AWS Cheatsheet - Part 3

Author: Bhaskar S
Version: 1.8
Date: 03/02/2024
Blog: PolarSPARC

Application Integration

- MQ supports ActiveMQ and RabbitMQ and enables lift-and-shift of on-prem apps that use messaging broker
- MQ support HA in a region with an active in one AZ and a standby in another AZ backed by EFS

- SQS is a pull-based, managed queue service with one consumer per queue
- SQS consumer can get up to 10 messages per pull and they need to explicitly delete the message to remove from queue
- SQS retains messages for up to 4 days (max of 14 days) with max message size of 256 KB
- SQS access can be controlled by IAM policies or SQS access policies (allows cross-account access)
- SQS encryption at-rest enabled by default using managed key
- SQS default message visibility timeout is 30 secs (not visible to other consumers once pulled)
- SQS long polling reduces the number of API calls and increases app efficiency - can be between 1 to 20 secs
- SQS Standard Queue has at least once delivery semantics with no message ordering guarantees
- SQS FIFO Queue has exactly once delivery semantics (using message deduplication id) with guaranteed message ordering (using message group id)
- SQS fifo queue throughput 300 messages per sec without batching AND 3000 messages per sec WITH batching of 10 messages

- SNS is a push-based, managed topic service with many subscribers listening on a topic (with **NO** persistence)
- SNS subscriber endpoints can be SQS, Lambda, Kinesis Data Firehose (**NOT** Kinesis Data Stream), http/s, email, sms
- SNS can receive messages from CloudWatch alarm, S3 bucket, Lambda, DynamoDB, etc
- SNS subcriber can filter messages using **Filter Policy**
- SNS encryption at-rest enabled by default using managed key
- SNS access can be controlled by IAM policies or SNS access policies (allows cross-account access)
- SNS can integrate with SQS for fan-out pattern with no data loss
- SNS FIFO Topic guarantees messages ordering with has exactly once delivery semantics, throughput 300 messages per sec

- Step Functions is a serverless workflow orchestration service that can integrate with on-prem and AWS services (Lambda, EC2, ECS, SQS, etc)
- Step Functions supports two types of workflows - **Standard** (for long running, at-most once execution) and **Express** (for at-least once, short running executions, up to 5 mins)

- Standard workflows can execute up to 2000 per sec, cost per state transition
- Express workflows can execute up to 100000 per sec, cost per state transition and duration of execution, can send execution history to CloudWatch

- EventBridge is a serverless event bus that can route events from many sources to many targets (like Lambda, SNS, Kinesis Data Streams, etc)
- EventBridge has a default event bus which gets all the state changes from various AWS resources
- EventBridge provides simple and consistent ways to ingest, filter, transform, and deliver events
- EventBridge supports backup of events for future replay, has support for Event Schema Registry

- API Gateway is a service for creating/publishing/maintaining/monitoring secure rest, http/s, websocket API at scale
- API Gateway can expose backend http/s endpoints, lambda, and other AWS services
- API Gateway can integrate with Lambda via Lambda Proxy to create serverless API services
- API Gateway can handle authn/authz and support throttling via rate limiting
- API Gateway supports 10000 requests per sec AND max concurrent requests of 5000 per sec
- API Gateway can validate/transform request and responses, cache API responses (with TTL)
- API Gateway supports canary release deployment strategy to deploy new version to a canary stage and direct a portion of the user traffic to the canary stage with gradual traffic increase and if successful promote the canary stage to production

- Kinesis Data Streams is a service that uses a collection of shards to process/aggregate a large continuous stream of data records (with partition key) in real-time
- Kinesis Data Streams data records with the same partition key go into the **SAME** shard, ensuring ordering within a shard
- Kinesis Data Streams supports up to 1000 data records per sec per shard
- Kinesis Data Streams has a retention period ranging from 1 day (default) to 365 days
- Kinesis Data Streams can have **MANY** consumers processing concurrently AND has two modes of consumption - **Standard** (pull) AND **Enhanced Fan-out** (push)
- Kinesis Data Streams supports the ability to replay (or reprocess) previously processed data records (in the same order)

- Kinesis Data Firehose is a managed/auto-scaled near-real-time data streaming service (buffers data in-memory up to a size/for a duration) before sending to destination (http, S3, Redshift, OpenSearch, 3rd-party like Splunk, DataDog, Dynatrace, New Relic, etc)
- Kinesis Data Firehose can leverage Lambda function for transformations before delivering to destination, supports many data formats/compression
- Kinesis Data Firehose data records are **NOT** persisted or stored, pay only for data that passes through

Databases

- RDS supports mysql/mariadb/postgres/oracle/mssql/aurora and runs on EC2 (choose instance type), manages common database admin tasks (auto provisioning/ patching/upgrades)
- RDS is useful for applications with **UNPREDICTABLE** workloads
- RDS supports automatic backups (daily full snapshot) with point-in-time restore (upto 5 mins ago)
- RDS allows read replicas (up to 15 with async replication) with support for multi AZ

- RDS multi AZ (one DNS name) automatically creates a primary DB instance and synchronously replicates data to a standby DB instance in a different AZ in the same region for auto fail-over
- RDS automatic backup has retention 1 to 35 days. For greater retention period, manual backup preferably from a standby
- RDS supports auto scale of storage up to the configured max threshold
- RDS backup restore always creates a new DB
- RDS unencrypted DB to encrypted means snapshot and then copy snapshot with encryption
- RDS Custom only for Oracle, Microsoft SQL Server, to enable native features, AND allows for the operating system (OS) and database customizations

- RDS creates SSL certs and installs them when the DB instance is created AND one can download and use those provided certs for making secure SSL/TLS connection to the DB
- IAM DB Authentication can authenticate to an RDS db instance and works with mysql/postgres. With this authentication method, one does **NOT** need to use a password when connecting to the db instance, instead use the auth token

- Aurora supports mysql and postgresql with distributed, fault-tolerant, <u>self-healing</u> storage that can grow to 128 TB
- Aurora <u>natively</u> supports HA with 6 copies of data across 3 AZs with one being the primary for writes
- Aurora replica is both a standby and a target for read traffic in a multi-AZ configuration
- Aurora supports self-healing through peer-to-peer replication
- Aurora can have up to 15 read replicas for better query performance (with auto scaling of replicas)
- Aurora Global DB can have one DB in one region and up to 5 replicas in a secondary region with cross-region replication taking less than 1 sec
- Aurora supports automated backups with retention between 1 to 35 days and **CANNOT** be disabled
- Aurora support DB cloning which is faster than snapshot and restore
- Aurora serverless **ONLY** accessible through VPC

- RDS Proxy is serverless, auto scaled, HA (across multiple AZs) DB connection pool service with support for mysql/mariadb/postgresql/mssql/aurora
- RDS Proxy makes apps more resilient to database failures by automatically connecting to standby, can enforce IAM authentication
- RDS Proxy is NEVER publicly accessible - **ONLY** in a VPC

- ElasticCache is a managed, distributed in-memory key/value store that runs on EC2 (choose instance type) and supports Redis AND Memcached
- ElasticCache Redis is multi AZ with auto fail-over and support for backup/restore, IAM auth and encryption
- ElasticCache Memcached only has multi node shards and multi-thread and **NONE** of the redis capabilities

- DynamoDB serverless nosql DB with predictable performance, seemless scalability, replicated to multi AZ
- DynamoDB supports <u>key-value</u> AS WELL AS <u>document</u> types with transaction support AND strong/eventual consistency models

- DynamoDB made up of tables (with items), max size of an item 400 KB, item can have a TTL after which auto deleted
- DynamoDB supports two capacity modes - provisioned (plan read/write capacity for predictable workloads) AND on-demand (auto scales for read/write for unpredictable workloads)
- DynamoDB auto scaling dynamically adjusts the provisioned throughput capacity in response to actual traffic patterns
- DynamoDB supports continuous (with per sec granularity to past 35 days) and on-demand (for long term retention) backup, on-demand can integrate with AWS Backup service
- DynamoDB Streams captures time-ordered events of item-level CRUD changes into a table (with 24 hr retention) that can be used for cross-region replication using Lambda
- DynamoDB Accelerator is an in-memory cache with TTL of 5 mins, supports read-through and a write-through modes, no changes to app
- DynamoDB Global Tables is multi-region AND multi-master solution with two-way async replication using DynamoDB Streams

- Redshift is a managed, petabyte scale data warehouse, with postgresql based columnar storage, running on EC2 (choose instance type)
- Redshift maintains 3 copies of data with support for multi-AZs
- Redshift uses a leader node (for query planning and result aggregation) and a number of compute nodes (perform actual query and send results back to the leader node)
- Redshift supports automatic backup as snapshots (incremental) every 8 hours OR after every 5 GB, with configured retention
- Redshift can get data ingested via Kinesis Data Firehose or S3 Copy

Analytics

- Athena is a federated SQL query engine that helps one perform BI on data (unstructured, semi-structured, structured) stored in S3, RDS, DynamoDB
- Athena supports compressed/uncompressed formats for csv, json, columnar (parquet, orc)
- Athena can be used for BI/reporting/analysis of logs from services (VPC Flow, ELB, CloudTrail)

- OpenSearch is a petabyte scale search (with support for partial field search) and analytics (with visualization) engine for unstructured data such as log/clickstream analytics, real-time monitoring
- OpenSearch can get data from Kinesis Data Firehose, CloudWatch Logs, etc

- Elastic Map Reduce (EMR) is a cluster platform on EC2 for hadoop/spark to process, transform, analyze vast amounts of data for BI

- QuickSight is a serverless BI visualization service that can connect to various data sources on-prem as well as AWS (Athena, RDS, Aurora, Redshift, S3, OpenSearch) for creating interactive dashboards
- QuickSight support in-memory computation engine which can responds with blazing speeds if data is imported into the service

- Glue is a serverless ETL service (using spark) that can discover/prepare/move data from multiple sources and load into Redshift
- Glue runs the ETL jobs on fully managed, scaled-out Apache Spark environment, can be used to convert data to columnar format (parquet, orc)

- Lake Formation a governed central data lake (structured/unstructured) in S3 with its own permissions model (augments IAM) for fine grained access which can be used for analytics and ML usecases
- Lake Formation helps one discover the data sources, catalog, cleanse, transform, and ingest data into the data lake

- Managed Streaming for Apache Kafka (MSK) is a managed Apache Kafka service that enables one to build and run data stream processing apps

- Managed Service for Apache Flink also referred to as Kinesis Data Analytics for Flink
- Managed Service for Apache Flink can source data from either Kinesis Data Streams or Kafka (MSK) for time-series analytics, real-time dashboards

- Batch is a regional service for running batch workloads that auto scales based on demand

Machine Learning

- Rekognition helps analyze images/videos to detect/label objects, people, scenes, activities, text AND detect inappropriate content

- Transcribe is a multi-lang Automatic Speech Recognition service that converts audio to text with automatic PII redaction

- Polly is a multi-lang text to speech conversion service with the ability to customize word pronounciation using Pronounciation Lexicons AND additional customizations like word emphasis using Speech Synthesis Markup Language

- Translate helps convert from one language to another that can be used for content localization

- Comprehend uses NLP to extract key phrases/sentiments from documents and automatically organize them by categories/topics

- SageMaker is a ML service for data scientists/developers to build/train/deploy ML models for production use

- Forecast uses statistical/ML algorithms to forecast domain specific metrics using historical data

- Textract helps extract text information from documents including the ones that are handwritten, from forms, and govt id documents

Security

- Directory Service provides a way to use AD with EC2, RDS for MSSQL, FSx for Windows, and IAM Identity Center
- Directory Service Managed AD runs Domain Controllers in different AZs in a region, can have trust relationship/SSO with on-prem AD via VPN/Direct Connect (for > 5000 users)

- Directory Service AD Connector is a proxy for existing on-prem AD via VPN/Direct Connect, directory requests redirected to on-prem AD
- Directory Service Simple AD is AD compatible service by AWS using Samba 4 with Kerberos SSO

- Cognito is a user directory for web and mobile app authn/authz, is OIDC provider supporting oauth, oidc, saml
- Cognito User Pool allows sign-up/sign-in for apps, with user identity store as well as federated identity with external providers (google, facebook), supports MFA
- Cognito Identity Pool allows temporary access to specific AWS resources via IAM policies based on user identity

- KMS uses HSM to create/manage/protect/validate cryptographic keys used for encryption & decryption and is region scoped
- KMS key can be symmetric (cannot get access) or asymmetric (cannot get access to private key)
- KMS key usage audited via CloudTrail, integrated with IAM for authorization
- KMS - AWS owned/managed keys no cost, customer managed **COSTS** $ 1.00 per month
- KMS automatic key rotation for AWS owned/managed keys every 1 year
- KMS custom Key Policy for cross account access
- KMS keys for specific resources enforced by **kms:ViaService** condition in KMS Key Policy
- KMS multi-region key implies primary key in a region replicated to another (has same id, key material) AND useful for global dynamodb or global aurora

- SSM Parameter Store provides secure, hierarchical storage of config/secrets info such as db urls, passwords, license codes, etc
- SSM Parameter Store can store details as either plain text or as encrypted data

- Secrets Manager helps one access/manage/auto-rotate db credentials AND is used by many AWS services
- Secrets Manager also helps access/manage oauth tokens and api keys, for auto rotation of these one must use a custom Lambda
- Secrets Manager can configure an automatic rotation schedule for the secrets, which will use a AWS Lambda function to update the secret
- Secrets Manager is integrated with RDS, Aurora, Redshift, DocumentDB for secrets management

- ACM allows one to create/store/manage private/public SSL/TLS certificates that protect secure websites and applications
- ACM supports both public and private SSL/TLS certificates
- ACM created public certs have AWS as the Certificate Authority and CAN auto renew
- ACM certs imported from external **CANNOT** auto renew - ONLY notify
- ACM integrates with ELB/CloudFront/API Gateway/Elastic Beanstalk, cannot be used with EC2

- WAF is a global service that helps protect web against common exploits (sql injection, xss) at layer 7
- WAF protects ALB, CloudFront, API Gateway, Graph API, Cognito User Pool
- WAF rules (region scoped) are defined using Web ACL AND specifies the inspection criteria and the actions to take if matched

- WAF web ACL can be based on geo-location (country), ip addr set, http headers/body/uri

- Shield is for protection against DDoS at layer 3/4 and layer 7
- Shield Standard enabled by default for protection from syn/upd flooding, reflection attack, and other layer 3/4 attacks
- Shield Advanced costs $3000, provides access to AWS DDoS response team, and protects ELB, EC2, CloudFront, Global Accelerator, Route 53
- Shield Advanced automatically creates, evaluates, and deploys WAF rules to mitigate Layer 7 attacks

- Network Firewall is a stateful, managed, network firewall and intrusion detection and prevention service for VPC

- Firewall Manager helps one administer/manage security policies (waf, shield advanced, security groups, network firewall, route 53) CENTRALLY across all the accounts/resources of an organization, has region scope, applies to current and future resources, cost $ 100.00 per month

- GuardDuty is an ML based continuous security monitor to detect unexpected, unauthorized malicious activity in an account
- GuardDuty can detect compromised EC2, ECS, EKS
- GuardDuty can be integrated with EventBridge to notify on findings
- GuardDuty event sources can be CloudTrail, VPC flow logs, DNS logs, RDS login activity, S3 logs, EBS volumes, EKS logs, Lambda network activity, etc
- GuardDuty service if disabled will delete all the data (incl findings, config, etc)

- Detective allows one to assess, investigate, and pinpoint the source of suspected security vulnerabilities or suspicious activity in the AWS environment

- Inspector is a scheduled vulnerability assessment service that detects deviations from best practices AND helps improve security/compliance of resources
- Inspector reports all findings via the Security Hub
- Inspector automatically discovers workloads, such as EC2 instances, ECS, EKS, and Lambda AND scans them for software vulnerabilities and unintended network exposure

- Macie is a security/privacy discovery/protection service for S3 to detect PII, PHI, sensitive data (api keys, secrets)

- Control Tower offers an easy way to set up, govern AWS multi-account environment, and orchestrate the capabilities of several AWS services, including AWS Organizations, Service Catalog, and IAM Identity Center, to build a landing zone in less than an hour

Logging/Monitoring

- CloudWatch Metrics gathers various metrics (**EXCEPT** RAM or Disk) for most services AS WELL AS on-prem systems
- CloudWatch Metrics runs every 5 mins for EC2, enable detailed monitoring for metrics every 1 min (incurs cost)

- CloudWatch Metrics can stream to Kinesis Data Firehose

- Unified CloudWatch Agent is the one that can collect both RAM and Disk info in additional to the other metrics AS WELL AS collect logs
- Unified CloudWatch Agent can gather/send custom metrics from apps using statsd or collectd protocols

- CloudWatch Alarms can initiate actions when a metric threshold is breached
- CloudWatch Alarms are of two types - Metric Alarm (single metric), Composite Alarm (multiple metrics with AND/OR expression)
- CloudWatch Alarms has three states - OK (within threshold), INSUFFICIENT_DATA (not enough data), ALARM (threshold breached)
- CloudWatch Alarms has three actions targets for EC2 - stop/terminate/reboot, auto scaling action, send notification to SNS

- CloudWatch Logs collects and stores system and applications logs in a single place with log expiration (never or 1 day to 10 yrs), encrypted by default
- CloudWatch Logs can collect logs from on-prem as well when an agent is installed
- CloudWatch Logs can send to S3 (via export), Kinesis Data Streams, Kinesis Data Firehose, Lambda, OpenSearch
- CloudWatch Logs supports query/streaming of log events across AWS services for AWS accounts

- CloudWatch Events (aka EventBridge) can stream events that describe changes to resources and can be used to trigger actions

- CloudWatch Insights has three types - Container Insights (collect/aggregate/summarize metrics and logs from containers), Lambda Insights (monitoring/ troubleshooting serverless apps using lambda incl cold starts/shutdowns), Contributor Insights (analyze logs to identify the top-N contributors)

- CloudTrail applies to all Regions by default, enabled by default, any API action by any entity recorded as events
- CloudTrail allows one to perform auditing/compliance on an account by looking at all the events related to who did what, when, and on what resources
- CloudTrail stores history of all events for 90 days. For greater retention period, move the events to CloudWatch Logs or S3
- CloudTrail data and insight events **NOT** enabled by default
- CloudTrail Insights Events allows one to detect unusual activity in an account once **ENABLED**, incurs cost

Deployment/Migration

- CloudFormation uses JSON/YAML template file to describe resources and allows one to create, update, delete the entire stack

- OpsWorks is a managed configuration management service for lift-and-shift of on-prem chef/puppet based config management

- Config extracts point-in-time view of all the config attributes/parameters of all the resources in an account for compliance reasons
- Config maintains the historical records of all the config changes for all resources
- Config Rules define the desired config settings which can be used to evaluate against the current settings AND notify on deviation

- Application Discovery Service helps collect the usage and static config of on-prem server/databases for cloud migration AND stores the details in Migration Hub
- Application Discovery Service - agentless for VMWare only WHILE agent-based for VMWare/HyperV/Physical
- Application Discovery Service agent-based **ALSO** collects info about network connections and processes

- Application Migration Service allows a firm to automatically lift-and-shift on-prem servers/VMs to the cloud with minimum down-time (in mins)
- Application Migration Service also supports apps such as SAP, Oracle, and SQL Server migrations

- Database Migration Service allows one to migrate on-prem DB to RDS on EC2 OR to any DB on AWS OR from AWS to on-prem
- Database Migration Service provides support for CDC
- Database Migration Service will need to use Schema Conversion Tool if source and target are different DBs
- Database Migration Service allows one to continuously replicate data with high availability to consolidate on-prem DBs into a petabyte-scale data warehouse by streaming data to Redshift and S3

- DataSync helps one securely transfer on-prem data (NFS/SMB/HDFS/S3) using an agent to AWS (EFS/S3/FSx) OR for region to region transfer
- DataSync preserves file permissions AND metadata
- DataSync is a scheduled service **NOT** a continuous/real-time service
- DataSync **CAN** move archival/historical data from on-prem to S3 Glacier or S3 Glacier Deep Archive as the destination

- Backup is a managed service for automated backups that are stored in S3 for both on-prem and aws services, eliminating the need for scripts and manual processes
- Backup plans define the frequency, window, retention period, transition to cold storage etc

- Snowcone comes with DataSync Agent and is for **TERABYTE** scale

- Snowball is for **PETABYTE** scale that comes with block and object storage

- Snowmobile is for **EXABYTE** scale

- Outposts is a fully managed service that extends AWS infrastructure, services, APIs, and tools to on-prem
- Outposts allows one to build and run applications on-prem using the same programming interfaces for EC2, EBS, ECS, EKS, S3, RDS, ALB, Route 53, etc