AWS Solutions Architect Study

Study Notes

These 3 topics are relevant but already covered in DevOps and Security Specialty. Review them again if there's enough time.

~ **AWS WAF** — FAQs - AWS WAF

~ **IAM Cross Account Policies** — Cross-account policy evaluation logic - IAM

~ **AWS ECS** — Amazon Elastic Container Service FAQs

✔ **Elastic Map Reduce** — Amazon EMR FAQs

Guess: No idea, a distributed data parsing tool that does map-reduce operations on data in various storage systems e.g. S3.
After Review: Runs standard Map Reduce based data analytics platforms like Spark/Hive etc. and manages the installation, configuration and maintenance of the platform. With EMR studio you can also run Spark/Hive jobs without building a cluster the job gets submitted and run "serverlessly".

✔ **Redshift** — Amazon Redshift FAQs

Guess: Reporting data tool that does columnar data storage. We ditched it for Snowflake because SF was cheaper and faster.
After Review: Data warehouse that runs SQL quickly, costs quite a bit at 5.6k/month for the default configuration which is a 2 node cluster supporting 256TB of storage. Has a basic query editor in the AWS console. SOC and PCI compliant.

✔ **S3 Select** — Amazon S3 FAQs

Guess: Never heard of it, some kind of S3 query tool similar to Athena?
After Review: S3 Select is an S3 API feature that allows you to retrieve portions of objects that are structured (CSV/json) by SQL-like queries. The max record size for input is 1MB so I believe that means a CSV line of 1MB is too big or a .
Filtering and retrieving data using Amazon S3 Select.

✔ **Athena** [Amazon Athena FAQs]

Guess: Query S3 data with a SQL-like language.
After Review: Also queries other data sources like Cloudwatch, Redshift, DynamoDB etc. To query S3 you need to now use Glue to build the data source.

✔ **Kinesis** [Amazon Kinesis Data Streams FAQs]

Guess: Data stream processor, throw data in one end and do basic manipulations and pass it out the other. Processes data by sharded processors which can each handle up to 1MB/s of input data including shard IDs. Can be used for doing things like redacting on the fly.
After Review: Shard limit is also 1000 messages/sec. Read capacity for each shard is basically double the input capacity and the stream can have a retention period set, order of events is preserved and multiple consumers can process the same stream concurrently. The service only provides data streaming, not the processing that can be done by stream consumers. The max record size is 1MB for a data blob. Data streams are one part of it but it also has Kinesis Firehose which is a managed consumer that can optionally process the data with a Lambda transformer and send the data to another system. Kinesis Data Analytics can be used to run Apache Flink/Beam data stream processing applications

✔ **Glue** [AWS Glue FAQs]

Guess: A tool similar to Prefect that connects together various data processing systems.
After Review: Yeah, AWS ETL. Also seems to be required for S3 data sources for Athena now.

✔ **Well Architected Framework** [AWS Well-Architected]

Guess: The AWS pillars of cloud architecture, focuses on Efficient, Scalable, Resilient, Cost-effective implementations of cloud based solutions.
After Review: Whoops, also being secure is important. Operational Excellence, Security, Reliability, Performance Efficiency, Cost Optimization and Sustainability are the pillars.

✔ **Elastic Container Registry** [Amazon ECR FAQs]

Guess: Dockerhub by AWS, a basic container registry similar to Azure Container Registry and Google Container Registry. Everyone has an xCR. Integrates with the Docker CLI with “aws ecr-login”. I believe there’s an AWS credential provider plugin for Docker and Kubernetes.
After Review: To allow ECS to pull from a private registry expose a VPC interface endpoint to ECS.

✔ **Step Functions** [AWS Step Functions FAQs]

Guess: Serverless orchestration service that handles triggering lambdas, passing data from one to another. Enables you to stitch together services with conditional logic, basic data manipulation like taking a field from a JSON response to send to a downstream service, parallelising parts of workflows and handling error conditions.
After Review: Nothing to add.

✔ **Cloudfront** Cloudfront FAQs

Guess: An AWS CDN that can be thrown in front of your web applications/buckets to provide an HTTP caching layer which reduces backend infrastructure costs and bandwidth while providing cached data closer to users because there are a heap of distributed PoPs across the globe - I assume one per AWS region, potentially some extra dedicated PoPs like in NZ where we don't have a data center yet.
After Review: Also provides field-level encryption for form values for situations where you only want certain services at your origin to be able to decrypt certain values e.g. credit card numbers. You can also run workers on the edge that are written in Javascript with CloudFront functions and more general purpose edge processes with Lambda@Edge.

✔ **CloudHub** AWS VPN CloudHub

Guess: No idea.
After Review: A hub and spoke VPN model where AWS provides a VPN gateway that multiple data centers/local sites are connected two that provides a route between local sites. For if you're too lazy for full mesh.

✔ **Global Accelerator** AWS Global Accelerator FAQs

Guess: Similar to Cloudflare Argo where you can route traffic over dedicated backbones to reduce latency and have predictable network conditions between your VPCs.
After Review: Does a bit more and provides health checking to the endpoints you are using e.g. an ALB and also smart-routes from the closest edge PoP to the closest regional endpoint.

✔ **Direct Connect** AWS Direct Connect | FAQ

Guess: ~~A site-to-site VPN between AWS VPCs and your local network.~~
After Review: Not a VPN it's a literal dedicated connection between you and AWS organized through your ISP or your data center. I was thinking of AWS Managed VPN with Virtual Private Gateways in my guess.

✔ **FSx** Amazon FSx

Guess: Some kind of distributed file system service.
After Review: Filesystems as a service. Easily configure ONTAP, OpenZFS, Windows File Server and Lustre file systems. These can be used as network file systems on other instances via NFS/SMB.

✔ **Lustre** Amazon FSx for Lustre FAQs

Guess: A distributed file system service not owned by AWS that can be presented through FSX APIs to AWS services/systems.
After Review: Lustre is a portmanteau (Linux Cluster). It's a high throughput parallel distributed file system. When integrated with FSx you can use FSx to replicate the bucket so it's available as a high throughput filesystem for some operational load and then dump it back to S3 and remove the filesystem. Lustre will do 100+GB/s where the other options are slower at 3GB/s (12.5GB/s for ZFS).

✔ **Storage Gateways** AWS Storage Gateway FAQs

Guess: Connect S3/EFS backed storage to your local environment via common network sharing protocols.
After Review: As I guessed but it doesn't do EFS it can do Windows File Server from FSx though and it supports local caching because it's deployed as an on-premise agent.

✔ **Private Link** AWS PrivateLink FAQs

Guess: No idea.
After Review: It's the magic behind VPC interface endpoints e.g. exposing an AWS service like SQS via a private IP address inside the VPC so your resources talk to it over an internal connection, not via the internet. This also configures DNS internal overrides, generally the internal DNS name is the same as the public name so no application reconfiguration is required.

✔ **EFS** Amazon Elastic File System (EFS)

Guess: Elastic File System, I'm not sure how it's used.
After Review: A simple shared file system that scales to the amount of data you throw at it. It's primarily intended for general purpose workloads but it can burst and also scale up to 10GB/s and half a million IOPS if the storage size is huge. Presented via NFS.

✔ **AWS MGN** AWS Server Migration Service FAQs

Guess: The old AWS Server Migration Service that's soon to be decommissioned in favor of AMS. This is used to migrate servers to the cloud with reduced downtime, the majority of your VM is transferred while it's still live on your local hypervisor platform and then when it's time to cut over you have a short period of downtime where the VM is stopped on prem to allow the final copy and then the VM goes live in the cloud.
After Review: SMS is based on snapshot replication and builds AMIs of the machines to be migrated to the cloud.

✔ **AMS** AWS Application Migration Service FAQs

Guess: Apparently AMS replaces SMS but I don't know what the difference is.
After Review: AWS seems to have bought CloudEndure Migration and brought it in-house. It uses an agent to replicate servers with block level replication which reduces the downtime required for cutover to the cloud server.

✔ **Cloud Data Migration** Cloud Data Migration on AWS

Guess: Some kind of agent that you get to upload your local data to the cloud.
After Review: This actually covers a few tools. The Snow* family of physical data moving: SnowCone for a suitcase of data, Snowball for PB and Snowmobile for EB. S3 Transfer Acceleration for increasing performance of S3 by uploading directly to a Cloudfront PoP. DataSync for common data migrations between AWS services and also between on-prem by deploying an agent to a VM on VMWare, Hyper-V hypervisor etc. that has access to the data to be transferred.

✔ **Database Migration Service** AWS Database Migration Service FAQs

Guess: An agent that aids you in migrating local databases to the cloud via the databases native backup and restore tooling e.g. with SQL Server it would do backup and restore the full backups and potentially log ship before cutting over.
After Review: Cool it also does automatic schema migrations to RDS/Aurora if that's the target system and highlights when something will require manual intervention. Seems to require a connection to the server so it will require a route to the source endpoint via VPN if it's on-prem.

✔ **AWS Managed VPN** AWS Managed VPN

Guess: A site-to-site VPN between AWS VPCs and your local network.
After Review: Created by setting up a Customer Gateway which represents the VPN gateway at the local network and then creating a Virtual Private Gateway linked to the Customer Gateway to create an IPSec VPN.

✔ **AWS Batch** AWS Batch FAQs

Guess: A batch computing job orchestrator.
After Review: A simple system where you can run jobs on ECS/Fargate by defining a task definition as a bash-like command to be executed on a container of your choice.

✔ **AWS Budgets** AWS Budgets FAQs

Guess: A way of managing and forecasting spending in AWS. Can be used as a trigger for running cost saving measures via things like Lambda.
After Review: You can directly run some simple actions, attaching IAM roles to identities, attach SCPs or stop instances.

✔ **AWS QuickSight** AWS QuickSight FAQs

Guess: A visual data analytics tool that can query arbitrary data or data sources from things like RDS/Athena.
After Review: Very similar to Looker, integrates with SSO/IAM users.

✔ **S3 Storage Classes** S3 Storage Classes

Guess: Standard, Infrequent-Access and Glacier are in order of decreasing expense. The tradeoff is the speed of data access. Infrequent access doesn't penalize you for infrequent accessing but will charge for large amounts of access.
After Review: Glacier comes in different flavors, instant, flexible and deep archive. Intelligent tiering allows for S3 to move storage objects to various object classes for efficient storage costs.

✔ **SSM Parameter Store** SSM Parameter Store User Guide

Guess: A way of storing configuration items that provides access controls, auditability, secure storage and encrypted credentials.
After Review: Limited to 4KB parameter size, 8KB if you want to pay for "advanced". Supports strings, stringlists and securestring which is just a string secured with a KMS key, the 4K limit is probably influenced by the max decryption payload size for KMS.

Testrun of the AWS Sample Questions

Prior to starting any study I attempt the sample questions to give myself a quick reality-check to find out the areas I'll need to put the most effort into.

Results (70%, 7/10)

1) Correct.
2) Correct.
3) Failed half of the multichoice question because I tried to set CORS headers on the bucket. The website origin is the API so it needs to allow the bucket as an allowed origin with CORS headers.
4) Correct.
5) Correct.
6) Correct.
7) Failed the question because I tried to do cross account policies with the role in the wrong account.
8) Failed the question because I had not heard of AWS SMS.
9) Failed the question because I didn't know the Kinesis data stream shard limits.
10) Correct.

Gotchas

- I failed two of these because I wasn't careful enough reading the question and answers.
- SMS is a server migration service but it is being replaced with AMS so the questions in the real exam might reflect that.
  https://aws.amazon.com/server-migration-service/
  https://aws.amazon.com/application-migration-service/
- A Kinesis shard has a limit of 1MB/s.
  https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-limits.html

Testrun of the AWS 20 Practice Questions

As a final test of what I've learnt from the study I redo the Sample Questions above to see if I still remember what's going on and if the questions seem easier to comprehend then I do the official 20 practice questions which shows me weak areas I need to do some extra study in.

Results (80%, 16/20) - Rerun (100%, 20/20)

The questions are in random order so I can't number these.

1. Failed half of the multichoice question because I didn't know RDS doesn't support auto scaling for read replicas.
2. Failed the entire question because I thought you could get more IOPS by increasing the DB storage size but beyond a certain point you stop getting more performance.
3. Failed the entire multichoice question because I thought S3TA had a limit on file size and I thought using regional buckets with replication would be faster because you could upload locally and replication would be async but the other half of this relied on route53 latency based routing to s3 regional endpoints which doesn't work.
4. Failed the entire question because I thought the Direct Connect connection was the single point of failure and by migrating the service to the cloud it would be resolved. In this case this was not allowed because on-premise services apparently need to use the database service but this was not specified in the question.

Gotchas

- DB instance volumes don't get any more IOPS once they're at ~5.34TB, to go faster you need provisioned IOPS.
  https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/CHAP_Storage.html
- Aurora is what you want if you need auto-scaling read replicas for reporting workloads.
  https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/Aurora.Integrating.AutoScaling.html
- S3TA is the go for speeding up S3 transfers, using regional buckets doesn't work well.
  https://stackoverflow.com/questions/21876483/how-to-configure-route53-to-nearest-s3-endpoint
- Direct Connect VIFs are like VIPs on F5 gear.
- I need to look more into the feature set of AWS Budgets.
- RDS maximum storage is 16GB for SQL Server and 64GB for other offerings.
- DynamoDB maximum item size is 400KB.
- I need to look more into SSM.
- I need to look more into Quicksight.
  https://aws.amazon.com/quicksight/resources/
- I need to look more into VPC flow log implementation and analysis.
- I need to look more into S3 storage class pricing.
  https://aws.amazon.com/s3/storage-classes/
- RDS does support transactions.
- I need to learn more about AWS Batch.
  https://aws.amazon.com/batch/faqs/

Real Exam Results

aws training and certification

AWS Certified Solutions Architect - Professional (Retiring Nov. 14, 2022)

Notice of Exam Results

Candidate: Shaun Lawrie	Exam Date: Aug 22, 2022
Candidate Score: 816	Pass/Fail: PASS

Congratulations! You have successfully completed the AWS Certified Solutions Architect - Professional (Retiring Nov. 14, 2022) and you are now AWS Certified.

Breakdown of Exam Results

The information in the table below details the composition of the AWS Certified Solutions Architect - Professional (Retiring Nov. 14, 2022) and your performance in each of the exam sections. The table includes the classifications of your performance at each **section level**.

This information is designed to provide general feedback concerning your examination performance. The examination is scored using a compensatory scoring model, which means you do not need to "pass" the individual sections. Please keep in mind that each section has a specific weighting on the examination, so some sections have more questions than others. This information is general in nature, highlighting your strengths and weaknesses.

Meets Competencies: Performance at this level demonstrates knowledge, skills, and abilities expected of a passing candidate.

Needs Improvement: Performance at this level does not demonstrate knowledge, skills, and abilities expected of a passing candidate.

	Score Performance		
Section	**% of Scored Items**	**Needs Improvement**	**Meets Competencies**
Domain 1: Design for Organizational Complexity			
Domain 2: Design for New Solutions			
Domain 3: Migration Planning			
Domain 4: Cost Control			
Domain 5: Continuous Improvement for Existing Solutions			

Disclaimer: AWS Certification exams are designed to make pass/fail decisions based on the total exam score. Section level results are designed to provide direction on areas where a candidate may be weak. Candidates should exercise caution when interpreting the above section level score information as it is less reliable than the total exam score and not intended to guide future test performance.

Gotchas from the Real Exam

- I didn't know enough about implementing Direct Connect in practice, just the concepts and that made any Direct Connect questions time consuming.
- I should have dusted off long division before going into this because I had a brain fart and forgot how to do it because I'm so used to doing bandwidth calculations on a calculator.
- I wish I had known more about the different gateway types https://www.megaport.com/blog/aws-vgw-vs-dgw-vs-tgw/
- Having practical experience with the Database Migration Tool and Schema Migration Tool would have helped a lot.
- Pretty much all real world solutions use multiple accounts so I would have struggled if I wasn't familiar with the permissions and AWS organizations.