AWS_SOLUTION_ARCHITECT_NOTES

ec2 is the elastic compute cloud.
Lambda provides serverless functions. Provide function to Lambda and runs when some action done on webpage.
Lightsail provision server with fixed ip address rdp access for windows and ssh access for Linux. Management console to manage server. worry about underlying system.
Batch for batch computing in cloud but not convered on certificate.
S3 oldest service storage. simple storage service. Upload to "buckets" in cloud.
EFS is elastic file system. Network attached storage. Store files on efs files and mount to multiple virtual machines.
Glacier archive data storage. Cheap.
Snowball brings large amounts of data into datacenter. Write physically to disk send to aws datacenter and import manually. Instead of transferring TBs over network.
Storage gateway are virtual appliances. Vms install in data center/Head office and replicate to S3.
RDS relational database service. MySQL, Aurora, Oracle, Relational Databases sit here.
DynamoDB non-relational databases.
Elasticache is a way of caching commonly queried things from database server (top 10 products of your store, etc.)
Red shift is data warehousing/ Business Intelligence. Complex queries. Instead of doing on production database, use REd Shift, used for data warehosuing.
AWS migration hub is a tracking service. for visualing migration
Application Discovery Service automation set of tools. what apps you have along with dependencies they have. (depending on domain controller, etc)
Database Migration Service migrate databases from on premise to AWS.
Server Migration Service migrate servers to AWS cloud.
Snowball also helps with Migration. (migrating large amounts of data TBs to the cloud)
VPC virtual private cloud. Virtual datacener. configure firewalls, network route table, ACLs, Understand VPC for Exams. KEY. Again, vpc-virtual private cloud.
CloudFront is amazon content delivery network. Video and image files, stores in a edge location closer to users.
Route53 is Amazons DNS service. Resolving names to IPv4/6 addresses.
API Gateway create api for services to talk to.
Direct Connect running a dedicated line from HO to Amazon to connect to VPC.
CodeStar way of project managing code, way of collaborating with developers.
CodeCommit way of storing code.
CodeBuild once code ready will compile/test against it and produce software packages ready to deploy.
CodeDeploy , deployment service and automates app deployments to EC2 instances/on-premise instance/Lambda instances.
Codepipeline , continuous delivery service to automate software.
X-Ray analyze software.
Cloud9 IDE integrated development environment, develop code in AWS console, develop in browser.
CloudWatch monitoring service. Bread/Butter of SysOps Admin exam
CloudFormation solutions architect uses this all the time. Way of scripting

infrastructure. Deploy anything, and reuse code to deploy in different locations (sydney, new york, etc.)
CloudTrail, creating an S3 bucket or new user, or new EC2 instance, cloud trail logs changes to AWS environment. Turned on by default, only stores records for 1 week, if you get hacked, can figure out how and where.
Config monitors configuration of entire AWS environment, and seeing configs at different times, to visualize environment.
OpsWorks similar to elastic beanstalk, alot more robust, way of automating environment configurations
Service Catalog, manage catalog of IT services approved for use, typically used by big corps for compliance.
Systems Manager interface for managing AWS resources. Use for patch maintenance, group resources by department or applications,
Trusted Advisor understand difference between Inspector and trusted advisor. Advise around security (leaving ports open, a trusted advisor)
Managed Services not want to worry about EC2 instances or auto scaling, can help out.
Elastic Transcoder takes video and resizes to look good on mobile or different screens.
MediaConvert create video and on demand content on scale
MediaLive creates highh quality video streams to broadcast tvs and etc.
mediapackage protects media over internet
Media store provides low latency storage for on demand contennt
mediatailor allows to do targeted advertising to media streams without compromsing quality of service.
none of media are on any AWS tests.
SageMaker for deep learning for developers.
Comprehend sentiment analysis around products.
DeepLens artifically aware camera. Create an app th
Lex powers Alexa service. Way of chatting to customers.
Machine Learning throw dataset to aws cloud and analyzes dataset and predict an outcome, Amazon uses this for their recommeneded products on their page.
Polly takes text and turns into speech.
Rekognition does both video and images. Upload file and tells you what is in the file. A dog picture, knows a dog is in the picture/video. **Note Amazon purchased ring doorbell company.
Amazon Translate like google translate but AWS version.
Amazon Transcribe is automatic speech recognization, speech to text.
Athena runs SQL queries against things in S3 buckets.
EMR is elastic map reduce. Used for processing large amounts of data. Chops data for analysis
CloudSearch/Elastic Search Service are search service for AWS.
Kinesis ingesting large amounts of data to AWS. Ex. social media feeds, tweets, hashtags relative to your company.
Kinesis Video Streams
QuickSight business intelligent tool. Fraction of cost of competitors.
Data Pipeline way of moving data between different AWS services.
Glue used for ETL. Extract Transform Load. Migrating large amounts of data not in

format that you want.
IAM identity access management.
Cognito device authentication. Using mobile apps. gives temporary access to AWS.
GuardDuty monitors for malicious activities.
Inspector is an agent to install on VM or instances and run tests on them, for security vulnerabilities, and can schedule this and give reports of vulnerabilities.
Macie scans S3 buckets for PII (personal identify info) and alerts you. (SSN, DL #)
Certificate Manager way of managing SSL certificateS (if using route53 for example)
CloudHSM hardware security module, stores keys (public and private) to access EC2 instances and other items. Per hour billing, used to be 5k setup fee, now $1.20/hour
Directory Service integrating Microsoft AD with AWS Active Directory
WAF web application firewall. Stops cross-site scripting, SQL injections, app level
Shield get by default for CloudFront/Load Balancers/Route53. DDOS mitigaton, helps prevent DDOS attacks.
Artifact is a portal for on demand access for client reports. their SOC controls, PCI reports, etc.
Mobile Hub is a management console. for Mobile App generates CloudConfig file etc.
Pinpoint use targeted push notification to users (near restuarant, etc).
AWS AppSync updates data in mobile and web app in real time and offline users as soon as they re-connect.
Device Farm way of testing apps on real life devices (android, ios)
Mobile Analytics, analytics for mobile.
No mobile services are on exam.
Sumerian for AR/VR
Step Functions, way of managing Lambda functions and steps to go through it.
Amazon MQ way of doing message qeues.
SNS/SQS/SWF very old.
SQS first service launched in 06.
SNS is a notification service. (over 10$ in account, etc.)
SQS decoupling infrastructure. Qeues for EC2, etc.
SWF is simple workflow service, can have humans as a component.
Connect and Simple Email Service for Customer engagement.
Connect, dynamic, natural customer engagements.
Simple Email Service, way of sending large amounts of emails.
Alexa for Business can be used for dial into meeting room, inform IT printer is broken, etc.
Chime is used for video conferencing, can record meetings and works with low bandwidth.
Work Docs is a dropbox for AWS for work related documents.
WorkMail like Outlook.
Work Docs (the dropbox service) only one covered in Solutions Architect Exam.
Workspaces is a VDI services, running windows/linux in AWS and streams to device.
AppStream 2.0 running apps in cloud and streaming to device. like citrix
iOT internet of things ways of having thousands of devices sending back sensor information. (audio, temperature, etc.)
iOT Device Management managing iot devices and information.
FreeRTOS is a OS for microcontrollers
Greengrass is software that allows for capabilities (interface ,machine learning,

caching) for iOT devices.
GameLift service to develop games, can be VR games as well, in the cloud.

IAM manage users access to console.
Gives centralized control of account
shared acces, granular permission, identiy federation (use different identity providers, fb, linkedin, active directory)
multifactor authentication
provide temp access for users/devices and services where necessary
set up own password rotation policy,
supports PCI DSS compliance
integrates with many diffferent AWS services
Policies is a document that defines one of more permisions, apply to users, groups, roles. Can share same policy document.

Services are not available in all regions (example glacier not available in a specific region)
Pick region that is closest to you.
IAM doesnt have a region. Under Global region automatically available everywhere.

Programmitic uses access key id and secret access key, such as a script running on S3 buckets from computer
Otherwise, users can login using password
If you do not note down the access key id and secret access key when generated, then you will need to regenerate it.
Will never get secret access key ever again unless regenerated, access key id is shown under users in IAM console.
Can only use access key id and secret access key when programmatically interating with console, wont be able to log in with access keys, need password.

Cant use username/password to programmtically interact.

New Group called HR, attach policy, search s3, click s3-readonly policy, create group.
Can attach permissions to groups and to users directly.

ROLES
Use IAM roles mainly for EC2 instances to be able to write files to S3.
Dont give admin access to roles, give s3 FULL ACCESS.

Cloudwatch..does monitoring.even for billing alerts.

IAM consists of users, groups, roles, and policy documents.
Javascript Notation (JSON) is how policy documents are written. remember how the system admin policy had a "action":"*", "Resource":"*" Dont need to program to pass exam.
IAM is universal, does not apply to regions at the time,

Root account is the account created when first setup AWS account. it has complete admin access.
New users have NO permissions when first created.
New users are assigned access key id and secret access keys when first created, can be regenerated.
Access keys not same as passowrd, cannot login with them , need password.
Always setup Multifactor Authentication on your root account.
Can create and customize your own passowrd roatation policies (longer than 8 characters, expire every 90 days, etc)
New access key id is also generated when secret access key is regenerated.

S3 is simple storage service. Secure, durable object storage.
Object based storage.
Data is spread across mmultiple devices and facilities, designed to withstand failure.
Object are video, pdf docs, word docs,photos, media, flat files, not for OS or databases. that is for block base.
files can be 0bytes to 5tb
unlimited storage available.
files are stored in buckets. buckets are basically folders.
S3 is a universal namespace, must be unique globally name for the s3 bucket.
When you upload a file to s3 you will recieve a http 200 code if the upload was succesful.

Data Consistency Model for S3
read after write consistency for PUTS of new Objects
Eventual Consistency for overwrite PUTS and DELETES (can take some time to propagate)
put new object, get immediate consistency, input text file, can read immediataly, updating or deleting file, takes some time, because must update all facilities.
might get some new or old data immediately after.
updates to s3 are atomic, so you will either get only the old or the new data, not partial. that is what atomic means.
Know this going into exam.
S3 is object based.
objects consists of key, value, versionID, metadata.
Key is name of object.
Value is the data and made up of bytes.
VersionId is important for versioning.
MetaData is data about data. example is date you uploaded/updated file.
Key is the filename of the object.
Add a salt to beginning of log file names to not create a bottleneck situation of date filenaems.
Subresources consist of Access Control Lists and Torrent. (supports BitTorrent protocol).
Built for 99.99% 4 9's availability for the S3 platform.
Amazon Gaurantee 99.9% 3's availabilty.
Amazon guarantees 11 9's durability for S3 information.

Tiered Storage Availability different types of storage options.
Lifecycle Management, after 30 days move data to different storage tier, then move to archive, etc.
Versioning, one object with multiple versions.
Encryption can encrypt on S3 and do it in different methodologies.
Secure your data using Access Control Lists and Bucket Policies.
Standard and Standard-IA are Designed to sustain the loss of 2 facilities concurrently.
RRS can only lose 1 facility.
S3-IA (infrequently accessed). for data that is accessed less frequently, but requires rapid access when needed. lower fee than s3, but you are charged a retrieval fee.
RRS, Reduced Redundancy Storage, designed to provide 99.99% durability and 99.99% availability of objects over a given year. used for files that you can regenerate, because of lower durability than s3.
Glacier is very cheap but used for archival only. It takes 3-5 hours to restore from Glacier.
Stores data for as little as a cent a month per GB. takes very long to retreive, 3-5 hours. no SLA for glacier. only have 11 9's durability.
charged for storage, requests, storage management pricing (able to tag which are related to which dept, and charges for that), data transfer pricing (charged for data replication), transfer acceleration.
Transfer Acceleration enables fast easy and secure transfers of files over long distances. Uses CLoudfront edge locations to route. accelerate upload of files.
url of bucket is region.amazonaws dot com and forward slash the globally unique bucket name.
Read S3 FAQ before taking exam.
S3 buckets are global.
Bucketname must be unique, cannot be named testbucket. because you access your bucket via url address
Select region for bucket.
By default all buckets are private.
once uploaded a file, recieve a 200 success http code.
Make object public, this way can access image from internet.
can use aes-256 encryption, server side encryption.
can tag bucket, but objects in bucket do not inherit bucket tag, can tag individual objects.
Minimin size is 0 bytes for upload to S3.
buckets are universal name
upload an object o s3 to receive an http 200 code.
Encryption, client side encryption and server side encryption.
Server side encryption with Amazon S3 Managed keys (SSE-S3)
server side encrypt with kms (sse-kms)

Once you enable versioning, cannot disable is, only suspend it, may have multiple versions for S3 bucket.
Can download old version, delete old version, delete latest version.
Every time you update a file, with versioning, will hold both files, will take a lot

of space if there is a lot of big files being updated. So versioning not the best for often updated big files..would increase storage costs.
MFA Delete can help with unauthorized delete of object, and unauthorized suspend/change of versioning.
Versioning Exam Tips
Stores all versions of an object (including all writes and even if you delete an object)
Great backup tool.
Once enabled, versioning cannot be disabled, only suspended.
Integrates with Lifecycle rules.
Versioning
s MFA Delete capabaility, which uses MFA, can be used to provide an additional layer of security against unauthorized changes.

Cross Region Replication
S3 bucket..management..replication..
Cross-region Replication enables automatic and asynchronous copying of objects across buckets in different AWS regions.
Destination bucket can be either in same aws account or different aws account destination.
can change storage class for destination bucket..i.e make it standard-ia (infrequently accessed)
can also change object ownership to destination owner.
Create new iam role..or use previous roles..however we use "Create new role"
Only new objects or objects that are changed will be replicated..existing files are not.
AWS CLI (command line interface)
create user account with admin access for programitic access
do aws configure and enter access key id and secret access key
type aws s3 ls and will be able to list the buckets in the s3.
type aws s3 cp --recursive s3://awes s3://mybucketamir to copy all current contents to newly created bucket.
Permissions do not copy across to copied files.
Will put delete mark, however if you remove delete marker on source bucket, must make change on destination bucket as well (push delete but change mind and want to keep)
Versioning must be enabled on both the source and destination buckets.
Regions must be unique for cross replication.
Files in existing bucket are not replicated automatically, subsequent files or updates are replicated automatically.
You cannot replicate to multiple buckets
Delete markers are replicated.
Delete individual versions or delete markers will not replicate.
Understand what Cross Region Replcation is at a high level.
Glacier has 90 days minimum storage.
Lifecycle management can be used in conjuction with versioning.
can be applied to manage current versions and previous verisons.
following actions can now be done: transition to the standard-infrequent acces

storage class (128kb and 30 days after the creation date).
and archive to the glacier storage class (30 days after IA, if relevant)
can immediately archive to glacier as well, must be 61 days minimum from object creation date).
CloudFront CDN Overview
CDN is a content delivery network with a system of delivery servers that delivers content depending on geographic location of user, webpage, and content delivery server.
Comes with different latencies for people from different geographic area for one server, no CD Network.
Edge location is where the content will be cached.
Origin is the origin of all the files that the CDN will distribute. can be either s3 bucket, ec2 instance, route53, or elastic load balancing. Cloud front also works with any non-aws origin server, which stores the original, definitive versions of your files.
distribution is the name given the CDN which consists of a collection of edge locations.
user clicks on distribution url and will route to edge location first and check cache, if there, sends it, otherwise, goes to home bucket.
First user has no difference in latency, second user benefits for it because of cache TTL.
Cloudfront can be used to deliver your entire website, including dynamic, static, streaming, and interactive content using a global network of edge locations.
Requests automatically routed to nearest edge location, so content is delivered with the best possible performance.
web distribution are typically used for websites.
RTMP is used for media streaming. example adobe flash
understand edge location, a location where content is cached.
origin is the origin of files can be many servers, or custom origin server.
distribution is the name given to the cdn which consists of edge locations.
has 2 types of distributions, web distribution and RTMP
Edge locations are not just READ only, you can write to them too. (put object will work, will then be updated to origin).
Objects are cached for the life of the TTL
You can clear non-relevant cached objects, but you will be charged if before TTL.
TTL always in seconds
Can restrict user access to distribution in CDN using signed urls or signed cookies
You can setup access control to your buckets using bucket policies and access control lists.
s3 buckets can log all access requests, as well as send the log to another bucket/even another aws account.
Encryption
In Transit; information to and from bucket itself
using ssl/tls (https)
At rest: has two kinds
server side encryption SSE has 3 kinds
sse-s3 has s3 managed keys, each object has a key with another encryptioned master key using 256 bit (clicing on object in bucket in s3 and saying encrypt)

SSE-KMS (key management service), provides audit trail of users who use
SSE-C, client manages key
Client Side Encryption, encrypt it on your device before uploading to s3
Storage Gateway
connects an on-premise software appliance with cloud-based storage
vm image installed on host on premise at data center
File Gateway (NFS) store flat files in s3, pictures, pdfs, videos
Volumes Gateway (iSCSI) using block based, used for operating system, virtual hard disk running vm or mysql on
volume gateway has 2 kinds, stored volumes-store an entire copy onsite, cached volumes-only recent stored on premise, rest in backed in amazon
Tape gateway (VTL) archiving tapes, create virtual tapes and send to s3
File Gateway- files store in s3 buckets, accessed through NFS mount point
Volume Gateway like a virtual hard disk, have a copy on site, as well as a backup to s3
Using Snowball addresses common challenges with large scale data transfers including high network costs, long transfer times, and security concerns. Tamper resistant encloresures, 256-bit encryption, etc
Snowball edge does both storage and compute capabilities
Snowmobile is a exabyte-scale data transfer service to mov data to AWS. up to 100PB per snowmobile
import export was the old snowball, people send their own hard disks, end up becoming difficult because of different formats.
snowball can import to S3 and export from S3
S3 Transfer Acceleration utilizes cloudfront edge network to accelerate uploads to S3.
Under properties of S3 bucket, click to enable, incurs additonal fee, new endpoint url. s3-accelerate.amazonaws.com
route53 with s3 make sure bucket is same as your domain name to use bucket as a static website.
Static website hosting under properties of S3 bucket.
static website url is bucket name. (dot) s3-website-region.amazonaws.com
set index document and error document (index.html/error.html)
upload html files to s3 bucket
grant public read access to uploaded files, not bucket
no php or asp or dynamic content
scales infinitely
normal bucket url is s3.region . (dot) amazonaws.com/(Slash)bucketname
read after write consistency for puts of new objects (Read immediately after writing)
eventual consistency for overwrite puts and deletes (can take some time to propagate)
key value store is s3.
cross region replication requires versioning to enabled on dest and source buckets.
origin is either s3 bucket, ec2 instance, elb or route53
distribution is collection of edge locations
read s3 faq before taking the exam.
EC2-elastic compute cloud

provides resizable compute capacity in the cloud. boot a new server instances to minutes.
on demand- pay by the hour or by the second (only linux instances are by the second)
reserved-apps with steady state/predictable usage provide with capactiy reservation and discount on hourly charge for an instace, 1 or 3 year terms
spot enable to bid prices for instance capacity, providing greater savings if applications have flexible start and end times, instances are terminated or stopped when prices goes back above mark.
dedicated hosts- physical ec2 server dedicated for your use. used for licensing, for server-bound software licenses. can be purchased on demand (hourly) or through a reservation for a discount
instance types
d2-dense storage-file servers/data warehouring d for density
r4-memory optimized-memory intensive apps/dbs
m4-general purpose-app servers
c4-compute optimized-cpu intensive apps/dbs
g2-graphic intensive-video encoding/3d app streaming
i2-high speed storage- nosql dbs, data warehousing
f1-field programmable gate array-hardware acceleration for your code
t2-lowest cost, general purpose- web servers/ small dbs
p2-graphics/general purpose GPU- machine learning, bit coing mining, etc
x1- memory optimized- sap hana/apache spart etc xtreme RAM, xtreme memory
DR MC GIFT PX
D for density
R for RAM
M for main choice for general purpose apps
C for compute
G for grpahics
I for IOPS , high speed storage
F for FPGA
T for cheap general purpose
P for graphics (pics)
X for extreme memory
ebs-elastic block store
EBS allows you to create storage volumes and attach them to ec2 instances. block devices
can run a file system, database, or any other way a block device would.
4 ebs volume types
GP2-general purpose SSD, balance both price and performance
IO1-provisioned iops ssd, designed for i/o intensive apps such as large relational or nosql dbs. use if you need more than 10,000 iops
ST1-throughput optimized HDD, sequential writes,big data, datawarehouses, log processes, cannot be a boot volume
SC1-cold HDD, lowest cost for infreqently access workloads, file server, cannot be a boot volumes.
Magnetic (standard)-lowest cost per gig of all ebs volume types that is bootable.
With spot instances, if you terminate the instance, you pay for the hour. if aws terminates the spot instance, you get the hour it was terminated in for free.

You cannot mount 1 EBS volume to multiple EC2 instances, instead use EFS.
EBS is like a hard disk in a laptop, 2 laptops cannot share that same disk.
One subnet equals one availability zone
can monitor ec2 instances with cloudwatch monitoring
By default, ebs volume (virtual hard disk) will delete on termination.
In security group setting (virtual firewall on ec2 instance) open http and https ports for web servers, and ssh for your ip only
Need to select public and private key.
Private key required to obtain password to login to windows / ssh into linux
Key things to remember
termination protection is turned off by default, must turn it on
on an ebs backed instance, the default action is for the root EBS volume to deleted when the instance is terminated.
EBS root volumes of your default ami's cannot be encrypted. you can use a third party tool (such as bit locker) to encrypt the root volume, or this can be done when creating ami's in the aws console or using the api.
additional volumes can be encrypted.

httpd is apache
cd /var/www/html
service httpd start
chkconfig httpd on so service will come on automatically
security group rules apply immediately
security group inbound rules will always be allowed out, that is stateful. vpc network access control lists, later in course are stateless by contrast.
cant deny ips or ports in security groups can only allow, can block in network access control lists
rdp is port 3389, for windows vm
can have two security groups for one instance
all inbound traffic is blocked by default.
all outbound traffic is allowed.
changes to security groups take effect immediately.
you can have any number of ec2 instances within a security group.
you can have multiple security groups attached to ec2 instances.
security groups are stateful.
if you create an inbound rule allowing traffic in, that traffic is automatically allowed back out again.
you cannot block specific ip addresses using security groups, instead use network access control lists.
you can specify allow rules, but not deny rules.
cannot have ebs volume in different regions, must be in same region (HDD or gp2 volumes)
cannot modify size of standard volume type, can modify sc1, st1, gp2
create snapshot of volume to create a copy of the volume and be able to move instance/volume to a different available zone
copy snapshot to another region, create an image of that snapshot in new region. to move regions
able to boot image as a new ec2 instance.

options-create image
also can copy ami to another region, to move ebs volumes to other regions.
snapshots stored on s3, cant view on s3 though,
snapshots are point in time copies of volumes
snapshots are incremental- this means only the block that have changed are moved to s3
first snapshot may take time to create
to create a snapshot, should stop instance before, but can do a snapshot while running
you can create ami's from both volumes and snapshots
you can change ebs volume sizes on the fly, including changing the size and storage type
volumes will always be in the same availability zone as the ec2 instance.
to move an ec2 volume from one az/region to another, take a snap or ami image of it, then copy to new az/region
snapshots of encrypted volumes are encrypted automatically.
you can share snapshots, but only if they are unencrypted. these shapshots can be shared with other aws accounts or be made public
volumes restored from encrypted snapshots are encrypted automatically
you can share snapshots, but only if they are unencrypted, can share will other aws accounts or made public
AWS never recommends raid 5
create raid to increase i/o
new striped volume, raid0
to take snapshot with raid array, freeze the file system, unmount the raid array, or shut down the instance and copying it
AMI Types (EBS vs Instance Store)
You can select your AMI based on region, OS, architecture (32 or 64 bit), launch permissions, and storage for root device (between instance store, which is ephermal storage, or ebs backed volumes.
Most are ebs root device types on amazon
instance store cannot stop instance, only terminate or reboot
instance store less durability
ebs, when you stop, and start, will go on another hypervisor, instance store cannot do that
All ami's are categorized as either backed by amazon ebs or backed by instance store
for ebs volumes: the root device for an instance launced from the ami is an amazon ebs volume created from an amazon ebs snapshot
for instance store volumes: the root device for an instance launched from the ami is an instance store volume created from a template stored in amazon s3.
fast provisioning times needed? then do ebs backed
Instance store volumes are sometimes called ephermeral storage, they do not move, ephemeral means lasting for a short time
instance store volumes cannot be stopped. if the underlying host fails, you will lost your data.
ebs backed instances can be stopped. you will not lose the data on this instance if it is stopped.
you can reboot both instance and ebs, you will not lose your data

by default, both root volumes will be deleted on termination, however with ebs volumes, you can tell aws to keep the root device volume
Elastic load balancing
classic load balancer is a layer 4 balancer, tcp/routing decisions, can also do layer 7 decisions (http apps)
there is an application load balancer, for http and https, but exam mainly on classic
to serve web traffic, external load balancer, not an internal load balancer.
instances monitored by elb are reported as inservice or outofservice, depending on health check settings
health checks check the instance health by talking to it
elastic load balancers have their own dns name. you are never given an ip address.
read the elb faq for classic load balancers
Metrics for ec2 default are cpu, disk, network, and status check related.
cloudwatch has dashboards,alarms,events,logs
basic monitoring period is 5 minutes, detailed monitoring is every 1 minute but more expensive
events can trigger actions (lambda functions)
standard monitoring always 5 minutes
detailed monitoring always 1 minute
dashboards-create with widgets to see aws environment metrics
cloudwatch vs cloudtrail
cloudwatch is monitoring, performance, et
cloudtrail completely seperate, for auditing, provides trail of access, etc. comes often on exam do not mix up
aws cli shortcut to copy s3 bucket to ec2 instance
create ec2 role for s3-admin-access, this way no need to configure aws commandline with secret key id, etc
type aws s3 cp --recursive s3://bucketname /home/ec2-user
Can put a bash script in advanced settings of ec2 instance, along with s3-admin-access iam role, put the following to automate a deployment of a webserver:
#!/bin/bash
yum
install httpd -y
yum update -y

aws s3 cp s3://YOURBUCKETNAMEHERE /var/www/html/ --recursive
service httpd start

chkconfig httpd on
for ec2 instance metadata
type curl http://169.254.169.254/latest/meta-data
need to remember that url for exam
remember, after the metadata for 169.254.169.254, never user data
for bash scripting, can do commands like curl
http://169.254.169.254/lastest/meta-data/public-ipv4 > mypublcip.html
for autoscaling, need to first setup a launch configuration
spread instances in auto scaling over many subnets (availability zones, this is the

reason for auto scaling, so that if one availability zone goes down, you have backups)
can grow and decrease group size automatically in auto scale group on a depency of different criteria, such as over 90% cpu utilization, add one instance, etc.
load balancer will send to one of the multiple instances set up
can also set up desired amount of instances wanted, and auto scale will do it, example 3 instances desired, 2 go down, so auto scale will restart 2 more instances according to launch configuration settings setup earlier
EC2 Placement Groups is a logical grouping of instances within a single availability zone. using placement groups enables apps to participate in a low latency, 10 gbps network. placement groups are recommended for apps that benefit frmo low network latency, high network throughput, or both. 10 gigs per second!!
A placement group cant span multiple availability zones
the name must be unique within the aws account
only certain types of instances can be launched in a placement group (compute optimized, gpu, memory optimized, storage optimized)
aws recommed homogenous instances (same size, same family) within placement groups
cant merge placements groups
cant move an existing instance into a placement group. you can create an ami from your existing instance, and then launch a new instance from the ami into a placement group
EFS
EFS is elastic file system, storage capacity is elastic, growing and shrinking automatically as you add and remove files, so apps have storage when they need it
cannot mount 1 ebs to two instances, efs you can
efs supports the nfsv4 protocol (network file system)
you pay for storage you use (no pre-provisioning reqiored)
can scale up to petabytes
can support thousands of concurrent nfs connections
data is stored across multiple az's within a region
read after write consistency
EFS is block based storage like ebs, not object based like s3
create one efs, multiple ec2 instances
assign ec2 instances security group to be in the same security group as the efs volume. ec2 instances can have multiple security groups
use case: efs is a file server, central repository, multiple ec2 instances to one efs, accessing same file
Lambda
AWS lambda is a compute service where you upload your code and create an lambda function.
No worry about OS, patching, scaling etc. use lambda for event-driven compute services, running in response to events, such as changes to s3 bucket or dynamodb table, or in response to http requests using amazon api gateway or api calls made using aws sdks.
lambda events can trigger other lambda events
2 users with 2 http requests, will invoke 3 lambda functions, never 3 requests and 1 lambda responding
Node.js Java python and c# are the 4 languages supported

First 1 million requests are free, twenty cents per 1 million requests thereafter, cheap to run a website
duration is calculated from the time code begins executing until it returns or terminates, price depends on amount of memory you allocate to your function, 0.00001667 for every gig/sec used
function cannot execute for more than 5 minutes
no db admins, network admins, sys admins, just focus on code
lambda scales instantly and scales out (not up) automatically
lambda functions are independent, 1 event= 1 function
lambda is serverless
know what services are serverless s3, api gateway, dynamodb, NOT EC2
lambda functions can trigger other lambda functions, 1 event can = x functions if functions trigger other functions
architectures can get extremely complicated for debugging, aws x-ray allows you to debug what is happening
lambda can do things globally, you can use it to back up s3 buckets to other s3 buckets etc
know your triggers, what services can trigger lambda
max function time is 5 minutes and the 4 languages are c#, python, node.js, and java
simple microservice permissions gives basic execution activity
Know what can and cannot trigger Lambda
RDS cannot trigger lambda, dynamodb and s3 can
CAN trigger Lambda functions: API gateway, AWS IoT, Alexa Skills Kit, Alexa Smart Home, CloudFront, CloudWatch Events, CloudWatch Logs, CodeCommit,Cognito Sync Trigger, DynamoDB, Kinesis, S3, SNS.
make sure route53 domain name is available in s3 bucket name (both need to be unique) so that you can connect the two
Polly is aws text to speech
cors-cross origin resource sharing s3 to access
Read Ec2 faq before attempting exam
Know the differences between on demand, spot, reserved, and dedicated hosts
With spot instances, if you terminate, you pay for the hour, if aws terminates the spot instance, you get the hour it was terminated in for free.
DR MC GIFT PX
E.B.S. consists of:
S.S.D., general purpose-gp2
S.S.D., provisioned iops- io1
H.D.D., throughput optimized -st1- frequently accesssed workloads
H.D.D., cold- sc1- less frequently accessed data
H.D.D., magnetic - standard- cheap, infrequently accessed storage, only bootable h.d.d.
You cannot mount 1 E.B.S. volume to multiple E.C.2 instances, instead use E.F.S.
Termination protection is turned off by default, you must turn it on.
on an e.b.s. backed instance, the default action is for the root E.B.S. volume to be deleted when the instance is terminated.
Root Volumes cannot be encrypted by default, you need a third party tool (such as bit locke etc) to encrypt the root volume.
Additional volumes can be encrypted

Volumes exist on E.B.S., which are basically virtual hard disks
snapshots exist on S3.
you can take a snapshot of a volume, this will store that volume on S3.
snapshots are point in time copies of volumes.
snapshots are incremental, this means that only the block that have changed since your last snapshot are moved to S3.
Snapshots of encrypted volumes are encrypted automatically.
volumes restored from encrypted snapshots are envcrypted automatically.
you can share snapshots, but only if they are unencrypted.
to create a snapshot for amazon E.B.S. volumes that serve as root devices, you should stop the device before taking a snapshot.
instance store volumes are sometimes called ephemeral storage.
instance store volumes cannot be stopped. if the underlying host fails, you will lose your data, cannot jump hypervisors (or vms)
E.B.S. backed instances can be stopped. You will not lose the data on this instance if it is stopped.
You can reboot both, you will not lose your data.
by default, both root volumes will be deleted on termination, however with E.B.S. volumes, you can tell A.W.S. to keep the root device volume, only for E.B.S.,not instance store
How can I take a snapshot of a RAID array? Because of cache, in a raid array, must stop the application from writing to disk and flush all caches to the disk. you can do this by freezing the file system, unmounting the RAID array, or the easiest solution, shutting down the associated EC2 instance.
A.M.I's are regional. you can only launch an ami in the region it was stored.
however you can copy A.M.I's to other regions
CloudWatch is for performance monitoring
CloudTrail is for auditing, providing trail of records
Standard Monitoring = 5 minutes
Detailed Monitoring = 1 minute
Cloudwatch, create dashboards, alarms, events, logs
Roles are more secure than storing your access key and secret access key on individual E.C.2 instances.
Roles are easier to manage, than access keys
Roles can be assigned to an E.C.2. instance AFTER it has been provisioned using both the command line and the A.W.S. console.
Roles are universal, you can use them in any region.
Instance metadata, used to get information about an instance (such as public ip)
remember this url for metadata
curl http://169.254.169.254/latest/meta-data/
E.F.S. supports the N.F.S.v.4. protocol.
you pay for storage you use (no pre-provisioning reqiored)
can scale up to petabytes
can support thousands of concurrent nfs connections
data is stored across multiple az's within a region
read after write consistency
You must deregister an AMI before being able to delete the root device it is registered to.

A placement group can NOT be deployed across multiple availability zones.
A placement group is ideal for EC2 instances that require high network throughput and low latency across a single availability zone.
Elastic Load balancer always only have a dns name, no ip address, no A record to resolve, must use Alias record in route53, again Elastic load balancer only has a dns name, never ip address
Drop TTL to 300 seconds before doing website migration to cloud, as normally set for 2 days.
AWS has Alias records, which work like cnames, except able to map resource sets in your hosted zone to elastic load balancers, cloudfront distributions, or s3 buckets configured as websites.
E.L.B.'s do not have a pre-defined ipv4 addresses, you resolve to them using a DNS name, (so it can balance amongst its list of i.p's)
understand the difference between an alias record and a cname
You are charged for using cnames requests, not charged for using alias records
given the choice, always choose an alias record over a cname
route53 is global service, similar to i.a.m., i.a.m. is global users
Route53 has 5 routing policies, simple, weighted, latency, failover, and geolocation
default routing policy is
Simple routing policy, used when having only one web server.
Weighted routing policy, 2 webservers, and can assign 20% to east and 80% of traffic to west for example, can be used in new website testing, sending 20% to test website, and 80% to normal website A/B testing..etc., can also be in same region or not
Latency routing policy, allows you to route your traffic based on lowest network latency for end user (which region will give them the fastest response time), 54 ms to eu-west and 300ms to ap-southeast, so route53 sends this traffic to eu-west
Failover routing policy, use to create an active/passive set up. for example, you may want your primary site to be in eu-west-2 and your secondary DR site in ap-southeast-2, route53 monitors using a health check, if health check fails on active site, then route53 will direct traffic to passive site
Geolocation routing policy, lets you choose where your traffic will be sent based on the geographic location (european customers go to euro ec2 for european site, us customers go to us site, by continent, country, or even by U.S. states)
Databases
understand between DyanmoD.B. and R.D.S.
R.D.S. is a relational database system.
R.D.S. types include S.Q.L. server, oracle, MyS.Q.L. server, postgreS.Q.L., aurora, and MariaD.B.
O.L.T.P. vs O.L.A.P
online transaction processing, trans#41, pulls up data, price, purchase, etc.
online transaction analytics processing, pulls in large numbers of records, net profit for product, need sum of multiple things, unit cost, sale price, profit, etc.
Elasticache is web service easy to deploy and scale an in-memory cache in the cloud.
improves performance of web apps, caches and takes load off of database
DMS is database migration service. allows to migrate production databse to aws. aws manages all complexities of migration process. AWS schema conversion tool automatically converts the source database schema to a format compatible with the

target database.
RDS(relational database system) is an OLTP (Transaction), including SQL, MySQL, PostgreSQL, Oracle, Aurora, MariaDB
DynamoDB- No SQL
RedShift - OLAP (analytical)
Elasticache -in memory caching, memcached and redis
DMS-database migration service

RDS instance
RDS is a database, relational database
used for mysql, oracle, etc
dont make it public, you will use securtiy group
create new security group and then edit/change inbound security rules for rds to come from mywebdmz (in source) security group on port 3306 on tcp protocol
run an ec2 instance, attach php bash script in bootstrap and put in webdmz security group
attach webdmz security group, which will allow to read/write
copy endpoint of rds mysql url and put in connect.php file under host name
point is have 2 security groups, one for ec2 instance, one for RDS instance, and allow web security group into rds group over port 3306
check rds instance allows port 3306 for successful connection to ec2 instances (common exam question)

Two types of backups: automated and database snapshots
automatic backups enabled by default, cannot access i/o during backup or experience heavy latency
 DB snapshots are done manually, stored even after you delete the original rds instance, unlike automated backups
when you restore from either backup, the restored version of the database will be a new rds instance with a new end point.
Encryption at rest is supported, cannot encrypt existing databases, can create a new rds instance encrypted

multi-az allows you to have a copy in another AZ and has a failover that will switchover. Used for disaster recovery, not scaling
read replica has asynchronas reads, used for scaling not for DR, if database is read heavy, use this
read replica is read only not supported in SQL or Oracle
my have auto backups turned on to deploy a read replica
can have up to 5 read replica copies of any datbase
can have read replicas of read replicas
each replica will have its own dns end point
you cannot have replicas that have multi az
can create replica of a multi-az source database however
replicas can be promoted to be their own databases to be be write to it, this breaks the replication
replica in a second region for mysql and mariadb, not postgresql
DynamoDB offers "push button" scaling, meaning that you can scale your database on

the fly, without any downtime.
RDS is not so easy and you usually have to use a bigger instance size or to add a read replica.

DynamoDB is a fast and flexible NoSQL database service for all application that need a consistent latency at any scale
stored on ssd storage
spread across 3 facilities (zones)
eventual consistent reads (default)-consistency across all copies reached within a second. repeating a read after a short time should return the updated data (best read performance)
strongly consistent reads- returns a result that reflects all writes that received a successful response prior to the read
dynamodb write capacity unit can handle 1 write per second, so for 11.6 writes/sec you need 12 write capacity units
writes more expensive than reads
to scale, just go to capacity tab and increase read/write capacity units (push button)

redshift-data warehouse, petabyte scale
more about columns, sums, etc. not about individual entries
single node starting at 160gb to multi-node (with leader node..communicating to other compute nodes, up to 128 compute nodes)
advanced compression, columnar data stores can be compressed much more than row-based because similar data is stored sequentially on disk
amazon automatically samples and compresses for you
massively parallel processing (mpp): redshift automatically distributes data and query load across all nodes. easy to add nodes
encrypted in transit using ssl, at rest using aes-256
by default redshift takes care of key management, can manage own keys through hardware security modules or aws kms
currently only available in 1 az
can restore snapshots to new avail zones in the event of an outage.
achieves speed because of column organization

elasticache-2 types of engines
elasticache is a web service that makes it easy to deploy, operate, and scale an in-memory cache in the cloud. the service improves the performance of web applications by allowing you to retrieve information from fast, managed, in-memory caches, instead of relying entirely on slower disk-based databases.
used to significantly improve latency and throughput for many read-heavy application workloads (social networking, gaming, media sharing, q/a portals) or compute-intensive workloads (recomendation engine)
cached information may include the results of i/o intensive database queries or the results of computationally-intensive calculations.
types of elasticache, memcached and redis
Exam tip: typically you will be given a scenario where a database is under a lot of stress/load. which service to alleviate this?

elasticache is a good choice if your database is particularly read heavy and not prone to frequent changing.
redshift (data warehouse) is a good answer if the reason your database is feeling stress is because management keeps running OLAP transactions on it.

Aurora is a mySQL-compatible, relational database engine that matches high-end commercial databases(enterprise-clas) while being 1/10th the price. competitor is oracle, and it is mySQL compatible.
start with 10gb and autoscale in 10gb increments
easily scale up
2 copies of data (data(harddrive) only) contained in each availability zone, with minimum of 3 availability zones, 6 copies of data, highly redundant
can lose up to two copies for write and up to 3 copies for read availability
self-healing, data blocks and disks are continuosly scanned for errors and repaired automatically.
2 types of replicas available, aurora replicas (currenlty 15, and automatic failover), and mysql read replicas (currently 5) of the aurora database
Assign failover priority, tier 0 highest priority for failover
always use cluster endpoint, aws will handle failover to instance endpoint, unless wanting to directly acces instance endpoint

Database Summary
RDS-OLTP, includes SQL, MySQL, PostgreSQL, Oracle, Aurora, MariaDB
DynamoDB- No SQL
RedShift - OLAP and used for data warehousing
Elasticache- In Memory Caching, with 2 services, Memcached and Redis
Multi-AZ versus Read Replica
multi-az for redundancy, read replica for throughput multiply
DynamoDb offers "push button" scaling, scale database on the fly, without any downtime
RDS is not so east and need to use a bigger instance size or to add a read replica
DynamoDB stored on SSD storage, spread across 3 geo distinct data centers
eventual consistent reads (default)
strongly consistent reads (choose if app needs data within less than 1 second of write)
Redshift configuration- single node(up to 160gb, for small business) and multi node, with a leader node for multi node

Read FAQ on RDS for exam

VPC-make sure can build vpc by memory before exam
think of a vpc as a virtual data center in the cloud
what can you do with a vpc?
-launch instances into a subnet of your choosing
assign custom ip address ranges in each subnet
configure route tables between subnets
create internet gateway and attach it to our VPC
much bettwe security control over your AWS resources

instance security groups
network access control lists (NACLS)
ONLY ONE GATEWAY PER VPC-be aware fore xam

Default VPC vs Custom VPC
Default VPC is user friendly, allowing you to immediately deploy instances
All subnets in default VPC have a route out to the internet, no private subnets
Each EC2 instance has both a public and private IP address

VPC peering
allows you to connect one VPC with another via a direct network route using private IP addresses
instances behave as if they were on the same private network
you can peer VPC's with other AWS accounts as well with othe VPC's in the same account
Peering is in a hub-spoke configuration, no transitive peering, i.e. 1 central vpc with 4 others

Think of a VPC as a logical datacenter in AWS.
consists of IGWs (or virtual private gateways), route tables, network access control lists, subnets, and security groups
1 subnet=1 availabilty zone
security groups are stateful (allow inbound, will inherit to outbound too); network access control lists are stateless (NACLs need to open both inbound and outbound ports if want connection)
no transitive peering

Must disable source/dest checks on NAT ec2 instance, not the nat gateway
attach nat instance to webdmz
attach private vpc to have an inbound rule of 0.0.0.0/ for the nat instance created
hard to manage nat instances, if a small compute, will need to scale, gateways are newer/easier

Exam Tips NAT Instances
When creating a NAT instance, Disable Source/ Destination Check on the Instance
NAT instances must be in a public subnet
There must be a route out of the private subnet to the NAT instance, in order for this to work.
The amount of traffice that NAT instances can support depends on the instance size. If you are bottlenecking, increase the instance size.
You can create high availability using Autoscaling Groups, multiple subnets in different AZs, and a script to automate failover.
Behind a security group

place nat gateway in webdmz group
Exam Typs NAT Gateways
Preferred by the enterprise

Scale automaticaly up to 10gpbs
no need to patch
not associated with security groups
automatically assigned a public ip address
remember to update your route tables, need to have in multiple AZs, and route tables should be updated accordingly to connect all for failover
no need to disable source/dest checks like nat instances
more secure than nat instaces, no need for security checks, cant ssh into it like a nat instance, much more secure. AWS will manage it for you, always want a gateway over a nat instance.

Exam Tips-Network ACLs
Your VPC automatically comes with a default NACL, and by default it allows all outbound and inbound traffic
You can create custom NACLs. By default, each custom NACL denies all inbound and outbound traffic until you add rules.
Each subnet in your VPC must be associated with a network ACL. If you dont explicitly associate a subnet with a NACL, the subnet is automatically associated with the default NACL.
You can associate a network ACL with multiple subnets; however, a subnet can be associated with only one NACL at a time. When you associate a NACL with a subnet, the previous association is removed.
Network ACLs contain a numbered list of rules that is evaluated in order, starting with the lowest numbered rule.
Network ACLS have seperate inbound and outbound rules, and each rule can either allow or deny traffic.
Allow outbound rules for ephermeral ports only (NEVER inbound rules for ephermeral ports)
Network ACLs are stateless, responses to allowed inbound traffic are subject to the rules for outbound traffic (and vice versa).
Block IP Addresses using network ACLs not Security Groups.

VPC Flow Logs Exam Tips
You cannot enable flow logs for vpcs that are peered with your vpc unless the vpc is in your accound
you cannot tag a flow log
after you've created a flow log, you cannot change its configuration; for example, you can't associate a different IAM role with the flow log.
Not all IP traffic is monitored:
traffic genereated by instances when they contact the amazon dns server. if you use your own dns server, then all traffic to that dns server is logged.
traffic generated by a windows instance for amazon windows license activation
traffic to and from 169.254.169.254 for instance metadata
DHCP traffic
Traffic to the reserved IP address for the default VPC router.

NAT instances always behind a security group, nat gateways are not

Exam Tips -ALB's
You will need at least 2 public subnets in order to deploy an application load balancer.
Exam Tips- NAT vs Bastions
A NAT is used to provide internet traffic to EC2 instances in private subnets
A Bastion is used to securely administer EC2 instances (using SSH or RDP) in private subnets.

SQS-oldest AWS service, read FAQs
a web service that gives you access to a message queue that can be used to store messages while waiting for a computer to process them
A temporay repository for messages that are awaiting processing.
SQS is always a pull based system. The Ec2 pulls from the Simple Query Service.
you can decouple the components of an application so they run independently, with sqs easing message management between compenents.
Acts as a buffer, resolves issues that arise is produces is producing work faster than the consumer can process it, can also automatically scale instances according to queue size, bringing elasiticity
There are two types of queue: standard queues(default) and FIFO queues
Standard queues lets you have a nearly-unlimited number of transactions per second.
Standard queues provide best-effort ordering, cannot gaurantee will be in same order sent, gaurantee sent at least once.
FIFO queues are exactly that and duplicates are not introduced in the queue(Sent only once) and limited to 300 transactions per second.
Key Facts
SQS is pull based, not push base
messages are 256 KB in size and any text
Messages can be kept in the queue from 1 minute to 14 days the default is 4 days.
Visibility Time Out is the amount of time that the message is invisible in the SQS queue after a reader picks up that message. This could result in the same message delivered twice.
Visibility time out maximum config is 12 hours.
SQS guarantees that your messages will be processed at least once, especially if VTO is low and process takes long
SQS long polling is a way to retrieve messages from your SQS queues. While the regular short polling returns immediately, even if the message queue being polled is empty, long polling doesnt return a response until a message arrives in the message queue, or the long poll times out.

SWS- Simple Workflow Service
Makes it easy to coordinate work across distributed application components.
Can involve human applications as well
SQS has a retention period of 14 days
SWF up to 1 year for workflow executions
SWF presents a task-oriented API, whereas SQS offers a message-oriented API (to decouple infrastructure for example)
SWF ensures that a task is assigned only once and is never duplicated. SQS, you may need to handle duplicated messages and may also need to ensure that a message is

processed only once.
SWF keeps track of all the tasks and events in an application. With SQS, you need to implement your own application-level tracking, especially if your application uses multiple queues.
SWF Actors (3)
Workflow Starters- an app that can initiate a workflow, could be your app searching for bus times.
Deciders- control the flow of activity tasks ina workflow execution, decides what to do next if failed or finished.
Activity Workers- carry out the activity tasks
12 month retention period! remember compared to SQS which is 14 days

SNS-Simple Notification Service
makes it easy to set up, operate, and send notifications from the cloud.
Push notifications to apple, google, windows devices, email, http endpoints, text, SQS queues, trigger Lambda functions.
Push service
sns allows you to group multiple recipients using topics.
To prevent messages from being lost, all messages published to SNS are stored redundantly across multiple availability zones.
Instantaneous, push-based devilery (no polling)
Simple APIs and easy integration with apps
Flexible message delivery over multiple transport protocols
Inexpensive, pay-as-you-go model with no up-front costs
Web-based AWS management console offers the simplicity of a point-and-click interface.
SNS vs SQS
Both messaging services in AWS
SNS- Push
SQS - Pull (Polls)
SNS Pricing
0.50 per 1 million SNS requests
0.06 per 100,000 notif devliveries over HTTP
0.75 per 100 notif deliveries over SMS
2.00 per 100,000 notif deliveries over Email

Elastic Transcoder
Media Transcoder in the cloud
Convert media files from their original source format in to different formats that will play on smartphones, tables, PC's, etc.
Provides transcoding presets for popular output formats, which means that you don't need to guess about which settings work best on particular devices
Pay based on the minutes that you transcode and the resolution at which you transcode.

API Gateway
can have your own api in aws to call lambda functions or ec2 instances
API gateway has caching capabilities to increase performance.

low cost/scale automatically
can throttle gaetway to prevent attacks.
can log results to cloudwatch
if using multiple domains with API gateway, ensure you have CORS on API gateway
CORS (cross-origin resouce sharing) is one way the server at the other end (not the client code in the browser) can relax the same-origin policy,
CORS is a mechanism that allows restricted resources (fonts) on a web page to be requested from another domain outside of the domain from which the first resource was shared.
Error- "origin policy cannot be read at the remote resource" you need to enable CORS on API Gateway

Kinesis
Streaming data is data generated continuously by thousands of data soruces, which typically send in the data records simultaneously, and in small sizes (order of Kilobytes)
Online stores, geospatial data, iot sensor data, etc.
Kinesis is a platform where you send your data to
Kinesis Streams, Firehouse, and Analytics,
Kinesis Streams, default 24 hours, can increase to 7 days retention, stored in Shards, moves from producers to stream (shards) to consumers
Kinesis Streams consist of shards, 5 trans per sec for reads, etc. The data capacity of your stream is a function of the number of shards that you specify for the stream. The total capacity of the stream is the sume of the capacities of its shards.
Kinesis Firehouse, either analyzed in real time using Lambda, or sent to S3, then to Redshift, or to Elasticsearch Cluster, no retention period,
Firehouse is automated, no need to worry about management of shards or data retention.
Analytics sits on top of both and allows to run SQL queries to both firehouse and streams, and then use the query to store data in S3, redshift, or elasticsearch cluster
Know the difference between streams and firehouse. and high level of analytics

side note: draw.io to build aws architect diagrams

6 advantages of cloud
trade capital expense for variable expense
benefit from massive economies of scale
stop guessing about capacity
increase speed and agility
stop spending money running and maintaining data centers
go global in minutes

11 regions with each region consisting of multiple AZs most regions have 3 availability zones, not all
Access is authorized on a 'least privilege basis'

PCI DSS Level 1, able to take credit card info, still need QSA auth

Shared Security (responsibility) model, AWS manages global infrastructure (hardware), you are responsible for the services you put (s3 services, etc)
IaaS, such as EC2, VPC, S3, completely under your control and require you to perform all the security configuration and management tasks.
Managed services (SaaS), AWS is responsible for patching, antivirus, etc. however you are responsible for account management and user access.
Recommened that MFA is implemented, communicate to these services using SSL/TLS and that API/user activity logging be setup with CloudTrail.
You must request a vulnerability scan in advance. cannot do port scans of ec2 instances unless giving notice

Instances have no access to raw disk devices, but instead are presented with virtualized disks.
Memory allocated is zero'd by the hypervisor when it is unallocated to a guest. memory not returned to pool under complete.
AWS does not have any access rights to your instances or the guest OS
Firewall-EC2 provides a complete firewall solution; this mandatory inbound firewall is configured in a default deny-all mode and You must expilicity open the ports needed to allow inbound traffic.
AES 256 encryption on EBS (elastic block storage) volumes, encryption occurs on ec2 instance so encrypts as data moves between EC2 and EBS
ELB-elastic load balancing- SSL termination on the load balancer is supported (unencrypted web servers possible, to allow more traffic)
Also allows you to identify the originating IP address of a client connecting to your servers, whether you're using HTTPS or TCP load balancing.

AWS management re-evaluates the strategic business plan at least every 6 months.

Storage Gateway is a software appliance for download as a VM that you install on a host in your site's datacenter.
2 types; Gateway cached volumes and
Gateway cached volumes utilizes S3 as primary data, while caching locally via iSCSI attached at the on-premisis datacenter. up to 32 TBs can be created Gateway-stored volumes store your primary data locally, while backing up that data to AWS, up to 1TB in size, in the form of EBS snapshots

Rule of thumb: design for failure
decouple your components (just choose SQS), build components that do not have tight dependencies on each other
in the case of batch-processing architecture, you can create asynchronous components that are independent of each other. (SQS)
Implement elasticity, in 3 ways
proactive cyclic scaling (fixed interval, monthly, daily)
proactive event-based scaling (marketing campaigns)
auto-scaling based on demand (cpu utilization %)

*well-architected framework
General Design Principle
stop guessing your capacity needs
test systems at production scale
automate to make architectural experimentation easier
allow for evolutionary architectures
data-driven architectures (using cloudwatch to inform)
improve through game days

Security Pillar
apply at all layers
enable traceability
automate responses to security events (add sns notifs)
focus on securing your system
automate security best practices (deploy hardened images instead of reg ami images)
Shared responsbiliy between aws and customer

Security in the cloud consists of 4 areas.

data protection-least privilege accesss ystem, organize and classify data into segments such as publicly avail, avail to only members of organization, available only to board, etc., encrypt everything when possible, whether it be at rest or in transit
maintain full control over data,
versioning helps by protecting against accidental overwrites, deletes,

privelege management-ensures only authorized and authenticated users are able to access resources, includes ACLs, role based access controls, and password management (such as password rotation policies)
use groups, how are you protecting access to and use of aws root account credentials
how are limiting automated access
how are you managing keys/credentials

infrastructure protection-How are you protecting your VPC
how are you enforcing network and host-level boundary protection
how are you enforcing AWS service level protection
how are you protecting the integrity of the OS on your EC2 instances

detective controls-identify a security breach
cloudtrail, cloudwatch, config, s3, glacier
how are you capturing and analyzing AWS logs
cloudtrail is regional (are you operating in each region)

Key AWS Services
Data protection-can encrypt your data both in transit and at rest using; ELB, EBS, S3 & RDS
Privilege management-IAM, MFA

Infrastructure protection-VPC
Detective controls-CloudTrail, Config, Cloud Watch

Reliability pillar
covers ability of a sys to recover from disruptions as well as ability to dynamically acquire computing resources to meet demand
3 sections
Foundations-IAM, VPC
How are you managing service limits
how are you planning your network topology on AWS
do you have an escalation path for issues
Change Management- Cloudtrail
how does your system adapt to changes in demand?
how are you monitoring AWS resources?
how are you executing change management?
Failure Management- CloudFormation
How are you backing up your data?
How does your system withstand compenent failures?
How are you planning for recovery/

Performance Efficiency Pillar
focuses on how to use computing resources efficiently and how to maintain that efficiency as demand changes and technology evolves
(Lambda released at Re:invent few years ago..)
4 areas
Compute-autoscaling
storage-ebs, s3, glacier
database-rds, dynamodb, redshift
space-time trade-off; cloudfront, elasticache, direct connect, rds read replicas (goal to lower latency)

Cost Optimization Pillar
reduce costs to a minium and use savings for other parts of business, still achieving business objectives
matched supply and demand-autoscaling
cost-effective resources-ec2 (reserved intances), aws trusted advisor
expenditure awareness-cloudwatch alarms, sns
optimizing over time-aws blog, aws trusted advisor

example for oot is MySQL RDS, and Aurora being launced in re:invent 2014
aurora may be better now because of its performance and redundancy, keep track of changes made to AWS

Operation Excellence Pillar
includes operational practices and procedures sued to manage production workloads.
includes how planned changes are executed, as well as responses to unexpected operational events.

should be automated
PREPARATION;
CONFIG provides detailed inventory of resources/configuration and continuously records configuration changes
SERVICE CATALOG helps to create a standardized set of service offerings that are aligned to best practices
designing workloads that use automation with services like auto scaling and SQS are good methods to ensure continuous operations in the event of unexpected operationals events
what best practices for cloud operations are you using?
how do you configure management for your workload?
OPERATION;
AWS CodeCommit, CodeDeploy, CodePipeline can be used to manage and automate code changes to AWS workloads
use AWS SDKs or third party libraries to automate operational changes
use Cloudtrail to audit and track changes made to AWS environments
how are you evolving your workload while minimizing the impact of change?
how do you monitor your workload to ensure it is operating as expected?
RESPONSES;
take advantage of all the CloudWatch service features for effective and automated responses.
Cloudwatch alarms can be used to set thresholds for alerting and notification, and cloudwatch evetns can trigger notifications and automated responses
how do you respons to unplanned operational events?
how is escalation managed when responding to unplanned operational events?

Additional Exam Tips
Kinesis-used to consume big data
stream large amounts of social media, news feeds logs THINK KINESIS

process large amounts of data;
REDSHIFT for business intelligence
EMR (elastic map reduce) for big data processing

EC2-EBS backed vs. instance store
ebs backed volumes are persistent, can be detached and reattached to other EC2 instances
instance store backed volumes are not persitent (ephermal, meaning short time) (exist only for life of instance)
ebs volumes can be stopped; data will persist (not lose data)
instance store volumes cannot be stopped - if you do this the data will be wiped
ebs backed- store data long term
instance store- shouldnt be used for long-term data storage

OpsWorks-orchestration service that uses Chef
chef consists of recipes to maintain a consistent state
look for term "chef"/"recipes"/"cook books" and think OPSWORKS!!

Elastic Transcoder
media transcoder in the cloud
convert media files from their source format to diff formats (including mobile)
dont need to guess about which settings work best on devices
pay based on minutes that you transocde and the resolution

SWF Actors
workflow starters-app that starts a workflow (mobile app searching for bus times)
deciders-control flow of activity tasks
activity workers-carry out activity taks

EC2- Get Public IP Address
need to query the instances metadata
curl http://169.254.169.254/latest/meta-data
get http://169.254.169.254/latest/meta-data
key thing to remember is that its META DATA not user data

wordpress- run as a crontab, so it will automatically run, root aws s3 sync --delete ec2 directory to /wp-content/uloads s3://buckname (sync that ec2 folder to the s3 buckets, including deletes)
.htaccess file to do cloudfront on s3 bucket files

Cloudformation is needed expertise for Solution architect. create templates or use made ones(Wordpress, already will create all needed resources,, etc. database, ec2 instances, alb,)

Consolidated Billing
allows you to get volume discounts on all your accounts.
unused reserved instances for ec2 are applied across the group
cloudtrail is on a per account and per region basis but can be aggregated into a single bucket in the paying account.

resource groups to find resources

VPC peering
You cannot create a vpc peering connection between vpcs that have matching or overlapping cidr blocks.
you cannot create a vpc peering connection between vpcs in different regions.
vpc peering does not support transitive peering relationships. (a connected to b, b connected to c.. a cannot talk to b through c....need to have direct connection)

ECS (elastic container service)
what is docker?
application->dependecies>guest OS
docker works like c containers (standard shipping container)
Docker is a software platform that allows you to build,test, and deploy apps quickly
docker is highly reliable:quickly deploy and scale apps into any environment and

know your code will run
docker is infinitely scalable: running docker on AWS is a great way to run distributed apps at any scale
docker packages software into standardized units called containers
containers allow you to easily package an apps code, configs, and dependencies into easy to user building blocks that delvier environmental consistency, operational efficiency, dev productivity, and version control.

traditional vms contain apps, dependency, and guest os
container only has apps, and dependencies
docker achieves higher density and portability
escape dependency hell
docker components-docker image, docker container,layers/union file system, dockerFile, docker Daemon/engine, docker client, docker registries/docker hub
EC2 Container Service (ECS) is a highly scalable, fast, container management service that makes it easy to run, stop, and manage Docker containners on a cluster of EC2 instance.
ECS lets you launch and stop container-based applications with simple API calls, allows you to get the state of your cluster from a centralized service, and gives you access to many familiar EC2 features.

ECS is a regional service that you can use in one or more AZs
ECS eliminates the need for you to operate your own cluster management and configuration management systems, or to worry about scaling your management infrastructure
ECS can also be used to create a consisten deployment and build experience, manage and scale batch and ETL workloads, and build sophisticated application architectures on a microservices model

Containers are a method of operating system virtualization that allow you to run an application and its dependencies in resource-isolated processes.
containers have everything the software needs to run- including libraries, system tools, code, and runtime
containers are created from a read-only template called an image.

an image is a read only template with instructions for create a docker container
an image is created from a dockerfile, a plain text file that specifies the components that are to be included in the container
images are stored ina registry, such as DockerHub or ECR
ECR is amazon EC2 container registry
ECR is a managed AWS docker registry service that is secure, scalable, and reliable.
ECR supports private docker repos with resource-based permissions using AWS IAM
Developers can use the docker CLI to push, pull, and manage images.

parameters may include which docker images to use with the containers in your task, how much cpu and memory to use with each container, whether containers are linked together in a task

ECS exam tips
ECS-amazons managed ec2 container service. allows you to manage docker containers on a cluster of ec2 instances.
containers are a method of os virtualization that allow you to run an application and its dependencies in resource-isolated processes
containers are created from a read-only template called an image
an image is a read-only template with instructions for creating a docker containter
images are stored in a registry, such as dockerhub or AWS ECR
Amazon EC2 container registry (amazon ECR) is a managed aws docker registry service
A task definition is required to run docker containers in ECS
task definitions are text files in JSON format that describe one or more containers that form your applicaion
think of a task definition as a cloud formation template but for docker.
configure things such as the amount of cpu, ram, etc
ECS allows you to run and maintain a specified number of instances of a task definition simultaneously in a ecs cluster
think of services like auto-scaling groups for ECS
An ECS cluster is a logical grouping of container instances that you can place tasks on
Clusters can contain multiple different container instance types
clusters are region-specific
container instances can only be part of one cluster at a time
you can create IAM policies for your clusters to allwo or restrict users' access to specific clusters
you can schedule ecs in two ways: service scheduler, customer scheduler
ecs agent to connect EC2 instances to yuor ECS cluster. linux only
IAM with ECS to restrict access
security groups operate at the instance level, not at the task or container level