Other Services in AWS

# S3:

## S3 Select

is a new Amazon S3 capability designed to pull out only the data you need from an object, which can dramatically improve the performance and reduce the cost of applications that need to access data in S3.

Please note that with Amazon S3 Select, you can scan a subset of an object by specifying a range of bytes to query using the **ScanRange parameter**. This capability lets you parallelize scanning the whole object by splitting the work into separate Amazon S3 Select requests for a series of non-overlapping scan ranges. Use the Amazon S3 Select ScanRange parameter and Start at (Byte) and End at (Byte).

## Amazon Redshift Spectrum Using this you can efficiently query and retrieve structured and semi-structured data from files in Amazon S3 without having to load the data into Amazon Redshift tables

# Kinesis

## Amazon Kinesis Data Firehose

is the easiest way to reliably load streaming data into data lakes, data stores, and analytics tools. It is a fully managed service that automatically scales to match the throughput of your data and requires no ongoing administration. It can also batch, compress, transform, and encrypt the data before loading it, minimizing the amount of storage used at the destination and increasing security.

## Kinesis Agent

is a stand-alone Java software application that offers an easy way to collect and send data to Kinesis Data Streams or Kinesis Firehose.

# VPC

## AWS Resource Access Manager (RAM)

is a service that enables you to easily and securely share AWS resources with any AWS account within your AWS Organization. You can share AWS Transit Gateways, Subnets, AWS License Manager configurations, and Amazon Route 53 Resolver rules resources with RAM. RAM eliminates the need to create duplicate resources in multiple accounts, reducing the operational overhead of managing those resources in every single account you own. You can create resources centrally in a multi-account environment, and use RAM to share those resources across accounts in three simple steps: create a Resource Share, specify resources, and specify accounts. RAM is available to you at no additional charge.

The correct solution is to share the subnet(s) within a VPC using RAM. This will allow all EC2 instances to be deployed in the same VPC (although from different accounts) and easily communicate with one another.

How Resource Access Manager Works: Text

Description automatically generated with medium confidence via - <https://aws.amazon.com/ram/>

# AWS Glue

AWS Glue is a fully managed extract, transform, and load (ETL) service that makes it easy for customers to prepare and loa d their data for analytics. AWS Glue job is meant to be used for batch ETL data processing. AWS Glue does not offer the same storage and processing speed as FSx for Lustre. So it is not the right fit for the high-performance workflow scenario.

# Storage

## Amazon EMR

Amazon EMR is the industry-leading cloud big data platform for processing vast amounts of data using open source tools such as Apache Spark, Apache Hive, Apache HBase, Apache Flink, Apache Hudi, and Presto. Amazon EMR uses Hadoop, an open-source framework, to distribute your data and processing across a resizable cluster of Amazon EC2 instances. EMR does not offer the same storage and processing speed as FSx for Lustre. So it is not the right fit for the given high-performance workflow scenario.

## Amazon FSx for Windows File Server

Amazon FSx for Windows File Server provides fully managed, highly reliable file storage that is accessible over the industry-standard Service Message Block (SMB) protocol. It is built on Windows Server, delivering a wide range of administrative features such as user quotas, end-user file restore, and Microsoft Active Directory (AD) integration. FSx for Windows does not allow you to present S3 objects as files and does not allow you to write changed data back to S3.

# Exam Tip

Multi-AZ vs Read replica

Please review this comparison vis-a-vis Multi-AZ vs Read Replica for RDS: Graphical user interface, text, application

Description automatically generated via - <https://aws.amazon.com/rds/features/multi-az/>

# EC2

## EC2 Placement Group

**Partition placement group**

**—>** groups of instances in one partition do not share the underlying hardware with groups of instances in different partitions

—> used by large distributed and replicated workloads, such as Hadoop, Cassandra, and Kafka

**Cluster placement group**

—> packs instances close together inside an Availability Zone

—> the low-latency network performance necessary for tightly-coupled node-to-node communication that is typical of HPC applications

—> not suited for distributed and replicated workloads such as Hadoop

**Spread placement group**

—> strictly places a small group of instances across distinct underlying hardware to reduce correlated failures

—> not suited for distributed and replicated workloads such as Hadoop.

**Hadoop =>** distributed and replicated workloads

# AWS Organizations

This helps you centrally govern your environment as you grow and scale your workloads on AWS. Using AWS Organizations, you can automate account creation, create groups of accounts to reflect your business needs, and apply policies for these groups for governance. You can also simplify billing by setting up a single payment method for all of your AWS accounts. Through integrations with other AWS services, you can use Organizations to define central configurations and resource sharing across accounts in your organization.

To migrate accounts from one organization to another, you must have root or IAM access to both the member and master accounts. Here are the steps to follow: 1. Remove the member account from the old organization 2. Send an invite to the new organization 3. Accept the invite to the new organization from the member account

# AWS OpsWorks

* This is a configuration management service that provides managed instances of Chef and Puppet. Chef and Puppet are automation platforms that allow you to use code to automate the configurations of your servers.
* OpsWorks lets you use Chef and Puppet to automate how servers are configured, deployed and managed across your Amazon EC2 instances or on-premises compute environments.

# Load Balancer

# WAF

AWS WAF is a web application firewall that helps protect your web applications or APIs against common web exploits that may affect availability, compromise security, or consume excessive resources. AWS WAF gives you control over how traffic reaches your applications by enabling you to create security rules that block common attack patterns, such as SQL injection or cross-site scripting, and rules that filter out specific traffic patterns you define.

The correct answer is to use WAF (which has integration on top of your ALB) and define a rate-based rule.

# AWS Shield

This is a managed Distributed Denial of Service (DDoS) protection service that safeguards applications running on AWS. AWS Shield provides always-on detection and automatic inline mitigations that minimize application downtime and latency, so there is no need to engage AWS Support to benefit from DDoS protection. There are two tiers of AWS Shield - Standard and Advanced.

AWS Shield Advanced provides enhanced resource-specific detection and employs advanced mitigation and routing

# AWS DataSync

This is an online data transfer service that simplifies, automates, and accelerates copying large amounts of data to and from AWS storage services over the internet or AWS Direct Connect.

AWS DataSync fully automates and accelerates moving large active datasets to AWS, up to 10 times faster than command-line tools. It is natively integrated with Amazon S3, Amazon EFS, Amazon FSx for Windows File Server, Amazon CloudWatch, and AWS CloudTrail, which provides seamless and secure access to your storage services, as well as detailed monitoring of the transfer.

# The AWS Transfer Family

This provides fully managed support for file transfers directly into and out of Amazon S3

# AWS Global Accelerator

This is a networking service that helps you improve the availability and performance of the applications that you offer to your global users. AWS Global Accelerator is easy to set up, configure, and manage. It provides static IP addresses that provide a fixed entry point to your applications and eliminate the complexity of managing specific IP addresses for different AWS Regions and Availability Zones. AWS Global Accelerator always routes user traffic to the optimal endpoint based on performance, reacting instantly to changes in application health, your user’s location, and policies that you configure. Global Accelerator is a good fit for non-HTTP use cases, such as gaming (UDP), IoT (MQTT), or Voice over IP.

# AWS Database Migration Service

This helps you migrate databases to AWS quickly and securely. The source database remains fully operational during the migration, minimizing downtime to applications that rely on the database. With AWS Database Migration Service, you can continuously replicate your data with high availability and consolidate databases into a petabyte-scale data warehouse by streaming data to Amazon Redshift and Amazon S3

# Database

## **DocumentDB**

Amazon DocumentDB is a fast, scalable, highly available, and fully managed document database service that supports MongoDB workloads. As a document database, Amazon DocumentDB makes it easy to store, query, and index JSON data. Although DocumentDB is fully managed, it does not have an in-memory caching layer.

## **Amazon Neptune**

Amazon Neptune is a fast, reliable, fully managed graph database service that makes it easy to build and run applications that work with highly connected datasets. The core of Amazon Neptune is a purpose-built, high-performance graph database engine optimized for storing billions of relationships and querying the graph with milliseconds latency. Neptune powers graph use cases such as recommendation engines, fraud detection, knowledge graphs, drug discovery, and network security.

Amazon Neptune is highly available, with read replicas, point-in-time recovery, continuous backup to Amazon S3, and replication across Availability Zones. Neptune is secure with support for HTTPS encrypted client connections and encryption at rest. Neptune is fully managed, so you no longer need to worry about database management tasks such as hardware provisioning, software patching, setup, configuration, or backups.

Amazon Neptune can quickly and easily process large sets of user-profiles and interactions to build social networking applications. Neptune enables highly interactive graph queries with high throughput to bring social features into your applications. For example, if you are building a social feed into your application, you can use Neptune to provide results that prioritize showing your users the latest updates from their family, from friends whose updates they ‘Like,’ and from friends who live close to them.

## **DynamoDB Streams**

A DynamoDB stream is an ordered flow of information about changes to items in a DynamoDB table. When you enable a stream on a table, DynamoDB captures information about every modification to data items in the table. Whenever an application creates, updates, or deletes items in the table, DynamoDB Streams writes a stream record with the primary key attributes of the items that were modified. A stream record contains information about a data modification to a single item in a DynamoDB table.

DynamoDB Streams will contain a stream of all the changes that happen to a DynamoDB table. It can be chained with a Lambda function that will be triggered to react to these changes, one of which is the developer's milestone. Therefore, this is the correct option.

# AWS OpsHub

Software that you install on your laptop or computer to manage your snow family device