Geoscience in the Cloud

Dr Carina Kemp
Director, Scientific Data

Co-authors: Laura Stanford, James Kingsmill & Tom Butler
The Cloud Enablement Team

Geoscience Australia
Geoscience Australia (GA)
Nation’s trusted advisor on the geology and geography of Australia
Geoscience Australia (GA)

The GA Science Principles:

- Relevance to Government
- Collaborative science
- Quality science
- Transparent science
- Communicated science
- Sustained science capability

‘A Framework for eResearch to enable Science’
Digital Science is driving change at GA

• Scientific information is expected to be available on-demand.
• Scientific questions are becoming increasingly complex.
• Scientific data is growing exponentially.
• Digital technologies have reached unprecedented capabilities.
1. Access the right mix of skills
2. Establish an agile way of working
3. Improve planning and program delivery processes
4. Establish a quantitative science platform
5. Formalise Digital Science leadership
Cloud: the journey so far

On premise:

• Limited scalability
• Vanilla
• Specialised teams

Cloud provider:

• Scalable
• Reliable
• Customisable
• Quick time to production
Cloud Enablement Team (Autobots)

James Kingsmill, Tom Butler & Laura Stanford

- Collaboration on projects
- Technical advice and workshops
- Automation and patterns
- Cloud governance
- Support GA developer teams and Scientists in the Science Divisions
Challenges…

Amazon EC2

"I think you should be more explicit here in step two."
Infracode: Infrastructure as Code

- Source control
- Configurations
- Divide into reusable modules
- Infrastructure patterns & examples
- Code reviews
- Automated testing
- Continuous delivery and deployment
AMI: Amazon Module Interface
Server Patterns

Example builds for common usages

Security heavily baked into AMIs.

Automated tests

<table>
<thead>
<tr>
<th>Repository</th>
</tr>
</thead>
<tbody>
<tr>
<td>buildbox</td>
</tr>
<tr>
<td>codedeploy</td>
</tr>
<tr>
<td>Docker</td>
</tr>
<tr>
<td>jumpbox</td>
</tr>
<tr>
<td>shibboleth</td>
</tr>
<tr>
<td>simple-webserver</td>
</tr>
<tr>
<td>tomcat-packer</td>
</tr>
</tbody>
</table>

```
"variables": {
"aws_access_key": "{{env 'AWS_ACCESS_KEY'}}",
"aws_secret_key": "{{env 'AWS_SECRET_KEY'}}"
},

"builders": {
  "type": "amazon-ec2",
  "region": "ap-southeast-3",
  "source_ami": "ami-4d3b862e",
  "instance_type": "t2.medium",
  "ssh_username": "ubuntu",
  "iam_instance_profile": "fuseki-jena",
  "ami_name": "FUSEKI/AM Server {{timestamp}}",
  "associate_public_ip_address": true,
  "tags": {
    "application": "fuseki-packer",
    "version": "current"
  }
},

"provisioners": {
  "type": "file",
  "source": "config",
  "destination": "/tmp"
},

"type": "shell",
"script": "scripts/pre-install.sh"
```

HashiCorp

Packer

Geoscience in the Cloud - Kemp et al
Terraform
Infrastructure as code:

• Codifies APIs into declarative configuration files
• Modular
• Platform agnostic
• Allows state locking

Example architecture:

```hcl
module "public" {
  source = "modules/public_layer"
  region = "${var.region}"
  availability_zones = "${var.availability_zones}"
  stack_name = "${var.stack_name}"
  environment = "${var.environment}" 
  owner = "${var.owner}"
  key_name = "${var.key_name}"
  enable_jumpbox = true
  nat_gw_count = 1
  vpc_id = "${module.shared.vpc_id}"
  igw_id = "${module.shared.igw_id}" 
  listeners {
    "instance_port" = 80
    "instance_protocol" = "HTTP"
    "lb_port" = 80
    "lb_protocol" = "HTTP"
  }
}
```

Available at: https://bitbucket.org/account/user/geoscienceaustralia/projects/TF
Bitbucket pipelines

Pipeline #25

Build

- Build setup
- export TF_VAR_environment=test
- terraform init -backend-config="key=$TF_VAR_stack_name=$TF_VAR_env...
- terraform apply
- cd tests
- bundle install
- bundle exec rake spec
- cd ../
- terraform destroy -force

Logs

18 min 25 sec • a month ago
geoscienceaustralia/autobots-terraform
Push by Tom Butler
View configuration

Commit

- 90795a2
  ASG Health check bugfix, ASG was setting health check to be EC2 which resulted in some undesirable effects thanks - @scelton
- integration

Download raw
Monitoring & Logging

• Cloud Trail
• Account logs
• Security logs
• CloudWatch

Security

• AMIs have security baked in
• Security well architected into our examples
• SSO functionality
• Central account access – Organisations
Cost Management

• Cloudability – to monitor costs on AWS
  • Views for each area
  • Allows quick responses to issues
  • Lambda Scripts:
    • To shut down dev/test servers overnight and on weekends
  • Reserved instances used
    • RDS
    • EC2
Exemplars
AUSCORS NTRIP Broadcaster

Geoscience Australia (GA) provides 1 Hz data streaming from our Global Navigation Satellite System (GNSS) stations throughout Australia, Antarctica and the Pacific. The data is distributed via the AUSCORS Ntrip Broadcaster. Access to the data is free, however a username and password is required. The real-time data is made available through an RTCM standard transmitted over the Internet using the NTRIP (Network Transport of RTCM via Internet Protocol).

To connect you will need:
- Username and password,
- NTRIP client and internet access,
- Connection details.

Register for an Account

Connection Details

Host Domain: auscors.ga.gov.au
Port: 2101

Mountpoint Information

Legend:
- Red: NTRIP
- Green: Other GNSS
- Orange: Other Streaming

Australian Government
Geoscience Australia
All the good things: empowerment

• Internal:
  • Empowering the Scientists
  • Empowering the Developers
  • Upskilling
  • Continuous Improvement

• External
  • Enabling access to GA’s data
  • Meeting our stakeholder needs
Thank you

The Cloud Enablement Team

Laura.Stanford@ga.gov.au
James.Kingsmill@ga.gov.au
Tom.Butler@ga.gov.au

Available at: https://bitbucket.org/account/user/geoscienceaustralia/projects/TF

Phone: +61 2 6249 9228
Web: www.ga.gov.au
Email: carina.kemp@ga.gov.au
Address: Cnr Jerrabomberra Avenue and Hindmarsh Drive, Symonston ACT 2609
Postal Address: GPO Box 378, Canberra ACT 2601