Event-Driven ECS

By Jason Myers

Why event-driven?

Technologies

- Infrastructure as Code (Terraform)
- Docker (ECS)
- Cloud Functions (Lambda, Apex)
- Events (CloudWatch)
- Event Stream (Kinesis)
- DB (DynamoDB)

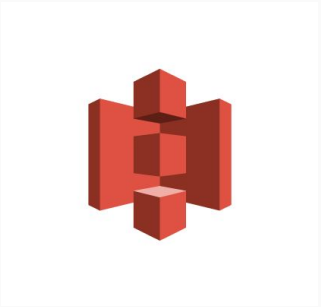

MUCH HYPE!
VERY EXCITE
memegenerator.net

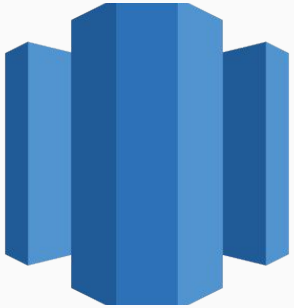
But I prefer...

The Experiment

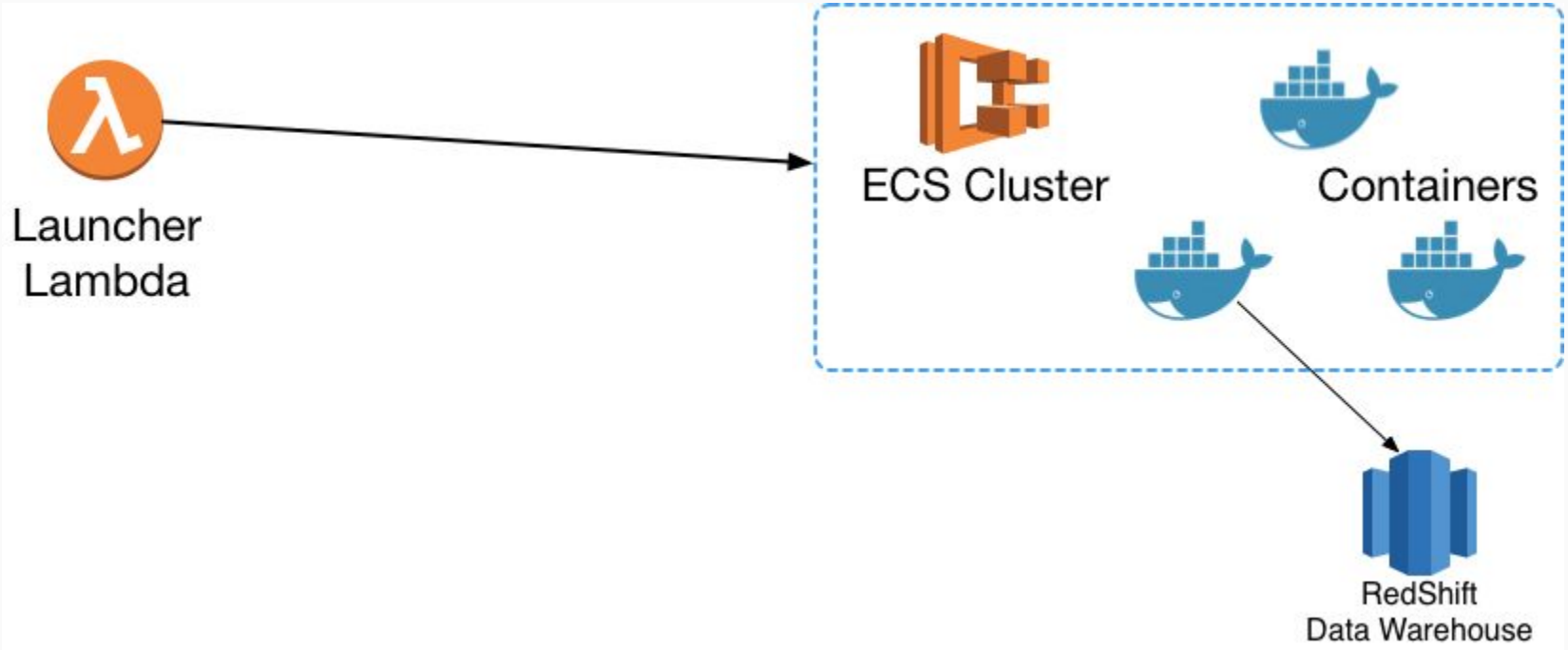
Our Containerized Task

- Get data from CSV files in S3
- Perform some transformations on the data within
- Output the results into Redshift (a columnar datastore)





Solution Overview



Building our infrastructure

Terraform

ECS Terraform Highlights

- Uses an AMI data source to ensure we are on latest AMI.
- user_data
  - Set ECS_CLUSTER
  - Sets ECS_ENABLE_TASK_IAM_ROLE to true
  - Makes sure we have instance metadata and credentials
  - starts logspout
  - maps an EFS volume
- Watches CPU and Memory Reservations to scale ECS up and down as needed

ECS Terraform Highlights - Scheduled Scaling

Don’t forget that you can schedule scaling events!

```
resource "aws_autoscaling_schedule" "ramp_up" {
  scheduled_action_name  = "ramp_up"
  min_size               = 0
  max_size               = 2
  desired_capacity       = 1
  recurrence             = "40 10 * * *"
  autoscaling_group_name = "${module.pipelines_ecs…"
}
```

ECR Terraform

We separate read only and read write roles.

Container Task Definition

Make sure to use data template files!

```
data "template_file" "pipeline_task_template" {
  template =
"${file("task-definitions/pipeline.json.tpl")}"
  vars {
    image_name = "${var.aws_account_id}.dkr.ecr.u..."
    env        = "${var.env}"
  }
}
```

Launching Scheduled Tasks

CloudWatch Scheduled Event (CRON)

```
resource "aws_cloudwatch_event_rule" "pipeline-daily" {
    is_enabled = "false"
    name = "pipeline-daily"
    description = "pipeline ETL daily load"
    schedule_expression = "cron(30 11 * * ? *)"
}
```

Apex

- Multiple Environment Support\
- Function Deployment
- Infrastructure as code via Terraform
- Lots of Language support!
  - Java
  - Python 2.7 and 3.6
  - Nodejs 4.3 and 6.10
  - Golang (any version)
  - Clojure (any version)
  - Rust (any version)

Invocation Function

Just runs the ECS task overriding it command.

Starts anywhere from 1 to 11 tasks in our world.

CloudWatch Lambda Permissions

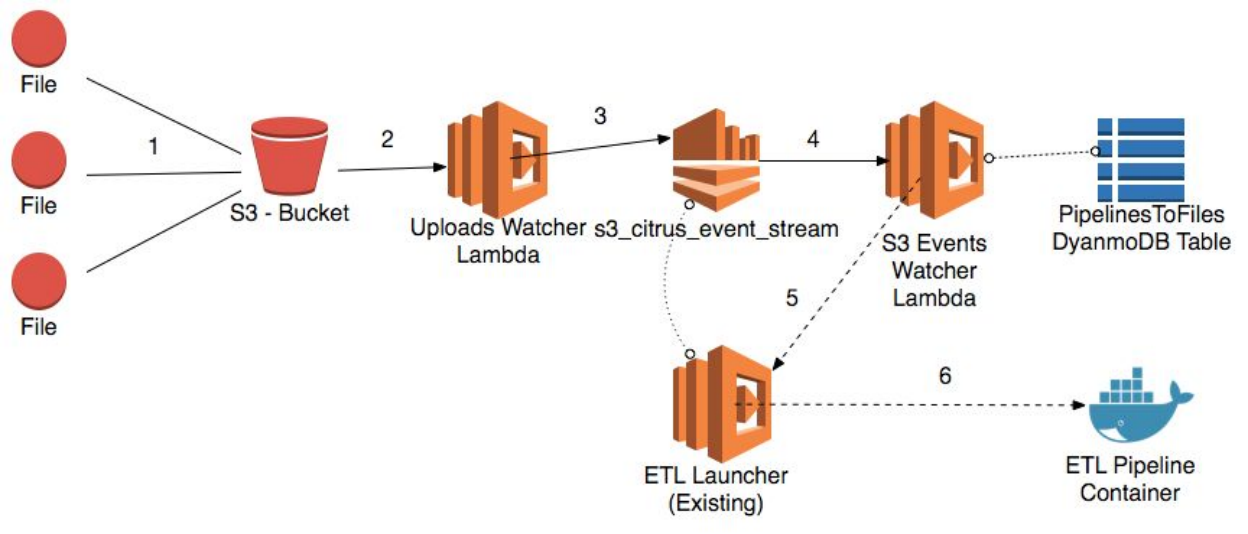
```
resource "aws_lambda_permission" "allow_cloudwatch" {
  statement_id   = "AllowExecutionFromCloudWatch"
  action         = "lambda:InvokeFunction"
  function_name  = "TriggerPipelineLambda"
  principal      = "events.amazonaws.com"
  source_account = "${var.aws_account_id}"
  source_arn     =
"${aws_cloudwatch_event_rule.pipeline-daily.arn}"
}
```

Lambda Role Terraform

Apex will create a role for the lambda function make sure to give it permissions to invoke ECS tasks.

Event Driven Invocation - S3 Uploads

Event-Driven Overview



Uploads Watcher Lambda

Receives each S3 event via (aws_s3_bucket_notification)

Decides if it’s important

Adds the event into a Kinesis stream

Kinesis Event Handler (S3 Events Watcher)

Triggered by an aws_lambda_event_source_mapping from Kinesis

Gets the full s3 event grabs out the bucket and the key

Looks up the client, key, bucket and gets the pipeline lambda function that should be run

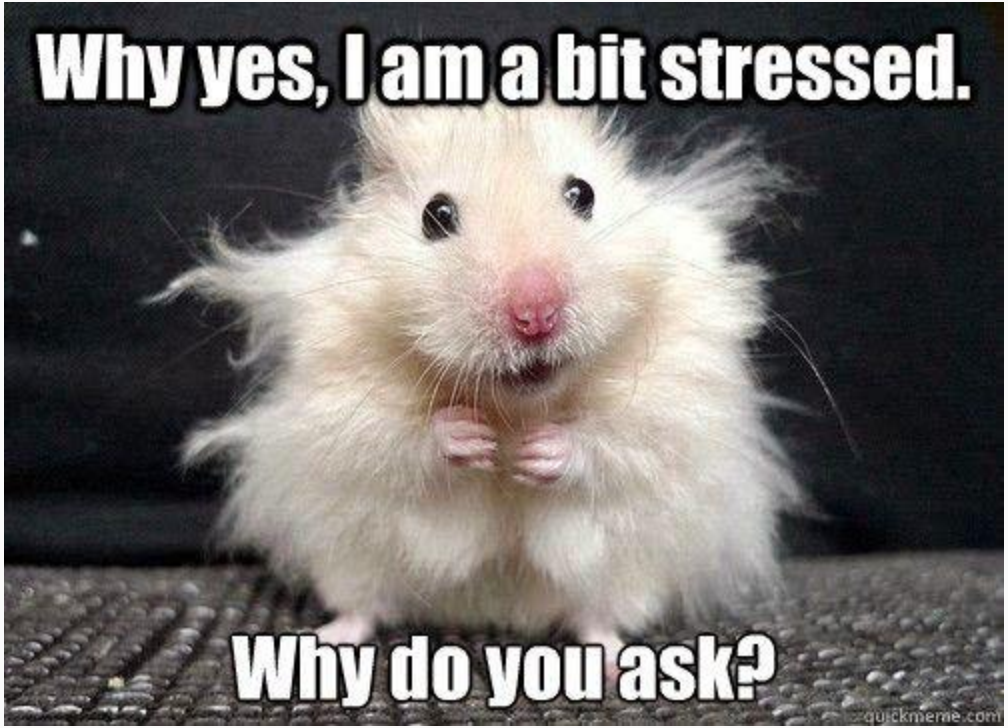
Invokes the pipeline lambda function

Updated Invocation Handler

No changes needed it if only takes 1 file, but it takes multiple…

Scans last hour or so in kinesis to see if all the needed files were uploaded

If so it triggers the job it not it exists


Why yes, I am a bit stressed.
Why do you ask?
quickmeme.com

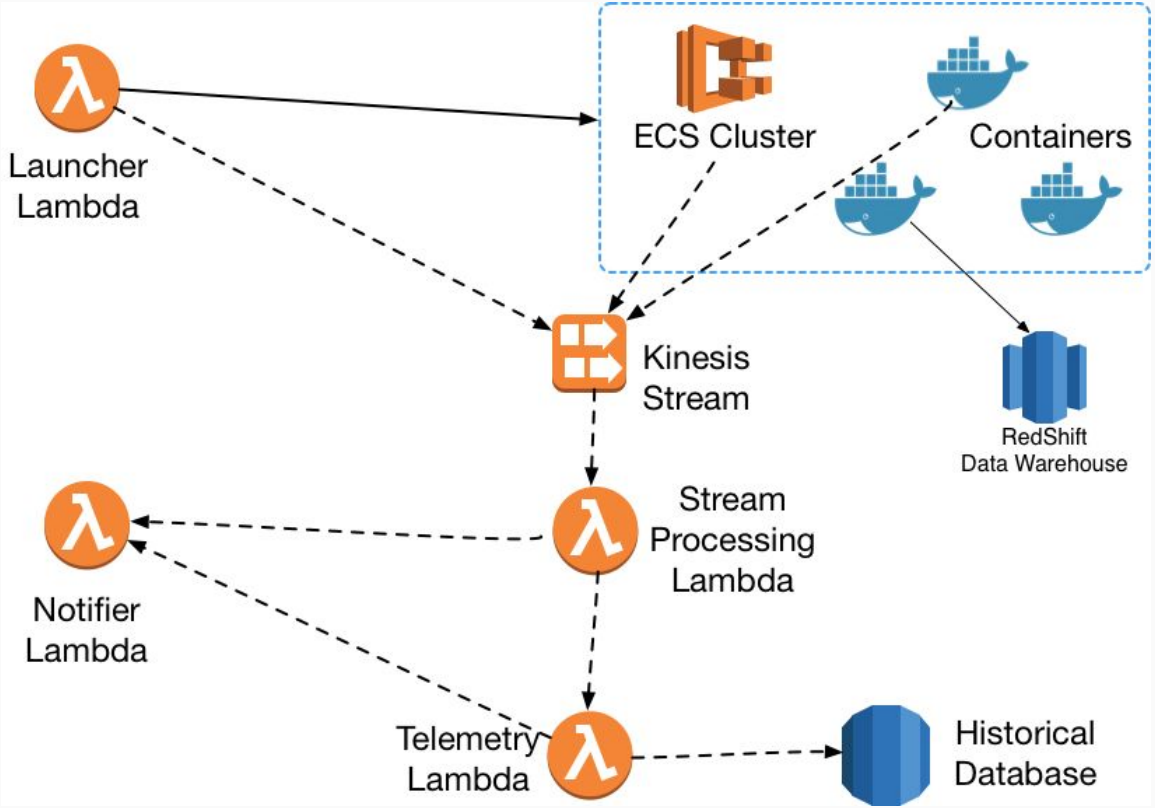
Updated Invocation Handler

Events will get triggered at least once…

Scanning events in Kinesis takes time and the the multiple invocations of the Pipeline Lambda function could try to trigger the container task multiple times

Observability and Autonomous Operation

Observability Overview



Notifications

Telemetry (💗)

Track runtimes by pipeline tasks and subtasks within the container

Track error rates

Track records affected

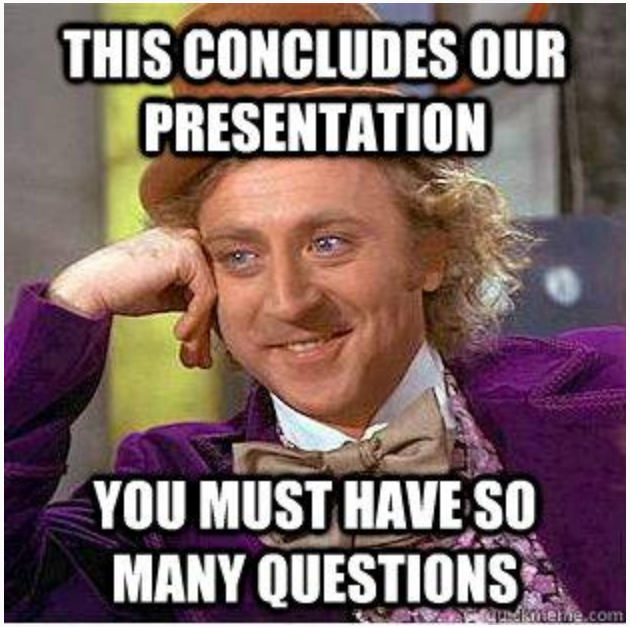
Anomaly Notifications

Ran too long

Updated too many Records

Detailed information about Errors

Autonomous Operation



http://bit.ly/jam-event-ecs - @jasonamyers