BUILDER SESSION - WPS201

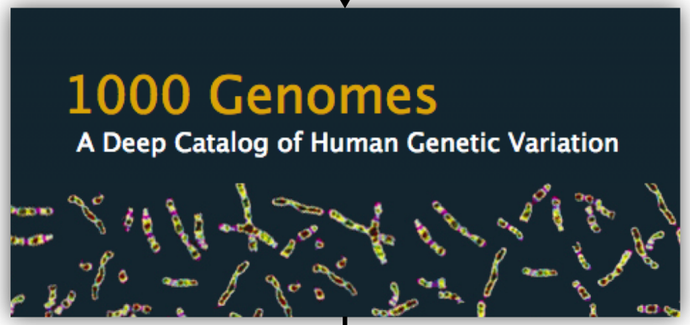
Building a Genome Clustering Model with Amazon Sagemaker

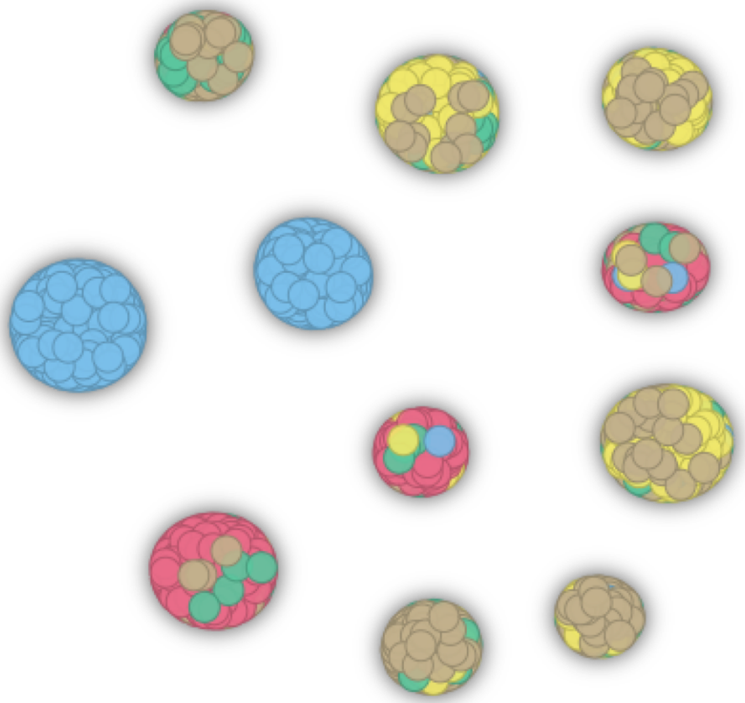
W. Lee Pang, PhD
Genomics & Life Sciences Specialist Architect
Amazon Web Services / WW Public Sector

AWS re:Invent

aws

What we are building





Geographic Population Clusters

Tools and services we will use

Infrastructure Automation

AWS CloudFormation

Data Storage

Amazon S3

Data Processing

AWS Batch

Data Science

Amazon SageMaker

Let’s start building!
http://bit.ly/reinvent2018-wps201

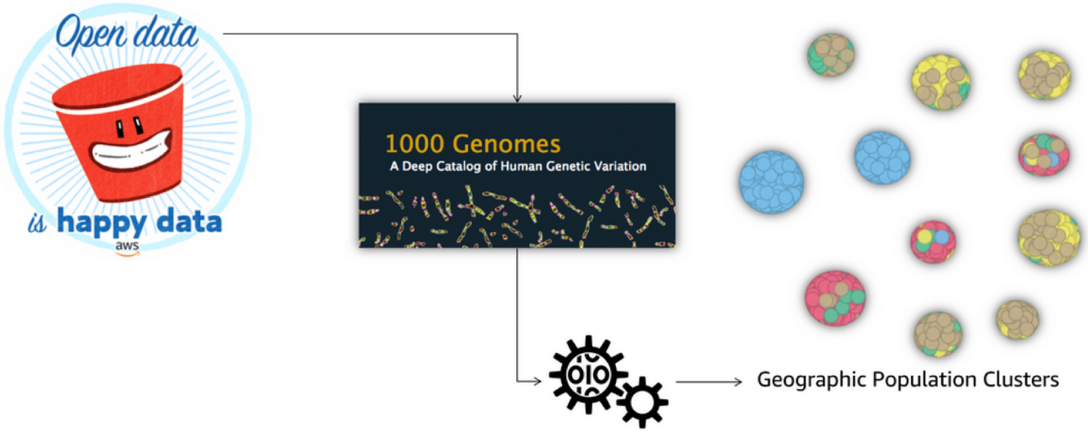
AWS re:Invent

aws

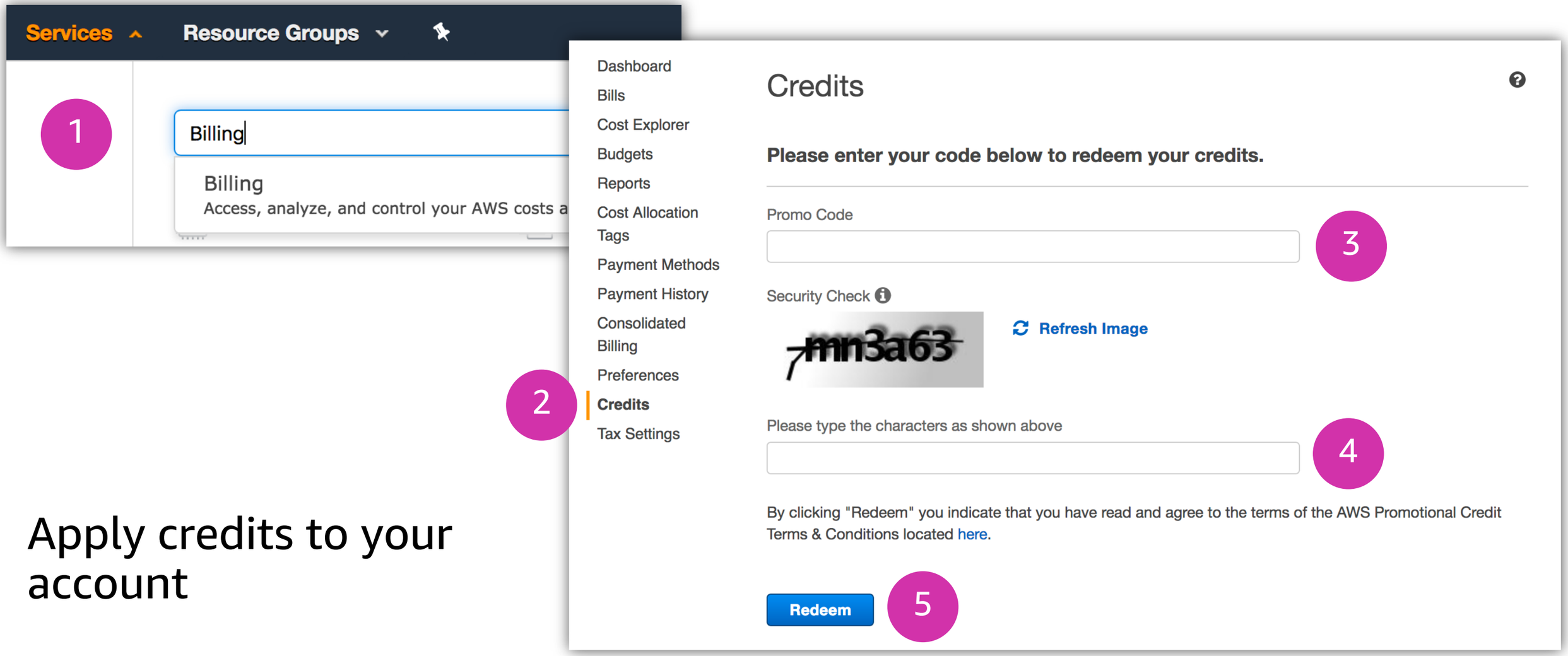
http://bit.ly/reinvent2018-wps201

re:Invent 2018 - WPS201 Search GitHub

Builder Session - WPS201



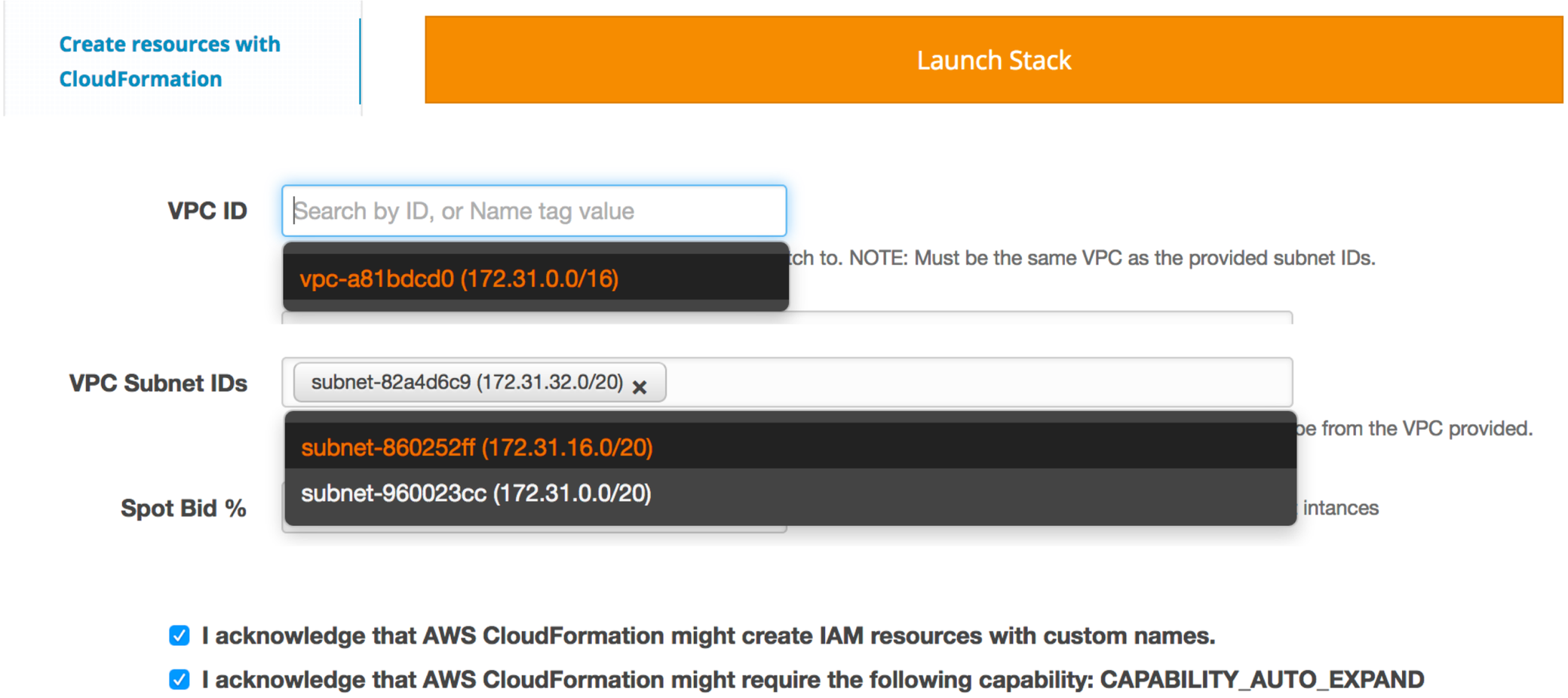
http://bit.ly/reinvent2018-wps201

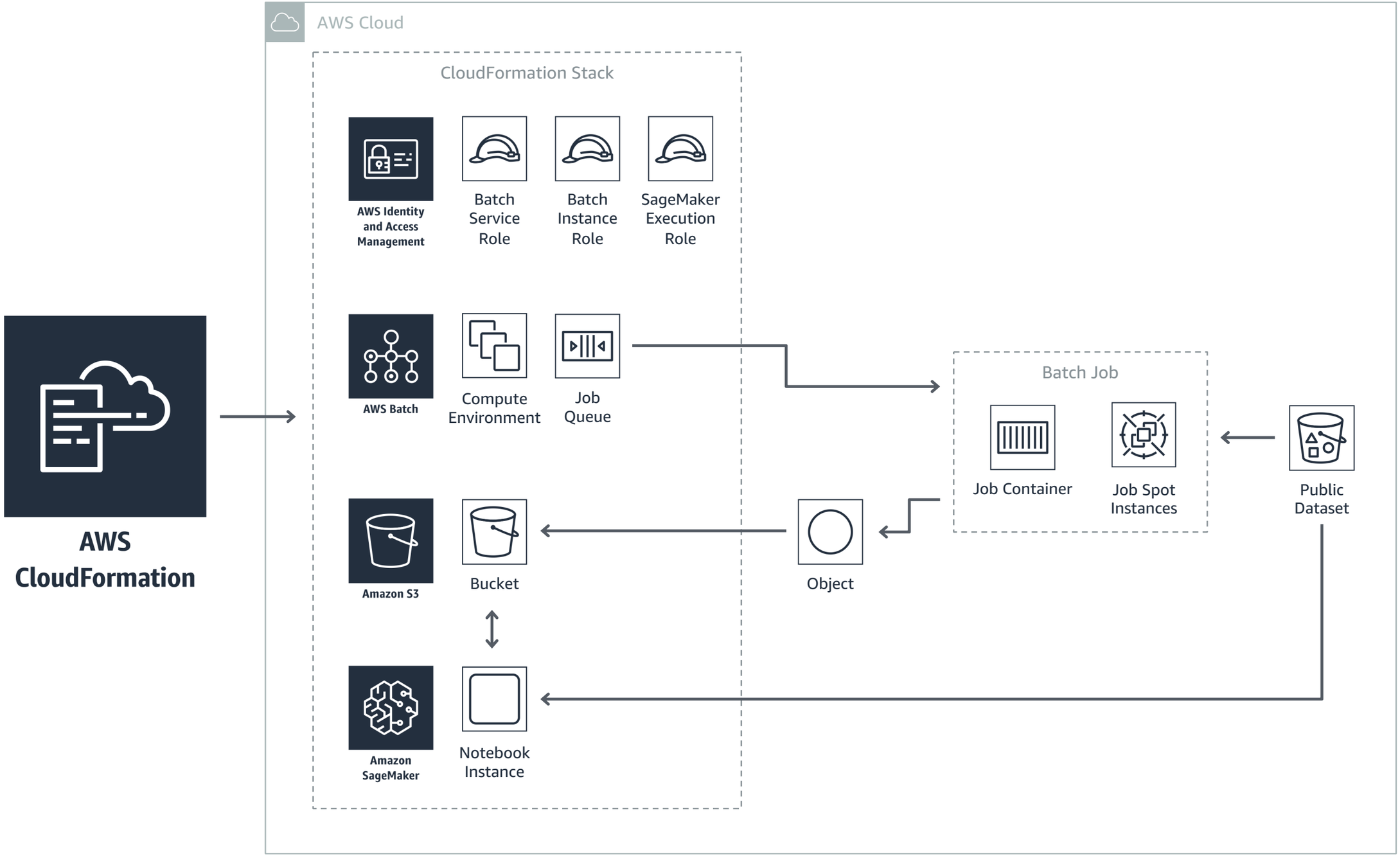


Apply credits to your account

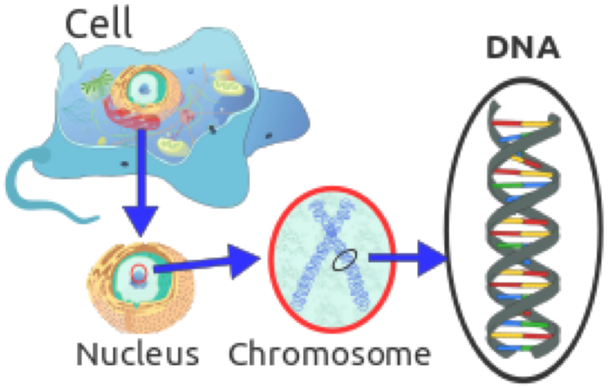
http://bit.ly/reinvent2018-wps201

Create resources with CloudFormation




AWS Cloud
CloudFormation Stack
AWS Identity and Access Management
Batch Service Role
Batch Instance Role
SageMaker Execution Role
AWS Batch
Compute Environment
Job Queue
Batch Job
Job Container
Job Spot Instances
Public Dataset
AWS CloudFormation
Amazon S3
Bucket
Object
Amazon SageMaker
Notebook Instance

Data background



atgatcggtgctg**a**tcgctgtgtcgc... reference

atgatcggtgctg**t**tcgctgtgtcgc... sample

Single Nucleotide Polymorphism (SNP)

samples..

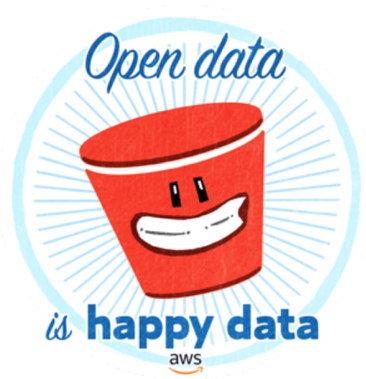
Genomic analysis pipeline

.vcf

Variant Call File

CHR	POS	REF	ALT	SAMPLES..
1	12345	G	G	0/0
1	12678	T	A	1/0

Data preparation



.vcf

Chr1-22

For each Chr:

get first 100 SNPs

AWS Batch

.csv

Chr1-22

.csv

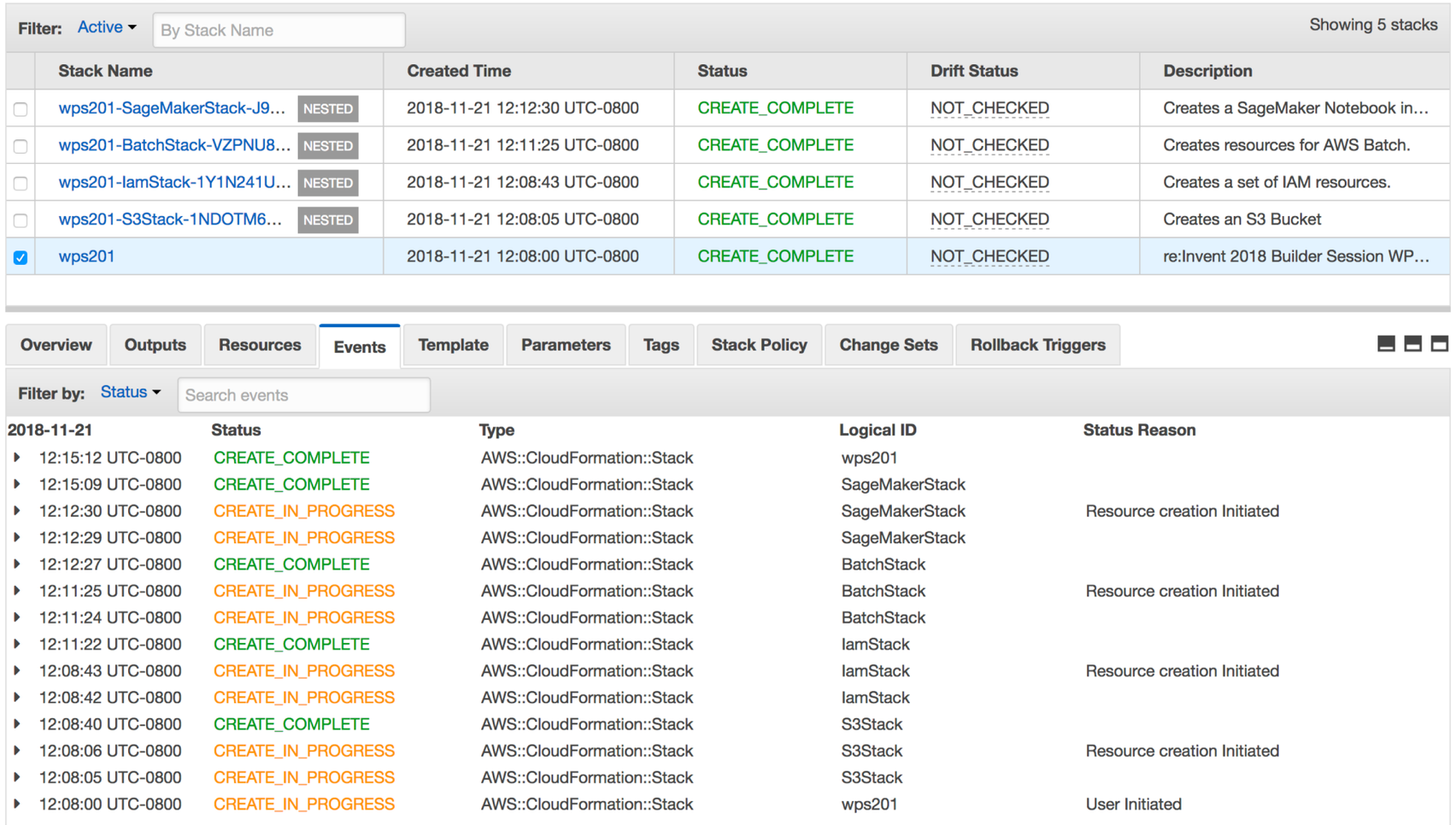
Pop. Labels

SampleID	POP	CHR1-SNP1	CHR-SNP2	…
HG00123	GBR	1	0	…
HG00124	CHB	2	1	…

Amazon SageMaker

http://bit.ly/reinvent2018-wps201

Create resources with CloudFormation



Filter: Active — By Stack Name — Showing 5 stacks

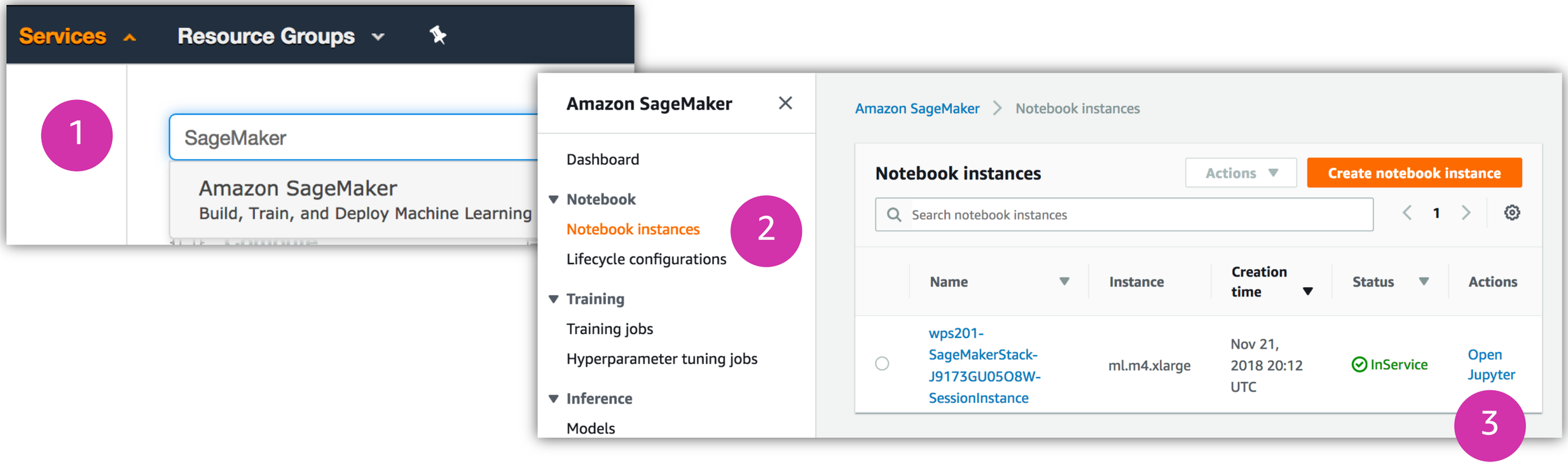
Stack Name	Created Time	Status	Drift Status	Description
wps201-SageMakerStack-J9… NESTED	2018-11-21 12:12:30 UTC-0800	CREATE_COMPLETE	NOT_CHECKED	Creates a SageMaker Notebook in…
wps201-BatchStack-VZPNU8… NESTED	2018-11-21 12:11:25 UTC-0800	CREATE_COMPLETE	NOT_CHECKED	Creates resources for AWS Batch.
wps201-IamStack-1Y1N241U… NESTED	2018-11-21 12:08:43 UTC-0800	CREATE_COMPLETE	NOT_CHECKED	Creates a set of IAM resources.
wps201-S3Stack-1NDOTM6… NESTED	2018-11-21 12:08:05 UTC-0800	CREATE_COMPLETE	NOT_CHECKED	Creates an S3 Bucket
wps201	2018-11-21 12:08:00 UTC-0800	CREATE_COMPLETE	NOT_CHECKED	re:Invent 2018 Builder Session WP…

Overview | Outputs | Resources | Events | Template | Parameters | Tags | Stack Policy | Change Sets | Rollback Triggers

Filter by: Status — Search events

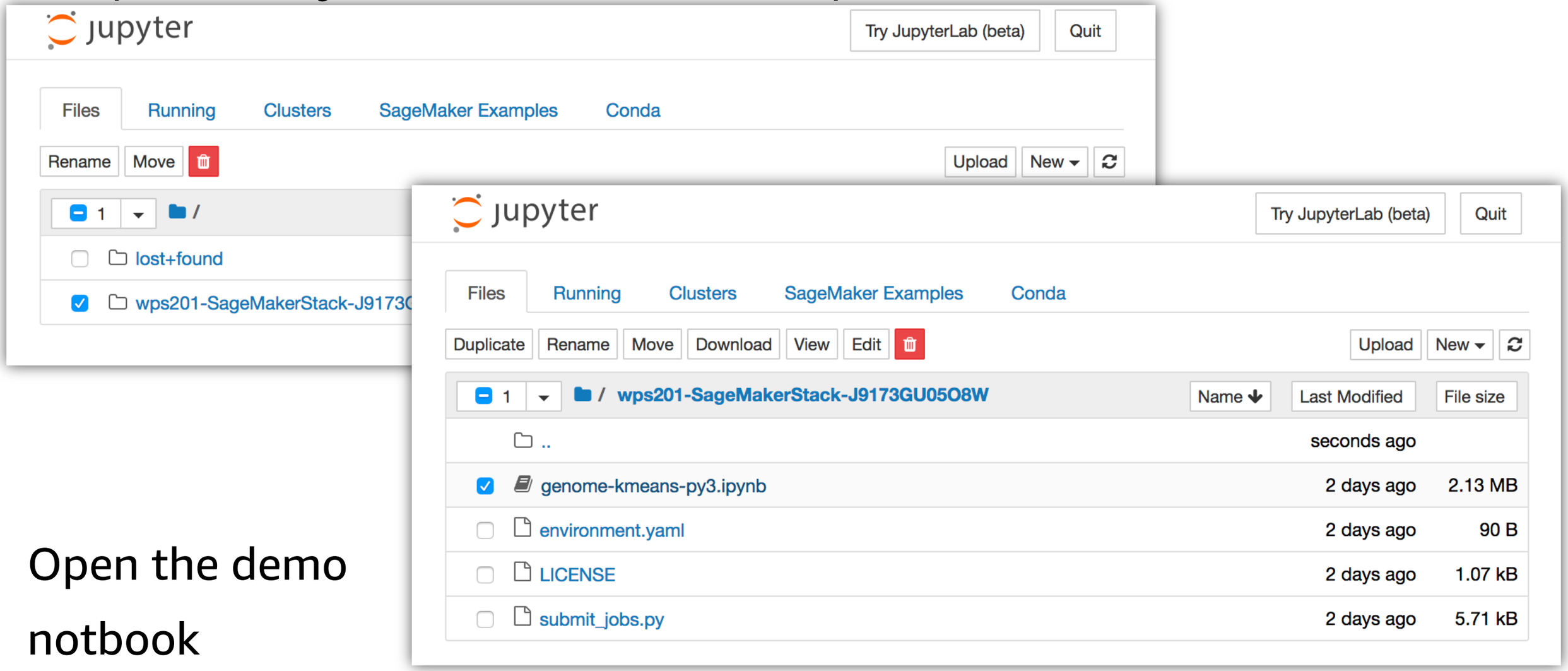
2018-11-21	Status	Type	Logical ID	Status Reason
12:15:12 UTC-0800	CREATE_COMPLETE	AWS::CloudFormation::Stack	wps201	
12:15:09 UTC-0800	CREATE_COMPLETE	AWS::CloudFormation::Stack	SageMakerStack	
12:12:30 UTC-0800	CREATE_IN_PROGRESS	AWS::CloudFormation::Stack	SageMakerStack	Resource creation Initiated
12:12:29 UTC-0800	CREATE_IN_PROGRESS	AWS::CloudFormation::Stack	SageMakerStack	
12:12:27 UTC-0800	CREATE_COMPLETE	AWS::CloudFormation::Stack	BatchStack	
12:11:25 UTC-0800	CREATE_IN_PROGRESS	AWS::CloudFormation::Stack	BatchStack	Resource creation Initiated
12:11:24 UTC-0800	CREATE_IN_PROGRESS	AWS::CloudFormation::Stack	BatchStack	
12:11:22 UTC-0800	CREATE_COMPLETE	AWS::CloudFormation::Stack	IamStack	
12:08:43 UTC-0800	CREATE_IN_PROGRESS	AWS::CloudFormation::Stack	IamStack	Resource creation Initiated
12:08:42 UTC-0800	CREATE_IN_PROGRESS	AWS::CloudFormation::Stack	IamStack	
12:08:40 UTC-0800	CREATE_COMPLETE	AWS::CloudFormation::Stack	S3Stack	
12:08:06 UTC-0800	CREATE_IN_PROGRESS	AWS::CloudFormation::Stack	S3Stack	Resource creation Initiated
12:08:05 UTC-0800	CREATE_IN_PROGRESS	AWS::CloudFormation::Stack	S3Stack	
12:08:00 UTC-0800	CREATE_IN_PROGRESS	AWS::CloudFormation::Stack	wps201	User Initiated

http://bit.ly/reinvent2018-wps201



Open a Sagemaker notebook instance

http://bit.ly/reinvent2018-wps201

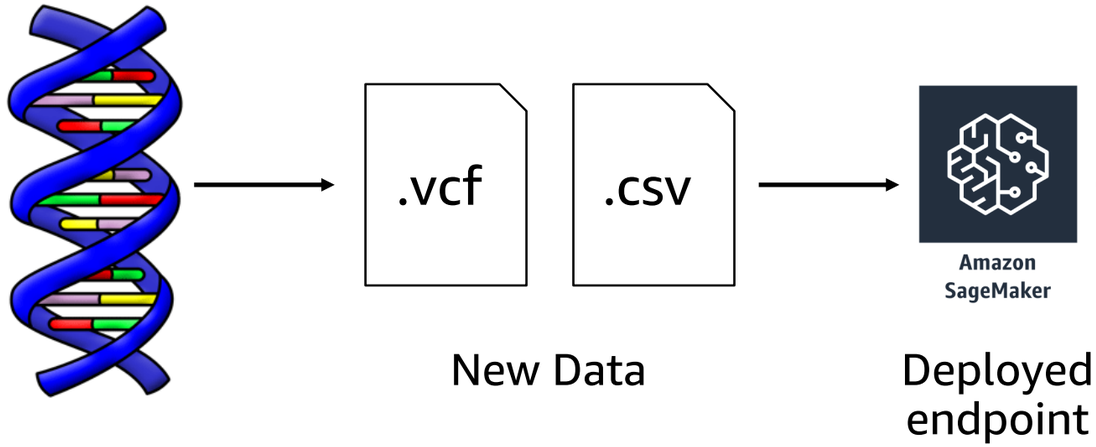


Open the demo notbook

Now what?

AWS re:Invent

aws



Where this DNA sample came from:

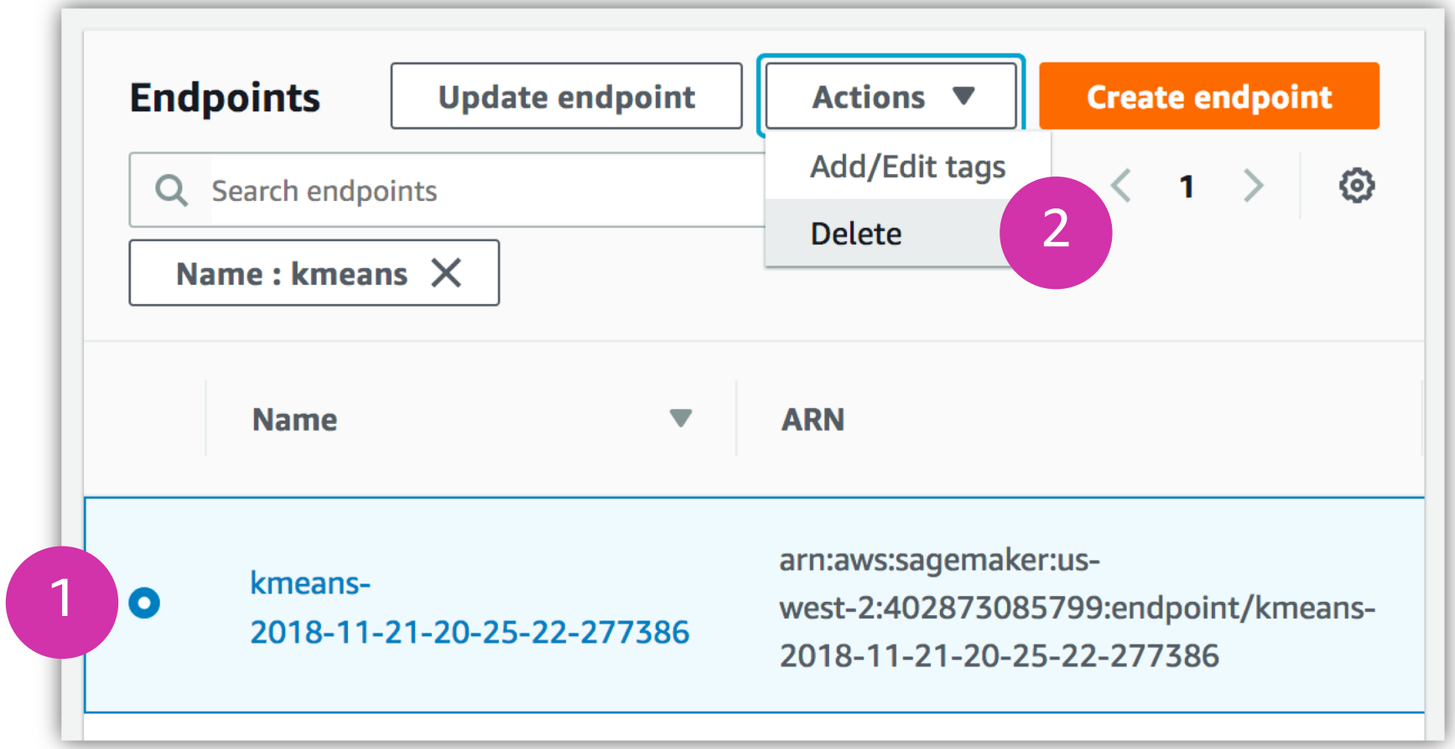
- **Geographic location**
- Cancer tumor type
- Species of origin

Clean-up!

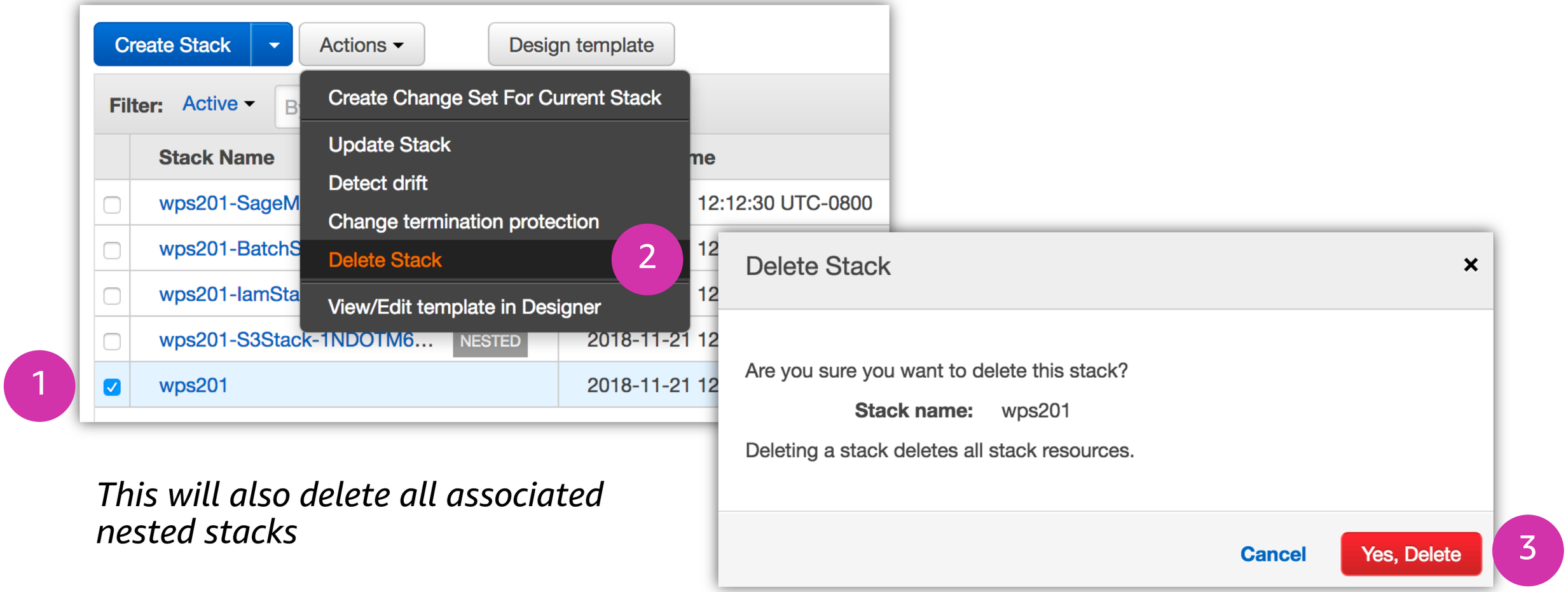
AWS re:Invent

aws

Clean-up: delete Sagemaker Endpoints

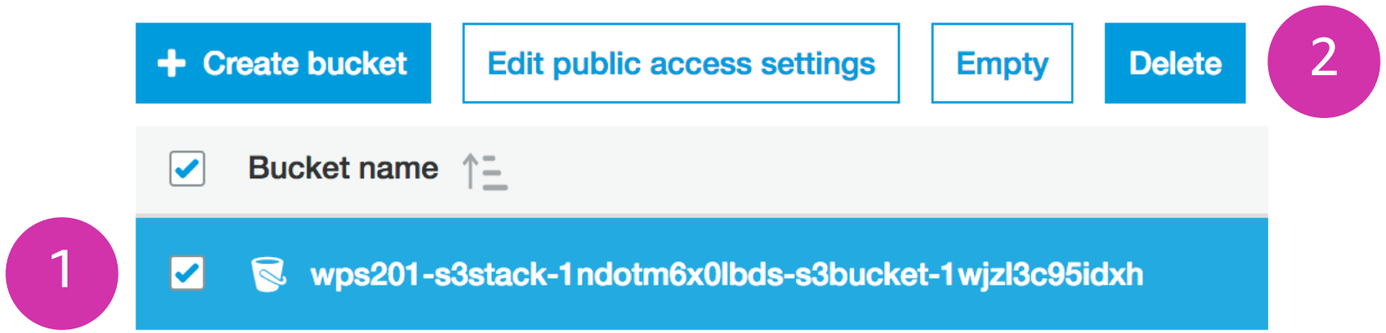


Clean-up: delete CloudFormation stack



This will also delete all associated nested stacks

Clean-up: delete S3 bucket



Thank you!

AWS re:Invent

aws