
AWS
re:Invent

OPN307

Pronto: An IaC suite for managing Cassandra at scale

Achal Kumar
Dev Manager
Intuit Persistence Services

Ben Covi
Staff Software Engineer
Intuit Persistence Services

Aaron Elman
Principal Account Executive
AWS

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Intuit, AWS, and open source

Intuit migrated its business to AWS to accelerate innovation – and open source is key to that innovation

AWS advocates for and partners with our customers in using, growing, and innovating on open source

Open source and cloud, together, are the future



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.



INTUIT powering prosperity

turbotax quickbooks proconnect mint

Who we serve:

Consumers
Small businesses
Self-employed

What do we do?

We provide multi-tenant data persistence as a service

We provide easy-to-use client SDKs, logical data modeling, encryption, access control, and polyglot persistence to our clients

For our fully managed NoSQL service, we use Cassandra as a metadata store and Amazon S3 as our document store

We are the data store supporting TurboTax



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why are we here? Our story

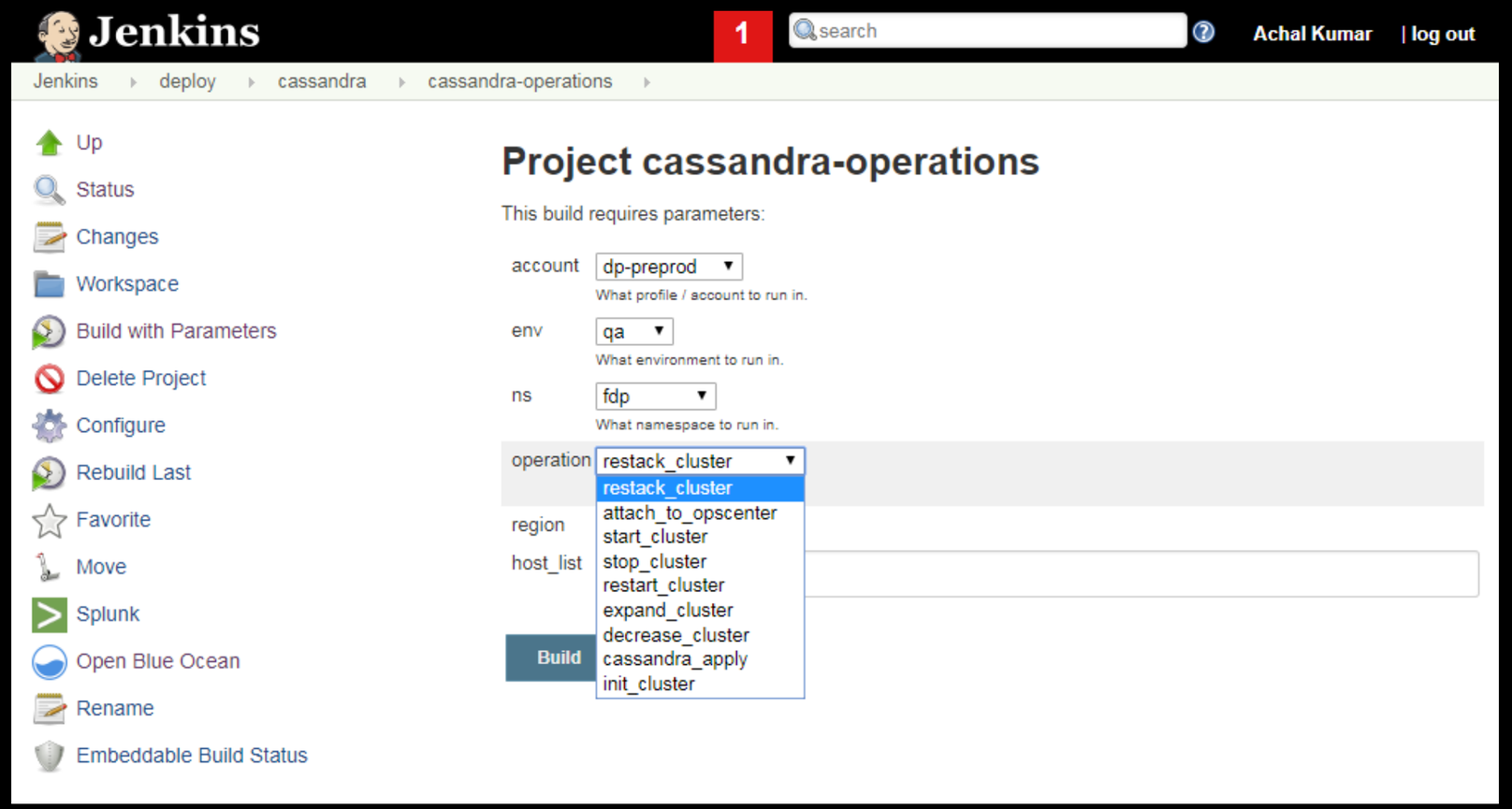
Operational struggle

- Our installation and configuration were manual and prone to human errors
- Dependent on a single Cassandra expert to perform DBA activities
- Slow pace of upgrades

Basic principles we set

- Automate everything about Cassandra
- Nodes going down should be part of life in the cloud
- Developers should be automating and managing it

Operational excellence



intuit

Pronto: What and how



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What is Pronto?

And why should you care?

- Suite of tools to deploy and maintain a Cassandra cluster in the cloud
- Best practices simplified with scripting (bash and python)
- Infrastructure as code (IaC)

What is Cassandra?

And why is it such a challenge?

- Distributed database
- Nodes operate as a cluster
- With great power comes great complexity



"Cluster" by Bruno Cordioli (CC BY 2.0)

How did Pronto happen?

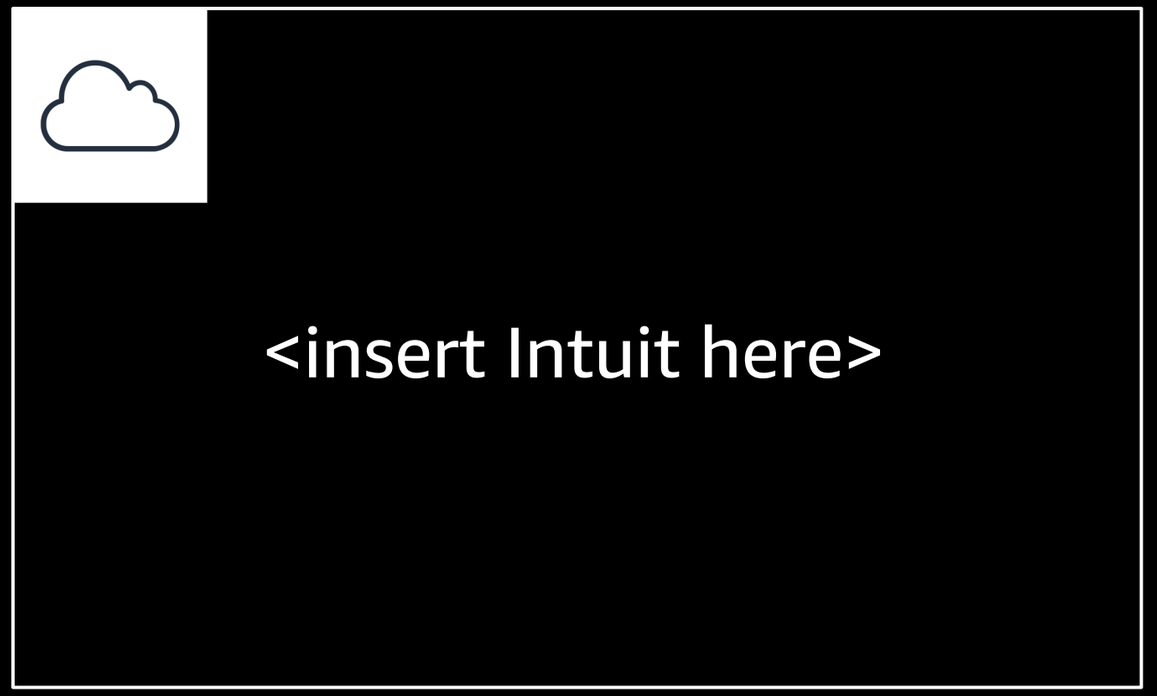
How to accidentally start an open source community

- Problems solved
- Patterns identified
- Audience attracted

Problem #1

We’re moving to the cloud

Is that really a problem?



While you're at it . . .

We'd like it to be:

- Faster
- Cheaper (?)
- More resilient
- Easier to maintain
- Deployable within hours (commit-to-production)
- Soon

Applications in the cloud



"Clouds" by Daniel Spiess (CC BY 2.0)

intuit

Cassandra in the Cloud

Fully managed:







"We customize and optimize the configuration
of your cluster so you can focus on your applications."
-Instaclustr

Cassandra in the cloud

Self-managed:

- Do it yourself



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Managing your own Cassandra cluster

First, decide what your standards are:

- Multi-Region, multi-AZ
- Data subnets
- Orchestrate from a build platform

Managing your own Cassandra cluster

Then, choose your tools:

- Configuration management
  - Old: Chef
  - New: Packer
- Provisioning
  - Old: Racks of hardware
  - New: Terraform
- Runtime operations
  - Old: Human operators
  - New: Ansible

Why Terraform?

- Separates the planning and execution phases
- Notices and handles configuration drift
- Does not attempt rollback on failure
- Cloud agnostic
- Open source and has a huge community
- Feels like programming

The DevOps divide

Dev

Likes:

- Abstractions
- Code re-use



"Railroad" by Julian Carvajal (CC BY 2.0)

Ops

Likes:

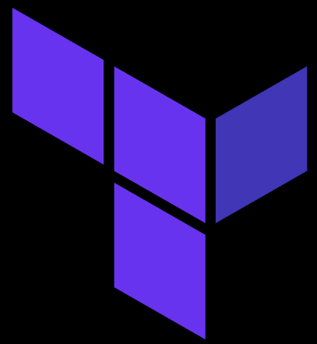
- Procedures
- Sleeping 1 hour / night

intuit

A quick review

This is how we’re going to operate in the cloud:



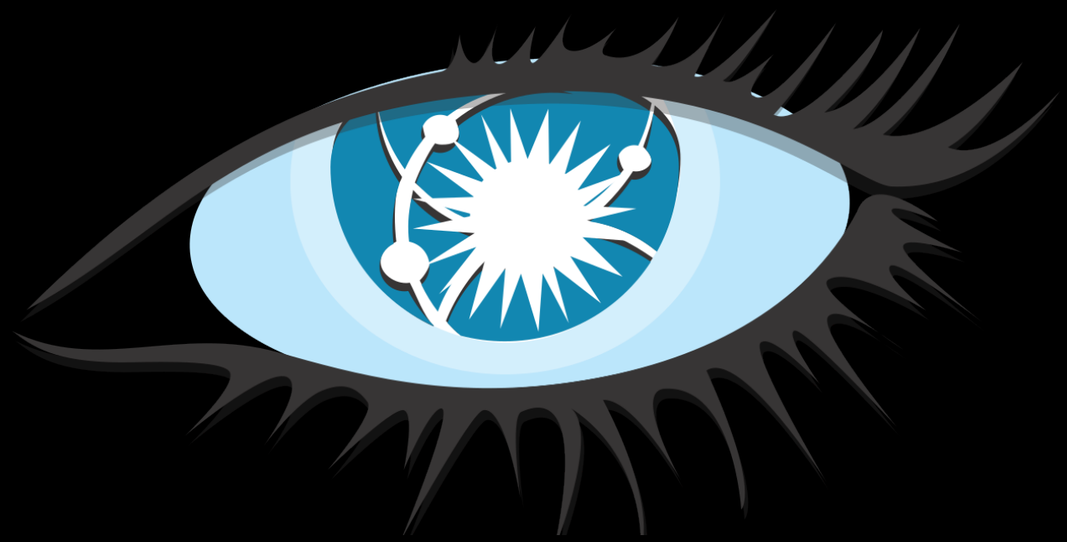




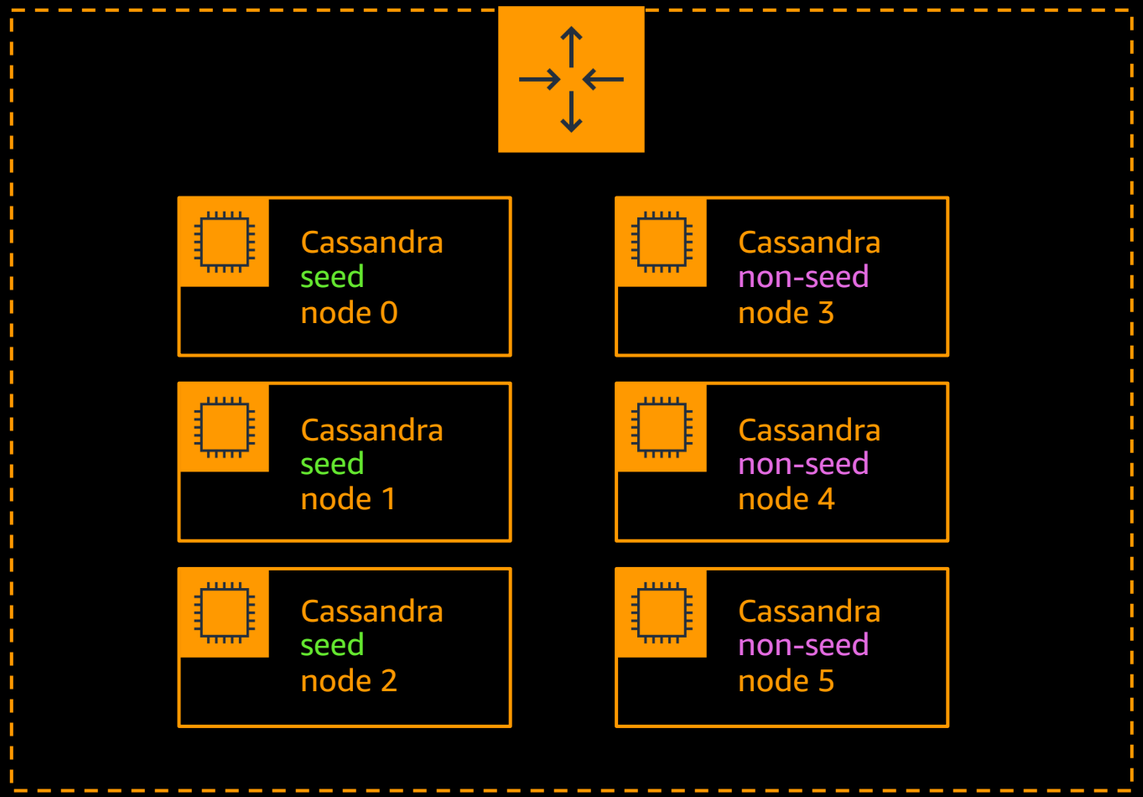
Problem #2

Stateful databases are hard

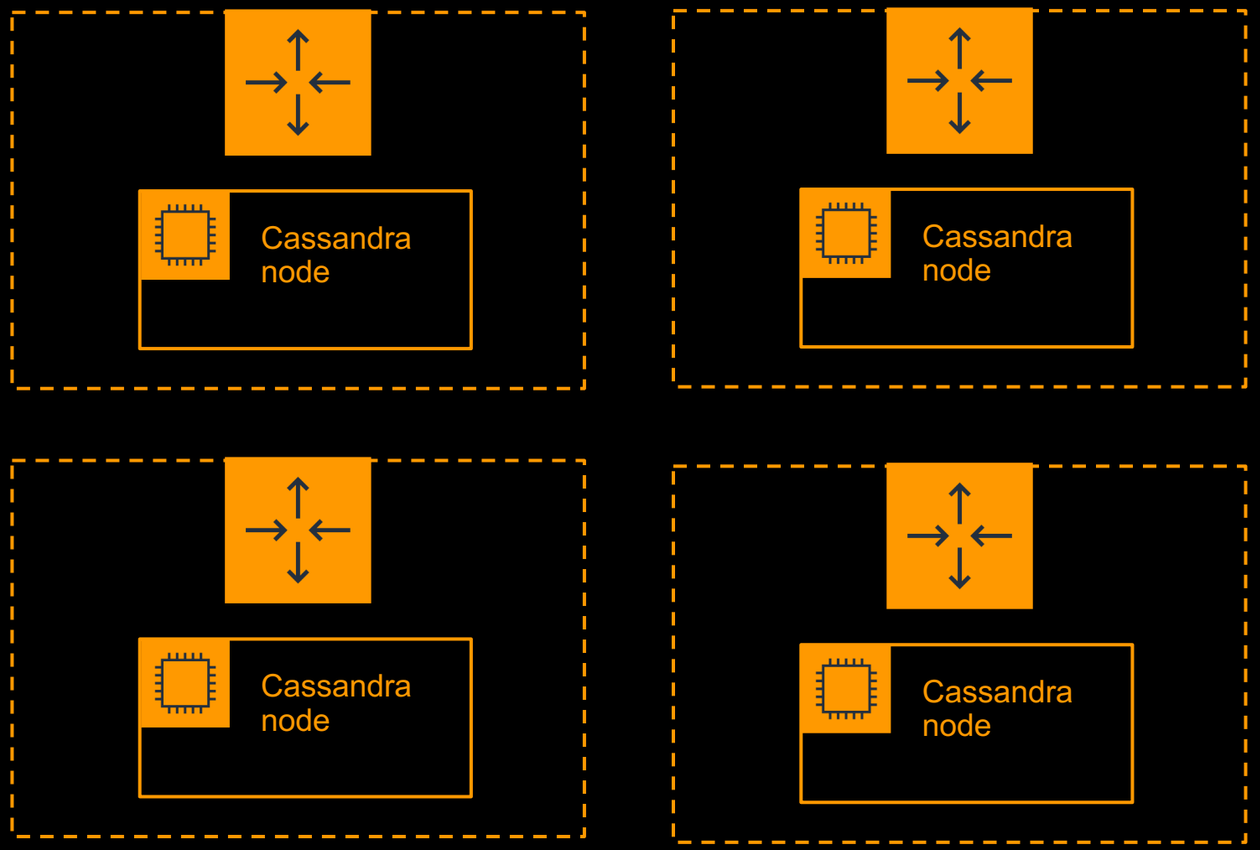
They don’t lend themselves to an elastic, cloud-based environment



The way it doesn’t work



But we can still auto scale, sort of



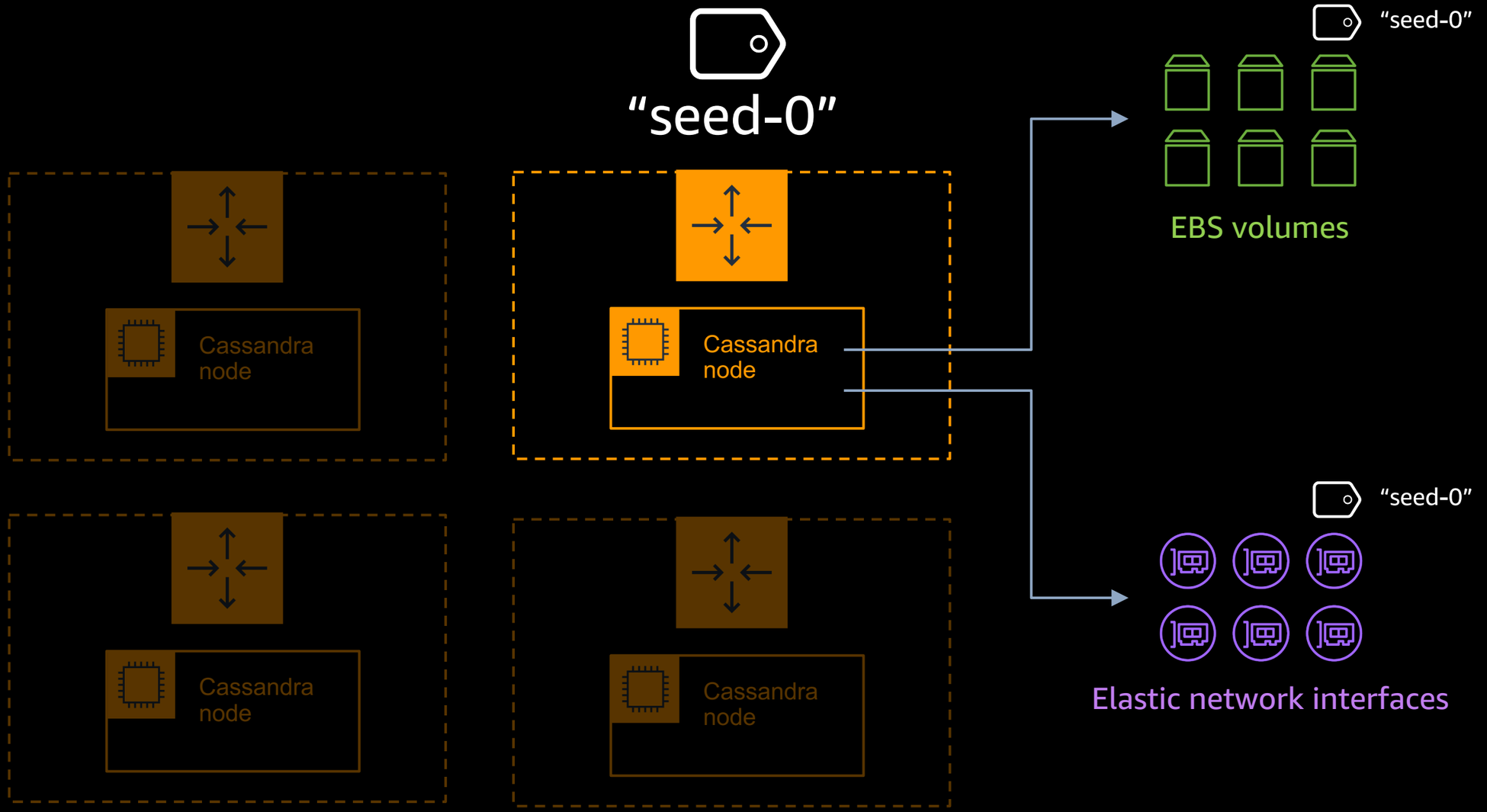
Min: 1

Max: 2

Desired capacity: 1

Scaling policy: n/a

But we can still auto scale, sort of



intuit

Another quick review

Packer



AMI

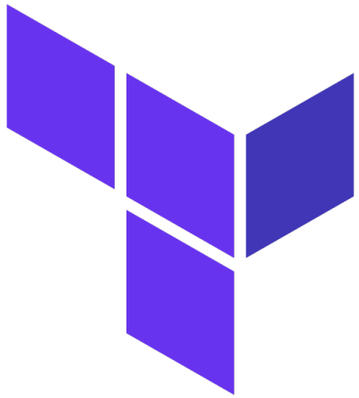
Cassandra



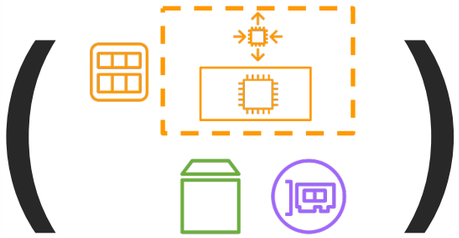
Python (boto3)



Terraform



Resources



x N

Ansible



Configuration management

- Bake a golden image
- Include boto3 scripts

Provisioning

- Deploy the golden image
- Deploy everything else too

Runtime ops

- Start / stop
- Restack the cluster

Problem #3

It exists – now other people want to use it



"Marrakesh" by Miguel Discart (CC BY 2.0)

intuit

Educating new users



Educating new users



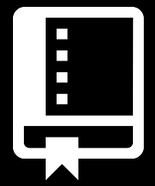
- aws-definitions
- aws-topology
- aws-topology-modules
- cassandra-utils
- df_pipeline
- utilities
- chef
- data-fabric-cookbooks

Consolidate, simplify, improve

- Revisit (relearn) the code and what it does
- Evaluate the code's sanity level (do not take sanity for granted)
- Refactor as necessary

Consolidate, simplify, improve

- Revisit (relearn) the code and what it does
- Evaluate the code's sanity level (do not take sanity for granted)
- Refactor as necessary



github.com/intuit/dse-pronto



Welcoming contributions

- Innersourcing works
- Maybe this open source thing works too
- Let’s build it together:
  - Support for tiered storage
  - Support for vanilla Apache Cassandra
  - Support for other cloud providers
  - ?

Thank you!

Achal Kumar

achal_kumar@intuit.com
linkedin.com/in/kumarachal

Ben Covi

ben_covi@intuit.com
linkedin.com/in/bencovi

Aaron Elman

elman@amazon.com

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Please complete the session survey

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.