CloudMan: Galaxy on the Cloud

Enis Afgan

Galaxy Community Conference May 26, 2011 - Lunteren

When to use the cloud?

- Don't have informatics expertise or infrastructure to run/maintain
- Have extended or particular resource needs
- Cannot upload data to a shared resource
- Need for customization
- Have oscillating data volume

The big picture



Galaxy CloudMan

- Complete solution for instantiating, running and scaling cloud resources with automatically configured Galaxy
 - Scope of tools and reference datasets exceed Galaxy Main
- Deployment on Amazon Web Services Cloud
 - Wizard-guided setup: requires no computational expertise, no infrastructure, no software
- Automated configuration for machine image, tools, and data
- Self-contained deployment
- Dynamic persistent storage
- Sharing of derived cluster instances
- Elastic resource scaling: manual or automatic
- Deploy a Galaxy cluster in minutes!

Deploying Galaxy on the AWS Cloud

- 1. Create an AWS account and sign up for EC2 and S3 services
- 2. Use the AWS Management Console to start a master EC2 instance
- 3. Use the Galaxy CloudMan web interface on the master instance to manage the cluster

2. Start an EC2 Instance

	Y Community Y Supp	ort Accour	*		
Your Account	🙀 aws.amazon.com 🛛 ANS (Pr	oducts Developers Comm	unity : Support : Account	Welcome, Enis Afgan	Settings Sign Out
Account Activity Personal Informatie View current charges and account activity; by service and using type.	Amazon S3 Amazon EC2	Amazon Elastic Ar MapReduce Cio	nazon udFront Amazon RDS		
Consolidated Billing	Navigation Amag	ron EC2 Console Dash	board		1
Euge up to receive one bill the multiple starts. Security Credentials accesseds, and add or remeave accesses from Access of a syster bill.	Region: 10 US Cast + Ge	tting Started	8	My Resources	
DevPay Activity View revenue and costs for your manepoints Arrane DevPey products. Psymet Method Was and with correct access institutions as	> EC2 Dashbaard Distances la > Instances la	start using Amazon EC unch a virtual server, kn stance.	2 you will want to own as an Amazon EC2	You are using the following Amazon EC2 resources in the US East (Virginia region:	5) 📿 Helmah
and as all the payment methods.	Spot Requests	[🗃 0 Running Instances 🍷 0 Eli	istic IPs
	DNAGES	Launch Ins	tance 🔛	> 6 EBS Volumes - 12 E	BS Snapshots
	+ AMIa + Bundle Taska	Nets: Your Instances w (Virgina	Request Instances W	izard	
Introduc g Amazon Relational Database Sei	ELASTIC BLOCK STORE > Volumes > Snapshots	rvice Health	CROSSERIE AND INCLUSION Please review the informat	OCHES OLAT BY Feet Collegent of tion below, then click Launch.	ACTUAL REALES
• Learn Hore	NETWORKING & SECURITY	Amazon EC2 (US East Virginia)	AHI	Other Linux AMI ID ami-ed03ed84 (x	86_64) Edit AMI
	Security Groups Key Pairs Load Balancers	+ Vew	Number of Instances: Availability Zone: Monitoring: Instance Type: Instance Class:	No Preference Disabled Large (m1.large) On Demand	Edit Instance Details
			Kernel ID: Ramilisk ID: User Data:	Use Default Use Default testGC1 AKIAJKQE3RT	Edit Advanced Deta
			Key Pair Name:	galaxy_keypair	Edit Key Pair

Launch

3. Configure Your Cluster

← → C ③ ec2-50-16-1-149.compute-1.amazonaws.com/cloud

Galaxy Cloudman





→ C ③ ec2-50-16-1-149.compute-1.amazonaws.com/cloud

Galaxy Cloudman

Galaxy Cloudman Console

Welcome to Galaxy Cloudman. This application will allow you to manage this cloud instance and the services provided within. If this is your first time running this cluster, you will need to select an initial data volume size. Once the data store is configured, default services will start and you will be able to add and remove additional services as well as 'worker' nodes on which jobs are run.

	Terminate	cluster	Add nodes v	Remove nodes	Access Galaxy
St	atus				
	Cluster name:	Heteroplasn	ny study 🛃		
	Disk status:	50M / 1000	Autoscaling is off.		
1	Worker status:	Idle: 0 Ava	ilable: 0 Requested: 0		Turn on?
1	Service status:	Applications	😐 Data 👄		
1	External Logs:	Galaxy Log			
	Cluster status	log			•



Components -> deployment



Building CloudMan machine image

- Dependencies and tools
 - Automated the configuration process
 - Go from bare bones OS to Galaxy-ready machine
- Applicable outside the cloud
 - Build-a-machine (yourself)
- Based on Python Fabric
 - Easily extensible: add your own tool



- Built a Galaxy pre-configured VM image
 - Galaxy and a range of tools
 - Reference genomes will need to be downloaded (but it's automated)
- Available from S3 bucket `usegalaxy`

CloudMan instances

- Self-contained configuration: CloudMan, Galaxy & tools, reference genomes, data
- Thus:
 - Reproducible
 - Customizable
 - Extensible

Data

- It is only yours
- Preserved while the cluster is off
- Can grow as the analysis grows
 - Up to 1,000GB currently

The importance of sharing

- Share entire Galaxy CloudMan cluster instances
- Publish an analysis
 - In progress or otherwise
- Use CloudMan as PaaS
 - Deploy your own tool and make it available
- Snapshot your instance
 - Data
 - Configuration

Deployment sharing

Galaxy Cloudman

Info: report bugs | wiki | screencest

×

Currer Ga	alaxy Cloudman		Info: report.bupi i wiki i screensi	and the second se		
Share	Galaxy C	loudman Console	×	X		
This f Instan You m saving While Then, 'share snapsi enable only b actual O P Spec Boti Thes Iden AWS AWS	Currently These are t Also, for ref Share-an We Ser Ext	Salaxy Cloudman Galaxy Cloudm Galaxy Clou Welcome to Galaxy within. If this is yt store is configured 'worker' nodes on Terminat Status Cluster name Disk status: Worker statu Service statu External Logs	Information of the services provided within. To get started, choose the type of cluster and the services provided within. To get started, choose the type of cluster you'd like to work with and specify the size of your persistent data storage, if any. Image: Image	provided he data vell as		
			Hide extra options (Start Cluster)			

Grow and Shrink



→ C () ec2-50-17-119-106	5.compute-1.amazonaws.com/cloud		☆ ▲		
Galaxy Cloudman		Info: p	eport bugs wiki screencast		
Galaxy Clo Welcome to Gala data st	oudman Console	Salaxy CloudMan. Your previ	aus .	6 3	
G	alaxy Cloudman		Infot p	sport. buga wiki screencast	
Stat	Galaxy Cloudman Console				
Clu Dis We	Autoscaling Configuration				
Ser	Autoscaling attempts to automate the elasticity offered to autoscaling takes over the control over the size of Autoscaling is simple, just specify the cluster size limits y do. The cluster will not automatically shrink to less than grow larger than the maximum number of worker nodes	y doud computing for this p your cluster, ou want to want to work wit the minimum number of wo you specify.	nticular cluster. Once turned o hin and use your cluster as you ker nodes you specify and it will	ne normally never	
	While respecting the set limits, if there are more jobs the automatically add compute nodes; if there are cluster no nodes reducing the size of the cluster and your cost. Once turned on, the cluster size limits respected by auto	in the cluster can comfortabl des sitting idle at the end of scaling can be adjusted or as	y process at a given time autosc an hour autoscaling will terminat toscaling can be turned off.	aling will a those	
	← → C (© ec2-50-17-119-106.compute-1.am	zonaws.com/cloud		Info: most buss I wil	्रि Li serrero
	Galaxy Cloudman Co Welcome to Galaxy Cloudman. The data store has been reconnected. nodes for running jobs.	is application allows you t Once the cluster has init	o manage this instance of G alized, use the controls below	alaxy CloudMan. Your previous v to add and remove 'worker'	
_	Terminate cluster	Add nodes	Remove nodes	Access Galaxy	
	Status Cluster name: share-an-ins Disk status: 84M / 10G (Worker status: Idle: 0 Ava Service status: Applications External Logs: Galaxy Log	tance demo 😭 1%) 😘 ilable: 0 Requested: 0 • Data •		Autoscaling is on. Turn off? Min nodes: 0 Max nodes: 15 Adjust limits?	
	Cluster status log			0	

+

Don't waste the resources

- Once the need for a given cluster subsides,
 Terminate cluster
 you can always start it
 back up
 - Data is preserved while a cluster is down

Summary of the CloudMan architecture

- Minimum setup time and cost
 - No need for an external broker
- Automated configuration
- Data persistence
- Built-in support for managing the oscillating data volume
- Self-contained deployment
 - Customizable instances: CloudMan as PaaS
 - Versioning of tools, data, and configurations



http://usegalaxy.org/cloud