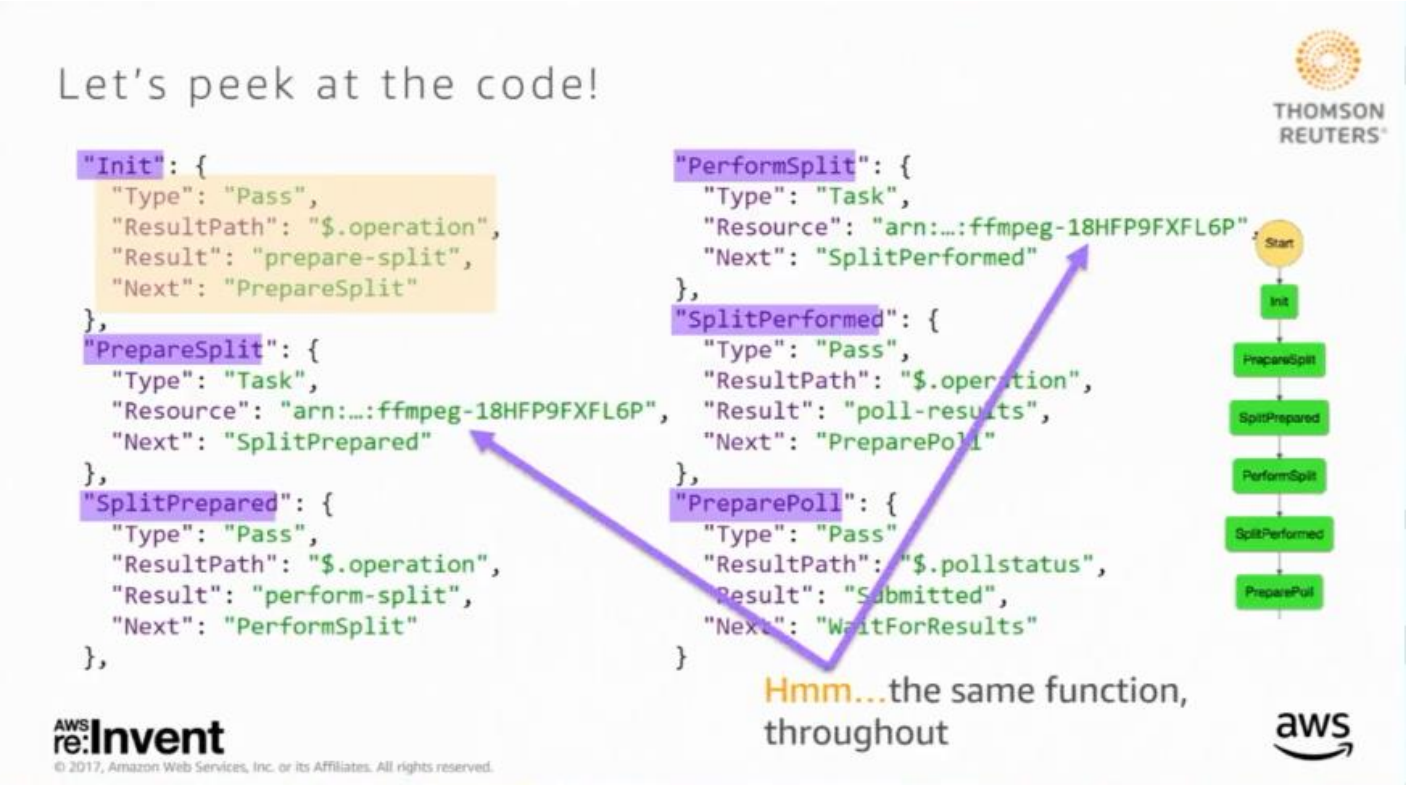
If you look inside a video, there are certain frames that are special called key frames while the other frames are deltas.



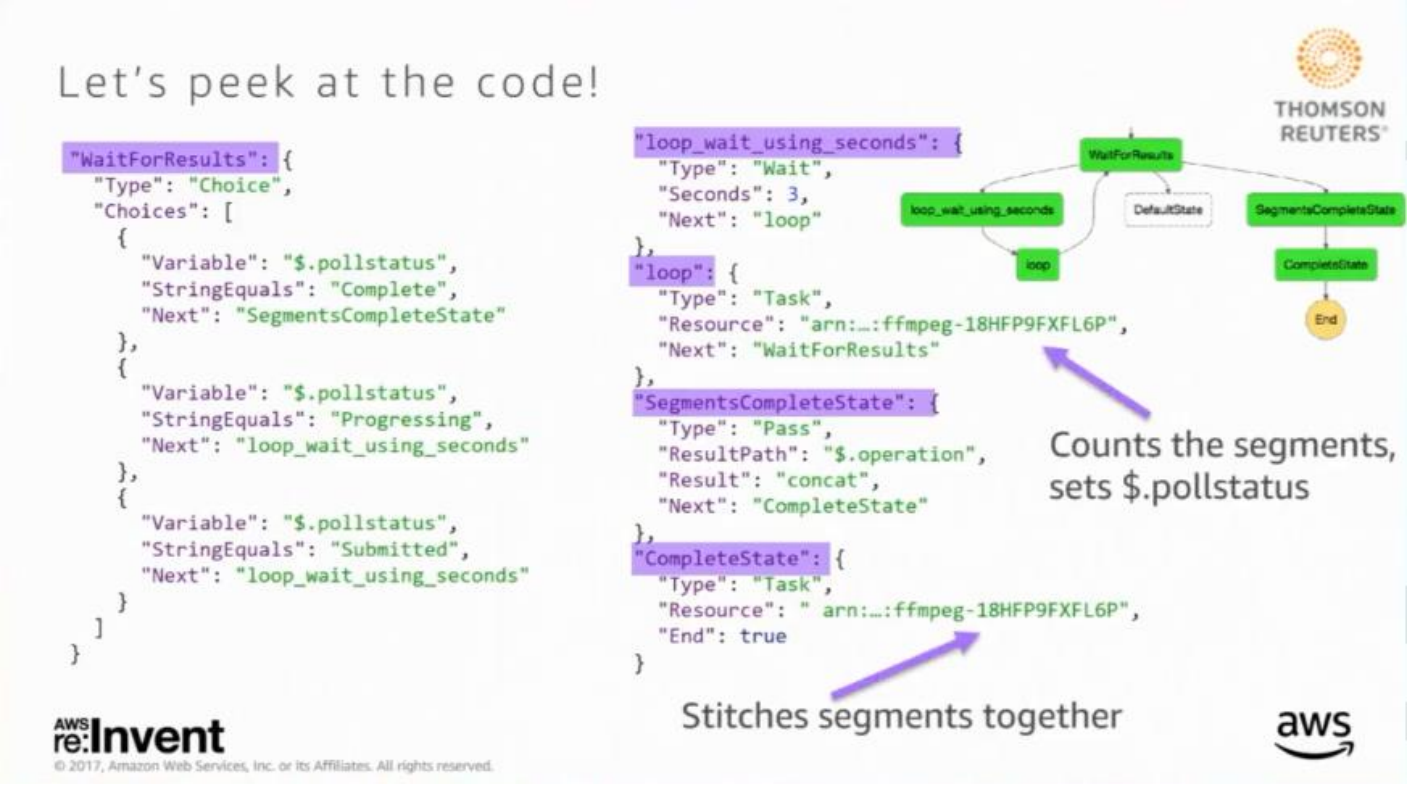
Since FFmpeg wants to write into a file, they told FFmpeg to write it to a file which is actually a named pipe that pipes into an AWS CLI command that dumps it out into an S3 bucket.



They break the video into frames and blast them all into S3 as little object Puts, then they wire up S3 to fire up a lambda for each one. A 12-minute video splitted into 1200 parts is sent to 1200 lambdas that each take like 3 seconds to transcode a 0.5 second video frame. Then they just have to glue back the transcoded pieces into the final processed clips.



Here are the 6 states that help prepare and do the video splits



This is the loop state machine part that does transcoding and then stitch back together

Takeaways

THOMSON REUTERS

Process video segments in parallel: win

Use Amazon S3 URLs to stream video to Lambda: win

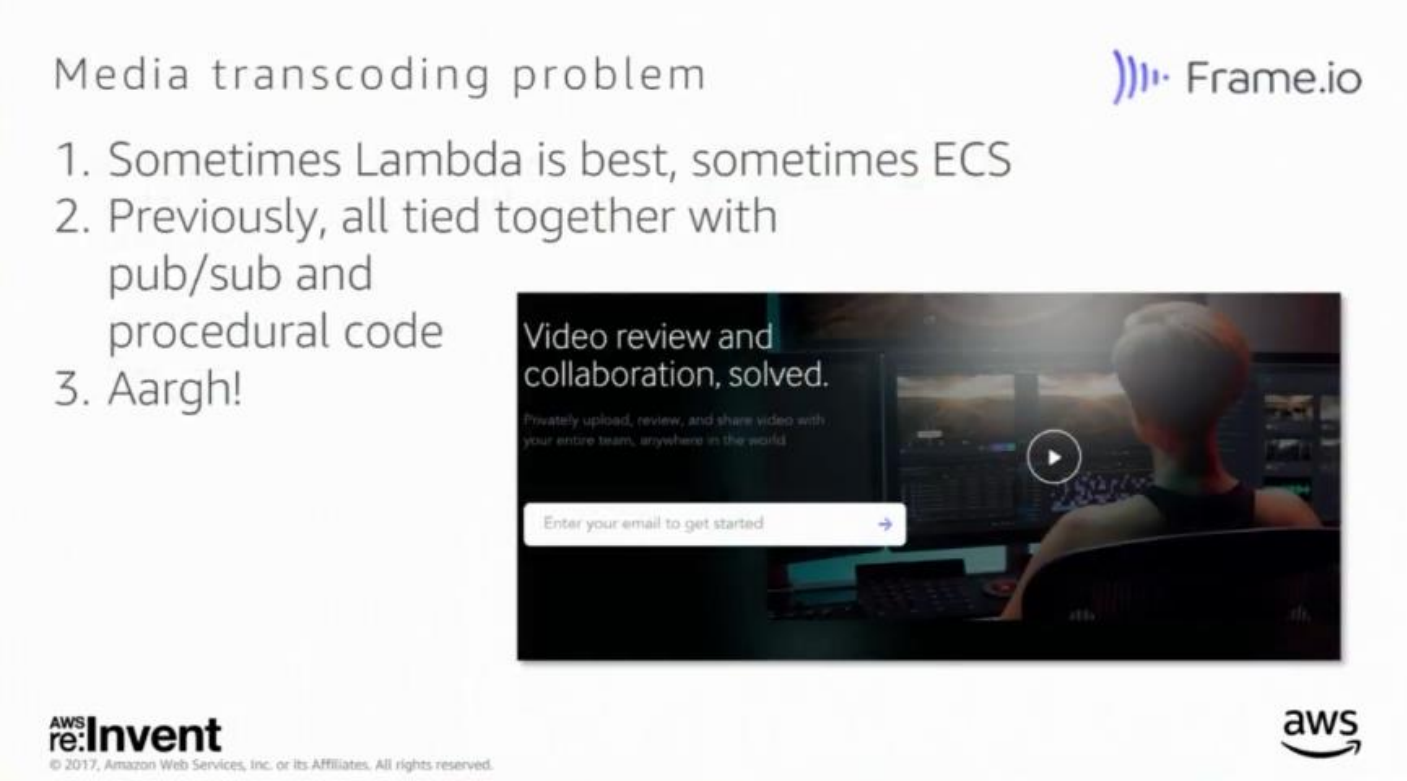
Use Amazon S3 data events to trigger parallel Lambda processing: win

Scaling to 1,000 Lambdas, rather than 1,000 EC2 instances: win

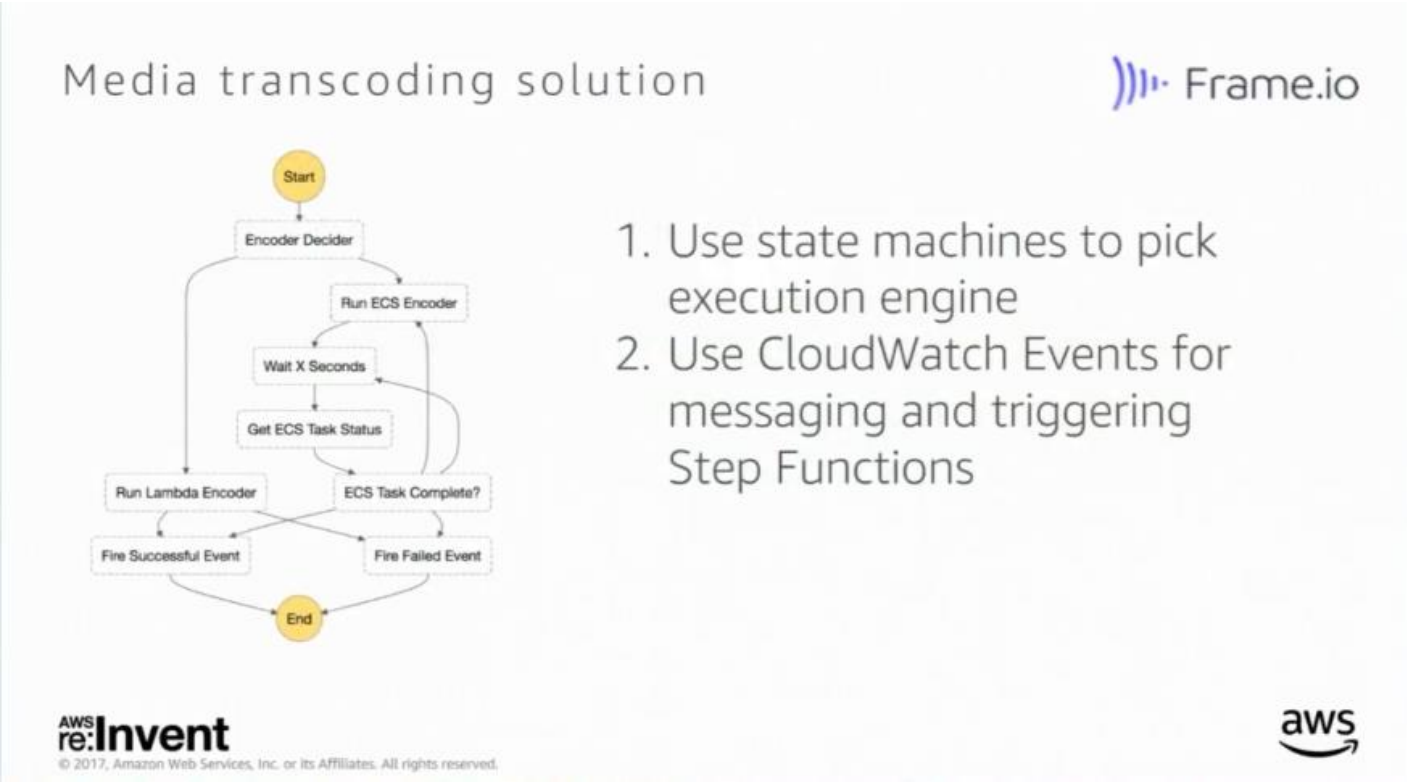
About Frame.io

Frame.io

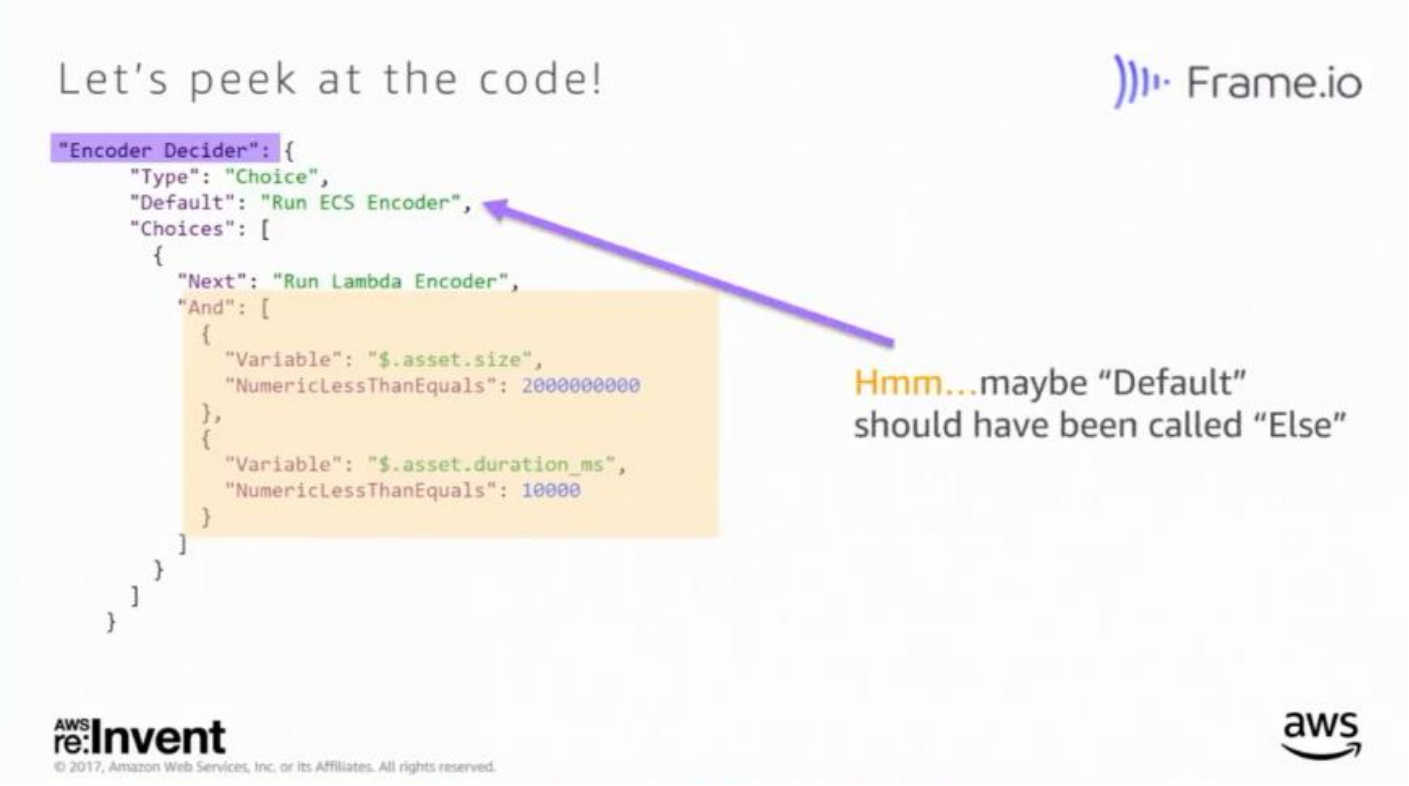
- Frame.io is the world's leading workflow management platform for video teams
- From small production agencies to major broadcast media companies, video teams of all sizes rely on Frame.io to streamline their media review and collaboration process
- Frame.io uses Step Functions to process media, transcode to different formats, create thumbnails, and much more



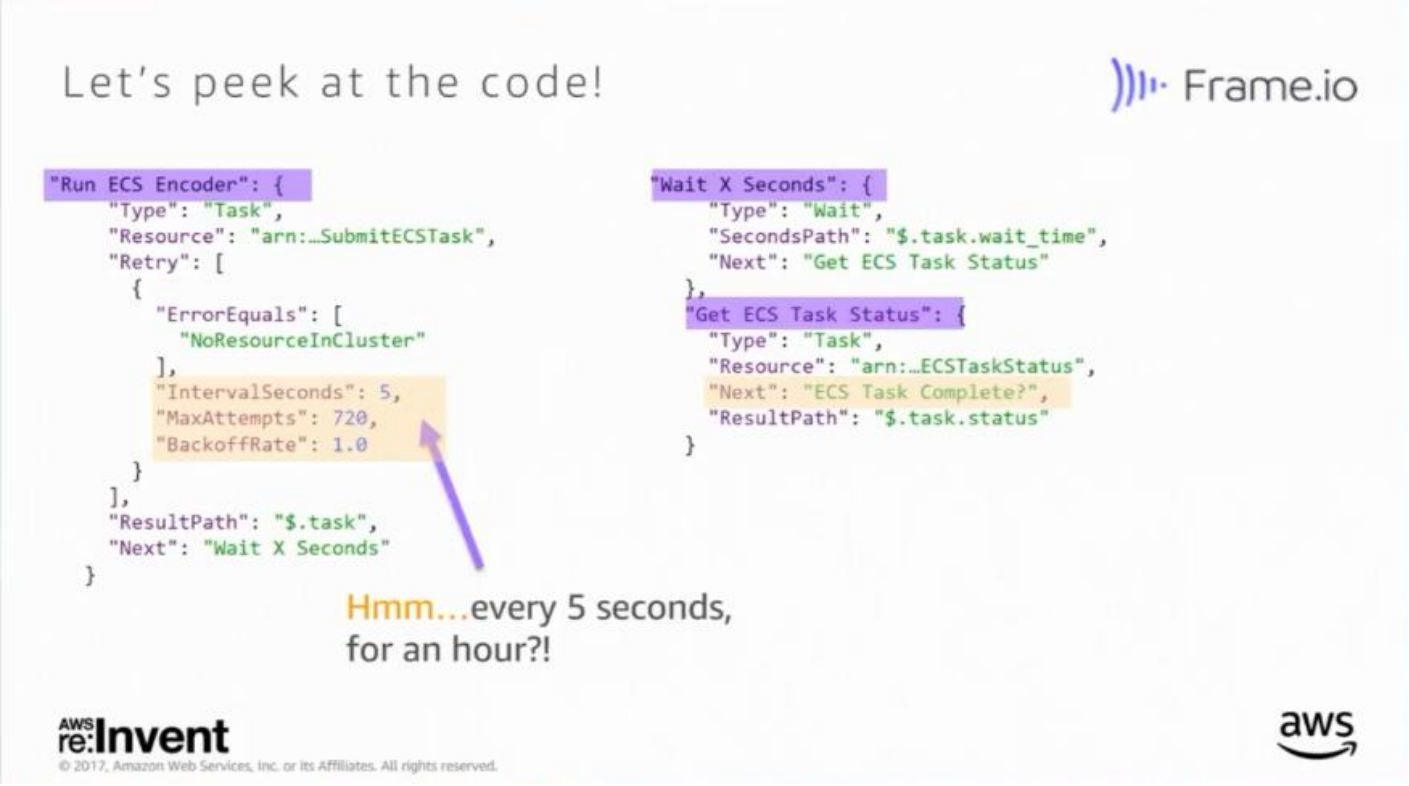
They are using the lambda approach for shorter video processing and using a Docker container when dealing with larger video processing tasks.



They decided to build the state machine above to pick which execution engine thy should use for a video processing task.



This is how they decide to use ECS or Lambda depending on the asset size in GB. The lambda task is a single task state machine that invokes a lambda that does a task.



This is their ECS decoder task path, they are using ***the ECS RunTask API*** to do the interval checking because if your cluster does not have anything available to run your container, it will say it can't run the container when you query the API.

Let's peek at the code!

Frame.io

```
"ECS Task Complete?": {
    "Type": "Choice",
    "Choices": [
      {
        "And": [
          {
            "Variable": "$.task.status",
            "StringEquals": "FAILED"
          },
          {
            "Variable": "$.task.attempt",
            "NumericGreaterThanEquals": 3
          }
        ],
        "Next": "Fire Failed Event"
      },
      {
        "Variable": "$.task.status",
        "StringEquals": "FAILED",
        "Next": "Run ECS Encoder"
      },
      {
        "Variable": "$.task.status",
        "StringEquals": "SUCCEEDED",
        "Next": "Fire Successful Event"
      }
    ],
    "Default": "Wait X Seconds"
  },
  "Fire Successful Event": {
    "Type": "Task",
    "Resource": "aws:…SendSuccessfulEvent",
    "End": true
  },
  "Fire Failed Event": {
    "Type": "Task",
    "Resource": "aws:…SendFailedEvent",
    "End": true
  }
```

These fire CloudWatch Events

Takeaways

Frame.io

Passing a correlation ID through a complex serverless app: win

SAM to make code changes auditable: win

CloudWatch Events pattern matching: win

Code Pipeline for CI/CD: win

More state machines

https://aws.amazon.com/step-functions/getting-started

EBS Snapshot Management

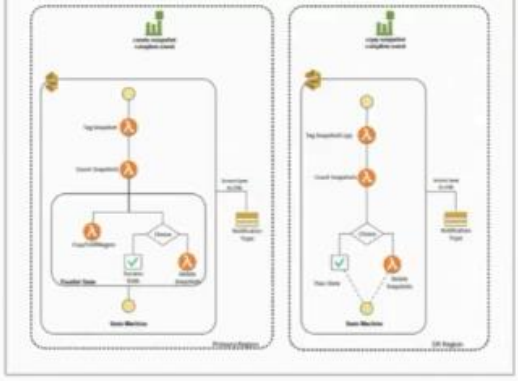
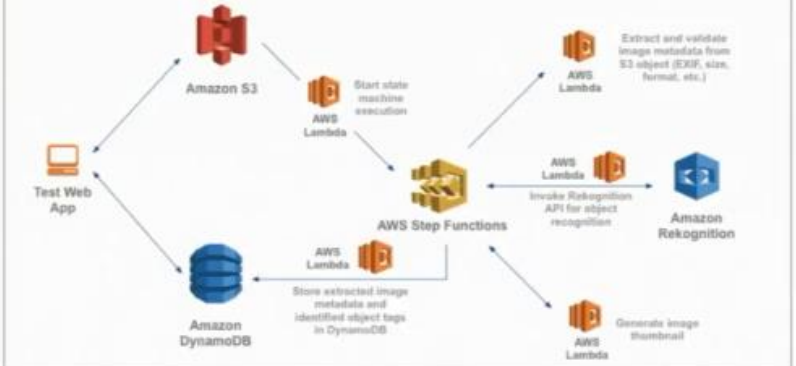


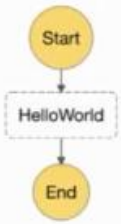
Image Recognition and Processing Backend



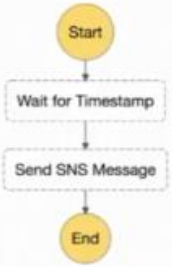
What are you building? #stepfunctions

https://console.aws.amazon.com/states/

Hello world



Task timer



Job status poller