AWS re:Invent

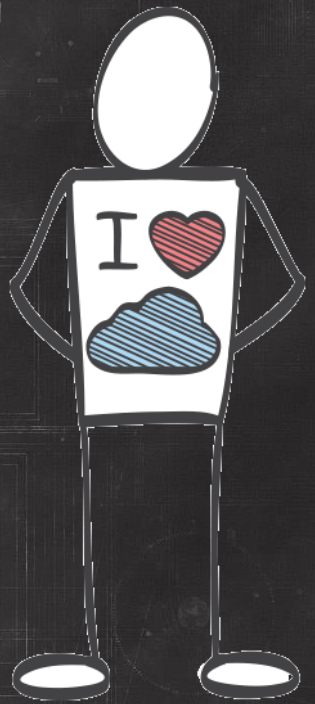
From One to Many: Evolving VPC Design

Robert Alexander, AWS Solutions Architect

November 14, 2013

amazon web services

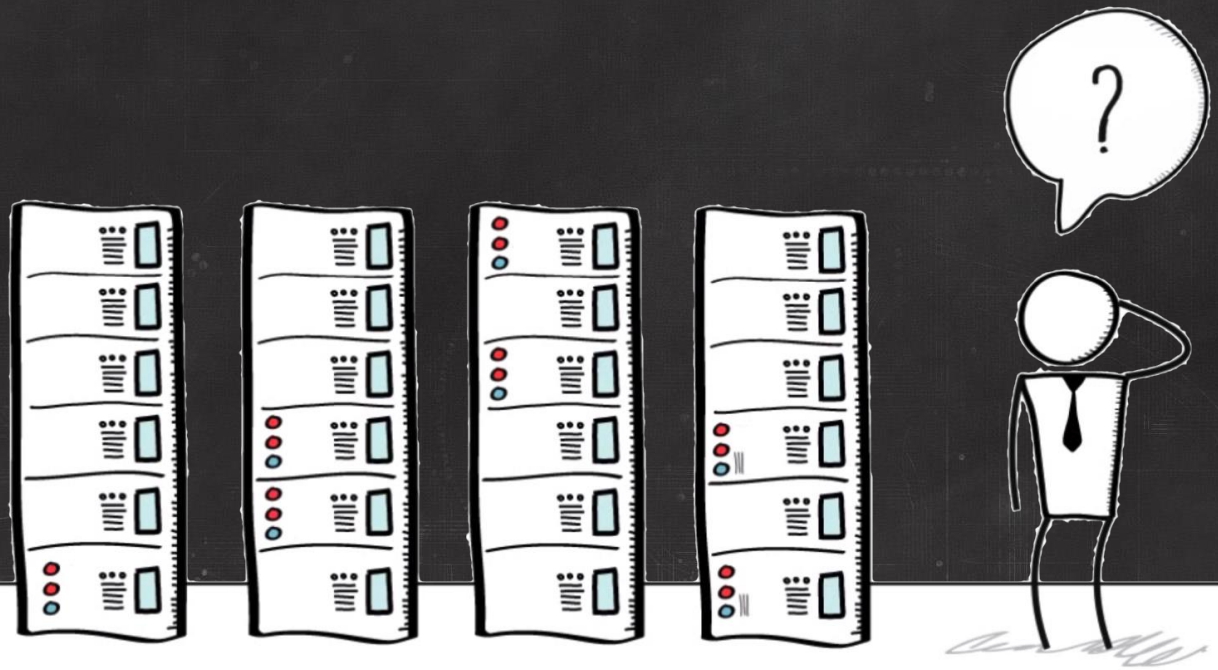
Disclaimer:



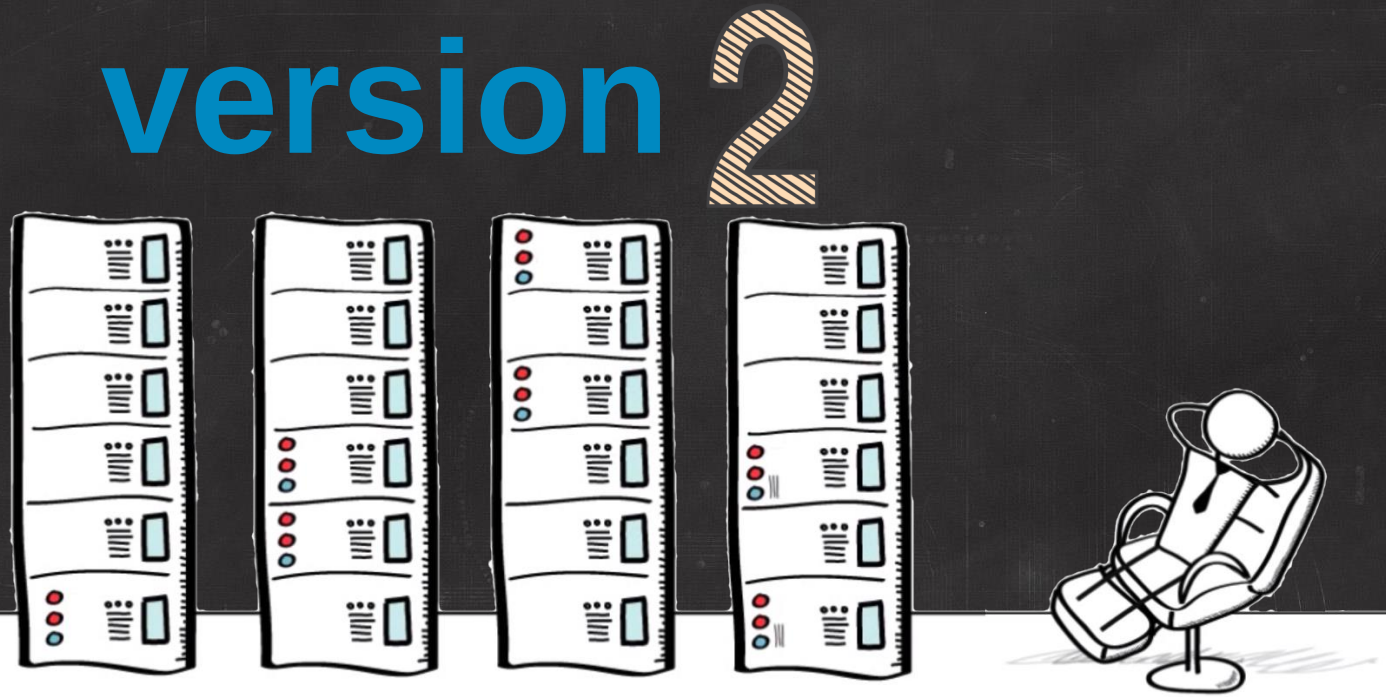
Do Try This at Home!

All these designs are in use by customers

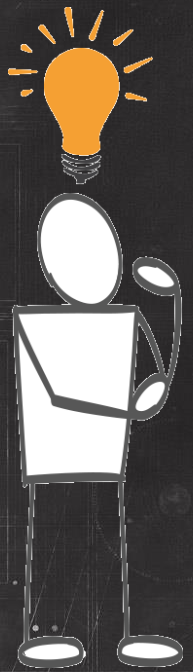
In a physical world you Design your network infrastructure… then spend a lot of time building and deploying



With Amazon Virtual Private Cloud, build and deploy virtual datacenters as fast as you design them



VPC Tip 1



Get to know AWS CloudFormation

- Source control and version control your datacenter
- Deploy infrastructure with one command
- Reproduce anywhere in the globe in minutes
- Segregation of Duties (SoD) between infrastructure and application owners

Elements of VPC Design


Amazon VPC


Router

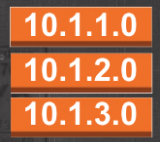

Internet Gateway


Customer Gateway


Subnet

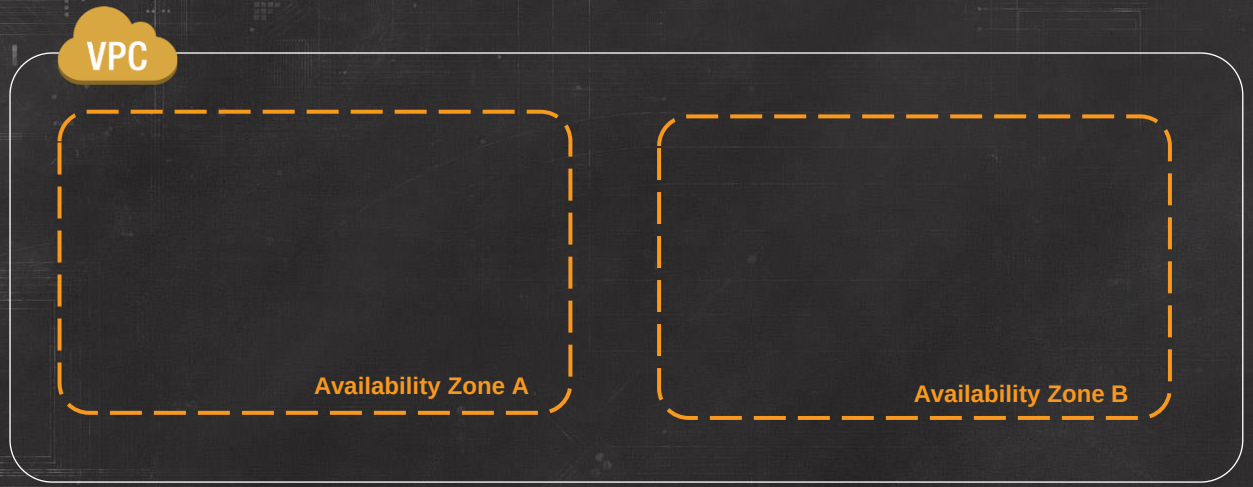

Virtual Private Gateway


VPN Connection



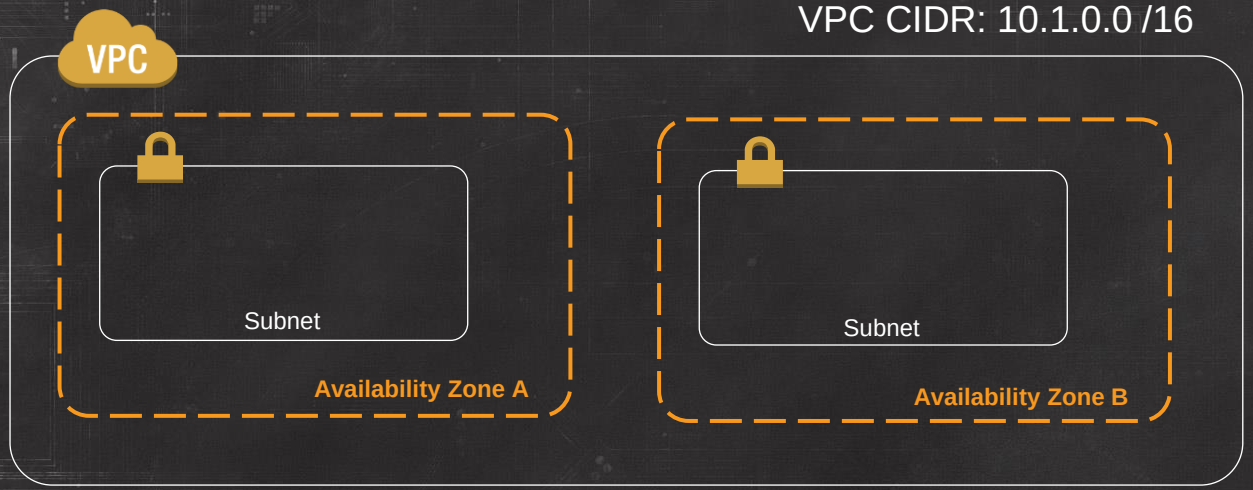
Route Table


Elastic Network Interface



- VPC is a private, isolated section of the AWS Cloud where you define the networking within
- A VPC can span all AZ's in an AWS Region
- Only one decision upon VPC creation:

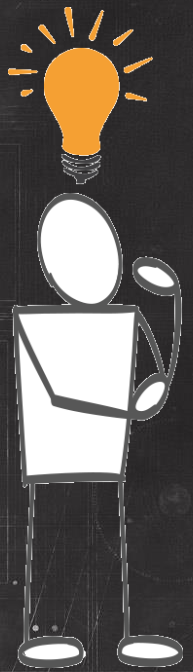
What IP CIDR block to assign?



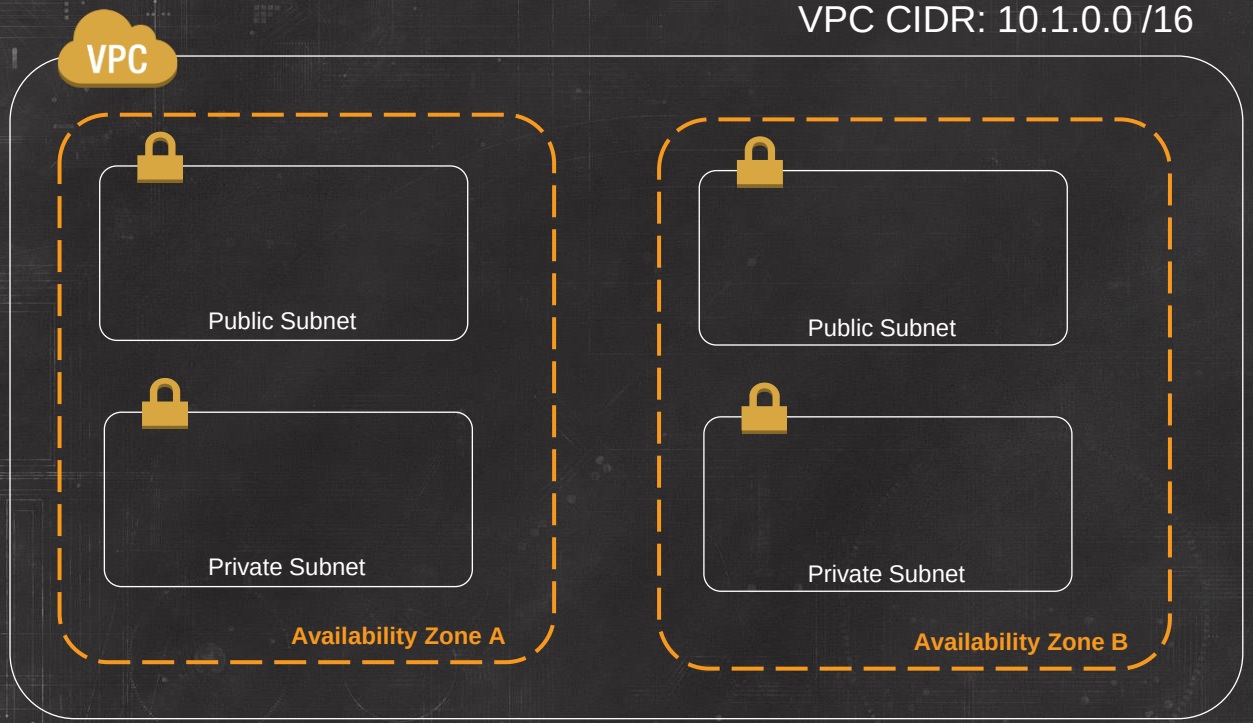
- Subnets are AZ specific
- On subnet creation only AZ, VPC and CIDR block designated
- Modifying a Subnet's Routing Table or Network Access Control Lists is done after creation

VPC Tip 2

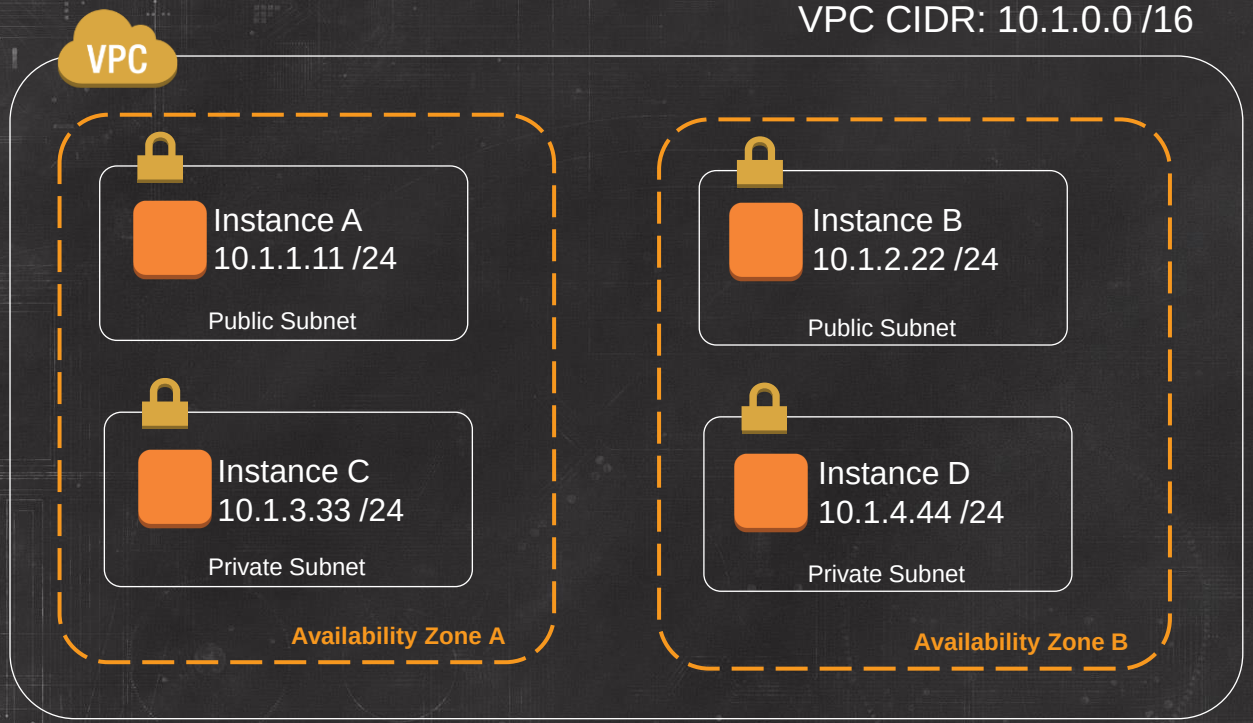
Plan your VPC IP space before creating it



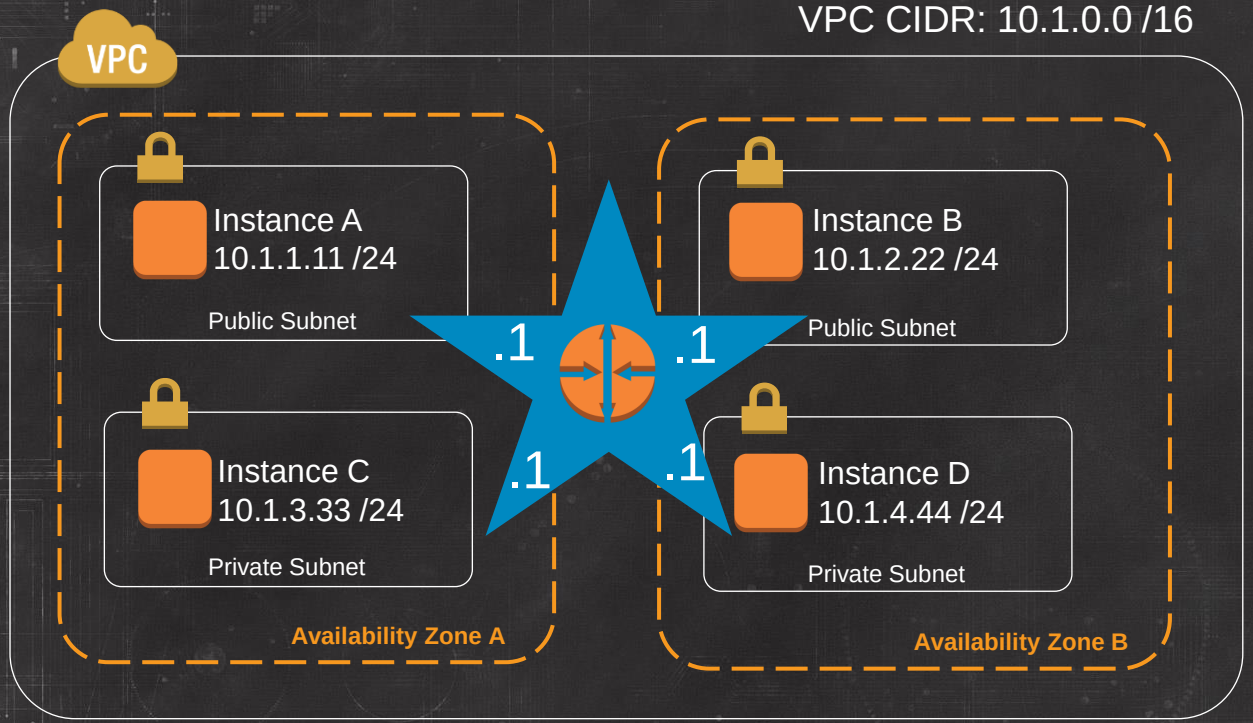
- Consider future AWS region expansion
- Consider future connectivity to corporate networks
- Consider subnet design
- VPC can be /16 down to /28
- CIDR cannot be modified once created
- Overlapping IP spaces = future headache



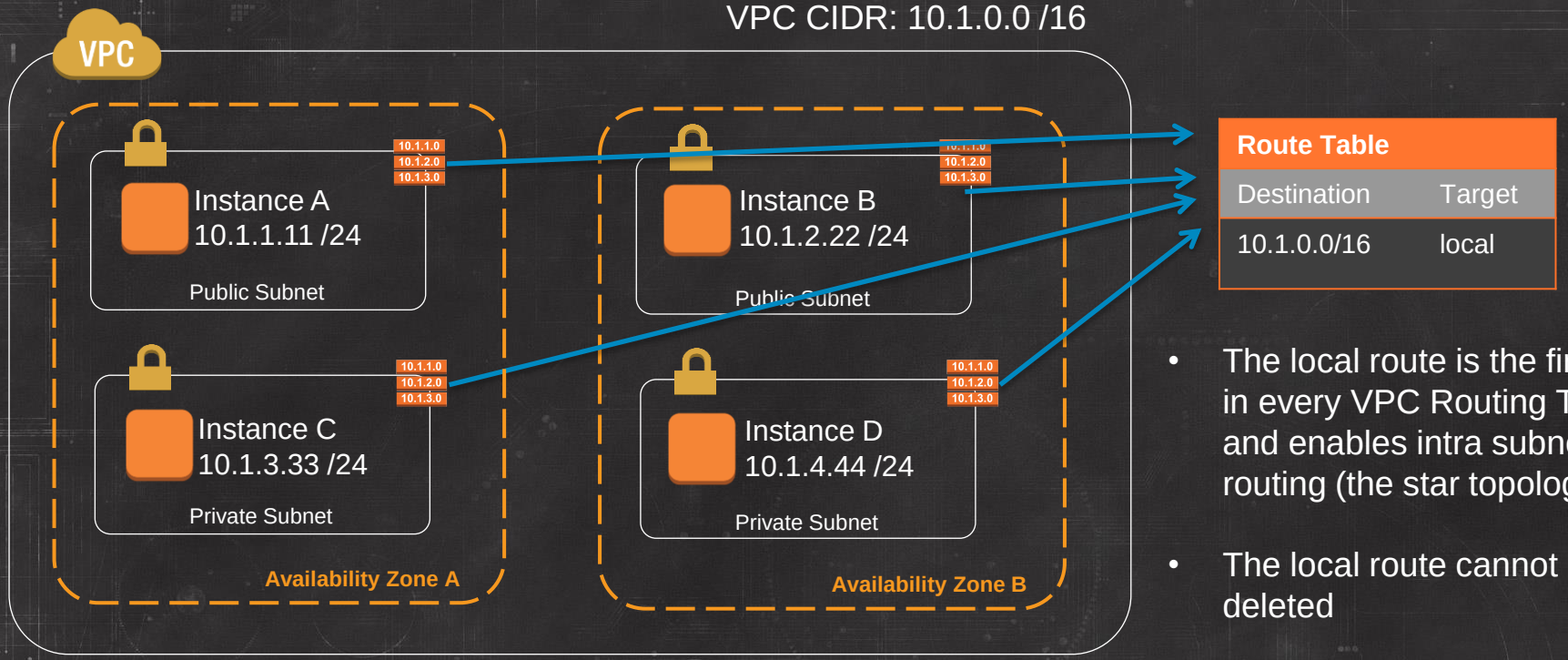
- Public and Private subnets are a common logical isolation
- At this point in VPC configuration, Public and Private are just indicators of the subnet purpose
- Several additional elements must be configured before traffic can egress the VPC



- Subnet size should be considered relative to subnet purpose and not the Layer 2 limits of traditional switched networks
- For subnets containing large, dynamic workloads, subnet size might be many 1000s of instances
- Traditional subnet constraints such as broadcast domain limits do not apply in VPC



- By default, every subnet can route to every other subnet in a VPC
- A virtual router forms this star topology between all VPC subnets
- The VPC DHCP Service hands out the virtual router address as the default gateway to every instance booting in a VPC subnet
- Virtual Router always takes the .1 address of every VPC subnet



- The local route is the first entry in every VPC Routing Table and enables intra subnet routing (the star topology)
- The local route cannot be deleted

VPC Tip 3

Leave the Main Route Table Alone

Route Table ID	Associated With	Main	VPC
rtb-39ca9d52	0 Subnets	Yes	vpc-3bca9d50 (10.1.0.0/16)

Route Table: rtb-39ca9d52

Routes | **Associations** | Route Propagation | Tags

Subnet	Actions
Select a subnet	Associate

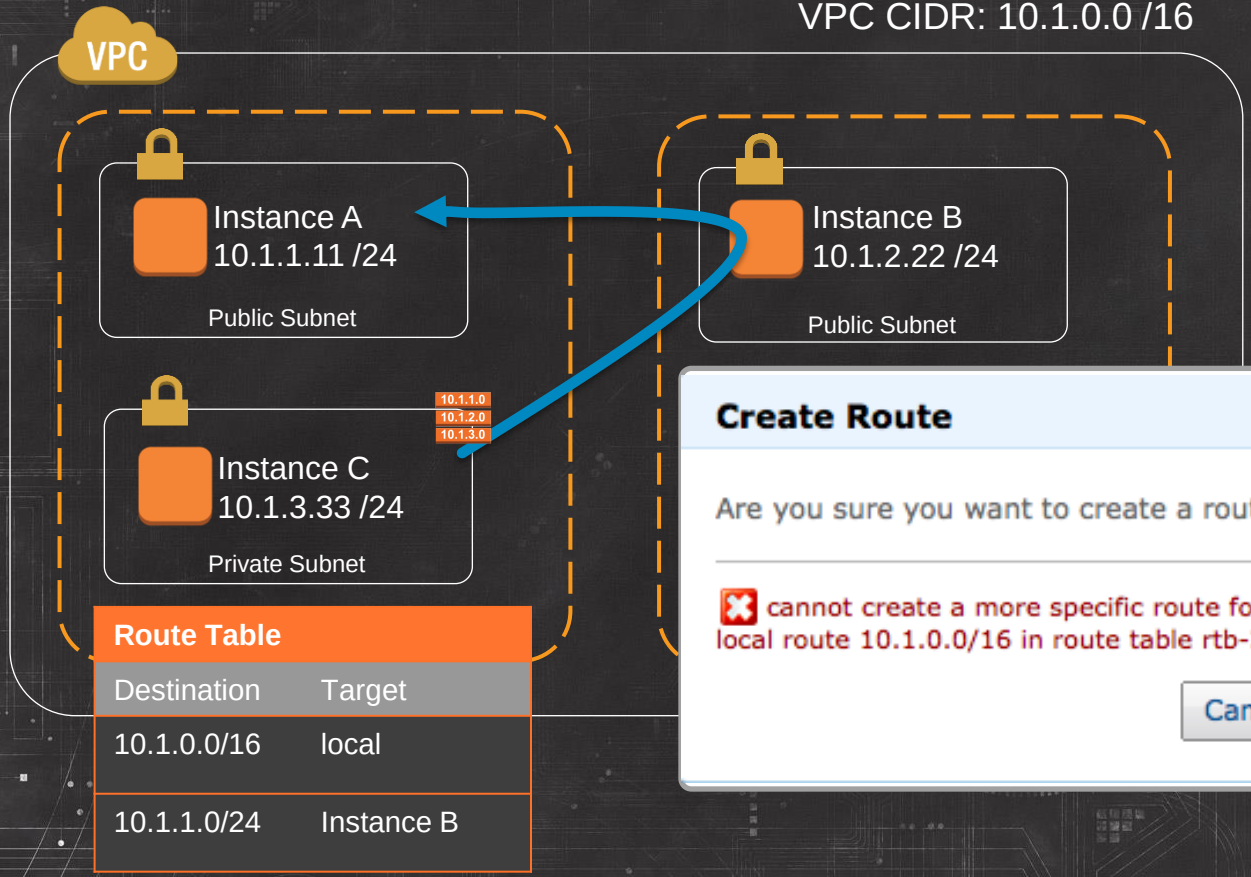
The following subnets have not been associated with any route tables and are therefore using the Main table routes:

- subnet-6af6a101 (10.1.4.0/24)
- subnet-2ff7a044 (10.1.1.0/24)
- subnet-8ef7a0e5 (10.1.3.0/24)
- subnet-d4f7a0bf (10.1.2.0/24)

VPC Tip 3

Leave the Main Route Table Alone

- Upon creation, every subnet is associated with the Main Route Table
- Only after subnet creation can you modify the Route Table assigned to a subnet
- So leave Main Route Table with only the local route and eliminate the possibility of a subnet being given routes it shouldn’t

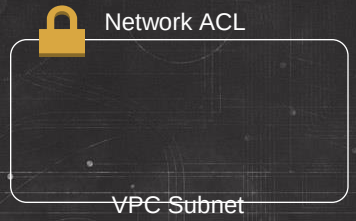


- You cannot create a route more specific than the local route
- VPC Routing Tables are for defining ways OUT of a VPC and not for defining Intra-VPC routes

Network ACLs vs Security Groups

NACLs

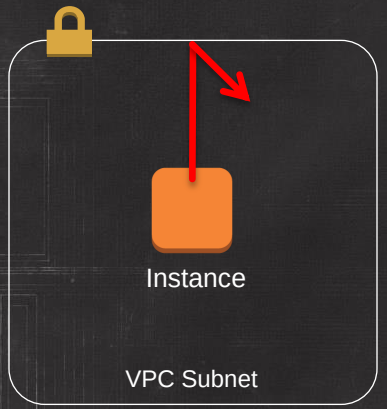
- Applied to subnets (1 per)
- Stateless
- Allow & Deny (blacklist)
- Rules processed in order



Security Groups

- Applied to instance ENI (up to 5 per)
- Stateful
- Allow Only (whitelist)
- Rules evaluated as a whole
- SGs can reference other SGs in same VPC

VPC Network ACLs: What are they good for?



- Enforcing baseline security policy
  - Example:

    “No TFTP, NetBIOS or SMTP shall egress this subnet”
- Catch all for holes in instance security groups
- Segregation of security between network ops and dev ops

VPC Network ACLs: Best Practices

- Use sparingly, keep it simple
- Egress security policies are best
- Create rule #’s with room to grow
- Use IAM to tightly control who can alter or delete NACLs

Pushing this will Hurt!

Default Network ACL:

Rule #	Port (Service)	Protocol	Source	Allow/Deny	Action
100	ALL	ALL	0.0.0.0/0	ALLOW	Delete
*	ALL	ALL	0.0.0.0/0	DENY	

VPC Tip 4

Create an IAM VPC Admin Group

Examples of “High Blast Radius” VPC API calls that should be restricted:

New Support Resource Permissions

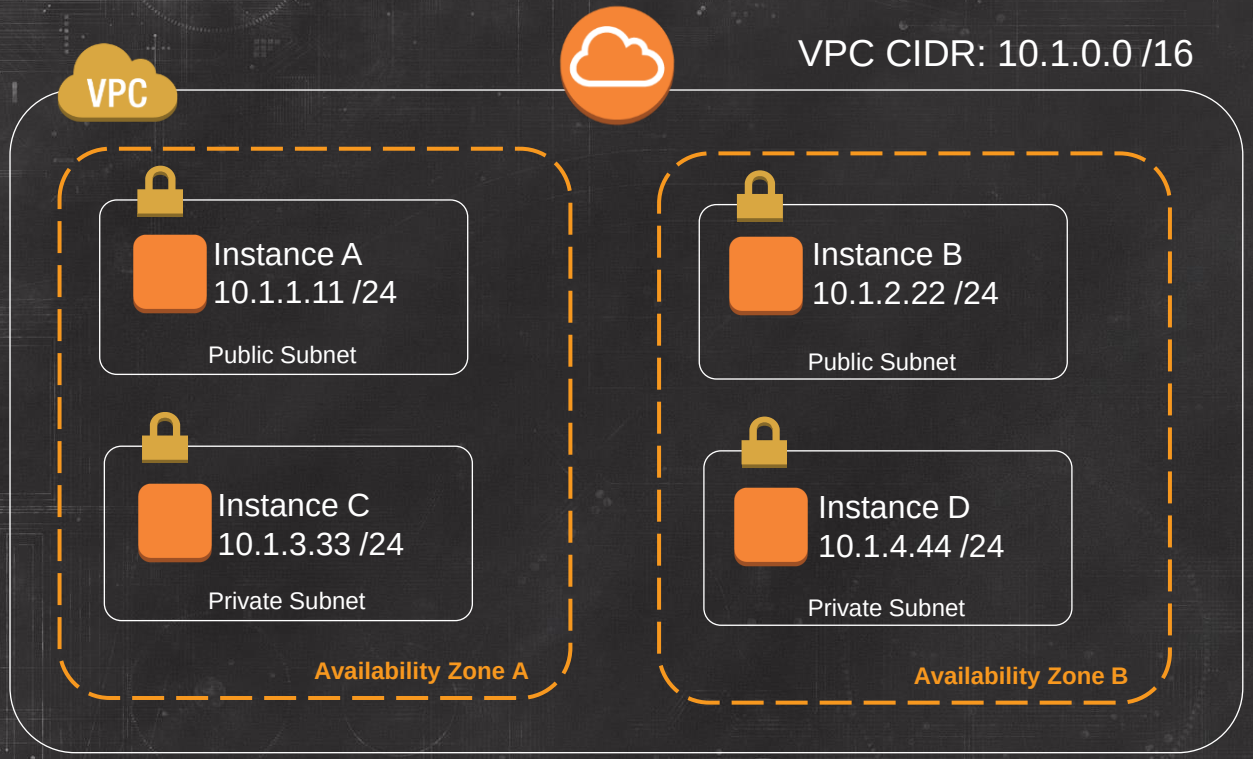
- AttachInternetGateway
- AssociateRouteTable
- CreateRoute
- DeleteCustomerGateway
- DeleteInternetGateway
- DeleteNetworkAcl
- DeleteNetworkAclEntry
- DeleteRoute
- DeleteRouteTable
- DeleteDhcpOptions
- ReplaceNetworkAclAssociation
- DisassociateRouteTable

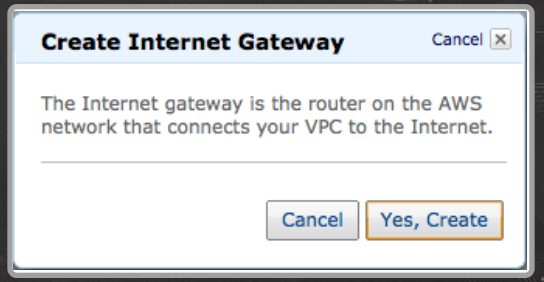
Example IAM Policy for NACL Admin

```
{
   "Version": "2012-10-17",
   "Statement": [
     {
      "Effect": "Allow",
      "Action": [
        "ec2:DeleteNetworkAcl",
        "ec2:DeleteNetworkAclEntry"
      ],
      "Resource": "arn:aws:ec2:us-west-2:123456789012:network-acl/*",
      "Condition": {
        "StringEquals": {
          "ec2:ResourceTag/Environment": "prod"
        },
           "Null": {
            "aws:MultiFactorAuthAge": "false"
        }
      }
   }
  ]
}
```

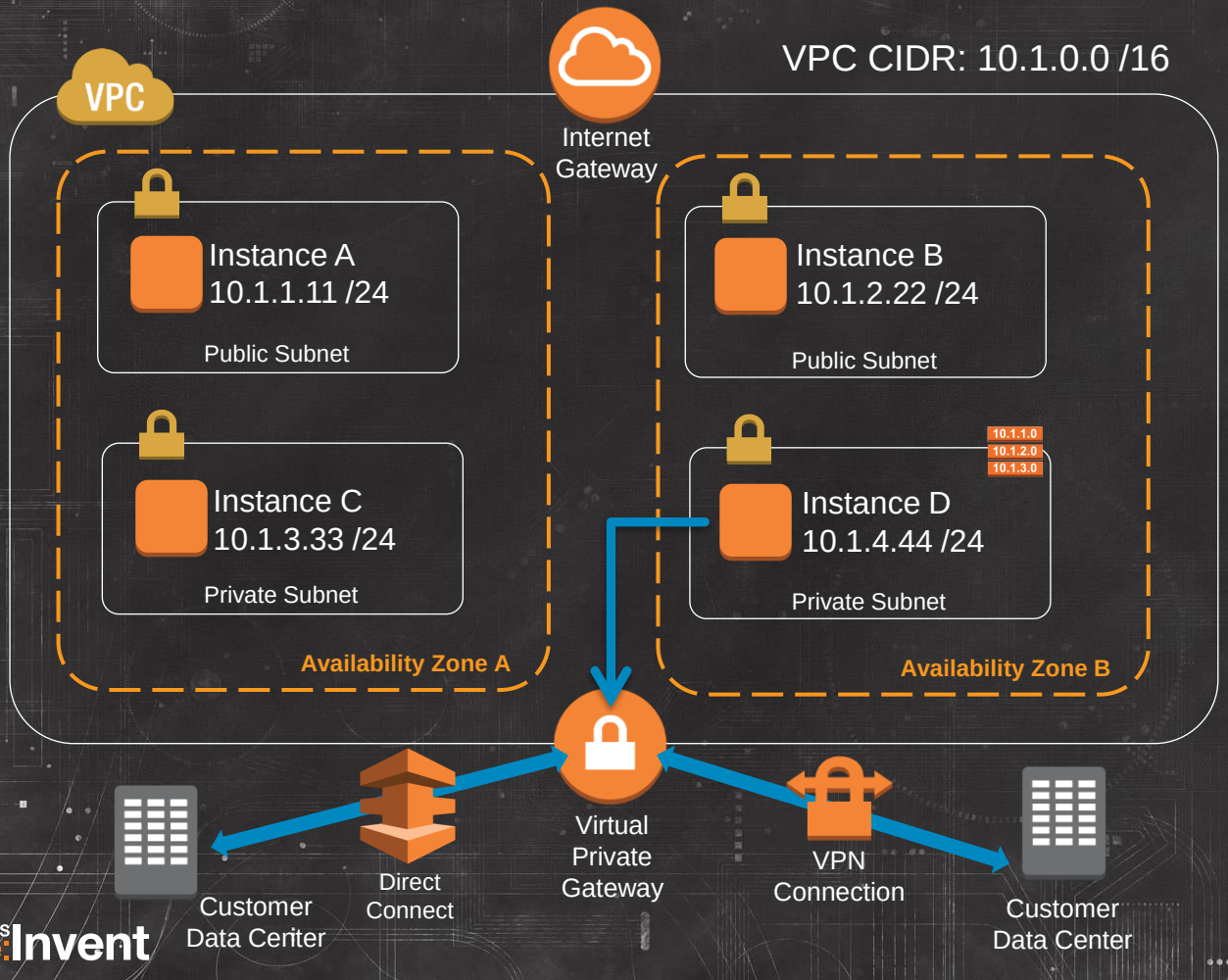
Multi Factor Authentication required for Actions in Policy

Creating Ways "Out" of a VPC

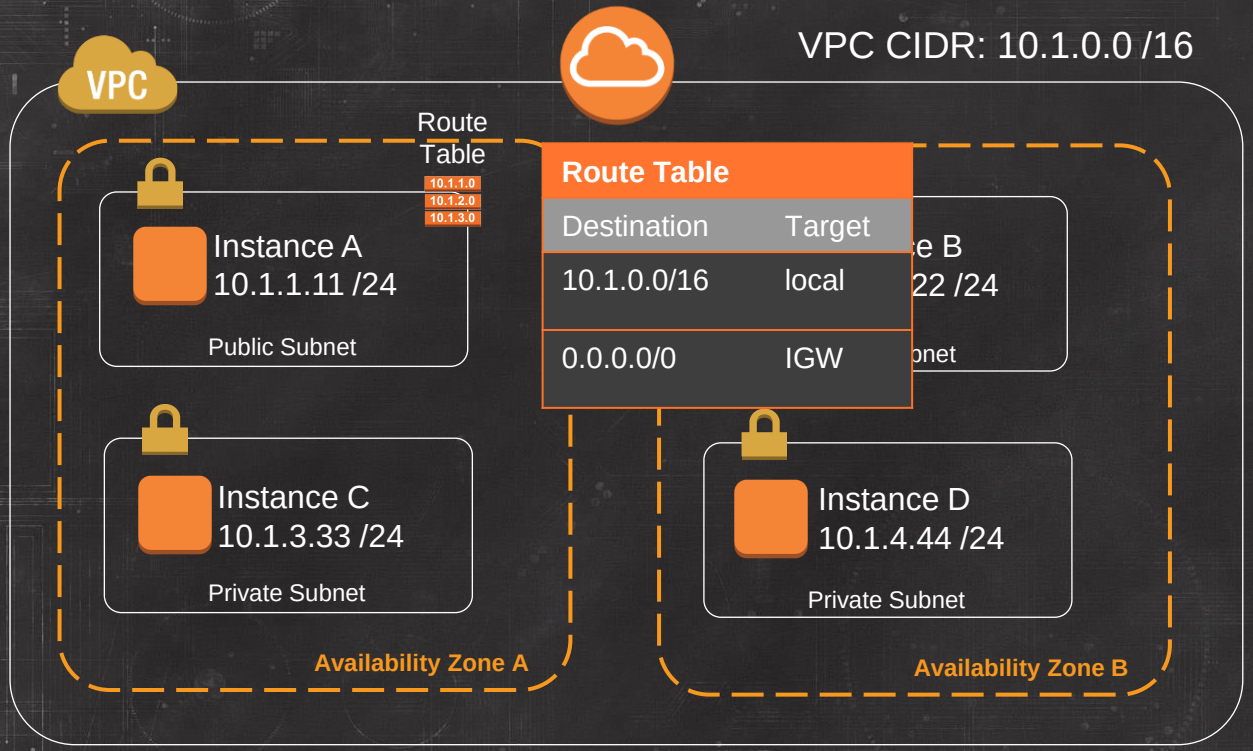




Only 1 IGW and 1 VGW per VPC



Route Table	
Destination	Target
10.1.0.0/16	local
Corp CIDR	VGW



Three Elements Required to Egress VPC from IGW:

1. Internet Gateway must be associated to VPC
2. Subnet must be associated to a Routing Table with a route to the IGW
3. Instances in the subnet that will egress VPC must be associated with a Public IP

Ways to Assign Public IPs

Elastic IP (EIP)

- Associated with AWS account and not a specific instance
- 1 Public IP to 1 Private IP static NAT mapping
- Instance does not “see” an EIP associated to it
- Persists independent of the instance
- Can be assigned while instance is stopped or running
- Can be moved, reassigned to other ENIs

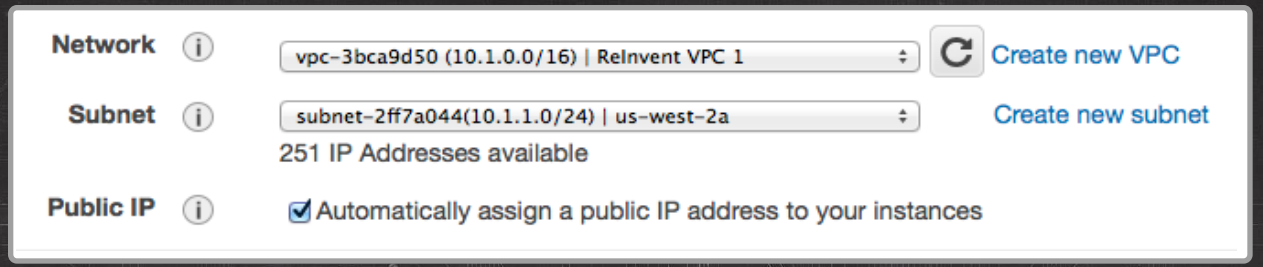
Ways to Assign Public IPs

New

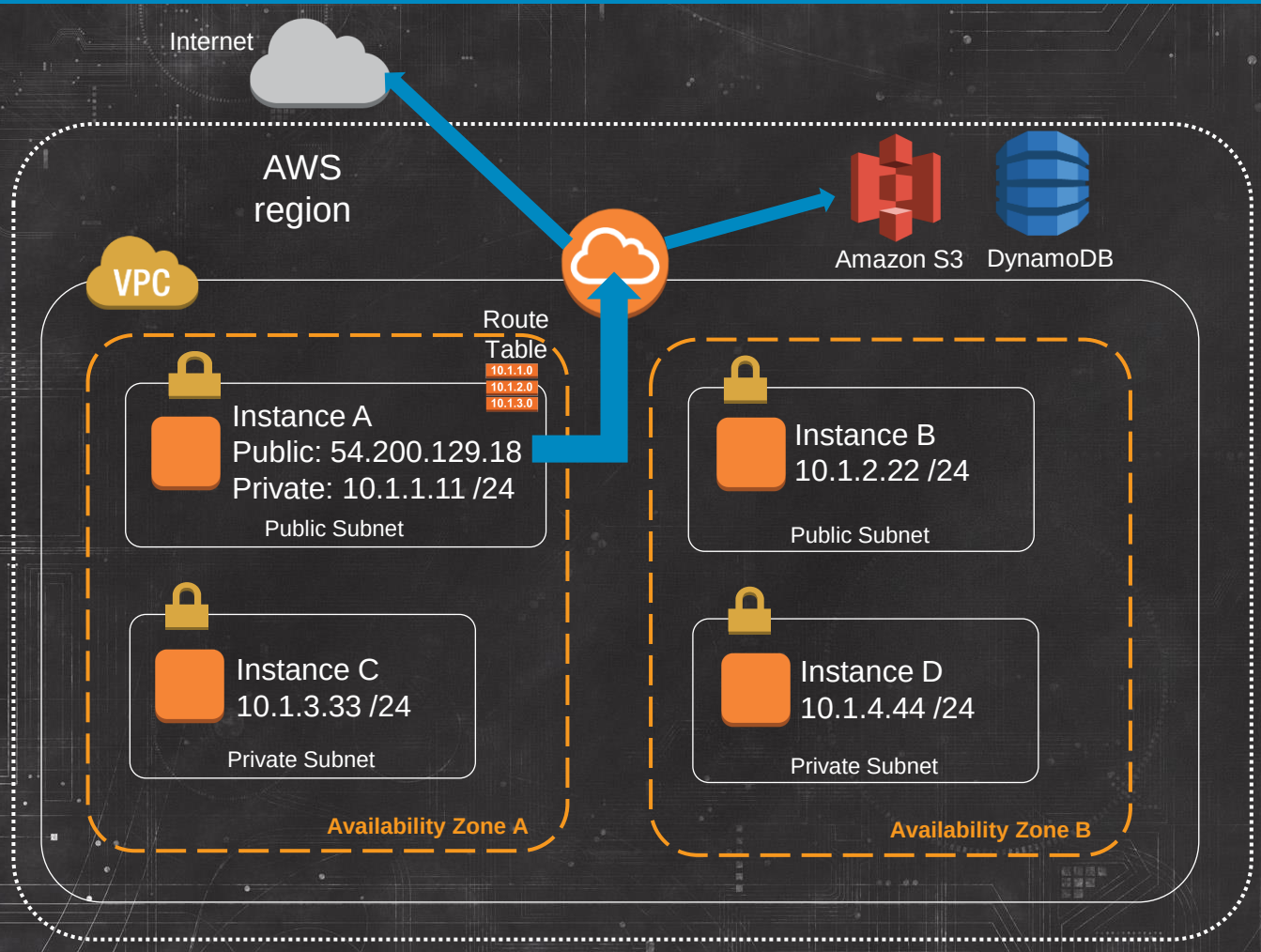
2

Automatic dynamic Public IP assignment

- Done on instance launch into VPC subnet
- Public IP is dynamic and could change if instance is stopped and restarted
- Does not count against AWS Account EIP limits
- Works only on instances with a single ENI



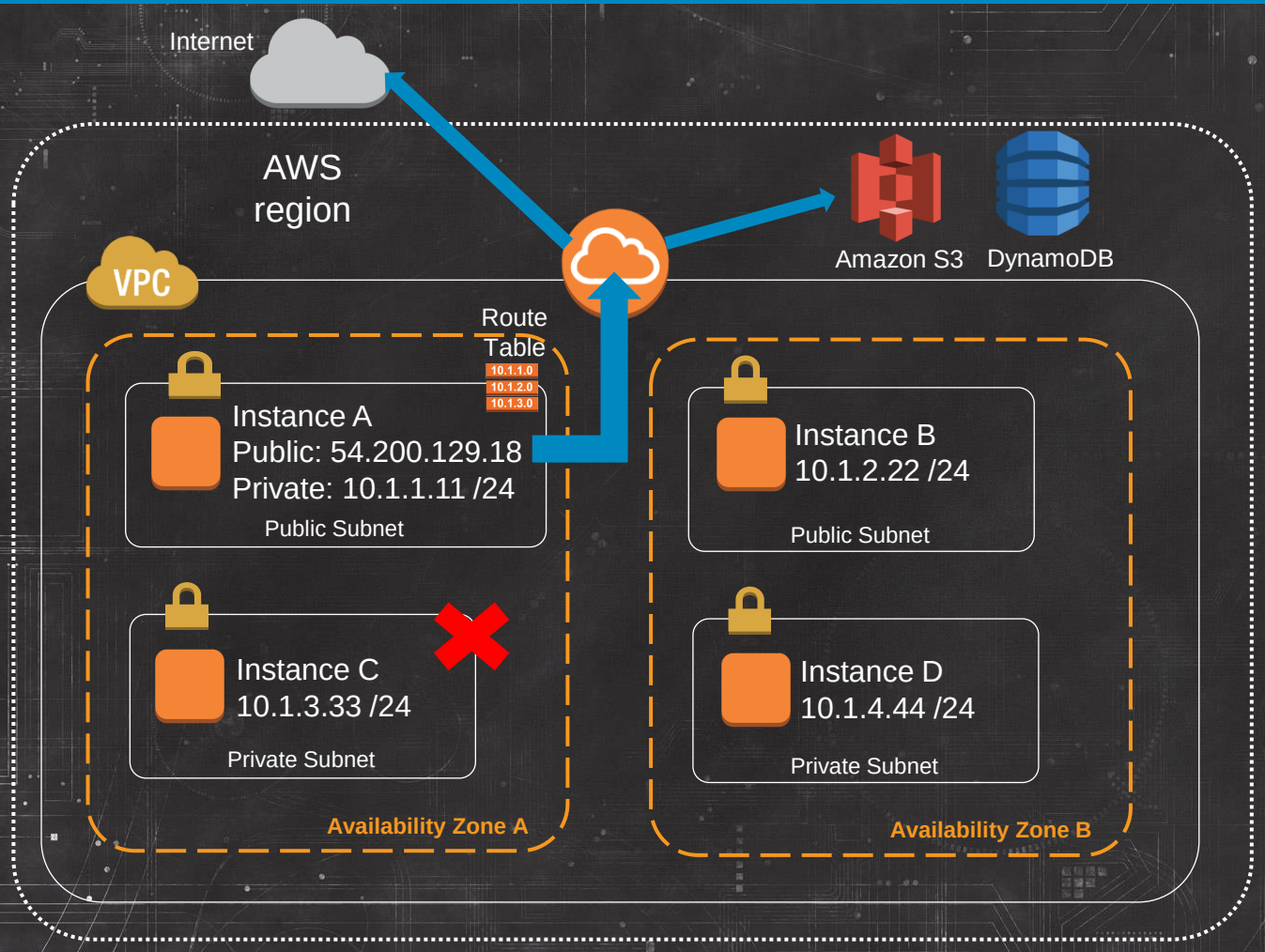
AWS outside the VPC



- Services such as S3 and Dynamo DB are Regional and accessible only via Public End Points
- Resources in a VPC requiring access to Regional services must be able to egress the VPC into the Public AWS network

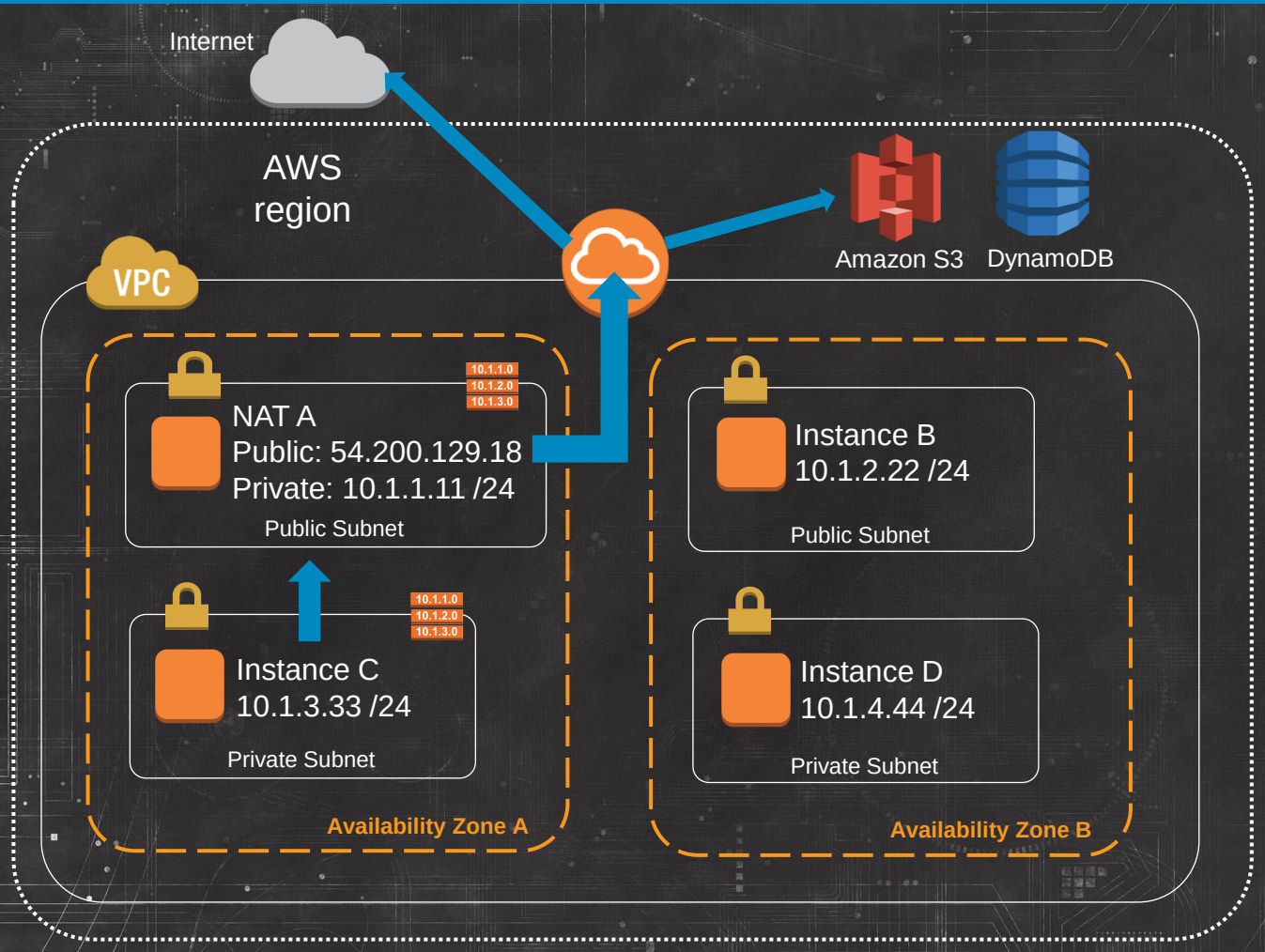
Examples of AWS outside the VPC

- AWS API Endpoints
  - Think about which APIs you might be calling from instances within the VPC
  - Good examples: Amazon EC2, AWS CloudFormation, Auto Scaling, Amazon SWF, Amazon SQS, Amazon SNS
- Regional Services
  - Amazon S3
  - Amazon Dynamo DB
- Software and Patch Repositories
  - Amazon Linux repo allows access only from AWS public IP blocks



And what if instance C in a private subnet needs to reach outside the VPC?

It has no route to the IGW and no public IP.



Deploy an instance that functions as a
N etwork
A ddress
T ranslat(or)

Route Table	
Destination	Target
10.1.0.0/16	local
0.0.0.0/0	NAT instance

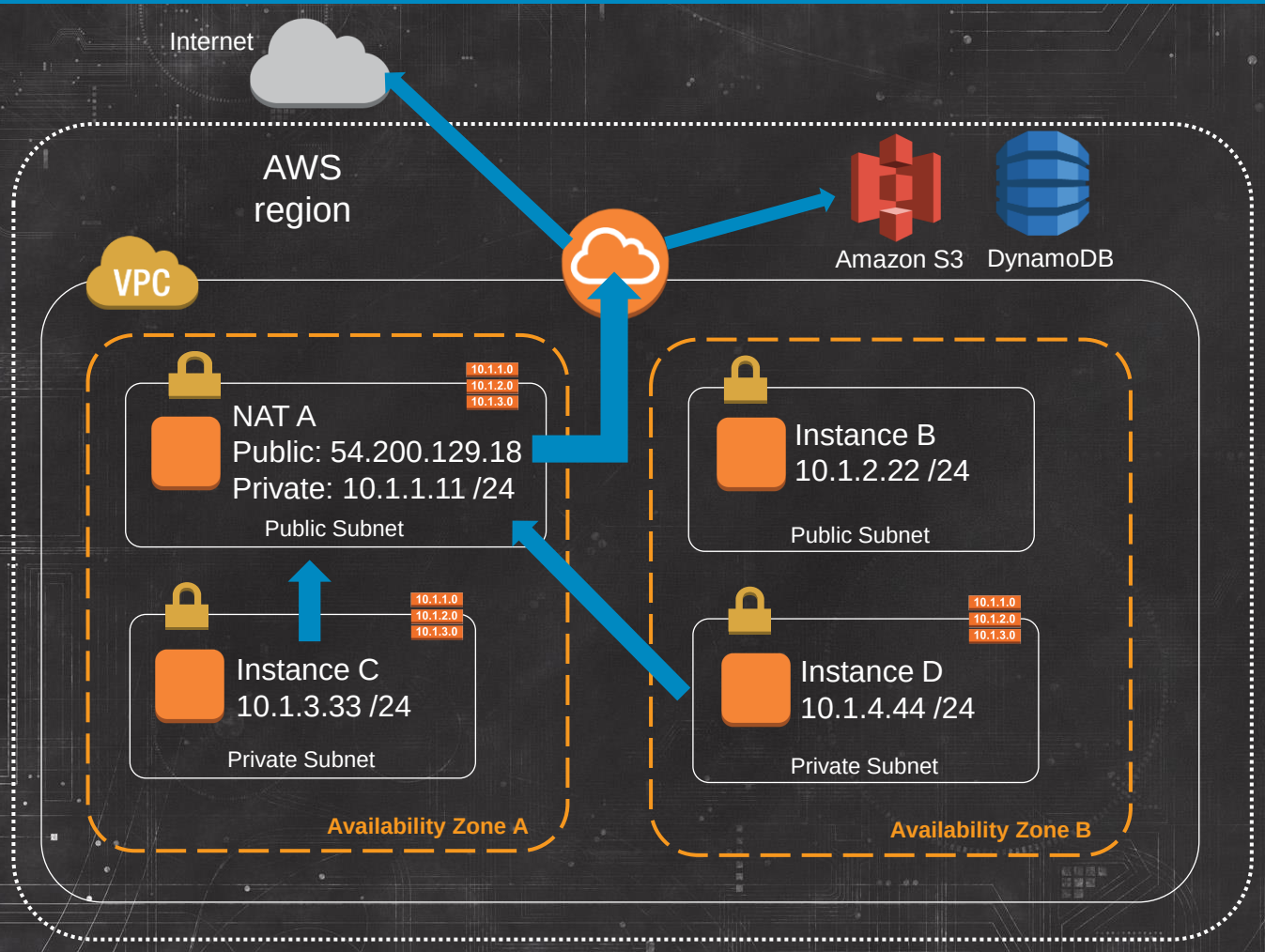
VPC Tip 5

What makes up the Amazon Linux NAT AMI?

Not much to it:

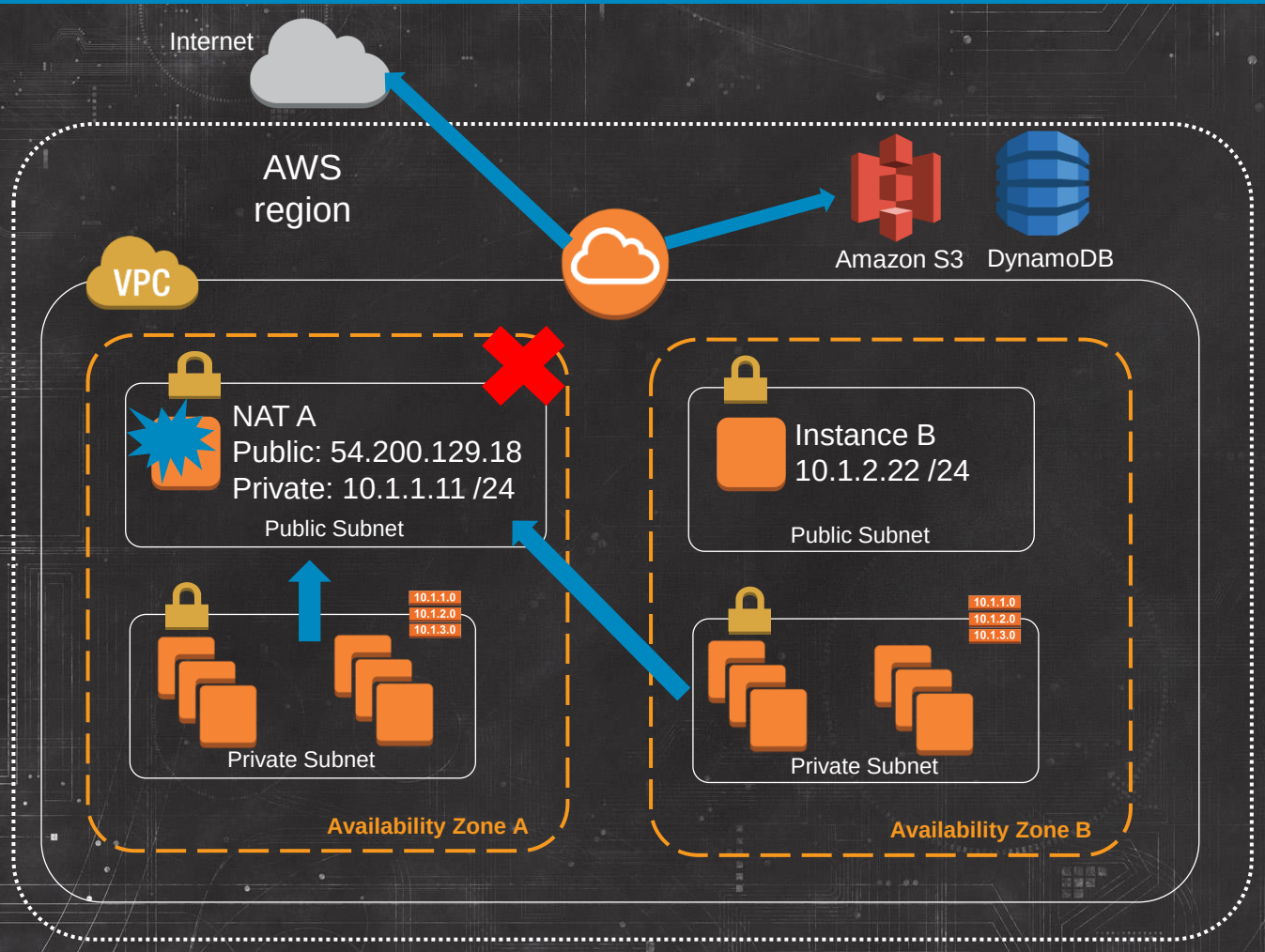
1. IP forwarding enabled
2. IP NAT Masquerading enabled in iptables
3. Source / Destination check is turned off on the instance

```
$echo 1 >  /proc/sys/net/ipv4/ip_forward
$echo 0 >  /proc/sys/net/ipv4/conf/eth0/send_redirects
$/sbin/iptables -t nat -A POSTROUTING -o eth0 -s 10.1.0.0/16 -j MASQUERADE
$/sbin/iptables-save
$aws ec2 modify-instance-attributes –instance-id i-xxxxxxxx –source-dest-check “{\”Value\”:false}”
```



Other private subnets can share the same routing table and use the NAT

But…

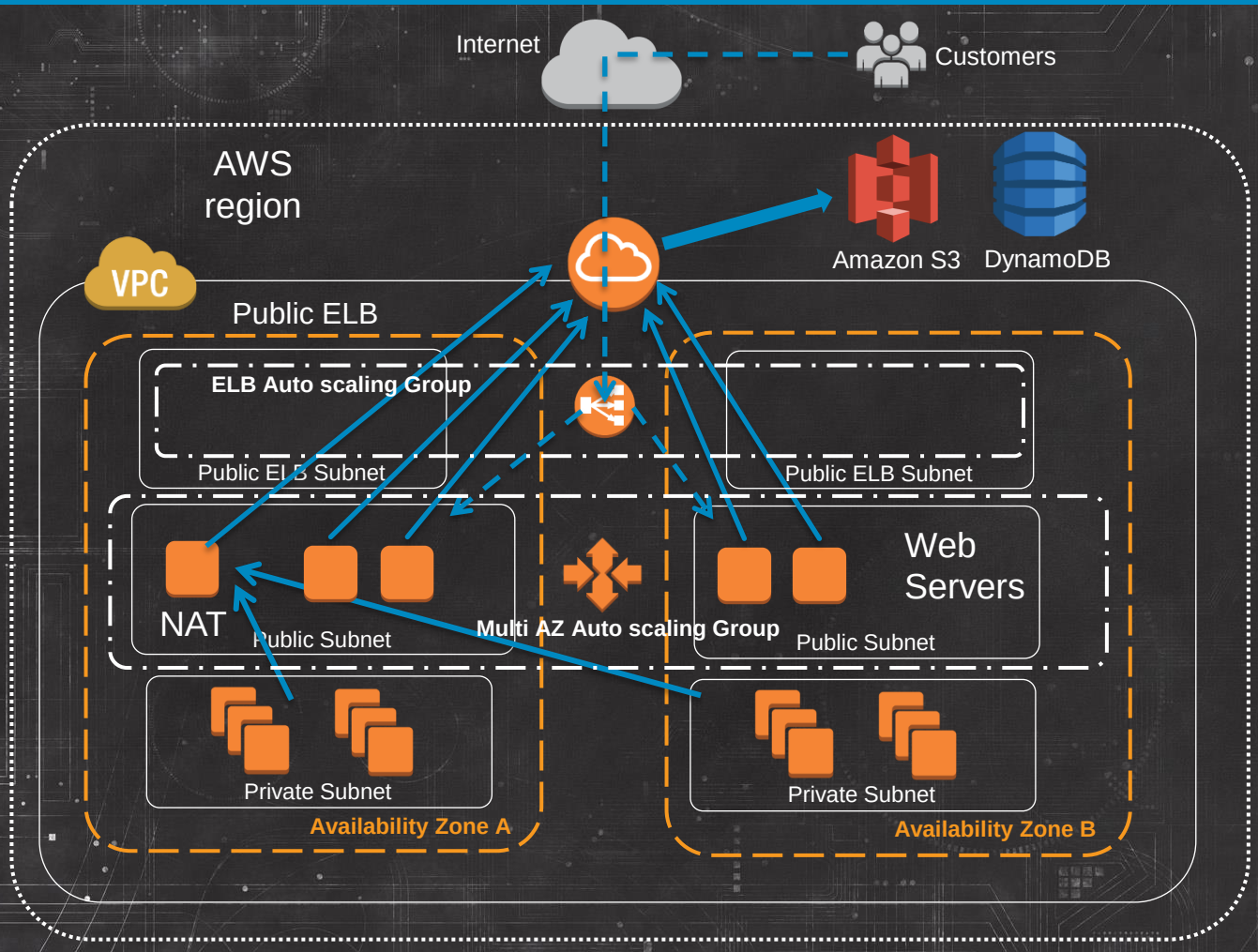


… you could reach a bandwidth bottleneck if your private instances grow and their NAT bound traffic grows with them.

Scalable and Available NAT

Do bandwidth intensive processes need to be behind a NAT?

- Separate out application components with bandwidth needs
- Run components from public subnet instances
- Goal is full instance bandwidth out of VPC
- Auto Scaling with Public IP makes this easy
- NAT still in place for remaining private instances
- Most Common use case:
  Multi-Gbps streams to Amazon S3



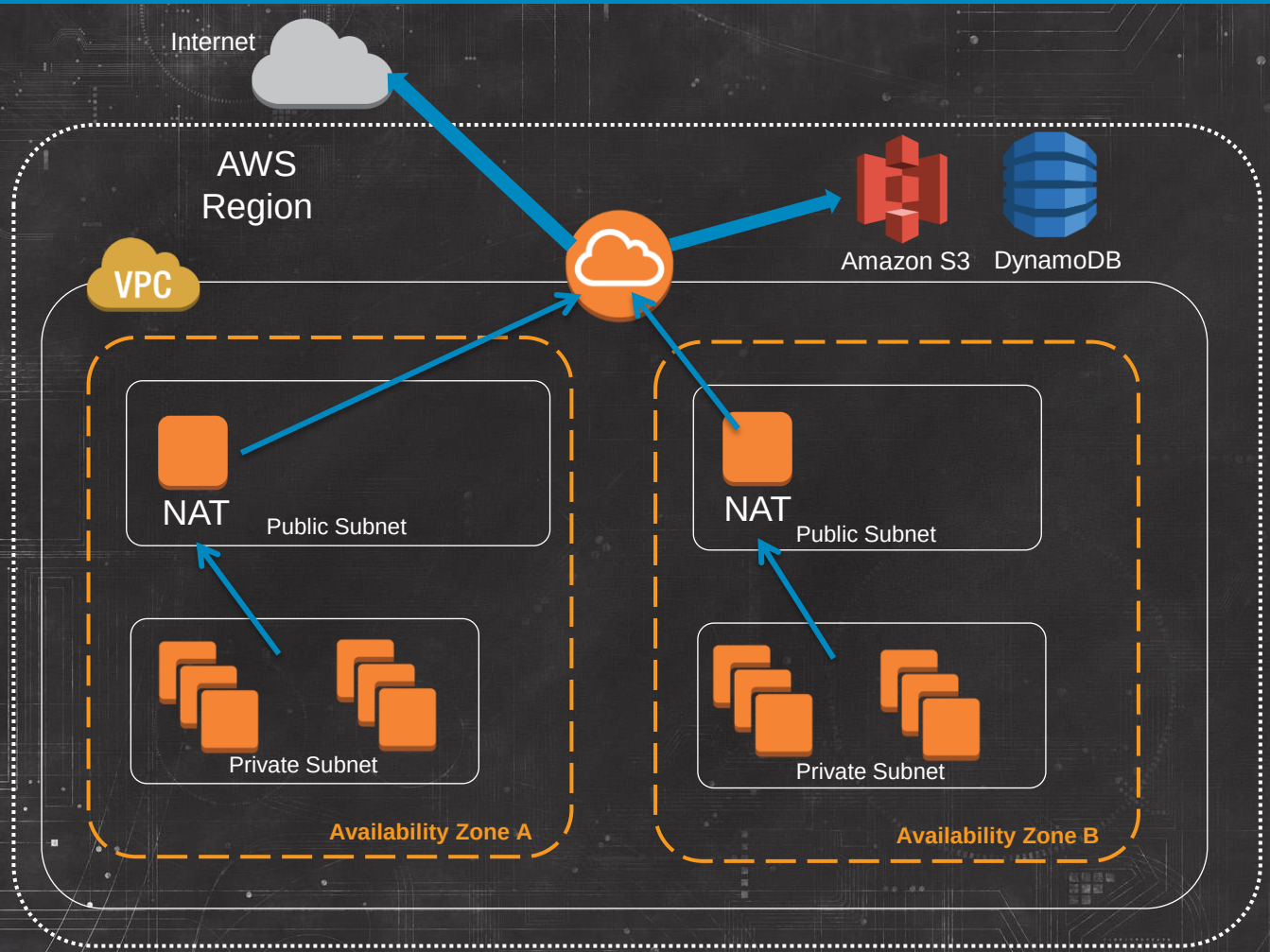
Direct to Amazon S3

- Image processing app with high outbound network to Amazon S3
- Public ELB receives incoming customer HTTP/S requests
- Auto Scaling assigns public IP to new web servers
- With public IPs, web servers initiate outbound requests directly to Amazon S3
- NAT device still in place for private subnets

Autoscaling Support for Automatic Public IP Assignment

Sample Launch Configuration (named “hi-bandwidth-public”):

```
$aws autoscaling create-launch-configuration --launch-configuration-name hi-bandwidth-
public --image-id ami-xxxxxxxx --instance-type m1.xlarge --associate-public-ip-address
```



Autoscale HA NAT

- Use Auto Scaling for NAT Availability
- Create 1 NAT per AZ
- All private subnet route tables to point to same AZ NAT
- 1 Auto Scaling group per NAT with min and max size set to 1
- Let Auto Scaling monitor the health and availability of your NATs
- If NAT fails, user data script in Autoscaling Launch config programmatically updates private subnet route tables to point to new NAT instance ID

Auto Scaling for Availability

Sample HA NAT Autoscaling group (named “ha-nat-asg”):

```
$aws autoscaling create-auto-scaling-group --auto-scaling-group-name ha-
nat-asg --launch-configuration-name ha-nat-launch --min-size 1 --max-size
1 --vpc-zone-identifier subnet-xxxxxxxx
```

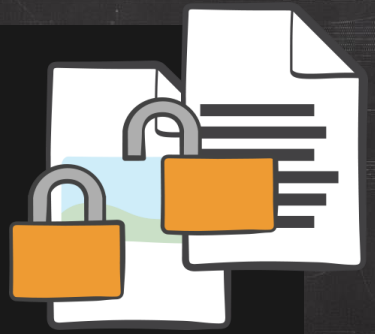
Automating HA NAT with EC2 User Data

Latest version of the HA NAT User Data script on GitHub:

https://github.com/ralex-aws/vpc

IAM EC2 Role for HA NAT Instance

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ec2:DescribeInstances",
        "ec2:ModifyInstanceAttribute",
        "ec2:DescribeSubnets",
        "ec2:DescribeRouteTables",
        "ec2:CreateRoute",
        "ec2:ReplaceRoute"
      ],
      "Resource": "*"
    }
  ]
}
```



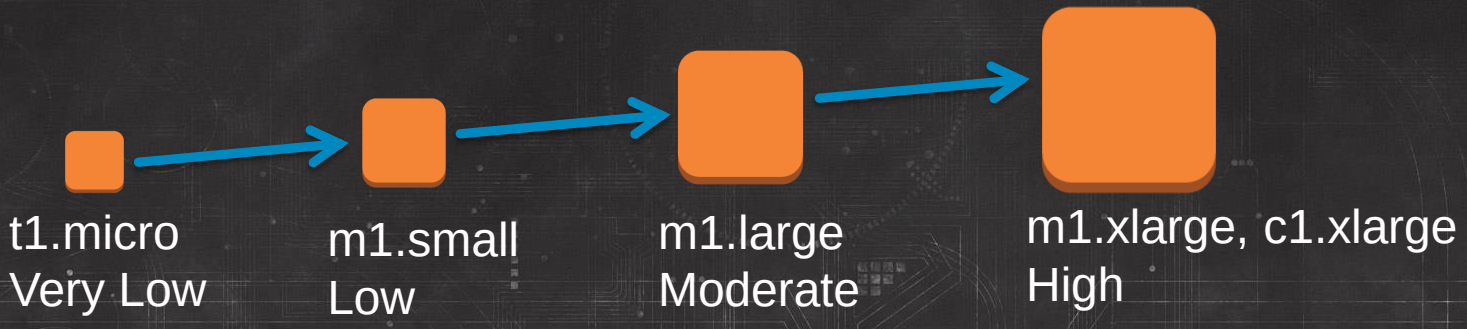
VPC Tip 6

Tag Early, Tag Often!

- Tagging strategy should be part of early design
- Project Code, Cost Center, Environment, Team, Business Unit
- Tag resources right after creation
- Tags supported for resource permissions
- AWS Billing also supports tags
- Tight IAM controls on the creation and editing of tags

Finally, if design requirements keep high bandwidth streams behind NAT:

- Use the 1 HA NAT per AZ design
- Vertically scale your NAT instance type to one with a High Network Performance rating
- Keep a close watch on your network metrics



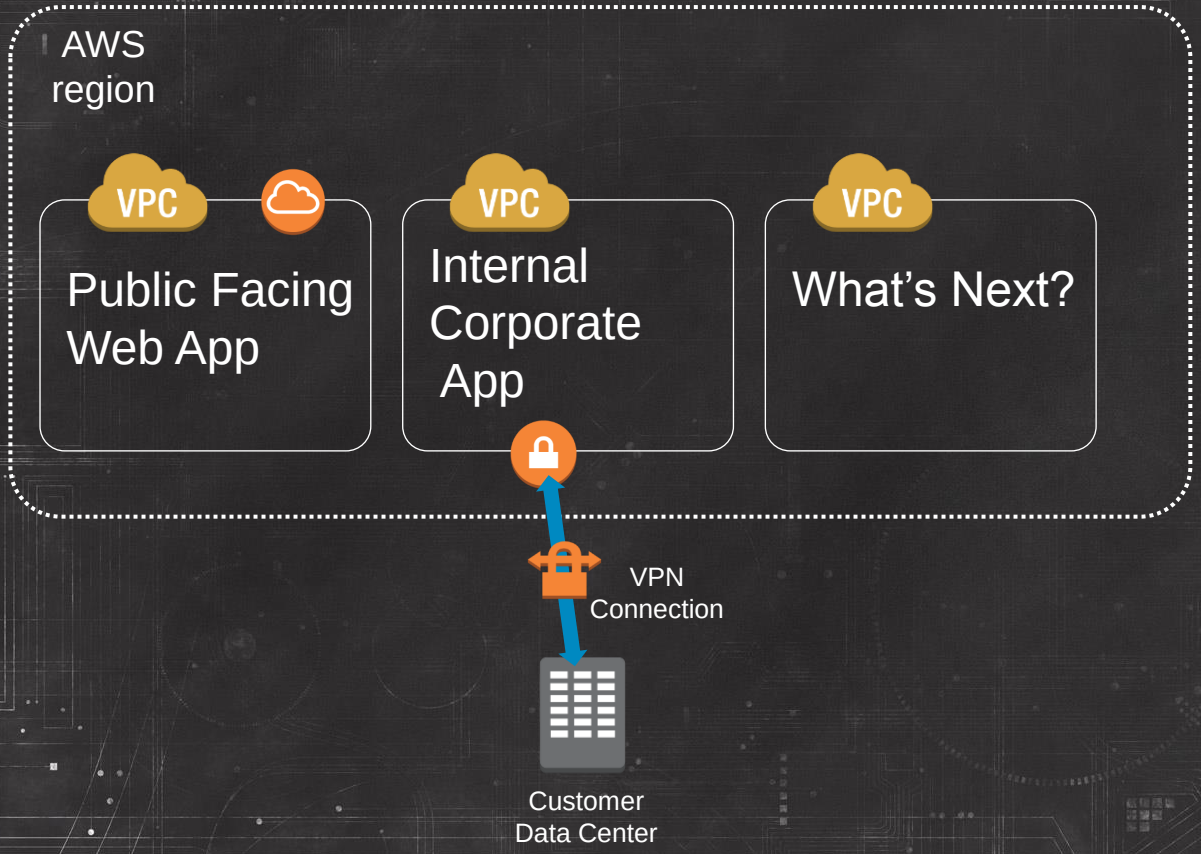
For further HA NAT design alternatives, please see:

High Availability for Amazon VPC NAT Instances
http://aws.amazon.com/articles/2781451301784570

Deploy HA NAT With Cloudformation Templates:
http://aws.amazon.com/articles/6079781443936876

One VPC, Two VPC

Considering Multiple VPCs



- Public Facing Web App deployed in own VPC
- Now want to deploy an internal only Corporate App connected to Corporate Datacenter via VPN
- New VPC created in the Region for Corporate app to keep the external and internal applications isolated from each other

Common Multi-VPC Customer Use Cases:

- Application isolation
- Scope of audit containment
- Risk level separation
- Separate production from non-production
- Multi tenant isolation
- Business unit alignment

Considerations for One or Many VPCs:

- Know your inter-VPC traffic
- Separate AWS accounts by definition means separate VPCs
- IAM / resource permissions and controls
- VPC limits:
  http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_Appendix_Limits.html

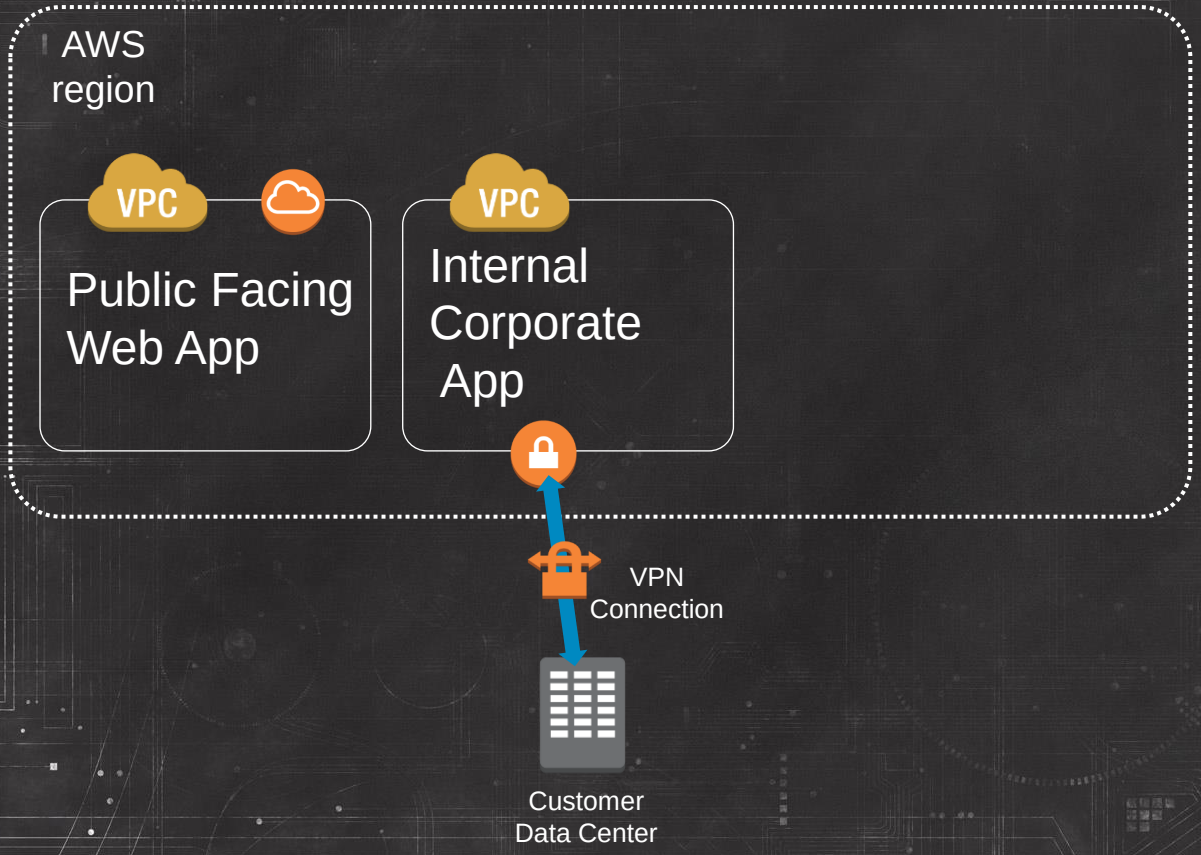
There is a whole talk on this one!

CPN208

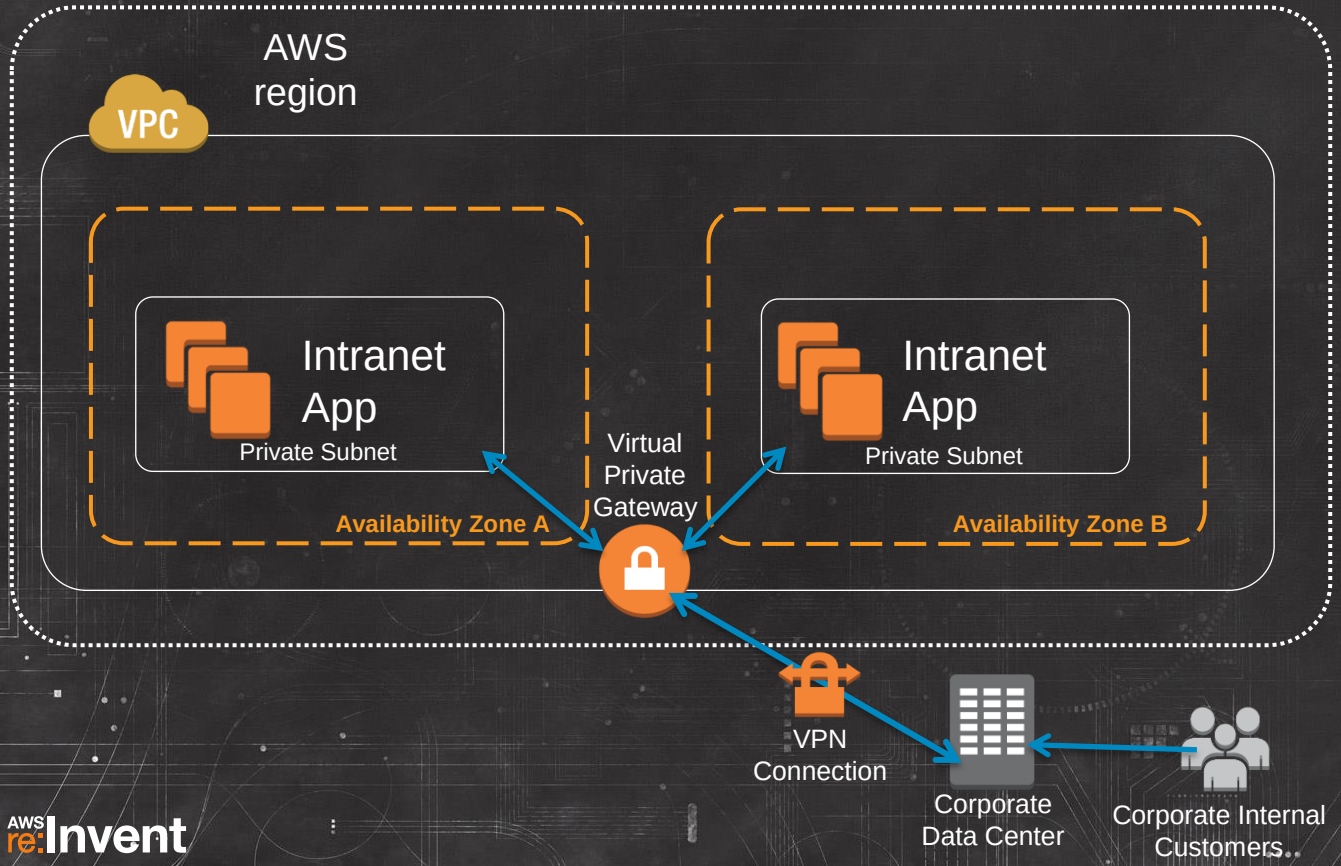
Selecting the Best VPC Network Architecture

Controlling the Border

Internal Application to VPC



Internal Application to VPC

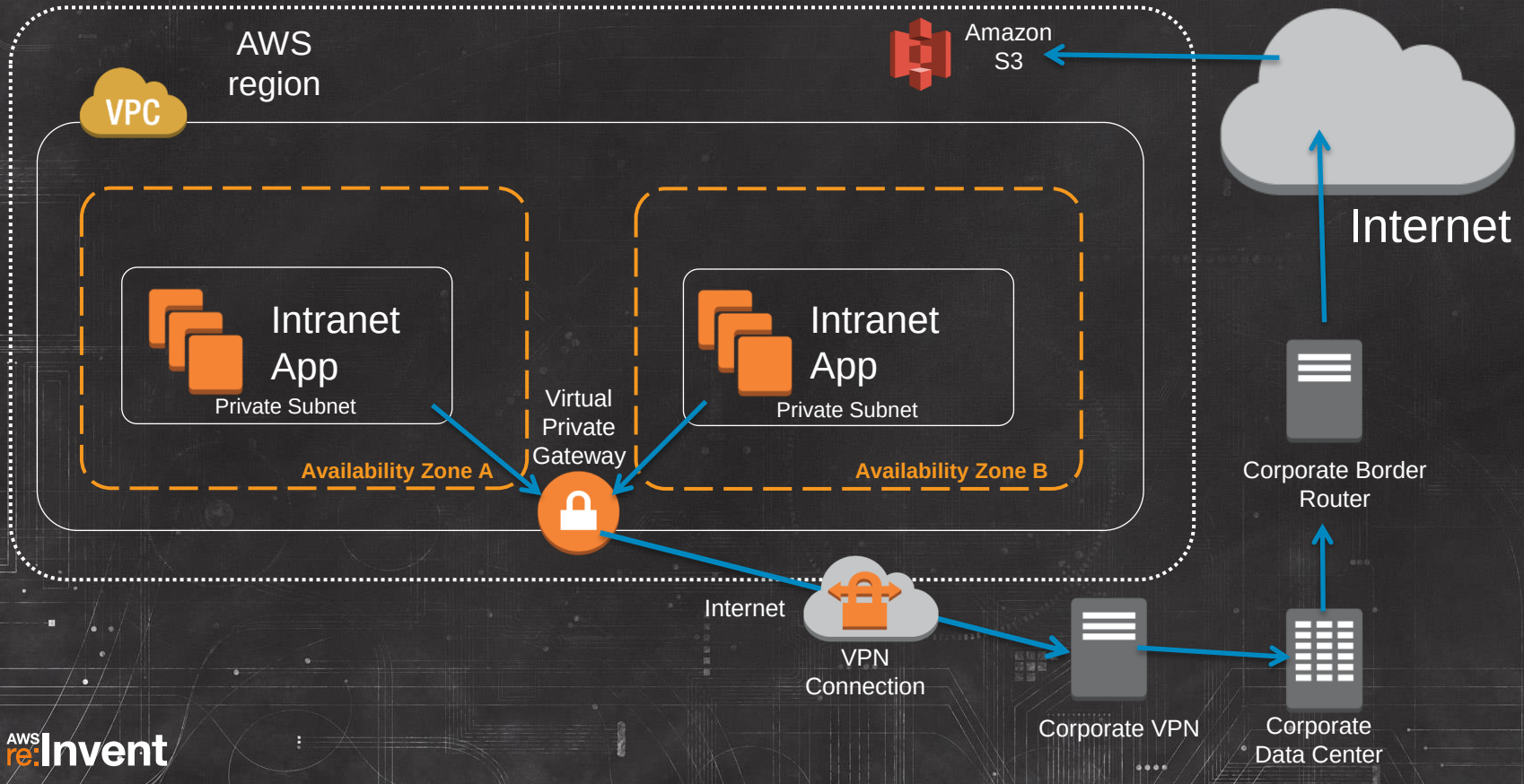


Route Table	
Destination	Target
10.1.0.0/16	local
Corp CIDR	VGW

But… app will leverage this for storing data:

Amazon S3

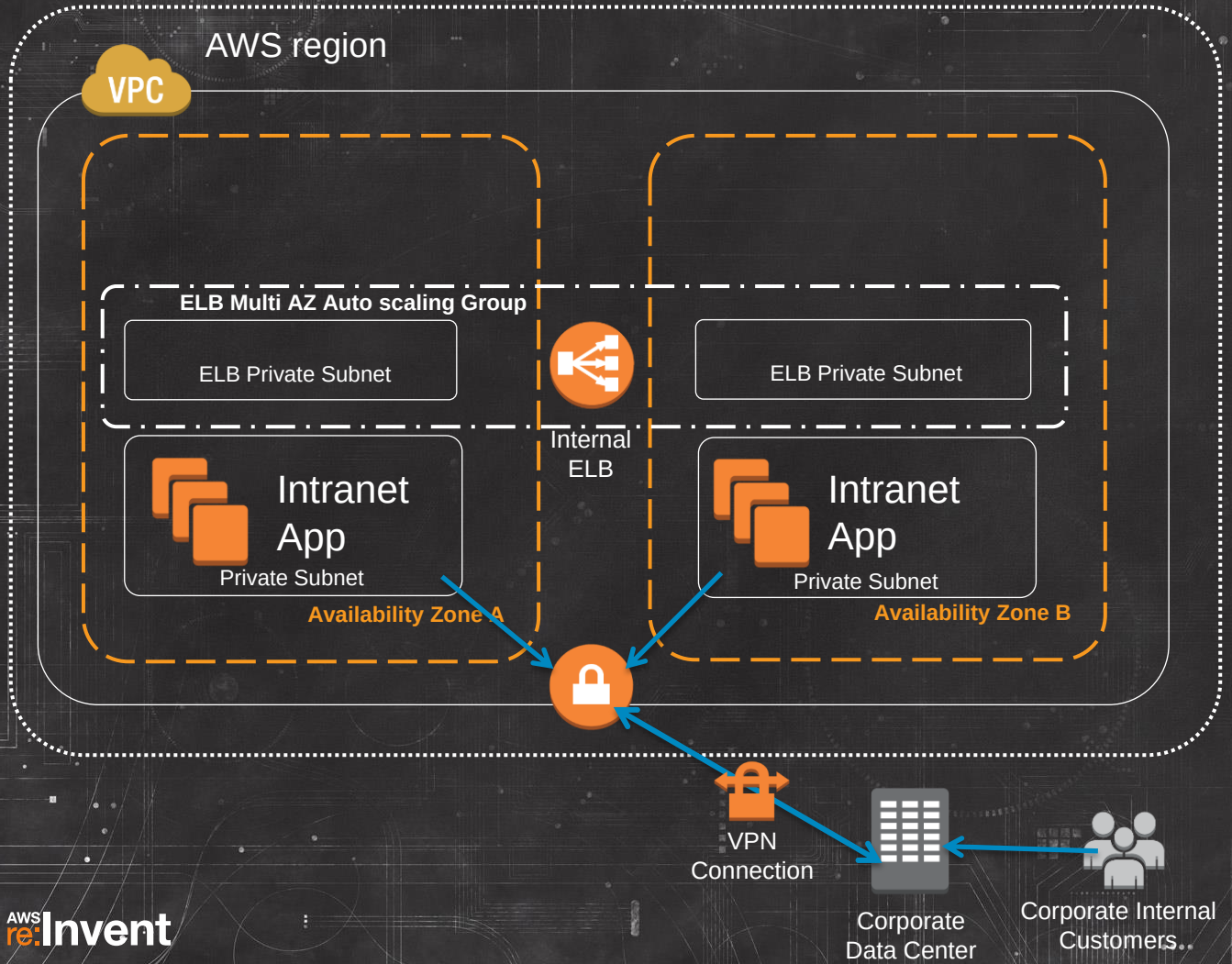
And you don't really want to do this:



Control IGW Access through a Proxy Layer

- Deploy a proxy control layer between application and IGW
- Restrict all outbound HTTP/S access to only approved URL destinations like Amazon S3
- No route to IGW for private subnets
- Control access to proxy through security groups
- Must configure proxy setting in OS of instances

Controlling the Border

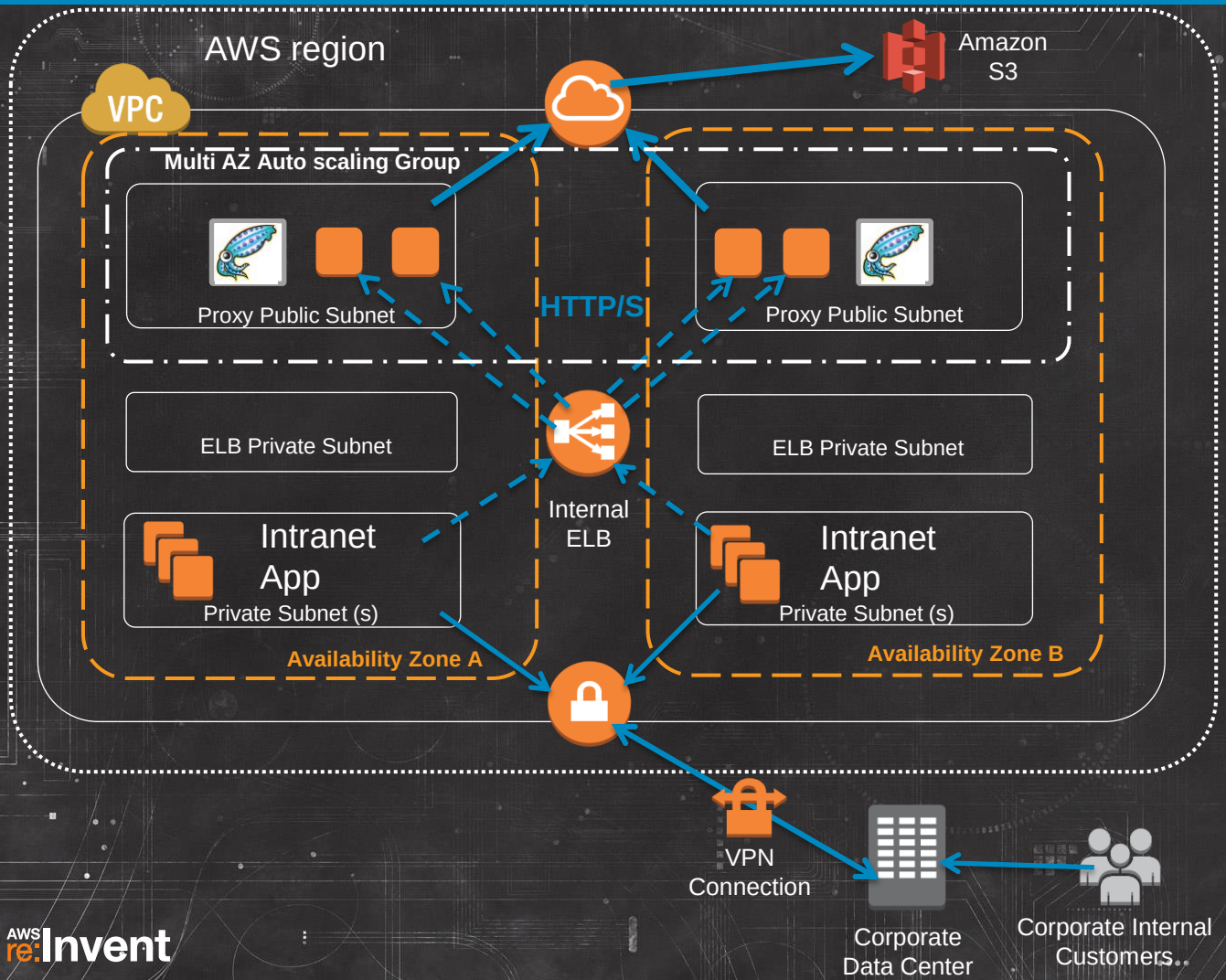


- Deploy internal ELB layer across AZs
- Add all instances allowed outside access to a security group
- Use this security group as the only source allowed access to the proxy port in the load balancer's security group

VPC Tip 7

Put load balancers in their own subnets

- Elastic Load Balancing is Amazon EC2 in your subnets
- Elastic Load Balancing is using your private addresses
- Separate subnets = separate control
- Distinguish LB layer from app layers



Controlling the Border

- Squid Proxy layer deployed between internal load balancer and the IGW border.
- Only proxy subnets have route to IGW.
- Proxy security group allows inbound only from Elastic Load Balancing security group.
- Proxy restricts which URLs may pass. In this example, s3.amazonaws.com is allowed.
- Egress NACLs on proxy subnets enforce HTTP/S only.

Squid.conf Sample Config:

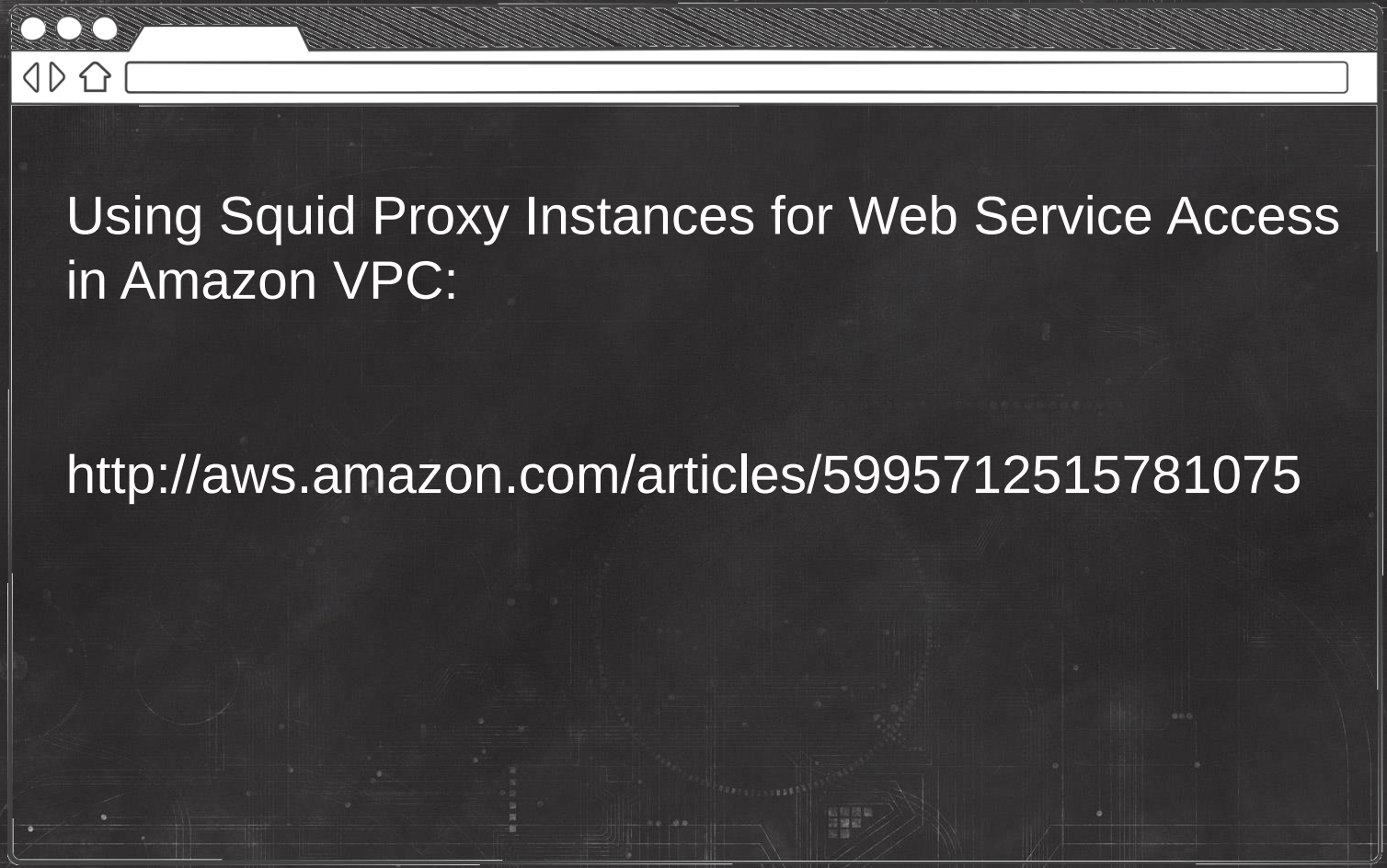
```
# CIDR AND Destination Domain based Allow

# CIDR Subnet blocks for Internal ELBs
acl int_elb_cidrs src 10.1.3.0/24 10.1.4.0/24

# Destination domain for target S3 bucket
acl s3_v2_endpoints dstdomain $bucket_name.s3.amazonaws.com

# Squid does AND on both ACLs for allow match
http_access allow int_elb_cidrs s3_v2_endpoints

# Deny everything else
http_access deny all
```



Using Squid Proxy Instances for Web Service Access in Amazon VPC:

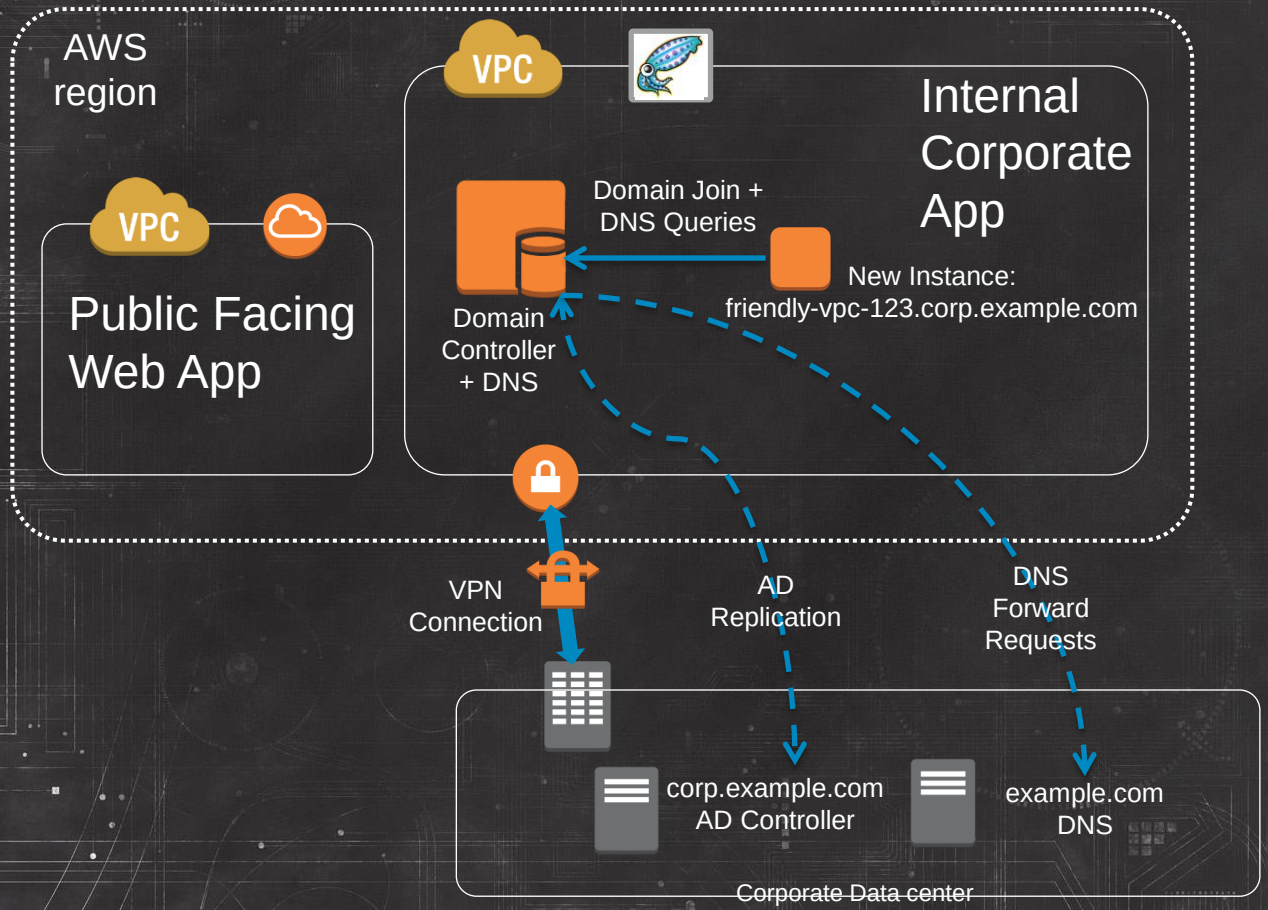
http://aws.amazon.com/articles/5995712515781075

… and this design could also be an option to our earlier NAT bandwidth discussion if outbound traffic requirements are HTTP only.

Directory and Name Services in the VPC

…or what do you mean ip-10-1-1-57.us-west-2.compute.internal isn’t a “friendly” name?

Active Directory + DNS in the VPC



- Domain Controllers launched in internal VPC
- Internal VPC instances join domain upon launch
- Instances use Dynamic DNS to register both A and PTR records
- Domain controller replicates with Corporate AD servers
- VPC DNS forwarding to corporate DNS

DNS in the VPC

- Enable automatic DNS hostname creation and resolution with these 2 options:

- Automatic hostname creation
- Private only instances assigned private hostname
- Public instances assigned public and private

Split DNS Resolution

Example hostnames for Public VPC instance:

ec2-54-200-171-240.us-west-2.compute.amazonaws.com

ip-10-1-1-87.us-west-2.compute.internal

From outside VPC:

```
a82066136617:~ ralex$ nslookup ec2-54-200-171-240.us-west-2.compute.amazonaws.com
Server:         192.168.1.1
Address:        192.168.1.1#53

Non-authoritative answer:
Name:   ec2-54-200-171-240.us-west-2.compute.amazonaws.com
Address: 54.200.171.240
```

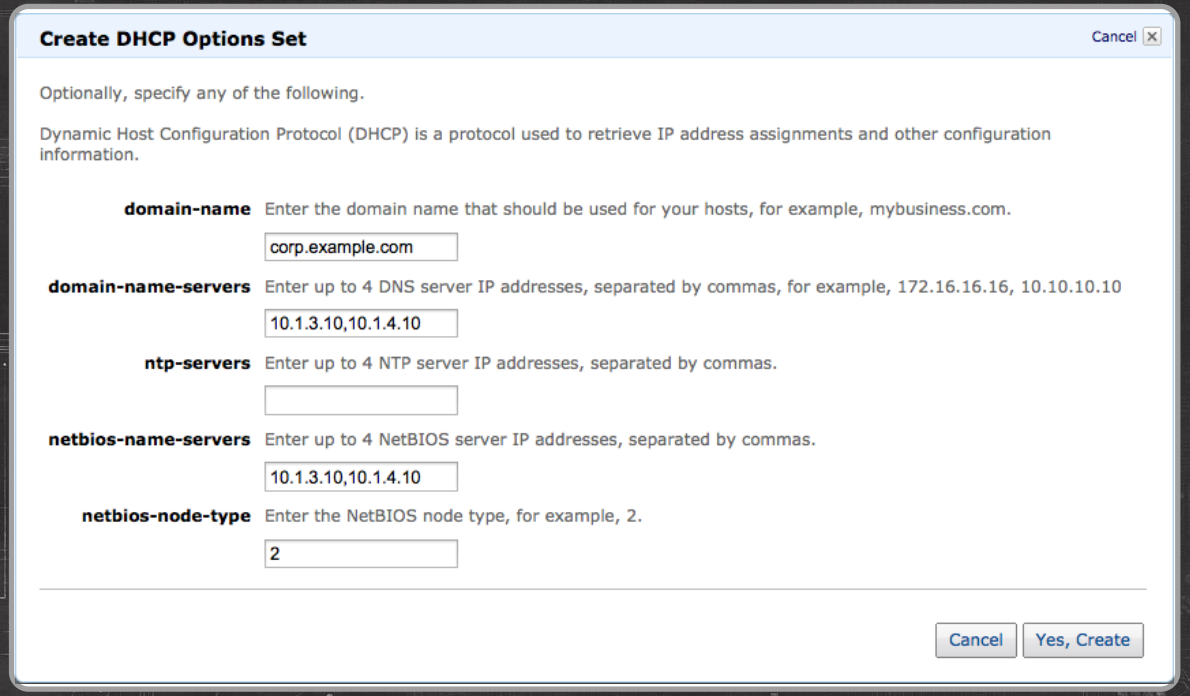
From inside VPC:

```
[ec2-user@ip-10-1-1-199 ~]$ nslookup ec2-54-200-171-240.us-west-2.compute.amazonaws.com
Server:         10.1.0.2
Address:        10.1.0.2#53

Non-authoritative answer:
Name:   ec2-54-200-171-240.us-west-2.compute.amazonaws.com
Address: 10.1.1.87
```

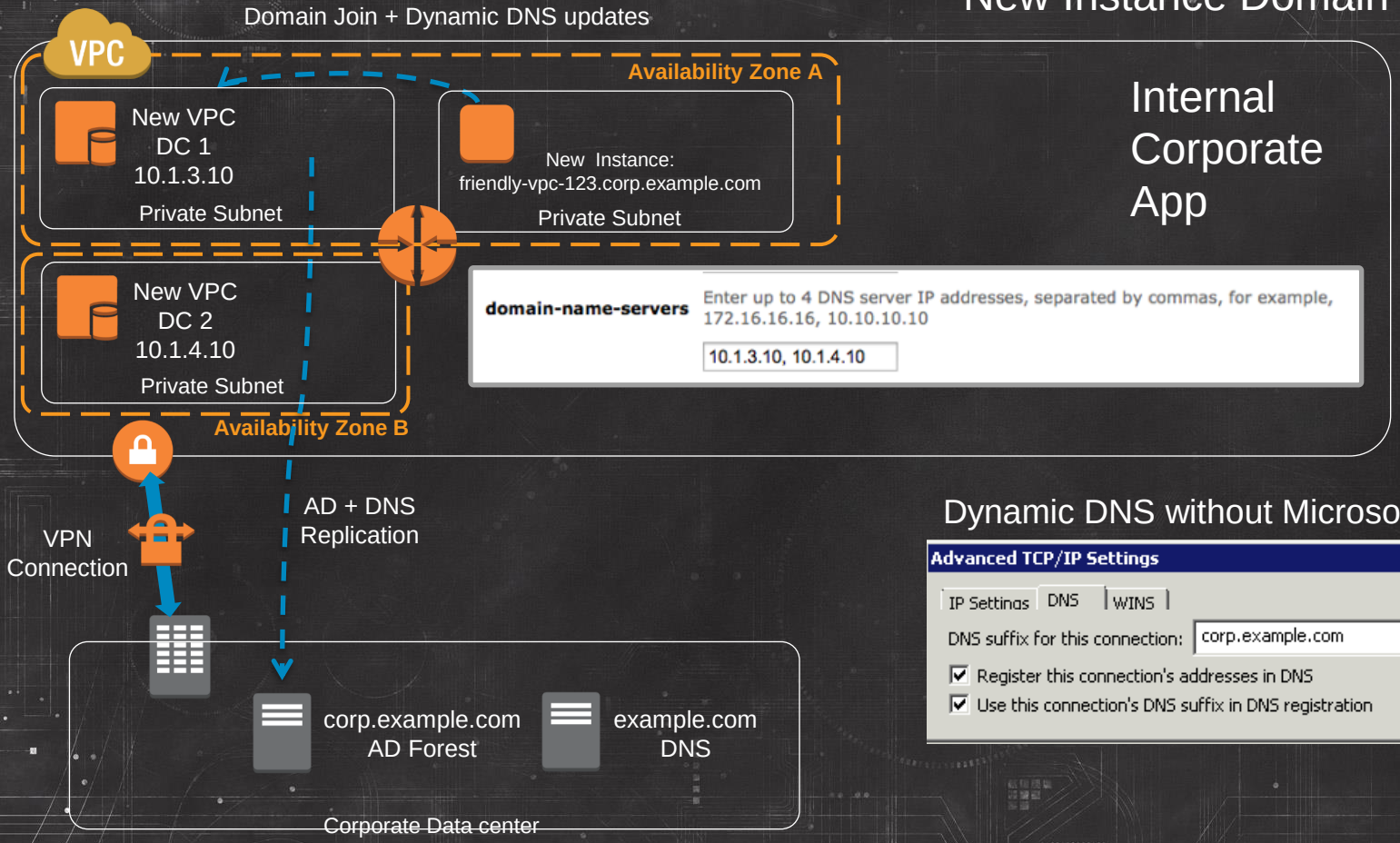
- Private hostnames only resolvable within VPC
- Public hostnames will resolve to private IP addresses within the VPC
- 10.1.0.2 represents the VPC Virtual DNS Service and will always take the .2 address of your VPC CIDR block
- VPC Virtual DNS Service is also called “AmazonProvidedDNS” and enables instances in a VPC to resolve public DNS names

DHCP Option Sets



- Not possible to replace the VPC DHCP service with your own
- But it is possible to customize what VPC DHCP hands out
- Default option set only contains DNS = “AmazonProvidedDNS”
- 1 option set assigned per VPC
- Changing option set dynamically applies the next time an instance requests a lease refresh

New Instance Domain Registration

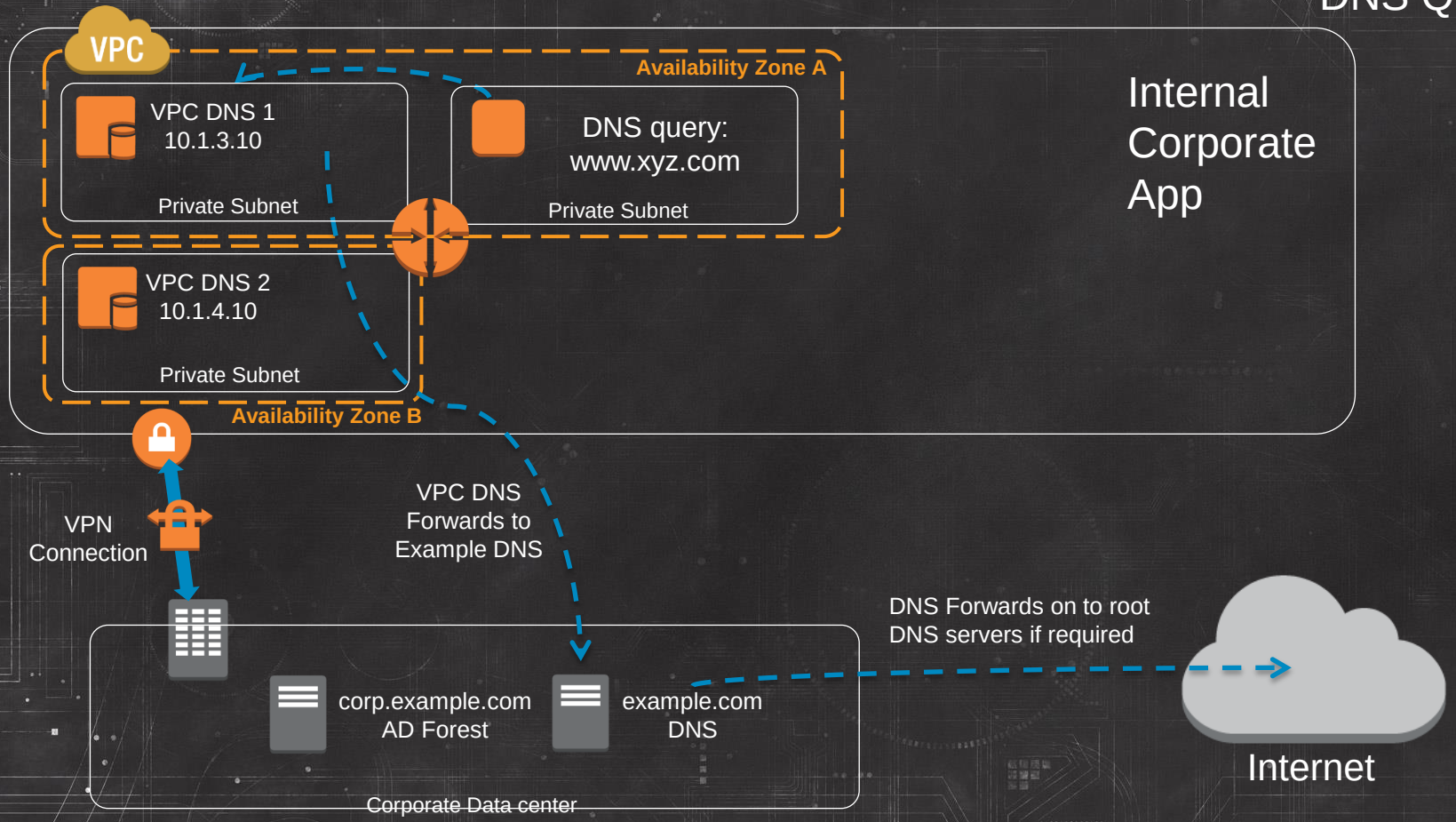


Dynamic DNS without Microsoft DHCP:

Sample Powershell user-data for AD instance add and rename:

```
<powershell>
$secstr = convertto-securestring -string "Password123" -AsPlainText -Force
$cred = new-object -typename System.Management.Automation.PSCredential -argumentlist domain-add, $secstr
$instanceId = (Invoke-WebRequest -Uri http://169.254.169.254/latest/meta-data/instance-id).Content
$servername = (Get-EC2Tag -Region us-west-2 | Where-Object {$_.ResourceId -eq $instanceId -and $_.Key
-eq "Name"}).Value
Add-Computer -DomainName "corp.example.com" -NewName $servername -Credential $cred -Restart
</powershell>
```

DNS Query Path



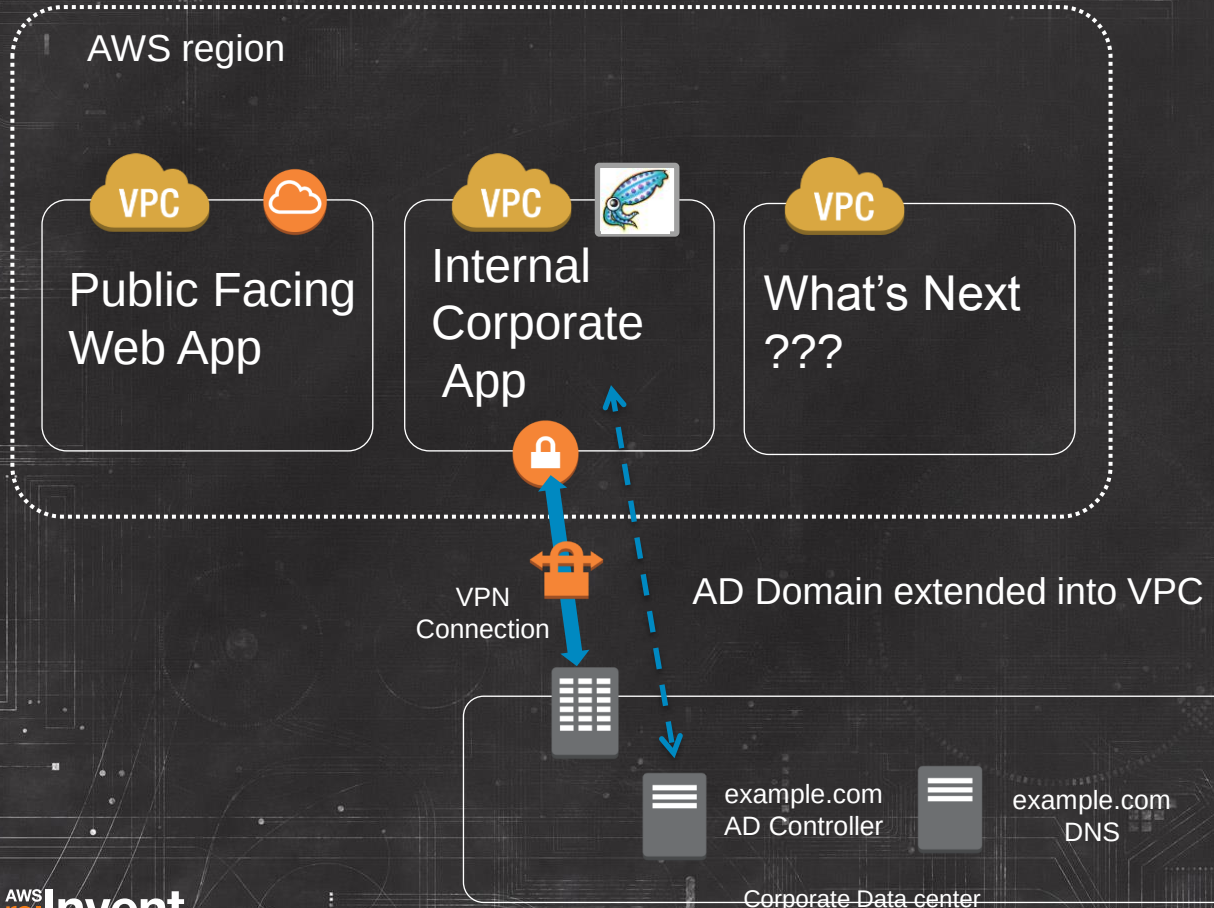
For AWS CloudFormation templates and guides to setting up Microsoft AD domains in VPC, please see:

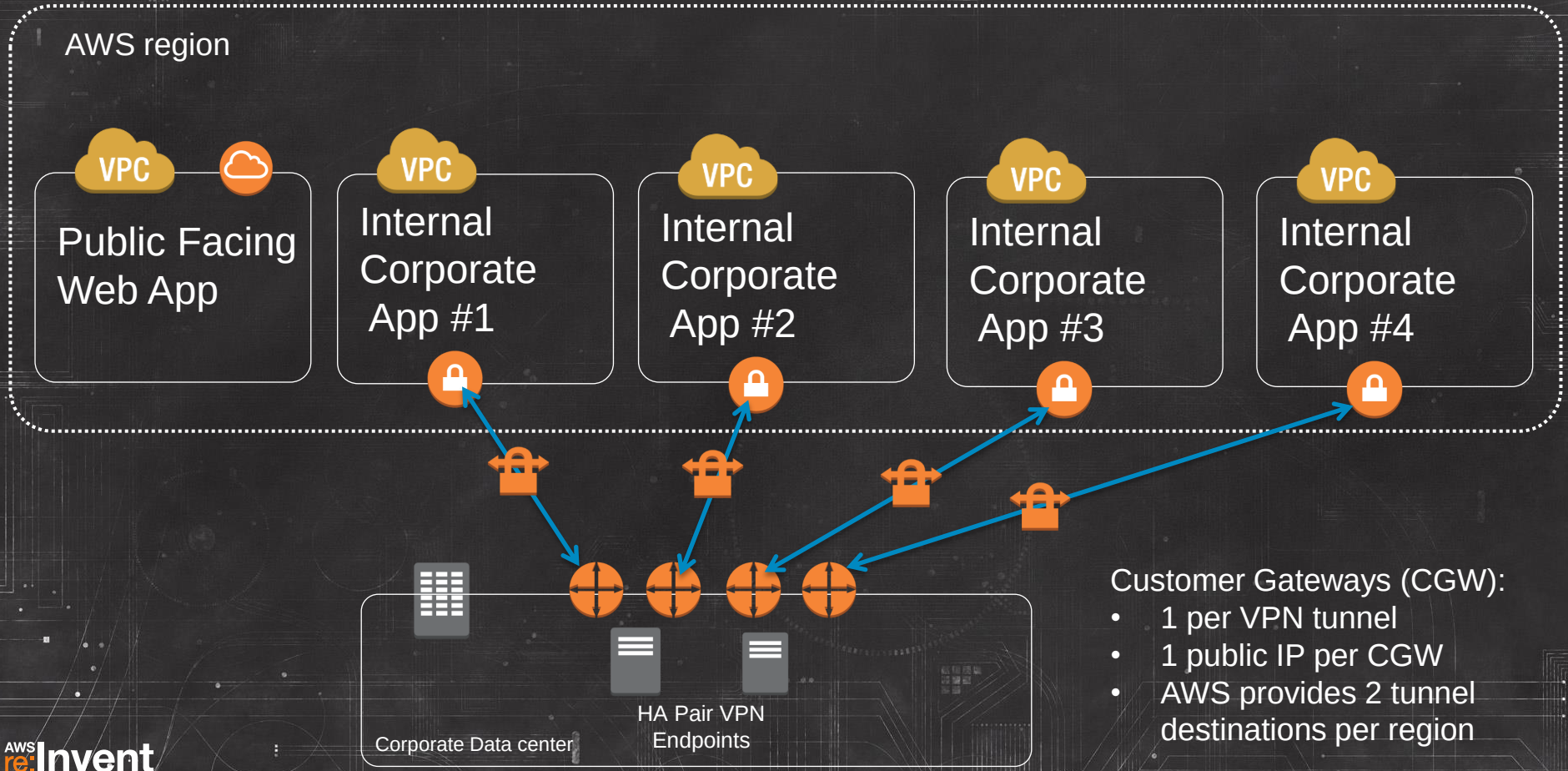
Deploy a Microsoft SharePoint 2010 Server Farm in the AWS Cloud in 6 Simple Steps:
http://aws.amazon.com/articles/9982940049271604

Implementing Microsoft Windows Server Failover Clustering (WSFC) and SQL Server 2012 AlwaysOn Availability Groups in the AWS Cloud
http://aws.amazon.com/whitepapers/microsoft-wsfc-sql-alwayson/

Microsoft Exchange Server 2010 in the AWS Cloud: Planning & Implementation Guide:
http://media.amazonwebservices.com/AWS_Exchange_Planning_Implementation_Guide.pdf

Bringing It All Back Home

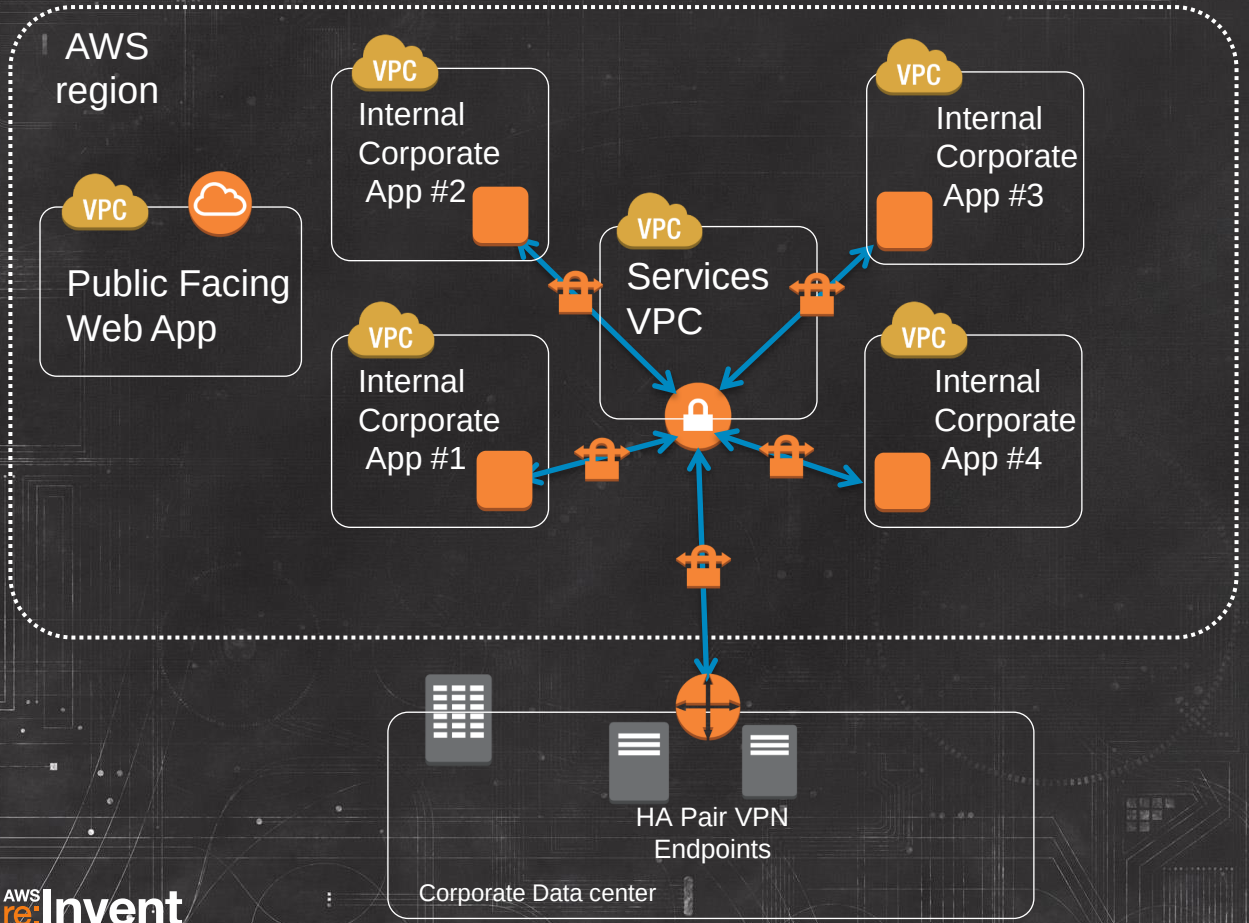

AWS region
VPC
Public Facing Web App
VPC
Internal Corporate App
VPC
What's Next ???
VPN Connection
AD Domain extended into VPC
example.com AD Controller
example.com DNS
Corporate Data center



Customer Gateways (CGW):

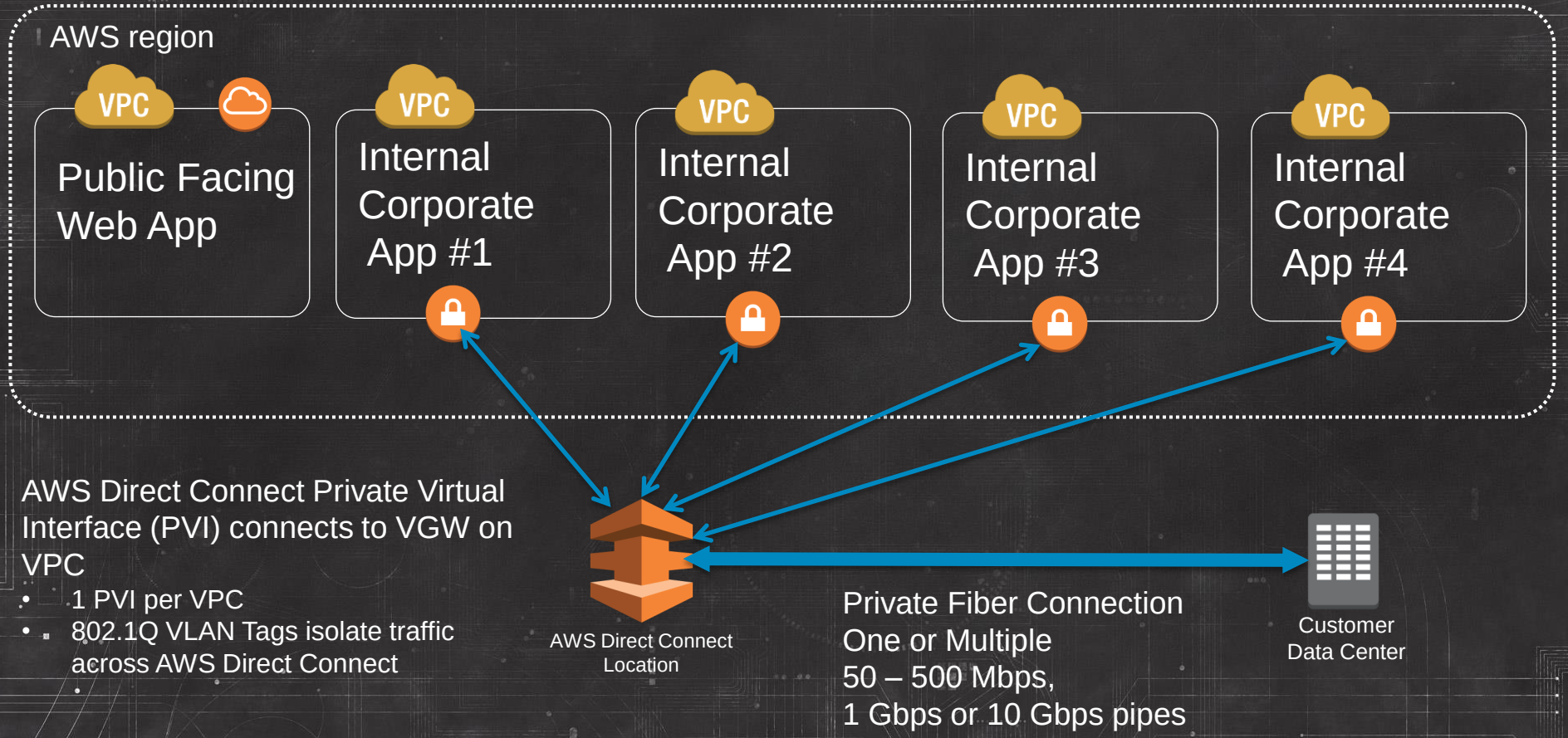
- 1 per VPN tunnel
- 1 public IP per CGW
- AWS provides 2 tunnel destinations per region

VPN Hub and Spoke an option…



- Amazon EC2 VPN instances to central virtual private gateway
- For HA, 2 Amazon EC2-based VPN endpoints in each spoke
- Control VPC contains common services for all app VPCs
- Dynamic Routing protocol (BGP, OSPF) between Spokes and Hub
- If multi Gbps traffic flow to Corporate Datacenter, then IPSec tunnels could become bandwidth bottleneck

… or simplify with AWS Direct Connect



AWS Direct Connect Private Virtual Interface (PVI) connects to VGW on VPC

- 1 PVI per VPC
- 802.1Q VLAN Tags isolate traffic across AWS Direct Connect

Private Fiber Connection
One or Multiple
50 – 500 Mbps,
1 Gbps or 10 Gbps pipes

A few bits on AWS Direct Connect…

- Dedicated, private pipes into AWS
- Create private (VPC) or public interfaces to AWS
- Cheaper data out rates than Internet (data in still free)
- Consistent network performance compared to Internet
- At Least 1 location to each AWS region (even GovCloud!)
- Recommend redundant connections
- Multiple AWS accounts can share a connection

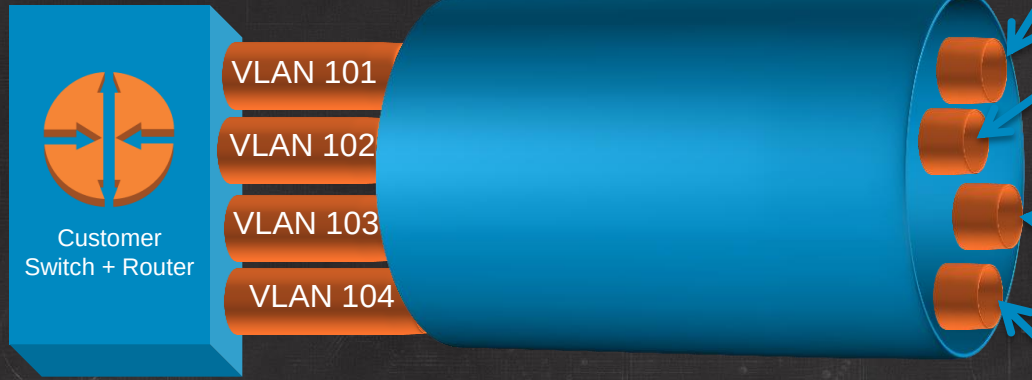
Multiple VPCs Over AWS Direct Connect

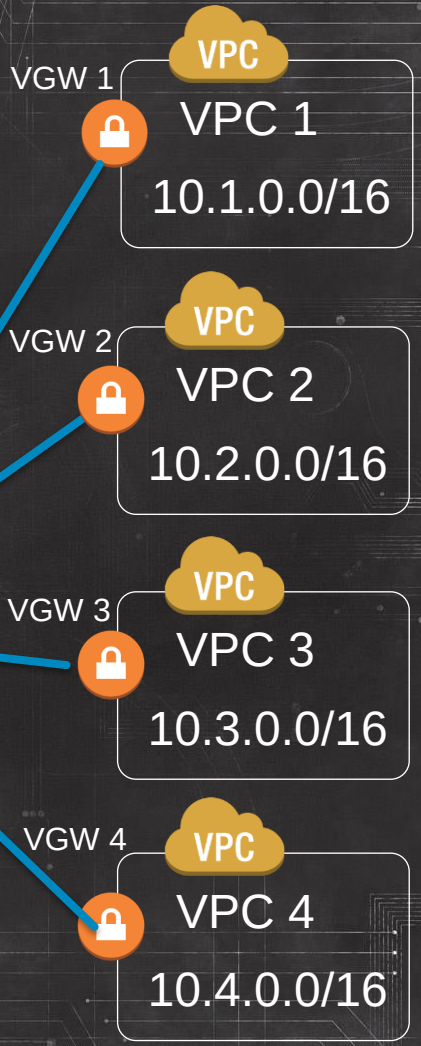
Customer Interface 0/1.101	
VLAN Tag	101
BGP ASN	65001
BGP Announce	Customer Internal
Interface IP	169.254.251.6/30

Private Virtual Interface 1	
VLAN Tag	101
BGP ASN	7224
BGP Announce	10.1.0.0/16
Interface IP	169.254.251.5/30

Customer Internal Network

Route Table	
Destination	Target
10.1.0.0/16	PVI 1
10.2.0.0/16	PVI 2
10.3.0.0/16	PVI 3
10.4.0.0/16	PVI 4

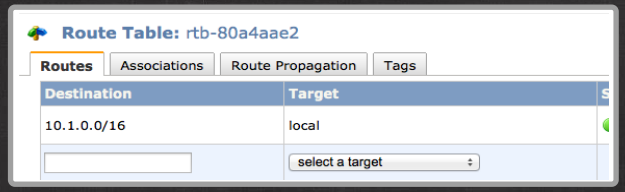




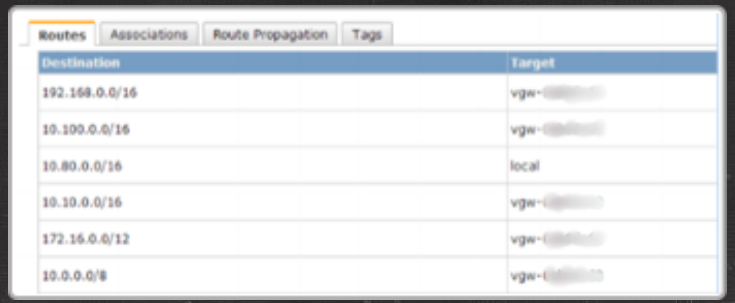
VPC Tip 8

Know Your Routing Database

Before:



After:



Enable:

VPC Tip 8

Know Your Routing Database

- Keep track of all incoming BGP announcements into your VPCs
- Create a new Routing Table, unassigned to any subnet, and enable Route Propagation
- Routing Table will show all routes the VGW has learned through BGP announcements
- See what the VGW sees

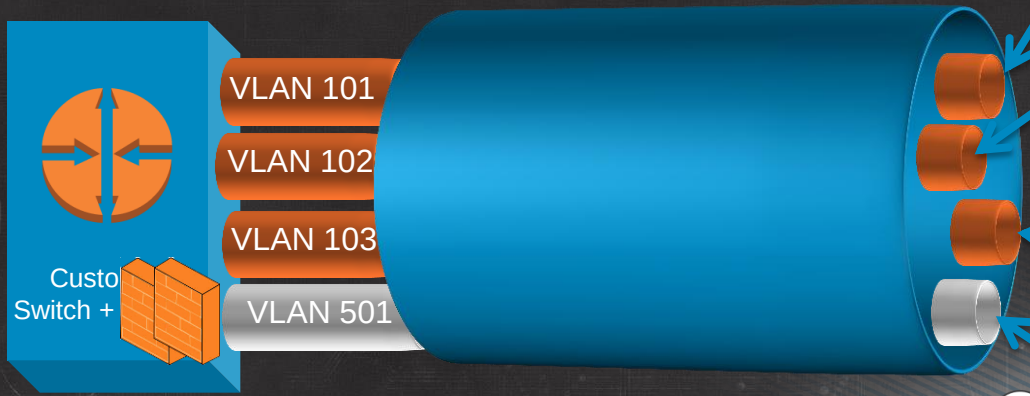
Public AWS + VPCs Over AWS Direct Connect

Customer Interface 0/1.501	
VLAN Tag	501
BGP ASN	65501 (or Public)
BGP Announce	Customer Public
Interface IP	Public /30 Provided

Public Virtual Interface 1	
VLAN Tag	501
BGP ASN	7224
BGP Announce	AWS Regional Public CIDRs
Interface IP	Public /30 Provided

Customer Internal Network

Route Table	
Destination	Target
10.1.0.0/16	PVI 1
10.2.0.0/16	PVI 2
10.3.0.0/16	PVI 3
10.4.0.0/16	PVI 4
Public AWS	PVI 5



NAT / PAT Security Layer

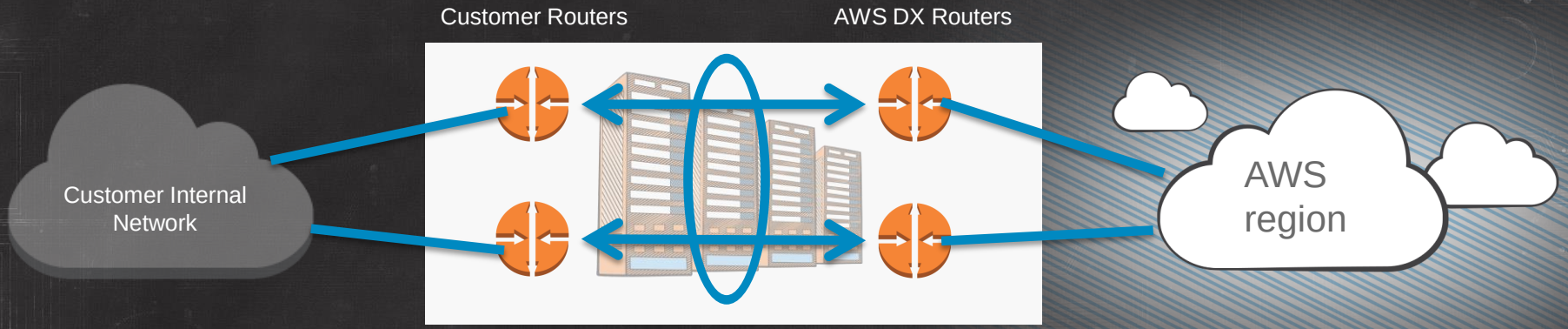
VGW 1
VPC
VPC 1
10.1.0.0/16

VGW 2
VPC
VPC 2
10.2.0.0/16

VGW 3
VPC
VPC 3
10.3.0.0/16

Public AWS Region

AWS Direct Connect Location



Multiple Physical connections:

- Active / Active links via BGP multi-pathing
- Active / Passive also an option
- BGP MEDs or local preference can influence route
- Bidirectional Forwarding Detection (BFD) supported

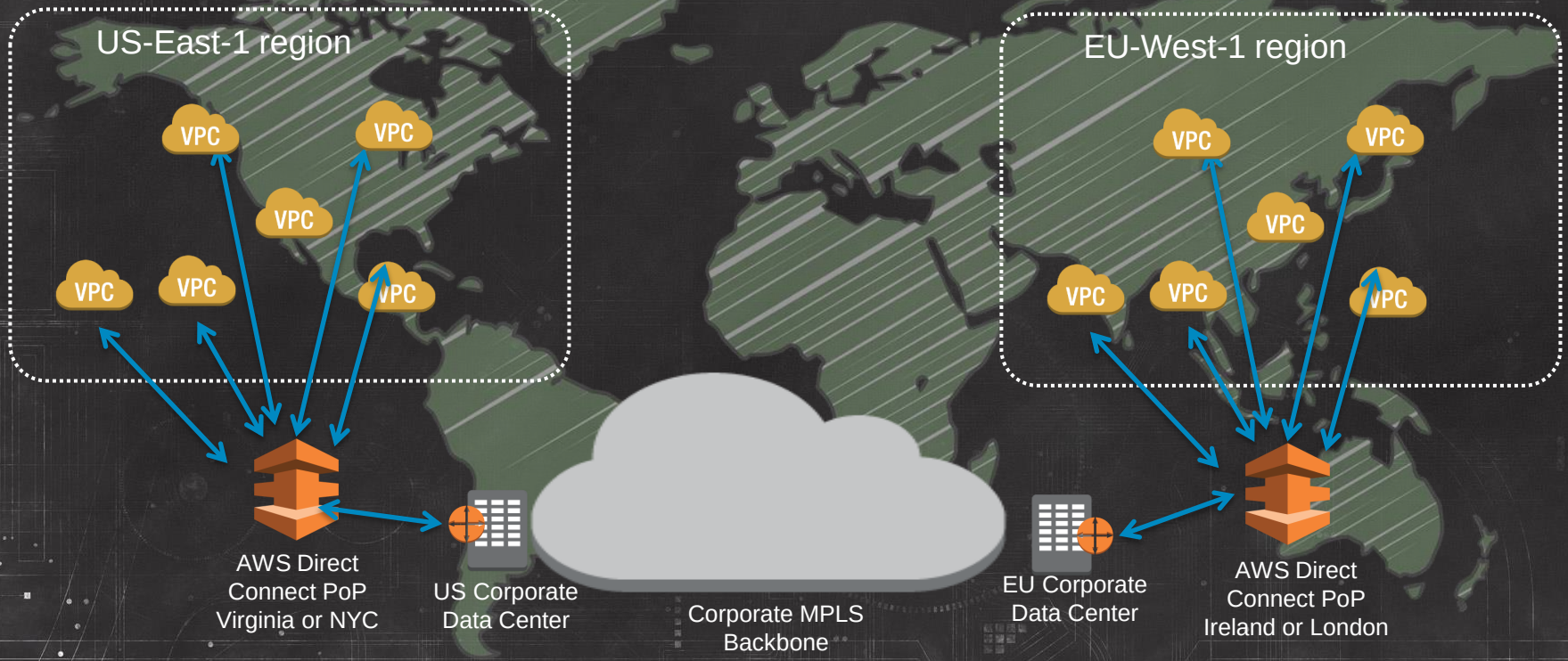
Going Global

US-East-1
AWS
region

Customer Global
MPLS Backbone
Network

EU-West-1
AWS
region

With AWS regions just another spoke on your global network,
it's easy to bring the cloud down to you as you expand around the world.



Evolving VPC Design: Recap

- Elements of VPC Design
- Scalable and Available NAT
- One VPC, Two VPC
- Controlling the Border
- Directory and Name Services in the VPC
- Bringing It All Back Home

Related re:Invent Sessions:

ARC202 High Availability Application Architectures in Amazon VPC

ARC304 Cloud Architectures with AWS Direct Connect

CPN205 Securing Your Amazon EC2 Environment with AWS IAM Roles and Resource-Based Permissions

CPN208 Selecting the Best VPC Network Architecture

DMG201 Zero to Sixty: AWS CloudFormation

DMG303 AWS CloudFormation under the Hood

AWS re:Invent

Please give us your feedback on this presentation

ARC401

As a thank you, we will select prize winners daily for completed surveys!

amazon web services