



# VISUALIZING

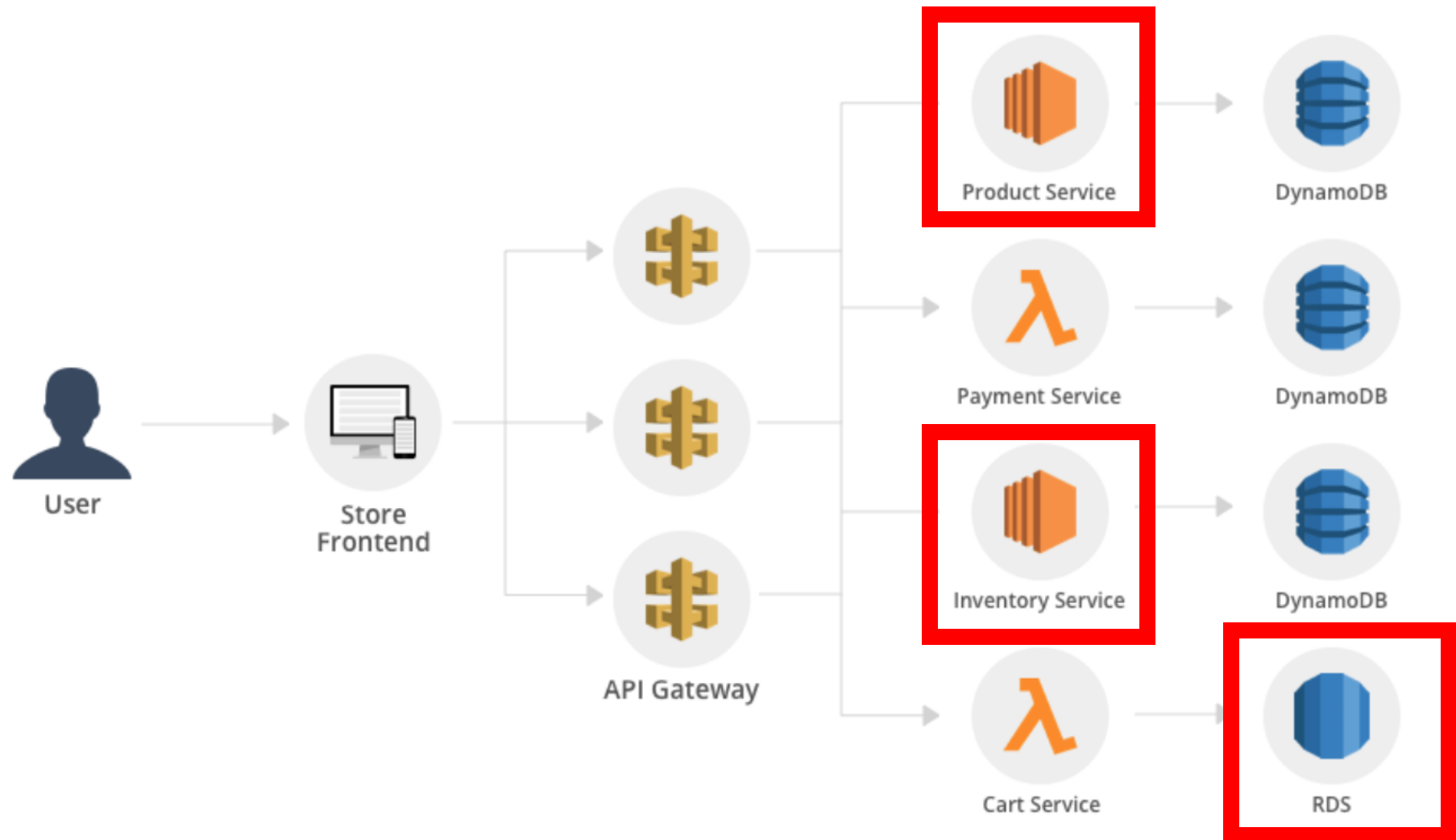
Cloud Systems

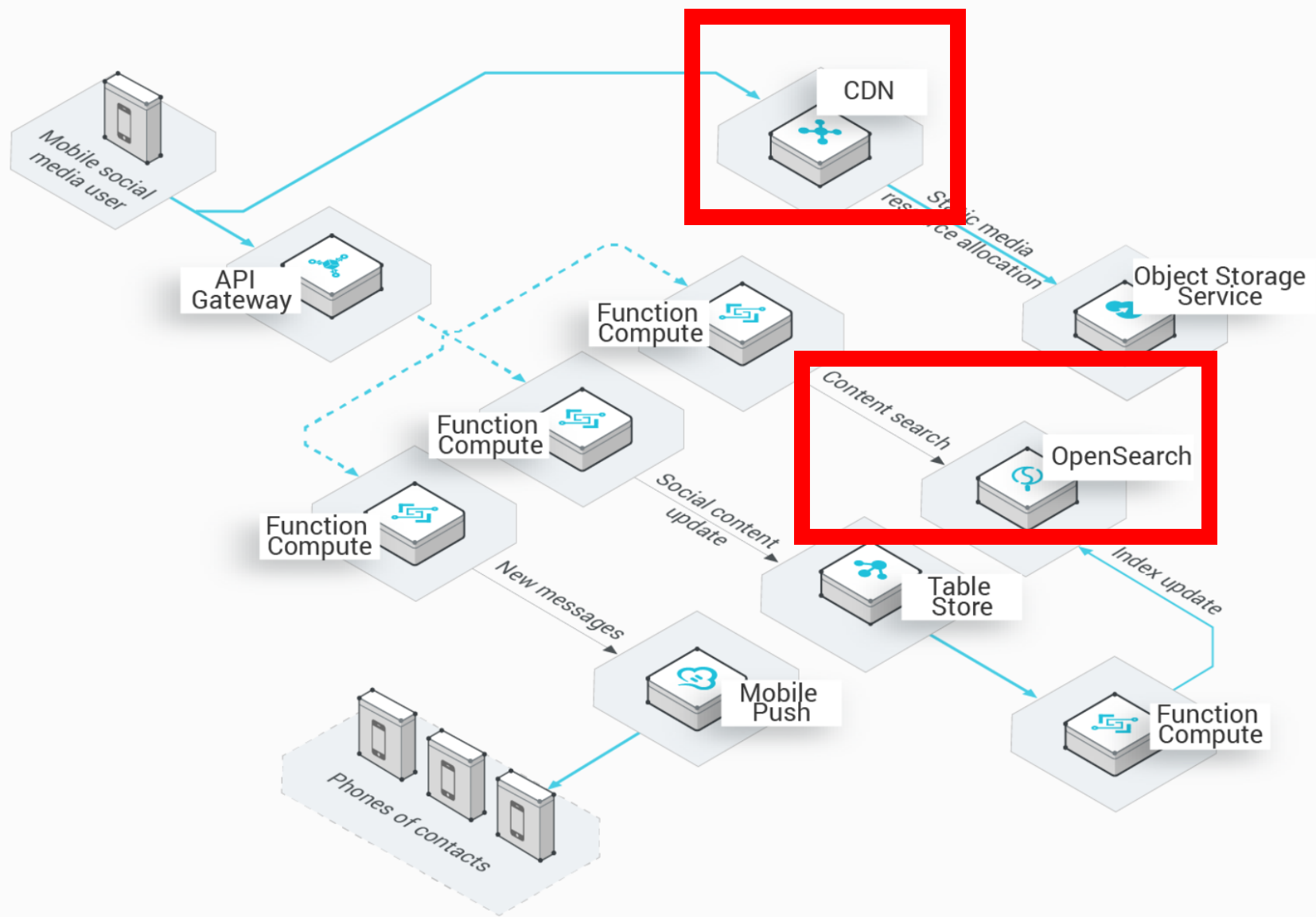
Lynn Langit

WHY DO SYSTEM  
PICTURES  
MATTER?


# CLOUD INFRASTRUCTURE

TYPE	METHOD	INFO
WebUI	click to create	AWS, GCP, Alibaba Cloud
Script	bash-like Pythonic	awscli, gcloud, aliyun
Deployment	Template YAML/JSON visual tool	AWS CloudFormation, GCP Deployment FUN template
Infrastructure	language visual tool	Terraform
Vertical bioinformatics	language	WDL, CWL Nextflow
Vertical WebUI	visual tool	Terra.bio on GCP Galaxy Project on AWS





## Compute

EC2  
Lightsail   
ECR  
ECS  
EKS  
Lambda  
Batch  
Elastic Beanstalk  
Serverless Application Repository

## Storage

S3  
EFS  
FSx  
S3 Glacier  
Storage Gateway  
AWS Backup

## Database

RDS  
DynamoDB  
ElastiCache  
Neptune  
Amazon Redshift  
Amazon QLDB  
Amazon DocumentDB

## Migration & Transfer

AWS Migration Hub  
Application Discovery Service  
Database Migration Service  
Server Migration Service  
AWS Transfer for SFTP  
Snowball  
DataSync

## Networking & Content Delivery

VPC

## Developer Tools

CodeStar  
CodeCommit  
CodeBuild  
CodeDeploy  
CodePipeline  
Cloud9  
X-Ray

## Robotics

AWS RoboMaker


## Blockchain

Amazon Managed Blockchain

## Satellite

Ground Station

## Management & Governance

AWS Organizations  
CloudWatch  
AWS Auto Scaling  
CloudFormation  
CloudTrail  
Config  
OpsWorks  
Service Catalog  
Systems Manager  
Trusted Advisor  
Managed Services  
Control Tower  
AWS License Manager  
AWS Well-Architected Tool  
Personal Health Dashboard   
AWS Chatbot


## Media Services

Elastic Transcoder


## Machine Learning

Amazon SageMaker  
Amazon Comprehend  
AWS DeepLens  
Amazon Lex  
Machine Learning  
Amazon Polly  
Rekognition  
Amazon Transcribe  
Amazon Translate  
Amazon Personalize  
Amazon Forecast  
Amazon Textract  
AWS DeepRacer

## Analytics

Athena  
EMR  
CloudSearch  
Elasticsearch Service  
Kinesis  
QuickSight   
Data Pipeline  
AWS Glue  
AWS Lake Formation  
MSK

## Security, Identity, & Compliance

IAM  
Resource Access Manager  
Cognito  
Secrets Manager  
GuardDuty  
Inspector  
Amazon Macie   
AWS Single Sign-On  
Certificate Manager  
Key Management Service  
CloudHSM  
Directory Service  
WAF & Shield

## Mobile

AWS Amplify  
Mobile Hub  
AWS AppSync  
Device Farm

## AR & VR

Amazon Sumerian


## Application Integration

Step Functions  
Amazon EventBridge  
Amazon MQ  
Simple Notification Service  
Simple Queue Service  
SWF

## Customer Engagement

Amazon Connect  
Pinpoint  
Simple Email Service

## Business Applications

Alexa for Business  
Amazon Chime   
WorkMail

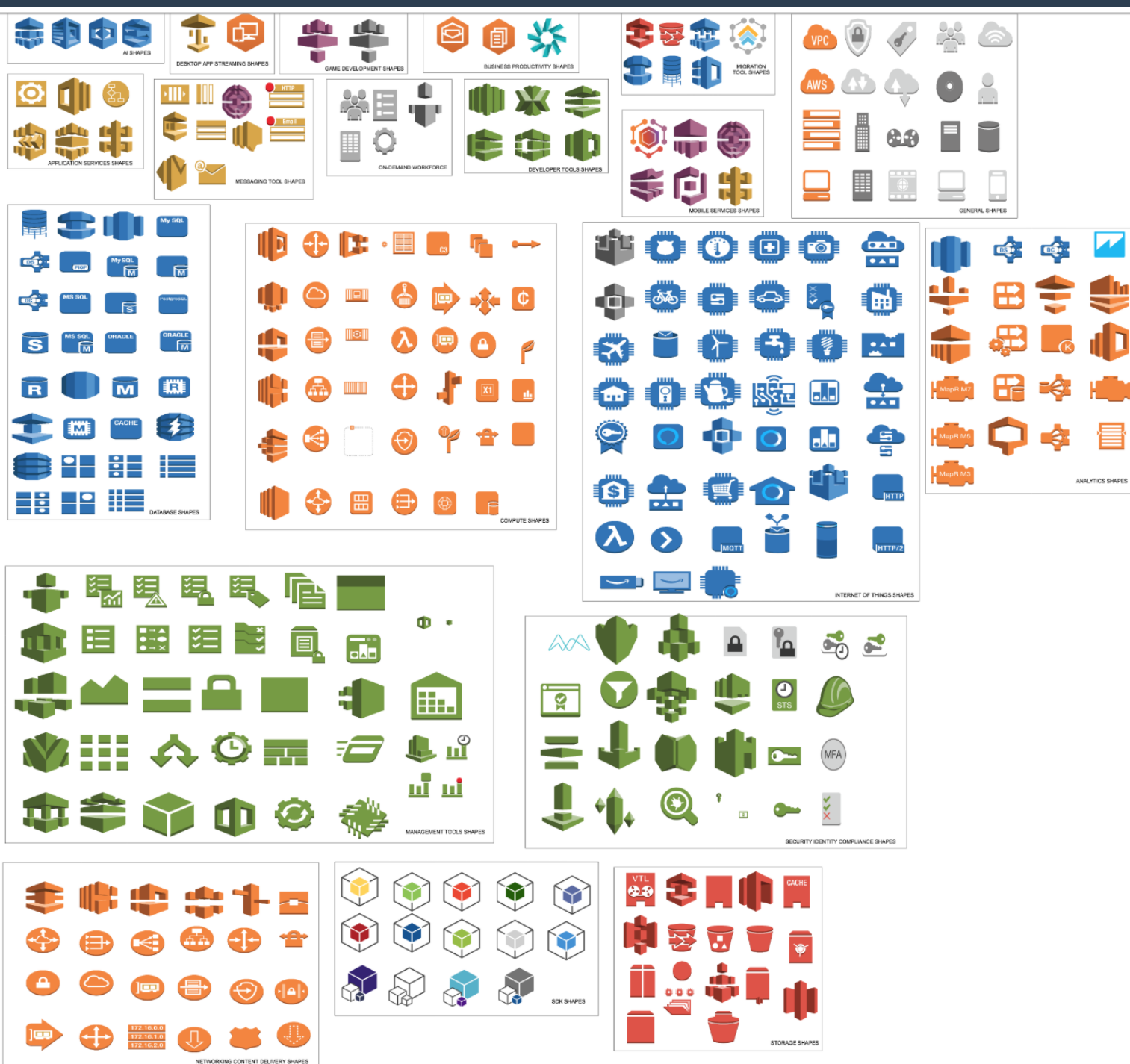
## End User Computing

WorkSpaces  
AppStream 2.0  
WorkDocs  
WorkLink

## Internet of Things

IoT Core  
Amazon FreeRTOS  
IoT 1-Click

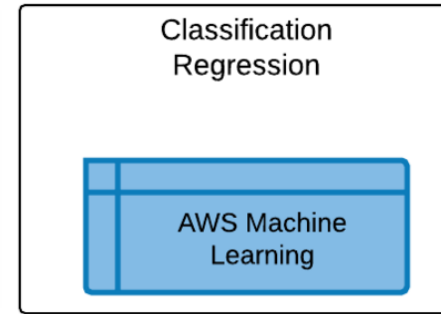
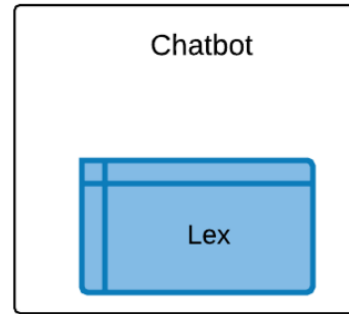
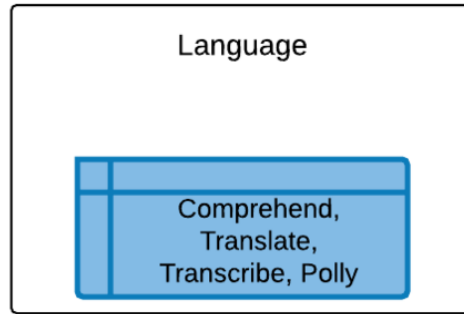
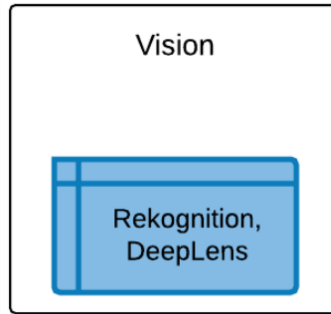




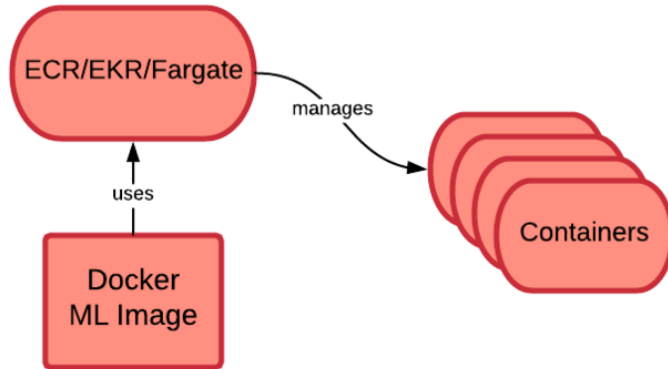
# BIAS

do icons matter?

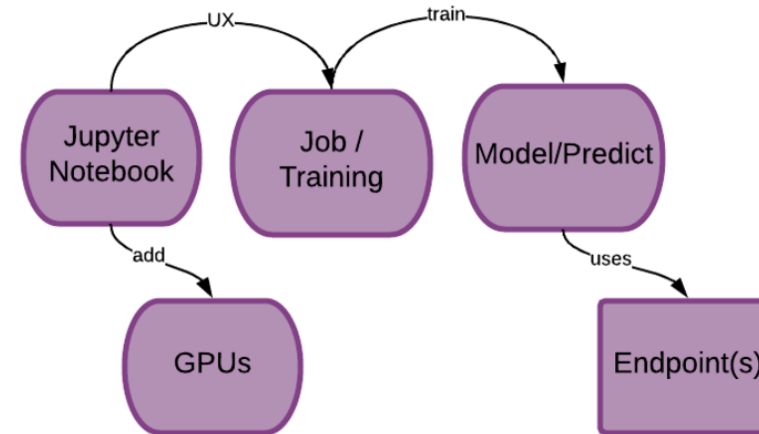
SaaS - APIs



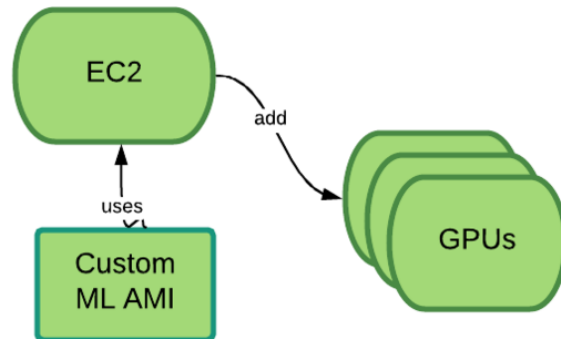
PaaS - Containers



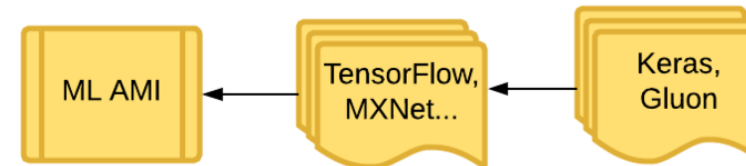
Managed Containers - SageMaker



IaaS - EC2



Supplied AMI



SaaS - APIs

Vision

Vision, Video

Language

Natural Language, Translate,  
Speech-to-Text, Text-to-Speech, Talent

Chatbot

Dialogflow

Classification  
Regression

BigQuery ML

AutoML for Vision, Natural Language, Translation - supply labels and labeled data.

PaaS - Containers

Kubernetes

uses

Docker  
ML Image

manages

Containers

Managed Containers -

ML Engine

add

GPUs

UX

train

Job /  
Training

Model/Predict

uses

Endpoint(s)

Colabs

IaaS - Virtual Machines

GCE

uses

Custom  
ML AMI

add

add

GPUs

TPUs

Supplied AMI

ML AMI

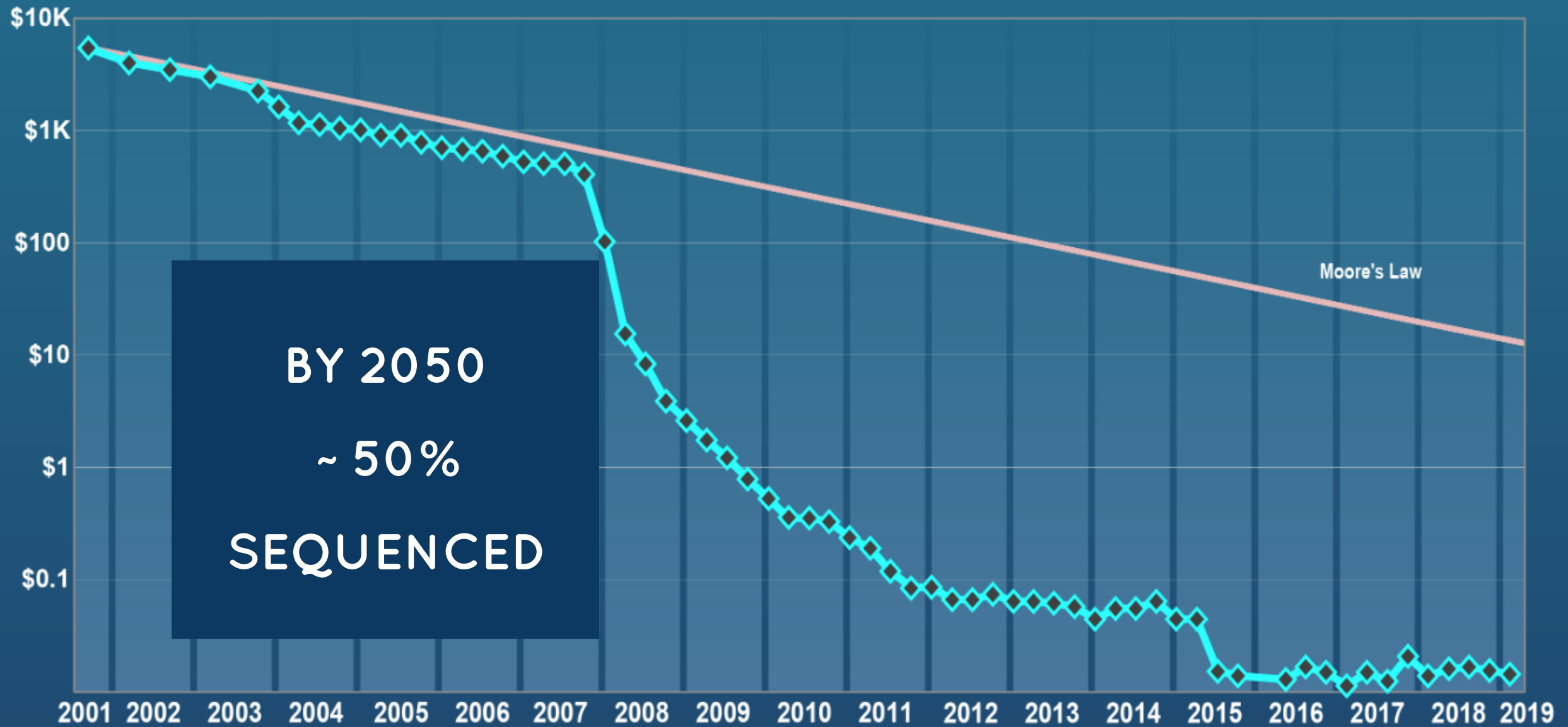
TensorFlow,  
Keras

Datalab

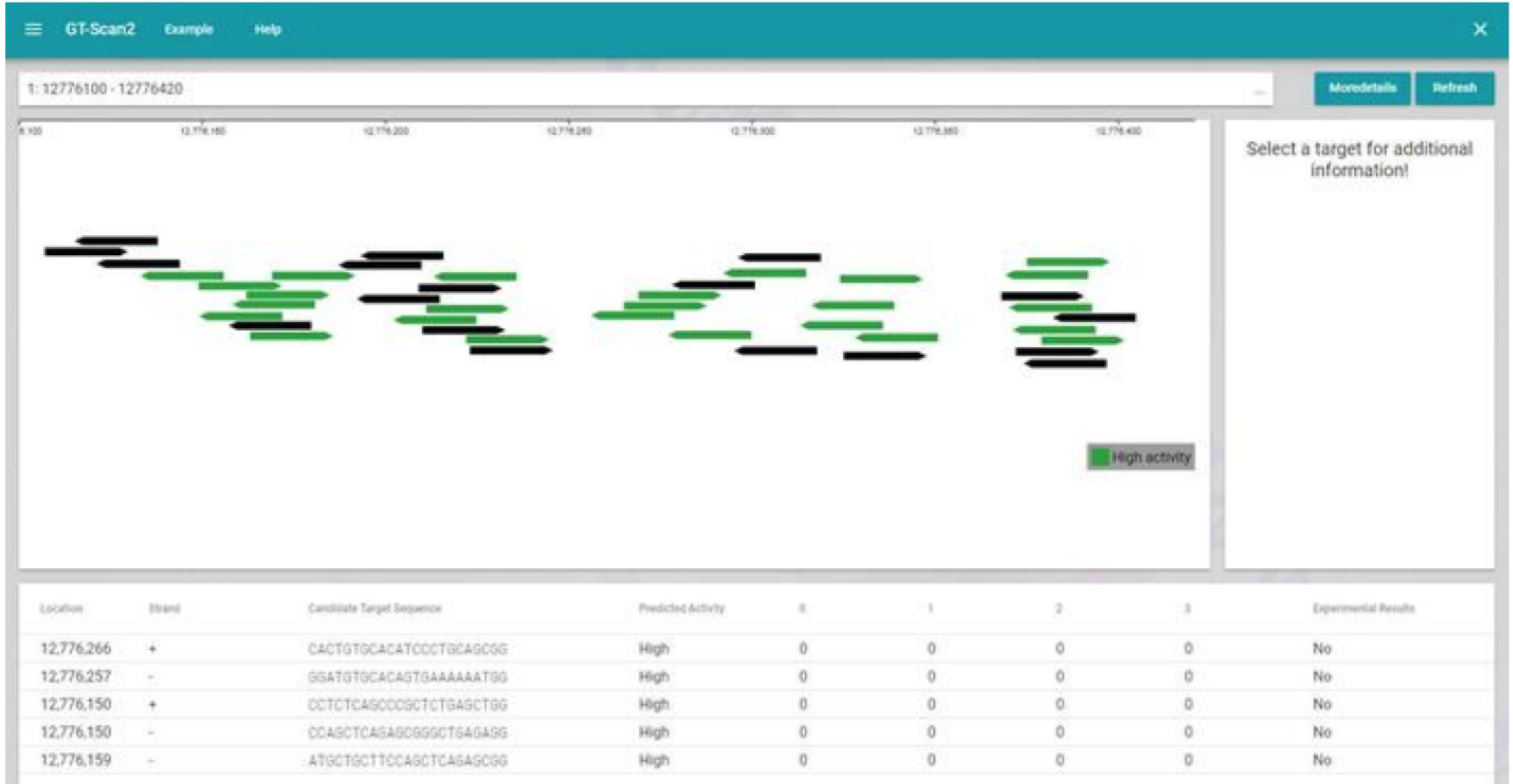
To learn more go to -- [bit.ly/learnfromlynn](https://bit.ly/learnfromlynn)

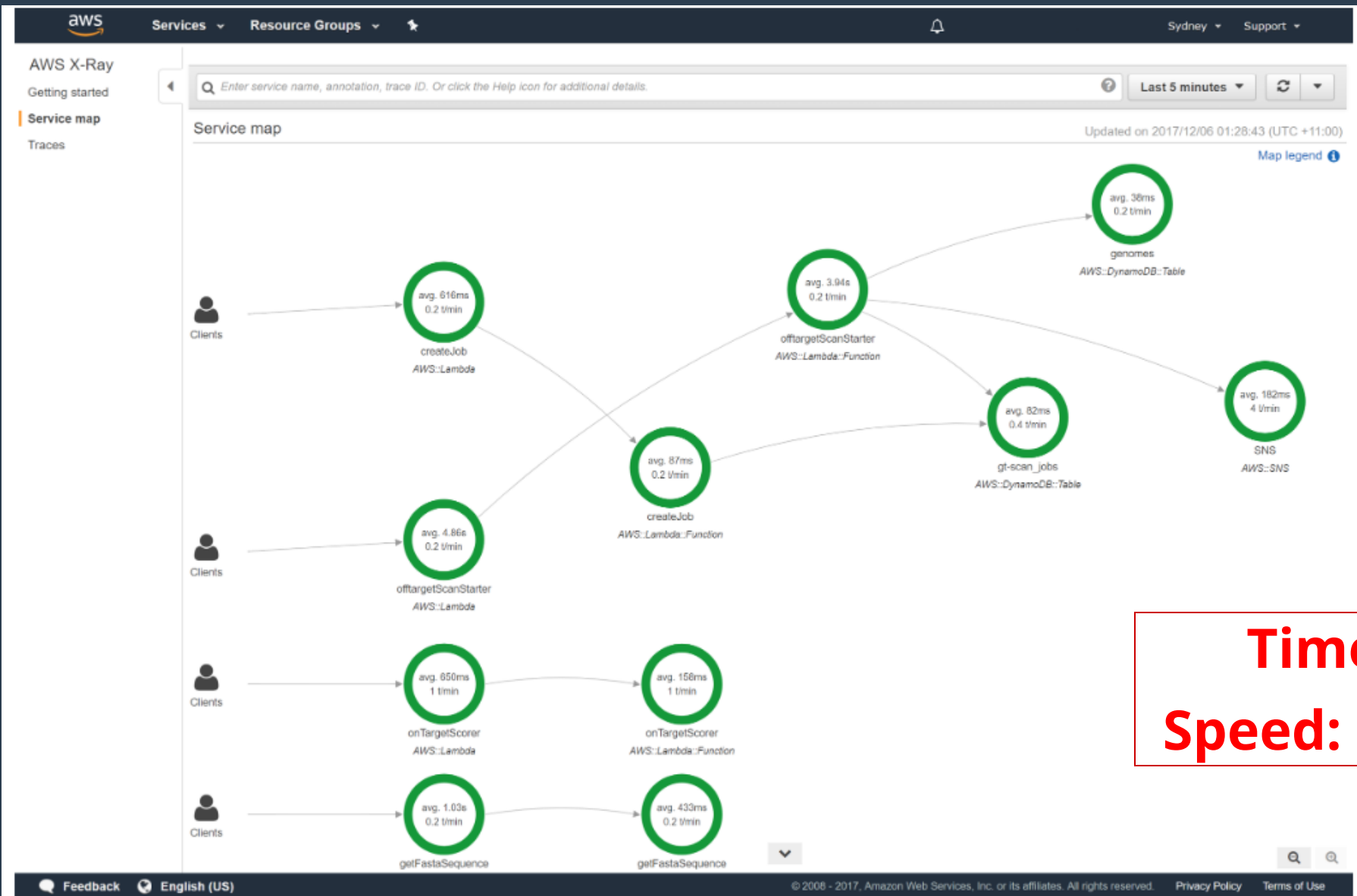


# Cost per Raw Megabase of DNA Sequence



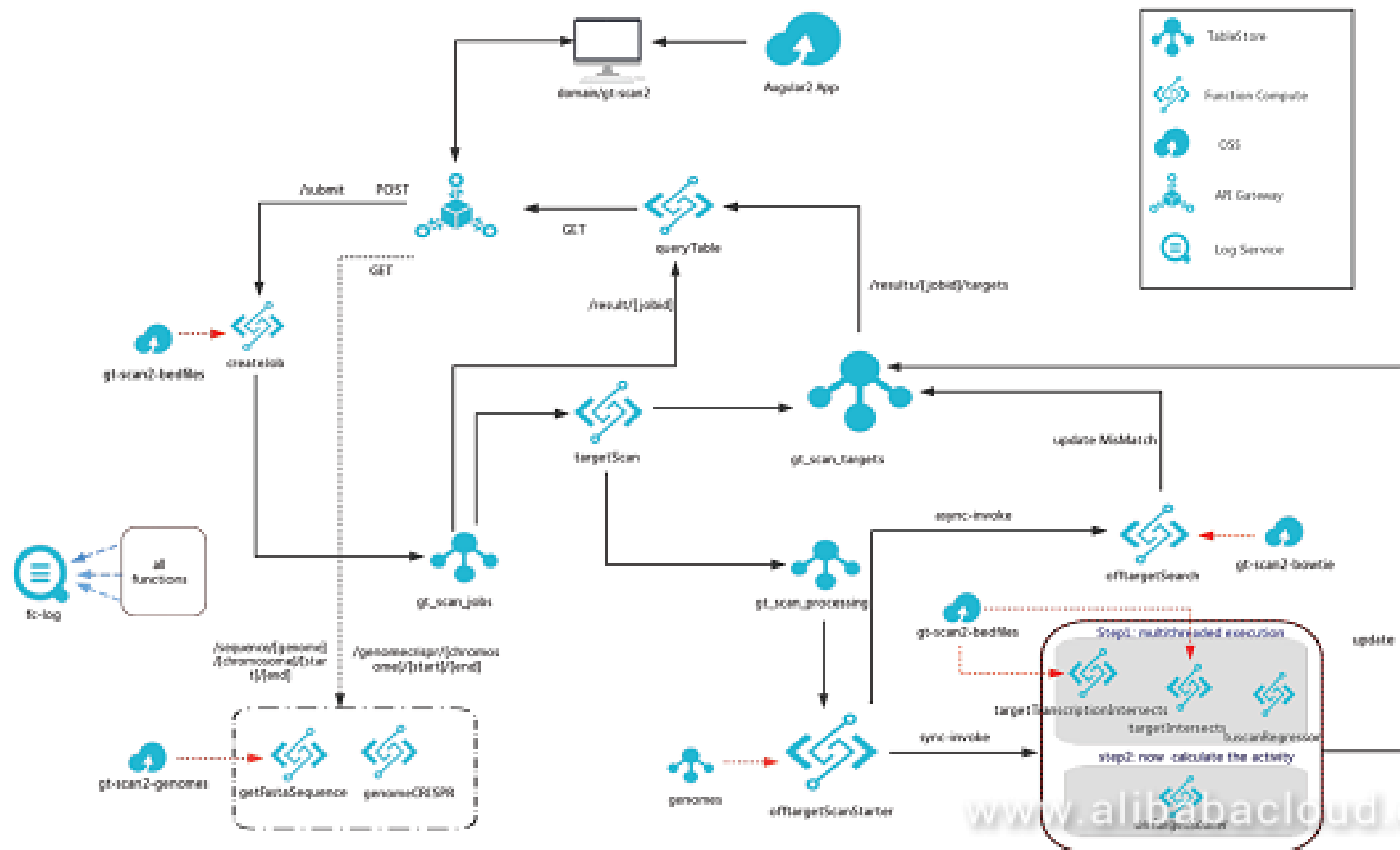
# TOOLS FOR RESEARCH - GT-SCAN2





**Time: 1 Day**  
**Speed: 80% faster**

Screenshot of the X-Ray dashboard listing the average runtimes.



# NEXT

**can we solve a more difficult challenge?**



Search or jump to...



Pull requests

Issues

Marketplace

Explore



aeherc / VariantSpark

Watch

10

Unstar

37

Fork

10

Code

Issues 25

Pull requests 0

Projects 2

Wiki

Insights

Branch: master










Create new file

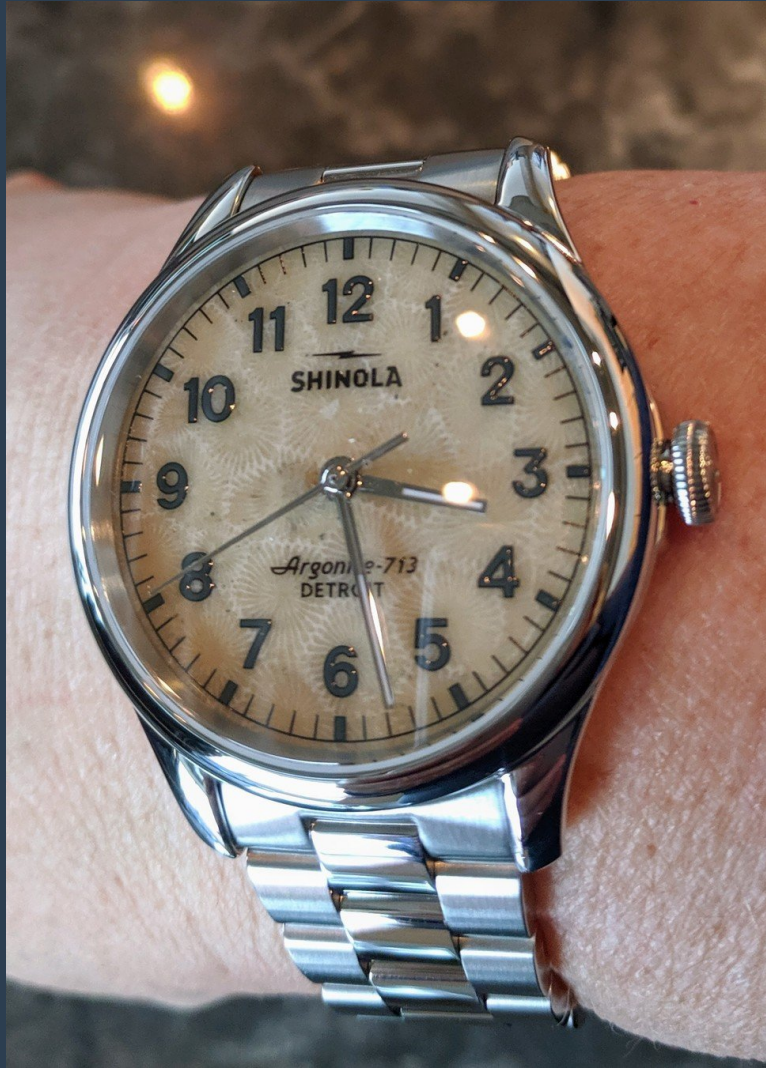
Upload files

Find file

History

VariantSpark / src / main / scala / au / csiro / variantspark / algo /

 piotr szul	[#30] Fix: Make results reproducibe (#86)	Latest commit fb95fae on Jun 26
..		
 metrics	[#43] Add API for Pairwise Operations (#50)	11 months ago
 CanSplitTypes.scala	code cleaning and fix typos in variable names (#3)	a year ago
 DecisionTree.scala	[#30] Fix: Make results reproducibe (#86)	4 months ago
 PairwiseOperation.scala	[#48] Add API for Pairwise Operations (#50)	11 months ago
 PredictiveModel.scala	code cleaning and fix typos in variable names (#3)	a year ago
 RandomForest.scala	[#29] Make -rre (randomize equal) option enabled by default (#36)	a year ago
 WideKMeans.scala	adding ScalaDocs to ML files	a year ago
 package.scala	Switch importance to work on bytes	2 years ago



TEN  
MINUTES

# BIAS

**solving difficult problems**



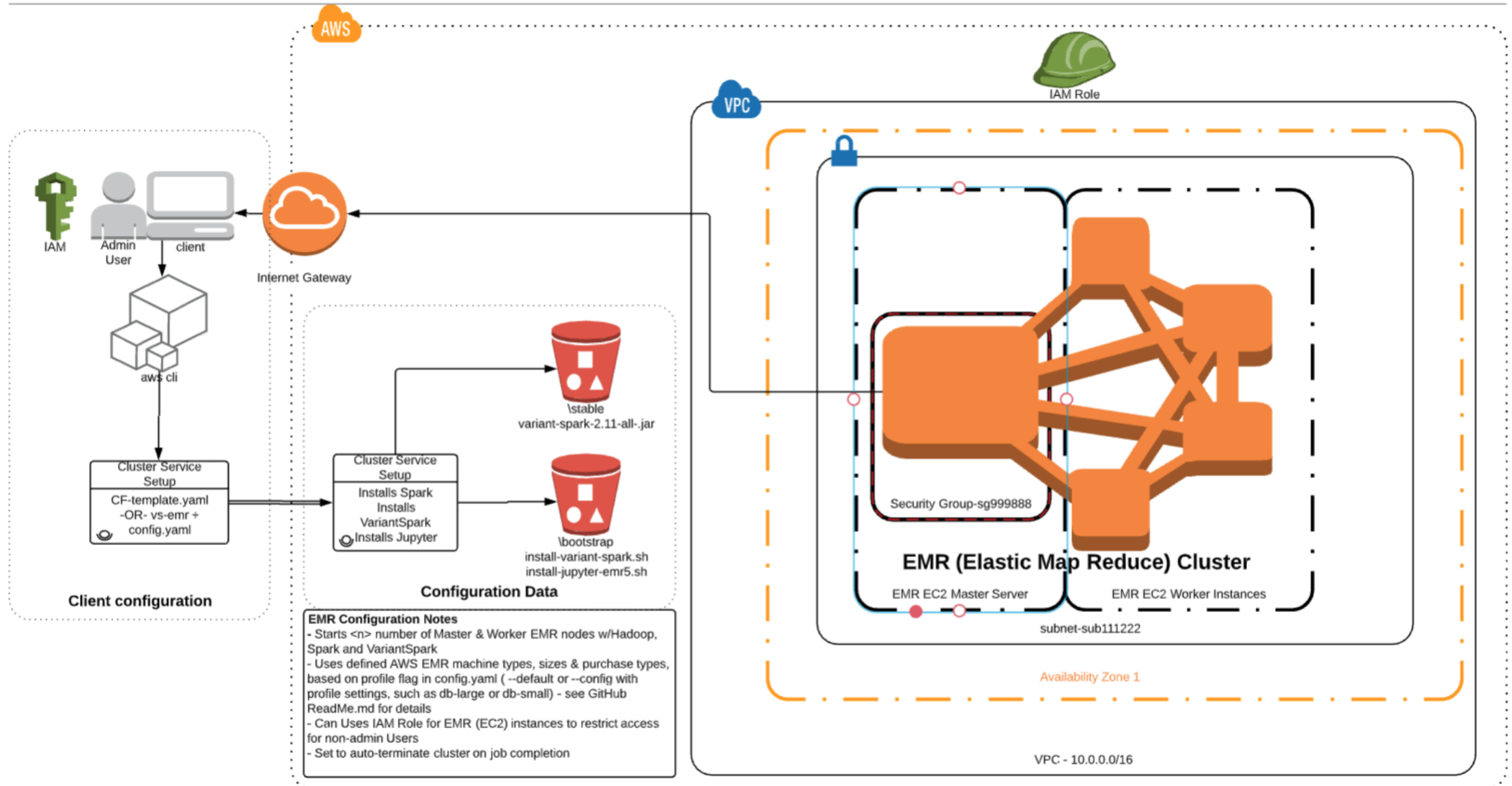
Lynn Langit  
Cloud/Big Data Architect  
Oct 1 · 8 min read

# Scaling Custom Machine Learning on AWS

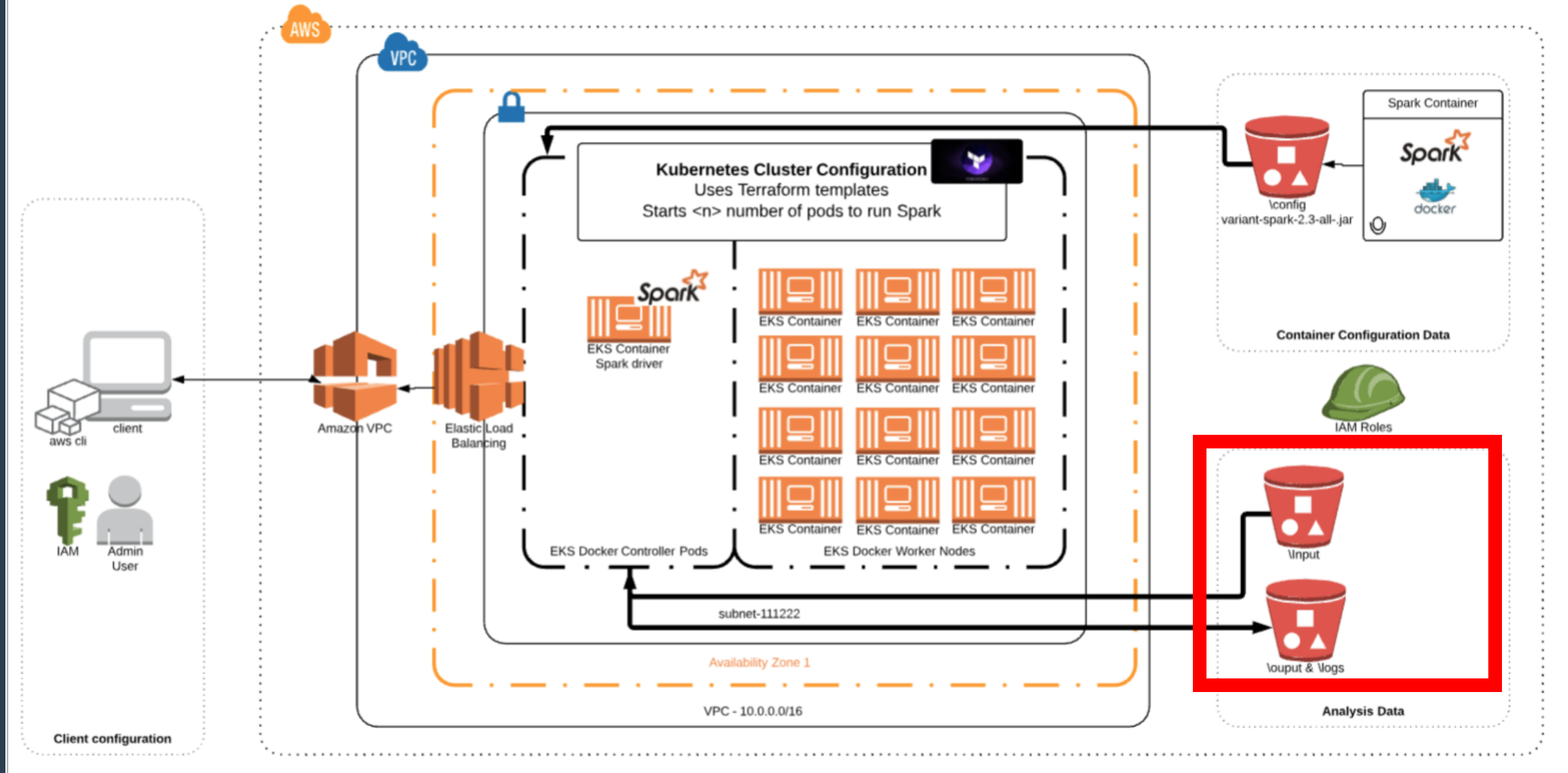
## Understanding the Challenge

Bioinformatics is one of the most interesting and challenging areas to work on scaling big data machine learning solutions. These challenges include not only the size and scale of genomic data (3 Billion DNA ‘letters’ per person). They also include the potential to improve feedback loops for important research in human health, such as understanding significant variants in genomic data for potential CRISPR-Cas9 research. This research can have profound impact on diseases such as cancer.





# FIRST ARCHITECTURE -> HOURS



# 2ND ARCHITECTURE -> MINUTES

Genomic  
Sequence  
Files

$\lambda$ -driven

Machine Learning - GWAS

$\lambda$ -eval  
 $\lambda$ -compress  
 $\lambda$ -convert  
 $\lambda$ -partition

Elastic Cluster

🗄️🗄️🗄️ Data Lake

Web tools  
Views  
local tools  
Jupyter

Collect

PREPARE

PROCESS

VIEW/USE

Jupyter  
Notebook

3RD ARCHITECTURE -  
FULL PIPELINE

WHAT

WENT

WRONG?

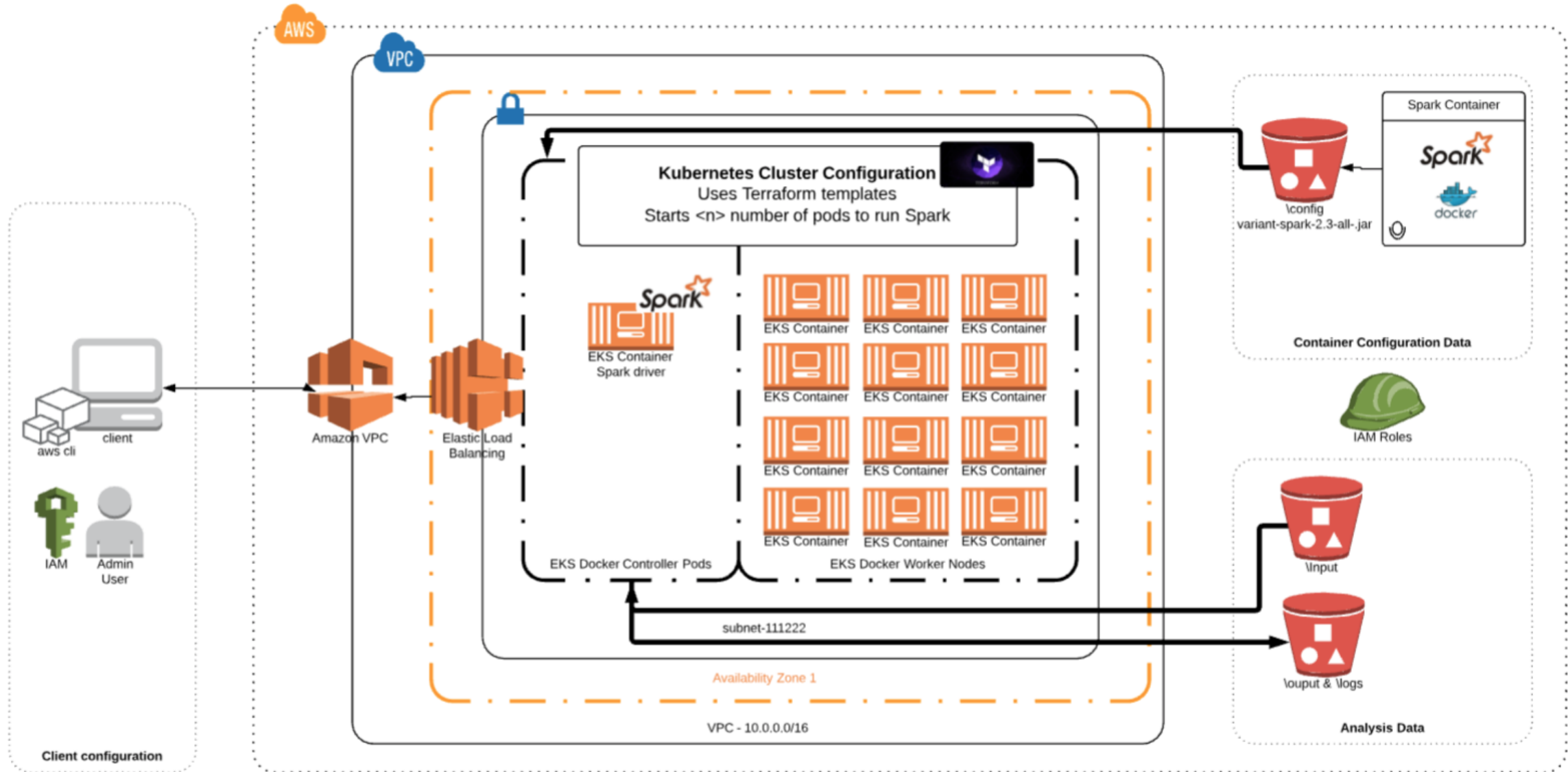
```
terraform {  
  backend "s3" {  
    bucket  = "variant-spark-july"  
    key     = "variantspark-k/tfstate"  
    region  = "us-west-2"  
    profile = "default"  
  }  
}  
  
provider "aws" {  
  profile = "${var.profile}"  
  region  = "${var.default_region}"  
}  
  
provider "aws" {  
  alias    = "us-east-1"  
  profile  = "${var.profile}"  
  region   = "us-east-1"  
}  
  
module "eks-vpc" {  
  source      = "modules/eks-vpc"  
  "cluster-name" = "${var.cluster-name}"  
}  
  
module "eks-master-role" {  
  source = "modules/eks-master-role"  
}  
  
module "eks-master-security-group" {  
  source = "modules/eks-master-security-group"  
  vpc_id = "${module.eks-vpc.vpc_id}"  
}  
  
module "eks-worker-role" {  
  source = "modules/eks-worker-role"  
}
```

# What does this "code" do?

Warning: This Pie Chart can elicit visually-induced seizures.

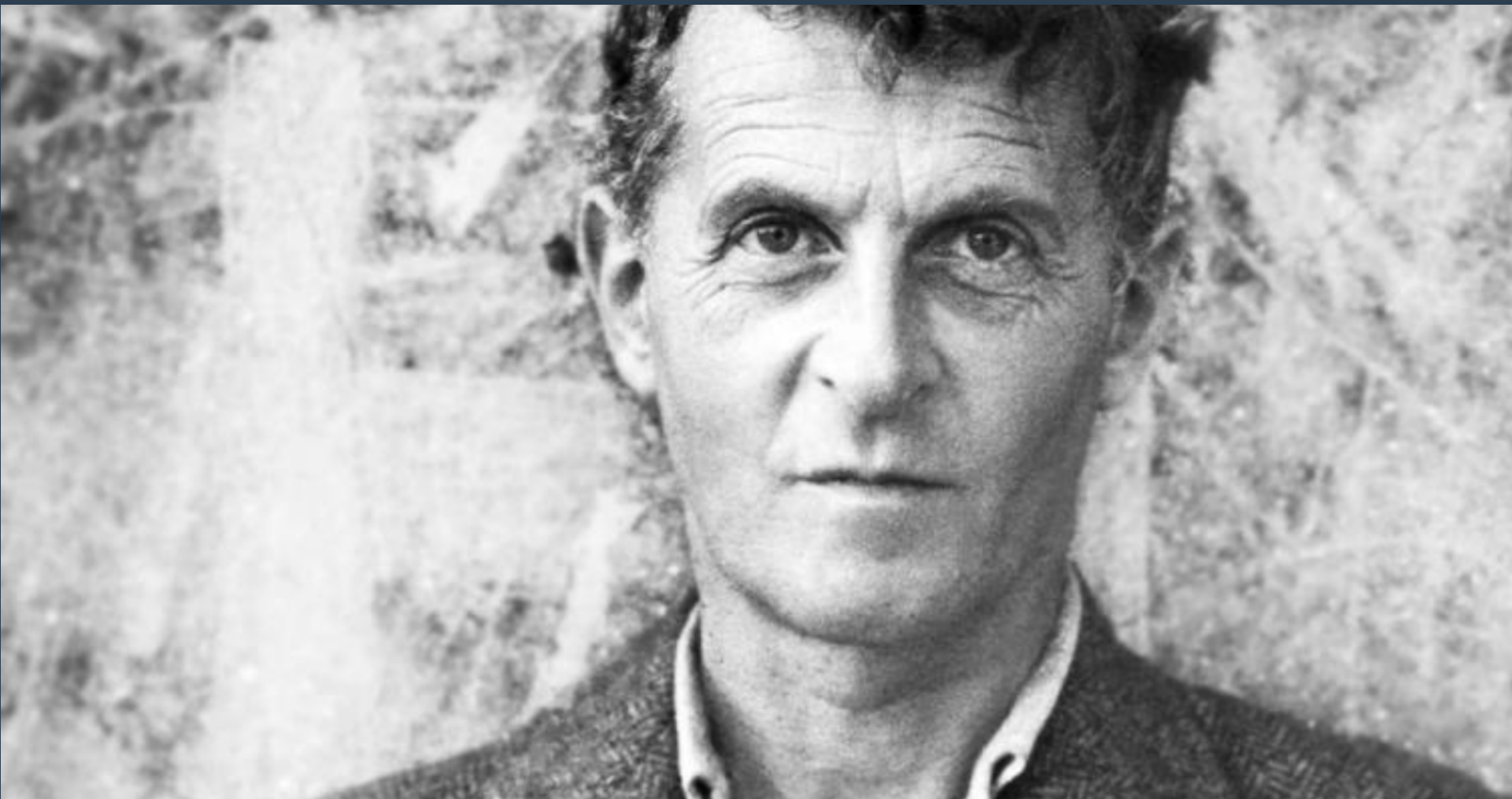
## AWS EKS SPARK

Lynn Langit



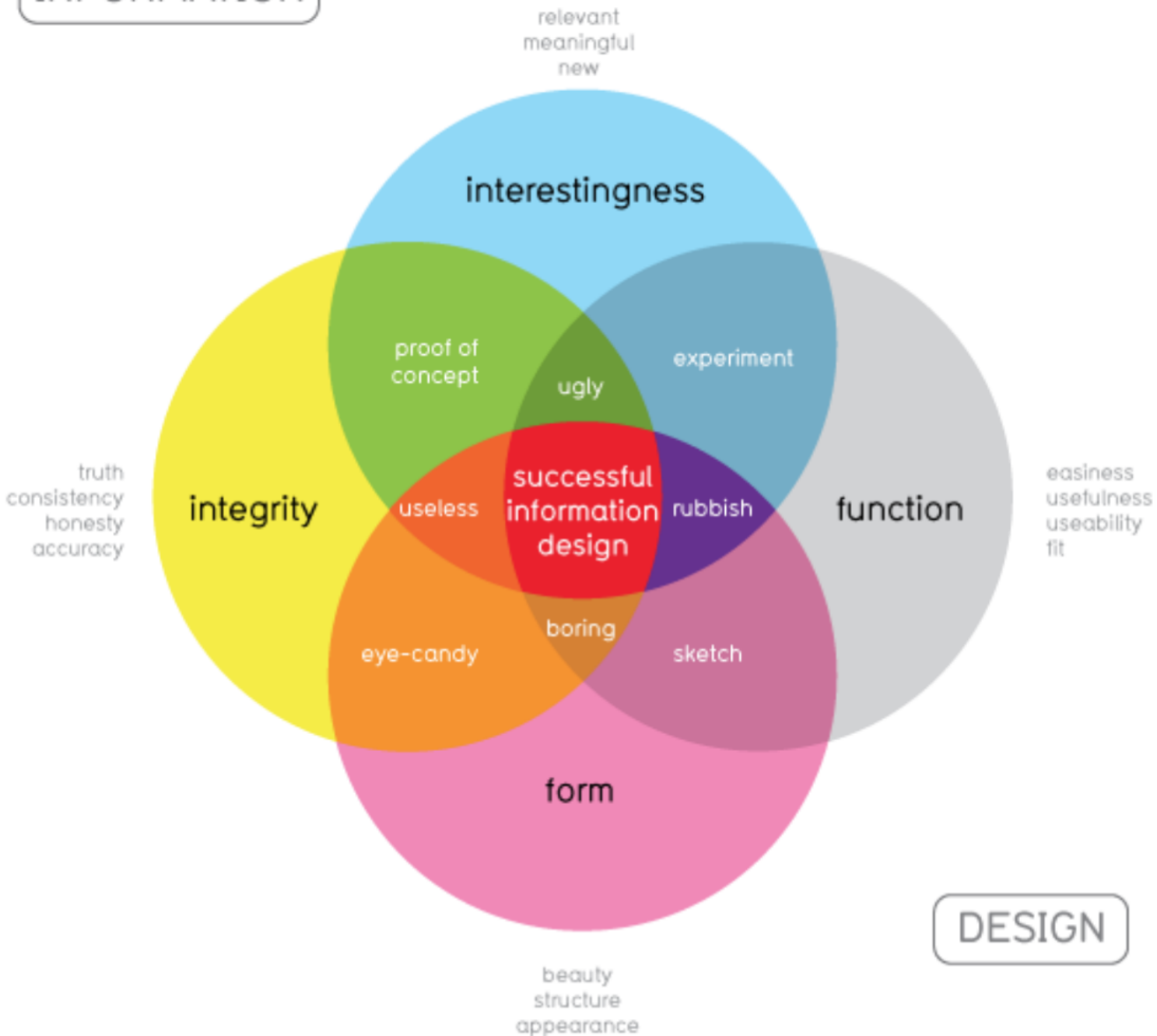
# BIAS

toward adding complexity



*" If we spoke a different language, we would  
perceive a somewhat different world." -  
Wittgenstein*

INFORMATION

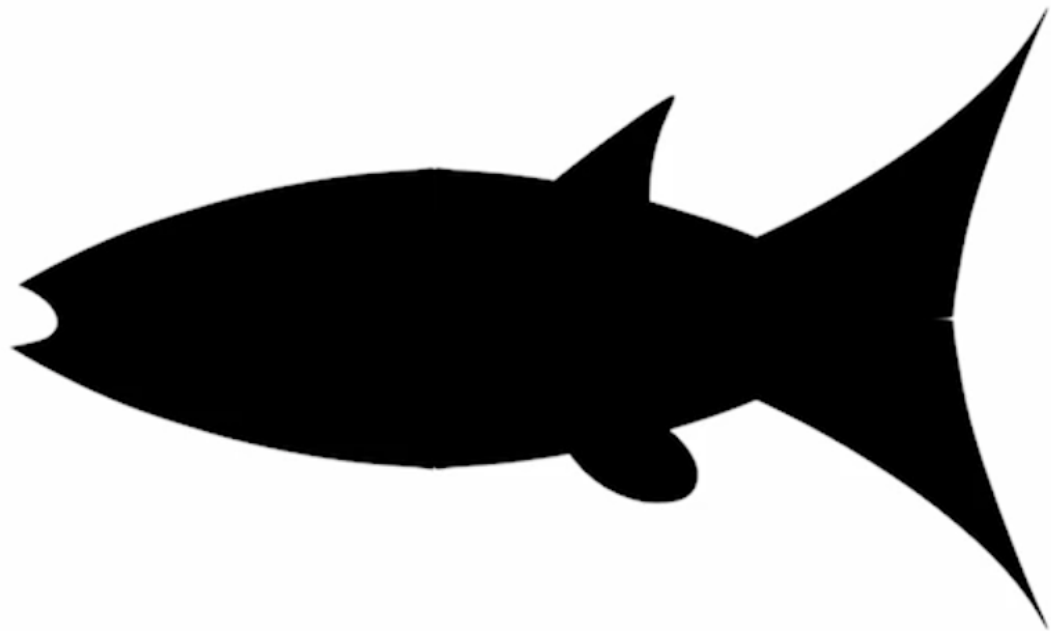


DESIGN

# WHAT MAKES GOOD INFORMATION DESIGN?

# READING

- *"The **Dimensionality** of Visual Space"*
- *"A Visual Interaction Framework for **Dimensionality Reduction** Based Data Exploration"*
- *"Understanding Visualization: A Formal Approach using **Category Theory** and Semiotics"*
- *"Introduction: The **Geometry** of the Visual Field—Early Modern and Contemporary Approaches"*
- *"Visualizing **MNIST**: An Exploration of Dimensionality Reduction"*



*This is not a fish.*



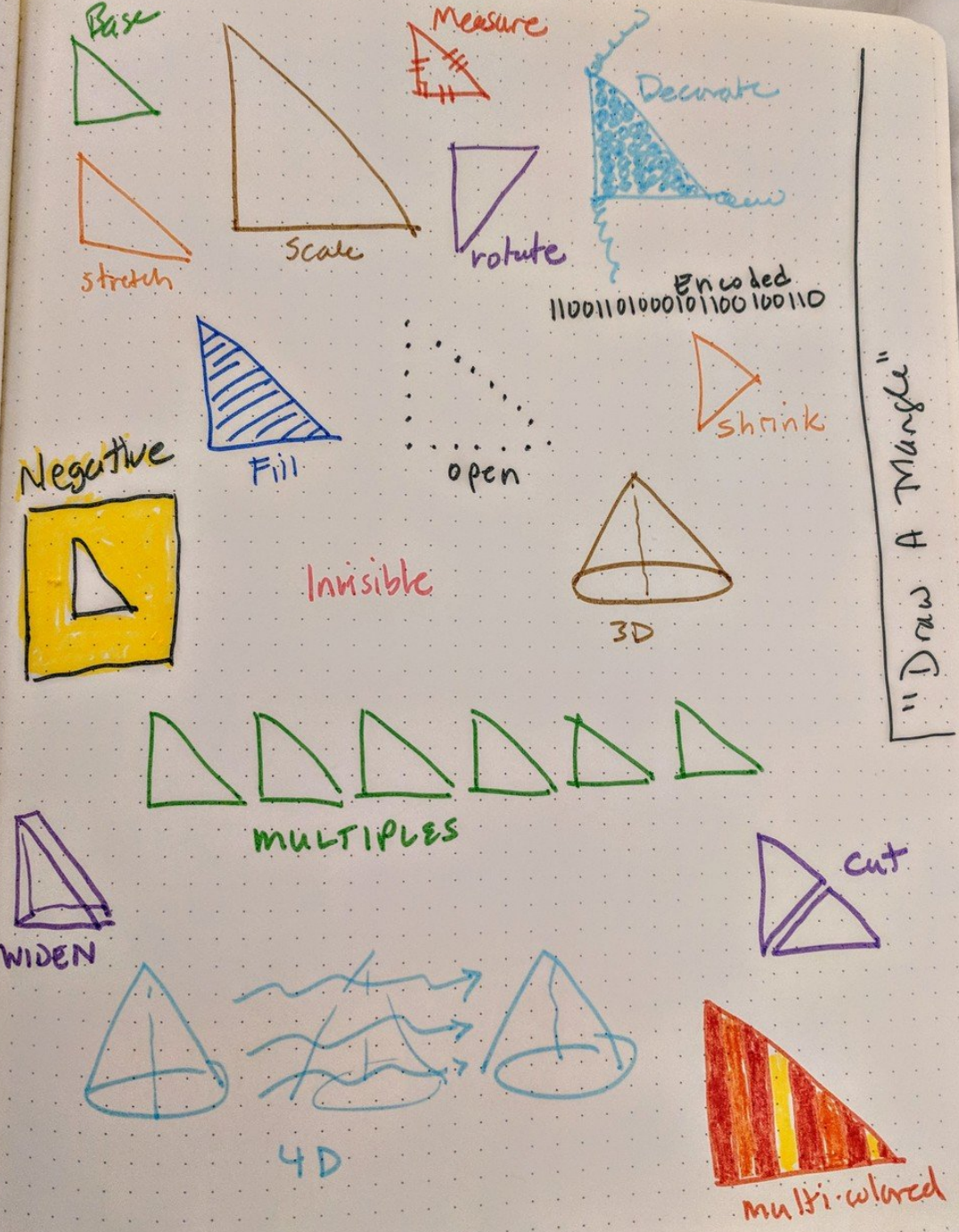
*"Artists draw because they want to convey something they can't describe"*  
- Bret Victor

*"If you can walk you can dance,  
If you can talk you can sing. "*

- African Proverb



**TRIANGLES**



# HOW TO DRAW A TRIANGLE



# BIAS

**not looking outward,  
using only your own context**

# VISUAL GRAMMARS

for Cloud Systems

# DIALECT FOR SYSTEMS

*a particular form of a language  
that is peculiar to a specific region  
or social group*

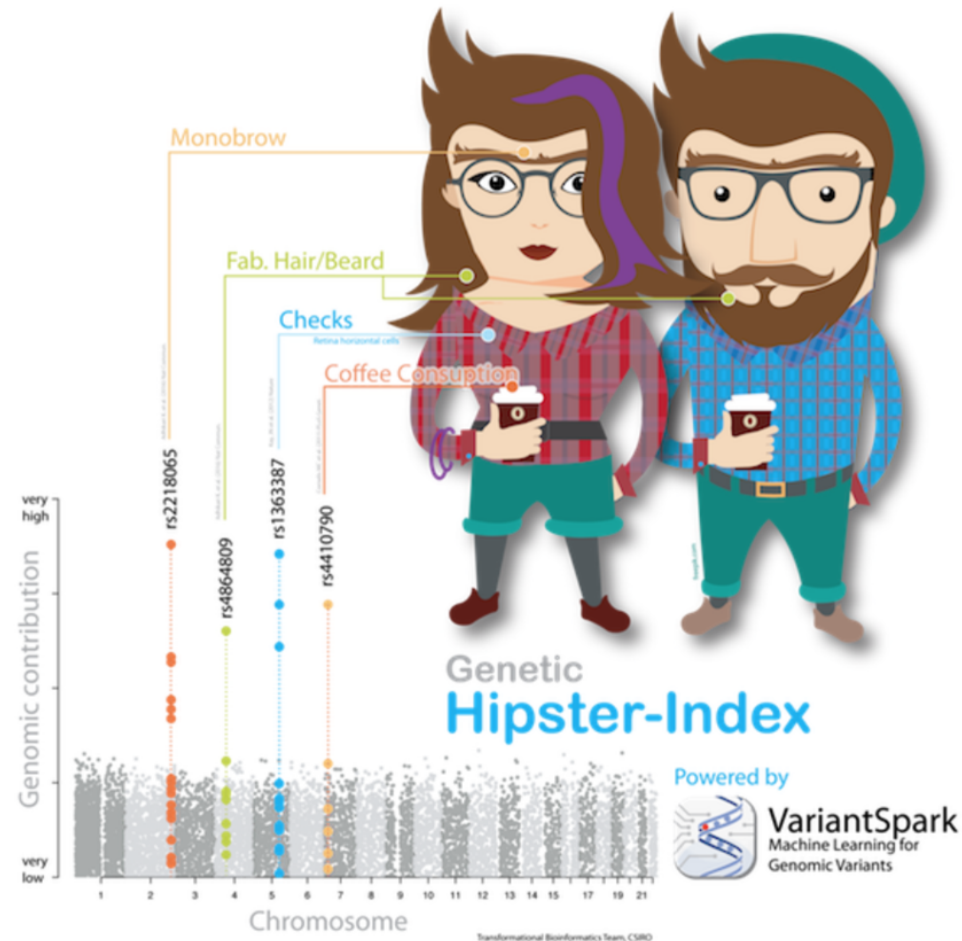
synonyms: regional language,  
local language,  
local speech,  
vernacular, idiom;  
regionalisms,  
localisms;  
*informal lingo*



# ENTRY POINT

**Executable Pathways**

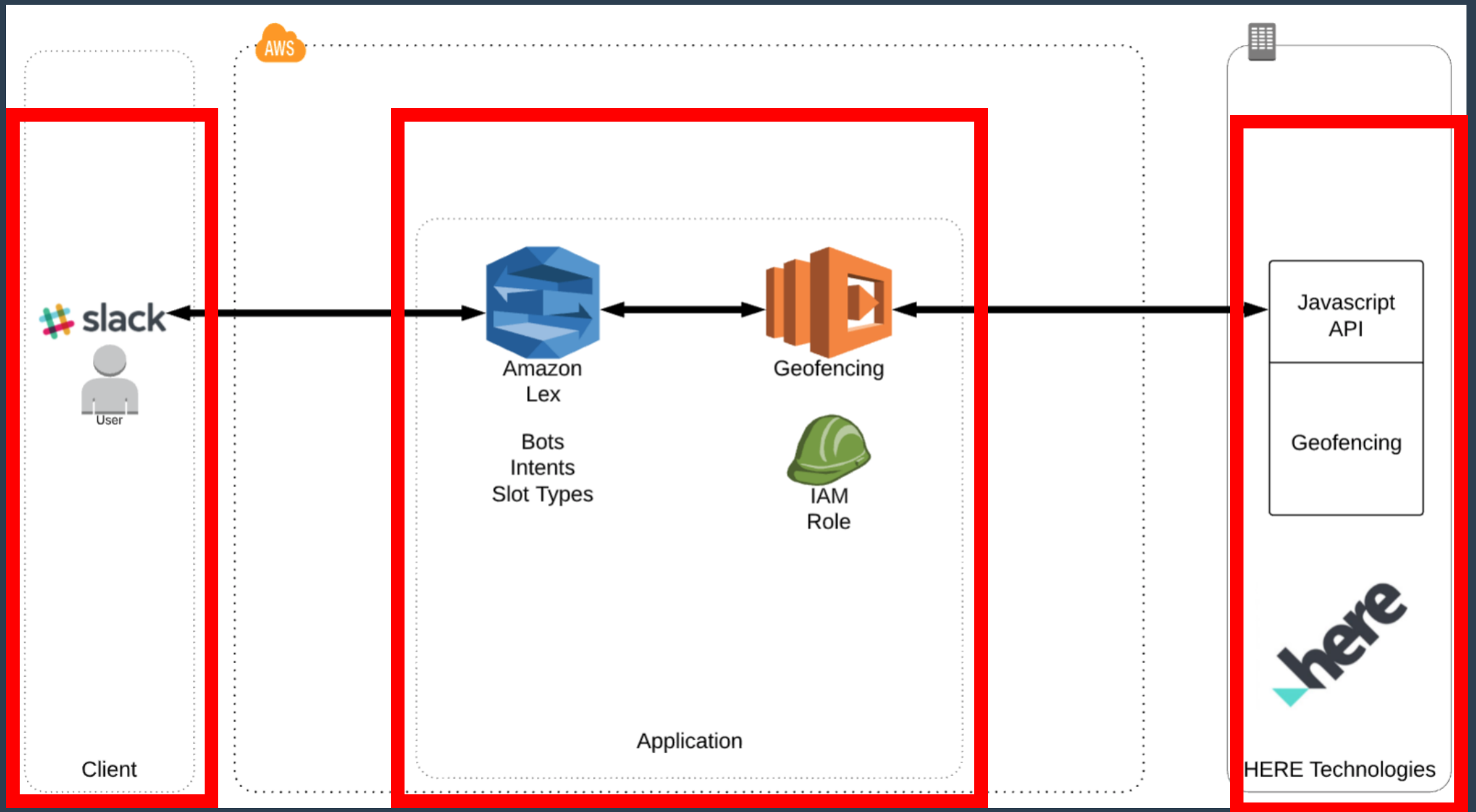
The plot above shows that VariantSpark has recovered the correct genotypes of this multivariate phenotype with interacting features (multiplicative and additive effects).



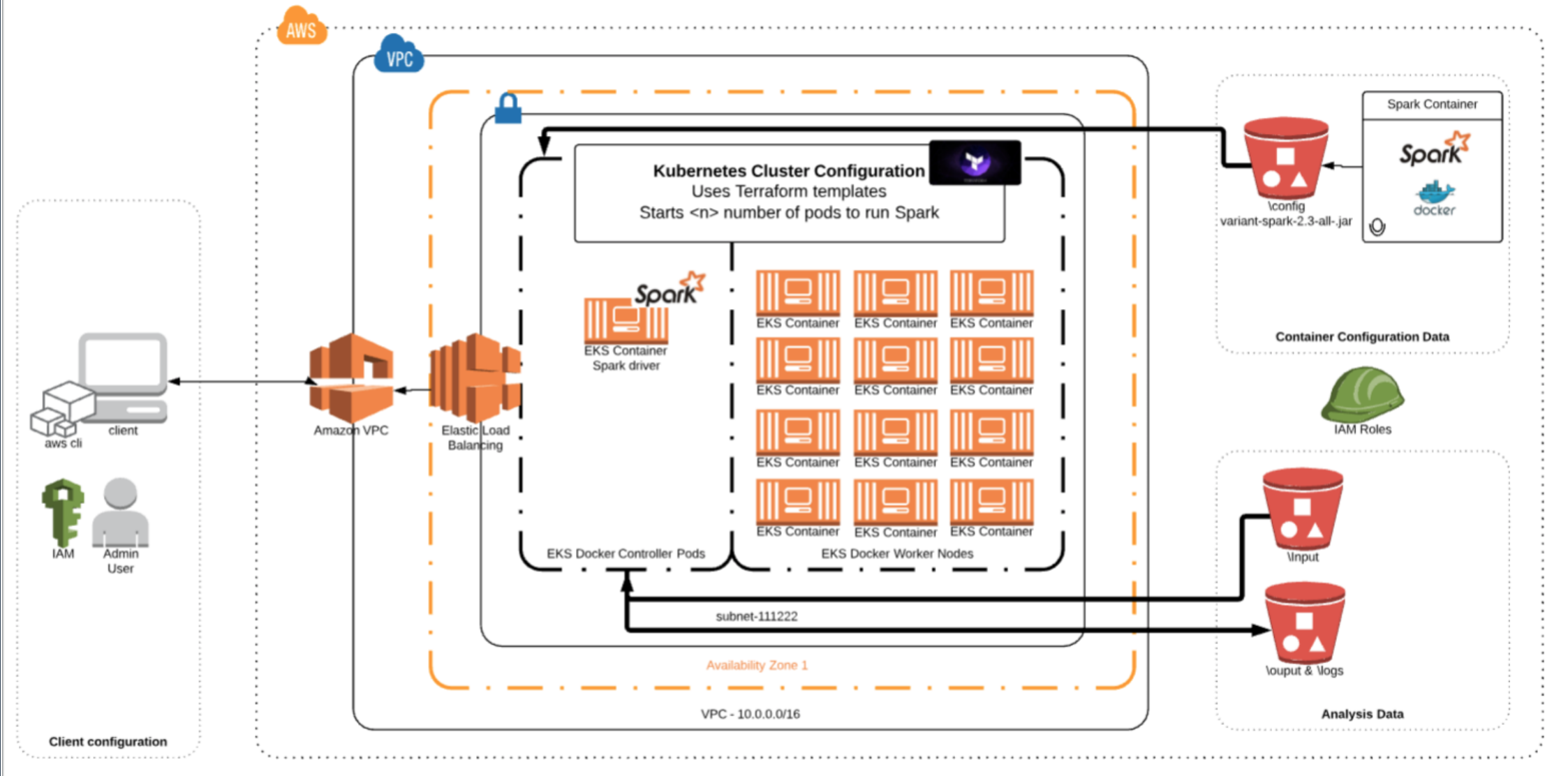
1. **chr2\_223034082** (rs2218065) encoding for monobrow is the most important feature
2. a group of SNPs encoding for the MEGF10 gene (**chr5\_126626044**), which is involved in Retina horizontal cell formation as the second most important marker, explaining why hipsters prefer checked shirts
3. **chr7\_17284577** (rs4410790) the marker for increased coffee consumption is ranked third
4. **chr4\_54511913** (rs4864809) the marker for beards is fourth

# DIMENSIONALITY

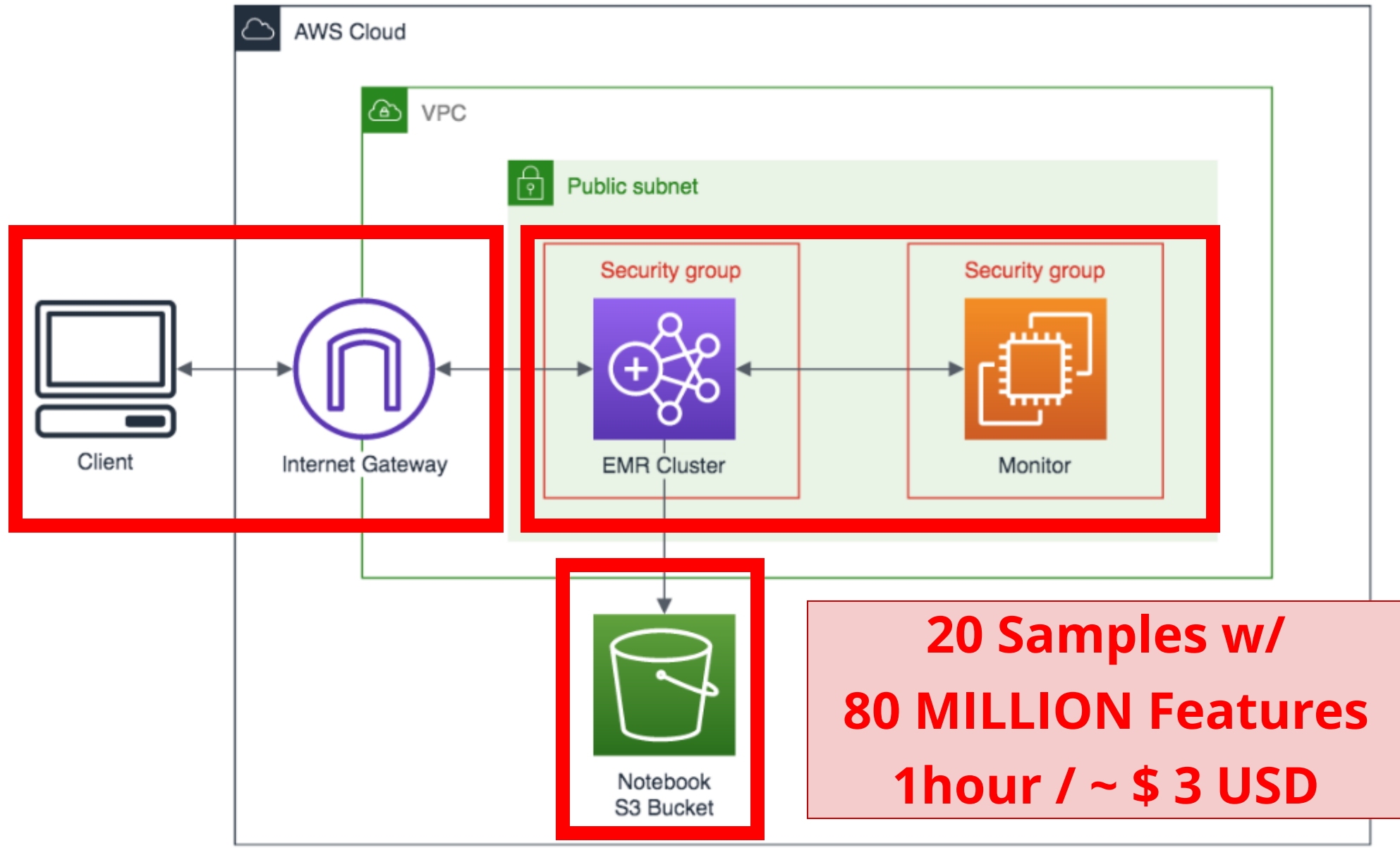
**Reduction**



# '3-TO-5 ASPECT' ARCHITECTURES



# SECOND ARCHITECTURE -> MINUTES



3RD ARCH -> BROAD REACH



CSIRO's Australian e-Health Research Centre (AEHRC) has led the way by becoming the first public sector organisation in the world to publish a machine learning-based health product on the AWS Marketplace.

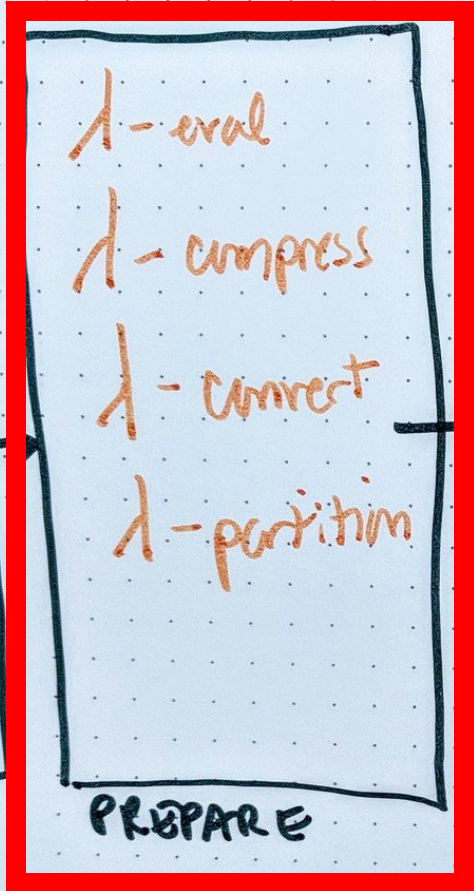
Genomic  
Sequence  
Files

$\lambda$ -driven

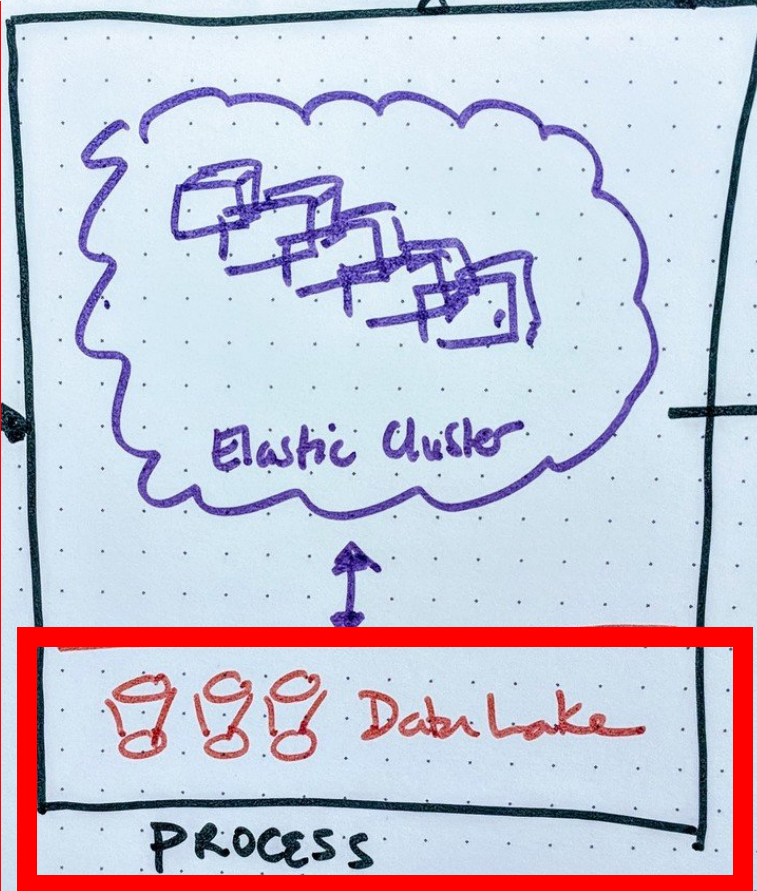
Machine Learning - GWAS



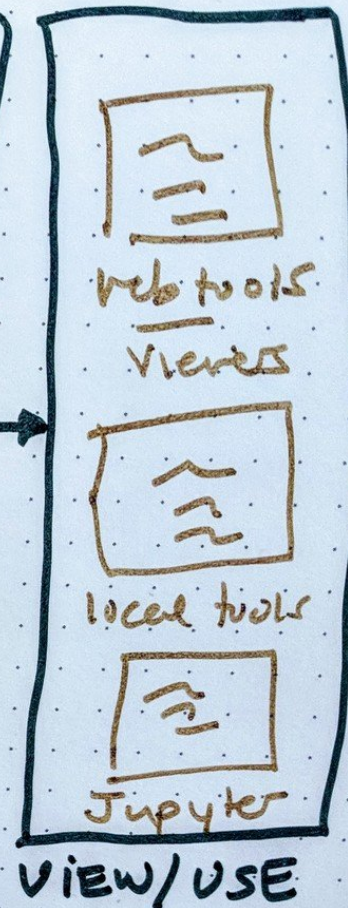
Collect



PREPARE



PROCESS



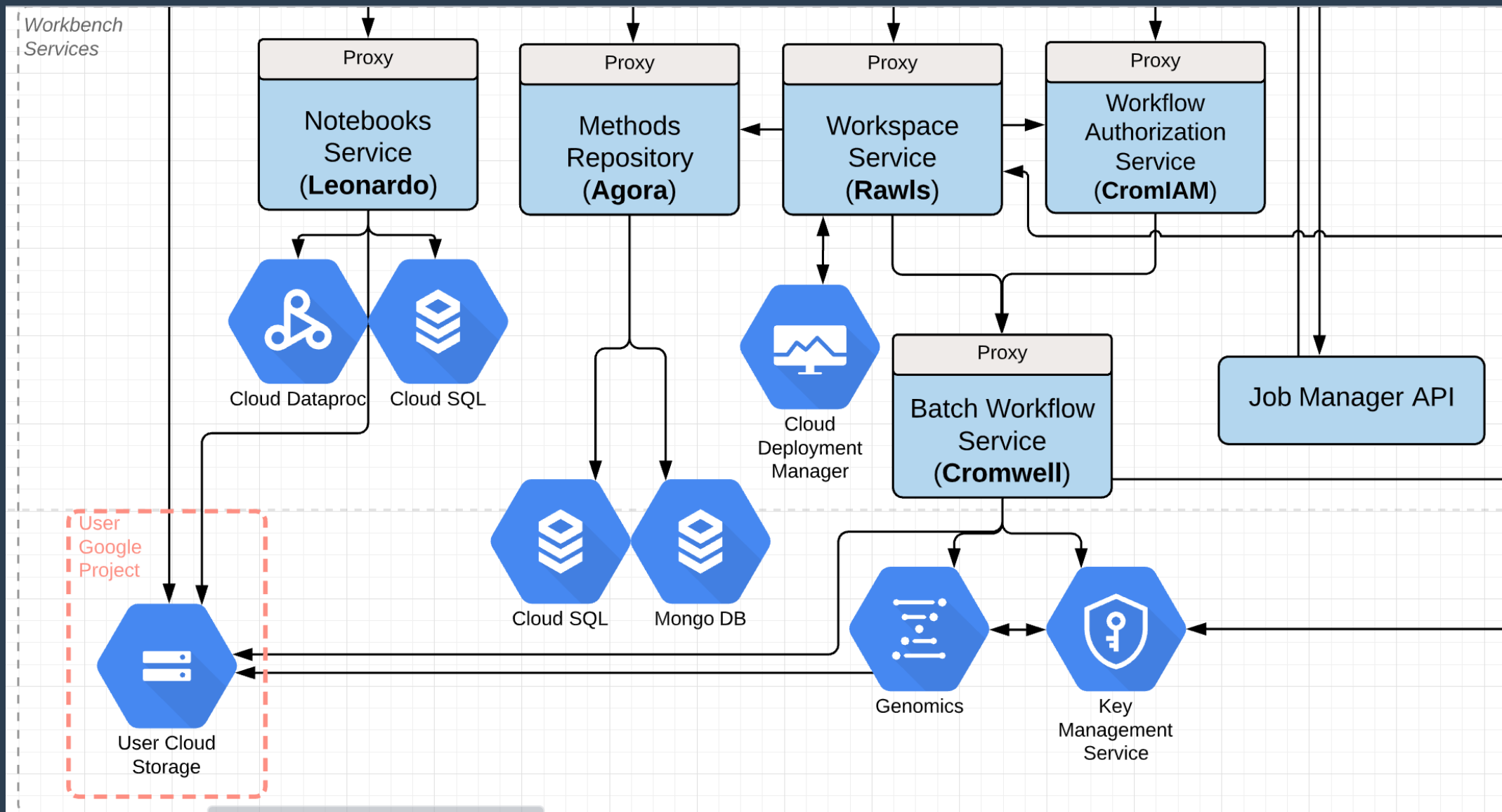
VIEW/USE



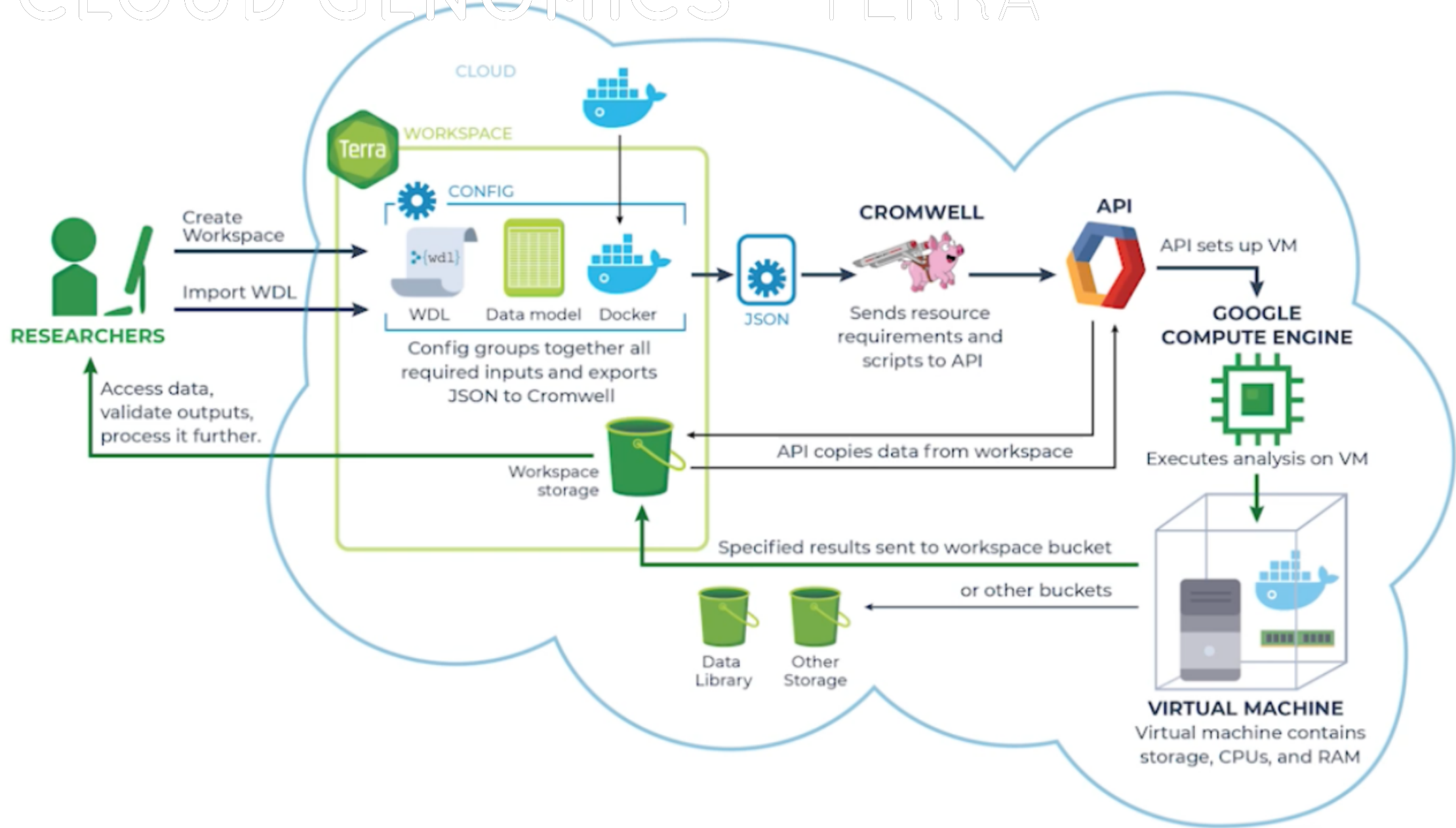
Jupyter  
Notebook(s)

# PERSPECTIVE

User | Developer | DevOps



# CLOUD GENOMICS - TERRA



```
import "https://raw.githubusercontent.com/broadinstitute/viral-pipelines/master/pipes/WDL/tasks/tasks_taxon_filter.wdl" as taxon_f
import "https://raw.githubusercontent.com/broadinstitute/viral-pipelines/master/pipes/WDL/tasks/tasks_assembly.wdl" as assembly

workflow assemble_denovo_with_deplete {

  call taxon_filter.deplete_taxa


  call taxon_filter.filter_to_taxon {
    input:
      reads_unmapped_bam = deplete_taxa.cleaned_bam
  }

  call assembly.assemble {
    input:
      reads_unmapped_bam = filter_to_taxon.taxfilt_bam
  }

  call assembly.scaffold {
    input:
      contigs_fasta = assemble.contigs_fasta,
      reads_bam = filter_to_taxon.taxfilt_bam
  }

  call assembly.refine_2x_and_plot {
    input:
      assembly_fasta = scaffold.scaffold_fasta,
      reads_unmapped_bam = deplete_taxa.cleaned_bam
  }
}
```

What does this "code" do?



BETA

LIBRARY

DATASETS

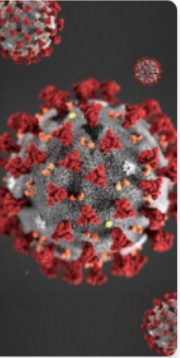
SHOWCASE & TUTORIALS

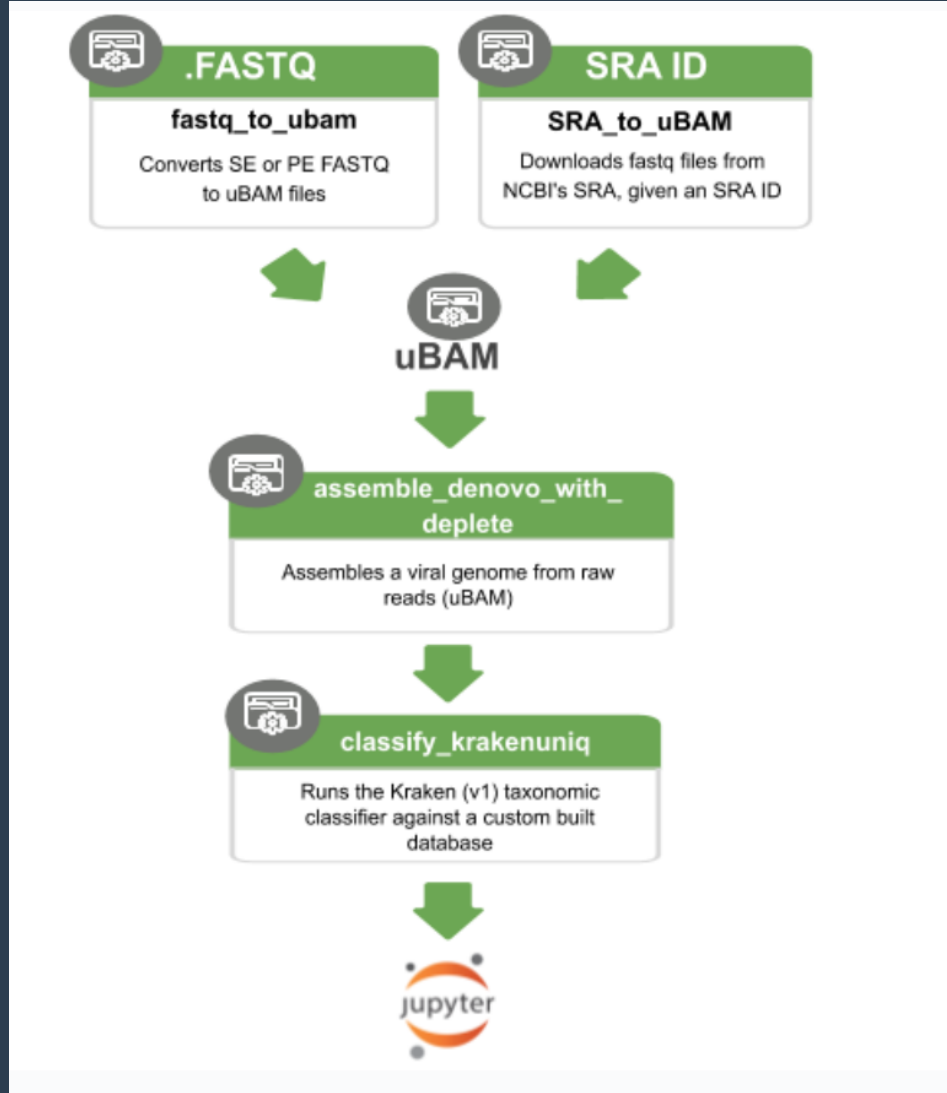
CODE & W

## New and interesting

COVID-19

This workspace contains COVID-19 genomic data and workflows that will enable you to perform viral genomic analysis. This workspace will be routinely updated with new, additional data as it





# //TODO #NOW

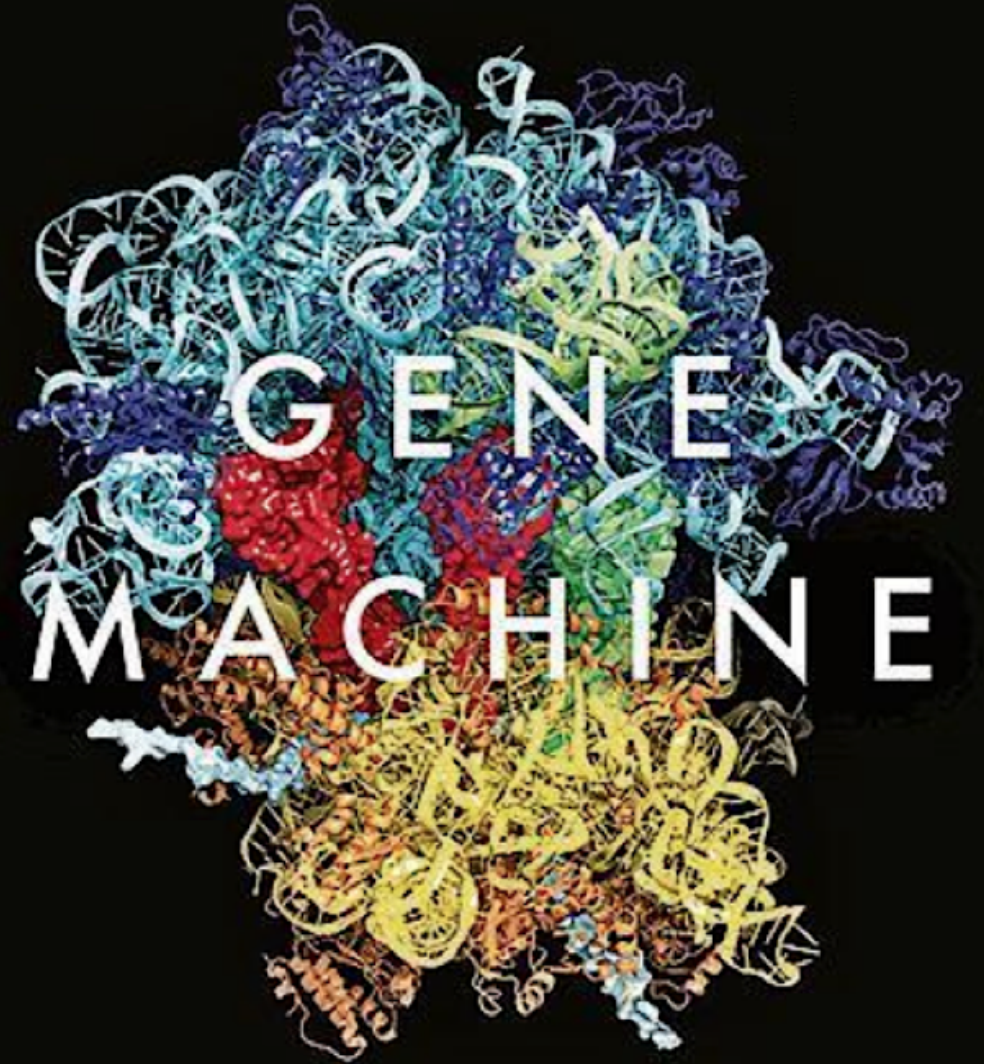
ACTION	TARGET	INFO
Visualize	Entire System	Code, Config & Data Entry point (Hello...)
Use	Visual Grammars	Account for bias, audience
Reduce	Dimensionality	Show 3-5 aspects max
Show	System State	Over time
Verify	Pictured Objects	Reproducibility

# VISUALIZING

Cloud Systems  
@lynnlangit

Presentation Links

VENKI RAMAKRISHNAN  
WINNER OF THE NOBEL PRIZE IN CHEMISTRY



# LOOKING FORWARD

**Promising Directions**

Google Cloud PlatformGoogleComputeEngine

Create an instance

To create a VM instance, select one of the options:

New VM instance

New VM instance from template

Marketplace

Name

instance-1

Region

us-central1 (Iowa)

Zone

us-central1-a

Machine configuration

Machine family

General-purposeMemory-optimized

Machine types for common workloads, optimized for cost and flexibility

Generation

Second

Powered by Cascade Lake CPU platform

Machine type

Custom

Cores

12 vCPU 2 - 80

Memory

48 GB 6 - 96

☐ Extend memory

CPU platform

Automatic

GPUs

+ Add GPU

\$357.51 monthly estimate

That's about \$0.49 hourly

Pay for what you use: No upfront costs and per second billing

Item	Estimated costs
12 vCPUs + 48 GB memory	\$446.39/month
10 GB standard persistent disk	\$0.40/month
Sustained use discount	-\$89.28/month
<b>Total</b>	<b>\$357.51/month</b>

[Compute Engine pricing](#)

[Less](#)

COMBINING VIEWS -  
DYNAMIC GCP 'SLIDERS'



## System Context

The system plus users  
and system dependencies



## Containers

The overall shape of the architecture  
and technology choices



## Components

Logical components and their  
interactions within a container



## Classes

Component or pattern  
implementation details

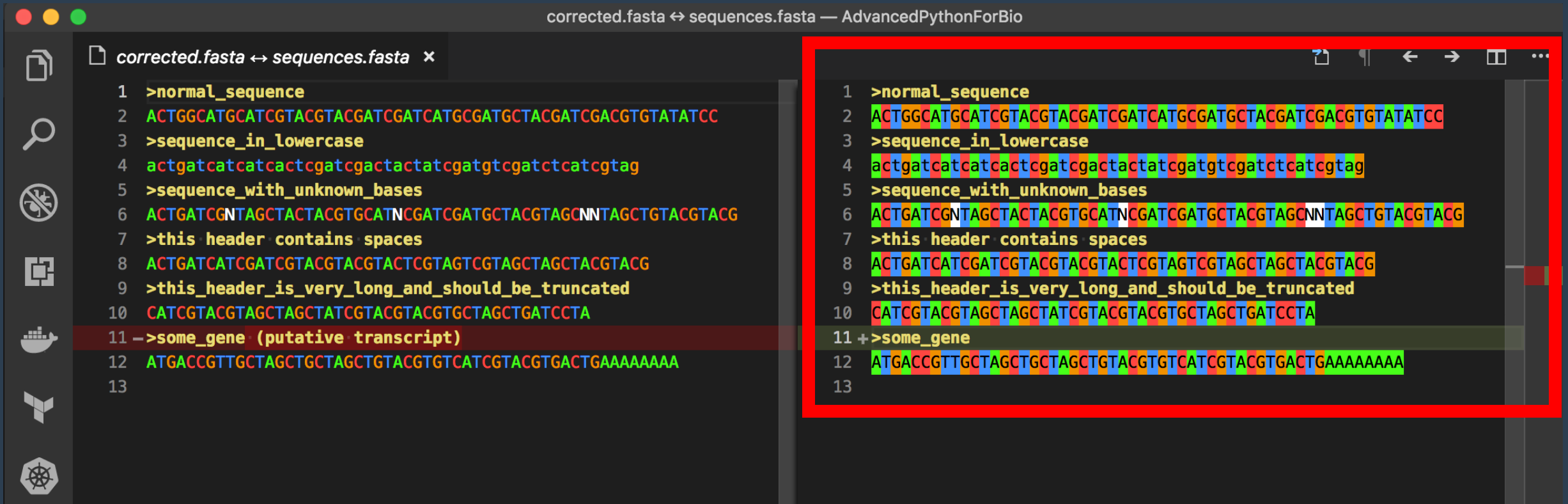
**Overview  
first**

**Zoom and  
filter**

**Details  
on demand**

# ADD LEVELS - C4 MODELS

# EXTENDABLE IDE - VSCODE



corrected.fasta ↔ sequences.fasta — AdvancedPythonForBio

```
corrected.fasta ↔ sequences.fasta x
1 >normal_sequence
2 ACTGGCATGCATCGTACGTACGATCGATCATGCGATGCTACGATCGACGTGTATATCC
3 >sequence_in_lowercase
4 actgatcatcatcactcgatcgactactatcgatgtcgatctcatcgtag
5 >sequence_with_unknown_bases
6 ACTGATCGNTAGCTACTACGTGCATNCGATCGATGCTACGTAGCNNTAGCTGTACGTACG
7 >this_header_contains_spaces
8 ACTGATCATCGATCGTACGTACGTACTCGTAGTCGTAGCTAGCTACGTACG
9 >this_header_is_very_long_and_should_be_truncated
10 CATCGTACGTAGCTAGCTATCGTACGTACGTGCTAGCTGATCCTA
11 ->some_gene (putative transcript)
12 ATGACCGTTGCTAGCTGCTAGCTGTACGTGTCATCGTACGTGACTGAAAAAAAAA
13
```

```
1 >normal_sequence
2 ACTGGCATGCATCGTACGTACGATCGATCATGCGATGCTACGATCGACGTGTATATCC
3 >sequence_in_lowercase
4 actgatcatcatcactcgatcgactactatcgatgtcgatctcatcgtag
5 >sequence_with_unknown_bases
6 ACTGATCGNTAGCTACTACGTGCATNCGATCGATGCTACGTAGCNNTAGCTGTACGTACG
7 >this_header_contains_spaces
8 ACTGATCATCGATCGTACGTACGTACTCGTAGTCGTAGCTAGCTACGTACG
9 >this_header_is_very_long_and_should_be_truncated
10 CATCGTACGTAGCTAGCTATCGTACGTACGTGCTAGCTGATCCTA
11 +>some_gene
12 ATGACCGTTGCTAGCTGCTAGCTGTACGTGTCATCGTACGTGACTGAAAAAAAAA
13
```

VISUALIZE DATA FILES (AND CODE)

# DESIGN TO BUILD - STACKERY

The image displays the Stackery 'Design to Build' interface, which bridges visual resource modeling with infrastructure-as-code templates.

**Left Panel (Visual Model):**

- Shows a resource titled `AWS::CLOUDFRONT::DISTRIBUTION` with the label `Cdn`.
- Below the resource, there are five expandable sections: `Origin`, `Viewer Requests`, `Origin Requests`, `Origin Responses`, and `Viewer Responses`.
- A green `+ Add Resource` button is located in the top right of the visual editor.




**Right Panel (Template Editor):**

The corresponding CloudFormation template is displayed in a dark-themed editor. The visible code is as follows:







```
1 Resources:
2   Cdn:
3     Type: AWS::CloudFront::Distribution
4     Properties:
5       DistributionConfig:
6         DefaultCacheBehavior:
7           Compress: true
8           ForwardedValues:
9             Cookies:
10              Forward: none
11              QueryString: false
12            TargetOriginId: StackeryCDN
13            ViewerProtocolPolicy: redirect-to-https
14          Enabled: true
15          PriceClass: PriceClass_100
16   EdgeFunction:
17     Type: Custom::StackeryEdgeFunction
18     Properties:
19       ServiceToken: !Sub arn:aws:lambda:${AWS::Region}:${AWS::AccountId}:function:stac
20       FunctionName: !Sub ${AWS::StackName}-EdgeFunction
21       Description: !Sub
22         - Stack ${StackTagName} Environment ${EnvironmentTagName} Function ${ResourceN
23         - ResourceName: EdgeFunction
24       Runtime: nodejs8.10
25       CodeUri: src/EdgeFunction
26       Handler: index.handler
27       MemorySize: 3008
28       Timeout: 30
29       Tracing: Active
30       Role: !GetAtt EdgeFunctionRole.Arn
31   EdgeFunctionRole:
32     Type: AWS::IAM::Role
33     Properties:
34       RoleName: !Sub ${AWS::StackName}-EdgeFunction
35       AssumeRolePolicyDocument:
36         Version: 2012-10-17
37         Statement:
38           Effect: Allow
39           Principal:
40             Service:
41               - lambda.amazonaws.com
42               - edgelambda.amazonaws.com
43           Action: sts:AssumeRole
44       ManagedPolicyArns:
45         - arn:aws:iam::aws:policy/service-role/AWSLambdaBasicExecutionRole
46         - arn:aws:iam::aws:policy/AWSXrayWriteOnlyAccess
```

# GENERALIZATIONS- CELL-BASED ARCHITECTURE

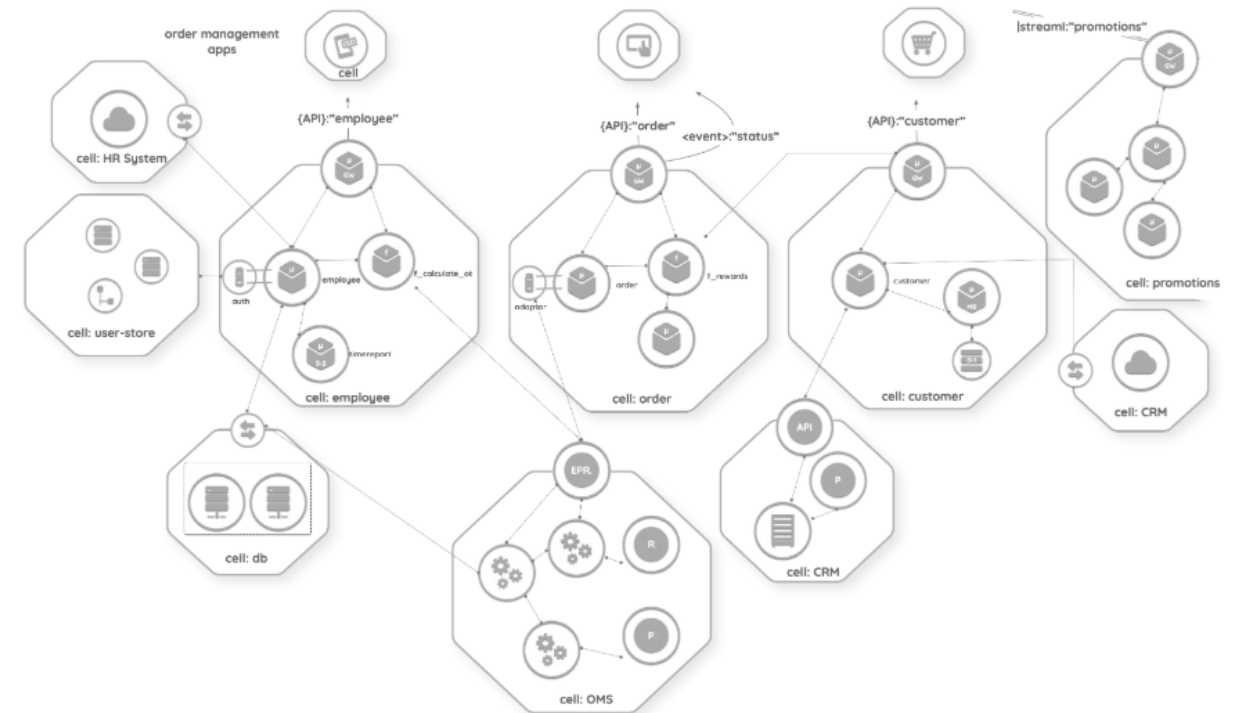
## Section 1: Abstractions

Icon	Name	Description
	User	A user is a human interacting with the system. Users can be external customers of the organization, internal employees, or partners.
	Component	A <b>component</b> is the <b>atomic unit</b> of this reference architecture. A component represents a process or business logic running in a container, serverless environment, or an existing runtime. This can then be categorized into many subtypes based on the functional capabilities. A component is designed based on a specific scope, which can be independently run and reused at the runtime. Runtime requirements and the behavior of the component vary based on the component type and the functional capabilities. The user may decide to build and run the code as a service, function, or microservice, or choose to reuse an existing legacy service based on the architectural need.
	Cell	Cell - an architecture block represents by an <b>octagon</b> .

## Component Types: Main Categories

Icon	Category	Component
	Legacy and data services	Databases, Existing systems, Registries and repositories, User stores, Business processes
	Microservices and serverless components	Core business logic, Aggregation and service composition, Transformation
	Gateways and brokers	Exposed APIs, events, and streams, Policy enforcement points, Message brokers, Identity brokers and identity gateways, Sidecars and bridges
	External endpoint	Access using APIs, events, and streams, Cloud systems and SaaS
	Frontend clients	Mobile apps, Reactive apps, API consumers
	Governance and utilities	Registry, Observability, Automation tools

## Mapping an Agile Enterprise to the Real-World: An Order Management System Reference Implementation



# DETECT DRIFT - CLOUD FORMATION

**Differences (4)** < 1 >

<input checked="" type="checkbox"/>	Property	Change	Expected value	Current value
<input checked="" type="checkbox"/>	AttributeDefinitions.1	ADD	-	{"AttributeName": "stuff", "AttributeType": "S"}
<input checked="" type="checkbox"/>	ProvisionedThroughput.ReadCapacityUnits	NOT_EQUAL	5	6
<input checked="" type="checkbox"/>	ProvisionedThroughput.WriteCapacityUnits	NOT_EQUAL	5	4
<input checked="" type="checkbox"/>	Tags	REMOVE	[{"Key": "test", "Value": "test"}]	-

**Details**

**Expected**

```
{
  "AttributeDefinitions": [
    {
      "AttributeName": "Name",
      "AttributeType": "S"
    }
  ],
  "KeySchema": [
    {
      "AttributeName": "Name",
      "KeyType": "HASH"
    }
  ],
  "ProvisionedThroughput": {
    "ReadCapacityUnits": 5,
    "WriteCapacityUnits": 5
  },
  "Tags": [
    {
      "Key": "test",
      "Value": "test"
    }
  ]
}
```

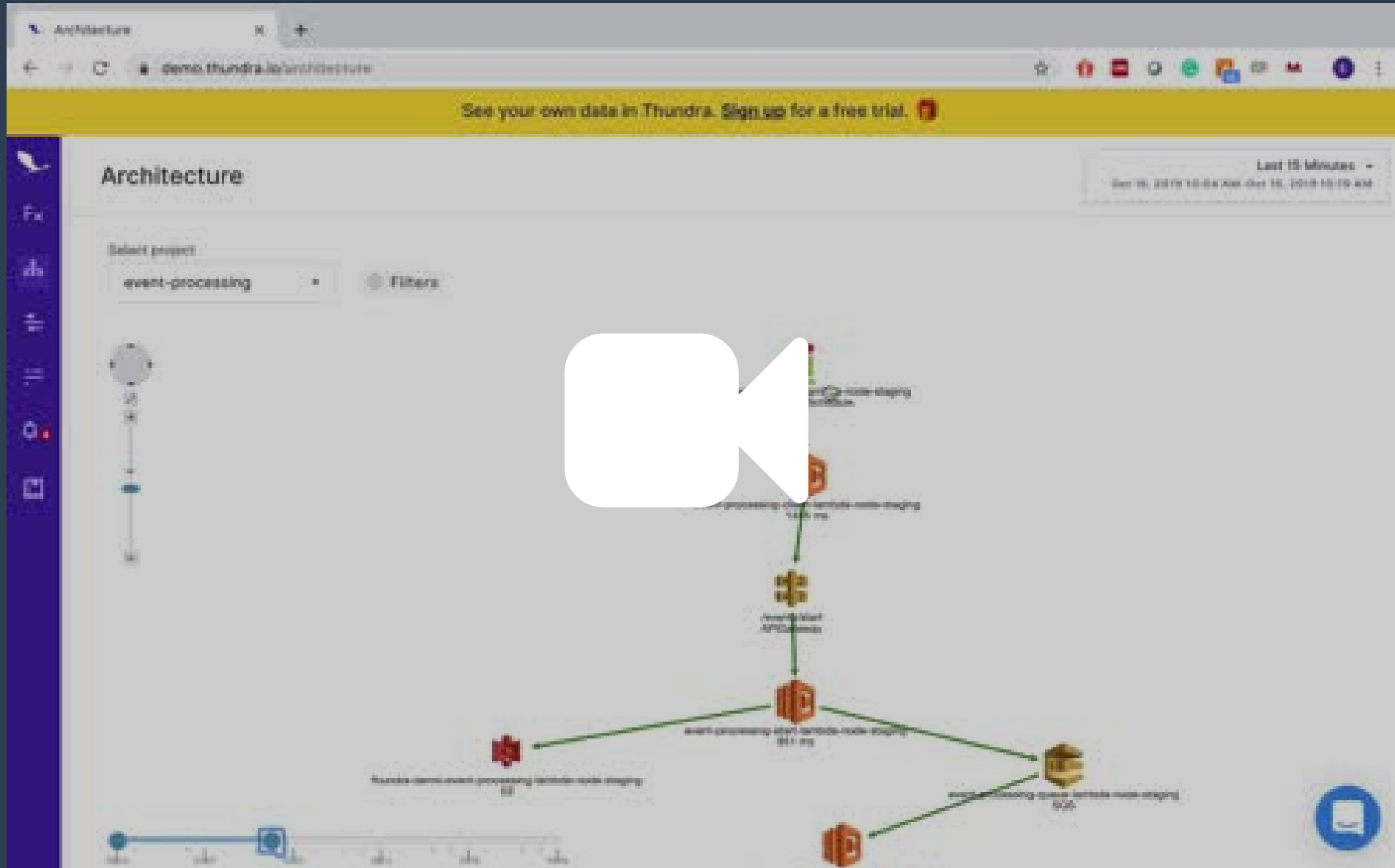
**Actual**

```
{
  "AttributeDefinitions": [
    {
      "AttributeName": "Name",
      "AttributeType": "S"
    },
    {
      "AttributeName": "stuff",
      "AttributeType": "S"
    }
  ],
  "KeySchema": [
    {
      "AttributeName": "Name",
      "KeyType": "HASH"
    }
  ],
  "ProvisionedThroughput": {
    "ReadCapacityUnits": 6,
    "WriteCapacityUnits": 4
  }
}
```

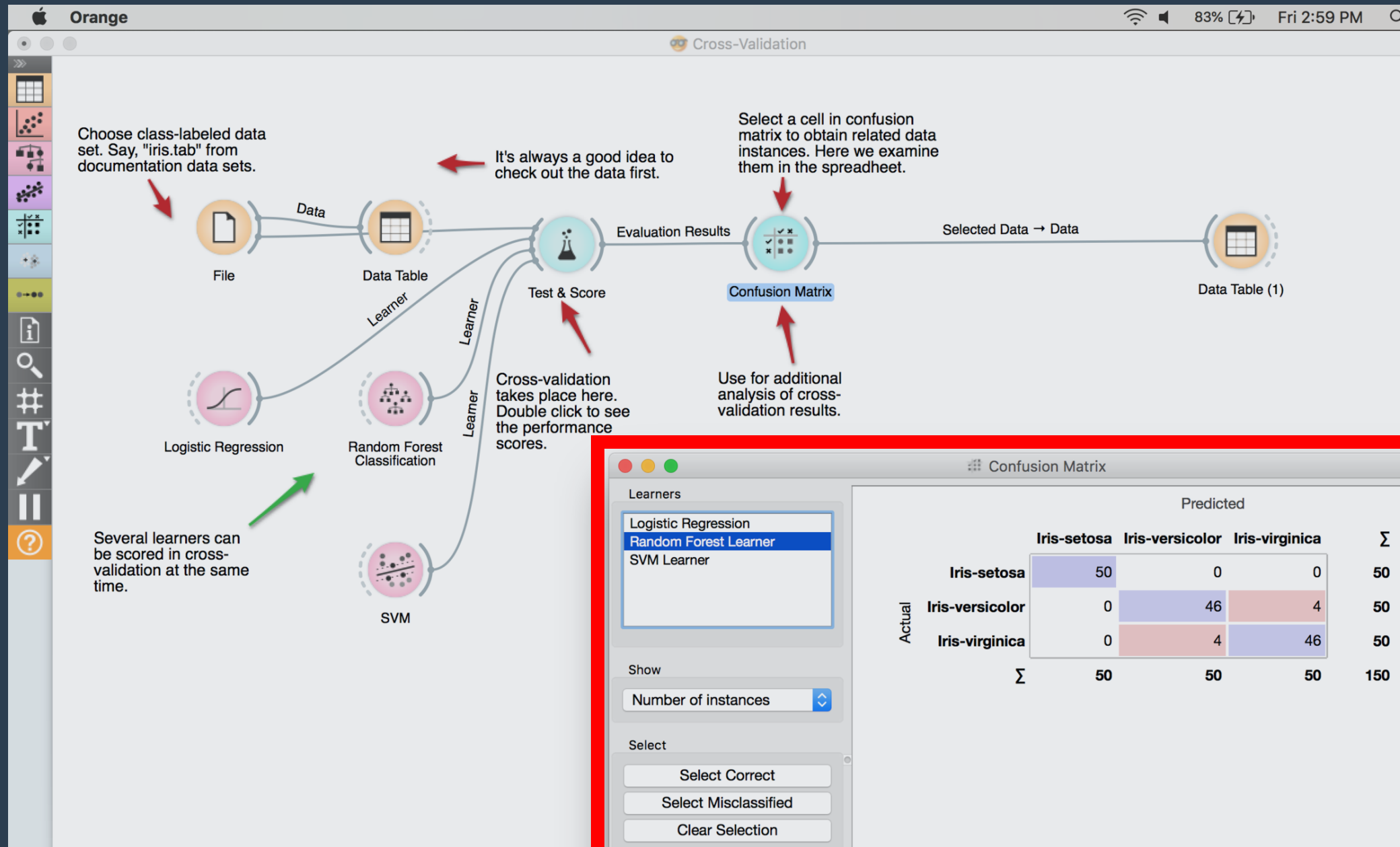
# CLOUD MONITORING

TYPE	METHOD	INFO
WebUI	log viewer	AWS CloudWatch, GCP Operations, Alibaba Logs
Alerts	bash-like Pythonic	awscli, gcloud, aliyun
Aggregations	script	AWS Log Metrics, GCP Operations Metrics
WebUI	language visual tool	Honeycomb Thundra
Vertical WebUI	language visual tool	Terra.bio JobRunner on GCP Nextflow Tower Galaxy Project Job Executions on AWS

# INTEGRATED OBSERVABILITY - THUNDRA

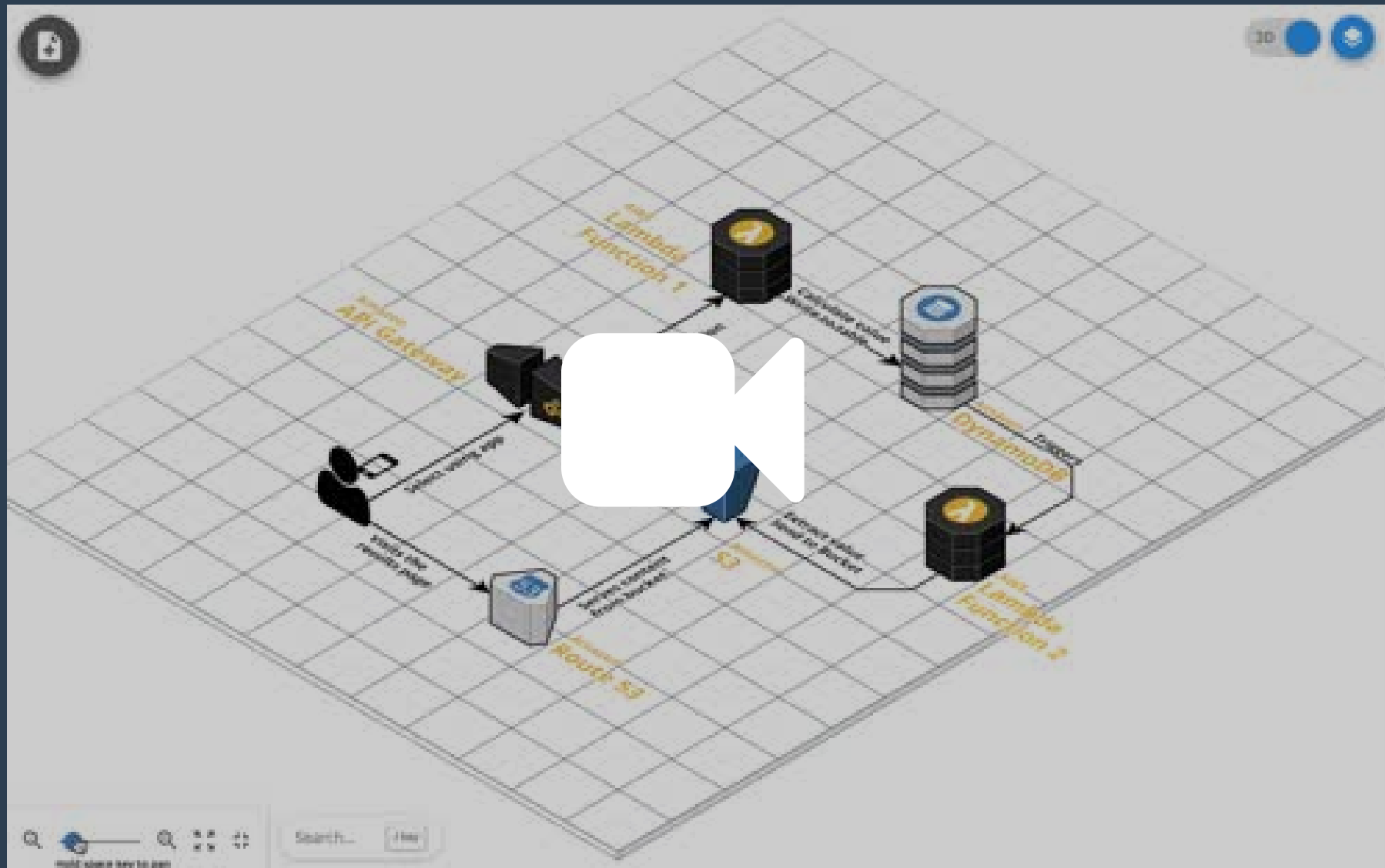


# EXECUTABLE PIPELINE- ORANGE

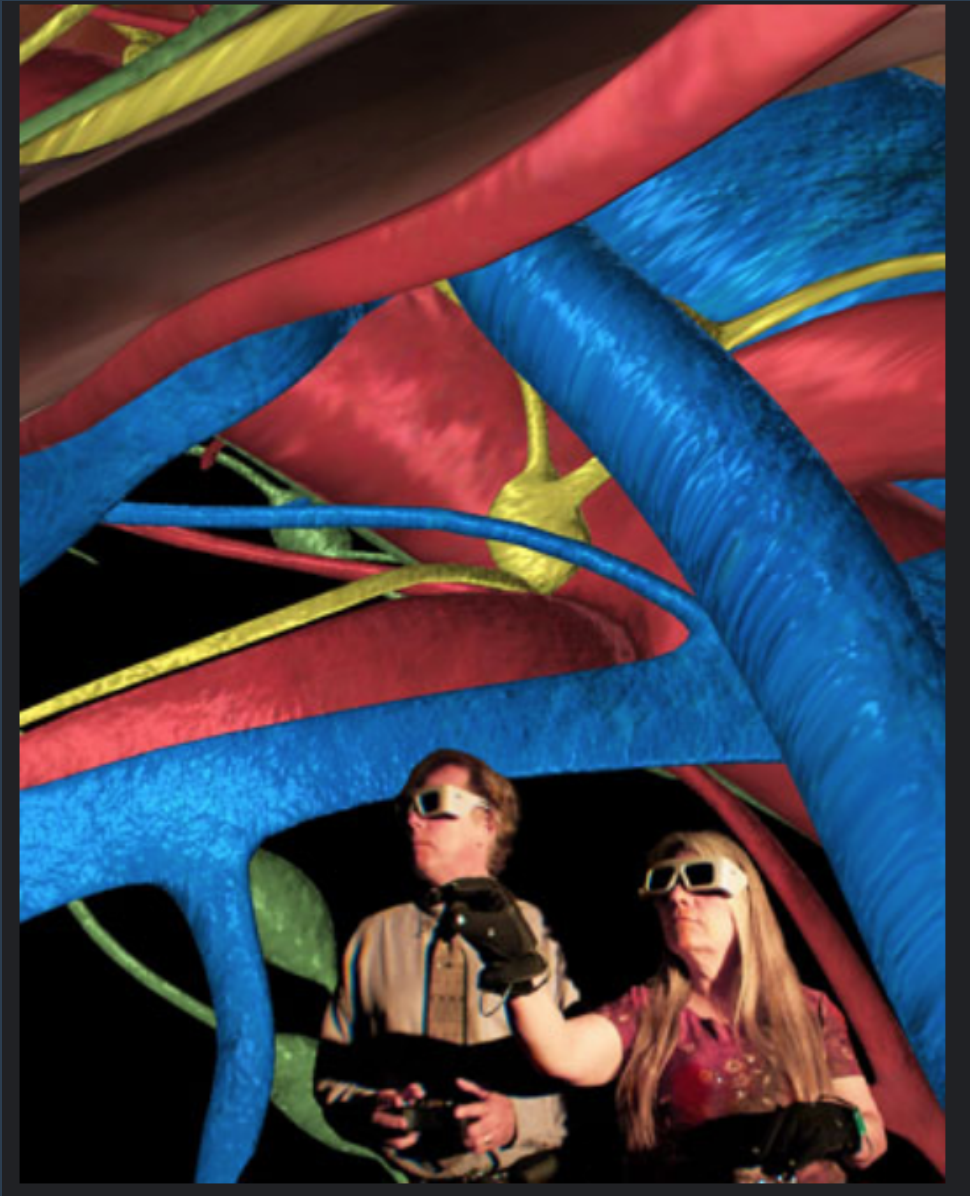


EXECUTABLE  
DIAGRAMS?

# DESIGN/COSTS - CLOUDCRAFT



# DATA IMMERSION - ALLOSPHERE



"a new way to see, hear  
and interpret scientific  
data"

# PRODUCTS TO WATCH

Company	Product/Service	Notable	Lifecycle
AWS	Cloud Formation X-Ray	Drift Detection Log Visualization	ALL
GCP	GCP console	Integration of estimated service costs	ALL
Stackery CloudCraft	Draw Config	Generates YAML Monitors costs	BUILD/DEPLOY
Thundra Epsagon	Log Aggregation	Includes external services Application / transactions	OPERATE
Datadog	ARM for Serverless	Log Visualization	OPERATE
Honeycomb	Honeycomb	Query and visualize event patterns	OPERATE

//TODO #NEXT  
VISUALIZE LIVING  
SYSTEMS



# ARTIST IN RESIDENCE - THE BROAD INSTITUTE

<https://www.youtube.com/embed/5CjYo3h7akc?enablejsapi=1>