Building Open Source platforms on AWS

Julien Simon
Principal Technical Evangelist
Amazon Web Services

julsimon@amazon.fr
@julsimon

amazon web services

Agenda

- VMs & OSes
- Language SDKs
- Docker
- Development infrastructure
- Databases
- Analytics
- Conclusion

Virtual Machines & Operating Systems

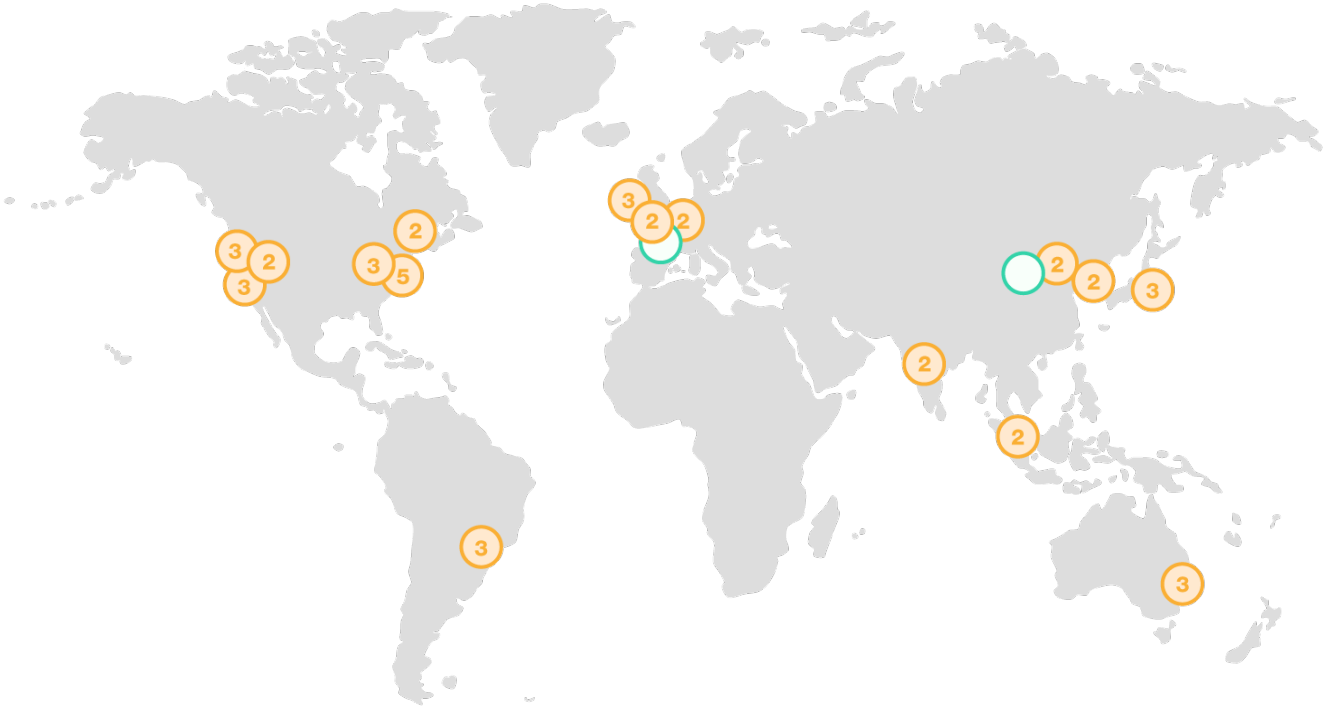
amazon web services

AWS Global Infrastructure

16 Regions

42 Availability Zones

73 Edge Locations



amazon web services

Amazon EC2 AWS Free Tier

- Infrastructure as a Service, launched in 2006
- Virtual machines (EC2 instances) and images (Amazon Machine Images)
- AMIs: Amazon, 3rd party (EC2 Marketplace), community or your own

- All-inclusive: networking (Virtual Private Cloud), storage (Elastic Block Storage), firewalling (Security Group), load balancing (Elastic Load Balancing), high availability (Availability Zones), automatic scaling (Auto Scaling groups), monitoring (Cloudwatch)

- Pay on an hourly basis
- Or use Reserved Instances and Spot Instances for large savings

https://aws.amazon.com/ec2/
http://aws.amazon.com/free/
https://aws.amazon.com/ec2/pricing/reserved-instances/
https://aws.amazon.com/ec2/spot/

amazon web services

Instances Types

<*Family*><*Generation*>.<*Size*>, e.g. m4.xlarge

General purpose: t2 (burstable), m4
Compute-optimized: c4
Storage-optimized: i3 (I/O), d2 (Density)
Memory-optimized: r4, x1
GPU: g2, p2
FPGA: f1

t2.nano: 1 vCPU, 512MB RAM, EBS storage
x1.32xlarge: 128 vCPU, 2TB RAM, 4TB SSD, 10Gb network

https://aws.amazon.com/ec2/instance-types/

Amazon Linux

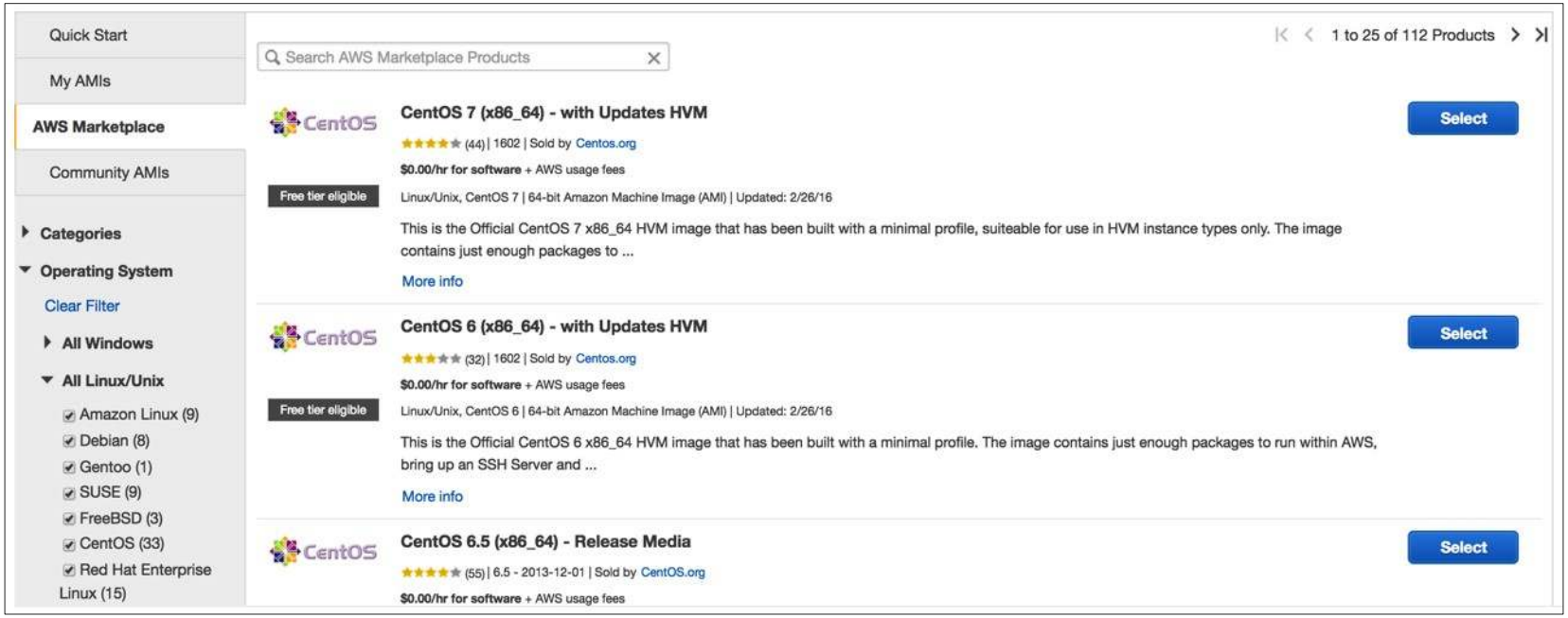
- Evolved from CentOS
- Secure configuration
- Pre-installed AWS tools
- Integrated with AWS repositories & security updates
- Available as Docker container
- Easy incremental updates

  ```
  yum clean all && yum update && reboot
  ```

- Latest version: Amazon Linux 2017.03 (kernel 4.9.x)

https://aws.amazon.com/amazon-linux-ami/
https://aws.amazon.com/amazon-linux-ami/2017.03-release-notes/ https://docs.aws.amazon.com/AmazonECR/latest/userguide/amazon_linux_container_image.html

100+ Linux/BSD AMIs



FreeBSD

NetBSD

OpenBSD

CentOS debian redhat SUSE ubuntu

We adapt. You succeed.

https://aws.amazon.com/marketplace/

amazon web services

Baking your own AMI

- Create a golden image to speed up deployment and Auto Scaling
- AWS CLI: `aws ec2 create-image`
- Aminator: Netflix tool, EC2 only for Red Hat and CentOS
- Packer: Hashicorp tool, more features

http://docs.aws.amazon.com/cli/latest/reference/ec2/create-image.html
https://github.com/Netflix/aminator
https://www.packer.io

Software Development Kits

amazon web services

10 programming environments

- Android
  https://github.com/aws/aws-sdk-android
- C++
  https://github.com/aws/aws-sdk-cpp
- Go
  https://github.com/aws/aws-sdk-go
- iOS
  https://github.com/aws/aws-sdk-ios
- Java
  https://github.com/aws/aws-sdk-java
- Javascript / Node.js
  https://github.com/aws/aws-sdk-js
- .NET
  https://github.com/aws/aws-sdk-net
- Python
  https://github.com/boto/boto3
- PHP
  https://github.com/aws/aws-sdk-php
- Ruby
  https://github.com/aws/aws-sdk-ruby

https://github.com/aws/aws-toolkit-eclipse
https://aws.amazon.com/visualstudio/

Docker

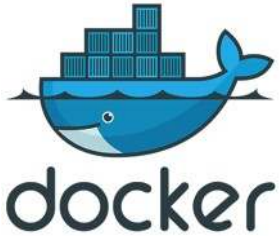
amazon web services

Running Docker on AWS

- Use docker-machine to start EC2 instances
- Start Docker-enabled AMIs
- Build Docker clusters
- Use Docker-based PaaS platforms

https://convox.com/blog/docker-in-production/

Docker-machine




```
docker-machine create \
        --driver amazonec2 \
        --amazonec2-region region \
        --amazonec2-zone az \
        --amazonec2-instance-type type \
         instance_name

docker-machine ssh instance_name
```

https://docs.docker.com/machine/drivers/aws/

amazon web services

Docker-enabled AMIs

Available on the AWS Marketplace

RancherOS

Rancher Server runs as a container

CoreOS

CloudFormation template
on CoreOS page

EC2 REGION	AMI TYPE	AMI ID	CLOUDFORMATION
eu-central-1	PV	ami-72867d1d	Launch Stack
	HVM	ami-27877c48	Launch Stack

Amazon ECS-optimized AMI

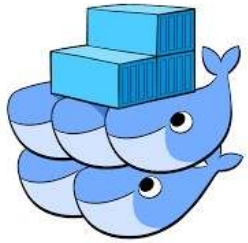
Amazon Linux + Amazon ECS Agent + Docker 1.12.6
More on ECS in a minute ☺

http://docs.rancher.com/os/running-rancheros/cloud/aws/
https://coreos.com/os/docs/latest/booting-on-ec2.html
http://docs.aws.amazon.com/AmazonECS/latest/developerguide/ecs-optimized_AMI.html

amazon web services

Docker orchestrators

Docker Swarm



- Installation with docker-machine
- Docker for AWS

  Create Swarm with CloudFormation
- Docker Datacenter
  - End-to-end platform for container management
  - Ready in 20-30 minutes

http://docs.docker.com/swarm/install-manual/
http://blog.scottlowe.org/2016/03/25/docker-swarm-aws-docker-machine/
https://docs.docker.com/docker-for-aws/ https://aws.amazon.com/about-aws/whats-new/2016/06/docker-datacenter-on-the-aws-cloud-quick-start-reference-deployment/

amazon web services

Kubernetes

kubernetes

```
$ kube-aws init \
--cluster-name=my-cluster-name \
--external-dns-name=my-cluster-endpoint \
--region=region-name \
--availability-zone=az-name \
--key-name=keypair-name \
--kms-key-arn=key-arn
```

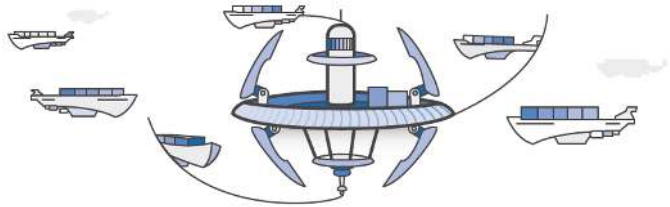
```
$ kube-aws render
```
→ CloudFormation template

```
$ kube-aws up
```

https://coreos.com/kubernetes/docs/latest/kubernetes-on-aws.html

amazon web services

Amazon ECS and ECR AWS Free Tier

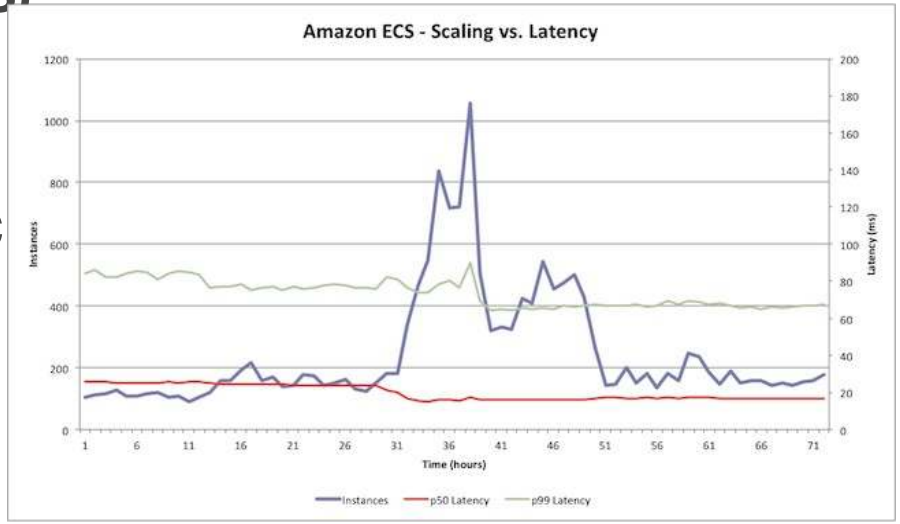


Amazon EC2 Container Service (ECS)

- Orchestration of Docker clusters
- No charge

Amazon EC2 Container Registry (EC

- Managed private Docker Registry
- Free tier available
- $0.10 / GB / month + outgoing traffic



http://www.allthingsdistributed.com/2014/11/amazon-ec2-container-service.html
http://www.allthingsdistributed.com/2015/04/state-management-and-scheduling-with-ecs.html
http://www.allthingsdistributed.com/2015/07/under-the-hood-of-the-amazon-ec2-container-service.html

https://aws.amazon.com/ecs/
https://aws.amazon.com/ecr/

amazon web services

The Amazon ECS CLI in one slide

```
ecs-cli configure --cluster cluster_name --region region_name
ecs-cli up --keypair keypair --capability-iam –size nb_nodes

ecs-cli compose service up
ecs-cli compose service ps
ecs-cli compose service scale nb_containers

ecs-cli compose service stop
ecs-cli compose service delete
ecs-cli down cluster_name --force
```

https://github.com/aws/amazon-ecs-cli

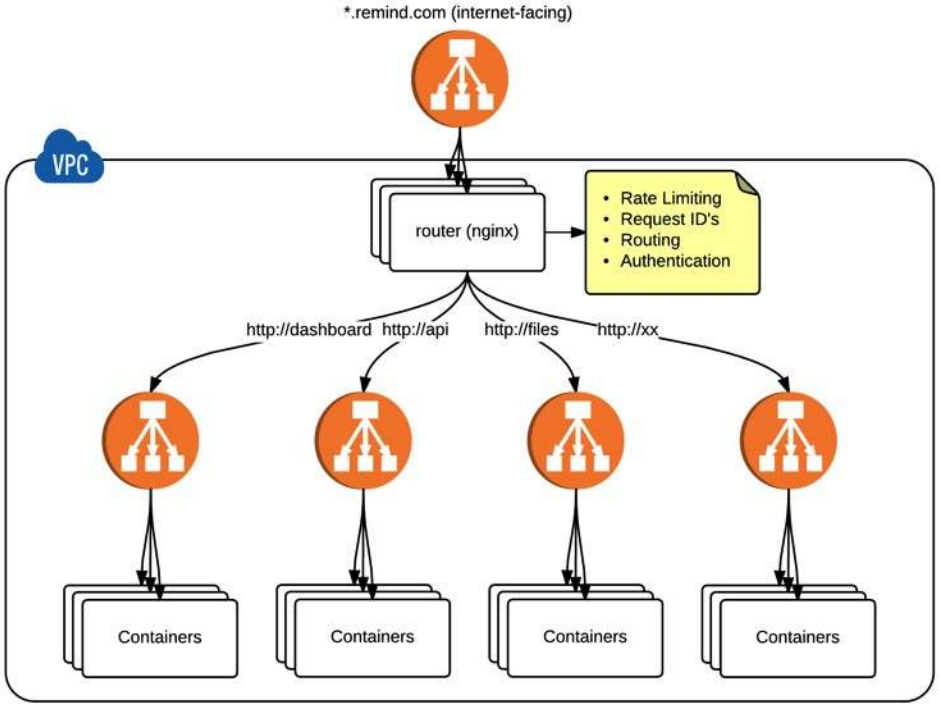
amazon web services

Docker PaaS

amazon web services

Empire

- Simple Open Source PaaS built by Remind
- Based on Amazon ECS
- Well-suited for 12-factor platforms

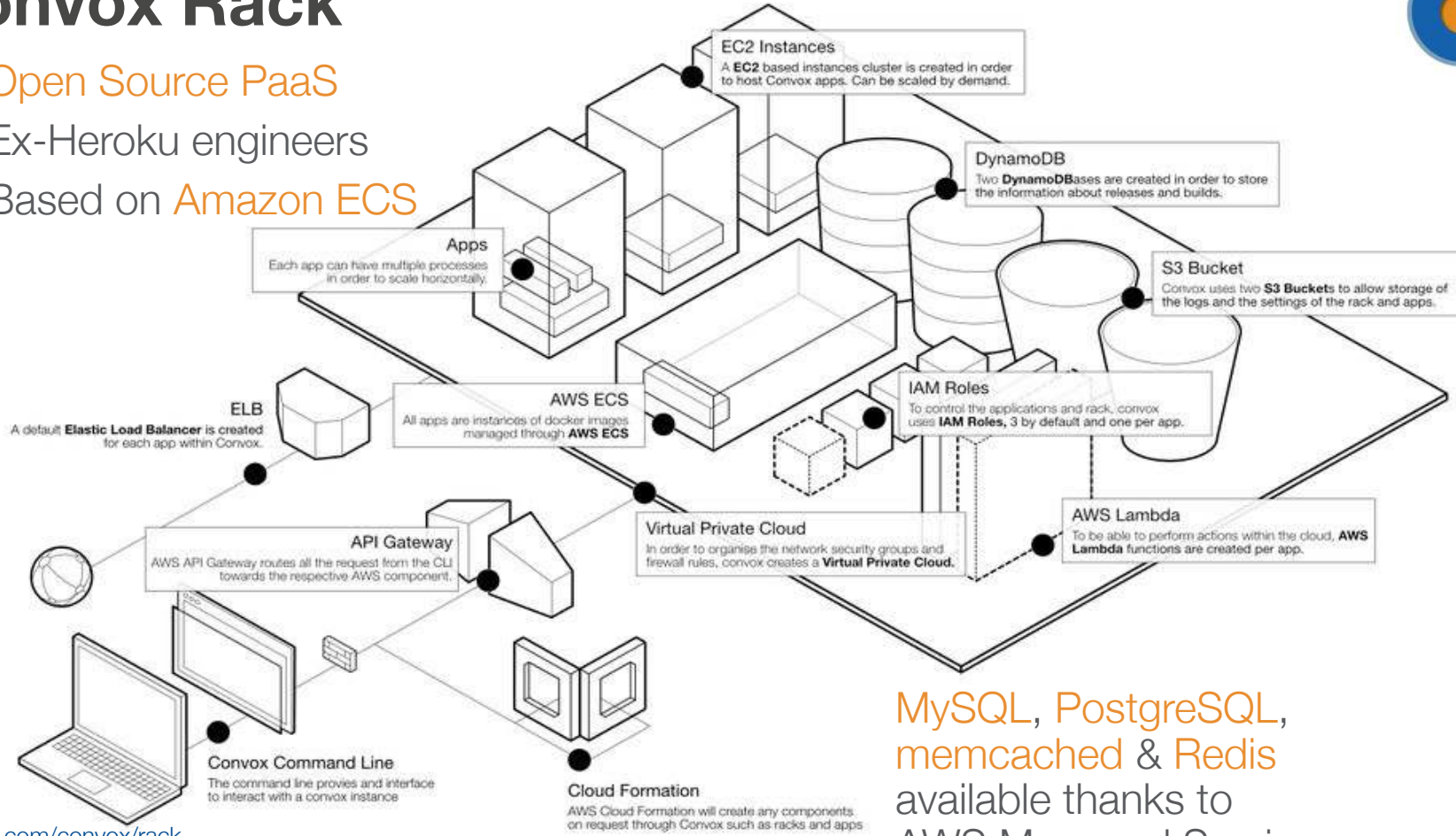


http://engineering.remind.com/introducing-empire/
https://github.com/remind101/empire
https://www.youtube.com/watch?v=8zbbQkszP04
https://12factor.net

amazon web services

Convox Rack

- Open Source PaaS
- Ex-Heroku engineers
- Based on Amazon ECS



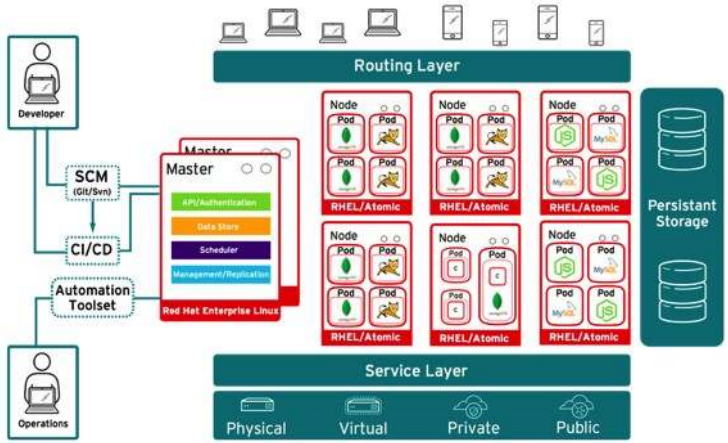
MySQL, PostgreSQL, memcached & Redis available thanks to AWS Managed Services

https://github.com/convox/rack
http://www.slideshare.net/nzoschke/minimum-viable-infrastructure

Openshift

OPENSHIFT

- Open Source PaaS built by Red Hat
- AWS Quick Start reference guide
  - CloudFormation template
- Openshift Dedicated
  - Platform hosted in the AWS cloud



https://github.com/openshift/origin
https://aws.amazon.com/about-aws/whats-new/2016/06/red-hat-openshift-on-the-aws-cloud-quick-start-reference-deployment/
https://www.openshift.com/dedicated/

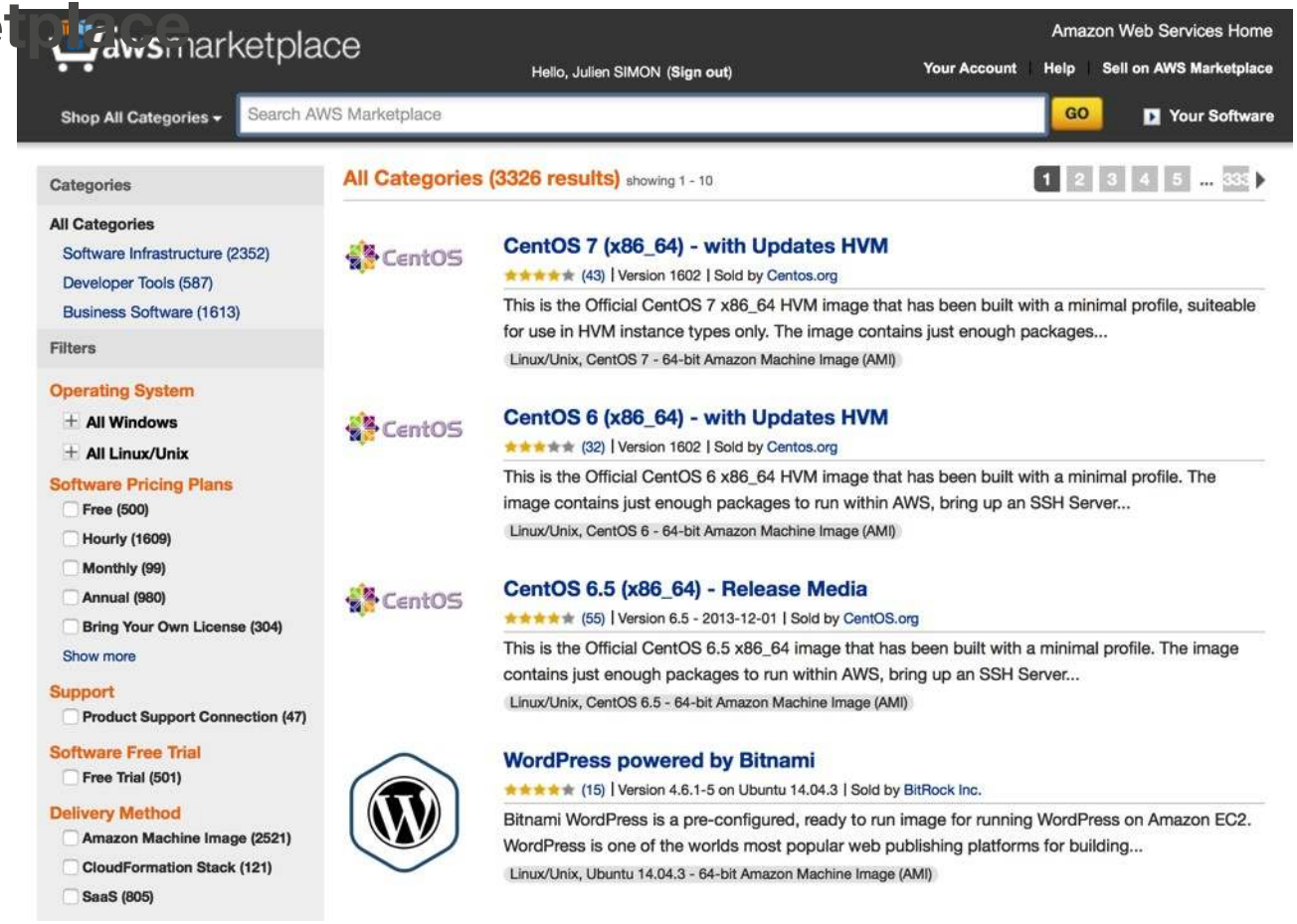
amazon web services

Development Infrastructure

amazon web services

Your CI/CD tools, right?

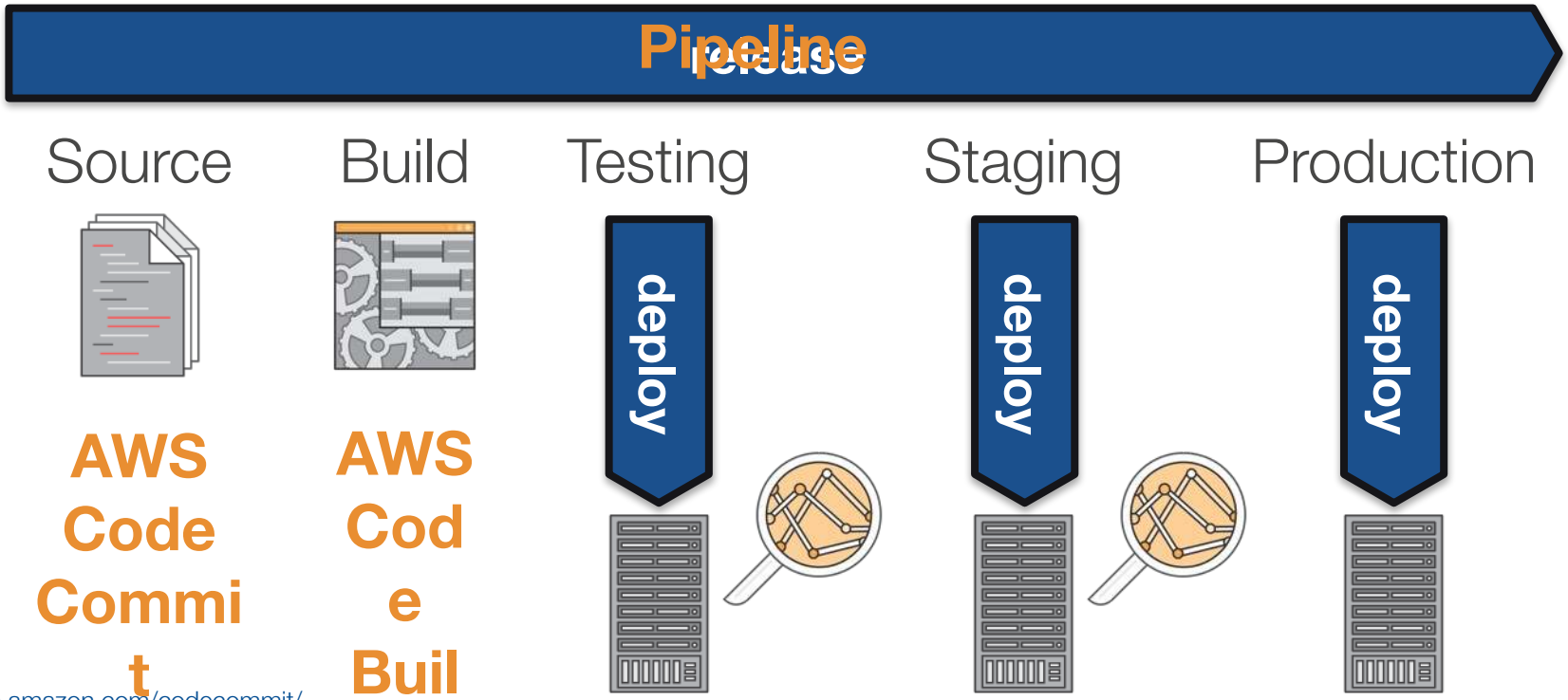
Most of them (and many more) are available on the Marketplace



amazon web services

They all work with our DevOps tools (aka Code*)

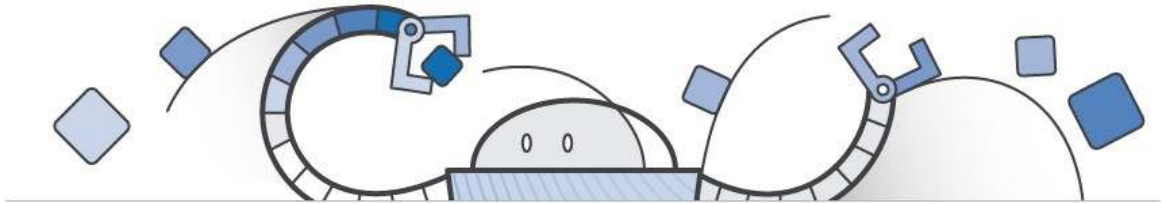
AWS Free Tier



https://aws.amazon.com/codecommit/
https://aws.amazon.com/codepipeline/
https://aws.amazon.com/codedeploy/
https://aws.amazon.com/codebuild/

amazon web services

AWS OpsWorks



- Managed Chef server
  - Chef 12, 11.10, 11.4 and 0.9 for Linux stacks
  - Chef 12.2 for Windows stacks
- Auto Healing
- Automatic Instance Scaling
- Monitoring
- Support for on-premises servers
- Permissions and policy management with IAM
- No charge for EC2

CHEF™

https://aws.amazon.com/opsworks/ http://docs.aws.amazon.com/opsworks/latest/userguide/best-practices-server-migrate.html

amazon web services

Relational Databases

amazon web services

Amazon Relational Database Service (RDS)

AWS Free Tier

- Managed infrastructure
  - SLA >= 99.95% for multi-AZ setups
  - Automatic backups & minor upgrades
  - No access to the database host operating system
- Scalable compute & storage
  - No downtime in most cases
  - Max storage for MySQL, MariaDB, PostgreSQL: 6 TB
- 3 Open Source engines
  - MySQL: 5.5.40a → 5.7.16
  - MariaDB: 10.0.17 → 10.1.19
  - PostgreSQL: 9.3.12-R1 → 9.6.1-R1

MySQL

MariaDB

PostgreSQL

https://aws.amazon.com/rds/
https://aws.amazon.com/releasenotes/Amazon-RDS
https://aws.amazon.com/rds/sla/

amazon web services

Amazon RDS: the small print ☺

- Using the rds_superuser Role
- Supported PostgreSQL Database Versions
- Supported PostgreSQL Features and Extensions
- Limits for PostgreSQL DB Instances
- Upgrading a PostgreSQL DB Instance
- Using SSL with a PostgreSQL DB Instance
- Creating Roles
- Managing PostgreSQL Database Access
- Working with PostgreSQL Parameters
- Working with PostgreSQL Autovacuum on Amazon RDS
- Audit Logging for a PostgreSQL DB Instance
- Setting up PostGIS
- Using pgBadger for Log Analysis with PostgreSQL
- Viewing the Contents of pg_config

http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/CHAP_PostgreSQL.html#PostgreSQL.Concepts.General.FeatureSupport
http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Appendix.PostgreSQL.CommonDBATasks.html

- Killing a Session or Query
- Skipping the Current Replication Error
- Working with InnoDB Tablespaces to Improve Crash Recovery Times
- Managing the Global Status History

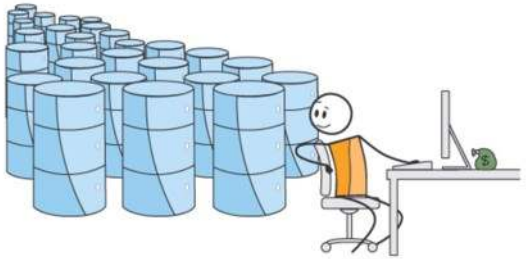
http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Appendix.MySQL.CommonDBATasks.html

Appendix: Parameters for MariaDB

http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Appendix.MariaDB.Parameters.html

Amazon Aurora

fastest growing service in AWS history



- Compatible with MySQL 5.6 (with some 5.7 features)
- Scalable compute & storage (up to 64TB)
- 5x the throughput of MySQL on the same hardware: up to 500K reads and 100K writes per second
- 10 ms latency on up to 15 read replicas
- Availability > 99.99%, failover < 30s
- 6 copies of your data in 3 AZs + continuous backup to S3
- Customers end up using smaller instances and/or less instances than MySQL → Aurora can be less expensive!

https://aws.amazon.com/rds/aurora/
https://d0.awsstatic.com/product-marketing/Aurora/RDS_Aurora_Performance_Assessment_Benchmarking_v1-2.pdf

amazon web services

NoSQL

amazon web services

Memcached & Redis

AWS Free Tier

- Amazon ElastiCache
- Managed service for in-memory data
- Memcached
  - 1.4.5 → 1.4.33
  - Automatic node discovery
- Redis
  - 2.6.13 → 3.2.4
  - Scale up without losing data thanks to replication
  - Scale out: up to 15 shards with 5 read replicas each → 3.5 TB, 20M reads and 4.5M writes per second
  - Backups & restores

https://redislabs.com/blog/5-tips-for-running-redis-over-aws
https://aws.amazon.com/elasticache/
https://d0.awsstatic.com/whitepapers/performance-at-scale-with-amazon-elasticache.pdf

amazon web services

MongoDB

- AWS Quick Start reference guide
  - CloudFormation template for v2.6, v3.0 or v3.2
  - Build a sharded cluster running on Amazon Linux in 15 minutes

- MongoDB Cloud Manager

  Provision and monitor instances in AWS

- MongoDB Atlas: MongoDB as a Service on AWS

https://docs.mongodb.com/ecosystem/platforms/amazon-ec2/
http://docs.aws.amazon.com/quickstart/latest/mongodb/welcome.html
https://www.mongodb.com/cloud/cloud-manager
https://www.mongodb.com/cloud/atlas

Apache Cassandra

- Build your own on EC2

  Please read our whitepaper for guidelines and best pratices

- ~~Use the Datastax AMI~~

  http://www.techrepublic.com/article/the-battle-for-apache-cassandra-highlights-major-problem-with-open-source-projects/

- Instaclustr: Cassandra as a Service on AWS
- Alternative: Amazon DynamoDB

```
aws dynamodb create-table
```

https://d0.awsstatic.com/whitepapers/Cassandra_on_AWS.pdf
https://docs.datastax.com/en/cassandra/2.1/cassandra/install/installAMI.html
https://github.com/riptano/comboami

amazon web services

Analytics

amazon web services

Elasticsearch

AWS Free Tier

- Amazon Elasticsearch Service
- Managed service (v1.5, v2.3, v5.1)
- Scale compute and storage without downtime
- Automatic detection and replacement of failed nodes
- Snapshots to S3
- Integration with LogStash and CloudWatch Logs
- Built-in Kibana 3 & 4

https://www.elastic.co/blog/running-elasticsearch-on-aws
https://aws.amazon.com/elasticsearch-service/
https://github.com/awslabs/logstash-output-amazon_es
http://docs.aws.amazon.com/AmazonCloudWatch/latest/logs/CWL_ES_Stream.html

amazon web services

RabbitMQ

- Build your own on EC2
- CloudAMQP: RabbitMQ as a Service on AWS
- Alternative: Amazon SQS

```
aws sqs create-queue --queue-name name
```

https://www.rabbitmq.com/ec2.html
https://www.cloudamqp.com/blog/2014-11-14-amazon-vpc-peering.html
https://aws.amazon.com/sqs/

amazon web services

Apache Kafka

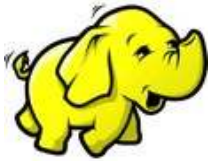
- Build your own on EC2
  - Quick start for Confluent platform
- Cloudkafka: Kafka as a Service on AWS
- Alternative: Amazon Kinesis

```
aws kinesis create-stream --stream-name name --shard-count nb_shards
```

https://aws.amazon.com/blogs/big-data/real-time-stream-processing-using-apache-spark-streaming-and-apache-kafka-on-aws/
https://www.confluent.io/blog/design-and-deployment-considerations-for-deploying-apache-kafka-on-aws/
https://www.cloudkafka.com
https://aws.amazon.com/about-aws/whats-new/2017/04/new-quick-start-deploys-confluent-platform-on-the-aws-cloud/
https://aws.amazon.com/kinesis/

amazon web services

Apache Hadoop and friends



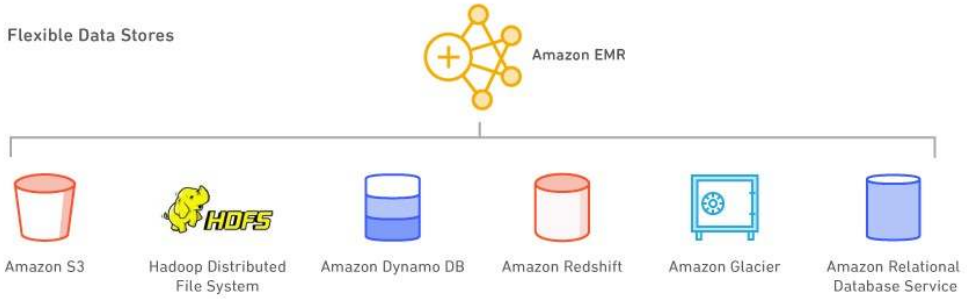
- Cloudera
  - AWS Quick Start reference guide: CloudFormation template

- Hortonworks
  - Hortonworks Data Cloud: CloudFormation template on AWS Marketplace

http://blog.cloudera.com/blog/2013/03/how-to-create-a-cdh-cluster-on-amazon-ec2-via-cloudera-manager/
http://docs.aws.amazon.com/quickstart/latest/cloudera/welcome.html
http://fr.hortonworks.com/blog/deploying-hadoop-cluster-amazon-ec2-hortonworks/
https://fr.hortonworks.com/products/cloud/aws/

Amazon Elastic Map Reduce (EMR)

- Apache Hadoop, Spark, Hive,etc.
- Managed service
- Easy to start, resize & terminate clusters
- Cost-efficient, especially with Spot Instances
- Integration with backends



https://aws.amazon.com/emr/

Conclusion

amazon web services


CentOS
debian
GNU/Linux
redhat.
SUSE
We adapt. You succeed.
ubuntu
FreeBSD
NetBSD
OpenBSD
iOS
Java
node
Microsoft
.NET
python
php
docker
kubernetes
RANCHER
CoreOS
OPENSHIFT
MySQL
ANSIBLE
GitHub
SALTSTACK
puppet
labs
CHEF
Jenkins
MariaDB
mongoDB
redis
cassandra
RabbitMQ
PostgreSQL
Spark
presto
HIVE
APACHE
HBASE
amazon
web services

Notable Open Source projects @ AWS

AWS Labs https://github.com/awslabs

- aws-shell https://github.com/awslabs/aws-shell
- s2n https://github.com/awslabs/s2n
- chalice https://github.com/awslabs/chalice

Blox: tools for custom Docker scheduling https://blox.github.io

Mxnet: Deep Learning library http://mxnet.io

Bugfixes and Feature Contributions

- Linux & Drivers
- Xen
- Apache Tomcat
- PostgreSQL
- Docker
- Boto
- Apache Hadoop
- Apache Hive
- Apache Bigtop
- Apache Oozie
- Apache Drill
- Apache Zeppelin
- Apache Pig
- Cloudera HUE
- Apache Lucene
- Apache Solr
- Kuromoji
- ElasticSearch
- CBMC
- Moses
- Apache Joshua

LINUX FOUNDATION
SILVER MEMBER

CORE INFRASTRUCTURE INITIATIVE

OPEN CONTAINER INITIATIVE

amazon web services

AWS is a rich and lively environment for Open Source platforms

Your choice: DIY, Marketplace, Partners, AWS Managed Services

The tools & projects you love, with less or no infrastructure drama

Built-in high availability, scalability, security & compliance

Focus on creativity and productivity, not on plumbing

Thank you!

Julien Simon
Principal Technical Evangelist
Amazon Web Services

julsimon@amazon.fr
@julsimon

amazon web services